Klaus Urich

Comparative Animal Biochemistry

Translated from the German by Patrick J. King

With 248 Illustrations by Charlotte Urich and 87 Tables

Springer-Verlag Berlin Heidelberg GmbH

Author
Professor Dr. Klaus Urich
Universität Mainz
Institut für Zoologie
Niklas-Vogt-Straße 25
D-55131 Mainz

Translator
Dr. Patrick J. King
Stutzackerweg 8
CH-3110 Münsingen

Title of the German Edition
Klaus Urich, Vergleichende Biochemie der Tiere
ISBN 978-3-642-08181-1

Illustration: Dr. Charlotte Urich, Niklas-Vogt-Straße 25, D-55131 Mainz
Cover Illustration: The illustration on the front cover shows the symbiosis between the sea anemone Radianthus kueckenthali and the fish Amphiprion perideraion: The signal substance amphikuemine that is produced by the sea anemone attracts the fish even at a concentration of 10^{-10} mol/L.

Library of Congress Cataloging-in-Publication Data
Urich, Klaus, 1926– [Vergleichende Biochemie der Tiere. English] Comparative animal biochemistry / Klaus Urich ; translated from the German by Patrick J. King. p. cm.
Includes bibliographical references and index.
ISBN 978-3-642-08181-1 ISBN 978-3-662-06303-3 (eBook)
DOI 10.1007/978-3-662-06303-3
1. Biochemistry. 2. Physiology, Comparative. I. Title.
QP514.2.U7513 1994 591.19'2–dc20 94-7808

Originally published by Springer-Verlag Berlin Heidelberg New York in 1994
Softcover reprint of the hardcover 1st edition 1994

Cover: E. Kirchner, Heidelberg
Typesetting: Mitterweger Werksatz GmbH, Plankstadt bei Heidelberg
SPIN 10041846 31/3130 – 5 4 3 2 1 0 – Printed on acid-free paper

Dedicated to the memory
of Prof. Dr. Bernt Linzen
(8. 6. 1931 – 5. 8. 1988)

Preface

The plan for this book goes back almost 20 years. Already, at that time, it was possible to recognize an extraordinary variation in metabolites and processes superimposed upon the basic biochemical system of animals. Each species, each individual, in fact each type of cell of the multicellular organism possesses its own biochemical character, and this molecular variety, its biological significance, and its evolutionary development throw up many interesting questions. The comparative approach that has been so productive at the higher levels of complexity of morphology and physiology can also be used to great effect at the molecular level.

There are countless papers, reviews and other articles on the subject of comparative biochemistry and evolution, but no actual textbook. This is remarkable in so far as each area of science is only really defined by its comprehensive representation in the literature. In the present book, I have given my view of the starting points and the content of comparative biochemistry. I would be glad if further, related efforts by other authors would follow this attempt at a definition of the subject. In the widest sense, a comparative biochemistry should consider all organisms, but my knowledge of the biochemistry of the bacteria, algae, lower fungi and higher plants is unfortunately not sufficient for this purpose. In any case, most of the available comparative biochemistry data refer to animals.

This book is directed at zoologists as well as biochemists. Biochemists may discover that molecular variation, though perhaps a complication for general biochemical research, is in itself a fascinating scientific problem. Zoologists will find that the molecular building blocks of animals display just as much colourful variation as do microscopic or macroscopic structures. They will see that molecular variants can be understood partly as adaptations to particular environmental conditions and partly as the result of diversification along separate evolutionary lines. Finally, it will be seen that biochemical characters can contribute greatly to understanding the origins of organisms.

To provide the biochemist with a ready overview of the structural diversity of animals, the book includes a simplified version of animal systematics; for further information on the classification, structure and life of particular animal species, the reader should consult the relevant textbooks. It is assumed that the zoologist reader has a basic knowledge of biochemistry; important general biochemical facts are in any case given for many of the subjects covered.

I had already completed several chapters of this book by the beginning of the 1970s. Because of far-reaching changes in my employment conditions and duties as a high-school teacher, my transfer from Berlin to Mainz, and the task of setting up a working group there, work on this book was interrupted and could only be resumed in the 1980s. In the meantime, the data to be considered had increased enormously; molecular biology, in particular, had produced a gigantic body of relevant data. Today, a single person can no longer keep track of all aspects of comparative biochemistry.

And so arises the well-known dilemma: the portrayal of a large area of science by one or a few authors inevitably leaves gaps and deficiencies; one book with contributions from many authors generally leaves much to be desired in uniformity and consistency. When I decided, according to the original plan, to attempt as a single author to bring together the whole of comparative biochemistry, the unavoidable deficiencies were very clear to me. I would be very grateful for any critical comments on gaps and mistakes in this presentation, but I would ask biochemically oriented readers to judge the whole work and not only those sections most familiar to them. I have researched the literature up to the middle of 1988, but the reference lists given for each chapter do not include the whole of the consulted literature; this would require too much space. Rather, I have referred to the latest

works and this should allow the reader access to earlier and original papers.

There are particular problems associated with biochemical terminology and the names of organisms, as well as with the use of abbreviations and symbols. Here, I have attempted on the one hand to be consistent, but on the other hand to diverge as little as possible from the terminology of the original authors. Some remarks on biochemical nomenclature and the use of abbreviations are to be found at the end of the book; the letter symbols for amino acids, other common abbreviations, and the genetic code are reproduced on the bookmark.

Many areas of comparative biochemistry are still in the fact-collecting stage, and the present book is a compromise between the desires to achieve a certain completeness and to avoid boring lists. The reader will notice that special cases, outside of the norm, and entertaining exceptions give me particular pleasure. The colourful diversity of molecular structures and processes in animals has always fascinated me. If I succeed in conveying just a part of this fascination to the reader, then this book will have fulfilled its purpose.

Finally, I would like to thank all those who directly or indirectly have contributed to the production of this book. My family, my students and my examination candidates have shown much patience with me during the years of work on the manuscript. My wife has, in addition, contributed directly to the book by preparing the figures. Particular thanks go to my secretary, Inge Pullig, and to Dr. Jens Beyer and Prof. Erhard Thomas, who took over many of my duties to allow me time for this literary effort. I would also like to thank the many colleagues who have given me reprints or who have critically read parts of the manuscript.

Prof. Bernt Linzen played a decisive role in the creation of the book through his constructive criticism, intelligent advice and encouraging comments but did not live to see its completion. I dedicate this work to the memory of this eminent researcher who contributed so much to comparative biochemistry.

Mainz, Autumn 1989 KLAUS URICH

Preface to the English Edition

The basic idea of the book and the disposition of chapters has not been changed from that of the German edition of 1990. The rapid progress in general and comparative biochemistry, however, had made it necessary to actualize the text and incorporate the many new facts and ideas that have emerged since summer 1988 when the German edition was completed.

In preparing the revised version, I have researched completely the original literature up to the end of 1991. As in the German edition, the references given in each chapter always refer to the latest works, and this could allow the reader access to earlier papers. Of a total of more than 4600 references, about 2100 were published later than July 1988 and so were not contained in the German version. As it was not possible to augment the total number of references, because of space limitations, a corresponding number of older references have been omitted.

I wish to thank Springer Verlag for generous cooperation during the preparation of the English edition. I feel particulary grateful to Dr. Patrick J. King who has succeded admirably in translating the author's sometimes rather individual German into good English and so has helped to propagate further ideas of comparative biochemistry.

Mainz, Spring 1993 KLAUS URICH

Contents

1 The Subject Matter and Methods of Comparative Biochemistry

1.1 Historical Development of the Comparative Approach in Biology

The diversity of living organisms, the abundance of forms, colours and phenotypes, is one of the most impressive aspects of our world. Since ancient times, science has concerned itself with the ordering and classification of this diversity. At first, only those structures and processes which were recognizable without any special apparatus could be compared; the invention of the light microscope and the electron microscope introduced new dimensions and new opportunities for comparison. Finally, progress to the molecular level became possible with the development of methods of biochemical analysis. Different comparative biological disciplines arose, each with particular goals and methods. The oldest, **taxonomy**, is concerned with the classification of species; this was originated by Aristotle, and is represented most importantly by the work of Carolus Linnaeus (1707–1778).

The binomial nomenclature suggested by Linnaeus, in which each organism is denoted by a genus and species name in Latin, is today considered obligatory. The denotation is completed by attachment of the name of the first user; the "L." that one finds at the end of many plant and animal names stems from the descriptions in Linnaeus' *Systema Naturae* (1735). Until the eighteenth century, the philosophical basis of taxonomy was the idealistic concept of a "Scala Naturae", a ladder reaching from the "lower organisms" up to man as the "crown of creation". In the nineteenth century, the idea of evolution, the assumption of genealogical relationships between organisms, found increasing acceptance.

The second comparative biological discipline, comparative **anatomy**, analyses the outer and inner structures of animals. This originated from medicine where comparisons were made between animal organs and those of man; the term "comparative anatomy" was coined by the doctor Nehemiah Grew in 1675. Since Cuvier (1769–1832), comparative anatomy has become a discipline of zoology which was remarkably consolidated in the nineteenth century and, with the distinction between homology and analogy, provided the most important tool for phylogenetic analysis. There were similar developments in botany but these will not be covered in this present zoological text.

The **comparative physiology** of animals also had its origins in medicine. Because, of course, few experiments are possible on people, human physiology relied, and still relies today, on various animal model systems. The extrapolation of results from animals to the human situation is ensured by the growing appreciation of the universal validity of the "ground rules" of general physiology. With the extent to which invertebrates were included in physiological experiments, there grew the realization that the conformity of life processes to several basic principles is overlayered by a confusing variety of detail. In this respect, a significant contribution was made with the improved possibilities for experiments on marine invertebrates at the Naples Zoology Station, founded in 1870 by Anton Dohrn.

The application of chemical concepts and methods for investigations of the life processes led to "physiological chemistry" or "chemical physiology", which soon became an independent discipline. The term "biochemistry" was first introduced by Neuberg in 1903. Also, in **biochemistry** the comparison of different species served at first to establish the consistency of basic principles; examples include Otto von Fuerth (1903) in *Comparative Chemical Physiology of Lower Animals* [9], and E. Baldwin (1937) in *Introduction to Comparative Biochemistry* [1]. Noch (1958) quote Dixon and Webb in their monograph on enzymes [6]:

"It is a remarkable fact that in general the catalytic properties, specificity, activity, affinities, etc., of a given enzyme vary little with the source. Although there may be slight physical differences in a given enzyme when it is produced by different cells they are usually unimportant, and the enzyme remains essentially the same enzyme."

In 1895 Emil Fischer had already pointed out differences between enzymes of different origins. At first, however, this, together with the species and tissue specificity of other biochemical characters, was considered to be a methodical impediment to biochemical studies rather than a unique scientific problem. Credit for emphasizing the importance of phylogenetic considerations in biochemistry must go to Marcel Florkin with his book *L'évolution biochimique* (1944). In his later works this commendable scientist was also concerned with establishing "comparative biochemistry", as an illustration of the biochemical diversity of organisms, alongside "general biochemistry", as lessons from the basic principles of biochemical organization; examples are his book *Unity and Diversity in Biochemistry* (1960) and as editor of the handbooks *Comparative Biochemistry* (7 volumes 1960–1964, with H. S. Mason) and *Chemical Zoology* (11 volumes 1967–1979, with B. T. Scheer).

In the last two decades, the comparative and phylogenetic approach to biochemistry has become rather fashionable. Monographs dealing with particular animal groups almost invariably contain sections on comparative biochemistry, and new monographs and reviews on comparative biochemical subjects are constantly appearing. The 13 volumes of *Comprehensive Insect Physiology, Biochemistry and Pharmacology*, published simultaneously in 1985, a remarkable achievement by Kerkut and Gilbert, contain many chapters on biochemical themes [20]. Papers with comparative or phylogenetic aims are not only to be found in specialist journals, e.g. *Comparative Biochemistry and Physiology: Part B (Biochemistry)* (December 1992 to Vol. 103), *Journal of Molecular Evolution* (up to Vol. 35), *Insect Biochemistry* (up to Vol. 22), and *Molecular Biology and Evolution* (up to Vol. 9); today, almost all biochemical journals contain work of this sort, as the references in this present book testify.

1.2 Uniformity and Diversity in Biochemistry

All living organisms show extensive correspondence between the structure of their component molecules and active chemical processes. A part of this chemical similarity may stem from the fact that only so is life at all possible. Lehninger refers vividly to the "molecular logic of the organism". Even between such different organisms as bacteria and mammals there are many similarities at the molecular level which cannot be satisfactorily explained either by functional necessity or by chance. This is particularly clear in the comparison of nucleic acid and protein sequences. This homology in the macromolecules carrying information is the strongest evidence for the common origin of all living organisms. It will be shown that the enormous number of different proteins can be arranged into just a few hundred groups of proteins with significant homology (protein superfamilies) which apparently came into being in a common ancestor during the early development of life on earth.

In addition to the correspondence arising from the basic necessities of life and ancestral relationships, there is in fact much variety among all characters of living organisms. On the one hand this is historically determined: only that which originates de novo always according to the same rules can be the same. On the other hand the variety of organisms is related to their complexity: the more complex living or non-living systems are, the less likely they are to resemble one another. In all aspects the variety of living things defies the imagination. The number of animal species on the earth can no longer be even reasonably estimated. Over a million animal species have been described scientifically so far, three-quarters of them are insects (see Appendix). Until recently, the real number of species was estimated by most zoologists to be 3–5 million, but following investigations of the tropical fauna [24] the figure is now believed to be ten times this. Fur-

thermore, the members of a species or of a reproductive population are never completely identical. With the help of the newly developed methods of isoenzyme separation by gel electrophoresis, it was shown at the end of the 1960s that the genetic polymorphism in natural populations was far larger than previously assumed, so large that the individuals in a bisexual, reproductive population could never be genetically identical. Even genetically identical individuals arising by asexual reproduction, such as twins or members of a clone, vary phenotypically due to environmental factors [23].

The characteristic variety of living organisms is apparent at all levels of complexity, from individuals through organs, tissues, cells and cell organelles down to individual macromolecules. For example, the human body is made up of 200 different cell types. Also, single cells of the same type in an organ or tissue show large differences [12]. The heterogeneity of the mitochondria in a cell or a cell type can be so large that they may be classified into several populations with different characters [10]. Basically, the variety of species, individuals, cells and cell components is related to molecular differences: to differences in the structure of the molecular building blocks and/or to differences in their arrangement. This molecular variation is the central theme of the present book.

1.3 The Subjects of Comparative Biochemistry

Biochemistry textbooks usually do not present the composition and metabolism of particular animals or cells; more often they describe a basic scheme of biochemical organization, a biochemical "archetype" as Pette called it [27]. Because the origins of biochemistry lie in human medicine, this basic scheme is mostly only valid for man and mammals, even when research results from yeast and bacteria are included.

It should never be assumed that the statements in biochemical textbooks are completely valid for lower vertebrates or for invertebrates. It is more often the case that the composition and metabolism of different animal groups or species must be specifically examined and presented; thus there exist, besides human and mammalian biochemistry, further presentations of "special biochemistry", e.g. the biochemistry of insects, crustaceans and molluscs. The "comparative biochemistry of animals" uses the same data but through the comparative approach comes to new conclusions. It can contribute to general biochemistry both by revealing similarities, and by pointing to experimental models that are suitable for investigating the general laws. The thesis known as the "August Krogh principle", named after the great Danish physiologist, proposes that for every problem there is an ideal animal model; this has found many applications in biochemistry [22]. The main task of comparative biochemistry is to describe the molecular variety of organisms and to explain both their biological importance and their development during evolution.

The present book restricts its view of comparative biochemistry to animals. The **classification of living organisms in the animal kingdom** usually presents no problem, but in specific cases may be difficult. Thus, the classification of flagellate species to the plant "phytoflagellates" or to the animal "zooflagellates" on the basis of the presence or absence of photosynthetic capacity has long been considered questionable. The classification by several authors of the "slime moulds" *Physarum* and *Dictyostelium* in the animal kingdom is also undoubtedly controversial. *Paramecium*, belonging to the Ciliophora, is today a favourite object of first-year undergraduate, zoology practical classes. However, because of the fundamental characteristics of their nucleic acid and protein sequences, the Ciliophora assume such a special position amongst the eukaryotes that they cannot logically be classified as either animals or plants. There are compelling reasons to assume that their divergence from the eukaryotic evolutionary line occurred before the separation of the lower fungi, higher plants and animals (see pp. 13 and 170). Only a very small number of the known animal species has ever been used for biochemical investigations; biological, methodological and economic factors play a role in the **choice of investigated species** [19]. Species-rich groups that have been studied physiologically have been preferred; in addition to the mammals, examples include the amphibians, fish, insects, crustaceans, molluscs and echinoderms. Because biochemical investigations require large amounts of genetically uniform material, which in most cases cannot be obtained from the field, species which can be reared in the laboratory have been used. Particular attention has been paid to species of practical importance, such as domestic animals, farm animals, pests and pathogens. Comparative biochemistry concerns itself with differences in composition and metabolism not only between members

of different species but also within a species, e.g. between male and female, between consecutive generations, between the different castes of the Hymenoptera, between the members of different races or populations, and between the individuals of a population.

If one defines the task of comparative biochemistry in general as the scientific analysis of molecular variety, this also covers **variety within individuals**, i.e. differences between organs, tissues and cells, between organelles, and between molecules of the same function or origin. Not only are the differences "in space" between various parts of an organism the subject of comparative biochemistry, but also the differences with time. The molecular changes in the course of embryo development are a part of "molecular embryology" [2]. The molecular viewpoint has also entered into the study of post-embryonic development and senescence [25]. Rhythmic processes, e.g. diurnal rhythms and seasonal periodicity, also belong to these time differences [7]. Only the qualitative aspects of such time differences and not their quantitative aspects, will be covered in this book.

1.3.1 Comparison of Low Molecular Weight Substances

Comparative investigations of small biological molecules in different species, tissues and developmental stages were particularly prominent during the initial phase of comparative biochemistry, before the appearance of methods for studying the nucleic acids and proteins. The expected quantitative differences between low molecular weight substances that are the substrates or intermediates of cell metabolism give important information about the physiological state of cells but lead to no phylogenetic conclusions. The large number of experiments carried out before the 1960s, for example to interpret the spectrum of free amino acids, were in this sense unsuccessful. However, qualitative differences in the low molecular weight products of secondary metabolism, e.g. pigments, defence molecules or pheromones, *were* found to be useful as biochemical markers. The spectrum of small molecules not produced de novo by the animal in question is dependent upon the composition of the food, e.g. the presence of carotenoids or, in many marine invertebrates, the sterols. Otherwise, the spectrum of low molecular weight compounds is both qualitatively and quantitatively determined by the enzymes active in the cell. Thus, indications of the presence or absence of specific enzymes or biosynthetic pathways may also present direct phylogenetic information.

1.3.2 Comparison of Information-Carrying Macromolecules

Evolution consists primarily of changes in the genetic information in succeeding generations, and its course can, therefore, best be read in the genes themselves or their products.

One can statistically determine whether compared DNA or protein sequences are derived from a common ancestral sequence. Sequences originating from a common ancestral gene are termed homologous; it will be necessary to discuss whether, and in what way, this concept in **molecular evolution** departs from the classical concept of homology (p. 117). The analysis of homologous DNA and protein sequences has produced much new knowledge on the mechanisms of evolution; as an example one can point to the frequency and outstanding importance of the transposition of mobile genetic elements. It will be shown that this new knowledge, rather than placing any doubt on the Darwinian synthetic theory of evolution, actually enriches and broadens the concept. In the near future, the genetically determined quantitative differences in gene expression, i.e. the control of mRNA or protein concentration, will also be available for phylogenetic assessment, as soon as more comparative data become available on the as yet only primitively understood control mechanisms of gene expression.

The molecular data make available completely new methods for the analysis of phylogenetic relationships. Proteins with clear homology have been organized by Dayhoff into "protein superfamilies" [4]. So far more than 200 such families have been identified, many of them are found in all organisms from prokaryotes to man; it is likely that 500–1000 such protein super-families exist. However, they number far less, for example, than enzymes with different reaction or substrate specificities, 2477 of which were listed by the International Union of Biochemistry (IUB) Commission in 1984 [17]. Proteins with very different functions can be found in the same super family, i.e. they are homologous. On the other hand, proteins of very different structure can have practically identical enzymic properties and are then denoted as "isoenzymes". Isoenzymes encoded by the alleles of a gene are called "alleloenzy-

mes". The analysis of alleloenzymes has become one of the most important tools in population genetics, with the help of which it is possible, for example, to decide whether two groups of individuals are exchanging genes, i.e. they constitute a population, or are sexually isolated and form two species.

By the use of molecular characters, it has been possible to recognize very similar animals as members of different species (sibling species), and to demonstrate the hybrid nature of apparently bona fide species. Unfortunately, it has proven more difficult than expected to use molecular data to solve taxonomic problems above the species level; in particular, the old problem of the relationships between the animal phyla has not been solved, despite the initial optimism. It is still problematic to derive a quantitative measure of evolutionary distance from nucleic acid- or protein-sequence differences. Even if this should be achieved, the application of a distance matrix for constructing a family tree of sequences is still fraught with difficulties. Such a tree may describe the evolution of molecules but does not necessarily reflect relationships between the species from which the molecules came. This is because gene duplication and subsequent divergent development (diversification) can result in different genes existing side by side in the same individual. In spite of such difficulties, there are still convincing arguments in favour of molecular characters as superior to all others for phylogenetic reconstruction.

1.4 Chance and Necessity in Molecular Evolution

1.4.1 Non-Adaptive ("Neutral") Differences

Research in comparative biochemistry attempts to describe structural or functional differences in biochemical organization: the presence or absence of particular compounds, differences in structural characteristics of low or high molecular weight substances, differences in enzyme activity or in the concentration of metabolites, or differences in the direction or rate of metabolic processes. Normally, structural and functional differences are considered to be the result of adaptive events, but one can actually also use the reverse argument: that adaptation exploits already available differences. Here it is usually quietly assumed that structural difference always means functional difference. Such assumptions arise from the idealistic view that, as a result of selection, all the characters of an organism are optimal for a particular environment. This may generally be true for complex anatomical, physiological and behavioral characters, even when it is rather awkward to imagine, for example, that the different arrangements of bristles on the bodies of closely related insects have adaptive significance.

For molecular evolution, the assumption that each structural difference represents a selective functional difference can be taken as categorically disproven; in DNA and subsequent amino acid substitutions in the resulting polypeptides there are quite clearly genetic differences that have no selective significance. Such genetically determined differences can be retained for long periods because they do not disturb, and can be spread by chance through a population. This is the central premise of the **neutral theory**, which will be dealt with extensively in Chapter 4. Of course, the concept "neutral" only applies under particular selection conditions; changes in the environment can confer on characters a positive or a negative value. The existence of a "reserve" of neutral variation in a population is a direct prerequisite for molecular evolution. Today, there can no longer be any doubt of the existence of neutral variation; the only discussion can be about the extent of the genetically determined differences which lack significant adaptive importance. However, as with all negative statements, the opinion that a particular molecular difference has no selection value is basically not provable. The existence of neutral mutations is important not only in the comparison of different individuals of one evolutionary line, but also in the comparison of distantly related species. When two evolutionary lines separate, a range of neutral variation can accumulate in both such that very large differences without selective significance can arise between species. Thus, for example, proteins with different amino acid sequences can be functionally equivalent. Epigenetically based differences between products of the same gene, e.g. the results of variable mRNA processing or post-translational changes in proteins, do not necessarily have adaptive significance, but are more likely to be the result of chance differences in enzyme activity. Such indications of the role of chance in molecular evolution should not deny the significance of molecular adaptation; this will in fact be one of the main themes in the following chapters. It should only be emphasized that it is not obligatory to enquire each time about the

adaptive sense of differences in molecular structure or activity found in comparative investigations.

1.4.2 Molecular Adaptation

Adaptive molecular evolution consists, on the one hand, of the adaptation of macromolecules, such as RNAs, enzymes and proteins, by various external and internal conditions and, on the other hand, of evolutionary changes in the unfortunately poorly understood mechanisms regulating gene expression, upon which the adaptive evolution of the complex characters of form, metabolism and behaviour is based. The evolutionary process involves thousands of generations; other events of molecular adaptation occur within the lifetime of an individual, requiring anything from a few milliseconds to several weeks. These processes also finally affect macromolecules: even when biochemical adaptation ends in changes in the concentration of small molecules or factors, these are always the result of changes in enzyme activity. Conversely, perturbations in the concentration of substrates or allosteric effectors have effects upon enzymes. In individual, physiological adaptation events, the properties or amounts of enzymes or proteins are modified. Alterations in the properties of proteins can arise through allosteric effects or chemical modifications, such as substitution or limited proteolysis, and changes in protein quantity arise mainly through the regulation of gene expression; here the distinction must be drawn between alterations in protein concentration and the appearance of new proteins or isoforms following induction.

Biochemical adaptation can serve two purposes: it can compensate for change in external or internal conditions (compensatory adaptation), or allow colonization of a new ecological niche (exploitative adaptation). Hochachka and Somero indicate the relevant problems and give examples in their book *Biochemical Adaptation* [15]; because of its close association with ecology, biochemical adaptation has been the subject of several recent reviews [8, 16, 29, 30]. Processes of biochemical adaptation occurring in single individuals have major consequences for the regulation of metabolism. In the field of comparative biochemistry, these events are themselves of minor interest compared with the underlying mechanisms, their variety, and changes in evolution. In the adaptive, molecular evolution of an enzyme, changes in regulatory characters play at least as great a role as changes in catalytic and physicochemical characters.

Many examples of regulatory and evolutionary **molecular adaptation events** will be described in particular chapters of this book, among them, for example, regulatory changes in connection with reproduction, exercise [28], hunger, and during inflammation and defence processes or other stress reactions, but most of all in connection with regulatory and evolutionary adaptation to various external conditions. Here is included, for example, the **osmolality of the extracellular fluids** to which the cell must adapt by isosmotic intracellular regulation. Amongst other changes occurring during this process is that of the intracellular concentration of specific osmotically active, organic substances, in particular the free amino acids and other nitrogen compounds, and thus the environment of intracellular proteins. The cartilaginous fish enrich their blood and other tissues with osmotically active substances, such as urea and trimethylaminoxide, to such an extent that specific protein adaptation is necessary. High extra and intracellular urea concentrations also occur during aestivation in the lungfish and certain frogs. The biochemical adaptations to transient or continuously low oxygen partial pressure are well known; the properties of the respiratory pigments may be changed or, in some animals, there are variedly effective mechanisms for anaerobic energy recovery; these mainly concern catabolism of carbohydrates and will be dealt with in Chapter 14. The use of different energy-yielding substrates in the muscles of various animals should also be considered as an adaptive mechanism to specific external and internal conditions (p. 581).

One external factor of critical biological importance is temperature [3, 5]. The homeothermal mammals and birds are able to maintain a constant body temperature by regulation of energy-producing metabolism. Many mammals possess a specialized tissue for heat production, the brown fat, in the mitochondria of which the Krebs cycle and ATP production are uncoupled by the action of a specific uncoupling protein (p. 702). In mammals and birds that can reduce their body temperature to the level of their surroundings during sleep (bats, hummingbirds, young swifts) or hibernation (bats, several rodents), the biochemical organization of the cells must be adapted to the combination of low temperature, reduced metabolism, acidification, and often increased urea concentration. The poikilothermal (cold-blooded) animals are at the mercy of the tem-

perature fluctuations in their surroundings and must, therefore, possess compensatory biochemical mechanisms. The stabilization of membrane viscosity through the "homeoviscous" regulation of the lipid composition should be mentioned here (p. 598), but most important of all is the temperature compensation of enzyme activity. As the substrate affinity of an enzyme declines with increasing temperature (i.e. the K_m increases) but at the same time k_{cat} and V_{max} increase, the quotient k_{cat}/K_m, which is an approximate measure of the in vivo catalytic activity, will be temperature independent and constant [3, 15, 31]. Many poikilothermal organisms possess isoenzymes with differing temperature optima, and the proportions of these can be variably adjusted. Specific adaptations to low temperature may be found in many animals, whilst heat-resistant animal species, comparable to the thermophilic bacteria, are unknown [11, 15]. Particularly interesting from a comparative biochemistry point of view are the various cryoprotective substances of animals: these include the accumulation of glycerol and sorbitol in many insects (p. 466) and, above all, the independent occurrence of cryoprotective proteins in arctic and antarctic fish and in several terrestrial arthropods (p. 207).

An especially enticing theme is the biochemical adaptations of deep-sea animals, which withstand hydrostatic pressures of up to 1160 bar [13, 18, 26]. Many pressure experiments on enzymes produce only indications of the stability of protein structures [21]. In 1973, however, a special expedition was mounted to research the biochemical-ecological problems of deep-sea creatures [14]. All effects of pressure are related to volume changes: if a chemical process is accompanied by an increase in volume, it will be negatively affected by increased pressure but promoted by a fall in pressure. Volume changes in proteins arise more from alterations in the state of hydration than from alterations in conformation. The association of protein subunits results in the release of water at the contact points and, thus, in an increase in volume; pressure, therefore, promotes the dissociation of oligomeric enzymes. The evolution of deep-sea animals, however, does not go in the direction of preference for monomeric enzymes, as was believed for a time, but rather towards a strengthening of nonpolar interactions between the subunits. Ligand binding and the catalytic activity of enzymes are also accompanied by volume changes, and are thus pressure dependent and the subject of adaptive evolutionary processes. Membrane structure is also pressure dependent but the corresponding adaptive changes in membrane lipids have, as yet, been little investigated.

References

1. Baldwin E.: An introduction to comparative biochemistry, 4. ed. Cambridge Univ. Press, Cambridge 1964
2. Brachet J., Alexandre H.: Introduction to molecular embryology, 2. ed. Springer, Berlin 1986
3. Crockett E.L. and Sidell, B.D.: Some pathways of energy metabolism are cold adapted in Antarctic fishes. Physiol. Zool. 63: 472–488 (1990)
4. Dayhoff M. O.: Atlas of protein sequences and structure, Vol. 5 and Suppl. 1–3. Nat. Biomed. Res. Foundation, Washington 1972–78
5. DiPrisco G.: Life under extreme conditions-Biochemical adaptation. Springer, Berlin 1991
6. Dixon M. and Webb E. C.: Enzymes. Longmans, London 1958
7. Engelmann W.: Biorhythmen. Quelle & Meyer, Heidelberg 1983
8. Evered D. and Collins G. M. (eds.): Origins and development of adaptations. Pitman, London 1984
9. von Führt O.: Vergleichende chemische Physiologie der niederen Tiere. G. Fischer, Jena 1903
10. Goto Y. et al.: Heterogeneity in the liver mitochondria of the tadpole, Rana catesbeiana. Comp. Biochem. Physiol. Pt. B 72: 637–640 (1982)
11. Grout B. W. W. and Morris G. J. (eds.): The effect of low temperature on biological systems. E. Arnold, London 1984
12. Haeussinger, D.: Nitrogen metabolism in liver: structural and functional organisation and physiological relevance. Biochem. J. 267: 281–290 (1990)
13. Henessy J. P. jr. and Siebenaller J. F.: Inactivation of NAD-dependent dehydrogenases from shallow- and deep-living fishes by hydrostatic pressure and proteolysis. Biochim. biophys. Acta 913: 285–291 (1987)
14. Hochachka P. W.: Biochemistry at depth. Comp. Biochem. Physiol. Pt B 52: 1–199 (1975)
15. Hochachka P.W. and Somero G.N.: Biochemical adaptation. Princeton Univ. Press, Princeton 1984
16. Hoffmann K. H.: Environmental physiology and biochemistry of insects. Springer, Berlin 1984
17. International Union of Biochemistry: Enzyme nomenclature 1984. Acad. Press, New York 1984
18. Jaenicke R.: Biochemical processes under high hydrostatic pressure. Naturwissenschaften 70: 332–341 (1983)
19. Kerkut G. A.: Which insects are most used in physiological and biochemical research? Comp. Biochem. Physiol. Pt. A 81: 705–706 (1985)
20. Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology. 13 vol. set. Pergamon Press, Oxford 1985
21. Kornblatt M. J. and Hoa G. H. B.: The pressure-induced inactivation of mammalian enolases is accompanied by dissociation of the dimeric enzyme. Arch. Biochem. Biophys. 252: 277–283 (1987)
22. Krebs H. A.: The August Krogh principle: „For many problems there is an animal on which it can be most conveniently studied". J. exp. Zool. 194: 221–226 (1975)

23. Lewontin R.: Human diversity. Scientific American Library, New York 1982
24. May R. M.: How many species are there? Nature 224: 514–515 (1986)
25. Mitchell D. H. and Johnson T. E. (eds.): Invertebrate models in aging research. CRC Press, Boca Raton USA (1984)
26. Pequeux A. J. R. and Gilles R. (eds.): High pressure effects on selected biological systems. Springer, Berlin 1985
27. Pette D.: Plan und Muster im zellulären Stoffwechsel. Naturwissenschaften 52: 597–616 (1965)
28. DiPrampero P. E. and Poortmans J. (eds.): Physiological chemistry of exercise and training. S. Karger, Basel 1981
29. Prosser C. L.: Adaptational biology – Molecules to organisms. Wiley & Sons, Chichester UK 1986
30. Schlee D.: Ökologische Biochemie. Springer, Berlin 1986
31. Somero G. N.: Environmental adaptation of proteins: strategies for the conservation of critical functional and structural traits. Comp. Biochem. Physiol. Pt. A 76: 621–633 (1983)

2 Nucleic Acids and Nuclear Proteins

2.1 The Structure of Chromosomal DNA

DNA deserves the most attention in any book dealing with molecular variety in animals. The complete genetic information of the organism is encoded in the order of the bases, and with it also is the whole spectrum of genetically determined variation within and between individuals. One can view DNA as a text in which each of the four letters at each position has a unique meaning. Some information and variety is lost en route from the DNA via RNA to the proteins and the complex morphological and physiological characters because only part of the DNA is transcribed into RNA, and not all RNA codes for proteins; furthermore, the genetic code is degenerate and the 64 possible triplet codons define only 20 amino acids. Sequence differences between proteins may also have no consequences for complex phenotypic characters. On the other hand, new (epigenetic) variability can arise during the information transfer process as a result of environmental factors.

The genetic information in the DNA that is passed from generation to generation is not as rigidly and inalterably inscribed as the text in a book. Spontaneous reactions, external factors and mistakes during replication, so long as they are not corrected by the cell's repair system, produce persistent changes in the genetic information. When this occurs in somatic cells, malfunction or neoplastic growth may result, and the survival chances of the individual are reduced. Mutations in germ cells, however, may spread through the progeny and instigate the evolutionary process. The latest results of molecular biology research have shown genetic alteration to be far more frequent than was previously assumed. The animal genome appears today as a dynamic system whose structure changes from one cell generation

to the next, not only through single base substitutions but also by rearrangement, transposition, amplification and the deletion of longer sequences. In accordance with the aims of this book, the main concern of the following chapter will be the structural variety in the nucleic acids that results from such processes. The reader is referred to the relevant literature for general information on molecular biology and related methods.

2.1.1 Conformation and Composition

The DNA that is organized together with histones and other nuclear proteins into the supramolecular structure of chromatin does not always have the well-known Watson and Crick conformation of a right-handed double helix (B form); the left-handed Z form is also quite often found in chromosomes. Longer sequences with purines on one strand and pyrimidines on the opposite strand (Y_n/R_n) can also form triple helices [256]. The double helix of the Z form has a diameter of 1,8 nm compared with 2.0 nm for the B-DNA, 12 instead of 10.5 bases per turn, and one turn per 4.46 instead of per 3.4 nm. Whereas the bases of B-DNA all have the same (antiparallel) polarity in terms of the sugar-phosphate chain, the orientation in Z-DNA is alternatively parallel and antiparallel. As the parallel orientation is more stable for purines than for pyrimidines, the Z conformation is favoured where purines and pyrimidines alternate. Thus, the widespread middle repetitive DNA components of mammals with the sequence $(CA/GT)_n$, i.e. with $(CA)_n$ on one strand and $(GT)_n$ on the other, should readily assume the Z conformation; this may perhaps be related to their function [186, 358]. Such sequences are never found in bacteria and are seldom found in protozoans; a comparative investigation into their origin would be interesting [308]. An equilibrium exists between B- and Z-DNA, with the B form being favoured thermodynamically. The Z conformation is stabilized by negative supercoiling and by methylation of CG to m^5CG. Changes from B to Z lead to far-reaching changes in the structural and functional character of the chromatin, e.g. in protein binding. Changes in DNA conformation are possibly involved in the regulation of transcription [358].

In contrast to B-DNA, Z-DNA is immune reactive. Z-DNA-specific fluorescent antibodies have shown a positive reaction with e.g. the polytene chromosomes of the dipterans *Drosophila* and *Chironomus*, the interphase nucleus of the ciliate *Stylonycha mytilus*, and the metaphase chromosomes of the primates. However, the solvents used (ethanol and acetic acid) promote the transformation of B to Z, and thus the in vivo existence of Z-DNA is questionable. In some cases, immunological evidence for Z-DNA has been obtained with more suitable methods, e.g. in transcriptionally active chromosomes of *Drosophila hydei* and in $(TG)_n$ sequences in the third intron of the α-lactalbumin gene of the rat [249, 297]. A biological role for Z-DNA is, furthermore, suggested by the existence of specific Z-DNA-binding proteins, e.g. in SV40 virus, *Escherichia coli*, wheat germ, *Drosophila* and rats. Three Z-DNA-specific proteins of 31–58 kDa have been isolated from bull testis, and a protein of 56 kDa has been isolated from HeLa cells [170, 241, 261, 367].

In addition to such far-reaching structural differences between different DNA regions, there is finer, more localized variation in helix parameters that probably represents the functional subdivision of the DNA into discrete units with characteristic properties. DNA molecules are not rigid and undergo thermal fluctuation in conformation that can be shown, for example, by the tritium exchange method. Such internal movements of the DNA and other macromolecules have been vividly described as "breathing reactions" [293, 444].

In order to describe the base composition of the DNA, it is only necessary, according to the Chargaff rule which states that A/T and G/C exist in equal proportions, to know the G+C content. Whilst in mammalian DNA this is always 40–50 %, values for the DNA of the Protozoa and those of the prokaryotes and lower eukaryotes vary widely (Table 2.1). The highly repetitive satellite DNAs separated as a minor band during density-gradient centrifugation have an extreme

Table 2.1. The G+C content of various DNAs [6, 99, 159, 168]

Malarial agent *Plasmodium falciparum*	18 %
Plasmodium berghei	20 %
Slime mould *Dictyostelium*	22 %
Ciliate *Tetrahymena thermophila* (macronucleus)	25 %
Yeast *Saccharomyces cerevisiae*	39 %
Rat (liver)	40 %
Wheat germ	43 %
Chicken (liver)	43 %
Mouse (spleen)	44 %
Escherichia coli	51 %
Neurospora crassa	54 %
Herpes simplex virus	72 %

composition. Further regional differences can be demonstrated in the base composition of the vertebrates. The genome is apparently split into internally homologous segments of up to 300 kb that differ in their G+C content (isochores); these correspond to the microscopically visible chromosome bands. The differing composition of the isochores has consequences not only for the sequences of the coding and non-coding DNA regions, but also for the transcripts and proteins. A high proportion of G+C-rich isochores is found only in birds and mammals; in lower vertebrates they are seldom found, if they occur at all. From human DNA, for example, various fractions with G+C values of between 36.7 and 49.2 % can be isolated [147, 311].

The transfer of methyl groups to single DNA bases results in the formation of 5-methylcytosine (m^5C) and other methylated bases. In animals and other eukaryotes, cytosine is methylated only when in the sequence CG. The resulting m^5CG sequence promotes the methylation of the complementary sequence GC, so that methylated residues are often found in pairs as m^5CG/Gm^5C [7]. In contrast to the situation in prokaryotes such as *E. coli*, some methylatable bases always remain unmethylated in the eukaryotes. The degree of methylation is species and tissue specific and dependent on development. In vertebrates about 3–6 % of cytosine residues are methylated, whereas in invertebrates the values are mostly lower; many insects and unicellular organisms contain no methylcytosines. In contrast, up to 30 % of the cytosine residues are methylated in plants. Relatively high degrees of methylation are to be found in the GC-rich 5′ non-translated (NT) regions of genes [7]. The DNA of embryos and tumour cells is always heavily methylated, and the transcriptionally inactive DNA of spermatozoa is more heavily methylated than that of active cells [5, 7, 79]. Using immunological methods, other methylated DNA residues, e.g. 6-methyladenine or 7-methylguanine, can be found in molar proportions of several percent of the corresponding bases; this occurs, e.g., in the citrus scale *Planococcus citri* or, in somewhat lower proportions, in *Drosophila*, human placenta, calf thymus and rat sperm [3]. 6-Methyladenine, but not 5-methylcytosine, is also present in the macronucleus of several ciliates [35, 159]. Methylation is apparently of importance for transcription of the DNA; for this reason the process is presently the subject of intensive study [5, 79]. However, the relationship between gene methylation and activity is by no means straightforward; active genes, as a whole or at particular points, are often hypomethylated, but they may also be completely methylated, or there may be no correlation between methylation and gene expression [113].

2.1.2 Base Sequence and Gene Structure

The information contained in the nucleic acids is encoded in the order of the subunits, the base or nucleotide sequence. The base sequence in the DNA of any given organism is the result of chance mutation and selection during evolution related to various functional requirements: for the encoding of polypeptides and RNAs with specific functions; for initiation, termination and regulation of transcription; for the coiling and packing of the DNA in the chromosomes; for the regulation of chromosome pairing during meiosis; and for recombination. In modern biology, one of the most fascinating possibilities for gaining new insight lies in our ability to read the DNA text and to trace therein the changes that occurred during the millions of years of evolution.

Automatic DNA sequencing methods have developed to such an extent that up to 50 000 bp can be sequenced per day [284], and the large amount of accumulated data can now be stored and evaluated only with the computer [459]. The DNA databanks established in Los Alamos and Heidelberg contained 6600 sequences in 1985, amounting to almost 6 million bp, and the problem of coping with the flood of data has since become acute [254]. Known sequences from the human genome, ribosomal RNAs, transfer RNAs and their genes, histones and histone genes, etc. are now regularly compiled in the journal *Nucleic Acids Research*. Approximately 60 % of the genome of the nematode *Caenorhabditis elegans*, which with its 80 million bp is one of the smallest known eumetazoan genomes, has already been cloned, as has the genome of baker's yeast, *Saccharomyces cerevisiae* [94]. Given the present sequencing capacity, the plan to sequence the complete 3000 million bp of the human genome no longer sounds utopian [468]. In 1989, about 6500 human DNA sequences were known, including approximately 1600 protein genes; in the order of 1000 new sequences are added each year [494]. DNA sequencing is much more efficient than protein sequencing and amino acid sequences are today almost entirely derived from the corresponding DNA sequences. The starting

sequence may be that of either a DNA complementing an RNA sequence (cDNA) or genomic DNA; the amounts of DNA required for sequencing are obtained through cloning reactions [30, 204]. The multiplication of complete genes from genomic DNA is now possible using the polymerase chain reaction (PCR), i.e. a DNA-polymerase-catalysed DNA synthesis using a large excess of short DNA primers that contain a partial sequence for the gene of interest [39, 267].

The gene concept has undergone many changes as our knowledge of mode of gene action has increased. After the classical definition of the gene as a unit of mutation and recombination, and then the "one gene, one polypeptide" principle, came the description in the 1960s of a gene as a stretch of DNA that is transcribed as a unit and that codes for a functional RNA (messenger RNA, ribosomal RNA, transfer RNA) [168]. Although, in the meantime, in animal genomes phenomena have been observed which do not fit this definition, e.g. overlapping and superimposed genes or the formation of multiple RNAs from a single gene, there is still much to be said for retaining the definition of a gene as a transcription unit (Fig. 2.1). The transcribed region of the DNA is always longer than the sequence coding for the finished product; in addition, in front of ("upstream", 5′ end) or behind ("downstream", 3′ end) the transcribed region lie DNA sequences of importance as signals for the initiation and termination of transcription. The question thus arises of how the gene as a functional unit is to be defined on the DNA. For this, one can exploit the changes in chromatin structure, shown by active genes, that make them more susceptible to DNase I digestion. Thus, the fibroin gene of the silkworm *Bombyx mori* is very sensitive to DNase when it is actively transcribed in the distal part of the silk gland of fifth-stage larvae, but is essentially insensitive to DNase I in its inactive state in the middle part of the silk gland of such larvae, or in later developmental stages [236]. In the chicken, the DNase-I-sensitive DNA region of the ovalbumin gene and the closely related genes "X" and "Y" has a length of 100 kb, that of the lysozyme gene is 24 kb, and for glyceraldehyde-3-phosphate dehydrogenase it is 12 kb; in all cases, the boundaries at the 3′ end are sharp and those at the 5′ end are more variable [213]. The signals for initiation and termination of transcription have different positions and structures according to the gene product and the relevant RNA polymerase, i.e. the Pol I transcribed ribosomal RNA genes, the Pol II transcribed protein-coding genes, and the Pol III transcribed genes for small RNAs. The transcription signals will be discussed in connection with the process of transcription. The start and finish of translation are also marked by signals on the DNA and mRNA, respectively. Translation begins at the first AUG of the mRNA and ends at the first UAA, UAG or UGA stop codon.

Normally the genes of the eukaryote genome are arranged in a linear fashion. In prokaryotes, on the other hand, overlapping genes are not uncommon; these lie either on different DNA strands and are read in opposite directions, or on the same strand and have interlocking reading frames. This situation, although it allows a higher density of information on the DNA and is advantageous for small genomes, hinders the evolution of DNA sequences. It is also particularly surprising to find such phenomena in animals that have a large surplus of DNA. The DNA flanking the protamine gene of the rainbow trout, *Salmo gairdneri*, includes an extra TATA box in front of the promoter and the region is transcribed into two different functional mRNAs. One of the mRNAs encodes protamine, and the other, which arises from a shifted reading frame and is found in large amounts only in the brain, produces the proline-rich Y protein (Fig. 2.2). The function of the Y protein is not known; it is significantly homologous to a proline-rich phosphoprotein from human saliva, the avian sarcoma virus (ASV) protein P19, and the product of the myc oncogene. The Y gene breaks the "first AUG rule" in so far as translation begins only at the third ATG after the TATA box. Interestingly, the protamine gene is flanked by long terminal repeats (LTRs) that are characteristic of ASV and other viruses. There is also a region in the ASV genome that codes for three different proteins using shifted reading frames, and it is easy to imagine that the anomalies in animals originated with the insertion of a viral genome [212]. At the

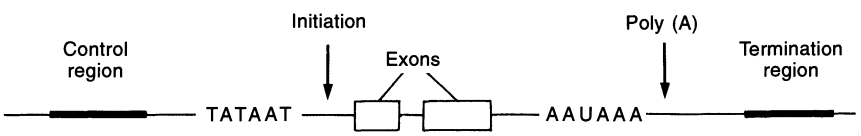


Fig. 2.1. A plan of the transcription unit of a protein-coding gene trancribed by polymerase II

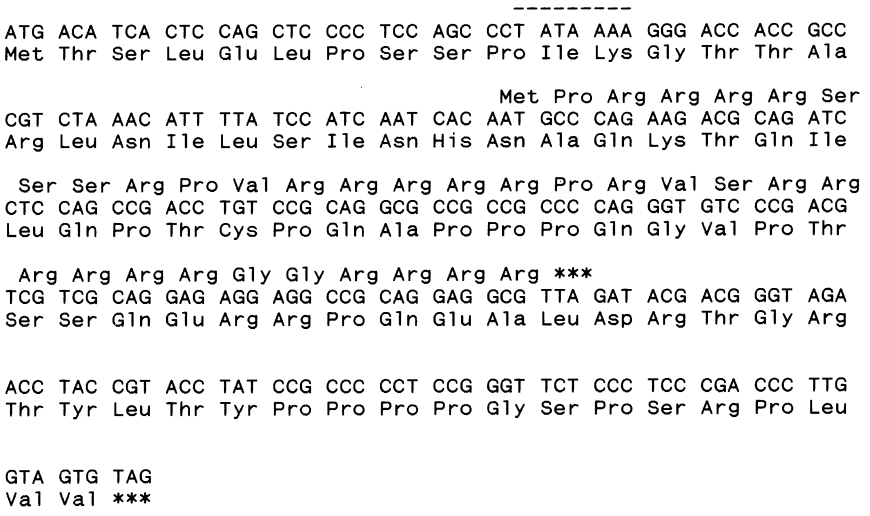


Fig. 2.2. The protamine gene of the rainbow trout *Salmo gairdneri* codes not only for protamine (*lower sequence*) but also, using a shifted reading frame for a second protein "Y" (*upper sequence*). The TATA box of the Y gene is marked with the *dashed line*; the termination codons are indicated by *asterisks* [212]

GnRH (gonadotropin-releasing hormone) locus of the rat, the complementary strand includes a gene (SH) of unknown function whose exons partly overlap those of the GnRH gene. The GnRH gene is transcribed in the brain and the SH gene in heart muscle [8]. Transcribed sequences that overlap on both strands are also observed in the mouse and in the region of the Dopa decarboxylase gene in *Drosophila* [415, 477].

2.1.3 The Genetic Code

The triplet code by which the protein-coding DNA sequences and their mRNA transcription products determine the amino acid sequence applies universally from *E. coli* to man *(see bookmark)*. It makes sense that the code was fixed very early in evolution as each change in the code would affect all the proteins of an organism, with catastrophic consequences. However, there are exceptions to this universal code in animals, which the codon reassignment hypothesis attempts to explain. This assumes that specific codons and relevant anticodons can disappear from the repertoire during evolution. If the codon reappears later, the tRNA with the appropriate anticodon may now have a different function [334]. Exceptions to the universal genetic code are to be found, on the one hand, in mitochondrial DNA (p. 53 and 55) and, on the other hand, in the Ciliophora. A long time ago it was observed that the mRNAs for the surface antigen of *Paramecium* were not accurately read in heterologous in vitro systems because they contain many TAA and TAG codons, which mean "stop" in the genetic code. TAA and TAG in the gene sequence and UAA and UAG in the mRNA code for glutamine or glutamic acid here and in other ciliates (*Oxytricha, Stylonchia* and *Tetrahymena*); only UGA remains as a stop codon. This alternative code is not valid in all ciliates; the tubulin and actin genes of *Euplotes* use UAA as the stop codon [139, 180, 298]. Three glutamine-specific tRNAs have been sequenced in *Tetrahymena thermophila*. One of them has the anticodon 5′-UmUG (where mU represents 2′-methyluridine) that recognizes the standard Gln codons CAA/G; both of the other tRNAs have 5′-UmUA (for the codon UAA) and 5′-CUA (for the codon UAG) [175]. Because of the size and complexity of the ciliate genome, it is difficult to imagine that this deviation from the universal genetic code occurred at a relatively late stage of evolution. The separate evolutionary line of the Ciliophora probably arose before the separation of the fungi, plants and animals [200]. Other molecular data speak in favour of this hypothesis (p. 170).

With the exception of tryptophan and methionine, each of the 20 primary amino acids is encoded by several synonymous codons that, arginine excluded, all differ in the third position *(see bookmark)*. The synonymous codons of each amino acid are not uniformly used; it is routine in the analysis of newly sequenced genes to state the codon usage [462]. The codon usage differs significantly between groups of organisms and species, and also between single genes of a species. The use of synonymous codes is limited by the base composition of the total genome or DNA segment (isochore) in which the gene is found. As the tRNAs specified by each synonymous codon are generally present in the cytoplasm in different amounts, the rate of protein synthesis should also

be dependent upon codon usage; in fact, in many cases codon usage, tRNA frequency and gene expression are correlated. Even where tRNA availability is uniform, some codon-tRNA combinations are translated faster but less accurately, e.g. in the cell-free systems of *E. coli* UUC (for phenylalanine) is used twice as frequently as UUU [109].

2.1.4 Introns

A basic difference between eukaryote and prokaryote genes is that the coding sequences of the former are not continuous but are interrupted to give partial sequences (exons) by non-coding, intervening sequences (IVS or introns) (Fig. 2.1). The introns are also initially transcribed but are removed by slicing during processing of the RNA (see Fig. 2.9). The presence of the introns dramatically increases the chances of successful intrachromosomal recombination of genetic information during crossing-over. Variations in the splicing of particular genes allows the formation of different mRNAs and proteins from the same gene (p. 40). Only in a very few eukaryotic genes are introns absent, e.g. histone genes, interferon genes, most heat-shock genes, and the genes of the trypanosomes. In contrast, an exception to the rule that prokaryotes do not possess introns has recently been found: the gene for the 23S ribosomal RNA of the archebacterium *Desulfurococcus mobilis* has a 622-bp intron that is excised during rRNA processing [230].

Most genes contain one or just a few introns, but there may be more than 10; e.g. there are 13 in the serum albumin gene of the rat, at least 14 in the β-crystallin gene and 16 in the ovtransferrin gene of the chicken, 33 in the vitellogenin gene of *Xenopus*, and more than 50 in the procollagen II gene of the chicken [45, 317, 446]. Ciliate genes are normally poor in introns. Recently, however, a gene was found in *Tetrahymena* with about 30 introns within 15 kb [287]. A relationship exists between the length of coding and non-coding sections such that protein-coding genes in excess of 550 bp are invariably interrupted by introns; the intron:exon ratio greatly increases with increasing exon length [45, 317]. Extremes in exon size are known: e.g. exon 17 in the troponin-T gene of the chicken has 6 bp (equivalent to two amino acid codons), and an exon in the human renin gene is only 9 bp long [90]; in contrast, there is a 14 000-bp exon in the fibroin gene of the silkworm *B. mori* and one of 7572 bp (exon 26) in the human apolipoprotein-B gene [37]. In general, introns have lengths of up to 10 000 bp, although very large introns of 60 and 79 kb have been found in the antp and dunce genes, respectively, of *Drosophila* [69, 150]. In contrast to exon lengths, mean intron lengths differ considerably between the genomes of different groups of organisms. Introns in the middle of coding sequences of vertebrates have a mean length of 1127 bp, in insects 622 bp, in the lower fungi only 86 bp, and in higher plants 249 bp [183]. Few introns are less than 75 bp long [406]. Experiments with artificially modified globin genes have shown that shortening the introns to 69 bp disturbs the splicing process [446].

Exons and introns, together with the non-translated regions in front of and behind the coding sequences, make up the typical eukaryotic transcription unit (Fig. 2.1). In addition, each gene requires further regions on both sides to ensure optimal expression. Thus, with the presence of introns, and non-translated and non-transcribed regions, eukaryotic genes require stretches of DNA many times the length of the actual protein-coding sequence. So, for example, in the gene for the 28-kDa large ovomucoid of the chicken, the sequence coding for the finished protein is 561 bp long, with another 260 bp for the signal peptide, the 5′ and 3′ non-translated regions, and the initiation and termination signals; these 0.82 kb, together with a total of 4.7 kb of introns, make up a gene of 5.5 kb; with the addition of about 4 kb of territorial DNA on both sides, it is seen that a coding sequence of 0.56 kb requires not less that 13.5 kb of DNA [45, 317]. In *Drosophila* there are genes with even larger introns; e.g. ubx, with a total length of 75 kb, or antp, at 100 kb, both code for relatively small proteins [189]. Introns lie mostly within coding regions, sometimes in the 5′ transcribed but not translated region, and seldom in the 3′ region. In homologous genes of different species or in multiple genes resulting from duplication, they are generally to be found at homologous positions; they can, however, disappear or appear during the course of evolution. Particularly large variation in intron position is found in the actin genes (see Fig. 10.4; p. 341). In length and sequence, introns show especially rapid evolutionary change and distinctive polymorphism. Thus, in the actin genes, differences in the number, size and position of introns are found not only between different species but even between different members of this multi-gene family in a single individual. In *Drosophila* and other dipterans, introns are

found in only some of the multiple 28S rRNA genes. The insulin genes of the vertebrates consistently have two introns, one in the 5′ non-coding region and one in the region of the C-peptide; in rats, however, the C-peptide intron is missing from one of the two insulin genes present [45].

In some introns there are open reading frames (ORFs) without stop codons that can code for proteins. The first example of an exon in an intron was the maturase gene, discovered within the yeast cytochrome-b gene in 1981; this plays a role in the splicing of the cytochrome-b transcript. Two examples are known in *Drosophila:* The first is the gart gene, which codes for three enzymes of purine synthesis and contains in its first 4142-bp-long intron on the complementary strand, i.e. being read in the opposite direction, the fully functional gene for pupa cuticular protein (PCP) with 184 amino acids, a homologue of four known larval cuticular proteins (LCPs). In contrast to the gart gene, the PCP gene is expressed only in the epidermis of the pre-pupa. The PCP gene situated in the gart intron itself contains an intron of 71 bp between codons 4 and 5. A sequence comparison with the relevant yeast gene clearly shows that the PCP gene was inserted later [189]. Sequence comparisons suggest that the PCP gene arose about 70 million years ago by duplication of the then single LCP gene and is, therefore, much older than the present *Drosophila* species [307]. The second example is the molecular organization of the dounce gene which codes for a cAMP phosphodiesterase and has proven to be particularly complex, with three divergently transcribed genes nested within its introns. Two of these, Sgs-4 and Pig-1, are nested within the extremely large 79-kb intron that separates exon 3 from exon 2 [69, 143]. Corresponding examples are also to be found in mammals and man. A gene of unknown function is located in the first intron of the mouse β-glucoronidase gene; introns 1 and 3 of the human pre-albumin gene contain ORFs that could code for proteins of 37 to 69 amino acids [442, 465].

There are two basically different hypotheses about intron evolution that relate to the localization of introns within coding sequences and the splicing process. According to the "association" hypothesis, introns and splicing were already present in the first organisms but were lost in the prokaryotes when DNA content was reduced in favour of more rapid replication. The second hypothesis ("divisive" intron origin) assumes that coding sequences originally were continuous and lacking in introns, and that only during eukaryote evolution were introns introduced and the splicing mechanism developed. Here one can think of introns arising from transposable elements that carried on their ends signals specific for insertion into DNA sequences that were already present and for the splicing process. The first hypothesis assumes that longer coding sequences arose during the early evolution of organisms by the joining of shorter sequences. Speaking in favour of this is the existence in prokaryotic genomes of a primitive, RNA-dependent splicing mechanism and the albeit rare occurrence of introns. A particularly strong argument for the associative origin of introns is that in many genes single exons represent subregions of polypeptide chains that either fold in a compact manner (modules) or are structurally and functionally autonomous (domains). On the other hand, for many genes there is no correlation between intron position and domain structure. The appearance of new introns can be demonstrated in the evolution of the serine proteases (see Fig. 3.6; p. 91). The tubulin and actin genes of eukaryotes vary enormously in the number and position of introns: α- and β-tubulin are homologous and agree in about 40 % of their amino acids, but only one of the 17 possible intron positions in α-tubulin and the 19 possible intron positions in β-tubulin coincide. All these positions, however, regardless of whether they contain an intron or not, show characteristics of the typical intron-flanking "junction sequence"; they appear to be "proto-splice sites" predestined for the insertion of introns [107, 155, 421].

The rearrangement of exons (exon shuffling) has played a large role in molecular evolution [366, 421]. One finds, for example, in genes of very different function and origin, one or more exons that code for an amino acid sequence homologous to epidermal growth factor (EGF) (see Fig. 3.6; p. 91). That introns may be very old is nicely demonstrated by the genes of triosephosphate isomerase; these are homologous in all organisms. In *E. coli* and baker's yeast these genes have no introns, but they have five introns in the fungus *Aspergillus nidulans*, six in vertebrates, and eight in maize. Five of the introns have the same position in maize and man. Many of the introns lie in regions of the triosephosphate isomerase genes whose amino acid sequence is highly conserved in all organisms; this speaks against them being inserted later and at random [155].

The idea that the genomes of the first organisms contained non-coding sequences can be

supported both theoretically and statistically. Computer simulation shows that in random sequences of the four bases, a stop codon appears after a maximum of 600 bp, and longer sequences can be obtained only by excision of the region of the stop codon and splicing of the resulting fragments; this coincides with the fact that only genes of less than 600 bp have no introns [392]. If one assumes a length of 20 kb for the primordial nucleic acid molecule of the first organisms, then a stop codon-free sequence (ORF) of 550 bp or more would be expected in a maximum of 4.6 % of all molecules; according to this, the excess of non-coding sequences found today in most eukaryotes represents the original situation [318]. The correspondence between exon and domain boundaries, used as an argument for the associative origin of introns, does not apply to all proteins; furthermore, even homologous genes do not always have the same number or location of introns. Thus, the origin of introns through the insertion of transposable sequences is by no means excluded; perhaps the two mechanisms exist side-by-side.

2.2 Multiple Genes and Pseudogenes

Many genes occur in the haploid genome in several, often very many, more or less identical copies. This was already postulated two decades ago because it was noted that there are many cases where different proteins (globin isoenzymes) with the same or a similar function exist in the same individual. Today, one can detect all the similar sequences in the genome by hybridization with a DNA probe. Such experiments have made it clear that there are further copies of almost all genes somewhere in the genome, although in many cases these may be non-functional. Multigene families make biological sense in two ways.

1. Dose repetition: In the biosynthesis of almost all proteins, one gene is sufficient for a high rate of synthesis because of the two amplifying effects of transcription and translation. In the formation of different RNA types, however, including the 5S rRNA and the different tRNAs, the second amplifying step is missing. Many identical genes are required in this case to satisfy requirements. The only example of dose repetition of a protein-coding gene is that of the histones, which must be produced in a short period during the cell cycle in an amount similar to that of the DNA. Here again, the amplifying effect of protein biosynthesis is insufficient. Only gene duplication or multiplication in the germline is appropriate, i.e. a process of molecular evolution that affects all the cells of an animal. The multiplication of the genome or of single genes can also be limited to particular somatic cells. This also serves to cover increased requirements for gene products but it is a process of individual biochemical adaptation and will be dealt with elsewhere (p. 27).
2. Variant repetition: When multiple genes produced by gene amplification develop independently in evolution (diversification), there arises a family of related but significantly different gene products that may be adapted to various conditions in different cell types or in different developmental stages. This book contains many such examples.

Multiple genes arise through gene duplication, the genetic mechanism of which is most likely unequal crossing-over. The duplicated genes are at first arranged one after the other (a tandem cluster); the genes of a family can also be spread throughout the genome (dispersed genes); and finally, there are multi-gene families with most of the members together in one or more clusters and single copies (orphons) at other positions in the genome. This phenomenon was discovered in 1981 in the histone genes of the sea urchin *Lytechinus pictus*, where more than 50 orphons have been found, but is apparently quite widespread. In cases of dose repetition, the genes of a gene cluster are generally regulated together and, through particular evolutionary mechanisms (horizontal or concerted evolution), remain similar in sequence (homogeneous). Neither of these is valid for cases of variant repetition. Orphons are excluded from horizontal evolution and, therefore, show great diversification.

In all multi-gene families there are probably members that are not transcribed because of the presence of some sort of defect. Such genes are rather misleadingly termed pseudogenes. The name was coined in 1977 by Jacq, who discovered the first such gene within the 5S rRNA genes of the clawed frog, *Xenopus laevis*. Many pseudogenes are now recognized, e.g. in the gene families of the globins, the immunoglobulins and the tubulins of various mammals, cytochrome c of rats, actin and metallothionein of man, and the cuticula proteins of *Drosophila*. In some gene families, the number of pseudogenes greatly

exceeds that of functional genes, the proportions being, for example, 4:1 for the cytochrome c gene of rats, 8:3 for the tubulin genes, 10:1 for the U1 RNA genes of man, and between 6:1 and 32:1 for various ribosomal protein genes of the mouse [464]. Two categories of pseudogenes are recognized according to their origins: the "traditional" and the "processed" pseudogenes. The traditional pseudogenes arose through the duplication of gene copies that, after a period of normal function, became defective, e.g. by single base substitutions that introduce intrinsic stop codons (nonsense mutations) or by the insertion or deletion of one or two bases that shift the reading frame (frameshift mutations). Genes may also be inactivated by changes in transcription or translation signals or by substitutions in the control region.

The pseudogenes of the second category are distinguished from functional genes by the absence of introns and by their remote location in the genome, far from their functional counterparts. Many possess a 3′ terminal poly(A) sequence of 11–38 bp; many are flanked by identical, similarly oriented sequences (direct repeats). All these characteristics speak for a mode of origin whereby a DNA copy produced by reverse transcription of a finished (processed) mRNA is inserted somewhere in the genome. The processed pseudogenes thus belong to the large number of RNA-dependent transposed sequences (retroposens), and are also termed retropseudogenes. This type of pseudogene is found in many multi-gene families of mammals but, up to now, has been found only rarely in other vertebrates and the invertebrates. Only by sequence analysis will it be possible to decide whether the widely spread orphons are really pseudogenes of this category [452, 464, 470]. The retropseudogenes and the other retroposons will be discussed more thoroughly below.

2.3 Repetitive Sequences and Mobile Elements

Whereas the small genomes of the prokaryotes consist almost entirely of unique sequences, the much larger eukaryote genomes consistently contain a proportion of repetitive sequences, which in unicellular or multicellular animals may be between 10 and 50 % of the DNA. The proportion of repetitive, information-deficient sequences in DNA fragments can be estimated from the rate with which single strands reassociate to duplexes (reassociation kinetics). The product of the starting concentration and half-life, c_0t, increases with increasing complexity. The c_0t value distinguishes highly repetitive DNA with about 10^6 copies, middle repetitive DNA with 10^2–10^5 copies, and single-copy (unique) DNA with one or a few copies. The multi-gene families of the ribosomal RNAs and histones fall into the category of middle repetitive DNA, although only with proportions of a few percent. The Protozoa, with their small genomes, have only 10–20 % repetitive DNA [203]; amongst the vertebrates, the teleost *Arothron diadematus* has the lowest proportion of repetitive DNA with 13 % [345]; in contrast, a relatively large proportion, particularly of highly repetitive DNA, is found in the extremely large genomes of the Urodela [43].

Many of the highly repeated sequences form tandem clusters in the heterochromatin of the centromere and telomere regions of chromosomes (satellite DNA). Tandem repeats of short sequences are also found in the telomeres themselves, the specialized structures at the ends of chromosomes that play an important role through their replication, stabilization and interaction with the nuclear membrane. The molecular structure of the telomeres was first investigated in the Ciliophora where 10^4–10^7 of such chromosome ends are to be found in the macronucleus. Telomeric repeats of $(TTGGGG)_n$, were found in *Tetrahymena*, and of $(TTTTGGGG)_n$ in the hypotrichous Ciliophora *Oxytricha, Euplotes* and *Stylonchia*. The telomeres have a single-stranded extension to which specific proteins are bound [192]. In addition to the repeats (TTGGGG) and (TGAGGG), one finds especially (TTAGGG) in humans, all classes of vertebrates, and the flagellate trypanosomes. The combined length of the tandem repeats in the lower eukaryotes amounts at the most to 1 kb, but in man it is 10–15 kb, and in the mouse it is up to 100 kb [36, 491]. Short, highly repeated sequences of only 2–3 bp are also widely scattered in the genome, e.g. in introns, in the spacers between genes, and also within longer repeated sequences. All possible types of such short, dispersed sequences are found in very different classes of eukaryotes (man, *Drosophila*, sea urchin, the micronucleus of the ciliate *Stylonchia*, and yeast); examples include repeats such as AA/TT (i.e. poly(A) on one strand and poly(T) on the other), GT/CA or CAG/GTC. Unlike similar satellite-DNA sequences, these dispersed sequences are transcribed [432]. *Droso-*

phila proteins that bind specifically to the sequence $(GT/CA)_n$ have now been found [453]. The proportion of highly repetitive DNA, determined with a radiolabelled DNA probe for example in rats, is higher than the estimate from reassociation experiments [480]. In contrast to highly repeated sequences, middle repetitive DNA is always widely distributed throughout the genome. It is now clear that many of these sequences are transposable, i.e. they can be introduced into other parts of the genome.

2.3.1 Satellite DNA

The name satellite DNA derives from the specific bands that form during density-gradient centrifugation due to the deviations in density caused by the typical composition of highly repetitive DNA. The term is today applied to all tandemly clustered, highly repeated sequences, even when the base composition, and thus the density, is no different from that of the rest of the DNA (cryptic satellites) [31]. With the exception of the large copy number and the tandem arrangement satellite DNAs in animals show considerable variety. The length of the repeat, for example, can vary between 2 and more than 1000 bp. Satellite DNA is mainly to be found in the highly condensed, strongly staining areas of the nucleus (heterochromatin) bound to specific proteins [13]. The amount may differ even between closely related species. The short satellites of *Drosophila melanogaster*, for example, make up 21 % of the genome, those of *D. simulans* constitute 5 %, but in *D. erecta* the value is only 0,5 % [269]. In the flour beetle, *Tenebrio molitor*, no less than 49 % of the genome is made up of satellite DNA, corresponding to the very large amount of heterochromatin seen in this species [343]. Satellite DNA is completely absent from the somatic cells of several species, which leads to the question of whether it has any somatic function at all. It is interesting in this connection that such DNA can be highly methylated in somatic cells but unmethylated in the germline; this is very different to the genes that are normally more highly methylated in the germline and are inactive [179].

The relationship between base composition and density can be exploited for DNA fractionation. Depending upon the molar fraction of guanosine and cytidine (G+C), DNA in a caesium chloride solution at 25 °C has a density of

$$p = 1.660 + 0.098\ (G+C).$$

In mammals and birds, density-gradient centrifugation results in four bands; two have densities less than 1.703, and two are GC-rich, "heavy" bands with densities of more than 1.704. Lower vertebrates usually show a complex pattern in the lower density region; only a very few species also have "heavy" DNA [345]. The simplest satellites are found, for example, in *Cancer pagurus* and other decapod Crustacea, where 10–30 % of the total genome consists of poly(AT). Although satellite DNA is mostly absent from lower eukaryotes, very complicated satellites with repeats of 177 and 196 bp are to be found in the flagellates *Trypanosoma brucei* and *T. cruzi*, respectively [400]. In *Drosophila melanogaster*, four satellites may be isolated from density gradients; one satellite is made up of repeats of more than 300 bp, and the other three consist of 10 or 11 different repeats of 5–9 bp [269]. Unusually complex repeats of 1460 bp are seen in the satellite DNA of the sea snail, *Rapana thomasiana* [282].

Satellite DNA shows rapid evolution, and even closely related species have different repeated sequences [111]. There are, however, exceptions to this rule but these have not yet been explained. For example, the same 6-bp repeat is found in the kangaroo rat *Dipodomys* and the guinea-pig; seven of the ten short satellites of *Drosophila melanogaster* are found in the twin species (sibling species) *D. simulans*, whereas all ten, in fact, occur in the more distantly related *D. erecta* [269]. The mechanisms of "horizontal evolution" lead, on the one hand, to more sequence uniformity within individual tandem clusters, but, on the other hand, to the rapid spread of single-point mutations. Thus, satellite DNA shows a higher degree of variability between individuals of a species than was suspected earlier from measurements of denaturation temperature or density: in both *Drosophila* and mammals, over 10 % sequence difference is found between individual repeats. In addition to sequence alterations by substitution, addition/deletion and rearrangement, changes in the amount through amplification and the elimination of whole regions appear to be frequent events in the evolution of satellite DNA [31].

The multiple satellites of a species often develop one from the other. In man and other primates, one finds a family of repeats of about 171 bp, called α-satellite or alphoid-DNA, in the centromere region of all chromosomes. The alphoid sequences of humans deviate, on average, by 16 % from a common consensus sequence and may be assigned to five groups of closely related

sequences [73]. In cattle, there are eight main satellites that together make up 23% of the DNA. These are apparently all composed of similar 23-bp repeats that probably arose from the same 11- to 12-bp ancestral sequence [430]. The domestic rat, *Rattus rattus*, possesses two satellite DNAs: type I is a 370-bp dimer, whose 185-bp a and b subunits show 60% sequence agreement and apparently arose by duplication of a common 92- to 93-bp ancestral sequence. The other satellite, type I, is a tandem of 185-pb repeats (a′) that show 85% agreement with a. Both types are polymorphous. The brown rat, *R. norvegicus* has only type I [125].

2.3.2 Middle Repetitive DNA and Genome Organization

Middle repetitive DNA sequences are widely dispersed in the genome and show two patterns of organization. *Xenopus* shows the typical short-period interspersion pattern. The largest part of the euchromatin consists of single-copy sequences of about 1–2 kb alternating with short repetitive sequences of 0.1–0.3 kb. The *Xenopus* pattern is apparently widespread and, in addition to its occurrence in fish, amphibians, reptiles and mammals, is found in acrania, sea urchins and various insects [385]. The long-period interspersion pattern, of which the *Drosophila* pattern is typical, consists of repeated sequences of average length 5 kb interspersed with single-copy sequences of up to 35 kb. This pattern is possibly an adaptation to a relatively small genome [153]. The *Drosophila* pattern is also present in the dipterans *Chironomus, Anopheles* and *Sarcophage* as well as in the honey bee, *Apis mellifera;* in contrast, one finds the short-period interspersion pattern in the dipterans *Aedes, Musca* and *Stomoxys*, in the butterfly *Antherea* and in the Thysanura *Thermobia*. An intermediate pattern occurs in the mosquito, *Culex* sp. [81]. Unlike other vertebrates, all birds that have been examined show the *Drosophila* pattern [456]. There are regions of the genome in all animals that deviate from the basic type. Thus, short dispersed repeats are to be found in *Drosophila*, although less often than in mammals [153]. In humans, DNA sections of more than 50 kb, with the basic type of short repetitive sequences, alternate with sections in which long repeats are also to be found [422].

2.3.3 Transposition of Middle Repetitive Sequences

The existence of mobile genetic elements (jumping genes) was postulated in the 1940s by Barbara McClintock on the basis of experiments with maize. The molecular-genetic evidence was first obtained from *Drosophila*, where the sequence "copia" and similar sequences can be detected at various positions in the genomes of different natural populations or laboratory strains, and even in different individuals of the same population or on the homologous chromosomes of individuals [220]. This shows that the transposition and deletion of mobile elements are extraordinarily frequent events, much more frequent even than base substitutions.

The transposition rate of copia and similar sequences in *Drosophila melanogaster* is about 10^{-3} per generation, whilst the insertion frequency of inducer-(I)-factors and hobo elements is even higher [176]. Extreme environmental factors, for example, heat shock in *Drosophila* can increase the transposition rate [221]. For various technical reasons, the direct detection of a transposition in mammals is more difficult. However, in the region of the β-globin cluster of the mouse extensive reorganization due to insertion of an L1 element is quite convincing, as is the allelic polymorphism resulting from the insertion of an Alu element in the region of the prolactin gene of the rat [56, 389].

The duplication, transposition and dispersion of mobile elements and the resulting changes in genome organization (rearrangements) are important mechanisms in molecular evolution as well as the main cause of the dynamic state of the eukaryotic genome [428]. On the other hand, the transposition of mobile sequences may, in many cases, be of no consequence for the phenotype. The fact that the majority of the mobile elements distributed throughout the genome appear to have no biological significance led to the concept of selfish DNA; this concept was formulated in 1980 by Doolittle and Sapienza and, in a similar way, by Orgel and Crick, and has since been the subject of active controversy [116]. "Selfish DNA" is understood to mean DNA elements that, on the basis of specific sequence characters, show the potential for transposition; they multiply and move around the genome and so avoid elimination. These elements "survive", replicate and mutate in the genome like parasites in a host. It is quite possible that later in evolution they may attain a biological significance, just as a

parasite may become a symbiont. Non-functional DNA, without the ability to transpose ("ignorant DNA" or "junk DNA"), would, in contrast, be relatively quickly eliminated during evolution. It must be argued that the concept of "selfish DNA" does not take sufficient account of the dynamic flexibility of the genome: undoubtedly, the genome is a mosaic of sequences that contribute to the phenotype (genes) and those that do not; however, there is a constant exchange between these two compartments [470].

The mechanisms of transposition have not yet been completely explained. Whereas transposition involving extrachromosomal DNA is important in the prokaryotes, transposition via RNA is limited to the eukaryotes and is distinguished by the term retroposition; sequences transposed in this way are called retroposons. The direct transposition of DNA elements can happen either by replication of the transposed element and insertion in the target site (replicative transposition), or by excision of the element and its transfer to a new site. Such a "cut-and-paste" mechanism is assumed for the P element of *Drosophila*. The enzyme required ("transposase") is coded by the P element itself. The intermediate products of the transposition are extrachromosomal, circular DNA molecules, such as are often found, for example, in tissue cultures of *Drosophila* and various vertebrates [347].

Retroposition begins with the transcription of the transposable elements. According to the type of retroposon, this is apparently carried out with RNA polymerase II or III; in any case, the ubiquitous, long LINEs contain a Pol II promoter, and the shorter SINEs contain a Pol III promoter. The RNA intermediate is temporarily included in a ribonucleoprotein particle. Finally, a DNA copy is produced from the RNA by reverse transcription and is reinserted into the genome [325, 353, 470]. A "staggered cut" occurs at the insertion site, as does a short sequence duplication (Fig. 2.3). All intact retroposons code for their own reverse transcriptase, which has significant homology to the enzyme of the retroviruses [488]. Little is known about the choice of the insertion site, although several retroposons show definite insertion specificity. Thus, the CRE-1 elements of the flagellates *Trypanosoma* and *Crithidia* are found exclusively in the genes of the mini-exons that are transpliced into all mRNAs [144]. The insertion elements R1 and R2 in the rRNA genes of several insects are site-specific mobile elements. In the silkworm, *Bombyx mori*, it has been shown that the R2 element codes for an endonuclease that is highly specific for the 28S rRNA gene and produces a staggered cut at the insertion site with a 4-bp sticky end [325].

Two types of retroposons are recognized according to structure and origin: "viral" and "non-viral". The viral retroposons resemble the tumorigenic RNA viruses (retroviruses) of mammals and birds. These are of special interest because they are able to incorporate particular host genes into their genome and then transfer these to other regions of the host genome. The host genes involved are known as cellular oncogenes, and their equivalents in the viral genome are termed viral oncogenes. At present, about 20 viral oncogenes are known and these are named after the virus in which they were found: *v-src* (Rous sarcoma), *v-myc* (myelocytomatosa), *v-ras* (sarcoma), etc.; the corresponding cellular oncogenes are accordingly termed *c-src, c-myc* and *c-ras* [89]. Sequences that are homologous to particular cellular oncogenes have been found throughout all classes of vertebrates and in *Drosophila*, the nematode *Caenorhabditis* and yeast [27, 157]. Most oncogenes are aberrant forms of components of the intracellular signalling system, i.e. growth hormones and their receptors, ligand-dependent protein tyrosine kinases, GTP-binding proteins, cytoplasmic serine/threonine kinases, and proteins involved in gene regulation. The DNA of the non-transducing retroviruses carries in its central region structural genes, in particular for reverse transcriptase, integrase (pol) and viral structural components (gag, env); it also carries LTRs and short inverted repeats at both ends and these are recognized by the mechanism for insertion into the genome (Fig. 2.4). The transcription signals (promotor, enhancer, poly(A) signal) are organized in the LTRs such that transcription begins at the anterior LTR and ends at the poste-

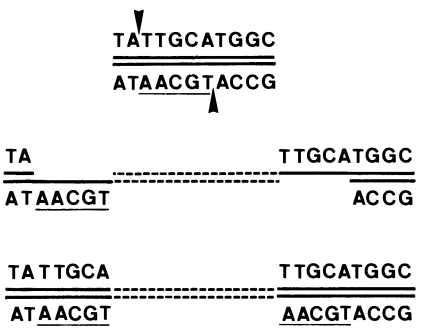


Fig. 2.3. During the insertion of a mobile element, the strands of the DNA double helix are cut at slightly different points (staggered cut), producing a short sequence duplication at the insertion site

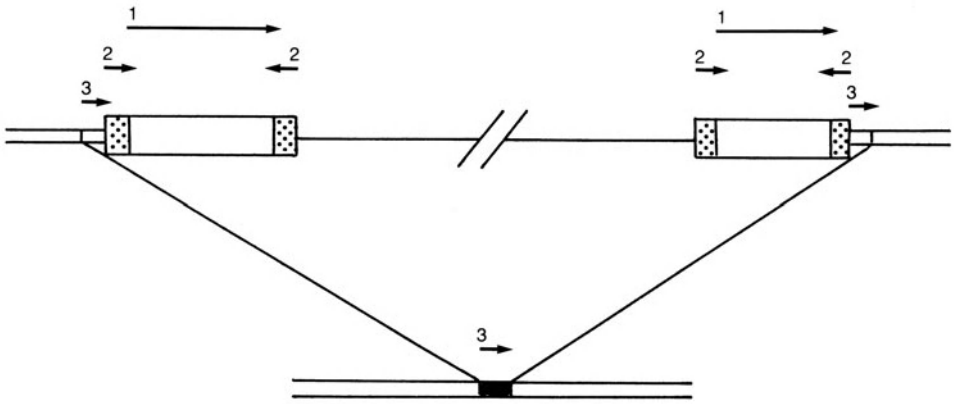


Fig. 2.4. Plan of retrovirus-provirus or retrotransposon [136]. *1,* Long sequence repeats with the same orientation (direct repeats); *2,* short repeats with the opposite orientation (inverted repeats); *3,* duplicated sequence at the insertion site. The *arrows* show the direction of reading

rior LTR. Transducing retroviruses mostly have defective viral genes which are partly substituted by oncogenes; such viruses need non-transducing helper viruses for their multiplication. Retroviruses are distributed throughout the vertebrates. The best-known retroposons are "copia" and the families similar to copia, which were discovered in *Drosophila* but are also found in many other animals and even in yeast and higher plants [461], and the similarly ubiquitous foldback (FB) element, "THE1" in humans and "VL30" in rats and mice [470]. The non-viral retroposons contain no LTRs or flanking inverted repeats; they usually have a 3′ terminal poly(A) sequence, and also cause a duplication of the target sequence that is longer (7–21 bp) than with the viral type (4–6 bp). Such non-viral retroposons have their origin in various cellular RNAs, with the exception of the rRNA and 5S rRNA, i.e. from mRNAs, tRNAs, 7SL RNAs and snRNAs. They correspond almost always to the fully spliced RNA, i.e. in contrast to the viral retroposons, they never possess introns but they are quite often shortened at the 5′ or 3′ end. With the exception of these few structural similarities, the non-viral retroposons show significant variation. Their length may be anything from 33 bp (retropseudogenes of the U2 snRNA of humans) to over 6 kb (unshortened L1). The processed pseudogenes, and the long and short middle repetitive sequences (LINEs and SINEs) of the mammals belong to this group. At first it was believed that non-viral retroposons were seldom to be found in other groups of animals [470], but it has since become clear that SINEs occur in fish, amphibians and reptiles [123], and LINE-like elements are distributed amongst not only different vertebrate classes but also insects and unicellular organisms [33, 301, 324, 325]. A large degree of variability is seen in all families of retroposons; mutations can spread and be present at high frequencies so long as they do not hinder transposition.

2.3.4 Mobile Sequences of Invertebrates

Approximately 12% of the genome of *Drosophila melanogaster* consists of middle repetitive sequences, of which about one-quarter concern dispersed tRNA genes, tandem-repeated rRNA and histone genes. The remaining three-quarters consist of about 50 families of widely dispersed elements, each with 10–100 copies and lengths of 5–10 kb. Other *Drosophila* species, such as *D. algonquin* and *D. affinis*, apparently have fewer repetitive elements but these have larger copy numbers [68, 191]. The best known are "copia" and the copia-like sequences (gipsy, 412, 297, 17.6, mdg1, B104, 3S18) [26, 179]. As shown by hybridization with the relevant DNA probes, copia and gipsy are widely distributed throughout the whole *Drosophila* genus, although both the sequence and number vary greatly between species [419]. Thus, in *Drosophila melanogaster* there are 20–60 copies of copia, in the very similar sibling species *D. simulans* there are only 2–4, and in *D. hydei* the element is not present [417]. The copia-like sequences correspond to vertebrate retroviruses in terms of their LTRs; they have transcription signals and flanking, inverted repeats (Fig. 2.4). The open reading frame (ORF) within copia codes for a polyprotein 1409 amino acids long that, like the retroviruses, encompas-

ses a nucleic acid-binding protein, a protease, an integrase and a reverse transcriptase. In cultured *Drosophila* cells, one finds retrovirus-like particles, with copia RNA and transcriptase activity [490].

Foldback (FB) elements in the genus *Drosophila* are found only in the species groups of *D. melanogaster, D. montium* and *D. takahashi*, but otherwise are widely distributed in the eukaryotes. They carry at both ends a long inverted repeat that results in the folding of the denatured element. In contrast to the complex terminal repeats, there is a tandem sequence of 31-bp repeats towards the middle [397]. Two flanking FB elements can mobilize large DNA sections and transport them to other genome positions; this has been shown for the loci "white" (w), "roughest" (rst) and "no-ocelli" (noc) in *Drosophila* [48, 71]. The element "mariner", 1.3 kb long and with terminal inverted repeats, is also confined to particular species of the *Drosophila* genus, where it produces, for example, eye-colour mutations. It is found in 20–30 copies in the genome of *Drosophila mauritiana*, in lower numbers in, e.g., *D. yakuba* and *D. simulans*, and not at all in *D. melanogaster* and *D. erecta* [53].

P elements have spread throughout the genus *Drosophila* by some as yet unknown mechanism. The first P elements were observed in the 1950s in North American populations of *D. melanogaster.* In the following two decades, these sequences spread into Europe, Asia and Australia, so that today P-free (M) strains are to be found in only a few places. Contrary to earlier assumptions, P elements are also present in other species of *Drosophila*, not as it happens in those most closely related to *D. melanogaster* but in the distantly related species group based on *D. willistoni*. This has led to the idea that P elements were transferred from a species of the *D. willistoni* group to *D. melanogaster* through some sort of infection process [103, 172]. P elements can be used as vectors to transfer foreign sequences into the genome of *D. melanogaster* [95]. Complete P elements, which are only transposed in cells of the germline, are 2.9 kb long, carry terminal repeats of 31 bp but no LTRs, and contain four open reading frames (ORF0 to ORF3) that together code for the 87-kDa enzyme (transposase) necessary for the transposition in germline cells. In somatic cells, splicing of ORF0, ORF1 and ORF2 results in a 66-kDa protein that acts as a transposition repressor [95, 361]. Crossing of a P male with an M female drastically increases the P-element transposition rate; this causes multiple mutations and chromosomal rearrangements and leads to over 90 % sterility. This effect is known as "hybrid dysgenesis". Reciprocal crosses between M males and P females, or intercrossing of P males and females, does not increase the P-element transposition rate; this stems most probably from the large amount of the 66-kDa protein that is found in the eggs of P females [95].

Hybrid dysgenesis can also be induced in *D. melanogaster* by the inducer (I) factors; these are 5.4 kb long, without terminal repeats but have a 3′ poly(A) sequence. They contain two open reading frames: ORF2, which encodes a reverse transcriptase, and ORF1, which encodes a cysteine-rich protein similar to the retroviral gag protein. During the transposition of I factors, an RNA is produced through polymerase II activity and forms a ribonucleoprotein complex with the products of ORF1 and ORF2. The genomes of flies from inductive I strains contain incomplete and about 15 complete I factors; only defective I factors are found in the responsive R strains. These factors are widely spread in the *Drosophila* genus, but have only relatively recently been transferred to *D. melanogaster.* The I factors of the closely related species *D. simulans* are more similar to those from *D. melanogaster* than those from *D. teissieri* [1, 454]. I factors closely resemble the LINEs of mammals, and this is true for a series of further insect retroposons such as F-, G-, and doc elements and jockey in *Drosophila*, or T1Ag in *Anopheles gambiae* [33, 301, 325].

In comparison with those of *Drosophila*, the mobile elements of other invertebrates have been investigated to only a limited extent. The single exception is the Tc1 element of *Caenorhabditis elegans;* this small nematode can be easily cultured and has been extensively investigated at the molecular biology level. The Tc1 element has only 1610 bp with terminal inverted repeats of 54 bp, and produces a 2-bp duplication of the target sequence. Tc1 can be transposed directly; the ORF of 273 triplets codes for a transposase carrying an N-terminal DNA-binding sequence [388]. Only 25 copies of the element are present in the Bristol strain of *C. elegans*, compared with the 250 copies in the strains Bergerac and DH424. Contrary to previous opinion, the elements are not completely identical; 20 of the 250 elements in Bergerac and one in Bristol have a cutting site for the HindIII restriction enzyme; this is missing from all the others [369]. In most strains with few Tc1 copies, there is no or negligible transposi-

tion in the germline [18]. Thus, mutations caused by Tc1 insertion into the unc-54 gene for the heavy chain (HC) chain of myosin are frequently found in Bergerac but never in Bristol [121]. Reversion of the mutated gene to the intact unc-54^+ is 1000 times more frequent in somatic than in germline cells; the excised elements are detected as extrachromosomal DNA [370]. In addition to Tc1, the related elements Tc2 to Tc5 are also found in *C. elegans*. Other elements similar to Tc1 are TCb1 (previously "Barney") from *Caenorhabditis briggsae*, HB1 from *Drosophila melanogaster*, and Uhu from *Drosophila heteroneura* [47, 181].

The free-living nematode *Panagrellus redivivus* shows high mutation rates in the DNA region that is homologous to the unc-22 gene of *C. elegans*. This is due to the insertion of a PAT-1 element of 4.8 kb that appears in 10–50 copies at different locations in various strains of *P. redivivus*. PAT-1 can probably be used as a vector for gene-transfer experiments in *C. elegans* [266]. A retrovirus-like element (Tas), recently discovered in *Ascaris lumbricoides*, consists of an internal sequence of about 7 kb and two flanking LTRs of 256 bp. The germline genome contains a total of 50 Tas elements at about 20 gene locations, with a 2:1 distribution between the variants Tas-1 and Tas-2. Because only one-quarter of the Tas-1 but all the Tas-2 elements are eliminated during chromatin reduction in somatic cell formation, Tas-2 elements are restricted to the germline [9].

2.3.5 The Retroposons of Vertebrates

Many sequences found in the mammalian genome possess some of the characteristics of retroviruses: long direct repeats (LTRs), internal ORFs and short duplications of target sequences, but only to a very limited extent do they represent infectious viruses. The VL-30 and LTR-IS sequences of the house mouse and several other *Mus* species [384] may be mentioned in this respect, together with the "mys" family of the North American white-foot mouse *Peromyscus leucopus*; this family is not found in *Mus domesticus* [475].

The human "THE1" repetitive sequences have corresponding characteristics. These are viral retroposons with a length of 2.3 kb and 350-bp LTRs which duplicate a 5-bp sequence at the insertion site. They code for a poly(A)-RNA of about 2 kb that, however, shows no similarity to known retrovirus products; thus, THE1 is apparently transposed passively [104, 470]. The human genome contains approximately 10000 THE-1 sequences. The element is found in all primates, although in only a few copies in the prosimians. The THE-1 of the prosimian galago is only in part homologous to the human element; a new combination of existing genomic sequences apparently occurred in primate THE-1 evolution, probably with the participation of a retrovirus [383]. The CR1 element of the chicken, which is detectable in the neighbourhood of several genes, probably also belongs to the viral retroposons. This consists of sequences of about 300 bp with short terminal inverted repeats with opposite orientations, according to whether they are located 5′ or 3′ of the gene in question [424]. It is still not clear whether sequences like CR1 of the chicken or OAX in the clawed frog *Xenopus* are really capable of retroposition [470].

Two groups can be distinguished amongst the non-viral retroposons of mammals: the short SINEs with a length of 70–300 bp and usually more than 10^5 copies; and the longer LINEs with a length of more than 5 kb and never more than 10^4 copies. The SINEs consistently carry an internal Pol III promoter and mostly also a 3′ poly(A) tail. Only in the case of the C family of the ruminants can one find the more complicated sequence $(AGC)_n$ instead of $(A)_n$ [470]. The best-known SINE family is the Alu sequence of humans and other primates; this owes its name to the restriction site for the endonuclease Alu I. In humans, about 500000 Alu copies are present per genome and make up 5–6% of the DNA. Assuming a regular distribution, an Alu sequence would be found every 5–8 kb of human DNA; however, they can occur in clusters. Thus, the 4826-bp intron of the 5β tubulin gene contains ten Alu, and there are four in a 2.7-kb spacer between the rRNA genes [158, 290]. The Alu elements are still mobile. For example, in the β-globin cluster of the gorilla there exists an Alu element between δ-globin and β-globin, but this is not found in any other primate; conversely, the intron 4 of the human α-fetoprotein gene contains an Alu that is not present in the gorilla [373, 440]. Alu-like elements are widely distributed in mammals. They show extensive sequence similarity to the 7SL RNA which is found in large amounts in the cytoplasm and which is part of the signal recognition particle responsible for the membrane transport of secretory proteins; however, the central part of the 7SL sequence is missing from the Alu and Alu-like sequences (Fig. 2.5). The human 7SL RNA shows 87% similarity to that of the frog *Xenopus laevis* and 64% to that of *Droso-*

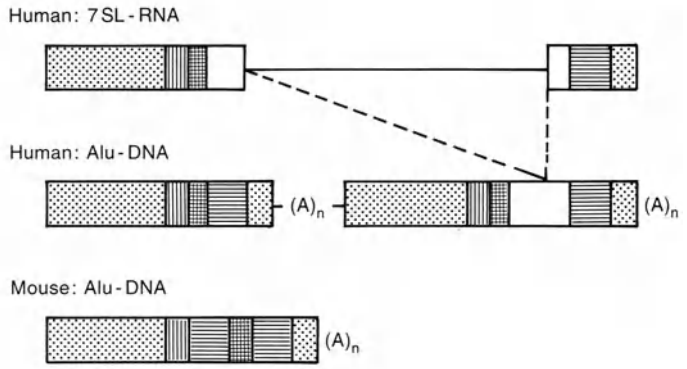


Fig. 2.5. A comparison of the Alu repeated elements of man and the mouse with the human 7SL RNA. The human Alu is a dimer of two homologous subunits, each ending in a poly(A) segment; the subunit in front is 31 bp shorter than that behind. The Alu subunits are derived from the 7SL RNA sequence by deletion of a 155-bp fragment. The mouse Alu (B1) corresponds to the left Alu subunit of man but has a 30-bp duplication [445]

phila melanogaster; therefore, Alu could quite easily be older than the mammals [445].

The Alu elements of the primates are dimers of about 300 bp, with 186 bp of the 7SL sequence missing from the left subunit and 155 bp missing from the right; thus, the right subunit is 31 bp longer (Fig. 2.5). The variation in length results mainly from the variable, 4- to 50-bp length of the 3′ poly(A) sequence. The human Alus are distributed between at least three sequence families that, on average, deviate by 15 % from the consensus sequence [222, 470, 476]. As a consequence of this intraspecific variability, no species-specific differences can be determined between Alus of humans and various primate species [380]. Especially aberrant Alu structures are found only in the prosimian *Galago crassicaudatus*. Here, there are two Alu types, I and II, of which I is very similar to the human Alu in both, its subunits, whereas II is similar only in the right subunit. The left, deviant subunit is also found as a monomer [102]. The Alu-like sequences of the rodents are all monomers of about 135 bp. The B1 element that occurs in about 100 000 copies in the mouse genome corresponds, like the primate Alu, to the 7SL RNA; the deletion of about 155 bp, however, is shifted by about 14 bp towards the 3′ end. Alu and B1 show more than 80 % similarity in their comparable regions [470]. In comparisons between the rat and mouse, B1 sequences showed 90 % similarity [120].

In contrast to the Alu-like elements, the other SINEs of the mammals are predominantly derived from tRNAs; this is true, for example, for the B2 and ID sequences found in addition to B1 in rodents, the C repeats of rabbits, and the unrelated C families of the ruminants, and perhaps also for the monomer of *Galago* [228, 299, 470]. Mobile elements derived from tRNAs are also found in fish, amphibians and reptiles. The majority of the SINEs contain an internal promoter for RNA polymerase III and are thus possibly transposed via an RNA intermediate. There are, however, tRNA-derived repetitive elements that can apparently not be transposed, for example, OAX from *Xenopus laevis* [123]. The sequences of the SINEs have been so altered during evolution that homology to particular tRNAs can no longer be determined. There are apparently also combined sequences derived from descendants of 7SL RNA and tRNA, for example, Alu type II in *Galago* and the bovine monomer family (BMF) [280].

The LINEs of the mammals are particularly individualistic and complex, non-viral retroposons: they can be 6–7 kb long but are mostly shortened at the 5′ end; they contain no LTRs but do have a 3′ terminal poly(A) sequence, several internal ORFs of together more than 3.5 kb, and a sequence duplication at the insertion site. Homologous RNAs are actually found in tumour cells and are possible intermediates of retroposition [131, 470]. The reverse transcriptases and other proteins coded by the LINE ORFs of various mammalian orders show so much similarity to the retroviruses, and also to the reverse transcriptases of the F elements and I factors of *Drosophila melanogaster*, that they must be related in evolution [130, 325]. Each mammalian species contains only one family of LINEs, for example, "Kpn1" of the primates, or "MIF1" ("BamH1") of the mouse. As the LINEs of different mammals show significant sequence homology and are apparently related, the general name "L1" should preferably be used, if necessary together with the beginning letter of the genus or species name, e.g. L1Hs for human LINEs, L1Ca for those from the long-tailed monkey *Cercopithecus aethiops*, and L1Md for the domestic mouse, *Mus domesticus*. The L1 sequences of different species of the mouse genus vary intraspecifically by about 14 % and show species-specific differences correspond-

ing to a divergence of 0.85% per million years [283]. The L1 of man, mouse and rats differ in about 33% of positions [408]. The conclusion that there is only one LINE family per mammalian species has become open to question since the discovery in man of the "L2H" LINE that deviates greatly from L1H and is found only at comparable frequencies in the gorilla; in the chimpanzee it occurs in at least 100 times fewer copies [315]. The human genome contains at least 40 000 L1 copies, and the numbers in other mammals are of the same order. The L1 elements are not uniformly distributed in the genome: in the approximately 60 kb of the β-globin cluster there are no less than nine L1, whereas with even distribution each L1 would be separated by about 150 kb [399, 470].

2.3.6 Retropseudogenes

The retropseudogenes or processed pseudogenes are derived almost exclusively from completely processed RNA and, hence, contain no introns and have a 3′ poly(A) tail. They are very frequent in mammals; the list of pseudogenes in man, rats and the mouse includes more than 30 multi-gene families, e.g. arginine succinate synthase, metallothionein, β-tubulin, β-actin, various immunoglobulin chains and oncogenes in humans, as well as cytochrome c, various ribosomal proteins, α-tubulin and α-globin in rodents [464, 470]. Like the other non-viral retroposons, the retropseudogenes are limited mainly to mammals and are almost unknown in other vertebrates or invertebrates. This is clear from the example of glyceraldehyde-3-phosphate dehydrogenase, for which in humans, rabbits, guinea-pigs and hamsters 25 pseudogenes are presently known; in rats and mice there are actually more than 200, but in the chicken there is not a single one [470]. Retropseudogenes are transposed passively, which means that the enzymes for reverse transcription and insertion must be available in the cell. Because, under these circumstances, each active gene could constantly produce additional DNA, the question arises as to what limits the number of processed pseudogenes; it can be assumed that an equilibrium is established between their formation and their spontaneous deletion [464]. An important difference exists here between Pol II- and Pol III-transcribed genes: because polymerase II uses transcription signals outside of the transcribed region, these are missing in the DNA copy of the RNA; pseudogenes derived from mRNA can therefore not be transcribed. In contrast, all the necessary transcription signals of most Pol-III transcribed genes lie within the transcription unit. Thus, DNA copies of Pol III genes contain all the information for the production of further processed pseudogenes [464]. In rate cases, pseudogenes arise from transcription of an initially incomplete or aberrantly processed RNA. If the transcription signals lying in front of the coding region are also included, then a functional gene copy can result. The only sure example of this at present is the gene for prepro-insulin in rats and mice. These rodents possess two insulin genes that are more or less equally strongly expressed. Gene I contains only one intron in the 5′ flanking non-coding region, whereas gene II contains an additional large intron in the sequence coding for the C peptide. Because the prepro-insulin gene of the chicken and various mammals also contains both introns, gene II may be looked upon as the original rodent gene; gene I apparently arose from a transcript initiated approximately 0.5 kb before the normal transcription start site [470]. A functional gene may also come about if the gene copy produced from processed mRNA is inserted close behind a Pol II promotor or later acquires such a promotor. The intron-free but functional calmodulin gene of the chicken probably arose in this way [167].

2.4 Size of the Genome

2.4.1 The DNA Content of Haploid Genomes (the C Value)

Since the first systematic investigations at the beginning of the 1950s, the DNA content in pg per haploid genome, the so-called C value, has been determined for more than 1000 species of prokaryotes and eukaryotes [62] (Table 2.2). The DNA content appears to increase with increasing level of organization. The lowest values in the animal kingdom are to be found in the Protozoa, e.g. *Plasmodium falciparum*, whose genome is only four to seven times larger than that of *E. coli* [168]; extremely large genomes are found amongst the Urodela. A plausible mechanism for the increase in genome size during evolution is genome doubling (polyploidization), followed by renewed "diploidization" and the diversification of gene sequences, new arrangements of chromosomes, and loss of parts of the DNA; according to this process, all eukaryotes would be diploidized

Table 2.2. The amount of DNA per haploid genome (C value). In DNA containing equal proportions of each base, 1 pg corresponds to approximately 10^9 base pairs, or a molecule 300 mm long

Animal group	DNA (pg)	Reference
Protozoa	0.02–0.38	[129, 159, 203]
Porifera	About 0.05	[17]
Cnidaria	0.35–0.73	[17]
Echinodermata	0.54–3.3	[17]
Mollusca	0.4–5.4	[17]
Annelida	0.09–5.3	[17]
Crustacea	0.09–15.8	[17]
Insecta	0.1–7.5	[17]
Acrania	1.28	[385]
Tunicata	About 0.45	[385]
Chondrichthyes	2.8–9.8	[17]
Osteichthyes	0.39–4.4	[17, 345]
Urodela	15–90	[17, 277]
Anura	1–17	[17, 277]
Reptilia	3.5–8.3	[17]
Aves	1.4–2.5	[455]
Mammalia	3.0–5.8	[17, 455]

polyploids. Animal groups in which no polyploidization has recently occurred should have more or less similar DNA contents; this is the case, for example, for most vertebrates, the sea urchins and starfish. Where the DNA content of the species within a group differs by more than 100 %, genome doubling within phylogenetically recent times may be assumed; in such groups there are tetraploid, and even octaploid, species in addition to diploids, as for example in the Selachei, certain Teleostei (Ostariophysi), Anura, insects and several other invertebrates (see Table 4.6; p. 126). Extreme DNA values of almost 100 pg, as found in the Dipnoi and Urodela, apparently result from amplification of particular sequences, especially repetitive DNA. Even here, the C values within a genus can vary greatly, for example, between 20 and 70 pg in the Urodela genua *Plethodon* [277]. Furthermore, Urodela with high C values have significantly larger cells; an increase in cell volume is accompanied by an increase in copy number of particular multi-gene families, such as histones and the various ribosomal RNAs, although this does not necessarily occur in direct proportion to the C value [195].

The exact number of genes cannot yet be defined for any animal organism. When making estimates, it is necessary to bear in mind that many genes are present in multiple copies. Hybridization of the mRNA, present in a tissue or organism, with the DNA allows an estimation to be made of the total length of unique DNA represented by the mRNA (the "complexity"). Technical objections to this method include, on the one hand, the fact that in no cell and at no time are all genes expressed together and, on the other hand, that the occurrence for non-paired DNA sections in the DNA/RNA hybrid molecule lead to overestimates. For the embryos of many species (mouse, *Xenopus, Triturus, Musca*, sea urchins), this method gives estimates of the mRNA complexity, independent of the C value, of $2–3 \cdot 10^7$ nucleotides (nt). In *Drosophila*, whose genome is actually quite small, the egg has a significantly smaller value of $1.2 \cdot 10^7$nt compared with the larval value of $2.18 \cdot 10^7$nt [202]. With an average mRNA size of 1500 nt, mRNA complexities of $1.2 \cdot 10^7$ and $2.18 \cdot 10^7$nt correspond, respectively, to about 8000 and 15000 genes. In *Drosophila*, it was initially assumed that each of the approximately 5000 chromosome bands (chromomeres) represented a gene; today, the estimate is of 15000–20000 genes. For man, figures of 20000 or 50000 genes are considered plausible. On the other hand, there is probably an upper limit to the number of individual genes that remain more or less constant despite the mutation rate observed in the eukaryotic genome; this upper limit could be about 30000 [11].

Whatever the case might be, relative to the total length necessary for encoding all mRNA, rRNA and tRNA sequences, the amount of DNA in haploid genomes is much too large (C value paradox). This fact has led to the assumption that large parts of the DNA are meaningless and without a function ("junk DNA"). However, several objections to this hypothesis can be raised: an extremely high rate of evolution would be expected for DNA sequences without selection pressure, i.e. 1–2 % of substitutions per million years instead of the 0.1 % found for coding sequences, as well as continuous changes in length through insertions and deletions. In reality, the introns and non-translated (NT) regions on both sides of each gene are surprisingly similar in homologous genes of different species or in different members of a multi-gene family. Such conservative evolution suggests, in fact, strong selection pressures and therefore important functions. Rather than the introns and DNA regions (spacers) between genes, it is the pseudogenes that correspond to "junk DNA". The repetitive sequences that make up so much of the DNA have also not yet been allocated a function. But as these sequences are mobile within the genome, they can be understood as "selfish" DNA that exists parasitically, so to speak, in the genome;

because of this ability they also can not be counted as "junk DNA" (see p. 20).

2.4.2 Increase and Decrease in the DNA Content of Individual Cells

Drastic changes in the DNA content take place during the development of somatic cells. Mitosis-like processes lead to the appearance of giant chromosomes (polyteny) or nuclei with multiple chromosomes (polyploidy). Enlargement of the DNA content can also come about by the multiplication of single genes (gene amplification). DNA multiplication in somatic eumetazoan cells is comparable with the formation of the (somatic) macronucleus in the Ciliophora. There are also several examples of reduction in the DNA content of eumetazoan cells (chromatin diminution).

Somatic polyploidy is widespread in animals [50], and high C values and somatic polyploidy can be understood as alternative evolutionary strategies. The ploidy level varies with the tissue; thus, in the waterbug *Gerris lacustris*, one finds diploid and tetraploid nuclei in muscle, 16-ploidy in the testis septa, and 32- or 64-ploidy in the Malpighian tubules. Also in mammals, for example, the liver, kidney and pancreas cells are often tetra- or octaploid, and the human megakaryocytes are up to 32-ploid. In particular cases, extremely high ploidy levels are found: approximately 500 000 (corresponding to 17–18 duplications) in the silk glands of *Bombyx mori*, 200 000 and 4096 in the giant neurons of the snails *Aplysia* and *Lymnaea*, and 4000 in rat trophoblasts [196, 316]. In the salivary gland cells of the Diptera, the mitotically doubled chromatids remain together and so form the giant polytene chromosomes [50]. At any one time, in *Drosophila melanogaster* these contain 1024 chromatids and 16 000 in *Chironomus* chromatids (corresponding to 9 and 13 replications, respectively); they are arranged so regularly that the individual chromatid loops lead to a cross-banding pattern. In *Drosophila*, about 5000 bands are recognizable in the haploid genome and each is equivalent to between 3 and 300 kb of DNA [268]. Somatic endomitosis in the Diptera produces different effects in different cell types: polyteny is the result in salivary glands, intestine, fat bodies, Malpighian tubules and circular glands; and polyploidy in the ovary, pericardium cells and nerve cells [316]. These processes do not lead to equal multiplication of all parts of the DNA; under-replication of specific satellite DNAs is found, for example, in polytene and polyploid cells of *Drosophila* [268].

There are basically two possibilities for ensuring large amounts of specific gene products at particular time-points in development; these are accumulation of the products and gene amplification. Stable mRNAs accumulated over a long period are to be found, for example, for globins in the reticulocytes or for ovalbumin in the oviduct cells. The large egg cells of the clawed frog, *Xenopus laevis*, require about 10^{12} ribosomes and corresponding amounts of tRNA for the protein synthesis that is initiated by fertilization. The immature oocytes contain reserves of 5S rRNA and tRNA in 7S and 42S nucleoprotein particles. In the smaller 7S particles, about half of the total 5S rRNA is assiociated with one protein species, and in the large 42S particles, the other half of the 5S rRNA and almost the total tRNA is associated with two further proteins. The protein of the 7S particle is identical to the transcription factor TFIIIA that is necessary for the transcription of the 5S rRNA. Of the two proteins in the 42S particle, the smaller one (43 kDa) binds to 5S rRNA; the larger one (48 kDa), on the other hand, is structurally and functionally very similar to the translation factor EF-1α [457]. The suddenly increased requirement for other rRNAs (28S, 18S and 5.8S rRNA) at the onset of development is satisfied by the existence of about 1 million extra copies of rRNA genes produced by gene amplification. This is in addition to the usual diploid complement of 1000, which exist in the form of extrachromosomal circular DNA molecules; each of these molecules carries 15–100 rRNA genes. Amplification of the rRNA genes is also found in other amphibians, fish, many insects, and to a lesser extent in other animals [420].

A different amplification mechanism occurs in the ovary follicle cells of *Drosophila;* within about 5 h these produce a secondary egg membrane (chorion) from a series of specific proteins. The chorion protein genes lie in two clusters on different chromosomes. The cluster on the X chromosome is amplified about 20 times and is expressed earlier than the 60- to 80-fold amplified cluster on chromosome 3. The region of DNA around the chorion protein locus is amplified by multiple replication events, starting at the same point, such that many parallel DNA helices are produced (Fig. 2.6). By the use of drugs, mutants with amplification of specific genes can be isolated from cell cultures of mammals or *Drosophila*. Up to now, about 20 such cases are known, e.g. the amplification of dihydrofolate-reducing

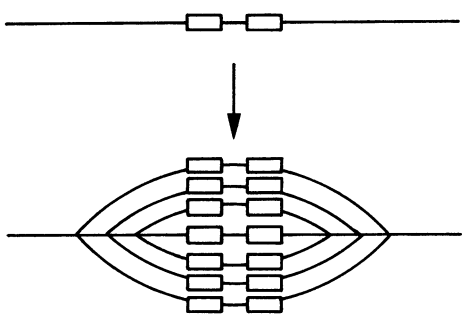


Fig. 2.6. Gene amplification via the "onion-skin" mechanism in the gene for the chorion protein of *Drosophila melanogaster* [418]

enzymes in mammalian cells or *Leishmania* through the use of methotrexate, or DNA polymerase α in *Drosophila* through use of aphidicoline. The frequencies of such mutations are around 10^{-4} to 10^{-7} per generation; the amplified DNA region of up to 1000 kb is usually much larger than the gene itself, and tandem clusters or extrachromosomal copies may be produced [193, 420, 483]. The first indication of a comparable phenomenon outside of the laboratory was from the mosquito *Culex quinquefasciatus*, where the organophosphorus-resistant strain Tem-R has 250 times more copies of the esterase I gene than does the sensitive strain S-Lab [310].

All ciliates show a typical nucleus dualism. The polyploid macronucleus has somatic functions and indulges in active RNA synthesis in vegetative cells; the diploid micronucleus has only generative functions and, after conjugation, it guides the production of the new macronucleus in complex division and recombination processes. Formation of the macronucleus occurs in three phases. In the first phase, the micronuclear chromosomes are converted into polytene chromosomes by multiple replications. In the following phase, the chromosomes are enclosed in vesicles and divided into many pieces, during which not only spacers between the genes but also internal sequences (IES) are eliminated. Repetitive elements are found amongst the IES; thus, the Tec1 element occurs in the micronucleus of *Euplotes crassus* in about 10^4 copies but is completely absent from the macronucleus. In the third phase, the linear DNA fragments are replicated several times and given telomeres; the macronucleus is then completed by reformation of the nuclear membrane. The macronuclear, linear DNA molecules of the hypotrichous ciliates *Euplotes, Stylonychia* and *Oxytricha* have lengths of from 400 bp to about 20 kb, and most likely correspond to single genes; according to the species, each DNA molecule is present in 1000–15 000 copies. Formation of macronuclei here leads to the elimination of tens of thousands of DNA elements with lengths of 14 to more than 500 bp. The macronuclei of holotrichous ciliates, such as *Paramecium* or *Tetrahymena*, contain only about 3000 types of linear DNA molecule with lengths of 20 to more than 100 kb. During the formation of the macronucleus of *Tetrahymena thermophila*, only about 5000–7000 DNA fragments are eliminated; however, these are longer than those in the hypotrichous species [15, 210, 431].

Chromatin diminution is the elimination of chromatin during the first somatic cell divisions such that all somatic cells contain less DNA than do the germline cells. This was discovered by von Boveri in 1887 in the roundworm *Parascaris equorum* and has since been detected in many nematodes as well as in several insects, crustaceans (Copepoda) and ciliates [431]. About one-quarter of the chromatin of *Ascaris lumbricoides* is eliminated. The eliminated DNA consists mainly of AT-rich, highly repetitive satellite sequences with repeat lengths of 121–123 bp which are combined in many variations in long tandems. There is no indication that this DNA is at any time transcribed [128]. Also, many transposable elements are eliminated at a higher than expected rate e.g. Tc1 in *Caenorhabditis elegans* and Tas-2 in *Ascaris lumbricoides* [9, 370].

2.5 Chromatin Proteins

The chromosomal DNA of a diploid mammalian cell has a total length of about 2 m. This DNA must be packed into a volume of a few cubic micrometres in the cell nucleus and still remain accessible to polymerases and other macromolecular effectors. This is made possible by the supramolecular structure of the chromatin in which the DNA is associated with various proteins. The largest fraction of these proteins consists of the very basic histones, which are present in an amount corresponding to that of the DNA. In many sperm, whose chromatin is particularly strongly condensed, the histones are partly or completely substituted by another type of basic protein, protamine. Histone is also missing from the dinoflagellates. In addition to the histones or protamines, the chromatin contains a heterogeneous mixture of other proteins, the non-histone proteins,

Table 2.3. Average length of DNA per nucleosome [5]

Species	Tissue	Length (bp)
Rabbit	Brain cortex	162
Man	Hela cells	188
Rat	Foetal liver	196
Rat	Cerebellum	200
Chicken	Erythrocytes	212

in amounts equivalent to one-tenth of the histone content [197]. The elementary units of the chromatin are the nucleosomes, which can be released from the chromatin by partial digestion with nuclease. These are made up of "core particles" that are about 11 nm in diameter and bound together with "linkers" of various lengths. The core particle is an octomer of two each of the histone molecules H2A, H2B, H3 and H4; around this is wound a 146-bp DNA fragment in two almost complete turns. A molecule of the fifth histone type, H1, is associated with the linker DNA. The length of linker DNA varies between about 20 and 80 bp, and the distance between the core particles shows species- and tissue-specific differences (Table 2.3). This "string-of-pearls" structure makes up chromatin threads, with a diameter of 30 nm, that are usually folded into loops about 0.4 μm in length and with approximately 20–80 kb of DNA. The molecular architecture of these chromatin structures, as well as the processes of their formation and alteration, are only incompletely understood. During nuclear division, the chromatin is condensed in the chromosomes; in interphase, it is more or less evenly dispersed throughout the whole nuclear volume. The heterochromatin remaining condensed during interphase is particularly rich in highly repetitive (satellite) DNA and appears, at the level of the nucleosome, to have a very similar organization to that of the euchromatin described above [197].

2.5.1 Structure and Evolution of the Histones

All histones are rich in the basic amino acids: H1, H2A and H2B predominantly in lysine, and H3 and H4 in arginine. H1 is so basic that it is soluble in 50 % perchloric acid. The core histones have lengths of 102 (H4), 121–129 (H2A, H2B), 135 (H3), and over 140 amino acids (sperm-specific H2B from sea urchins). They are clearly subdivided into an N-terminal region with many basic amino acids which is involved in DNA binding, and a C-terminal region, which is responsible for histone-histone binding and contains a large fraction of non-polar amino acids. The function of the linker histone H1 is not yet clear. It may determine the distance separating the nucleosomes, or it may be involved in the establishment of higher levels of chromatin structure; possible functions in the regulation of gene expression are also under discussion [485]. Most H1 histones have a length of between 189 (chicken erythrocyte H5) and 255 (*Drosophila melanogaster*) amino acids [234, 314]. The H1 molecule is made up of three domains, of which the N-terminus (27–44 amino acids) and the C-terminus (90–97 amino acids) are rich in basic amino acids and interact with the DNA; in contrast, the central region (72–76 amino acids) displays characters of a typical globular protein [485]. In *Platyneris dumerilii* there are two very small sperm-specific H1 histones with only 121 and 119 amino acids; the N- and C-termini of these are much shorter, whilst the 80-amino-acid central region has about the normal length [234]. A very unusual H1 is found in the macronucleus of the ciliate *Tetrahymena*. This H1, with only 163 amino acids, is also relatively small but, above all, it lacks the usual domain structure and the 62 basic amino acids are evenly distributed over the whole sequence. If the central region really is involved in the formation of the higher chromatin structure, it would be superfluous in the macronucleus, which divides amitotically without the dramatic changes in chromatin structure seen during mitosis in other cell types. The corresponding gene is also exceptional in that, in contrast to all other histone genes, it contains an intron [484]. The known sequences of histones and histone genes are regularly published in the journal *Nucleic Acids Research*.

The histones H1, H2A and H4 are irreversibly N-terminally acetylated during translation. In histone H4, at least one lysine-ε-amino group is acetylated in the cytoplasm and later deacetylated in the nucleus; this may be necessary for the correct assembly of the histone octomer. The corresponding acetyltransferases are found in the cytoplasm and nuclei of mammalian cells, but have been isolated only from the brine shrimp *Artemia salina* and characterized in detail with respect to specificity and regulatory properties [127]. In several histones, other blocked amino acids are found at the N-terminus; e.g. N-dimethylproline in H2B of the starfish *Asterias rubens*, and N-trimethylalanine in H2B of *Tetrahymena* [286].

The question of the homology of the histones is difficult to answer because of their monotonous amino acid composition; it is apparent, however, that at least histones H2A, H2B and H4 share a common origin. It is controversial whether H3 also belongs to this group. H1 is different to the core histones in almost every respect and is undoubtedly not homologous [486]. The core histones have changed only slowly during evolution, H3 and H4 being amongst the most conserved proteins (see Table 4.12; p. 161). The sequence of H4 is constant throughout the vertebrates, and differs in sea urchins, starfish and the nematode *Caenorhabditis elegans* by only one amino acid, and in higher plants by only two amino acids [451]. The H3 sequence of mammals differs by only one amino acid from that of the rainbow trout, *Salmo gairdneri*, and by only four amino acids from those of *Caenorhabditis elegans* and peas [87, 451]. The most unusual core histones are found in the Ciliophora: the H4 of *Tetrahymena thermophila* shows only 78% similarity with that of yeast, and only 80% similarity with the higher eukaryotes. This provides one of the arguments in support of the idea that the ciliates diverged from the general evolutionary line of the eukaryotes prior to the separation of the fungi, higher plants and animals [201].

Histones H2A and H2B show strong species specificity. Compared with the corresponding histones from calf thymus, the H2A from the rainbow trout differs in 6 positions, from the cephalopod *Sepia officinalis* and the worm *Sipcinculus nudus* in 19 positions, and from *Caenorhabditis elegans* in 23 positions; the H2B from the trout *Salmo trutta* differs in 8 positions, from *Drosophila* in 21 positions and from the limpet *Patella* in 26 positions [87, 233, 451, 479]. H1 is evolutionarily much less conserved than the core proteins and may be represented within a single animal by several cell-specific variants [485]. The C-terminal regions of histones H2A and H2B, in agreement with the requirements for stable protein-protein binding, show much less variation than the N-terminal regions, which must be adapted to different DNA sequences and activities in different cells, chromosomes or chromosome regions. For similar reasons, the central section of the H1 histone is much more highly conserved than are the terminal regions [451, 479, 485]. In addition to substitution and deletion of individual amino acids, the evolution of the histones has also involved the duplication, insertion or deletion of longer sequences, as will be described for several sperm-specific H1 and H2B histones.

2.5.2 The Histone Genes

In completely differentiated cells, the histones are subject to a turnover process, the rate of which is greater for the various H1 types than it is for the core histones. Thus, the half-lives of histones H1° and H1A in mouse kidney cells are 41 and 61 h, those of H2B and H2A are 90 and 135 h, and those of H3 and H4 are as high as 216 and 316 h, respectively [110]. The most intensive histone synthesis occurs more or less simultaneously with DNA replication in the S-phase of the cell cycle. Chromatin doubling requires large quantities of histones, e.g. in mammalian cells about 30 million molecules of each histone type are needed. Hence, all eukaryotes possess multiple genes for the five histone classes, with several hundred copies in sea urchins, 660 copies in the polychaete *Platynereis dumerilii*, 120–140 copies in *Drosophila hydei*, 100–110 copies in *D. melanogaster*, 11 copies in *Caenorhabditis elegans*, 145 copies in the rainbow trout *Salmo gairdneri*, 90 copies in the clawed frog *Xenopus laevis*, 10 copies in the chicken, 10–20 copies in mice, and only 2–3 copies in *Tetrahymena* [86, 240, 364, 391, 443, 485]. The different copy numbers may be explained by the fact that the successive cell divisions in the early embryo development of sea urchins and *Drosophila* follow much more rapidly than in vertebrates. The number of histone genes in the Urodela increases out of proportion to the C value; the crested salamander, *Triturus cristatus*, (C = 23 pg) has 636 H4 genes, and the axolotl, *Ambystoma mexicanum*, (C = 38 pg) has 2685; nevertheless, it is questionable whether all these genes are active [194].

The histone genes of *Tetrahymena* are dispersed throughout the genome, whereas in yeast they are organized into four gene pairs, H2A-H2B and H3-H4 [21]. In contrast, most of the histone genes of *Drosophila* and the histone genes expressed early in sea urchin embryo development are arranged in highly organized clusters that contain one gene of each of the five histone classes in a tandem array. The order of the genes within the clusters is species specific but uniform within each genome; the individual histone genes, however, may be read in different directions (Fig. 2.7). Gene clusters lacking H1 genes are found in the starfish (H2B-H2A-H4-H3) and the polychaete *Platynereis* (H4-H2B-H2A-H3) [88, 391]. The organization of the gene clusters in vertebrates is much less regular: the clusters mostly do not lie in tandem; the number and order of the genes within the cluster of any one animal may

	Length:	Order:
Various sea urchins	6.3-7.2	→ → → → → H1-H4-H2B-H3-H2A
Drosophila melanogaster	5.0	← → ← → ← H3-H4-H2A-H2B-H1
Salmo gairdneri	10.2	→ → → → → H4-H2B-H1-H2A-H3
Notophthalmus viridescens	9.0	→ → ← → → H1-H3-H2B-H2A-H4
Xenopus borealis	8.5	→ → → → → H4-H2A-H2B-H1-H3
		H1-H2B-H2A-H1-H4-H3
Xenopus tropicalis	10.5	H1-H3-H4-H2A-H2B
Xenopus laevis	8.5	→ ← → → → H4-H2A-H2B-H1-H3
	14.0	H4-H3-H2A-H3B +7 others
Gallus domesticus		→ ← ← → ← H4-H2A-H3-H2A-H4

Fig. 2.7. Histone gene repeats in various animals, showing repeat lengths (kb) together with the order and direction of reading of the genes [391, 467]

be different; some clusters contain several copies of specific genes (e.g. in *Xenopus* and chicken), or even lack some genes (Fig. 2.7); the repeat length itself varies within an individual due to differences in the length of the spacers between genes [443, 479]. In summary, it may be said that there was an increase in both number and level of organization of the histone genes during evolution from the lower to the higher eukaryotes, but this was later lost during the evolution of the vertebrates.

Despite the many conservative aspects of histone amino acid sequences, there are many indications of rapid evolutionary change in the organization and sequences of the histone genes. In the vertebrates, the organization can be very different, even in closely related species (e.g. *Xenopus laevis* and *X. borealis*) or in different individuals of the same species (*X. laevis*) [443]. Significant differences in the number and location of histone gene clusters are also found between various species in the genus *Drosophila*. In *D. virilis*, in contrast to *D. melanogaster*, 30–40 % of the clusters lack H1 genes [115]. In both sea urchins and *Drosophila*, some histone gene clusters may lie as "orphons" outside of the tandem array, at other places in the genome. "Horizontal evolution", which would otherwise ensure the similarity of neighbouring genes, is more difficult for orphons and the isolated genes of the vertebrates; the result is greater variability and even the development of pseudogenes [291, 434]. Histone genes coding for identical or very similar amino acid sequences may have drastic differences in nucleotide sequence: thus, although H3 of cattle, mice and rainbow trout differ by only one amino acid, these species have only 81 of 135 codons in common [86]. The amino acid sequence of H4 is identical in all vertebrates, but only 28 of 102 codons agree [479]. The two H4 genes of *Tetrahymena* differ in only 14 of the 104 codons, but the non-coding regions are completely different; the same is true for the H2B genes [201, 327]. Five H4 genes from four individual sea urchins (*Strongylocentrotus purpuratus*) were found to contain 59 variant nucleotides in the spacers, and in two genes there was, in addition, an inserted 195-bp element that is also found at other gene locations of this species. In the species *S. purpuratus* and *S. droebachiensis*, which separated 4–6 million years ago, the spacers between the H2B and H3 genes differ in 11.2 % of the nucleotides. The intraspecific variability is two- to fourfold less. This indicates, furthermore, that in the last million years there has been no sequence equalization between the five H4 genes through horizontal evolution [489].

Sea urchins like *S. purpuratus* and *Lytechinus pictus* have several histone gene families that are expressed in a development and tissue-specific manner; best known are the "early" and "late" genes of the embryo. The families of the early histone genes consist of many hundreds of clusters that, on the basis of minor sequence differences in the spacer, may be divided into subfamilies. The mRNAs of the early genes exist already in the unfertilized egg; their transcription increases during cleavage to reach a maximum in the early blastula. From the gastrula onwards, the early genes are replaced by the late type, which include

only 5–15 clusters spread throughout the genome. The nucleotide sequences of the early and late histone genes differ in the coding regions by about 20 % and in the non-coding regions by much more than this [291]. The large sequence differences suggest that the separation of the two types occurred a long time ago and is general for the sea urchins.

2.5.3 Variability of the Histones

Non-allelic variants (subtypes) of histones are now known in all histone classes except H4. Even H3, despite its conservative character in the mammals, is found as three somatic variants (H3.1, H3.2 and H3.3) and a sperm-specific variant (TH3). The differentiation into H3.1 and H3.2, where the only difference is Cys or Ser at position 96, occurs only in mammals; the classes H3.1 and H3.3, on the other hand, are found in birds and other vertebrates. The H3.3 gene contains introns and probably represents far more the original H3 gene; the intronless genes of the H3.1 class, however, possibly arose before the separation of plants and animals. Both H3.1 and H3.3 show no species-specific differences between humans and chickens but differ from one another at four positions. Histones of class H3.1 are only synthesized in vertebrates during preparation for nuclear division, but class 3.3 is synthesized throughout the whole cell cycle. The H3.1 class also resembles the early and late H3 of the sea urchin [471]. Subtypes of the core histone H2A are more than likely to be found in all vertebrates, but they also occur in the macronucleus of the ciliate *Tetrahymena*, in *Caenorhabditis elegans*, and in *Drosophila melanogaster* [98, 451]. Whereas the species-specific differences in histones H2A and H2B are usually localized in the N-terminal region, the variants found in the vertebrates also show differences in the C-terminal region; this probably results in changes in histone conformation and in structural and functional properties of the whole nucleosome. The existence of different variants of core histones thus facilitates heterogeneity amongst the nucleosomes, and this may be important for the variable expression of different DNA regions.

The linker histone H1 is the most variable of all histones, and the various subtypes appear to be correlated with metabolic activity and the degree of condensation of the chromatin. Up to seven H1 subtypes are found in somatic cells of vertebrates and many invertebrates. Immunohistochemical investigations have shown that different chromatin regions in the giant chromosomes of the midge *Chironomus thummi* have different spectra of H1 subtypes [303]. The chicken is the only animal in which a complete set of six H1 sequences of 217–224 amino acids has been determined [84]. Two variants of the particularly lysine-rich H1° are to be found in all mammalian organs; similar H1 variants are also present in all other vertebrate classes and even in the mussel *Anodonta cygnea* [300]. The proportion of the H1° type is not only cell-specific but also significantly higher in active regions of the genome: it amounts to 12 % in the (inactive) globin region of the liver cell genome of the mouse, 15–19 % in the satellite DNA, and 61 % in the region of the albumin gene [105]. Extremely aberrant histone genes have been found in the duck *Caraina moschata;* the coding sequence of H3 differs from that of the chicken in 10 of 135 amino acids, and of H1 in 22 of 217–218 amino acids [439].

Unique histone variants are present in the strongly condensed chromatin of nucleated erythrocytes and germline cells, and also in the transcriptionally inactive micronucleus of the Ciliophora. In bird erythrocytes, which retain their nuclei after differentiation, an H1 variant, H5, appears simultaneously with their genetic inactivation. H5 shows significant similarity to H1° in its central region and is also related to repression of DNA synthesis. The proportion of erythrocyte-specific histone increases after the end of DNA synthesis [177]. Specific, as well as non-specific, H1 variants may also be detected in the nucleated erythrocytes of many fish, amphibians and reptiles, but do not always correspond electrophoretically with the H5 of bird erythrocytes. A typical H5 predominates in the teleosts *Salmo gairdneri, Perca flavescens* and *Pomoxis nigromaculatus; Cyprinus carpio* has little H5 but a lot of H1; and *Catostomus commersoni* has no H5. The erythrocytes of *Xenopus laevis* also contain both H1 and H5; on the other hand, typical H5 was missing from the red blood cells of *Rana catesbeiana* and all reptiles that were examined [371].

The chromatin of mature male gametes is not transcribed; it is highly condensed and thus the genetic information is protected until the time of fertilization. Sperm-specific histone variants are responsible for the compact structure, as is protamine, the alternative type of nuclear protein. During spermatogenesis in rats, there are dramatic changes in the spectrum of nuclear proteins: already in the spermatogonium there is a large increase in the somatic forms H1a and

H2A.X and the appearance of the testis-specific H3 variant, TH3; the sperm-specific histones H1t, TH2A and TH2B are produced during meiosis. In the intermediate spermatid stages, all histone and non-histone proteins are replaced by the transitional proteins TP-1 and TP-2, which are no longer organized into nucleosomes. The protamines appear 2–3 days later; there is one form in the rat and several forms in other mammals, and these persist into the mature sperms [165, 379]. Similar alterations in nuclear proteins are seen during spermatogenesis in other vertebrates; thus, in the dogfish, *Scyliorhinus caniculus*, for example, two transitional proteins are followed by four protamines [32].

In sea urchin sperm there are no protamines but late H4 and H3, particular H2A subtypes and sperm-specific H1 and H2B variants. The latter are interesting in that the N-terminus is extended by 20–26 amino acids with highly basic, repetitive tetrapeptides (H1) or pentapeptides (H2B); the many positive charges neutralize the negative charges on the DNA and thus allow dense packing of the chromatin [348]. The sequence of the repeats shows species-specific variation [246]. H2B variants with N-terminal extensions appear to be limited to sea urchin sperm; the H2B from the sperm of the starfish, *Asterias rubens*, consists of only 121 amino acids without N-terminal repeats [283]; and the histones in the sperm of the holothurians are also not significantly different from these of the somatic cells [492]. The sperms of the polychaete *Platynereis dumerilii* contain the same core histones as other cells, but they have four specific H1 variants and two protamines. Two of the H1 variants are unusually small, with only 119 and 121 amino acids. Only one H1 and one protamine are found in *Chaetopterus variopedatus* [234]. The micronucleus of the Ciliophora contains genetically inactive chromatin that is, nevertheless, organized into nucleosomes. In *Tetrahymena* this includes, in addition to the histone H3S that is also present in the macronucleus, a special H3F and three polypeptides α–γ, which are not H1 variants, in the linker region [484]. The unusual structure of the H1 in the macronucleus has already been mentioned (p. 29); in *Tetrahymena*, the macronucleus also contains a specific H2A variant [474].

The variability of the histones is also increased by post-translational modification, such as phosphorylation, acetylation, methylation and covalent binding to poly-ADP-ribose. The significance of these processes is to be found in the fine adaptation of chromatin structure to changing internal and external conditions. uH2A is a particularly interesting post-translationally altered histone that is a conjugate of H2A with ubiquitin. This polypeptide, with a length of 76 amino acids (see Fig. 3.5; p. 88), occurs in all organisms from bacteria to mammals and apparently functions as a signal for protein degradation. In uH2A, the C-terminal glycine of ubiquitin is attached via an isopeptide bond to the ε-amino group of the H2A lysine-119, and the conjugated molecule has a forked "Y" configuration that is very unusual for a protein. About 10 % of H2A is found in this form. uH2A should probably not be considered an intermediate of histone degradation, and ubiquitin apparently has more than one function in the cell [320].

2.5.4 Protamines

The protamines are small, extremely basic nuclear proteins with 30–70 % arginine and were first discovered in mature fish sperm, although they are also be found in the sperm of birds, mammals, and several invertebrates. As has already been mentioned, protamines are first produced during sperm maturation and they replace the pre-existing histones and transition proteins. Fish protamines are known after the genus or species name: "scyliorhinin" from the dogfish, *Scyliorhinus canicula;* "clupein" from the herring, *Clupea harengus;* "salmin" from salmon of the genera *Salmo* and *Oncorhynchus;* "iridin" from the trout *Salmo irideus*, etc. The sperm of any one species contains several protamines which are designated, for example, as iridin Ia, Ib and II. Protamines have been sequenced from more than 50 species of cartilaginous and bony fish (Fig. 2.8). Due to their unusual amino acid composition, the protamine sequences are difficult to compare; analysis of their relationships and evolution is best carried out using the gene sequences [332].

The teleost protamines are, without exception, very small polypeptides of less than 35 amino acids, in which arginine predominates and cysteine is absent. In the particularly well-investigated rainbow trout, *Salmo gairdneri*, and in other teleosts, the protamines show pronounced heterogeneity; six different protamines, albeit with quite similar sequences, can be isolated from a single trout testis (Fig. 2.8) [295]. Protamines are not found in all teleosts; they are lacking, for example, in carp and in the winter flounder, *Pseudopleuronectes americanus*. The discovery that a protamine gene of the rainbow trout, *Salmo*

```
                10        20        30        40
Sc  MKCGrKrrrrrrHACKrKKrACKQrSSTIVrAHLVHrrrAArrCPr

                10        20        30
Sg  Prrrr-SSSrPVrrrrrPrVSrrrrrrGGrrrr
El  PrrrrASSSrPVrrrrrPKMSrrrrr-GGrrrr

                10        20        30        40        50        60
Gd  ArYrrSrTrSrSPrSrrrrrrSGrrrSPrrrrrYGSArrSrrSVGGrrrrYGSrrrrrrrY

                10        20        30        40        50
Bt  ArYrCCLTHSGSrCrrrrrrrCrrrrrrFGrrrrrrVCCrrYTVIrCTrQ
Ss  ArYrCCrSHSrSrCrPrrrrCrrrrrrCCPrrrrrAVCCrrYTVIrCrrC
Hs  ArYrCCrSQSrSrYYrQrQrSrrrrrrSCQTrrrAMrCCrPrYrPrCrrH
```

Fig. 2.8. The sequences of different protamines. *Sc,* Z2 from the shark *Scyliorhinus canicula* [285]; *Sg,* C2b from the rainbow trout *Salmo gairdneri* [295]; *El,* the protamine from the pike *Esox lucius* [413]; *Gd,* "gallin" from the chicken *Gallus domesticus* [332]; *Bt,* the protamine from the cow *Bos taurus* [292]; *Ss,* the protamine from the pig *Sus scrofa* [292]; *Hs,* human protamine P1 [296]. The protamines of both teleosts, the two mammals and man are arranged to show the maximum sequence similarity. For the sake of clarity, arginine is represented as *r*

gairdneri, is flanked by long terminal repeats, just like a viral retroposon, leads to the idea that the uneven distribution of protamines amongst the bony fish is the result of gene transfer between different species [212].

The protamines of the Elasmobranchii often deviate quite markedly from those of the teleosts. In *Scyliorhinus canicula*, the sperm contains four protamines (Z1, Z2, Z3 and S4) that vary in length from 27 (Z3) to 45 amino acids. Only Z3 corresponds in its composition to the typical picture of the arginine-rich teleost protamines; the others contain many cysteines, in addition to the arginines, and are associated through disulphide bridges. Sequence comparisons of protamines are difficult to undertake due to their odd amino acid composition; in any case, there are no indications of any relationship between the protamines of the dogfish and the teleosts [32, 70].

Protamines occur in the sperm of all birds and mammals. The protamines of mammals have 50–60 amino acids, those of the quail, *Coturnix japonica*, have 56, and "gallin" from the domestic fowl has 61 amino acids (Fig. 2.8); these are all significantly larger than those of the fish [332]. The protamines of higher mammals contain 8–18 % cysteine in addition to 47–61 % arginine, and are so cross-linked by disulphide bridges that their isolation and characterization is difficult. In contrast, the protamines of the marsupials are cysteine-free [134]. The sperms of most mammals contain only one protamine species; until recently, two protamines (P1 and P2) were identified in only man and the mouse. The other mammalian species do actually possess two protamine genes but the P2 gene has suffered many mutations and is only weakly expressed [279]. The protamine-like proteins from the sperm of invertebrates have not yet been systematically investigated. However, several proteins from mussels, snails, cephalopods and holothurians have been sequenced. Some of these, in contrast to the vertebrate protamines, contain large amounts of lysine in addition to arginine [14, 223, 288, 349].

2.5.5 Non-Histone Proteins

In addition to histones, the chromatin of all eukaryotes contains a heterogeneous mixture of other proteins that are equivalent to about one-tenth the quantity of the histones. These include, for example, RNA polymerases and other enzymes, proteins involved in the binding and maturation of transcripts, and regulatory proteins. The extraction of clean nuclei preparations releases non-histone proteins not only from the chromatin but also from the microfibrils of the nuclear matrix to which the chromatin is bound, and from the nucleoplasm, the nuclear membrane, the nucleoli and the heterogenous nuclear ribonucleoprotein (hnRNP) particles. The non-histone proteins are divided into two groups according to their electrophoretic behaviour: a "high mobility group" (HMG) and a "low mobility group" (LMG). The HMG has been more extensively investigated. The HMG proteins have an unusual composition of one-quarter each of acidic and basic amino acids. Free ubiquitin, as HMG-20, is also included in this category. One-dimensional elec-

trophoresis of HMG proteins routinely yields 20–40 bands, and 2-D electrophoresis of Hela cell preparations results in more than 450 HMG fractions, although in this case proteolytic artefacts cannot be excluded. The nuclear protein spectrum of invertebrates, such as the Mediterranean fruit fly *Ceratitis capitata* or the nematode *Caenorhabditis elegans*, have a comparably high degree of complexity [57, 450]. In the absence of information on the amino acid sequences, direct comparisons of the non-histone proteins of vertebrates and invertebrates are impossible; however it is, remarkable that antibodies against HMG-14 and HMG-17 of calf thymus show cross-reactivity with the chromosomes of *Drosophila*.

In calf thymus, the largest fraction of the HMG proteins is made up of the proteins numbered 1, 2, 14 and 17, the sequences of which are mostly already known. Each nucleosome carries two binding sites for HMG-14 or HMG-17. The sequences of these two HMGs agree, for example, in the chicken at only 44 % of positions. Compared with HMG-17, HMG-14 shows a much higher rate of evolution: the human HMG-14 (98 amino acids) and that of the chicken (104 amino acids) show only 51 % agreement, compared with 98 % agreement in the HMG-17 proteins [112]. HMG-1 and HMG-2 from mammalian thymus are identical in 79 % of their over 200 amino acids [395]. The larger HMG-1 and HMG-2 of calf thymus correspond to HMG-T1 and HMG-T2 of the rainbow trout *Salmo gairdneri*, and the smaller HMG-14 and HMG-17 correspond to the HMG-D and HMG-C; in addition, there are further HMGs in fish that are not readily comparable [51]. HMG-T1 and HMG-T2 of the trout are each encoded by four genes, whereas the human HMG-14 and HMG-17 proteins are represented by two multi-gene families, which in the case of HMG-17 has 40–50 members [250]. The nucleoli of all eukaryotes contain nucleolin, a multifunctional, 100-kDa protein that plays a key role in rRNA maturation and the biogenesis of ribosomes. The sequence of 713 amino acids, determined via cDNA, shows no similarities to other known proteins [251].

2.6 DNA Replication and Repair

DNA replication and repair in the eukaryotes has been nowhere near as well investigated as in the prokaryotes; the sparse comparative biochemical data that are available are restricted mainly to the DNA polymerases, which in fact play a central role in these processes. These enzymes link the free 3′-OH group of an RNA or DNA primer to the 5′-phosphate of a deoxynucleotide triphosphate. In the reactions of the polymerases, the enzyme complex is either translocated to the next nucleotide after each polymerization step (processive) or released (distributive). During the replication of the chromosomal double helix, one strand (the leading strand) is synthesized continuously, whereas the other (the lagging strand) is synthesized only in sections (Okazaki fragments). Thus, two different polymerase activities are required for DNA replication: a highly processive one for producing the leading strand, and one less tightly bound polymerase that is released after the replication of each Okazaki fragment. Five types of DNA polymerase (α–ε) are found in the cells of mammals and other eumetazoans. Polymerases α and δ take part in the replication of chromosomal DNA, with the δ-polymerase probably synthesizing the leading strand and the α-polymerase the lagging strand [20, 407, 427]. With only 10^{-9} to 10^{-12} errors per nucleotide per generation, the reliability of DNA replication is very high; in addition to the specificity of base-pairing, this accuracy can be attributed to "proof-reading", i.e. the excision of wrong nucleotides by a polymerase-associated 3′-5′-exonuclease activity [245]. The γ-polymerase is responsible for the replication of mitochondrial DNA, and the β-polymerase for the repair of chromosomal DNA [478].

The α-polymerases are characterized by their high molecular weight and by sensitivity to the inhibitor N-ethylmaleimide (NEM); their function is the replication of chromosomal DNA. The RNA primer required for the polymerase reaction is synthesized by a primase that is always found tightly associated with eukaryote α-polymerases. The best-known animal α-polymerase primase is that of *Drosophila melanogaster*. This is made up of four subunits, α–δ, of 182, 73, 60 and 50 kDa respectively; polymerase activity is associated with the 182-kDa subunit, primase activity with the 60- and 50-kDa subunits; and no catalytic activity is known for the 73-kDa subunit [92, 93].

The α-polymerases (replicases) of mammals have molecular masses of up to 1300 kDa and a highly complex structure that is not yet known in detail [160, 224]. The α-polymerase of mammals has no exonuclease activity but, in contrast, the enzyme of *Drosophila melanogaster* exhibits "cryptic" 3′-5′-exonuclease activity. The δ-

polymerase is always associated with an exonuclease. Whilst the α-polymerase incorporates less than 100 nucleotides before being released from the substrate, the δ-polymerase has an in vivo processive capacity of at least 4500 nucleotides. This is related to the effect of an accompanying 36-kDa protein that is only produced in the S-phase of the cell cycle and is known as proliferating cell nuclear antigen (PCNA) or cyclin. In the absence of PCNA/cyclin, the δ-polymerase is released from its substrate after only ten incorporation steps [407].

Many polymerase complexes are simultaneously active on the chromosomal DNA, with the gap between them (the replicon length) varying from 5 (in embryos) to 350 μm (in spermatogenesis). All new nucleosomes are apparently taken up on the same daughter strand. In the Protozoa and in the lower eukaryotes there are always multiple DNA polymerases available but these are only partially comparable with those of the Eumetazoa. They have, without exception, higher molecular masses, and in some cases display exonuclease activity [66, 146].

The β-polymerases responsible for DNA repair have no exonuclease activity and only a limited accuracy; the error rate is 1:1000. Their structure is much simpler than that of the other polymerases. The enzymes of man and the mouse are 39-kDa monomeric polypeptides with 335 amino acids [478]. Like the α-polymerases, these enzymes also appear to have changed only very slowly during evolution. The β-polymerases of different mammals are very similar immunologically and in their tryptic peptide patterns; antisera against the rat enzyme react with the corresponding bird and fish enzymes but not with those of the trypanosomes [64, 429]. Because of its much larger, 110-kDa subunits, the *Drosophila* β-polymerase belongs to a completely different category [375]. The γ-polymerase is responsible for replication of the mitochondrial DNA and, as it represents only about 1% of the total DNA polymerase activity of the cell, is known less accurately; in pure mitochondrial preparations, only the γ-polymerase is found. However, the enzyme is also found in cell nuclei, where it is apparently involved in virus replication and possibly in DNA recombination processes. γ-polymerase is characterized by a high molecular mass (150–300 kDa), NEM sensitivity and the exclusive use of DNA primers. The *Drosophila* enzyme is a heterodimer of 125- and 35-kDa subunits. The enzyme of the clawed frog, *Xenopus laevis*, has a catalytic subunit of 140 kDa and is associated, like the chicken enzyme, with a 3′-5′-exonuclease [206, 472].

Only rather sketchy information is available about the other enzymes and protein factors associated with the DNA polymerases of animals [59], enzymes such as those required for unwinding the DNA and stabilizing the single strands, for the synthesis of RNA primers, and nucleases and ligases. In fact, *Drosophila* is the only animal from which not only all three DNA polymerases (α–γ) but also primase, ligase and both topoisomerases have been isolated. Ligases create phosphodiester bonds between the free 5′-phosphate groups and the free 3′-hydroxyl groups of broken DNA strands. The ligase in the unfertilized eggs of *Drosophila melanogaster* has the highest activity; this enzyme consists of only one 80- to 83-kDa polypeptide [354, 355]. The topoisomerases catalyse the reversible interconversion of different DNA isoforms, e.g. supercoiling and relaxation, or the reversible chain-like linking of circular DNA (catenation and decatenation). The enzymes play important, but incompletely understood, roles in replication and transcription. There are two very different forms of topoisomerases [466]. The type I nicking-closing enzymes break one strand of the double helix and pull the other strand through the gap in such a way that the energy of the phosphodiester bond is conserved, and no further energy supply is required for the reannealing process. Such enzymes have been characterized in detail in various mammals, the chicken, *Xenopus, Drosophila* and *Trypanosoma cruzi*. The type II topoisomerases open and close both strands simultaneously and require ATP. These enzymes are also well described for mammals, *Xenopus* and *Drosophila* [356, 359, 362, 466, 487].

Spontaneous alterations in DNA constituents (point mutations) are of frequent occurrence in all organisms; in man, for example, about 5000 purine bases per day are released from their N-glycosidic bonds (depurination), and 100 cytosine residues are deaminated to uracil. A system of repair enzymes deals with these and other forms of damage; examples include DNA glycosylases, exonucleases, DNA polymerases and ligases. However, little is known about these from a comparative biochemical point of view [248, 376]. RNA-dependent DNA polymerases (reverse transcriptases) are required for the transposition of retrovirus-like DNA elements. High activities of a reverse transcriptase have been found in the ciliate *Paramecium teraurelia* but the biological significance of this is not clear [226].

2.7 Transcription and RNA Maturation

2.7.1 RNA Polymerases

The DNA-dependent RNA polymerases synthesize complementary RNA on the DNA matrix by the creation of phosphodiester bonds between the free 3′-OH groups of the RNA and the 5′-phosphates of ribonucleoside triphosphates. Whilst in bacteria, mitochondria and chloroplasts there is only one RNA polymerase, in the cell nucleus there are three classes (I–III) of this enzyme. Pol I is responsible for the formation of the pre-rRNA, Pol II for the transcription of protein-coding genes, and Pol III for the synthesis of the tRNA precursors, the 5S RNA and other smaller RNAs. Pol I is localized in the nucleolus and is activated by Mg^{2+}, whereas the other two enzymes are found in the nucleoplasm and require Mn^{2+}. Pol II is inhibited by as little as 10^{-7}M α-amanitin, Pol III by at least 10^{-4}M α-aminitin, and Pol I is completely insensitive. The little-investigated mitochondrial RNA polymerases, like those of the prokaryotes, are specifically inhibited by rifampicin [6, 37].

In spite of different specificities, the three classes coincide in their general functional and structural principles; the catalysed process always follows the order: DNA sequence recognition, binding, initiation, elongation, termination and release of the products. All polymerases have molecular masses between 500 and 700 kDa and are composed of two non-identical, large polypeptides of 120–240 kDa and several others of mostly less than 50 kDa [37, 96, 377]. Comparison of gene sequences in different eukaryotes reveals that the large subunits of the three polymerase types are homologous. Only the large subunit of Pol II carries a C-terminal extension made up of repeated heptapeptides, numbering 26 in yeast, 44 in *Drosophila* and 52 in mammals, with the consensus sequence YSPTSPS. In contrast to all eukaryotes, *Trypanosoma brucei* possesses two Pol II forms that are coded by different genes, differ in only four amino acids and have differently constructed, non-repetitive C-terminal domains. Quite surprisingly, the biosynthesis of the variable surface glycoproteins (VSG) of the trypanosomes is not inhibited by α-amanitin; ongoing studies of the three polymerases have as yet found no explanation for this observation [404].

2.7.2 Transcription

The initiation and termination signals for the three polymerase classes have different locations and structures. Around the initiation point (′) of the Pol I-transcribed rRNA genes (rDNA) there is a sequence that can be represented by the consensus sequence ATRT′A, where R stands for a purine; the 5′ end of the 18S rRNA gene of the silkworm *Bombyx mori* lies 909 bp downstream from the initiation point [142]. Transcription of the rDNA extends several hundred nucleotides beyond the 3′ end of the mature 28S rRNA; e.g. in *Xenopus* by 235 nt and in the mouse by 565 nt [346]. The control region for the Pol III-transcribed genes for the small RNAs lies in the middle of the coding sequence. In the gene for the 5S rRNA, for example, a control protein of 37 kDa is bound at this point and guides the polymerase to its binding site in front of the coding sequence; sequences in front of the coding region are also required for the transcription of the 5S rRNA gene of *Bombyx mori* [464]. Other Pol III-transcribed genes most probably have specific protein factors that correspond to the control protein of the 5S rRNA gene. The termination signal here consists, in the simplest case, of at least TTTT in a GC-rich region; this is found, for example, in the *Xenopus* and mammalian tRNA genes. However, there are also Pol III-transcribed genes without poly(T) [346].

By far the most thoroughly understood are the transcription units of the Pol II-transcribed, protein-coding genes and their signals for the initiation and termination of transcription, the formation of the 3′ end of the mRNA, the excision of the introns and the splicing of the exon ends (Fig. 2.1). The human β-globin genes may be considered as prototypes for Pol II-transcribed genes; the analysis of spontaneous mutations that lead to reduced synthesis of the β-globin chains (β-thalassemia) has contributed greatly to explaining the significance of different regions of the gene sequence [333]. The sequence TATAAA, the TATA box or Hogness-Goldberg box, is found at about –30 in the 5′ region upstream of the coding sequence of all Pol II-transcribed genes. There are, however, genes in which a typical TATA is missing, e.g. the gene for tropomyosin I in *Drosophila melanogaster* [23]. In many genes, there are further, less strongly conserved elements in the region between –130 and –50; mutation of these can influence transcription, e.g. the sequences CCAAT or GGGCG [54]. The histone genes of the ciliate *Tetrahymena*

have a CCAAT sequence but neither a TATA box nor GC elements; here again, it is clear that the ciliates stand apart from the other eukaryotes [52]. The DNA region at which Pol II begins transcription is termed the promoter; the TATA box is an element of the promoter. Pol II requires several transcription factors (TFII) for initiation. In the formation of the Pol II initiation complex, the factors TFIID and TFIIA first bind directly to the TATA box, and then to the factors TFIIB and TFIIE, which interact alternately with each other and with the polymerase. For Pol III there are the cofactors TFIIIA and TFIIIC [149, 342]. Enhanced, regulated gene expression is achieved by the binding of specific regulatory proteins (activators) to a sequence element (the enhancer) that can stimulate the transcription of the relevant gene from a distance of several thousand base pairs. Enhancers are also known for Pol I-transcribed, ribosomal RNA genes (see Fig. 2.15) [313, 381]. In front of the start site of translation at the AUG start codon there is a conserved sequence in the mRNA that probably corresponds to a complementary sequence in the 18S rRNA of the ribosome. In vertebrates, the consensus sequence here is GCCRCCAUG [239], and in *Drosophila* one finds the significantly different sequence C/AAAA/CAUG; thus, in *Drosophila* A is preferred, whereas in vertebrates C is preferred [62a]. In the 3′ region, the strongly conserved consensus sequence AAUAAA is the signal for splitting the primary transcript 10–30 nt downstream during RNA maturation, and for attaching the poly(A) tail. Due to the immediately ensuing processes of RNA maturation in vivo, the termination of Pol II transcription can only be investigated in vitro using purified enzymes. Then it is seen that particular T-rich sequences act as termination signals at which the polymerase is released from the DNA. However, the termination is apparently not very precise [227].

2.7.3 Maturation of the Primary Transcripts (RNA Processing)

Four types of intron can be distinguished on the basis of the splicing mechanisms. For intron types I and II, the splicing process requires no enzyme but is instead catalysed from the RNA itself. Self-splicing introns of group I are found in chromosomal and mitochondrial transcripts of the fungi and in chloroplasts and specific RNAs of the ciliate *Tetrahymena*. Group II introns, which undergo a different version of self-splicing, are to be found together with the group I type in mitochondrial transcripts of fungi and higher plants. The maturation of chromosomal mRNA takes place enzymatically in a protein-RNA complex called the spliceosome. Chromosomal tRNA is

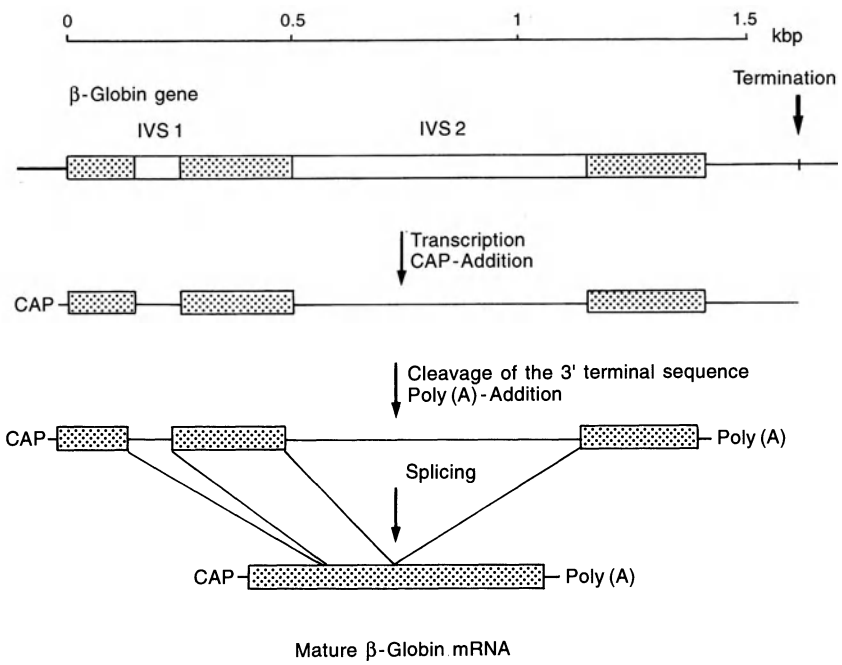


Fig. 2.9. mRNA maturation, as illustrated by β-globin mRNA. *IVS,* intervening sequence or intron

also spliced enzymatically, although the enzymes involved (endonuclease, ligase, etc.) are not organized in a spliceosome [63]. Soon after initiation, the 5′ ends of the primary Pol II transcripts receive the cap sequence m^7GpppXpmY... (Fig. 2.9); these protect the transcript from attack by 5′-exonucleases and bind the finished mRNA onto the large ribosomal subunit. The pre-mRNA is already linked to proteins during transcription to form the "string-of-pearls" type of hnRNP particle. The hnRNP particles contain a set of evolutionarily conserved polypeptides, amongst which are six so-called core proteins of 30–43 kDa [34, 119, 350, 357]. With the attachment of different, smaller nuclear RNAs (snRNAs, p. 52), the hnRNP particle is converted into a "spliceosome", on which the excision of the introns and the ligation of the exons (splicing) takes place [119]. Upon completion of transcription, excess nucleotides are removed from the 3′ end of the Pol II transcript, and a poly(A) sequence of initially 200–300 nt is added; the poly(A) tail of mature mRNA in the cytoplasm is usually shorter. Some mRNAs, e.g. most histone mRNAs, are not polyadenylated [264]. The function of the poly(A) chain is not clear, although it apparently increases the in vivo stability of the mRNA [281].

Accurate splicing requires particular signals at the intron-exon border (junctions), the consensus sequences of which are AG:GUAAGU at the 5′ end of the intron, and $(U/C)_{11}$NCAG:G at the 3′ end [208, 336]. Occasionally, one finds signals deviating from the "GU-AG rule" but which, nevertheless, function perfectly: e.g. GC- instead of GU- is found at the 5′ end of the second intron of the α^D-globin gene of the duck *Cairina moschata* [126]. An example of a modified 3′ signal with -CG instead of -AG is provided by the per gene of *Drosophila* [77]. The splicing process begins with a break at the 5′ end of the intron. The exposed 5′-G is linked by a 2′,5′-phosphodiester bond to an internal A residue near the 3′ end of the intron; this produces a circular structure with a tail (a lariat) (Fig. 2.10). Just how the intron becomes fully released and the exons ligated is not completely clear. The whole process requires ATP. All stages of RNA maturation take place in the nucleus; the finished mRNA is then transported into the cytoplasm by an as yet incompletely understood mechanism, but it is known that it becomes bound to a further set of proteins [119].

In the Kinetoplastida, like *Trypanosoma, Leishmania* or *Crithidia*, mRNA maturation does not involve linking of fragments of the pre-mRNA molecule, but rather the binding of the pre-mRNA to an independent transcript (trans-splicing). Consequently, all mature mRNAs of each species begin with the same 35- to 39-nt sequence, the "spliced leader" (SL) or "mini-exon". The SL sequences differ according to the species but are quite similar, e.g. between *Crithidia fasiculata* and *Trypanosoma brucei* they agree at 27 positions (77 %). In *T. brucei, Leptomonas collosoma* and most likely other Kinetoplastida, the SLs have an unusual terminal cap structure of pm^7G... followed by four $O^{2'}$-methylnucleotides. The SLs originate from the 5′ end of an SL RNA, with a length of 85–140 nt, which is encoded by its own specific gene family. SL or mini-exon genes have lengths of 0.4–1.35 kb and form tandem clusters of 200–250 copies. The primary transcripts of the Kinetoplastida are multicistronic and during maturation are cleaved at specific sequences into individual protein-coding pre-mRNAs; these are further processed at the 5′ end by trans-splicing and at the 3′ end by attachment of a poly(A) chain. The typical polyadenylation signal AAUAAA is most certainly missing

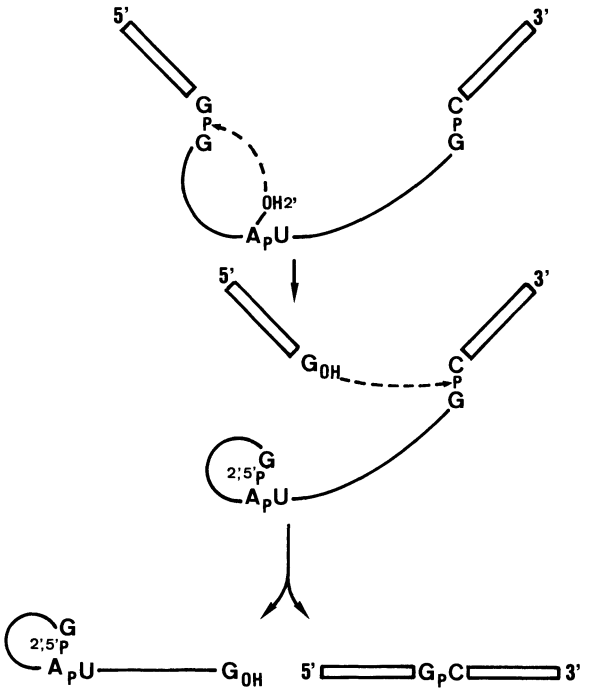


Fig. 2.10. The mechanism of the splicing process, during which an intron in the primary transcript is excised and the breakpoint resealed [394]. In the first step, the GpG bond at the 5' end of the intron is broken and the exposed 5'-terminal G of the intron is linked to an internal A by a 2', 5'-phosphodiester bond. The tailed, circular structure that results is known as a lariat. In the second step, the 3'-terminal OH group of the leading exon is linked to the 5' phosphate of the succeeding exon and the lariat is released

from the Kinetoplastida transcripts [10, 341, 493].

A spliced leader of 22 nt has also been found in 10–20 % of the mRNAs of various free-living and parasitic nematode species. As in the Kinetoplastida, the SL here also stems from an SL RNA coded for by a family of more than 100 genes. Any one pre-mRNA can be spliced either cis- or trans-. Trans-splicing requires a particular type of spliceosome. In contrast to the Kinetoplastida, the SL sequences of all nematode species are identical. However, in *Caenorhabditis elegans* there are two SLs that differ in 7 of the 22 nt and are distributed gene-specifically amongst the mRNAs. Thus, for example, the mRNAs of three out of four glyceraldehyde-3-phosphate dehydrogenase (GAPDH) genes, and many other mRNAs, carry SL1, whereas the fourth GAPDH gene carries SL2. Trans-splicing has also been found in trematodes [10, 101, 117, 161, 205].

2.7.4 Production of Multiple Transcripts of a Gene

There are many exceptions to the rule of "one gene, one polypeptide". Alternative initiation, polyadenylation or splicing signals lead to the formation of multiple mRNAs on a gene, and possibly also to the production of multiple proteins; these alternatives may be regulated in a tissue- or stage-specific manner. Thus, in addition to the diversification of duplicate genes and post-translational protein modification, this gives a further mechanism for the creation of variant proteins which can be adapted to the specific requirements of different cells or developmental stages. Alternative splicing has now been shown for many pre-mRNAs. In many cases, the alternatively spliced exons stem from the duplication of a common precursor; this mechanism for producing multiple proteins apparently arose early in the evolution of the eukaryotes. As yet, very little is known about the regulation mechanisms that determine the splicing processes preferred by different tissues or developmental stages [403].

Examples are known for all the possibilities for alternative splicing. The alternatives may depend upon the retention in the mature mRNA of a particular exon (Fig. 2.11 A), as in the tropomyosin I gene of *Drosophila* or the troponin T gene of the chicken. In many cases, more than one exon is involved in alternative splicing. The random combination of n exons produces 2^n mRNAs. Four variants of the basic myelin protein of the mouse are known to arise from the random combination of two exons (Fig. 2.12 c). In the troponin T gene of the rat, each of the five exons 4–8 can be excised during mRNA maturation, producing 32 tissue-specific and development-dependent isoforms. In many cases, particular exons may alternate but they never lie in tandem (Fig. 2.11 B). This is true for exons 16 and 17 of the 18 exons in the rat troponin T gene; in this way, the number of possible isoforms is increased to 64 [46]. The different C-terminal sequences of the membrane-bound and secreted immunoglobulins arise because the mRNAs are endowed with different exons, following differential splicing of the 3′ end; each has its own poly(A) signal (see Fig. 6.4, p. 224).

Facultative splicing signals within an exon result in differences between mature mRNAs due to the presence or absence of parts of exons (Fig. 2.11 C and D); for example, the fibronectin variants of the rat are produced in this way (Fig. 2.12a). If an intron is not excised (Fig. 2.11 E), then the encoded protein gains an extra domain. This is true, for example, for the γ-fibrinogen molecules of the rat, 10 % of which contain an amino acid sequence coded for by

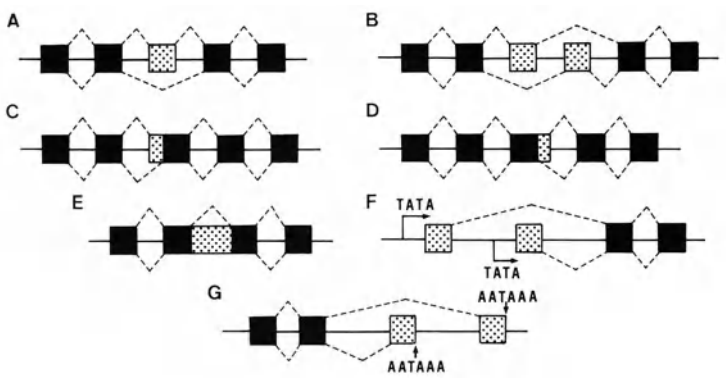


Fig. 2.11 A–G. Different mRNAs and proteins can be produced from the same gene by alternative splicing [46]. **A** Particular exons are spliced or excised; **B** alternating exons are excluded from the finished mRNA; **C** exons include several acceptor sequences; **D** exons include several donor sequences; **E** introns are not excised; **F** several different promoters are available; **G** several different poly(A) signals are available

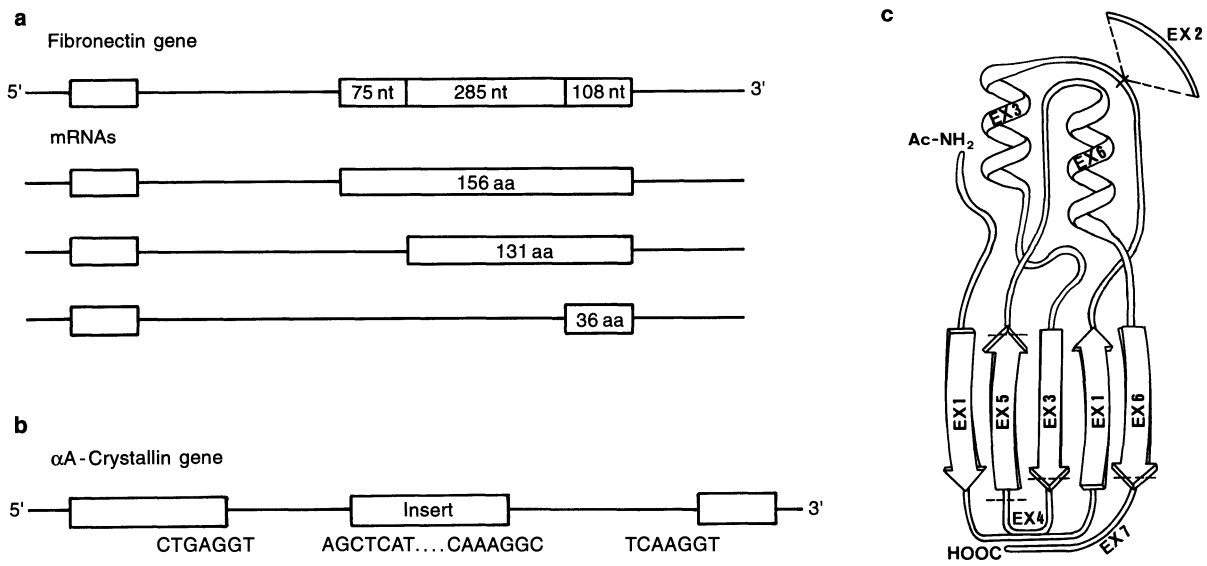


Fig. 2.12 a–c. Examples of alternatively spliced genes. **a** The three fibronection mRNAs in rat liver contain coding sequences of 36 or 131 (36 + 95) or 156 (36 + 95 + 25) amino acids [258]. **b** The intron in the αA-crystallin gene of the rat contains the coding sequence for an insert of 22 amino acids; the variant αA^{ins}, which makes up 10–20 % of the αA-crystallin molecules, arises by alternative splicing [229]. **c** The gene for the basic myelin protein (MBP) of the mouse includes seven exons, the boundaries of which in the amino acid sequence are shown by *dashed lines*. Four different proteins of 21.5, 18.5, 17 or 14 kDa arise by the splicing out of exon 2 and/or exon 6. The 18.5-kDa MBP is illustrated; in the 21.5-kDa MBP, the 26 amino acids coded by exon 2 are included at the site indicated. α-Helices are shown as *spirals*, and β-sheets as *broad arrows* [133]

intron 7. In αA-crystallin of the rat, the inserted sequence corresponds to only a part of the intron (Fig. 2.12b). Alternative promoters or poly(A) signals produce heterogeneity of mRNAs at the 5′ and 3′ ends (Fig. 2.11 F and G).

A particularly complicated situation is found in the "period" (per) gene of *Drosophila*, and this also deserves attention because of its interesting function. This gene is involved in establishing circadian rhythms; in per^o flies the rhythm is missing, and in per^s and per^1 flies it is accelerated or retarded, respectively; transformation of per^o flies with the per^+ gene restores the rhythm [238, 436]. The per gene has eight exons and produces three biologically active transcripts. Transcript A, which is only expressed in the brain and amounts to 70 % of per mRNA, is 4519 nt long and codes for a protein with 1218 amino acids and a mass of 127 kDa. In transcript C, the last three introns are not excised; a large exon of 2800 nt is formed. The C-terminal end of the amino acid sequence coded by this transcript is completely different from that of the A transcript, in that the reading of the intron between exons 5 and 6 shifts the reading frame of exon 6. Finally, in transcript B 288 bp have been removed during splicing; here, however, because the extra intron lies behind the stop codon, the amino acid sequence is unchanged. Transcript A is involved in the establishment of the circadian rhythm; C is possibly concerned with another rhythm, such as that involved in courtship [77].

2.7.5 Regulation of Transcription

Only a portion of the genes are transcribed in any one cell, and the transcription pattern changes during the course of cell differentiation. In the intact cell, regulatory mechanisms can affect the steps in protein biosynthesis that come after transcription: during RNA maturation, during mRNA transport from the nucleus to the cytoplasm, during mRNA degradation, and during translation. A particularly clear picture of tissue-specific and stage-specific differences in gene expression is obtained when the total mRNA of the investigated material is translated in vitro and the translation products are separated by 2-D gel electrophoresis [83]. It is expected that an explanation of the mechanisms controlling the space- and time-dependent patterns of gene

expression will provide decisive insight into the processes of cell differentiation and morphogenesis. The control mechanisms of gene expression, particularly of transcription, are at the centre of one of the most actively researched areas of modern biology. Investigations in invertebrates, especially *Drosophila melanogaster* and the nematode *Caenorhabditis elegans*, have become particularly important; the unique suitability of the latter for such studies is based on the fact that each individual contains the same number of cells and the developmental fate of each of these cells has been described.

The most important mechanism for transcription control is the interaction between regulatory proteins and particular DNA sequences, especially promoters and enhancers. The DNA-binding domains of the regulatory proteins have four different structural motifs [215, 253, 423]: the helix-turn-helix motif, which consists of two α-helices separated by a β turn, was discovered in gene activation and inhibition in prokaryotes, but is also found in vertebrates, e.g. in the form of the determination factors of muscle development [44]. Related structures are the homeodomains that are coded by the homeobox, a DNA element found in many very different eukaryotes. A second DNA-binding motif is the zinc finger, which was discovered in the transcription factor TFIIIA of the clawed frog, *Xenopus laevis*. This contains nine repetitive sequences of about 30 amino acids, each with two terminal cysteine and histidine residues that together bind a zinc molecule, whilst the amino acids in between project as a finger [76]. Many proteins with this motif are now known to exist in a variety of organisms from baker's yeast to man; they function as transcription activators or determination factors. In the vertebrates, the zinc fingers are encoded in multigene families [75, 235]. In *Drosophila melanogaster* the products of, for example, the segmentation genes "Krüppel" (Kr) and "hunchback" (hb) are zinc finger proteins, as are the products of the glass gene that is required for the differentiation of the photoreceptor cells [309]. The third DNA-binding motif, found in the steroid receptors of the vertebrates, is also a zinc finger in which, however, the zinc is bound coordinately by four cysteine residues. To this class belong, for example, the products of the *Drosophila* "knirps" (kni) gene, which is responsible for abdominal segmentation, and the "Seven-up" gene, the absence of which causes the photoreceptor cells R1, R3, R4 and R6 of the eye to assume the characteristics of the R7 type [302, 319]. The fourth motif, the leucine zipper, contains four or five leucine residues, each separated by exactly seven amino acids; as a result, the leucines all lie on the same side of the α-helix. Interaction between these leucine residues stabilizes a heterodimeric, DNA-binding structure such as is found, for example, in the products of the oncogenes jun, fos and myc [215, 423]. Active chromatin regions are always more sensitive to nucleases; the areas in which the DNA is free of nucleosomes, and is therefore easily accessible for regulatory proteins, show a tenfold higher sensitivity and are termed "hypersensitive sites". These areas are rich in enzymes such as topoisomerases I and II, RNA polymerase II, and transcription factors [166].

Still to be explained is the mechanism whereby the binding of a regulatory protein to a DNA sequence can influence the expression of a gene that is 100 bp or even 1000 bp removed. Four mechanisms have been suggested:

1. Two proteins, bound to different sites on the DNA, interact through the formation of a DNA loop ("looping").
2. The binding of one protein so alters the conformation of the DNA that the binding of another protein (e.g. an enzyme) is facilitated ("twisting").
3. The regulatory protein recognizes a specific DNA sequence and then moves from there to another site, at which it initiates transcription, for example, by interaction with the promoter ("sliding").
4. Binding of a protein to one site facilitates protein binding at a neighbouring site and so on until the promoter is reached ("oozing") [352].

Genes that in different cell types are activated at different times should possess several regulatory sequences that interact with different factors. There are various possible patterns of organization:

1. The individual members of a multi-gene family could carry different control elements; the best-known example here is the globin gene family.
2. One gene can possess different control elements for the same initiation signal, as found, for example, in the genes for the yolk protein, the heat-shock protein hsp26, and the "white" and "fushi tarazu" genes of *Drosophila*.
3. One gene can possess two independently regulated initiation signals, as is seen, for example, in the α-amylase gene of the mouse and the gene for alcohol dehydrogenase in *Drosophila melanogaster* [137].

The regulatory elements may be found more than once in a gene; the ovalbumin locus of the chicken, for example, carries many copies of the sequence to which the progesterone receptor can bind.

The protein-coding sequence of about 180 bp, termed a homeobox, was discovered in 1984 as homologous regions in various genes of *Drosophila melanogaster* that are involved in the growth and differentiation of body segments during embryo development; in particular, it involves the genes "fushi tarazu" (ftz), "Antennapedia" (Antp) and "bithorax" (btx). Homeobox sequences have since been identified in many different eukaryotes. The homeodomain encoded by the homeobox has a length of about 60 amino acids (Fig. 2.13) and normally lies close to the C-terminus of the protein. Although, in many cases, mutations of the homeobox lead to disturbances of morphogenesis and cell differentiation [152], the biological role of the homeoprotein lies most likely in its function as a transcription regulator. It can be seen in *Drosophila* that each homeobox is only expressed in specific embryonal cells and, conversely, each cell has a specific pattern of homeoproteins. This may be explained by the fact that any given homeobox gene can influence the expression of others. On the other hand, these genes should affect the expression of non-regulatory structural genes whose products are responsible for the shaping of specific cell types. The homeodomain is apparently responsible for DNA binding, although little is known about the sequence specificity and mechanism of this binding. The remaining nine-tenths of the homeoprotein sequence are not important for DNA binding but may be involved in determining the specificity of the effect through interactions with other transcription factors [185, 263].

Today, homeoboxes have been detected not only in more than 20 *Drosophila* genes but also in other arthropods, annelids, molluscs, nematodes, echinoderms and vertebrates; they are missing, however, from brachiopods, nemerteans and cestodes. The homeoproteins of the vertebrates normally have a length of only 250–300 amino acids,

```
0          10         20         30
ERKRGRTTYT RYQTLELEKE FHFNRYLTRR RRIEIAHALC

40         50         60
LTERQIKIWF QNRRMKWKKE NK
```

Fig. 2.13. The consensus sequence of the homeodomains from 29 species of insects and vertebrates. The homeodomain of the Antp gene of *Drosophila melanogaster* differs from the sequence shown only by Q instead of T at position 6 [151]

whilst the genes Antp and Ubx of *Drosophila* are unusually large and complicated [390]. Six of the homeogenes known in *Drosophila* can be identified in the honey bee, *Apis melifera;* surprisingly, however, the homologue of the segmentation gene "fushi tarazu" is missing [463]. The nematode *Caenorhabditis elegans* possesses about 60 homeogenes and the vertebrates have apparently more than 100, of which about 30 have been sequenced in man and about 26 in the mouse [2, 55, 114, 184, 198, 377, 469]. In extreme cases, the homeodomains in these genes have only about 20 % amino acids in common, although, in contrast, almost identical homeodomains can be found in proteins of man and *Drosophila*. Within a species, the homeodomains may be very different; thus, in *Drosophila melanogaster* a gene has been found that is expressed only in the muscles of the gut, and the homeodomain of this gene coincides with the next most related gene in only 26 of 66 positions [22]. The homeodomains can be divided into 11 families according to their amino acid sequences [469]; the particularly widespread genes of the Antp class are organized in clusters in man and the mouse, as they are also in *Drosophila*, whereas the genes of the engrailed (en) class are dispersed [271].

2.7.6 Heat-Shock Genes and Heat-Shock Proteins

In 1962 the observation was made that a temperature increase from 25 to 37 °C produced drastic alterations within a few minutes in the puff pattern of *Drosophila* polytene chromosomes. It has since been found that sudden increases in temperature cause marked, rapid but transient activation of particular genes in almost all investigated organisms, from bacteria to mammals. These genes encode new mRNAs and proteins that are termed **heat-shock proteins (hsps)** [265, 328]. Amongst the freshwater polyps of the genus *Hydra*, species such as *H. oligactis*, which show extreme sensitivity to heat stress, produce no hsps, in contrast to heat-tolerant species like *H. attenuata* [40]. The optimal temperature for the induction of hsps varies with the species: it is 28 °C in the trout, 35–37 °C in *Drosophila*, and over 40 °C in birds and mammals. Many hsp genes can also be induced by chemical or other stress factors; thus, the hsps apparently have a general protective function and are therefore also called "stress proteins" [218, 382]. The induction of these proteins presents an interesting model for gene regulation.

The hsp spectrum differs with the animal species, but in general is neither organ nor development dependent. However, it was recently observed in *Drosophila* that the complexity of the hsp spectrum increases with age: 10-day-old flies develop 14 hsps and 45-day-old flies develop at least 59 [138]. In general, two main classes of protein, of 68–110 and 15–30 kDa, can be distinguished, although the types hsp70 and hsp83 usually dominate; the large hsp100 and hsp110 of the mammals are not represented in *Drosophila*. Both the coding sequences and the control regions of hsp genes are extremely conservative. The sequence of hsp70 in *Drosophila* shows 73 % agreement with the human protein, 72 % with yeast, and 61 % with the malarial agent *Plasmodium falciparum* [244, 265]. Hsp83 is just as conservative, showing 63 % agreement between *Drosophila* and yeast. In many animals there are several hsp70 genes, often with different levels of inducibility. In *Leishmania major*, the gene family includes four members, in *Caenorhabditis elegans* there are six, in *Drosophila melanogaster* five or six, and in the vertebrates at least two [190, 257, 262]. The heterogeneity of the large hsps may be further increased by post-translational modification. The spectrum of the small hsps varies greatly with the species; there are even differences between different species of *Drosophila*. In *D. melanogaster* one finds in particular hsp28, hsp26, hsp23 and hsp22; *Caenorhabditis elegans* has hsp25, hsp18 and hsp16; and *Xenopus laevis* and several lungless salamanders have only hsp30. The small hsps of a species are all related; in *Drosophila*, for example, they have approximately 50 % identical amino acids. They are not as strongly conserved as the large hsps but, nevertheless, show significant agreement between the insects, nematodes and vertebrates [216, 265, 328].

The hsp70 gene from *Drosophila* is also heat-shock inducible in mouse cells, *Xenopus* oocytes, sea-urchin embryos and yeast cells. The control sequences, the heat-shock elements (HSEs), lie 13–100 bp in front of the initiation site and have the consensus sequence CNNGAANNTTCNNG. The HSE binds a specific protein (heat-shock transcription factor, HSTF) that, together with a further protein, covers about 130 bp of the control region of the hsp70 gene. HSTF is found in normal cells but is without activity; a heat-dependent alteration of the HSTF is apparently required for its induction. Most hsp genes have several HSEs, e.g. hsp70 of *Drosophila* has four, of which only two are sufficient for maximal induction [12]. In contrast, at least three HSEs are required for maximal expression of the *Drosophila* hsp26 [338]. Accumulation of an hsp causes a negative-feedback inhibition of transcription of the heat-shock gene [265, 328]. Using gene technology methods, other genes can be brought under the control of a heat-shock promoter. The hsp27 gene of *Drosophila* can be induced by the moulting hormone ecdysone, as well as by heat shock, using different regulatory elements; whilst the HSEs lie between bp –270 and –370, the binding sites for ecdysone receptors are between bp –579 and –455 [360].

Heat shock in *Drosophila* causes a 100- to 1000-fold increase in transcription of the hsp genes, whilst the transcription of other genes ceases; simultaneously, the translation of already available stable mRNAs stops, although they remain intact and translatable. The hsp mRNAs differ from other *Drosophila* mRNAs in having unusually long "leaders" in front of the coding region. Introns have been found in only two of the many hsp genes sequenced so far: hsp83 from *Drosophila* and a small hsp from *Caenorhabditis*. In contrast to *Drosophila*, there is no difference between the translation of hsp mRNAs and other mRNAs in the mammals; similarly, in the somatic cells of *Xenopus*, only the transcription of the various genes is differentially affected. In the oocytes, however, the hsp genes are usually already transcribed and only their translation is increased following heat shock. It is not known how these differences arose, but they do allow some biological interpretation. *Drosophila* achieves a much quicker heat-shock response through the simultaneous regulation of transcription and translation than is either possible or necessary in the somatic cells of *Xenopus;* on the other hand, the oocytes of *Xenopus* are so large that an increase in the transcription rate alone would require several days to produce an effective heat-shock reaction. Intracellular protein degradation following the binding of proteins to ubiquitin is intensified following heat shock; in fact, ubiquitin behaves as a heat-shock-induced gene in the chicken and also in yeast [265, 328].

2.8 Ribonucleic Acids and Ribonucleoproteins

Approximately 10 % of the total RNA of the cell is present in the nucleus; the remaining 90 % is in

the cytoplasm. The RNA of the cell nucleus consists mainly of the primary transcription products of the three RNA polymerases, i.e. pre-rRNAs, pre-tRNAs and the mRNA precursors; due to their variable length these are known as heterogeneous nuclear RNAs (hnRNAs). These precursors and intermediates of RNA maturation show a very high turnover rate. In addition, a group of smaller, metabolically more stable RNA molecules, the snRNAs, is also found in the nucleus. All the RNAs of the cell nucleus are organized, together with various proteins, into complex ribonucleoprotein (RNP) particles. The nuclear RNAs are, for the most part, degraded in the nucleus; only about 5% of them reach the cytoplasm in the form of ribosomes, tRNAs or mRNAs. The cytoplasmic RNA consists mainly of rRNAs; the tRNAs amount to about 10% and the mRNAs account for just a few percent. The latter are also bound to various proteins in the cytoplasm to form RNP particles, the informosomes. There are, on the one hand, free informosomes with non-translatable mRNA and, on the other hand, the active mRNAs that are bound to ribosomes (polysomes) and form loose associations with the protein factors of translation and also with other cytoplasmic proteins [276, 416]. In the oocytes of amphibians and sea urchins, about 70% of the poly(A)-RNA is bound to transcripts of repetitive DNA sequences; this combined RNA is not translated and disappears at the onset of embryo development. This possibly represents a reserve of immature mRNA [85].

2.8.1 The rRNA Genes and Their Transcription

In all cells there exist organelles of about 20-nm diameter, the ribosomes, on which protein biosynthesis takes place [178]. They are universally constructed from two subunits of different size that only combine with mRNA during protein synthesis. In the ribosomes of *E. coli*, the best-known ribosomes and the prototype for all the rest, the smaller subunit contains an SS-rRNA (small subunit) of 16S with 1540 nt and 21 different proteins, whereas the larger subunit has an LS-rRNA (large subunit) of 23S (2900 nt), a 5S rRNA (120 nt) and 32 proteins. The cytoplasmic ribosomes of the eukaryotes are generally somewhat larger; the RNA of the small subunit has 18S (as many as 1800 nt), and the larger subunit contains a 28S RNA (4800 nt), the 5S RNA, and an additional 5.8S rRNA (160 nt). The number of proteins in both subunits is also larger than in prokaryotes. The smallest eukaryotic ribosomes are found in the Microsporidia, a group of parasitic unicellular organisms that exist in the cells of many very different Eumetazoa and Protozoa; these apparently separated quite early from the evolutionary line of the eukaryotes. Their ribosomes are no larger than those of the prokaryotes [460].

Mitochondria and plastids possess their own systems for protein synthesis with a special type of ribosome. The ribosomes of the mitochondria are, without exception, smaller than prokaryotic ribosomes, both subunits contain only one RNA molecule each; the size of these mitochondrial rRNAs varies from 12S + 16S in the higher eukaryotes to 9S + 12S in the trypanosomes [49, 326]. In spite of these extreme differences in size and composition, the architecture of all ribosomes is so similar that they can be traced back to a common origin. Comparative investigations of ribosome structure have led several authors to classify the "Eocyta" as the fourth subkingdom of living organisms, in addition to the Archaebacteria, Eubacteria and Eukaryotes. This is a group of sulphur bacteria that live in hot springs and are closer than the other two subkingdoms to the eukaryotes in their ribosome structure [247]. According to other authors, however, ribosome morphology alone does not provide satisfactory evidence to justify such far-reaching conclusions [335].

The coding sequences of the 18S rRNA, 5.8S rRNA and 28S rRNA together make up the pre-rRNA genes (rDNA), which are organized in tandem repeats in the nucleolus. In all animals, in contrast to the situation in yeast, the genes for the 5S rRNA are separated from the other rRNA types. The number of pre-rRNA genes in the Eumetazoa lies between 45 and several thousand (Table 2.4), but the number of rRNA genes can differ quite markedly between different individuals of a species, as has been seen, for example, in *Drosophila* [156]. In the haploid genome of the malarial agent *Plasmodium berghei* there are only four rRNA genes, but these are, however, very different in both length and sequence and are expressed in a stage-specific manner [100]. The ciliate *Tetrahymena* has only one rRNA gene per haploid genome in the micronucleus, but in the non-chromosomal DNA of the macronucleus there are about 10 000 hairpin-like rDNA molecules that together amount to several percent of the total DNA [374].

Table 2.4. Number of rRNA, 5S rRNA and tRNA genes in the haploid genome of some animals [135, 140, 159, 242, 272, 340, 398, 417]

Species	rRNA	5S-RNA	tRNA
Plasmodium berghei	4		
Tetrahymena pyriformis (Micronucleus)	1	150	
Caenorhabditis briggsae	55		300
Schistosoma mansoni	100		
Urechis caupo	290		
Mytilus edulis	220		
Drosophila melanogaster	100– 240	100–200	590–900
Sciara coprophila	45		
Sarcophaga bullata	144		
Lytechinus variegatus	260		
Squalus acanthias	960		
Carassius auratus	240– 280		
Necturus maculosus	2700–4000		
Xenopus laevis	500– 760	24000	6500–7800
Xenopus borealis	500	9000	
Gallus domesticus	190– 200		
Rattus norwegicus	150– 170	830	6500
Homo sapiens	50– 200	2000	1310

The repeated units of the rDNA contain the genes for the different rRNA types always in the order 18S, 5.8S and 28S with transcribed spacers in front of and behind the coding sequences (ETS) and between the genes of the individual rRNA types (ITS); there are also non-transcribed spacers (NTS) between the repeats (Fig. 2.14). Control elements in the NTS stimulate rDNA transcription [164]. The length of the transcribed and, in particular, the non-transcribed spacers differs greatly according to the species; thus, the mouse has rDNA repeats of 43.5 kb, of which 14.5 kb is transcribed. The corresponding values for *Xenopus* are 12.5 and 7.5 kb [411]. The smallest eukaryotic rDNA unit, of only 5.4 kb, is found in the flagellate *Giardia lamblia;* the three rRNAs are also smaller than in any other eukaryote [39a]. The total length of the rDNA repeats, as well as the length of individual repeats, varies between different species; this is based on the fact that the NTS are mainly made up of short repetitive sequences, the number of which can vary (Fig. 2.15).

Approximately half of the rDNA repeats in *Drosophila melanogaster* contain insertions in the 28S gene, and these are apparently active retroposons. They contain the coding sequence for a reverse transcriptase and mostly create duplications of the target sequence. Two different types of insertion are found in the rDNA repeats: type I (or R1Dm) of 5.35 kb, and type II (or R2Dm) that is 3.6 kb long and inserted 74 bp further upstream. Both show agreement in 28–39 % of their sequences with the retroposons R1Bm and R2Bm found in the 28S rRNA genes of the silkworm, *Bombyx mori*. The type I of *D. melanogaster* corresponds to elements that are found in *D. virilis* and the bluebottle, *Calliphora erythrocephala*. There are insertions in the 28S rRNA genes of the mosquitos *Anopheles gambiae* and *A. arabiensis* that are inserted 634 bp further downstream than the type I of *D. melanogaster*, although these apparently have similar target sequences [211, 337].

In the ciliate *Tetrahymena*, whose rDNA repeats in the macronucleus can all be traced back to the two homologous genes of the diploid micronucleus, there is an intron of 403–413 nt in the 28S region of all repeats that has a similar sequence in all species and belongs to the class of self-splicing introns (p. 15). The intronless alleles found in *T. pigmentosa* are expressed to the same extent as the intron-containing alleles [321, 322].

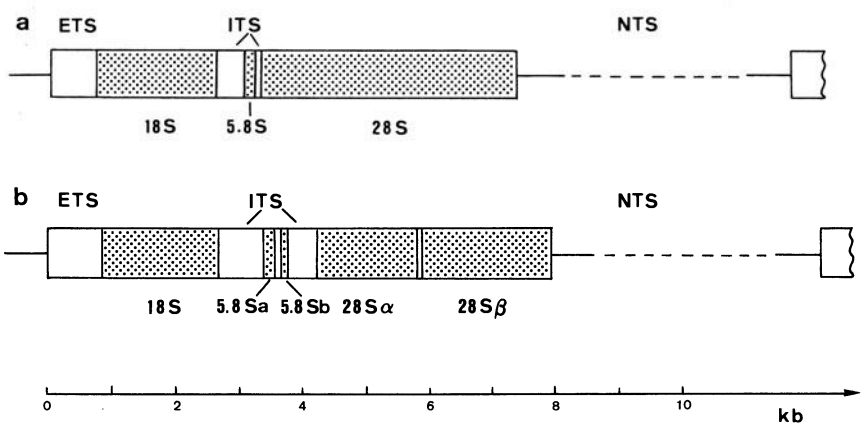


Fig. 2.14 a, b. The structure of the rDNA repeats of **a** *Xenopus laevis* and **b** *Drosophila melanogaster* [272]. *ETS* and *ITS,* transcribed spacers in front of or within the coding sequence; *NTS,* non-transcribed spacer of variable length. See text for further explanation

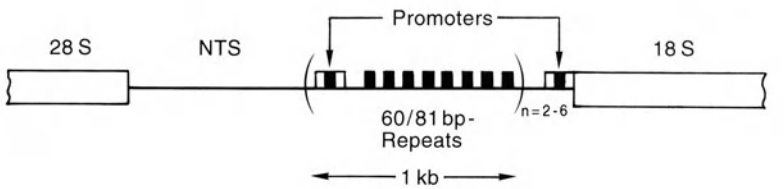


Fig. 2.15. The structure of the spacers between the rRNA repeats of *Xenopus laevis*. The rRNA possesses two promoters (*white*); the non-transcribed spacer (*NTS*) contains two to six segments of 1 kb, made up of repeated elements of 60 and 81 bp that function as enhancers (*black*). These segments contain further promoters [411]

Amongst the vertebrates, only the salamander *Triturus vulgaris* has an intron, and this occurs in 8% of the rDNA repeats; they are not found in other species of *Triturus* or in other amphibian genera like *Xenopus* [270]. In view of the extensive similarities in the structural and functional characters of the rRNAs, the variation in gene organization, with differences in location, number of genes, lengths of spacers, presence or absence of introns etc., can hardly be looked upon as adaptive evolution.

About one-half of the total transcriptional activity of the cell is devoted to the transcription of the rDNA, and this forms the single function of RNA polymerase I. One mammalian cell requires in the order of 2 million new ribosomes before each cell division; with a generation time of 15 h, a Pol I transcription rate of 30 nt/s and a distance of 100 bp between the Pol I molecules, a minimum of 50 rRNA genes per haploid genome is required to cover the rRNA needs of the cell, and, in fact, the number of rRNA repeats is mostly higher (Table 2.4). A closer study of rDNA transcription has only become possible within the last few years since in vitro and in vivo systems have become available for transcribing cloned rDNA. The primary transcript (pre-rRNA) includes significant parts of the spacers (ETS) from in front of and behind the coding sequence, as well as the coding sequence itself and the internal spacers (ITS). At 47S, this pre-rRNA in mammals is larger than was originally thought (45S), but the 5′-terminal 650 nt are quickly removed; on the other hand, in *Xenopus* and *Drosophila* the 5′-terminus of the pre-rRNA is identical to that of the 18S rRNA. Contrary to previous ideas, transcription, at least in mammals and *Xenopus*, proceeds far beyond the end of the 28S rRNA [411].

The Pol I promoters, in contrast to those of Pol II and Pol III, appear to be taxon specific. Thus, there is no transcription of mouse rRNA in extracts of human cells, or of *Drosophila virilis* rDNA in extracts of *D. melanogaster*; however, *Xenopus laevis* and *X. borealis* are compatible, as are the mouse and rat, and also man and several ape species. The reason for this appears to lie with proteinaceous transcription factors that show species-specific sequence recognition. However, positive results have been reported with the rDNA of the beetle *Dytiscus* in oocytes of the clawed frog *Xenopus*, and with *Xenopus* rDNA in mouse cells. The efficiency of Pol I-catalysed transcription is apparently strongly dependent upon the experimental conditions. Thus, for initiation of transcription in *Xenopus* the "proximal promoter", the region between –40 and +10 (where the initiation site is +1), is in many cases sufficient, whereas in other cases optimal transcription requires additional upstream promoters between –150 and –110. In *Xenopus* there are further promoters lying 1 or more kb in front of the initiation site; in addition, there are repetitive sequences of 60 or 81 bp that act as enhancers to increase transcription (Fig. 2.15) [411]. The trypanosomes, whose Pol II transcripts with the later added "spliced leader" sequence are so unusual (p. 39), apparently have quite normal rDNA transcripts [118].

2.8.2 Structure of rRNAs

All known sequences of SS-rRNAs, LS-rRNAs and 5.8S rRNAs, together with those of the 5S rRNAs, tRNAs and snRNAs, are published regularly in supplementary issues of the journal *Nucleic Acids Research*. The length of the SS-rRNA in the small ribosomal subunit varies between 9S (about 640 nt) and 18S (about 1870 nt), and that of the LS-rRNA in the large subunit varies between 12S (about 1230 nt) and 28S (about 4800 nt). Nevertheless, the primary and secondary structures of both rRNA classes are very similar. The differences in length are the result of insertions in otherwise homologous regions which do not significantly disturb the secondary structure of the central region. Open

loops in the rRNAs alternate with stretches of double helix in which unorthodox base-pairing, like G-U and G-A, may be seen (Fig. 2.16) [49, 326]. Based on the known SS-rRNA sequences, a family tree can be constructed; although this does not provide the conclusive solution to the important problems of systematics, it nevertheless shows how heterogeneous the unicellular organisms are. *Vairimorpha necatrix* (Microspora) branches off first from the evolutionary line, then the flagellates *Euglena* and *Trypanosoma*, next the ciliates, and only then do *Amoeba castellanii*, baker's yeast, higher plants and the animals separate from each other [409, 460].

After transcription, the pre-rRNA is broken down into the individual rRNAs. This post-transcriptional process happens rather differently in the various groups of organisms, and leads to a wealth of different rRNA types. Sequence comparisons show that the 5.8S rRNA of the cytoplasmic ribosomes corresponds to the 5′ terminus of the prokaryote 23S rRNA; the known 5.8S rRNA sequences show more than 50 % agreement with the corresponding RNA of *E. coli*. Therefore, a new cutting site must have appeared

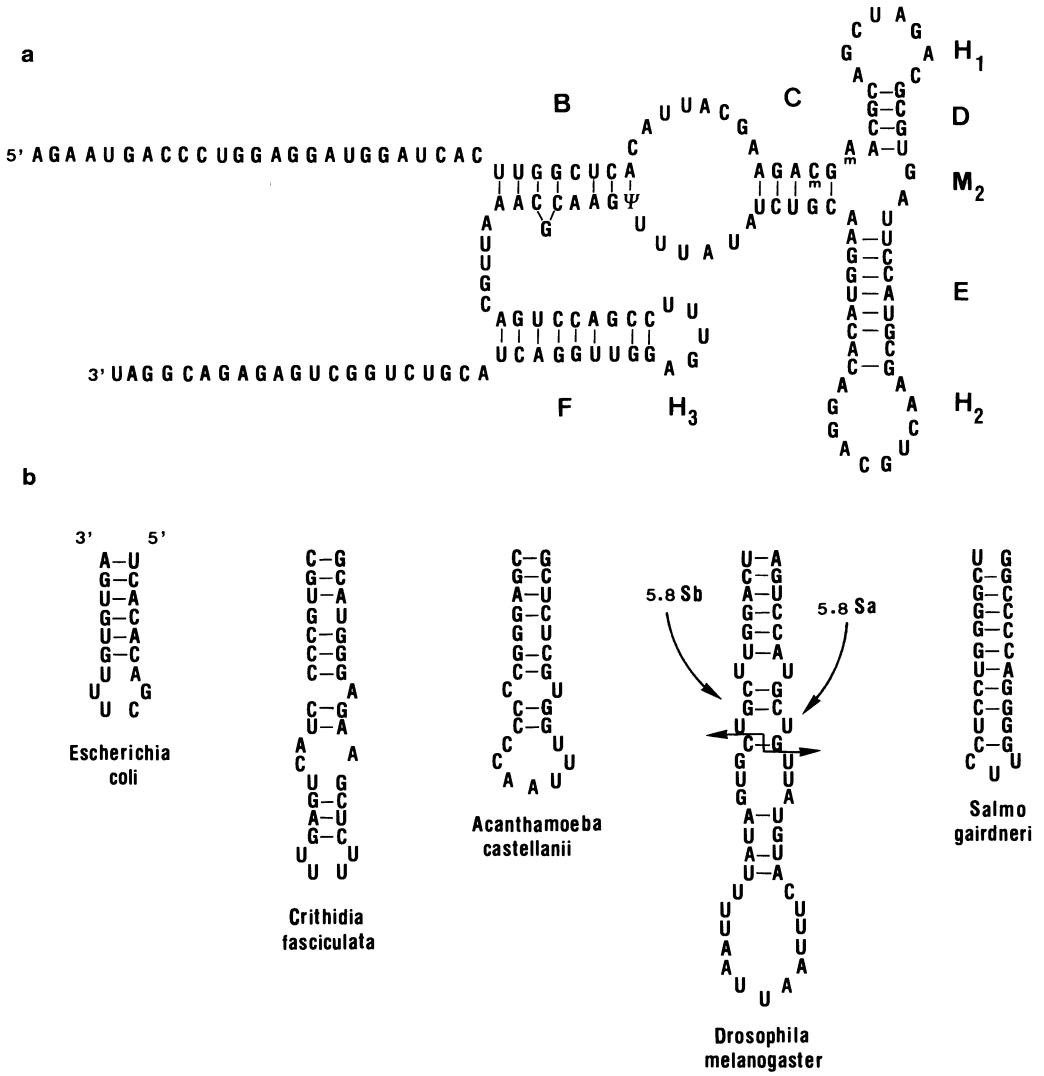


Fig. 2.16 a, b. The 5.8S rRNAs. **a** The sequence and structure of the 5.8S rRNA from *Artemia salina* [447]. *B–F are base-paired areas; M_2* a multi-branched loop; *H_1–H_3* hairpin loops. **b** The "hairpin" region F-H3 of different 5.8S rRNAs. The 5.8S rRNA of *Drosophila melanogaster* is made up of two fragments (S*a* and S*b*) that are held together non-covalently by base-pairing in the hairpin region. This configuration results from the excision of the fragment corresponding to the H3 loop out of the transcribed sequence (at the site shown by the *arrows*) without the subsequent joining of the free ends [386]; see also Fig. 2.14 b

during the evolution of the eukaryotic rRNA. The 5.8S rRNA is attached at both ends via hydrogen bonds to the 28S rRNA: the 3′ end joins with the 5′ terminus of the 28S rRNA, and the 5′ end attaches to a sequence 400 nt further away. The lengths of the 5.8S rRNAs are around 160 nt; those of the flagellate *Giardia lamblia* (140 nt) and the ciliate *Tetrahymena* (154 nt) are the shortest, whereas those of the silkworm *Bombyx mori* (167 nt) and the trypanosomes (170 nt) are the longest. Despite all the differences in length and sequence, the secondary structure always fits the same plan (Fig. 2.16a). The largest differences between the various 5.8S rRNAs lie in the (G+C)-rich region near the 3′ end (Fig. 2.16b). Compared with all the other 5.8S rRNAs, approximately 15 nt are missing from the 3′ end in *Giardia lamblia;* hence, the interaction with the 5′ end of the LS-rRNA is also absent [39a]. In the dipterans *Drosophila* and *Sciara*, a piece is cut out of the 5.8S rRNA after transcription, but the free ends are not spliced. Two short rRNAs result and these are held together only by base-pairing: 5.8Sa with 123 nt and 5.8Sb, or 2S rRNA, with 30 nt (Figs. 2.14 & 2.16b) [419, 433]. The largest variety of products is shown by the pre-rRNA of the flagellate *Crithidia fasciculata*, where, in addition to 5S rRNA and 5.8S rRNA, there are four further small rRNAs [49, 386].

In many animals, for example protozoans, molluscs, annelids, crustaceans and insects, there is a breakpoint in the middle of the 28S rRNA, and under denaturing conditions this gives rise to two fragments of about 18S ("28Sα and "28Sβ") (Fig. 2.14b). Investigations of *Drosophila*, the fungus gnat *Sciara coprophila* and the silkworm *Bombyx mori* have revealed not only the opening of a bond but also the excision of a short, extremely UA-rich, 19-nt fragment that forms a stem-loop structure. In the pea aphid, *Acyrtosiphon pisum*, and other aphids, this characteristic structure, and thus also the breakpoint, is missing [329]. The 5.8S rRNA of the silkworm *Bombyx mori* has an unusual length of 167 nt instead of about 160 nt, and is perhaps specifically adapted to form a stable complex with the 28S rRNA [141]. In several carp-related fish, preformed breakpoints that open under denaturing conditions are found in the majority of the 18S and 28S rRNA molecules of somatic cells, but not in germline cells. Ribosomes with such rRNAs have reduced translation activity in vitro. It is remarkable that such aberrant rRNAs are found only in tetraploid species, such as *Barbus barbus, Cyprinus carpio* and *Carassius auratus*, and do not occur in the diploid species of the family Cyprinidae or other fish [259, 260].

2.8.3 The 5S rRNAs and Their Genes

The 5S rRNA genes are organized in tandem sequences outside of the nucleolus and are transcribed by polymerase III (p. 37). In the clawed frogs of the genus *Xenopus*, and also in the teleost *Misgurnus fossilis*, there are two sets of 5S rRNA genes that are termed oocyte-specific or somatic, according to the cell type in which they are expressed; these differ in sequence by about 5% [289, 323]. In the clawed frogs, the gene family of the oocyte-specific 5S rRNA consists of 24 000 copies *(Xenopus laevis)* or 9000 copies *(X. borealis)* on chromosome 9, whereas that of the somatic type consists of just a few hundred copies dispersed over almost all the chromosomes (Table 2.4). Also, in the nemertean *Emplectonema gracile* there are two types of 5S rRNAs that differ by no less than 22 (14%) nucleotides [243]. Although in all other higher eukaryotes the 5S rRNA genes are independent of the rDNA repeats, in the plant-parasitic nematode *Meloidogyne arenaria* they are included [448]. The micronucleus of *Tetrahymena* contains, in fact, only one gene for the larger rRNAs but has about 30 clusters of altogether approximately 150 genes for the 5S rRNA; the number is not significantly larger in the macronucleus. In contrast, the macronucleus of the hypotrichous ciliate *Euplotes eurystomus* contains about 10^6 5S rRNA genes [340, 363].

More than 350 5S rRNAs have been sequenced. There have been continuous attempts to draw conclusions about phylogenetic relationships between the species from the species-specific differences amongst the 5S and 5.8S rRNAs, and to draw up family trees [199, 330, 412]. The results are, however, contradictory or even quite obviously absurd; clearly, these short sequences do not provide sufficient information for determining evolutionary distance [188]. Better results can be expected from comparisons of the significantly longer SS-rRNA and LS-rRNA sequences.

2.8.4 Ribosomal Proteins

The ribosomes of *E. coli* contain altogether 53 proteins in their two subunits; the number of ribosomal proteins in the eukaryotes is known

only approximately but is clearly larger than in the prokaryotes. For example, 84 proteins have been isolated from the cytoplasmic ribosomes of rat liver [64]; with the help of 2-D electrophoresis, 35–45 proteins were identified in the large subunits of the insects *Acheta, Drosophila* and *Bombyx* and 30–32 were identified in the small subunits [41, 278]. Of these, at least 26 of the *Drosophila* proteins have already been purified [74]. The ribosomal proteins from rat liver have molecular masses between 11 and 41 kDa, and those of *Drosophila* are between 11.5 and 61 kDa. There is apparently only one molecule of each protein species per ribosome, except for the protein type that corresponds to L7/L12 of *E. coli;* this is present in several copies. The proteins of the small subunit are designated "S" and those of the large subunit "L", and they are numbered rather variably by different authors according to their location in 2-D gels. Because comparisons between proteins isolated by different authors are so difficult, attempts have been made to develop a uniform nomenclature [294]. The 53 ribosomal proteins of *E. coli*, with altogether 7336 amino acids, have now been completely sequenced. The sequencing of all the ribosomal proteins of the rat, involving about 16000 amino acids, is now the subject of intense efforts; by 1989 more than 30 had been examined [65, 339, 426]. Although altogether about 500 ribosomal proteins have been sequenced so far, the information obtained is still insufficient to allow firm statements to be made about species-specific sequence differences and the evolution of the ribosomal proteins. However, it is already quite clear that they are very conserved proteins. Direct sequence comparisons show that, for example, the proteins S14 and L31 of the rat and man differ by less than 1%, protein S19 differs by just as little between *Xenopus* and the golden hamster, 22% difference exists between S26 in the rat and the homologous protein in *Drosophila*, 25% difference exists between S6 in the rat and yeast, and there is about 50% identity between L44 in *Trypanosoma brucei*, baker's yeast and man [19, 207, 231, 255, 435].

Although in vitro many ribosomal proteins are phosphorylated by cAMP-dependent or -independent protein kinases, in vivo one particular basic protein is phosphorylated in all eukaryotes: the protein from rat liver is designated S6 and has within its sequence of 249 amino acids 15 serine residues that possibly act as a phosphate receptor [187, 473]. S6-like proteins have been detected in various animals, even down to the Protozoa; the protein from ovaries of the clawed frog, *Xenopus laevis*, shows immunological cross-reactivity with the S6 from chicken liver. In all cells that have been examined, the phosphorylation of S6 increases in connection with cell division and with the association of ribosomes into polysomes [217]. The ribosomes of yeast, higher plants, the crustacean *Artemia salina, Drosophila melanogaster*, and various mammalian organs all contain two strongly acidic proteins with pI values less than 4.5 and molecular masses from 13.5 to 17 kDa; in *Artemia* and *Drosophila* these are singly phosphorylated, and in mammals they are multiply phosphorylated. The two proteins in *Artemia*, known as eL12 and eL12′, are found in several copies in the large ribosomal subunit and are possibly involved in the elongation step of protein biosynthesis. The sequences of eL12 and eL12′ are known; they are identical in the 22 amino acids of the C-terminus but otherwise show no significant homology. Although the protein L7/L12 from *E. coli* shows little similarity to eL12, it can replace eL12 in the ribosomes of *Artemia* [275].

The **proteins of the mitochondrial ribosomes** form a distinct group, although they are synthesized in the cytoplasm and only later introduced into the mitochondria. Comparisons, by 2-D gel electrophoresis, of mitochondrial and cytoplasmic ribosomal proteins from the same tissue reveal only very limited similarities between species from the vertebrates to *Neurospora* [393]. Just like the rRNAs, the proteins of the mitochondrial ribosomes show a much greater rate of evolution than do the components of the cytoplasmic ribosomes. In 2-D gels of rat and cow ribosomal proteins, only 15% of the spots are different for cytoplasmic ribosomes, whereas 85% differ for the mitochondrial ribosomes [344]. Even between the closely related clawed frog species *Xenopus laevis* and *X. mulleri*, the mitochondrial ribosomes show seven differences. It appears that only in the mammals are there multiple genes for the ribosomal proteins, e.g. 7–20 copies per protein in the mouse. In contrast, there are only one or two copies of each gene in yeast, insects and amphibians. Surprisingly, only 300–400 relatively stable mRNAs are present for each ribosomal protein in rapidly growing mouse cells; the reason for the large number of genes is therefore not clear [25, 304].

2.8.5 The Transfer RNAs and Their Genes

The cytoplasm of eukaryotic cells contains up to 10^8 tRNA molecules with lengths of 75–93 nt; together these make up 10–15 % of the total RNA. The number of different types of tRNA is greater than that of the amino acids, so that several different tRNAs (isoaccepting tRNAs) are available for the same amino acid; on the other hand, there is not a specific tRNA for each amino acid codon as, in accordance with the "wobble effect", the anticodon of some tRNAs recognizes different codons of the same amino acid. The mitochondria contain a special set of tRNAs that differ from those in the cytoplasm (p. 53). The homologous tRNAs of different animal species show only minor differences; e.g., the sequence identity between the $tRNA^{Lys/AAA}$ of *Drosophila melanogaster* and that of the rabbit is about 95 %, and the $tRNA^{Lys/AAA}$ of *Drosophila* and man has in fact the same sequence [242]. Regardless of the differences in length and sequence, tRNAs always have a similar 3-D structure with four helical regions and three or four loops (Fig. 2.17).

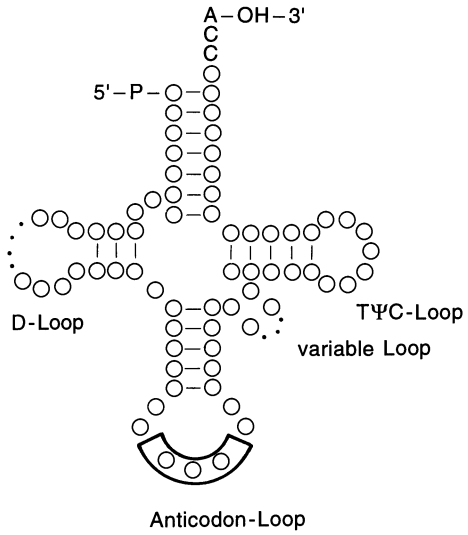


Fig. 2.17. A model of a tRNA molecule. The amino acid is attached to the 3' hydroxyl group by an ester bond

In bacteria and yeast there are nonsense suppressor tRNAs that can read through stop codons. The corresponding effect is achieved in higher eukaryotes through false reading of stop codons by normal cytoplasmic tRNAs. Thus, in rabbit reticulocytes the stop codon UGA (opal) of the β-globin mRNA is sometimes overlooked in vivo and a "read through" protein, which is longer than normal by 22 amino acids, is synthesized. The responsible tRNA here is either the normal $tRNA^{Trp}$, as has been detected in vitro, or a specific suppressor $tRNA^{Ser}$. Two opal-suppressive $tRNA^{Ser}$ molecules are known in the cow and chicken, and constitute about 1–3 % of the total $tRNA^{Ser}$ population; they are probably post-transcriptionally modified products of the same gene. Their anticodon CmCA (mC represents methylcytosine) recognizes the stop codon UGA rather than, as expected, UGG = Trp [182]. In the nematode *Caenorhabditis elegans*, mutations of CCA → CTA in the anticodons of five $tRNA^{Trp}$ genes resulted in suppressor tRNAs for the "amber" stop codon UAG [237].

All animals possess hundreds, if not thousands, of tRNA genes (Table 2.4) that are organised very differently between the various species: they may occur as clusters of the same or different tRNA genes, e.g. in the clawed frog *Xenopus*, the salamander *Taricha, Drosophila, Trypanosoma* and *Tetrahymena;* or are mainly dispersed in the genome as single genes, e.g. in man, the nematode *Caenorhabditis elegans* and the silkworm *Bombyx mori*. In *Xenopus* there are at least 43 different tRNA genes, each with 200 copies that are mainly arranged in long tandem clusters; in *Drosophila* there 12 copies each of about 60 different genes that are dispersed in many short clusters [58, 122, 312, 417, 437]. With few exceptions, the tRNA genes of animals, in contrast to those of yeast, contain no introns [163]. The tRNA genes are transcribed by polymerase III, and the formation of mature tRNA from the primary transcripts requires several steps: the removal of superfluous nucleotides from in front of the 5′ end and after the 3′ end, the addition of the sequence CCA to the 3′ terminus, and the modification of individual nucleotides [368]. The sequences of all the genes for a particular tRNA are kept constant by "horizontal evolution", which at the same time hinders adaptive evolution. In place of the latter, there appears to be adaptation through post-transcriptional modifications like methylation, acetylation and pseudo-uridylation; however, in all eukaryotes these are limited to about 20 specific positions. As a result, the tRNAs of the eukaryotes contain more unusual nucleotides than all other RNAs; over 50 modified nucleotides have already been discovered [242].

2.8.6 The Small Nuclear RNAs and Their Genes

A very important part of the spliceosome, the complicated molecular complex in the cell nucleus that converts the primary transcripts of protein-coding genes into mature, functional mRNAs in all organisms from the ciliate *Tetrahymena* to man, is the group of different types of small ribonucleoproteins (snRNPs). An snRNP is made up of one molecule of an snRNA, 56–217 nt long, together with about ten different proteins; some of these proteins occur in all snRNPs and others are limited to particular types. Thirteen different snRNAs are known so far and are designated U1 to U13; however, results with yeast and plant cells suggest that all eukaryotic cells probably contain 20–30 different snRNAs. U1 to U6 are present in about 10^5 copies per cell. The quantities of the other snRNAs are much smaller, and as a consequence they were found only more recently. With the exception of snRNA U6, the gene for which is transcribed by Pol III, all other U-snRNAs are products of Pol II. The functions of the individual snRNAs in the spliceosome are only partly known. The U1-, U2-, U4/U6- and U5-snRNPs are essential components of the spliceosomes; U7 is required for the formation of the 3′ end of histone mRNA [171, 273]. The snRNA sequences are apparently very conservative in terms of evolution. The partially sequenced snRNAs U1 to U6 of the silkworm *Bombyx mori* are 66–81 % identical with the corresponding snRNAs of the vertebrates, and U6 of *Trypanosoma brucei* is 62 % identical with that of the rat. In the nematodes, which possess trans-splicing as well as normal spliceosomes, there are no overall peculiarities in the snRNAs [4, 438, 441]. There are about 2000 genes present for each snRNA in the higher vertebrates, although the majority are actually pseudogenes. In man, the rat and the chicken, and also in *Drosophila*, the snRNA genes are dispersed in the genome, whereas in the sea urchin *Lytechinus* they are arranged tandemly. Variants of particular snRNAs occur in man and other mammals, and also during sea urchin development [171, 274, 378].

2.9 Mitochondrial DNA

The mitochondria of the eukaryotes possess their own system of protein synthesis, for which the rRNAs and tRNAs are encoded in the mitochondrial genome, and the required enzymes and protein components come from the cytoplasm [232]. The genetic code involved here deviates markedly from that of the chromosomal-cytoplasmic system. The discovery of a unique genetic apparatus in the mitochondria and plastids led to the idea that these organelles have their origin in symbiotic prokaryotes that entered primitive eukaryotic cells lacking mitochondria at a very early stage of evolution [162, 169]. Although many characters of the mitochondrial DNA (mtDNA) genome are identical in all eukaryotes, and they can be considered as homologues, there is a large variation in the molecular structure. In the Eumetazoa, and apparently in some Protozoa, the mtDNA has the form of a circular DNA double helix with a length of 15–20 kb (in extreme cases, up to 42 kb); in the mitochondria of the Ciliophora, however, one finds larger, linear molecules, and in the trypanosomes and relatives a complicated network of linked circles exists. During evolution there has apparently been a transfer of DNA between the mitochondria and the nucleus. At any rate, mtDNA-like sequences have been detected in the nuclei of many different kinds of eukaryotes (yeast, the migratory locust, sea urchin, mammals and man) and were probably integrated into the chromosomal DNA by illegitimate recombination [91].

As a rule, each mitochondrium contains about five to ten mtDNA molecules associated with the inner membrane; in the unfertilized eggs of *Drosophila* the number of mtDNA circles increases to 50 per organelle and declines again after fertilization [417]. Complete or nearly complete mtDNA sequences are now known for more than a dozen species of vertebrates, invertebrates and protozoans. Sequencing and restriction endonuclease analysis clearly indicate a higher rate of evolution of mtDNA than of chromosomal DNA. This may be related to the fact that the mitochondria lack a DNA repair system. As a result, mtDNA is a particularly useful system for molecular studies of relationships between species and the study of molecular evolution.

2.9.1 The mtDNA of Vertebrates

Human mtDNA will be used both as an example of vertebrate mtDNA and to illustrate mtDNA characteristics in general. It has a total length of 16 569 nt and codes for 2 rRNAs, 22 tRNAs and 13 mitochondrial proteins (Fig. 2.18). Five of these proteins were identified some time ago as

the subunits I–III of cytochrome oxidase, subunit 6 of ATPase, and cytochrome b; later, a further one was identified as subunit 8 of ATPase. As the products of the remaining seven protein-coding sequences were unknown, they were referred to as URFs (unassigned reading frames). In 1986 it became known that all seven URFs encode polypeptides for complex I of the respiration chain; they have since been designated "ND" (for NADH dehydrogenase) [154]. All other mitochondrial proteins originate in the cytoplasm and are transported into the mitochondria by an as yet incompletely understood mechanism.

The 12S rRNA for the smaller subunit and the 16S rRNA for the larger subunit of the mitochondrial ribosomes are significantly smaller than the cytoplasmic rRNAs; 5S and 5.8S rRNAs are not present. The number of tRNAs is notably very small and not compatible with the classical idea of the relationship between the number of mRNA codons and tRNA anticodons. According to the wobble theory, 31 tRNAs with different anticodons are required for the 61 amino acid codons. Thus, in the mitochondria a variant of the wobble mechanism must apply, whereby an unmodified U in the first position of the anticodon can be paired with any of the four bases in the third position of synonymous (coding for the same amino acid) codons, whereas a modified U can pair only with A or G. However, in this way the number of required tRNAs would only be reduced to 23. A further tRNA is spared because in vertebrate mitochondria the normal codons for arginine (AGA and AGG) function as stop codons, along with UAA and UAG; the standard stop codon UGA here codes for the amino acid tryptophan. As many mitochondrial genes end with -U or -UA, the stop codon UAA often only appears after polyadenylation of the mRNA. The translation start can be signified by each of the four AUN codons, all of which then apparently code for methionine; during elongation, AUU specifies isoleucine, as in the universal code [80, 139, 351].

The two strands of the mtDNA double helix are distinguished, according to density, as the H-strand (heavy) and L-strand (light). Most mtDNA molecules have a small section with a three-stranded structure in which a third strand, complementary to the L-strand (the D-loop, displacement loop), lies between the strands of the double helix (Fig. 2.18). Replication of mtDNA

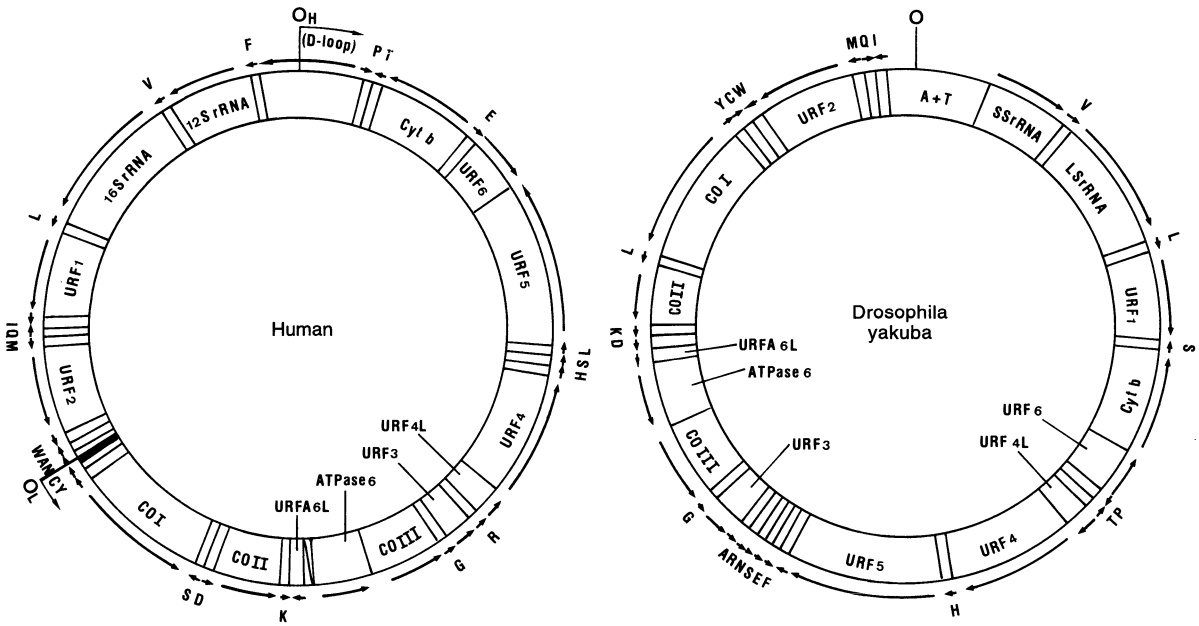


Fig. 2.18. The mitochondrial DNAs (mtDNAs) of man [80] and the fly *Drosophila yakuba* [481]. The starting points for replication of the heavy and light strands of the human mtDNA are indicated by O_H and O_L, respectively; *O* marks the start of replication in the *Drosophila* mtDNA. The *arrows* show the orientation of the genes. The tRNA genes are identified by the letter code of their respective amino acids. *12S rRNA* and *SS rRNA* indicate the RNAs of the small ribosomal subunits, and *16S rRNA* and *LSrRNA* those of the larger subunits. For the protein-coding genes, *COI–COIII* and *ATPase 6* signify the subunits of cytochrome oxidase and ATPase, respectively. Of the human URFs (unassigned reading frames), recent results show that URFA6L encodes the ATPase 8 subunits, and the other seven URFs encode subunits of NADH dehydrogenase [72]

begins in the D-loop region with the formation of a new H-strand, during which the old H-strand is displaced; the synthesis of the L-strand begins only when the tRNA cluster on the old H-strand, between ORF-CO1 and ND2 (URF2), becomes exposed. The transcription of the two strands progresses from the D-loop region and leads to the formation of two large transcripts, the processing of which has not yet been fully described.

The mtDNA molecule shows the greatest economy of space in that there are no introns and practically no non-coding sequences between the genes. The total number of nucleotides lying between genes is 87 in man, 64 in the mouse, and 57 in the cow. The tRNA genes lying between the individual rRNA and structural genes serve as signals to separate the transcripts. The sequence -CAA is attached post-transcriptionally to the 3′ end of the pre-tRNAs [80]. A peculiarity of human mtDNA is a 200-nt-long sequence lying in front of the D-loop on the L-strand; this codes for a 7S RNA of unknown function, but sequences homologous to this have not been found in other mtDNAs [80]. Complete sequences are also known for the mtDNAs of the cow, mouse, chicken and clawed frog, *Xenopus laevis*. These data, together with partial sequences and restriction patterns from the mtDNAs of amphibians and fish, show that the organization of the genes in the mtDNA of all vertebrates corresponds to that of man. The single known exception is in the chicken, where the genes for ND6 and cytochrome b have exchanged positions. Vertebrate mtDNAs are also very similar in sequence: the mtDNA of the cod fish, *Gadus morrhua*, and that of the clawed frog, *Xenopus laevis*, agree in their coding sequences by between 46 (ND4L) and 93% (CO I) [106, 145, 214, 365]. In many vertebrates one finds variant mtDNAs up to 500 bp longer, and sometimes these are in combination with the normal form. Cells containing more than one type of mtDNA are referred to as heteroplastic. In several species of the lizard genus *Cnemidophorus*, the mtDNA of certain populations or single individuals contains duplications with a length of 0.8–8 kb, resulting in increases in the total length of up to 25 kb. In the teleost *Alosa sapidissima*, the mtDNA can be increased in size by two to three repeats of 1.5 kb in the D-loop region. Similar length polymorphisms are known from other fish, amphibians and reptiles [16, 29, 306].

2.9.2 The mtDNA of Invertebrates

Complete sequences are known for the dipterans *Drosophila yakuba* and *Anopheles quadrimaculatus*, the nematodes *Ascaris suum* and *Caenorhabditis elegans*, and the sea urchins *Paracentrous lividus* and *Strongylocentrous purpuratus* [61, 78, 82, 209]. Partial sequences, or at least the order of the genes, are available for the migratory locust *Locusta migratoria*, three species of broad-nosed weevil (Curculionidae), the brine shrimp *Artemia salina*, the liver-fluke *Fasciola hepatica,* and the starfish *Pisaster ochraceus* [24, 42, 148, 405, 482].

The mtDNA of *Drosophila yakuba*, the first determined invertebrate mtDNA sequence, has a length of 16 019 bp. It contains the same genes as vertebrate mtDNA, although in a rather different order (Fig. 2.18). A peculiarity of the insects is the (A+T)-rich region that has a length of 1 kb in *D. yakuba, D. teissieri, D. erecta* and *D. orena* but about 5 kb in *D. melanogaster, D. simulans* and *D. mauritiana;* the total length of mtDNA in *Drosophila* varies correspondingly between 16.0 and 19.5 kb [410]. The basic structure of *Drosophila* mtDNA is valid for all other insects, the encoded protein sequences of *Anopheles* and *Drosophila* agreeing by 72–98% [24, 42, 82]. In the crustacean *Artemia*, only the tRNA genes are organized differently from those in the insects [24]. The mtDNA of the pig roundworm *Ascaris suum* has a length of 14 284 bp and contains the genes for the two rRNAs and all the proteins that are encoded in the mtDNAs of the vertebrates and the arthropods, except for the ATPase subunit 8. Distributed between the rRNA and the protein genes are genes for an aberrant type of tRNA. These lack the TψC loop and the variable loop that are found both in the mitochondrial and cytoplasmic tRNAs of vertebrates and insects (Fig. 2.17); instead there is a single loop of 4–12 nt. A similar situation is seen in the mtDNA of *Caenorhabditis elegans* and is thus typical for the nematodes [482]. The mtDNAs of the two sea-urchin species both have a gene order that is distinct from all other mtDNAs: between the genes for the SS-rRNA and the LS-rRNA lies a cluster of 15 of the 22 tRNAs and the genes for ND1 and ND2. The mtDNA of the starfish *Pisaster ochraceus*, in contrast, contains an inversion of the segment: -tRNA cluster-ND1-ND2-16S rRNA- [61, 209]. During the evolution of sea-urchin mtDNA, the coding sequence for the $tRNA^{Leu/CUN}$ fused with the gene for the NADH dehydrogenase subunit 5 and this now encodes a 24-

amino-acid, N-terminal extension of this protein. In parallel, duplication and diversification has produced a new $tRNA^{Leu/CUN}$ from the gene for $tRNA^{Leu/UUR}$ [60]. All in all, it is clear that multiple rearrangements have taken place in the evolution of the mtDNA so that the order of genes differs quite markedly between the individual animal groups. The tRNA genes are primarily affected.

The mtDNA genetic codes of the invertebrates have the same unique features as those of the vertebrates. In addition, AGA and AGG here code for serine instead of arginine, as in the standard code, or for STOP in the case of vertebrate mtDNA. In the mtDNAs of the starfish *Pisaster ochraceus*, the sea urchin *Strongylocentrous purpuratus* and the liver-fluke *Fasciola hepatica*, AAA appears to code for asparagine instead of lysine [24, 331]. The pilgrim scallop, *Placopecten magellanicus*, possesses the largest mtDNAs that have so far been recorded. The insertion of two to eight copies of a 1442-bp repeat lengthens the mtDNA to an average of 35 kb and, in extreme cases, to 42 kb. In various broad-nosed weevils (Curculionidae), the mtDNA has a total length of 30–36 kb due to increases in an (A+T)-rich region of 9–13 kb; individual beetles are heteroplastic, with two to five mtDNA size classes [42]. Size variants and heteroplasty, owing to differences in the number of repeats, are also found in *Gryllus*, various *Drosophila* species and nematodes [42, 252, 410].

2.9.3 The mtDNA of Ciliates

The mitochondrial DNA of *Paramecium aurelia* is a linear double helix with a length of 40 469 bp and deviates markedly from the circular mtDNA of the Eumetazoa in both gene content and order. The CO I and II genes of cytochrome oxidase are present but CO III is not; there are genes for cytochrome b and N1 to N5, but ND4L and ND6 are missing. Instead, there are two other genes that apparently code for subunits of NADH dehydrogenase and correspond to parts of the chloroplast genome. The genes for the ATPase subunits 6 and 8 are not found, but a gene for ATPase 9 is present; this has otherwise been reported only in the mtDNA of yeast and higher plants. *Paramecium* mtDNA includes genes for ribosomal proteins (L2, L14, S12 and S14) that are usually to be found only in the DNA of plant mitochondria and chloroplasts and in the mtDNA of the ciliate *Tetrahymena*. At least 31 mono- and polycistronic transcripts are produced from the mtDNA of *Paramecium* but their maturation has not yet been described. The only partially sequenced mtDNA of *Tetrahymena thermophila* carries genes for CO I, CO II and L14 in the same order as in *Paramecium*, but codes for at least eight tRNAs as against the three tRNAs of *Paramecium*. However, there do appear to be differences in the order of the genes between individual *Tetrahymena* species. The start of replication in *Paramecium* lies at one end of the mtDNA, and in *Tetrahymena* it is in the middle. The genetic code in the mtDNAs of the ciliates is in general closer to the standard code than is that of many other mtDNAs; the translation start codons here are AUN, GUG and perhaps GUA [305, 351].

2.9.4 Kinetoplast DNA

In stained preparations of trypanosomes and other flagellates of the group Kinetoplastida, a body may be seen at the base of the flagellum; it varies in size from 4–6 μm in the genus *Trypanosoma* to 25–20 μm in *Crithidia*. It is actually a modified mitochondrium but was at first mistakenly considered to be involved in flagellum movement and was thus termed the kinetoplast. This cell organelle contains up to 25% of the total DNA of these unicellular organisms. The kinetoplast DNA (kDNA) consists of several thousand smaller, circular DNA molecules (mini-circles) and up to 50 larger maxi-circles; these mesh with each other in a 3-D network like the links of a chain (catemers) [124]. Some strains of *Trypanosoma equiperdum* are known to lack kinotoplasts. They have neither a network nor mini-circles, but contain single circular DNA structures that apparently correspond to the maxi-circles of other strains [173].

Maxi-circles have a size of 20–40 kb, according to the species. The length variation stems from species-specific differences in a non-transcribed region made up of tandem repeats of various types. The transcribed region, which corresponds to the mtDNA of other eukaryotes, has a similar size of 15–17 kb in all species. The maxi-circles of *Leishmania tarentolae* and *Trypanosoma brucei* have been completely sequenced. They contain two rRNA genes, the structural gene for cytochrome b, CO I–III, ND1, ND4, ND5 and several UFRs; ATPase and tRNA genes have not been found [28, 174]. UUG, CUG and UUA, which in the standard code all signify leucine,

function here as methionine start codons, in addition to AUG [402]. The rRNAs in the kinetoplasts of *L. tarentolae, T. brucei* and *Crithidia fasciculata* are 9S (610–612 nt) and 12S (1141–1150 nt) and are the smallest known. In spite of their minimum size and unusual composition (83 % A+U), their secondary structure includes the highly conserved central regions found in all organisms. The sequences of the rRNAs of the three species agree at 77–84 % of positions [97, 401].

The transcripts of the maxi-circles are later altered by insertion and deletion of uridine residues before they become functional ("RNA editing"). This unique process was discovered through the uridine residues found in the transcripts that are not encoded by the maxi-circle DNA. These alterations are often restricted to a small region of the mRNA and cause, for example, the creation of an initiation codon AUG (in CO III and URF2 of *C. fasciculata*) or a shift in a reading frame (in the CO II gene of *C. fasciculata*). In contrast, there are 39 extra uridine residues in the cytochrome b mRNA of *C. fasciculata* and *L. tarentolae*. Finally, the CO III transcript in *T. brucei*, encoded in a maxi-circle of 450 nt, is extended to an mRNA of almost 1000 nt by the insertion of several hundred uridine residues, such that the final mRNA sequence bears little resemblance to the corresponding gene [28, 132, 414]. There are two competing ideas about the mechanism of RNA editing: one proposes that the edited mRNA is synthesized as a whole [458], and the other envisages a cyclic process from the 3′ to the 5′ end involving a family of small transcripts of the maxi- and mini-circles and known as gRNAs (guide RNAs). Mismatch repair between the mRNA–gRNA pairs continues until the mRNA perfectly matches the gRNA and is released. In *L. tarentolae*, four maxi-circle-coded gRNAs are known for maxi-circle genes; the gRNA for the less exactly edited CO III gene is encoded by mini-circles [425, 449].

The mini-circles make up about 95 % of the kDNA. They are uniform in size within an individual network, but may vary species-specifically between 900 and 2500 bp. The sequences of the mini-circles show contrasting levels of heterogeneity in different species. In *T. brucei* there are several hundred different classes of mini-circle; in *T. equiperdum*, on the other hand, little heterogeneity is found. The sequences of mini-circles vary dramatically, even between closely related species like *Crithidia luciliae* and *C. fasciculata*, or between the different strains of *T. brucei*. The species-specific variations in length and sequence, together with the sequence heterogeneity, speak for an extremely rapid rate of evolution of the mini-circles. Their function is, however, unknown. The open reading frames (ORFs) of the mini-circles would be sufficient for proteins of 20–70 amino acids and, in fact, mini-circles from *C. fasciculata* expressed in *E. coli* yielded immunoreactive products [124, 396]. The mini-circles are replicated as free molecules after release from the network, and the daughter molecules are reincorporated. Thus, the kinetoplast grows to about double its original size up to the moment of cell division, when it itself divides. The two daughter molecules produced during replication are rather different: one contains the discontinuously growing H-strand with fragments of 20–110 nt in length, and the other contains the continuously growing L-strand. The H-strand is afterwards partly repaired, whilst the L-strand becomes nicked; hence, during the reincorporation phase both daughter molecules possess nicks that are closed on incorporation into the network. A specific topoisomerase is responsible for the release of mini-circles (decatenation) from the network and for the reincorporation of the daughter molecules (catenation) [372].

References

1. Abad P. et al.: A long interspersed element – the I factor of Drosophila teissieri – is able to transpose in different Drosophila species. Proc. Nat. Acad. Sci. USA 86: 8887–91 (1989)
2. Acampora D. et al.: The human HOX gene family. Nucleic Acids Res. 17: 10385–402 (1989)
3. Achwal C. W., Iyer C. A. and Chandra H. S.: Immunochemical evidence for the presence of 5mC, 6mA and 7mG in human, Drosophila and mealybug DNA. FEBS Letters 158: 353–358 (1983)
4. Adams D. S. et al.: Isolation and partial characterization of U1-U6 small RNAs from Bombyx mori. Biochemistry 24: 117–125 (1985)
5. Adams R. L. P. and Burdon R. H.: Molecular biology of DNA methylation. Springer, Berlin 1985
6. Adams R. L. P., Leader D. P. and Knowler J. T.: The biochemistry of nucleic acids, 11th ed. Chapman & Hall, New York 1992
7. Adams R. L. P.: DNA methylation. The effect of minor bases on DNA-protein intereactions. Biochem. J. 265: 309–320 (1990)
8. Adelman J. P. et al.: Two mammalian genes transcribed from opposite strands of the same DNA locus. Science 235: 1514–17 (1987)
9. Aeby P. et al.: Structure and genomic organization of proretrovirus-like elements partially eliminated from the somatic genome of Ascaris lumbricoides. Embo J. 5: 3353–60 (1986)

10. Agabian, N.: Trans splicing of nuclear pre-mRNAs. Cell 61: 1157–60 (1990)
11. Alberts B. et al.: Molekularbiologie der Zelle. VHC Verlagsgesellschaft, Weinheim 1986
12. Amin J., Ananthan J. and Voellmy R.: Key features of heat shock regulatory elements. Mol. Cell Biol. 8: 3761–69 (1988)
13. Ashley C. T. et al.: Isolation and sequencing of cDNA clones encoding Drosophila protein D1. A repeating motif in proteins which recognize AT DNA. J. biol. Chem. 264: 8394/8401 (1989)
14. Ausio J. and McParland, R.: Sequence and characterization of the sperm-specific protein phi-3 from Mytilus californianus. Eur. J. Biochem. 182: 569–576 (1989)
15. Austerberry C. F., Yao M. C.: Sequence structures of two developmentally regulated, alternative DNA deletion junctions in Tetrahymena thermophila. Mol. Cell Biol. 8: 3947–50 (1988)
16. Avise J. C., Bowen B. W. and Lamb T.: DNA fingerprints from hypervariable mitochondrial genotypes. Mol. Biol. Evol. 6: 258–269 (1989)
17. Ayala F. J. (ed.): Molecular evolution. Sinauer Associates, Sunderland 1976
18. Babity J. M., Starr T. V. B. and Rose A. M.: Tc1 transposition and mutator activity in a Bristol strain of Caenorhabditis elegans. Mol. Gen. Genetics 222: 65–70 (1990)
19. Bagni C. et al.: Structure of Xenopus laevis ribosomal protein L32 and its expression during development. Nucleic Acids Res. 18: 4423–26 (1990)
20. Bambara R. A. and Jessee C. B.: Properties of DNA polymerases delta and epsilon, and their roles in eukaryotic DNA replication (Review). Biochim. Biophys. Acta 1088: 1–9 (1991)
21. Bannon G. A. et al.: Tetrahymena H4 genes: structure, evolution and organization in macro- and micronuclei. Nucleic Acids Res. 12: 1961–75 (1984)
22. Barad M. et al.: A novel, tissue-specific Drosophila homeobox gene. Embo J. 7: 2151–61 (1988)
23. Basi G. S. and Storti G. V.: Structure and DNA sequence of the tropomyosin I gene from Drosophila melanogaster. J. biol. Chem. 261: 817–827 (1986)
24. Batuecas B. et al.: Genome organization of Artemia mitochondrial DNA. Nucleic Acids Res. 16: 6515–29 (1988)
25. Beccari E. et al.: Sequences coding for the ribosomal protein L14 in Xenopus tropicalis; homologies in the 5' untranslated region are shared with other r-protein mRNAs. Nucleic Acids Res. 14: 7633–46 (1986)
26. Bell J. R. et al.: A new copia-like transposable element found in a Drosophila rDNA gene unit. Nucleic Acids Res. 13: 3861–71 (1985)
27. van Beneden R. J., Watson D. K. and Chen T. T.: Cellular myc (c.myc) in fish (rainbow trout): Its relationship to other vertebrate myc genes and to the transforming genes of the MC29 family of viruses. Proc. Nat. Acad. Sci. USA 83: 3698–3702 (1986)
28. Benne R.: RNA-editing in trypanosome mitochondria (Review). Biochim. biophys. Acta 1007: 131–139 (1989)
29. Bentzen P., Leggett W. C. and Brown G. G.: Length and restriction site heteroplasmy in the mitochondrial DNA of American shad (Alosa sapidissima). Genetics 118: 509–518 (1988)
30. Berger S. L. and Kimmel A. R.: Guide to molecular cloning techniques. Acad. Press, New York 1987
31. Beridze T.: Satellite DNA. Springer, Berlin 1986
32. Berlot-Picard F., Vodjani G. and Doly J.: Isolation and characterization of a cDNA clone encoding testis protamine Z1 from the dogfish Scylliorhinus caniculus. Eur. J. Biochem. 165: 553–557 (1987)
33. Besansky N. J.: A retrotransposable element from the mosquito Anopheles gambiae. Mol. Cell Biol. 10: 863–871 (1990)
34. Biamonti G. et al.: Isolation of an active gene encoding human hnRNP protein A1. Evidence for alternative splicing. J. mol. Biol. 207: 491–503 (1989)
35. Blackburn E. H., Pan W. and Johnson C. C.: Methylation of ribosomal RNA genes in the macronucleus of Tetrahymena. Nucleic Acids Res. 11: 5131–45 (1983)
36. Blackburn E. H.: Structure and function of telomeres (Review article). Nature 350: 569–573 (1991)
37. Blackhart B. D. et al.: Structure of the human lipoprotein B gene. J. biol. Chem. 261: 15364–67 (1986)
38. Blair D. G. R.: Eukaryotic RNA polymerases. Comp. Biochem. Physiol. Pt. B 89: 647–670 (1988)
39. Bloch W.: A biochemical perspective of the polymerase chain reaction. Biochemistry 30: 2735–47 (1991)
39a. Boothroyd J. C. et al.: An unusually compact ribosomal DNA repeat in the protozoan Giardia lamblia. Nucleic Acids Res. 15: 4065–84 (1987)
40. Bosch T. C. G. et al.: Thermotolerance and synthesis of heat shock proteins. These responses are present in Hydra oligactis. Proc. Nat. Acad. Sci. USA 85: 7927–31 (1988)
41. Bosselman R. A. and Kaulenas M. S.: Ribosomal proteins of the cricket, Acheta domesticus L. Insect Biochem. 10: 129–138 (1980)
42. Boyce T. M., Zwick M. E. and Aquadro C. F.: Mitochondrial DNA in pine weevils: size, structure and heteroplasmy. Genetics 123: 825–836 (1989)
43. Bozzoni I and Beccari E.: Clustered and interspersed repetitive DNA sequences in four amphibian species with different genome size. Biochim. biophys. Acta 520: 245–252 (1978)
44. Braun T. et al.: Myf-6, a new member of the human gene family of myogenetic determination factors: evidence for a gene cluster on chromosome 12. Embo J. 9: 821–831 (1990)
45. Breathnach R. and Chambon P.: Organization and expression of eukaryotic split genes coding for proteins. Annual Rev. Biochem. 50: 349–383 (1981)
46. Breitbart R. E., Andreadis A. and Nadal-Ginard B.: Alternative splicing: A ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annual Rev. Biochem. 56: 467–495 (1987)
47. Brezinsky L. et al.: The transposable element Uhu from Hawaiian Drosophila – member of the widely dispersed class of Tc1-like transposons. Nucleic Acids Res. 18: 2053–59 (1990)
48. Brierley H. L. and Potter S. S.: Distinct characteristics of loop sequences of two Drosophila foldback transposable elements. Nucleic Acids Res. 13: 485–500 (1985)
49. Brimacombe R. and Stiege W.: Structure and function of ribosomal RNA. Biochem. J. 229: 1–17 (1985)
50. Brodsky V. Y. and Uryvaeva I. V.: Genome multiplication in growth and development. Cambridge Univ. Press, Cambridge 1985

51. Brown E. and Goodwin G. H.: Comparison of the high-mobility-group chromosomal proteins of the rainbow-trout (Salmo gairdneri) liver and testis. Biochem. J. 215: 531–538 (1983)
52. Brunk C. F. and Sadler L. A.: Characterization of the promoter region of Tetrahymena genes. Nucleic Acids Res. 18: 323–329 (1990)
53. Bryan G., Garza D. and Hartl D.: Insertion and excision of the transposable element mariner in Drosophila. Genetics 125: 103–114 (1990)
54. Bucher P. and Trifonov E. N.: Compilation and analysis of eukaryotic POL II promoter sequences. Nucleic Acids Res. 14: 10009–26 (1986)
55. Buerglin T. R. et al.: Caenorhabditis elegans has scores of homoeobox-containing genes. Nature 341: 239–243 (1989)
56. Burton F. H. et al.: Transposition of a long member of the L1 major interspersed DNA family into the mouse beta globin gene locus. Nucleic Acids Res. 13: 5071–84 (1985)
57. Bustin M., Lehn D. A. and Landsman D.: Structural features of the HMG chromosomal proteins and their genes. Biochim. biophys. Acta 1049: 231–243 (1990)
58. Campbell D. A.: Tandemly linked tRNA-Gln, tRNA-Val and tRNA-Lys genes in Trypanosoma brucei. Nucleic Acids Res. 17: 9479 (1989)
59. Campbell J. L.: Eukaryotic DNA replication. Annual Rev. Biochem. 55: 733–771 (1986)
60. Cantatore P. et al.: Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. Nature 329: 853–855 (1987)
61. Cantatore P. et al.: The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of Paracentrotus lividus. J. biol. Chem. 264: 10965–75 (1989)
62. Cavalier-Smith T. (ed.): The evolution of genome size. Wiley & Sons, Chichester 1985
62a. Caverner D. R.: Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucleic Acids Res. 15: 1353–61 (1987)
63. Cech T. R.: Self-splicing of group I introns. Annual Rev. Biochem. 59: 543–568 (1990)
64. Chan Y. L. et al.: The primary structure of rat ribosomal protein L 19. J. biol. Chem. 262: 1111–15 (1987)
65. Chan Y. L. and Wool I. G.: The primary structure of rat ribosomal protein S 20. Biochim. biophys. Acta 1049: 93–95 (1990)
66. Chang L. M. S. et al.: DNA polymerases in parasitic protozoans differ from host enzymes. Science 208: 510–511 (1980)
67. Chang L. M. S., Plevani P. and Bollum F. J.: Evolutionary conservation of DNA polymerase beta-structure. Proc. Nat. Acad. Sci. USA 79: 758–761 (1982)
68. Charlesworth B. and Langley C. H.: The population genetics of Drosophila transposable elements. Annual Rev. Genet. 23: 251–287 (1989)
69. Chen C. et al.: At least two genes reside within a large intron of the dunce gene of Drosophila. Nature 329: 721–724 (1987)
70. Chevaillier P. et al.: Amino-acid sequence of scylliorhinine Z1 and comparison of the primary structure of the protamines of the dogfish Scylliorhinus caniculus. Biochim. biophys. Acta 914: 19–27 (1987)
71. Chia W. et al.: Spontaneous excision of a large composite transposable element of Drosophila melanogaster. Nature 316: 81–83 (1985)
72. Chomyn A. et al.: URF6, last unidentified reading frame of human mtDNA, codes for an NADH dehydrogenase subunit. Science 234: 614–618 (1986)
73. Choo K. H. et al.: A survey of the genomic distribution of alpha-satellite DNA in all the human chromosomes and derivation of a new consensus sequence. Nucleic Acids Res. 19: 1179–82 (1991)
74. Chooi W. Y.: Purification of Drosophila ribosomal proteins. Isolation of proteins S8, S13, S16, S19, S20/L24, S22/L26, S24, S25/S27, S26, S29, L4, L10/11, L12, L13, L16, L18, L19, L27, 1,7/8,9, and 11. Biochemistry 19: 3469–76 (1980)
75. Chowdhury K., Rohdewohld H. and Gruss P.: Specific and ubiquitous expression of different Zn finger protein genes in the mouse. Nucleic Acids Res. 16: 9995–10011 (1988)
76. Churchill M. E. A., Tullius T. D. and Klug A.: Mode of interaction of the zinc finger protein TFIIIA with a 5S RNA gene of Xenopus. Proc. Nat. Acad. Sci. USA 87: 5528–32 (1990)
77. Citri Y. et al.: A family of unusually spliced biologically active transcripts encoded by a Drosophila clock gene. Nature 326: 42–47 (1987)
78. Clary D. O. and Wolstenholme D. R.: The mitochondrial DNA molecule of Drosophila yakuba: Nucleotide sequence, gene organization, and genetic code. J. mol. Evol. 22: 252–271 (1985)
79. Clawson E. A. et al. (eds.): Nucleic acid methylation. Wiley-Liss, New York 1990
80. Clayton D. A.: Transcription of the mammalian mitochondrial genome. Annual Rev. Biochem. 53: 573–594 (1984)
81. Cockburn A. F. and Mitchell S. E.: Repetitive DNA interspersion patterns in Diptera. Arch. Insect Biochem. Physiol. 10: 105–113 (1989)
82. Cockburn A. F., Mitchell S. E. and Seawright J. A.: Cloning of the mitochondrial genome of Anopheles quadrimaculatus. Arch. Insect Biochem. Physiol. 14: 31–36 (1990)
83. Colbert R. A. and Young D. A.: Electrophoretic separation of in vitro translation products on giant two-dimensional gels allows detailed analysis of cellular mRNAs. J. biol. Chem. 261: 14733–39 (1986)
84. Coles L. S. et al.: Characterization of the chicken histone H1 gene complement. Generation of a complete set of vertebrate H1 protein sequences. J. biol. Chem. 262: 9656–63 (1987)
85. Collier J. R.: The molecular weight of ribosomal ribonucleic acids among the protostomia. Biol. Bull. 164: 428–432 (1983)
86. Connor W. et al.: Organization of the histone genes in the rainbow trout (Salmo gairdneri). J. mol. Evol. 20: 227–235 (1984)
87. Connor W. et al.: Organization and nucleotide sequence of the rainbow trout histone H2A and H3 genes. J. mol. Evol. 20: 236–250 (1984)
88. Cool D. et al.: Histone genes in three sea star species: Cluster arrangement, transcriptional polarity, and analyses of the flanking regions of H3 and H4 genes. J. mol. Evol. 27: 36–44 (1988)
89. Cooper G.: Oncogenes. Jones and Bartlett, Boston 1990

90. Cooper T. A. and Ordahl C. P.: A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternative splicing. J. biol. Chem. 260: 11140–48 (1985)
91. Corral M. et al.: DNA-sequences homologous to mitochondrial genes in nuclei from normal rat tissues and from rat hepatoma cells. Biochem. biophys. Res. Commun. 162: 258–264 (1989)
92. Cotterill S., Chui G. and Lehman I. R.: DNA polymerase-primase from embryos of Drosophila melanogaster. The DNA polymerase subunit. J. biol. Chem. 262: 16100–104 (1987)
93. Cotterill S., Chui G. and Lehman I. R.: DNA polymerase-primase from embryos of Drosophila melanogaster. DNA primase subunits. J. biol. Chem. 262: 16105–108 (1987)
94. Coulson A. et al.: Toward a physical map of the genome of the nematode Caenorhabditis elegans. Proc.Nat. Acad. Sci. USA 83: 7821–25 (1986)
95. Craig N. L.: P element transposition. Cell 62: 399–402 (1990)
96. Cruces J. et al.: Purification and subunit structure of RNA polymerase II from different stages of Artemia development. Eur. J. Biochem. 141: 279–282 (1984)
97. de la Cruz V. F. et al.: A minimal ribosomal RNA: Sequence and secondary structure of the 9 S kinetoplast ribosomal RNA from Leishmania tarentolae. Proc. Nat. Acad. Sci. USA 82: 1401–05 (1985)
98. van Daal A. et al.: Drosophila has a single copy of the gene encoding a highly conserved histone H2A variant of the H2A.F/Z type. Nucleic Acids Res. 16: 7487–97 (1988)
99. Dame J. B. and McCutchan T. F.: The four ribosomal DNA units of the malaria parasite, Plasmodium berghei. Identification, restriction map, and copy number analysis. J. biol. Chem. 258: 6984–90 (1983)
100. Dame J. B., Sullivan M. and McCutchan T. F.: Two major sequence classes of ribosomal RNA genes in Plasmodium berghei. Nucleic Acids Res. 12: 5943–52 (1984)
101. Dandekar T. and Sibbald P. R.: Trans-splicing of pre-mRNA is predicted to occur in a wide range of organisms including vertebrates. Nucleic Acids Res. 18: 4719–25 (1990)
102. Daniels G. R. and Deininger P. L.: A second major class of alu family repeated DNA sequences in a primate genome. Nucleic Acids Res. 11: 7595–7610 (1983)
103. Daniels S. B. et al.: Evidence for horizontal transmission of the P transposable element between Drosophila species. Genetics 124: 339–355 (1990)
104. Deka N. et al.: Human transposon-like elements insert at a preferred target site: evidence for a retrovirally mediated process. Nucleic Acids Res. 16: 1143–51 (1988)
105. Delabar J. M.: Nonrandom location of H1-H1° histones on chromatin of mouse liver and brain. J. biol. Chem. 260: 12622–28 (1985)
106. Desjardins P. and Morais R.: Sequence and gene organization of the chicken mitochondrial genome. A novel gene order in higher vertebrates. J. mol. Biol. 212: 599–634 (1990)
107. Dibb N. J. and Newman A. J.: Evidence that introns arose at proto-splice sites. Embo J. 8: 2015–21 (1989)
108. Dillon L. S.: The gene. Its structure, function and evolution. Plenum Press, New York 1987
109. Dix D. B. and Thompson R. C.: Codon choice and gene expression. Synonymous codons differ in translational accuracy. Proc. Nat. Acad. Sci. USA 86: 6888–92 (1989)
110. Djondjurov L. P., Yancheva N. Y. and Ivanova E. C.: Histones of terminally differentiated cells undergo continous turnover. Biochemistry 22: 4095–4102 (1983)
111. Dod B. et al.: Concerted evolution of light satellite DNA in genus Mus implies amplification and homogenization of large blocks of repeats. Mol. Biol. Evol. 6: 478–491 (1989)
112. Dodgson J. B., Browne D. L. and Black A. J.: Chicken chromosomal protein HMG-14 and HMG-17 cDNA clones: isolation, characterization and sequence comparison. Gene 63: 287–295 (1988)
113. Doerfler W. et al.: Eukaryotic DNA methylation – facts and problems. FEBS Letters 268: 329–333 (1990)
114. Dolecki G. J., Wang G. and Humphreys T.: Stage- and tissue-specific expression of two homeo box genes in sea urchin embryos and adults. Nucleic Acids Res. 16: 11543–58 (1988)
115. Domier L. L. et al.: Drosophila viridis histone gene clusters lacking H1 coding segments. J. mol. Evol. 23: 149–158 (1986)
116. Doolittle W. F.: Selfish DNA after fourteen months. In: Dover G. A. and Flavell R. B. (eds.): Genome evolution, pp. 3–28. Acad. Press, New York 1982
117. van Doren K. and Hirsh D.: Trans-spliced leader RNA exists as small nuclear ribonucleoprotein particles in Caenorhabditis elegans. Nature 335: 556–558 (1988)
118. Dorfman D. M. et al.: The 5.8 S ribosomal RNA gene of Trypanosoma brucei: structural and transcriptional studies. Nucleic Acids Res. 13: 3533–49 (1985)
119. Dreyfuss G., Swanson M. S. and Piñol-Roma S.: Heterogenous nuclear ribonucleoprotein particles and the pathway of mRNA formation. Trends biochem. Sci. 13: 86–91 (1988)
120. den Dunnen J. T. and Schoenmakers J. G. G.: Consensus sequences of the Rattus norwegicus B1- and B2 repeats. Nucleic Acids Res. 15: 2772 (1987)
121. Eide D. and Anderson P.: Transposition of Tc1 in the nematode Caenorhabditis elegans. Proc. Nat. Acad. Sci. USA 82: 1756–60 (1985)
122. Endoh H., Nagahashi S. and Okada N.: Tetrahymena pyriformis DNA fragment with a gene cluster for 3 putative serine tRNAs and an asparagine tRNA. Nucleic Acids Res. 17: 10122 (1989)
123. Endoh H., Nagahashi S. and Okada N.: A highly repetitive and transcribable seqeunce in the tortoise genome is probably a retroposon. Eur. J. Biochem. 189: 25–31 (1990)
124. Englund P. T., Hajduk S. L. and Marini J. C.: The molecular biology of trypanosomes. Annual Rev. Biochem. 51: 695–726 (1982)
125. Epstein, D. A., Witney F. R. and Furano A. V.: The spread of sequence variants in Rattus satellite DNAs. Nucleic Acids Res. 12: 973–988 (1984)
126. Erbil C. and Niessing J.: The primary structure of the duck αD-globin gene: an unusual 5'-splice junction sequence. Embo J. 2: 1339–43 (1983)
127. Estepa I. and Pestana A.: Isolation and partial characterization of three histone-specific acetyltransferases from Artemia. Eur. J. Biochem. 132: 249–254 (1983)

128. Etter A. et al.: Eleminished chromatin of Ascaris contains a gene that endodes a putative ribosomal protein. Proc. Nat. Acad. Sci. USA 88: 5593–96 (1991)
129. Fan J. B. et al.: Giardia lamblia – Haploid genome size determined by pulsed field gel electrophoresis is less than 12-Mb. Nucleic Acids Res. 19: 1905–08 (1991)
130. Fanning T. G. and Singer M. F.: The LINE-1 DNA sequences in four mammalian orders predict proteins that conserve homologies to retrovirus proteins. Nucleic Acids Res. 15: 2251–60 (1987)
131. Fanning T. G. and Singer M. F.: LINE-1, a mammalian transposable element (Review). Biochim. biophys. Acta 910: 203–212 (1987)
132. Feagin J. E.: RNA editing in kinetoplastid mitochondria (Minireview). J. Biol. Chem. 265: 19373–76 (1990)
133. de Ferra F. et al.: Alternative splicing accounts for the four forms of myelin basic protein. Cell 43: 721–727 (1985)
134. Fifis T., Cooper D. W. and Hill R. J.: Characterisation of the protamines of the tammar wallaby (Macropus eugenii). Comp. Biochem. Physiol. Pt. B 95: 571–575 (1990)
135. Files J. G. and Hirsh D.: Ribosomal DNA of Caenorhabditis elegans. J. mol. Biol. 149: 223–240 (1981)
136. Finnegan D. J. et al.: Transposable DNA sequences. In: Dover G. A. and Flavell R. B. (eds.): Genome evolution, pp. 29–40. Acad. Press, New York 1982
137. Fischer J. A. and Maniatis T.: Regulatory elements involved in Drosophila Adh gene expression are conserved in divergent species and separate elements mediate expression in different tissues. Embo J. 5: 1275–89 (1986)
138. Fleming J. E. et al.: Aging results in an unusal expression of Drosophila heat shock proteins. Proc. Nat. Acad. Sci. USA 85: 4099–4103 (1988)
139. Fox T. D.: Natural variation in the genetic code. Annual Rev. Genet. 21: 67–91 (1987)
140. French C. K., Fouts D. L. and Manning J. E.: Sequence arrangement of the rRNA genes of the dipteran Sarcophaga bullata. Nucleic Acids Res. 9: 2563–76 (1981)
141. Fujiwara H, Kawata Y. and Ishikawa H.: Primary and secondary structure of 5.8 S rRNA from the silkgland of Bombyx mori. Nucleic Acids Res. 10: 2415–18 (1982)
142. Fujiwara H. and Ishikawa H.: Structure of the Bombyx mori rDNA: initiation site for its transcription. Nucleic Acids Res. 15: 1245–58 (1987)
143. Furia M. et al.: A new gene nested within the dunce genetic unit of Drosophila melanogaster. Nucleic Acids Res. 18: 5837–41 (1990)
144. Gabriel A et al.: A rapidly rearranging retrotransposon within the miniexon gene locus of Crithidia fasciculata. Mol. Cell Biol. 10: 615–624 (1990)
145. Gadaletta G. et al.: The complete nucleotide sequence of the Rattus norvegicus mitochondrial genome: Cryptic signals revealed by comparative analysis between vertebrates. J. mol. Evol. 28: 497–516 (1989)
146. Ganz P. R. and Pearlman R. E.: Purification from Tetrahymena thermophila of DNA polymerase and a protein which modifies its activity. Eur. J. Biochem. 113: 159–173 (1980)
147. Gardiner-Garden M and Frommer M.: CpG islands in vertebrate genomes. J. mol. Biol. 196: 261–282 (1987)
148. Garey J. A. and Wolstenholme D. R.: Platyhelminth mitochondrial DNA: Evidence for early evolutionary origin of a tRNA-Ser(AGN) that contains a dihydrouridine arm replacement loop and of serine-specifying AGA and AGG codons. J. mol. Evol. 28: 374–387 (1989)
149. Species variation in transcription factor IIIA. Nucleic Acids Res. 17: 781–794 (1989)
150. Gehring W. J. and Hiromi Y.: Homeotic genes and the homeobox. Annual Rev. Genet. 20: 147–173 (1986)
151. Gehring W. J.: Homeotic genes, the homeobox, and the spatial organization of the embryo. The Harvey Lectures Series 81, pp. 153–172 . Alan R. Liss, New York 1987
152. Gehring W. J.: Homeoboxes in the study of development. Science 236: 1245–52 (1987)
153. Georgiev G. P.: Mobile genetic elements in animal cells and their biological significance. Eur. J. Biochem. 145: 203–220 (1984)
154. Gibb G. M. and Ragan C. I.: Identification of the subunits of bovine NADH dehydrogenase which are encoded by the mitochondrial genome. Biochem. J. 265: 903–906 (1990)
155. Gilbert W., Marchionni M. and McKnight G.: On the antiquity of introns. Cell 46: 151–153 (1986)
156. Glover D. M.: The rDNA of Drosophila melanogaster. Cell 26: 297–298 (1981)
157. Goddard J. M., Weiland J. J. and Capecchi M. R.: Isolation and characterization of Caenorhabditis elegans DNA sequences homologous to the v-abl oncogene. Proc. Nat. Acad. Sci. USA 83: 2172–76 (1986)
158. Gonzalez I., Petersen R. and Sylvester J. E.: Independent insertion of alu elements in the human ribosomal spacer and their concerted evolution. Mol. Biol. Evol. 6: 413–423 (1989)
159. Gorovsky M. A.: Genome organization and reorganization in Tetrahymena. Annual Rev. Genet. 14: 203–239 (1980)
160. Goulian M. and Heard C. J.: Intact DNA polymerase alpha/primase from mouse cells. J. Biol. Chem. 264: 19407–15 (1989)
161. Graham R. W. et al.: Maturation of the major ubiquitin gene transcript in Caenorhabditis elegans involves the acquisition of a trans-spliced leader. J. Biol. Chem. 263: 10415–19 (1988)
162. Gray M. W.: The evolutionary origin of organelles. Trends Genet. 5: 294–299 (1989)
163. Green C. J., Sohel I. and Vold B. S.: The discovery of new intron-containing human tRNA genes using the polymerase chain reaction. J. Biol. Chem. 265: 12139–42 (1990)
164. Grimaldi G. and Nocera P. P. D.: Multiple repeated units in Drosophila melanogaster ribosomal DNA spacer stimulate rRNA precursor transcription. Proc. Nat. Acad. Sci. USA 85: 5502–06 (1988)
165. Grimes S. R. jr.: Nuclear proteins in spermatogenesis (Minireview). Comp. Biochem. Physiol. Pt. B 83: 495–500 (1986)
166. Gross D. S. and Garrard W. T.: Nuclease hypersensitive sites in chromatin. Annual Rev. Biochem. 57: 159–197 (1988)

167. Gruskin K. D., Smith T. F. and Goodman M.: Possible origin of a calmodulin gene that lacks intervening sequences. Proc. Nat. Acad. Sci. USA 84: 1605–08 (1987)
168. Guntaka R. V. et al.: Organization of Plasmodium falciparum genome: 1. Evidence for a highly repeated DNA sequence. Nucleic Acids Res. 13: 1965–75 (1985)
169. Gupta R. S., Picketts D. J. and Ahmad S.: A novel ubiquitous protein „chaperonin" supports the endosymbiotic origin of mitochondrion and plant chloroplasts. Biochem. biophys. Res. Commun. 163: 780–787 (1989)
170. Gut S. H. et al.: Z-DNA-binding proteins from bull testis. Nucleic Acids Res. 15: 9691–9705 (1987)
171. Guthrie C. and Patterson B.: Spliceosomal snRNAs. Annual Rev. Genet. 22: 387–419 (1988)
172. Hagemann S, Miller W. J. and Pinsker W.: P-related sequences in Drosophila bifasciata – A molecular clue to understanding of P-element evolution in the genus Drosophila. J. mol. Evol. 31: 478–484 (1990)
173. Hajduk S. L. and Cosgrove W. B.: Kinetoplast DNA from normal and dyskinetoplastic strains of Trypanosoma equiperdum. Biochim. biophys. Acta 561: 1–9 (1979)
174. Hancock K. and Hajduk S. L.: The mitochondrial transfer RNAs of Trypanosoma brucei are nuclear encoded. J. Biol. Chem. 265: 19208–15 (1990)
175. Hanyu N. et al.: Dramatic events in ciliate evolution: alteration of UAA and UAG termination codons to glutamine codons due to anticodon mutations in two Tetrahymena tRNAs-Gln. Embo J. 5: 1307–11 (1986)
176. Harada K., Yukuhiro K. and Mukai T.: Transposition rates of movable genetic elements in Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 87: 3248–52 (1990)
177. Harborne N. and Allan J.: The resolution of five linker histone subtypes from chicken erythrocytes. FEBS Letters 194: 267–272 (1986)
178. Hardesty B. and Kramer G. (eds.): Structure, function and genetics of ribosomes. Springer, Berlin 1986
179. Hardman N.: Structure and function of repetitive DNA in eukaryotes. Biochem. J. 234: 1–11 (1986)
180. Harper D. S. and Jahn C. L.: Differential use of termination codons in ciliated protozoa. Proc. Nat. Acad. Sci. USA 86: 3252–56 (1989)
181. Harris L. J., Prasad S. and Rose A. M.: Isolation and sequence analysis of Caenorhabditis briggsae repetitive elements related to the Chaenorhabditis elegans transposon Tc1. J. mol. Evol. 30: 359–369 (1990)
182. Hatfield D.: Suppression of termination codons in higher eukaryotes. Trends biochem. Sci. 10: 201–204 (1985)
183. Hawkins J. D.: A survey on intron and exon lengths. Nucleic Acids Res. 16: 9893–9908 (1988)
184. Hawkins N. C. and McGhee J. D.: Homeobox-containing genes in the nematode Caenorhabditis elegans. Nucleic Acids Res. 18: 6101–06 (1990)
185. Hayashi S. and Scott M. P.: What determines the specificity of action of Drosophila homeodomain proteins. Cell 63: 883–894 (1990)
186. Hayes T. E. and Dixon J. E.: Z-DNA in the rat somatostatin gene. J. biol. Chem. 260: 8145–56 (1985)
187. Heinze H. et al.: The primary structure of the human ribosomal protein S6 derived from a cloned cDNA. J. biol. Chem 263: 4139–44 (1988)
188. Hendricks L. et al.: Primary structures of the 5 S ribosomal RNAs of 11 arthropods and applicability of 5 S RNA to the study of metazoan evolution. J. mol. Evol. 24: 103–109 (1986)
189. Henikoff S. et al.: Gene within a gene: nested Drosophila genes encode unrelated proteins on opposite DNA strands. Cell 44: 33–42 (1986)
190. Heschl M. F. P. and Baillie D. L.: The hsp 70 multigene family of Caenorhabditis elegans (Minireview). Comp. Biochem. Physiol. Pt. B 96: 633–637 (1990)
191. Hey J.: The transposable portion of the genome of Drosophila algonquin is very different from that in D. melanogaster. Mol. Biol. Evol. 6: 66–79 (1989)
192. Hicke B. J. et al.:Two versions of the gene encoding the 41-kilodalton subunit of the telomere bindings protein of Oxytricha nova. Proc. Nat. Acad. Sci. USA 87: 1481–85 (1990)
193. Hightower et al.: Extrachromosomal elements in the lower eukaryote Leishmania. J. Biol. Chem. 263: 16970–76 (1988)
194. Hilder V. A. et al.: Histone gene number in relation to C-value in amphibians. Nucleic Acids Res. 9: 5737–46 (1981)
195. Hilder V. A., Dawson G. A. and Vlad M. T.: Ribosomal 5 S genes in relation to C-value in amphibians. Nucleic Acids Res. 11: 2381–90 (1983)
196. Hoffmann J. and Porchet M. (eds.): Biosynthesis, metabolism and mode of action of invertebrate hormones. Springer, Berlin 1984
197. van Holde H.: Chromatin. Springer, Berlin 1988
198. Holland P. W. H. and Hogan B. L. M.: Phylogenetic distribution of antennapedia-like homoeo boxes. Nature 321: 251–153 (1986)
199. Hori H. and Osawa S.: Origin and evolution of organisms as deduced from 5 S ribosomal RNA sequences. Mol. Biol. Evol. 4: 445–472 (1987)
200. Horowitz S. and Gorovsky M. A.: An unusual genetic code in Tetrahymena. Proc. Nat. Acad. Sci. USA 82: 2452–55 (1985)
201. Horowitz S. et al.: Unusual features of transcribed and translated regions of the histone H4 gene family of Tetrahymena thermophila. Nucleic Acids Res. 15: 141–160 (1987)
202. Hough-Evans B. R. et al.: Complexity of RNA in eggs of Drosophila melanogaster and Musca domestica. Genetics 95: 81–94 (1980)
203. Hough-Evans B. R. and Howard J.: Genome size and DNA complexity of Plasmodium falciparum. Biochim. biophys. Acta 698: 56–61 (1982)
204. Howe C. J. and Ward E. S. (eds.): Nucleic acid sequencing. A practical approach. Oxford University Press, New York 1990
205. Huang X. Y. and Hirsh D.: A second trans-spliced RNA leader sequence in the nematode Caenorhabditis elegans. Proc. Nat. Acad. Sci. USA 86: 8640–44 (1989)
206. Insdorf N. F. and Bogenhagen D. F.: DNA polymerase gamma from Xenopus laevis. J. Biol. Chem. 264: 21491–503 (1989)
207. Itoh N. et al.: The nucleotide sequence of cDNA for a Drosophila ribosomal protein with homology to rat ribosomal protein S26. Nucleic Acids Res. 17: 441 (1989)
208. Jacob M. and Gallinaro H.: The 5' splice site: phylogenetic evolution and variable geometry of association with U1RNA. Nucleic Acids Res. 17: 2159–80 (1989)

209. Jacobs H. T. et al.: Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. mol. Biol. 202: 185–217 (1988)
210. Jahn C. L., Krikau M. F. and Shyman S.: Developmentally coordinated en masse excision of a highly repetitive element in E. crassus. Cell 59: 1009–1018 (1989)
211. Jakubczak J. L., Burke W. D. and Eickbush T. H.: Retrotransposable elements R1 and R2 interrupt the rRNA genes of most insects. Proc. Nat. Acad. Sci. USA 88: 3295–99 (1991)
212. Jankowski J. M. et al.: In vitro expression of two proteins from overlapping reading frames in a eukaryotic DNA sequence. J. mol. Evol. 24: 61–71 (1986)
213. Jantzen K., Fritton, H. P. and Igo-Kemenes T.: The DNAse I sensitive domain of the chicken lysozyme gene spans 24 kb. Nucleic Acids Res. 14: 6085–99 (1986)
214. Johansen S., Guddal P. H. and Johansen T.: Organization of the mitochondrial genome of Atlantic cod, Gadus morrhua. Nucleic Acids Res. 18: 411–419 (1990)
215. Johnson P.F. and McKnight S. L.: Eukaryotic transcriptional regulatory proteins. Annual Rev. Biochem. 58: 799–839 (1989)
216. Jones D. et al.: Structure, organization, and expression of the 16-kDa heat shock gene family of Caenorhabditis elegans. Genome 31: 690–697 (1989)
217. Jones S. W. et al.: A Xenopus ribosomal protein S6 kinase has two apparent kinase domains that are similar to distinct protein kinases. Proc. Nat. Acad. Sci. USA 85: 3377–81 (1988)
218. Joplin K. H., Yokum G. D. and Denlinger D. L.: Cold shock elicits expression of heat shock proteins in the flesh fly, Sarcophaga crassipalpis. J. Insect Physiol. 36: 825–834 (1990)
219. Jordan B. R., Latil-Damotte M. and Jourdan R.: Coding and spacer sequence in the 5.8 S – 2 S region of Sciara ribosomal DNA. Nucleic Acids Res. 8: 3565–73 (1980)
220. Junakovic N. and Angelucci V.: Polymorphisms in the genome distribution of copia-like elements in related laboratory stocks of Drosophila melanogaster. J. mol. Evol. 24: 83–88 (1986)
221. Junakovic N. et al.: Transposition of copia-like nomadic elements can be induced by heat shock. J. mol. Evol. 24: 89–93 (1986)
222. Jurka J. and Milosavljevic: Reconstruction and analysis of human alu genes. J. mol. Evol. 32: 105–121 (1991)
223. Kadura S. N. and Krapunov S. N.: Displacement of histones by sperm-specific proteins at different stages of spermatogenesis. Eur. J. Biochem. 175: 603–607 (1988)
224. Kaguni L. S. and Lehman I. R.: Eukaryotic DNA polymerase-primase: structure, mechanism and function. Biochim. biophys. Acta 950: 87–101 (1988)
225. Kalthoff H. et al.: Ribosomal protein S6 from Xenopus laevis ovaries. Isolation, phosphorylation in vivo and cross-reaction with heterologous anti-S6 antibodies. Eur. J. Biochem. 122: 439–443 (1982)
226. Katinka M. D.: RNA-dependent DNA polymerase activity in Paramecium tetraurelia: what for? Eur. J. Biochem. 163: 569–575 (1987)
227. Kerppola T. K. and Kane C. M.: Analysis of the signals for transcription termination by purified RNA polymerase II. Biochemistry 29: 269–278 (1990)
228. Kido Y. et al.: Shaping and reshaping of salmonid genomes by amplification of transfer RNA-derived retroposons during evolution. Proc. Nat. Acad. Sci. USA 88: 2326–30 (1991)
229. King C. and Piatigorsky J.: Alternative splicing of αA-crystallin RNA. Structural and quantitative analyses of the mRNAs for the αA-2 and αA-ins-crystallin polypeptides. J. biol. Chem. 259: 1822–26 (1984)
230. Kjems J. and Garrett R. A.: An intron in the 23 S ribosomal RNA gene of the archaebacterium Desulfurococcus mobilis. Nature 318: 675–677 (1985)
231. Klein H. and Suh M.: Nucleotide sequence of a hamster cDNA highly homologous to the Xenopus laevis S19 ribosomal protein. Nucleic Acids Res. 18: 3997 (1990)
232. Kleinow W.: Structure of the mitochondrial protein synthesizing apparatus in invertebrates. Comp. Biochem. Physiol. Pt. B 91: 247–255 (1988)
233. Kmiecik D. et al.: Primary structure of the histone H2A from nucleated erythrocyte of the marine worm Sipunculus nudus. Presence of two forms of H2A in the sipunculid chromatin. Eur. J. Biochem. 135: 113–121 (1983)
234. Kmiecik D. et al.: Primary structure of the two variants of a sperm-specific histone H1 from the annelid Platynereis dumerilii. Eur. J. Biochem. 150: 359–370 (1985)
235. Knoechel W. et al.: Evolutionary conserved modules associated with zinc fingers in Xenopus laevis. Proc. Nat. Acad. Sci. USA 86: 6097–6100 (1989)
236. Kondo K. et al.: Tissue-specific and periodic changes in the nuclease sensitivity of the fibroin gene chromatin in the silkworm Bombyx mori. J. biol. Chem. 262: 5271–79 (1987)
237. Kondo K., Hodgkin J. and Waterston R. H.: Differential expression of five tRNA-Trp(UAG) amber suppressors in Caenorhabditis elegans. Mol. cell. Biol. 8: 3627–35 (1988)
238. Konopka R. J.: Genetics of biological rhythms in Drosophila. Annual Rev. Genet. 21: 227–236 (1987)
239. Kozak M.: An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. 15: 8125–48 (1987)
240. Kremer H. and Hennig W.: Isolation and characterization of a Drosophila hydei histone DNA repeat unit. Nucleic Acids Res. 18: 1573–80 (1990)
241. Krishna P. et al.: Yolk proteins from nematodes, chickens, and frogs bind strongly and preferentially to left-handed Z-DNA. J. Biol. Chem. 263: 19066–70 (1988)
242. Kubli E.: The structure and function of transfer RNA genes of higher eukaryotes. Experientia 37: 1–9 (1981)
243. Kumazaki T., Hori H. and Osawa S.: The nucleotide sequences of 5 S rRNAs from ribbon worms: Emplectonema gracile contains two 5 S rRNA species differing considerably in their sequences. Nucleic Acids Res. 11: 7141–44 (1983)
244. Kun J. and Mueller-Hill B.: The sequence of a third member of the heat shock protein family in Plasmodium falciparum. Nucleic Acids Res. 17: 5384 (1989)
245. Kunkel T. A. and Bebenek K.: Recent studies of the fidelity of DNA synthesis. Biochim. biophys. Acta 951: 1–15 (1988)

246. Lai Z. C. and Childs G.: Isolation and characterization of the gene encoding the testis specific histone protein H2B-2 from the sea urchin Lytechinus pictus. Nucleic Acids Res. 14: 6845–56 (1986)
247. Lake J. A.: Evolving ribosome structure: Domains in Archaebacteria, Eubacteria, Eocytes and Eukaryotes. Annual Rev. Biochem. 54: 507–530 (1985)
248. Lambert M. W. and Laval J. (ed.): DNA repair mechanisms and their biological implications in mammalian cells. Plenum Press, New York 1989
249. Lancillotti F. et al.: Z-DNA in transcriptionally active chromosomes. Proc. Nat. Acad. Sci. USA 84: 1560–64 (1987)
250. Landsman D. et al.: Chromosomal protein HMG-17. Complete human cDNA sequence and evidence for a multigene family. J. biol. Chem. 261: 7479–84 (1986)
251. Lapeyre B., Bourbon H. and Amalric F.: Nucleolin, the major nucleolar protein of growing eukaryotic cells: An unusual protein structure revealed by the nucleotide sequence. Proc. Nat. Acad. Sci. USA 84: 1472–76 (1987)
252. LaRoche J. et al.: Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop Placopecten magellanicus. Mol. Biol. Evol. 7: 45–64 (1990)
253. Latchman D. S.: Eukaryotic transcription factors. Biochem. J. 270: 281–289 (1990)
254. Lawton J. R., Martinez F. A. and Burks C.: Overview of the LiMB database. Nucleic Acids Res. 17: 5885–99 (1989)
255. Lebeau M. C. et al.: Ribosomal protein L27 is identical in chick and rat. Nucleic Acids Res. 19: 1337 (1991)
256. Lee J. S. et al.: Triplex DNA in plasmids and chromosomes. Gene 82: 191–199 (1989)
257. Lee M. G. S. et al.: Structure and expression of the hsp 70 gene family of Leishmania major. Nucleic Acids Res. 16: 9567–85 (1988)
258. Leff S. E., Rosenfeld M. G. and Evans R. M.: Complex transcriptional units: diversity in gene expression by alternative RNA processing. Annual Rev. Biochem. 55: 1091–1117 (1986)
259. Leipoldt M., Kellner H. G. and Stark S.: Comparative analysis of ribosomal RNA in various fish and other vertebrate species: hidden breaks and ribosomal function in phylogenetically tetraploid species of Cyprinidae. Comp. Biochem. Physiol. Pt. B 77: 769–777 (1984)
260. Leipoldt M. and Kellner M.: Ribosomal RNA structure in the diploid and the phylogenetically polyploid amphibian species Hyla and Odontophrynus. Comp. Biochem. Physiol. Pt. B 79: 181–185 (1984)
261. Leith I. R., Hay R. T. and Russell W. C.: Detection of Z DNA-binding proteins in tissue culture cells. Nucleic Acids Res. 16: 8277–89 (1988)
262. Leng T. K. C. et al.: The human heat-shock protein family. Expression of a novel heat-inducible HSP70 (HSP70B') and isolation of its cDNA and genomic DNA. Biochem. J. 267: 125–132 (1990)
263. Levine M. and Hoey T.: Homeobox proteins as sequence-specific transcription factors. Cell 55: 537–540 (1988)
264. Lieber T. et al.: A histone H1 protein in sea urchins is endoded by a poly(A)+ mRNA. Proc. Nat. Acad. Sci. USA 85: 4123–27 (1988)
265. Lindquist S. and Craig E. A.: The heat-shock proteins. Annual Rev. Genet. 22: 631–677 (1988)
266. Link C. D., Graf-Whitsel J. and Wood W. B.: Isolation and characterization of a nematode transposable element from Panagrellus redivivus. Proc. Nat. Acad. Sci. USA 84: 5325–29 (1987)
267. Linz U. and Degenhardt H.: The polymerase chain reaction (in German). Naturwissenschaften 77: 515–530 (1990)
268. Lönn U.: DNA replication in polytene chromosomes. Trends biochem. Sci. 7: 24–26 (1982)
269. Lohe A. R. and Brutlag D. R.: Identical satellite DNA sequences in sibling species of Drosophila. J. mol. Biol. 194: 161–170
270. Lohmann K. and Kraus K.: Are there insertions in the ribosomal DNA of vertebrates? Nucleic Acids Res. 13: 5145–55 (1985)
271. Lonai P. and Orr-Utreger A.: Homeogenes in mammalian development and the evolution of the cranium and central nervous system. Faseb J. 4: 1436–43 (1990)
272. Long E. O. and Dawid I. B.: Repeated genes in eukaryotes. Annual Rev. Biochem. 49: 727–764 (1980)
273. Luhrmann R., Kastner B. and Bach M.: Structure of spliceosomal snRNPs and their role in pre-messenger RNA splicing. Biochim. biophys. Acta 1087: 265–292 (1990)
274. Lund E.: Heterogeneity of human U1 snRNAs. Nucleic Acids Res. 16: 5813–26 (1988)
275. Maassen J. A. et al.: Molecular cloning and analysis of cDNA sequences for two ribosomal proteins from Artemia. The coordinate expression of genes for ribosomal proteins and elongation factor 1 during embryogenesis of Artemia. Eur. J. Biochem. 149: 609–616 (1985)
276. Macgillivray A. J. and Birnie G. D. (eds.): Nuclear structures. Isolation and characterization. Butterworths, London 1986
277. Macgregor H. C.: Big chromosomes and speciation among amphibia. In: Dover G. A. and Flavell R. B. (eds.): Genome evolution , pp. 325–341. Acad. Press, New York 1982
278. Madjar J. J. and Fournier A.: Bombyx mori L. ribosomal proteins – Resolution, nomenclature, molecular weights and invivo phosphorylation. Mol. gen. Genetics 182: 273–278 (1981)
279. Maier W. M. et al.: The lack of protamine 2 (P2) in boar and bull spermatozoa is due to mutations within the P2 gene. Nucleic Acids Res. 18: 1249–54 (1990)
280. Majewska K. et al.: A new family of dispersed, highly repetitive sequences in bovine genome. Biochim. biophys. Acta 949: 119–124 (1988)
281. Manley J. N.: Polyadenylation of mRNA precursors. Biochim. biophys. Acta 950: 1–12 (1988)
282. Markova N. G. and Markov G. G.: Complex organization of a cryptic satellite DNA in the genome of the marine invertebrate Rapana thomasiana Grosse (Gastropoda). Biochim. biophys. Acta 741: 7–14 (1983)
283. Martin S. L. et al.: Tempo and mode of concerted evolution in the L1 repeat family of mice. Mol. Biol. Evol. 2: 127–140 (1985)
284. Martin W. J.: New technologies for large-genome sequencing. Genome 31: 1073–80 (1989)
285. Martinage A. et al.: Amino acid sequence of a cysteine-rich, arginine-rich sperm protamine of the dog-fish Scylliorhinus caniculus. Biochim. biophys. Acta 831: 172–178 (1985)

286. Martinage A. et al.: Primary structure of histone H2B from gonads of the starfish Asterias rubens. Identification of an N-dimethylproline residue at the aminoterminal. Eur. J. Biochem. 147: 351–359 (1985)
287. Martindale D. W. and Taylor F. M.: Multiple introns in a conjugation-specific gene from Tetrahymena thermophila. Nucleic Acids Res. 16: 2189–2201 (1988)
288. Martinponthieu A. et al.: Cullefish sperm protamines 1. Amino acid sequences of two distinct variants. Eur. J. Biochem. 195: 611–619 (1991)
289. Mashkova T. D. et al.: The primary structure of oocyte and somatic 5 S rRNAs from the loach Misgurnus fossilis. Nucleic Acids Res. 9: 2141–51 (1981)
290. Matera A. G. et al.: Recently transposed alu repeats result from multiple source genes. Nucleic Acids Res. 18: 6019–23 (1990)
291. Maxson R. et al.: Distinct organizations and patterns of expression of early and late histone gene sets in the sea urchin. Nature 301: 120–125 (1983)
292. Mazrimas J. A. et al.: A corrected primary sequence for bull protamine. Biochim. biophys. Acta 872: 11–15 (1986)
293. McCammon J. A. and Harvey S. C.: Dynamics of proteins and nucleic acid. Cambridge Univ. Press, Cambridge 1987
294. McConkey E. H. et al.: Proposed uniform nomenclature for mammalian ribosomal proteins. Mol. gen. Genetics 169: 1–6 (1979)
295. McKay D. J., Renaux B. S. and Dixon G. H.: Rainbow trout protamines – Amino-acid sequences of six distinct proteins from a single testis. Eur. J. Biochem. 158: 361–366 (1986)
296. McKay D. J., Renaux B. S. and Dixon G. H.: Human sperm protamines – Amino-acid sequences of two forms of protamine P2. Eur. J. Biochem. 156: 5–8 (1986)
297. Meera G., Ramesh N. and Brahmachari S. K.: Z-introns in rat alpha-lactalbumin gene. FEBS Letters 251: 245–249 (1989)
298. Meyer F. et al.: UGA is translated as cysteine in pheromone-3 of Euplotes octocarinatus. Proc. Nat. Acad. Sci. USA 88: 3758–61 (1991)
299. Miles C. and Meuth M.: G-repeats: a novel hamster sine family. Nucleic Acids Res. 17: 7221–28 (1989)
300. Miloshev G. et al.: Presence of a H1°-type histone in an invertebrate: the bivalve mollusc Anodonta cygnea. Comp. Biochem. Physiol. Pt. B 82: 759–761 (1985)
301. Mizrokhi L. J. and Mazo A. M.: Evidence for horizontal transmission of the mobile element jockey between distant Drosophila species. Proc. Nat. Acad. Sci. USA 87: 9216–20 (1990)
302. Mlodzik M. et al.: The Drosophila seven-up gene, a member of the steroid receptor gene superfamily, controls photoreceptor cell fates. Cell 60: 211–224 (1990)
303. Mohr E., Trieschmann L. and Grossbach U.: Histone H1 in two subspecies in Chironomus thummi with different genome sizes: Homologous chromosome sites differ largely in their content of a specific H1 variant. Proc. Nat. Acad. Sci. USA 86: 9308–12 (1989)
304. Monk R. J., Meyuhas O. and Perry R. P.: Mammals have multiple genes for individual ribosomal proteins. Cell 24: 301–306 (1981)
305. Morin G. B. and Cech T. R.: Phylogenetic relationships and altered genome structures among Tetrahymena mitochondrial DNAs. Nucleic Acids Res. 16: 327–346 (1988)
306. Moritz C. and Brown W. M.: Tandem duplications in animal mitochondrial DNAs: Variation in incidence and gene content among lizards. Proc. Nat. Acad. Sci. USA 84: 7183–87 (1987)
307. Moriyama E. N. and Gojobori T.: Evolution of nested genes with special reference to cuticle proteins in Drosophila melanogaster. J. mol. Evol. 28: 391–397 (1989)
308. Morris J., Kushner S. R. and Ivarie R.: The simple repeat poly(dT-dG) · poly(dC-dA) common to eukaryotes is absent from Eubacteria and Archaebacteria and rare in protozoans. Mol. Biol. Evol. 3: 343–355 (1986)
309. Moses K., Ellis M. C. and Rubin G. M.: The glass gene encodes a zinc-finger protein required by Drosophila photoreceptor cells. Nature 340: 531–536 (1989)
310. Mouches C. et al.: Amplification of an esterase gene is responsible for insecticide resistance in a California Culex mosquito. Science 233: 778–780 (1986)
311. Mouchiroud D., Gautier C. and Bernardi G.: The compositional distribution of coding sequences and DNA molecules in humans and murids. J. mol. Evol. 27: 311–320 (1988)
312. Mueller F., Clarkson S. G. and Galas D. J.: Sequence of a 3.18 kb tandem repeat of Xenopus laevis DNA containing 8 tRNA genes. Nucleic Acids Res. 15: 7191 (1987)
313. Mueller M. M., Gerster T. and Schaffner W.: Enhancer sequences and the regulation of gene transcription. Eur. J. Biochem. 176: 485–495 (1988)
314. Murphy T. J. and Blumenfeld M.: Nucleotide sequence of a Drosophila melanogaster H1 histone gene. Nucleic Acids Res. 14: 5563 (1986)
315. Musich P. R. and Dykes R. J.: A long interspersed (LINE) DNA exhibiting polymorphic patterns in human genomes. Proc. Nat. Acad. Sci. USA 83: 4854–58 (1986)
316. Nagl W.: DNA endoreduplication and polyteny understood as evolutionary strategies. Nature 261: 614–615 (1976)
317. Naora H. and Deacon N. J.: Relationship between the total size of exons and introns in protein-coding genes of higher eukaryotes. Proc. Nat. Acad. Sci. USA 79: 6196–6200 (1982)
318. Naora H., Miyahara K. and Curnow R. N.: Origin of noncoding DNA sequences: Molecular fossils of genome evolution. Proc. Nat. Acad. Sci. USA 84: 6195–99 (1987)
319. Nauber U. et al.: Abdominal segmentation of the Drosophila embryo requires a hormone receptor-like protein encoded by the gap gene knirps. Nature 336: 489–492 (1988)
320. Nickel B. E. and Davie J. R.: Structure of polyubiquinated histone H2A. Biochemistry 28: 964–968 (1989)
321. Nielsen H. and Engberg J.: Sequence comparison of the rDNA introns from six different species of Tetrahymena. Nucleic Acids Res. 13: 7445–55 (1985)
322. Nielsen H. and Engberg J.: Functional intron(+) and intron(-) rDNA in the same macronucleus of the ciliate Tetrahymena pigmentosa. Biochim. biophys. Acta 825: 30–38 (1985)
323. Nietfeld W. et al.: Oocyte and somatic 5S ribosomal RNA and 5S RNA encoding genes in Xenopus tropicalis. Nucleic Acids Res. 16: 8803–15 (1988)

324. di Nocera P. P.: Close relationship between non-viral retroposons in Drosophila melanogaster. Nucleic Acids Res. 16: 4041–52 (1988)
325. di Nocera P. P. and Sakaki Y.: Lines: a superfamily of retrotransposable ubiquitous DNA elements. Trends Genet. 6: 29–30 (1990)
326. Noller H. F.: Structure of ribosomal RNA. Annual Rev. Biochem. 53: 119–162 (1984)
327. Nomoto M. et al.: Characterization of two types of histone H2B genes from macronuclei of Tetrahymena thermophila. Nucleic Acids Res. 15: 5681–97 (1987)
328. Nover L.: Molecular cell biology of the heat stress response. Part I (in German). Naturwissenschaften 77: 310–316 (1990)
329. Ogino K. et al.: What causes the aphid 28S rRNA to lack the hidden break? J. mol. Evol. 30: 509–513 (1990)
330. Ohama T. et al.: Evolution of multicellular animals as deduced from 5 S rRNA sequences: a possible early emergence of the Mesozoa. Nucleic Acids Res. 12: 5101–08 (1984)
331. Ohama T. et al.: Evolution of the mitochondrial genetic code IV. AAA as an asparagine code in some animal mitochondria. J. mol. Evol. 30: 329–332 (1990)
332. Oliva R. and Dixon G. H.: Vertebrate protamine evolution I. Sequence alignments and gene structure. J. mol. Evol. 30: 333–346 (1990)
333. Orkin S. H. and Kazanian H. H. jr.: The mutation and polymorphism of the human β-globin gene and its surrounding DNA. Annual Rev. Genet. 18: 131–171 (1984)
334. Osawa S. and Jukes T. H.: Codon reassignment (codon capture) in evolution. J. mol. Evol. 28: 271–278 (1989)
335. Pace N. R., Olsen G. J. and Woese C. R.: Ribosomal RNA phylogeny and the primary lines of evolutionary descent. Cell 45: 325–326 (1986)
336. Padgett R. A. et al.: Splicing of messenger RNA precursors. Annual Rev. Biochem. 55: 1119–50 (1986)
337. Paskewitz S. M. and Collins F. H.: Site-specific ribosomal DNA insertion elements in Anopheles gambiae and A. arabiensis: nucleotide sequence of gene-element boundaries. Nucleic Acids Res. 17: 8125–33 (1989)
338. Pauli D., Spierer A. and Tissieres A.: Several hundred base pairs upstream of Drosophila hsp23 and 26 genes are required for their heat induction in transformed flies. Embo J. 5: 755–761 (1986)
339. Paz V. et al.: The primary structure of rat ribosomal protein L26. FEBS Letters 251: 89–93 (1989)
340. Pederson D. S. et al.: Sequence organization within and flanking clusters of 5 S ribosomal RNA genes in Tetrahymena. Nucleic Acids Res. 12: 3003–21 (1984)
341. Perry K. L., Watkins K. P. and Agabian N.: Trypanosome mRNAs have unusual „cap4" structures acquired by addition of a spliced leader. Proc. Nat. Acad. Sci. USA 84: 8190–94 (1987)
342. Peterson M. G. et al.: Functional domains and upstream activation properties of cloned human TATA binding protein. Science 248: 1625–30 (1990)
343. Petitpierre E., Gatewood J. M. and Schmid C. W.: Satellite DNA from the beetle Tenebrio molitor. Experientia 44: 498–499 (1988)
344. Pietromonaco S. F., Hessler R. A. and O'Brien T. W.: Evolution of proteins in mammalian cytoplasmic and mitochondrial ribosomes. J. mol. Evol. 24: 110–117 (1986)
345. Pizon V., Cuny G. and Bernardi G.: Nucleotide sequence organization in the very small genome of a tetraodontid fish. Eur. J. Biochem. 140: 25–30 (1984)
346. Platt T.: Transcription termination and the regulation of gene expression. Annual Rev. Biochem. 55: 339–372 (1986)
347. Pont G., Degroote F. and Picard G.: Illegitimate recombination in the histone multigenic family generates circular DNAs in Drosophila embryos. Nucleic Acids Res. 16: 8817–33 (1988)
348. Porter D. C., Moy G. W. and Vacquier V. D.: The amino terminal sequence of sea urchin sperm histone H1 and its phosphorylation by egg cytosol. Comp. Biochem. Physiol. Pt. B 92: 381–384 (1989)
349. Prats E., Cornudella L. and Ruiz-Carillo A.: Nucleotide sequence of cDNA for Phi_0, a histone to protamine transition protein from sea cucumber spermatozoa. Nucleic Acids Res. 17: 10097 (1989)
350. Preugschat F. and Wold B.: Isolation and characterization of Xenopus laevis C protein cDNA: Structure and expression of a heterogeneous nuclear ribonucleoprotein core protein. Proc. Nat. Acad. Sci. USA 85: 9669–73 (1988)
351. Pritchard A. E. et al.: Nucleotide sequence of the mitochondrial genome of Paramecium. Nucleic Acids Res. 18: 173–180 (1990)
352. Ptashne M.: Gene regulation by proteins acting nearby and at a distance. Nature 322: 697–701 (1986)
353. Quentin Y.: Successive waves of fixation of B1 variants in rodent lineage history. J. mol. Evol. 28: 299–305 (1989)
354. Rabin B. A., Hawley R. S. and Chase J. W.: DNA ligase from Drosophila melanogaster embryos – Purification and physical characterization. J. biol. Chem. 261: 10637–45 (1986)
355. Rabin B. A. and Chase J. W.: DNA ligase from Drosophila melanogaster embryos – Substrate specificity and mechanism of action. J. biol. Chem. 262: 14105–111 (1987)
356. Rasy S. D. et al.: ATP-independent type II topoisomerase from trypanosomes. Proc. Nat. Acad. Sci. USA 83: 7152–56 (1986)
357. Reed R.: Protein composition of mammalian spliceosomes assembled in vitro. Proc. Nat. Acad. Sci. USA 87: 8031–35 (1990)
358. Rich A., Nordheim A. and Wang A. H.: The chemistry and biology of lefthanded Z-DNA. Annual Rev. Biochem. 53: 791–846 (1984)
359. Richard R. E. and Bogenhagen D.F.: A high molecular weight topoisomerase I from Xenopus laevis ovaries. J. Biol. Chem. 264: 4704–09 (1989)
360. Riddihough G. and Pelham H. R. B.: Activation of the Drosophila hsp27 gene promoter by heat shock and by ecdysone involves independent and remote regulatory sequences. Embo J. 5: 1653–58 (1986)
361. Rio D. C.: Molecular mechanisms regulating Drosophila P-element transposition Annual Rev. Genet. 24: 543–578 (1990)
362. Riou G. F. et al.: A type I DNA topoisomerase from Trypanosoma cruzi. Europ. J. Biochem. 134: 479–484 (1983)
363. Roberson A. E. et al.: The 5S RNA gene minichromosome of Euplotes. Nucleic Acids Res. 17: 4699–4712 (1989)

364. Roberts S. B., Emmons S. W. and Childs G.: Nucleotide sequences of Caenorhabditis elegans core histone genes. Genes for diifferent histone classes share common flanking sequence elements. J. mol. Biol. 206: 567–577 (1989)
365. Roe B. A. et al.: The complete nucleotide sequence of the Xenopus laevis mitochondrial genome. J. biol. Chem. 260: 9759–74 (1985)
366. Rogers J. H.: The roles of introns in evolution. FEBS Letters 268: 339–343 (1990)
367. Rohner K. J., Hobi R. and Kuenzle C. C.: Z-DNA-binding proteins – Identification critically depends on the proper choice of ligands. J. Biol. Chem. 265: 19112–15
368. Rooney R. J. and Harding J. D.: Processing of mammalian tRNA transcripts in vitro: different pre-tRNAs are processed along alternative pathways that contain a common rate-limiting step. Nucleic Acids Res. 14: 4849–64 (1986)
369. Rose A. M. et al.: Tc1(Hin): a form of the transposable element Tc1 in Caenorhabditis elegans. Can. J. Biochem. Cell Biol. 63: 752–756 (1985)
370. Ruan K. S. and Emmons S. W.: Precise and imprecise somatic excision of the transposon Tc1 in the nematode C. elegans. Nucleic Acids Res. 15: 6875–81 (1987)
371. Rutledge R. G., Neelin J. M. and Seligy V. L.: Isolation and expression of cDNA clones coding for two sequence variants of Xenopus laevis histone H5. Gene 70: 117–126 (1988)
372. Ryan K. A. and Englund P. T.: Replication of kinetoplast DNA in Trypanosoma equiperdum. Minicircle H strand fragments which map at specific locations. J. Biol. Chem. 264: 823–830 (1989)
373. Ryan S. C. and Dugaiczyk A.: Newly arisen DNA repeats in primate phylogeny. Proc. Nat. Acad. Sci. USA 86: 9360–64 (1989)
374. Saiga H. et al.: Determination of the transcription initiation site of Tetrahymena pyriformis rRNA using in vitro capping of 35 S pre-rRNA. Nucleic Acids Res. 10: 4223–36 (1982)
375. Sakaguchi K. and Boyd J. B.: Purification and characterization of a DNA polymerase β from Drosophila. J. biol. Chem. 260: 10406–11 (1985)
376. Sancar A. and Sancar G. B.: DNA repair enzymes. Annual Rev. Biochem. 57: 29–67 (1988)
377. Sanford, T., Prenger J. P. and Golomb M.: Purification and immunological analysis of RNA polymerase II from Caenorhabditis elegans. J. biol. Chem 260: 8064–69 (1985)
378. Santiago C. and Marzluff W. F.: Expression of the U1 RNA gene repeat during early sea urchin development: Evidence for a switch in U1 RNA genes during development. Proc. Nat. Acad. Sci. USA 86: 2572–76 (1989)
379. Sautiere P. et al.: Comparison of the amino acid sequences of human protamines HP2 and HP3 and of intermediate basic nuclear proteins HPS1 and HPS2. Structural evidence that HPS1 and HPS2 are proprotamines. J. Biol. Chem. 263: 11059–62 (1988)
380. Sawada I., et al.: Evolution of Alu family repeats since the divergence of human and chimpanzee. J. mol. Evol. 22: 316–322 (1985)
381. Schena M.: The evolutionary conservation of eukaryotic gene transcription (Review). Experientia 45: 972–983 (1989)
382. Schlesinger M. J., Santoro M. G. and Garaci E. (eds.): Stress proteins. Induction and function. Springer, Berlin 1991
383. Schmid C. W., Wong E. F. K. and Deka N.: Single copy sequences in Galago DNA resemble a repetitive human retrotransposon-like family. J. mol. Evol. 31: 92–100 (1990)
384. Schmidt M. et al.: Evidence that a major class of mouse endogenous long terminal repeats (LTRs) resulted from recombination between exogenous retroviral LTRs and LTR-like elements (LTR-IS). Proc. Nat. Acad. Sci. USA 81: 6696–7000 (1984)
385. Schmidtke J., Epplen J. T. and Engel W.: Genome analysis of Amphioxus and speculation as to the origin of contrasting vertebrate genome organization patterns. Comp. Biochem. Physiol. Pt. B 63: 455–458 (1979)
386. Schnare M. N. and Gray M. W.: Nucleotide sequence of an exceptionally long 5.8 S ribosomal RNA from Crithidia fasciculata. Nucleic Acids Res. 10: 2085–92 (1982)
387. Schughart K., Kappen C. and Ruddle F. H.: Duplication of large genomic regions during the evolution of vertebrate homeobox genes. Proc. Nat. Acad. Sci. USA 86: 7067–71 (1989)
388. Schukink R. F. and Plasterk R. H. A.: TcA, the putative transposase of the C. elegans Tc1 transposon, has an N-terminal DNA binding domain. Nucleic Acids Res. 18: 895–900 (1990)
389. Schuler L. A., Weber J. L. and Gorski J.: Polymorphism near the rat prolactin gene caused by insertion of an Alu-like element. Nature 305: 159–160 (1983)
390. Scott M. P.: Complex loci of Drosophila. Annual Rev. Biochem. 56: 195–227 (1987)
391. Sellos D., Krawetz S. A. and Dixon G. H.: Organization and complete nucleotide sequence of the core-histone-gene cluster of the annelid Platynereis dumerilii. Eur. J. Biochem. 190: 21–29 (1990)
392. Senapathy P.: Origin of eukaryotic introns: A hypothesis, based on codon distribution statistics in genes, and its implications. Proc. Nat. Acad. Sci. USA 83: 2133–37 (1986)
393. Shannon M. F. and Duke E. J.: Comparison of mitochondrial and cytoplasmic ribosomal proteins in Drosophila. Comp. Biochem. Physiol. Pt. B 81: 683–686 (1985)
394. Sharp P. A.: On the origin of RNA splicing and introns. Cell 42: 397–400 (1985)
395. Shirakawa H., Tsuda K. I. and Yoshida M.: Primary structure of non-histone chromosomal protein HMG2 revealed by the nucleotide sequence. Biochemistry 29: 4419–23 (1990)
396. Shlomai J. and Zadok A.: Kinetoplast DNA minicircles of trypanosomatids encode for a protein product. Nucleic Acids Res. 12: 8017–28 (1984)
397. Silber J. et al.: Distribution and conservation of the foldback transposable element in Drosophila. J. mol. Evol. 28: 220–224 (1989)
398. Simpson A. J. G. et al.: The arrangement of ribosomal RNA genes in Schistosoma mansoni. Identification of polymorphic structural variants. Eur. J. Biochem. 139: 41–49 (1984)
399. Singer M. F. and Skowronsky J.: Making sense out of LINES: long interspersed repeat sequences in mammalian genomes. Trends biochem. Sci. 10: 119–122 (1985)

400. Sloof P. et al.: Characterization of satellite DNA in Trypanosoma brucei and Trypanosoma cruzi. J. mol. Biol. 167: 1–21 (1983)
401. Sloof P. et al.: Further characterization of the extremely small mitochondrial ribosomal RNAs from trypanosomes: a detailed comparison of the 9 S and 12 S RNAs from Crithidia fasciculata and Trypanosoma brucei with rRNAs from other organisms. Nucleic Acids Res. 13: 4171–90 (1985)
402. Sloof P. et al.: The nucleotide sequence of a 3.2 kb segment of mitochondrial maxicircle DNA from Crithidia fasciculata containing the gene for cytochrom oxidase subunit III, the N-terminal part of the apocytochrome b gene and a possible frameshift gene; further evidence for the use of unusual initiator triplets in trypanosome mitochondria. Nucleic Acids Res. 15: 51–65 (1987)
403. Smith C. W. J., Patton J. G. and Nadal-Ginard B.: Alternative splicing in the control of gene expression. Annual Rev. Genet. 23: 527–77 (1989)
404. Smith J. L. et al.: In trypanosomes the homolog of the largest subunit of RNA polymerase II is encoded by two genes and has a highly unusual C-terminal domain structure. Cell 56: 815–827 (1989)
405. Smith M. J. et al.: Nucleotide sequence of the nine protein-coding genes and 22 tRNAs in the mitochondrial DNA of the sea star Pisaster. J. mol. Evol. 31: 195–204 (1990)
406. Smith M. W.: Structure of vertebrate genes: A statistical analysis implicating selection. J. mol. Evol. 27: 45–55 (1988)
407. So A. G. and Downey K. M.: Mammalian DNA polymerases alpha and delta: Current status in DNA replication. Biochemistry 27: 4591–95 (1988)
408. Soares M. B., Schon E. and Efstratiadis A.: Rat Line 1: The origin and evolution of a family of long interspersed middle repetitive DNA elements. J. mol. Evol. 22: 117–133 (1985)
409. Sogin M. L., Elwood H. J. and Gunderson J. H.: Evolutionary diversity of eukaryotic small-subunit rRNA genes. Proc. Nat. Acad. Sci. USA 83: 1383–87 (1986)
410. Solignac M., Monnerot M. and Mounolou J. C.: Concerted evolution of sequence repeats in Drosophila mitochondrial DNA. J. mol. Evol. 24: 53–60 (1986)
411. Sollner-Webb B. and Tower J.: Transcription of cloned eukaryotic ribosomal RNA genes. Annual Rev. Biochem. 55: 801–830 (1986)
412. Specht T., Ulbrich N. and Erdmann V. A.: Nucleotide sequence of the 5 S rRNA from the Annelida species Enchytraeus albidus. Nucleic Acids Res. 14: 4372 (1986)
413. Speckert W. et al.: Primary structure of protamine from the northern pike Esox lucius. Eur. J. Biochem. 136: 283–289 (1983)
414. van der SPEK H. et al.: RNA editing in transcripts of the mitochondrial genes of the insect trypanosome Crithidia fasciculata. Embo J. 9: 257–262 (1990)
415. Spencer C. A., Gietz R. D. and Hodgetts R. B.: Overlapping transcription units in the dopa decarboxylase region of Drosophila. Nature 322: 279–281 (1986)
416. Spirin A. S. and Ajtkhozhin M. A.: Informosomes and polyribosome-associated proteins in eukaryotes. Trends biochem. Sci. 10: 162–165 (1985)
417. Spradling A. C. and Rubin G. M.: Drosophila genome organization – Conserved and dynamic aspects. Annual Rev. Genet. 15: 219–264 (1981)
418. Spradling A. C.: The organization and amplification of two chromosomal domains containing Drosophila chorion genes. Cell 27: 193–201 (1981)
419. Stacey S. N. et al.: Distribution and conservation of mobile elements in the genus Drosophila. Mol. Biol. Evol. 3: 522–534 (1986)
420. Stark G. R. and Wahl G. M.: Gene amplification. Annual Rev. Biochem. 53: 447–491 (1984)
421. Stone E. M. and Schwartz R. D. (eds.): Intervening sequences in evolution and development. Oxford University Press, New York 1990
422. Strayer D. et al.: Three organizations of human DNA. Proc. Nat. Acad. Sci. USA 80: 4770–74 (1983)
423. Struhl K.: Helix-turn-helix, zinc-finger, and leucine-zipper motifs for eukaryotic transcriptional regulatory proteins (Review). Trends biochem. Sci. 14: 137–140 (1989)
424. Stumph W. E. et al.: Genomic structure and possible retroviral origin of the chicken CRI repetitive DNA sequence family. Proc. Nat. Acad. Sci. USA 81: 6667–71 (1984)
425. Sturm N. R. and Simpson L.: Kinetoplast DNA minicircles encode RNAs for editing of cytochrome oxidase subunit III mRNA. Cell 61: 879–884 (1990)
426. Suzuki K. and Wool I. G.: The primary structure of rat ribosomal protein-L17. Biochem. biophys. Res. Commun. 178: 322–328 (1991)
427. Syvaeoja J. et al.: DNA polymerase alpha, delta and epsilon: Three distinct enzymes from HeLa cells. Proc. Nat. Acad. Sci. USA 87: 6664–68 (1990)
428. Syvanen M.: The evolutionary implications of mobile genetic elements. Annual Rev. Genet. 18: 271–293 (1984)
429. Tanabe K. et al.: Structural homology of DNA polymerase β from various mammalian cells. J. biol. Chem. 256: 3098–3102 (1981)
430. Taparowsky E. J. and Gerbi S. A.: Structure of 1.711 bovine satellite DNA: evolutionary relationship to satellite I. Nucleic Acids Res. 10: 5503–15 (1982)
431. Tausta S. L. and Klobutcher L. A.: Internal eliminated sequences are removed prior to chromosome fragmentation during development in Euplotes crassus. Nucleic Acids Res. 18: 845–853 (1990)
432. Tautz D. and Renz M.: Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucleic Acids Res. 12: 4127–38 (1984)
433. Tautz D. et al.: Complete sequences of the rRNA genes of Drosophila melanogaster. Mol. Biol. Evol. 5: 366–376 (1988)
434. Taylor J. D., Wellman S. E. and Marzluff W. F.: Sequences of four mouse histone H3 genes: Implications for evolution of mouse histone genes. J. mol. Evol. 23: 242–249 (1986)
435. Tebabi P., Halleux S. and Pays E.: Nucleotide sequence of a full-length cDNA coding for the ribosomal L44 protein of Trypanosoma brucei. Nucleic Acids Res. 18: 2809 (1990)
436. Thackeray J. R. and Kyriacou C. P.: Molecular evolution in the Drosophila yakuba period locus. J. mol. Evol. 31: 389–401 (1990)
437. Thomann H. U. et al.: Genes, variant genes and pseudogenes of the human tRNA-Val gene family. Expression and pre-tRNA maturation in vitro. J. mol. Biol. 209: 505–523 (1989)

438. Thomas J. et al.: The spliceosomal snRNAs of Caenorhabditis elegans. Nucleic Acids Res. 18: 2633–42 (1990)
439. Tönjes R. and Doenecke D.: A highly conserved sequence in H1 histone genes as an oligonucleotide hybridization probe: Isolation and sequence of a duck H1 gene. J. mol. Evol. 25: 361–370 (1987)
440. Trabuchet G. et al.: Recent insertion of an alu sequence in the beta-globin gene cluster of the gorilla. J. mol. Evol. 25: 288–291 (1987)
441. Tschudi C., Krainer A. R. and Ullu E.: The U6 small nuclear RNA from Trypanosoma brucei. Nucleic Acids Res. 16: 11375 (1988)
442. Tsuzuki T. et al.: Structure of the human prealbumin gene. J. biol. Chem. 260: 12224–27 (1985)
443. Turner P. C. et al.: The organization and expression of histone genes from Xenopus borealis. Nucleic Acids Res. 16: 3471–85 (1988)
444. Udvardy A. and Schedl P.: Structural polymorphism in DNA. J. mol. Biol. 166: 159–181 (1983)
445. Ullu E. and Tschudi C.: Alu sequences are processed 7SL genes. Nature 312: 171–172 (1984)
446. Upholt W. B. and Sandell L. J.: Exon/intron organization of the chicken type II collagen gene: Intron size distribution suggests a minimal intron size. Proc. Nat. Acad. Sci. USA 83: 2325–29 (1986)
447. Ursi D., Vandenberghe A. and de Wachter R.: The sequence of the 5.8 S ribosomal RNA of the crustacean Artemia salina. With a proposal for a general secondary structure model for 5.8 S ribosomal RNA. Nucleic Acids Res. 10: 3517–30 (1982)
448. Vahidi H. et al.: Unusual sequences, homologous to 5S RNA, in ribosomal DNA repeats of the nematode Meloidogyne arenaria. J. mol. Evol. 27: 222–227 (1988)
449. Vanderspek H. et al.: Conserved genes encode guide RNAs in mitochondria of Crithidia fasciculata. Embo J. 10: 1217–24 (1991)
450. Vanfleteren J. R. and Meheus L. A.: Analysis of the chromosomal proteins of Caenorhabditis elegans by two-dimensional electrophoresis, silver staining and immunodetection. Comp. Biochem. Physiol. Pt. B 91: 103–110 (1988)
451. Vanfleteren J. R., Vanbun S. M. and Vanbeeumen J. J.: The histones of Caenorhabditis elegans – No evidence of stage-specific isoforms- An overview. FEBS Letters 257: 233–237 (1989)
452. Vanin E. F.: Processed pseudogenes. Characteristics and evolution. Annual Rev. Genet. 19: 253–272 (1985)
453. Vashakidze R. P. et al.: Nuclear proteins from Drosophila melanogaster embryos which specifically bind to simple homopolymeric sequences poly-(dT-dG) · (dC-dA). Nucleic Acids Res. 16: 4989–94 (1988)
454. Vaury C. et al.: Molecular characteristics of the heterochromatic I elements from a reactive strain of Drosophila melanogaster. J. mol. Evol. 31: 424–431 (1990)
455. Venturini G., D'Ambrogi R. and Capanna E.: Size and structure of the bird genome – I. DNA content of species of Neognathae. Comp. Biochem. Physiol. Pt. B 85: 61–65 (1986)
456. Venturini G., Capanna E. and Fontana B.: Size and structure of the bird genome – II. Repetitive DNA and sequence organization. Comp. Biochem. Physiol. Pt. B 87: 975–979 (1987)
457. Viel A. et al.: Structural and functional properties of thesaurin a (42Sp50), the major protein of the 42-S particles present in Xenopus laevis previtellogenic oocytes. J. Biol. Chem. 266: 10392–399 (1991)
458. Volloch V., Schweitzer B. and Rits S.: Uncoupling of the synthesis of edited and unedited COIII RNA in Trypanosoma brucei. Nature 343: 482–484 (1990)
459. Vonheijne G.: Computer analysis of DNA and protein sequences (Review). Eur. J. Biochem. 199: 253–256 (1991)
460. Vossbrinck C. R. et al.: Ribosomal RNA sequence suggests Microsporidia are extremely ancient eukaryotes. Nature 326: 411–414 (1987)
461. Voytas D. F. and Ausubel F. M.: A copia-like transposable element family in Arabidopsis thaliana. Nature 336: 242–244 (1988)
462. Wada K. et al.: Codon usage tabulated from the GenBank genetic sequence data. Nucleic Acids Res. 19 (Suppl.): 1981–86 (1991)
463. Walldorf U., Fleig R. and Gehring W. J.; Comparison of homeobox-containing genes of the honeybee and Drosophila. Proc. Nat. Acad. Sci. USA 86: 9971–75 (1989)
464. Walsh J. B.: How many processed pseudogenes are accumulated in a gene family? Genetics 110: 345–364 (1985)
465. Wang B. et al.: Overlapping transcriptional units of the same strand within the murine β-glucuronidase gene complex. J. Biol. Chem. 263: 15841–44 (1988)
466. Wang J. C.: DNA topoisomerases – Why so many? (Minireview). J. Biol. Chem. 266: 6659–62 (1991)
467. Wang S. W. et al.: Inverted duplication of histone genes in chicken and disposition of regulatory sequences. Nucleic Acids Res. 13: 1369–87 (1985)
468. Watson J. D.: The human genome project: Past, present, and future. Science 248: 44–49 (1990)
469. Wedeen C. J. et al.: Evidence for a new family of evolutionary conserved homeobox genes. Nucleic Acids Res. 18: 1908 (1990)
470. Weiner A. M., Deininger P. L. and Efstradiadis A.: Nonviral retroposons. Annual Rev. Biochem. 55: 631–661 (1986)
471. Wells D., Bains W. and Kedes L.: Codon usage in histone gene families of higher eukaryotes reflects functional rather than phylogenetic relationships. J. mol. Evol. 23: 224–241 (1986)
472. Wernette C. M., Conway M. C. and Kaguni L. S.: Mitochondrial DNA polymerase from Drosophila melanogaster embryo: Kinetics, processivity, and fidelity of DNA polymerization. Biochemistry 27: 6046–54 (1988)
473. Wettenhall R. E. H., Nick H. P. and Lithgow T.: Primary structure of mammalian ribosomal protein S6. Biochemistry 27: 170–177 (1988)
474. White E. M. and Gorovsky M. A.: Localization and expression of mRNA for a macro-nuclear-specific histone H2A variant (hv1) during the cell cycle and conjugation in Tetrahymena thermophila. Mol. cell. Biol. 8: 4780–86 (1988)
475. Wichman H. A., Potter S. S. and Pine D. S.: Mys, a family of mammalian transposable elements isolated by phylogenetic screening. Nature 317: 77–81 (1985)
476. Willard C., Nguyen H. T. and Schmid C. W.: Existence of at least three distinct alu subfamilies. J. mol. Evol. 26: 180–186 (1987)

477. Williams T. and Fried M.: A mouse locus at which transcription from both DNA strands produces mRNAs complementary at their 3' ends. Nature 322: 275–279 (1986)
478. Wilson S., Abbots J. and Widen S.: Progress toward molecular biology of DNA polymerase β. Biochim. biophys. Acta 949: 149–157 (1988)
479. Winkfein R. J. et al.: Histone H4 and H2B genes in rainbow trout (Salmo gairdneri). J. mol. Evol. 22: 1–19 (1985)
480. Witney F. R. and Furano A. V.: Highly repeated DNA families in the rat. J. biol. Chem. 259: 10481–92 (1984)
481. Wolstenholme D. R. and Clary D. O.: Sequence evolution of Drosophila mitochondrial DNA. Genetics 109: 725–744 (1985)
482. Wolstenholme D. R. et al.: Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Nat. Acad. Sci. USA 84: 1324–28 (1987)
483. Wright J. A. et al.: DNA amplification is rare in normal human cells. Proc. Nat. Acad. Sci. USA 87: 1791–95 (1990)
484. Wu M. et al.: An intervening sequence in an unusual histone H1 gene of Tetrahymena thermophila. Proc. Nat. Acad. Sci. USA 83: 8674–78 (1986)
485. Wu R. S. et al.: Histones and their modifications. CRC Crit. Rev. Biochem. 20: 201–264 (1986)
486. Wuilmart C. and Wyns L.: An evolutionary scheme for the histones as derived from a study of internal repetitions and homologies among the different classes. J. theor. Biol. 65: 231–252 (1977)
487. Wyckoff E. et al.: Structure of the Drosophila DNA topoisomerase II gene. Nucleotide sequence and homology among topoisomerases II. J. mol. Biol. 205: 1–13 (1989)
488. Xiong Y. and Eickbush T. H.: Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. 5: 675–690 (1988)
489. Yager L. N., Kaumeyer J. F. and Weinberg E. S.: Evolving sea urchin histone genes – Nucleotide polymorphisms in the H4 gene and spacers of Strongylocentrotus purpuratus. J. mol. Evol. 20: 215–226 (1984)
490. Yioshioka K. et al.: Virus-like particle formation of Drosophila copia through autocatalytic processing. Embo J. 9: 535–541 (1990)
491. Zakian V. A.: Structure and function of telomeres. Annual Rev. Genet. 23: 579–604 (1989)
492. Zalenskaya I. A., Zalenskaya E. O. and Zalensky A. O.: Basic chromosomal proteins of marine invertebrates – II. Starfish and Holothuria. Comp. Biochem. Physiol. Pt. B 65: 375–378 (1980)
493. Zwierzynski T. A., Widmer G. and Buck G. A.: In vitro 3' end processing and poly (A) tailing of RNA in Trypanosoma cruzi. Nucleic Acids Res. 17: 4647–60 (1989)
494. Sequences Supplement. Nucleic Acids Res. 18: 2215–87 (1990)

3 The Structural Variety and Metabolism of Proteins

3.1 Structural Variety

Individual eukaryote cells contain in the order of 10^4 different proteins, and each animal species contains an even greater number due to differences between the tissues of an individual and between the individuals themselves; furthermore, the protein spectrum changes during the course of development. The number of different proteins to be found in extant organisms may be as high as 10^{12}. The description of this variety, its origin and biological significance is the most extensive theme in comparative biochemistry. This chapter will concern itself with the possibilities for structural variation and the general metabolism of proteins; further chapters will deal with comparative studies of individual proteins.

3.1.1 Protein Constituents

The primary products of protein biosynthesis are linear polymers of the **20 standard amino acids**. There has been much speculation about why the α-amino acids, out of the many possible amino-carbonic acids, should be used in the construction of proteins; why only those particular 20; and why only in the L-configuration. Undoubtedly, this has to do with the changing relationships between the evolution of amino acid metabolism and that of the genetic code, about which there is as yet no general agreement [245]. The analysis of amino acid composition is part of the routine work of the protein chemist, and innumerable tables of data are now available. Despite the considerable differences between individual proteins, the spectrum of the protein-bound amino acids varies much less than that of the free amino acids (Table 3.1). The number of codons in the genetic code for individual amino acids varies between one and six. If each of the 61 amino acid codons occurred at the same frequency, then each amino acid would be present in the total protein in proportion to the number of its codons. In actual fact, lysine, aspartic acid and glutamic acid in particular occur more frequently than would be expected; arginine, on the other hand, is markedly under-represented. This may be because arginine appeared as a protein constituent quite late in evolution (the "intruder hypothesis" of Jukes), or because a previously larger arginine fraction has, for some reason, been selected against in the course of evolution (the "selection hypothesis" of Wallis), e.g. because the peptide bonds involving arginine are preferred sites for protease attack, or the extremely basic property of arginine, in contrast to lysine, cannot be later modified by substitution. Although the basic amino acids arginine and lysine together have eight codons and the acidic amino acids aspartic acid and glutamic acid have only four, they are

Table 3.1. The spectrum of protein-bound amino acids (the fractions of individual amino acids in mol %)

	1	2	3	4	5a	5b	5c	5d
Ala	6.6	7.4	8.7	4.5	8.5	8.1	16.4	11.5
Arg	9.8	4.2	3.9	3.0	6.3	4.2	26.9	15.0
Asn	3.3	4.4	4.9	4.0	10.8	12.3	1.2	2.1
Asp	3.3	5.9	5.9	3.5			0.6	1.7
Cys	3.3	3.3	2.5	0	1.9	0.7	0.1	0.2
Gln	3.3	3.7	4.3	7.0	14.6	13.5	7.6	4.8
Glu	3.3	5.8	5.7	12.5			5.3	5.2
Gly	6.6	7.4	7.7	4.5	8.9	10.6	33.1	16.6
His	3.3	2.9	2.1	2.5	1.9	2.3	0.6	1.6
Ile	4.9	3.8	5.6	5.5	4.5	3.8	0.5	2.2
Leu	9.8	7.6	7.4	8.5	8.2	7.7	1.0	5.8
Lys	3.3	7.2	6.9	7.0	7.9	5.1	0.8	8.7
Met	1.6	1.8	1.6	2.5	2.4	1.6	0.3	2.0
Phe	3.3	4.0	3.8	4.0	3.4	3.4	0.3	3.1
Pro	6.6	5.0	3.8	8.5	3.5	5.2	1.1	2.8
Ser	9.8	8.1	6.9	8.0	5.9	6.6	1.8	5.6
Thr	6.6	6.2	5.9	2.5	4.9	6.0	0.9	3.8
Trp	1.6	1.3	1.3	1.0	–	Tr	–	Tr
Tyr	3.3	3.3	4.3	5.0	2.8	2.8	0.5	2.8
Val	6.6	6.8	6.9	5.5	5.2	5.6	0.9	4.7

1, According to the genetic code; 2, in the "average protein" according to King and Jukes; 3, the mean of 189 proteins from 81 protein families [110]; 4, in the bovine α-casein, which is frequently used as a standard protein in nutrition experiments; 5, in the crayfish *Astacus leptodactylus* [168]: a, protein-bound amino acids in abdominal muscle; b, protein-bound amino acids in the mid-gut gland; c, free amino acids in abdominal muscle; d, free amino acids in the mid-gut gland. Tr, trace

found in about the same proportions in the total protein [110]. Abnormal amino acid compositions of individual proteins always have adaptive value, in that they determine, for example, the surface charge, solubility or sensitivity to protease attack.

Most proteins contain other constituents in addition to the amino acids; however, the distinction from the early days of protein research between "proteid" and "protein", based upon the presence or absence of non-amino acid components, is now obsolete. Thus, membrane proteins and secreted proteins usually carry oligosaccharide chains, i.e. they are glycoproteins; many proteins are more or less tightly associated with metal ions, coenzymes, prosthetic groups, nucleic acids, lipids or pigments. Even amongst the amino acid components, the variety extends beyond the 20 standard residues. Although the aminoacyl-tRNAs are specific for the 20 standard amino acids, chemical modification of the amino acids is, in principle, possible at all times, but occurs mostly in the post-translational stages. In this way, more than 100 further amino acid components can arise through substitution or conjugation. The modified amino acids are released during degradation of proteins and can not be reincorporated. As most such amino acids are limited to particular protein types, they are termed **"rare" amino acids**, although they may be quite ubiquitous [248].

Post-translational changes can affect the terminal amino acids or the side-chains of internal regions. Modification of the **terminal carboxyl group** by amidation is relatively rare; it occurs, for example in some proteotoxins and peptide hormones. Methylation of the carboxyl group with adenosylmethionine as the donor is widely found, at least in the vertebrates. This increases the exocytosis of secreted proteins, marks the D-aspartic acid residues in ageing proteins, and reduces the enzyme-activating properties of calmodulin; it is still not clear how far these processes are important for the regulation of metabolism in living cells [287]. In many cytoplasmic proteins, the **terminal amino-group** is blocked with an acyl residue and is thus protected from aminopeptidase attack. Acetyl residues are widely found, especially on terminal serine and alanine residues, less frequently on threonine or glycine, and never on valine, cysteine or proline. Other acyl residues are found at the N-terminus: e.g. N-myristylglycine is found in the cAMP-stimulated protein kinases, and N-formylglycine is found in the melittin of the bee toxin [85, 177]. Pyroglutamyl residues arise by "internal acylation" (Fig. 3.1), e.g. in several peptide hormones. N-terminal

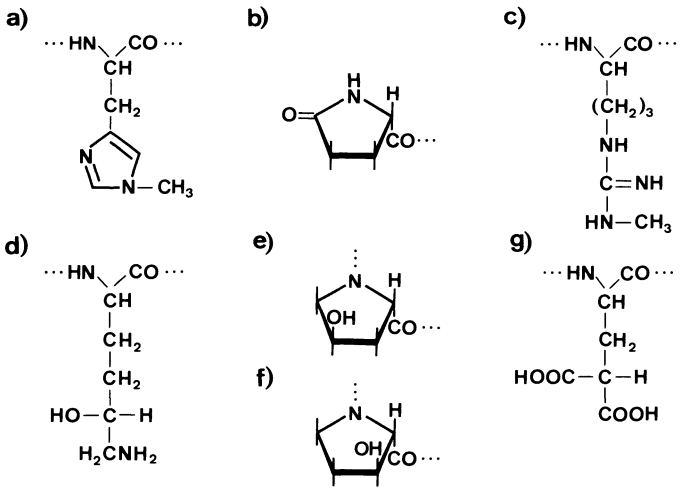


Fig. 3.1a–g. Post-translationally modified, protein-bound amino acids.
a N^T-methylhistidine; **b** pyroglutamic acid; **c** N^{ω}-methylarginine; **d** δ-hydroxylysine; **e** 4-hydroxyproline; **f** 3-hydroxyproline; **g** γ-carboxylglutamic acid

methylation has seldom been detected so far and is, in any case, undetectable with routine methods: N-trimethylalanine in histone H2B of the ciliate *Tetrahymena* and in the myosin chain A1 of rabbit skeletal muscle; N-dimethylproline in histone H2A of the starfish *Asterias rubens* and in cytochrome c-557 of the flagellate *Crithidia oncopelti* [169, 293].

Side-chain modifications are known for all protein-bound amino acids and include, for example, methylation, acylation, phosphorylation, oxidation, hydroxylation, halogenation, glycosylation or ADP-ribosylation, as well as covalent binding to non-amino acid components. There is almost no comparative biochemical information on the enzymes involved [75]; the single exception concerns the protein kinases and protein phosphatases, to which a separate section will be devoted. Methylation of amino acids is almost as widely found as phosphorylation (Fig. 3.1). A classical example of a multiply methylated protein is rabbit myosin, where two N^T-methylhistidines, two ε-N-monomethyllysines and four ε-N-trimethyllysines are found. An ε-N-trimethyllysine occurs at position 115 of most calmodulins, and three such residues are found in the elongation factor EF-1 of the brine shrimp *Artemia salina* [104]. Hydroxylated amino acids, such as δ-hydroxylysine or 4- and 3-hydroxyproline (Fig. 3.1), are typical of collagens, but also occur in other proteins. A curiosity that can be introduced here is that amongst the 22 amino acids of a small toxic peptide in the sea-snail *Conus geographus*, there are no fewer than three hydroxyproline residues (see Fig. 9.4b, p. 326). The prolylhydroxylases require molecular oxygen and ascorbic acid; the prolyl-4-hydroxylases of man, the chicken and the mussel *Mytilis edulis* are heterotetramers of two α-chains of 60–64 kDa and two β-chains of 57–60 kDa. Surprisingly, the β-subunit is identical to the enzyme protein-disulphide isomerase and a cellular thyroxine-binding protein [103, 171].

Tyrosine residues are particularly reactive. Thus, a whole series of halogenated tyrosines are found in the scleroproteins of marine invertebrates, e.g. 3-chloro-, 3,5-dichloro-, 3-bromo-, 3,5-dibromo-, 5-bromo-3-chloro-, 3-iodo-, and 3,5-diiodotyrosine; the two iodo-amino acids are also intermediates in the biosynthesis of the thyroid hormone. Tyrosine-O-sulphoproteins are detected in all cells, and have a cell-specific electrophoretic pattern [107]. The compounds methionine sulphoxide, ornithine and citrulline, which were previously only known from the pool of free amino acids, are now known for certain to be constituents of native proteins [293]; thus, the shell ligament protein (abductine) of several marine mussels contains 20–25 % methionine sulphoxide [141]. The late alterations of protein-bound amino acids also include the conversion of the aspartic acid residues of long-lived proteins into the D-isomer at the rate of about 0.1 % per year. A methyltransferase specific for protein-bound D-aspartic acid is ubiquitous in the eukaryotes; the isomeric alteration of aspartate, giving a non-functional protein, possibly represents a signal for the proteolytic destruction of such proteins [19?].

The **biological significance** of the post-translational modification of protein-bound amino acids is in most cases not known. Some derivatives serve as cross-bridges to stabilize the 3-D structure: Examples include the disulphide bridges arising by the oxidation of cysteine resi-

dues; the phosphodiester bonds; the thioether bonds of lanthionines or lysinolanines in α-keratin; the cross-bridges in collagen and elastin, formed by the oxidative deamination of lysine or hydroxylysine (see Fig. 11.3, p. 379); and the bis- and tertyrosines of resilins and other scleroproteins, which arise by the phenolic coupling of tyrosine residues (see Fig. 11.4, p. 380). Most phosphorylations, acetylations and methylations probably have a regulatory function; glycosylation serves not only for cell recognition but also for the marking of proteins and the regulation of protein degradation. Other modified protein amino acids have special functions, e.g. γ-carboxylglutamic acid (Fig. 3.1) for Ca^{2+} binding, or iodotyrosine in the synthesis of the thyroid hormone [293]. The variable post-translational glycosylation observed in some homologous proteins of related species raises the possibility that some modifications are only the coincidental result of the activities of available enzymes, and in particular cases have no special functional significance.

3.1.2 Phosphorylation and Dephosphorylation

The phosphorylation of the serine, threonine and tyrosine residues of cellular proteins by specific **protein kinases** plays a central role in the regulation of metabolism and differentiation. Several protein kinases may often be found to act sequentially in the form of a cascade; the effect of regulatory signals is amplified and dispersed through a network of protein phosphorylation and dephosphorylation reactions, the complexity of which is by no means fully understood [63]. The first protein kinases to be isolated and characterized were phosphorylase kinase in 1959 and a cAMP-regulated protein kinase in 1968. In the last 10 years, the number of known protein kinases has grown at an ever-increasing rate, particularly since it was discovered that the products of many cellular oncogenes belong to the protein tyrosine kinases. By 1987, there were already 50 protein serine/threonine kinases and 29 protein tyrosine kinases known in mammals, a total of 18 protein kinases known in *Drosophila*, and 14 in baker's yeast. It is possible that the mammalian genome codes for up to 1000 different protein kinases [37, 115].

All known protein kinases, including the oncogene products, show significant sequence similarity in their catalytic regions, and thus belong to the same protein super-family [96]. Particular sequence motifs are always found and these can be used for the identification of new protein kinases: at the N-terminus one finds the sequence GXGXXG-, followed 15–20 amino acids further on by a lysine residue; these amino acids belong to the ATP binding site. At a distance 80–180 amino acids closer to the C-terminus, there is a conservative region with the typical sequences RDL, DFG and APE, and this facilitates a more detailed classification of the protein kinases. Between the motifs DFG and APE there is always an autophosphorylatable amino acid; in this region a tyrosine surrounded by acidic amino acids distinguishes the protein as a protein tyrosine kinase [115]. Various types of protein kinase occur both as particle-bound and free in the cytoplasm [40, 217]. The phosphorylation of cell proteins can be reversed by the action of protein phosphatases. Although particular protein phosphatases are regulated, the majority are, in fact, always active. Therefore, the degree of phosphorylation of individual proteins, and thus their functional status, is mainly determined by the regulated activity of the protein kinases.

Many protein kinases are regulated by secondary messengers and are hence closely involved in signal transduction through the cell membrane. The **regulators** of protein serine/threonine kinases are, on the hand, the cyclic nucleoside-3′,5′-monophosphates cAMP and cGMP and, on the other hand, calcium ions in association with the protein calmodulin, with phospholipids and diacylglycerins (protein kinase C), or also as free ions. Recently, AMP-dependent protein kinases have been discovered that apparently play an important role in the regulation of lipid metabolism [25]. There are, however, a large number of protein serine/threonine kinases for which no specific regulators are known. Protein tyrosine kinases function as membrane receptors for insulin and various growth factors, and arise as products of various retroviruses; they may also be found not bound to membrane receptors in normal cells [72, 115, 148]. In only a few instances are the **physiological substrates** of the different protein kinases identifiable. The majority of the protein kinases show broad, overlapping substrate specificity in vitro, and it is often found that different protein kinases phosphorylate the same protein but at different positions. Many protein kinases can phosphorylate themselves (autophosphorylation) and thus influence their own substrate affinity [252].

The protein kinases that are activated by cyclic nucleoside monophosphates normally prefer one

of the two secondary messengers, but are mostly not absolutely specific; there are in fact enzymes in the insects that show equal affinity for cAMP and cGMP. **cAMP-specific protein kinases** are widely distributed from the mammals to yeast, but appear to be missing from prokaryotes and higher plants [274]. Their molecules consist of two each of the regulatory (R) and catalytic (C) subunits. On activation, which apparently requires four cAMPs per enzyme molecule, the subunits dissociate according to the following equation

$$R_2C_2 + 4\ cAMP = R_2(cAMP)_4 + 2\ C.$$

Two types of cAMP-specific protein kinases (I and II) are known in the mammals and differ mainly in the R subunits. The ratio of the two types is species and tissue specific: I and II occur in the ratio 4:1 in rabbit skeletal muscle and in the heart muscle of the mouse and rat, but occur in the ratio 1:1 in the heart muscle of the rabbit; only II is found in guinea-pig or bovine heart muscle. The C subunits show a lower rate of evolution; the mouse (351 aa) and bovine (349 aa) heart C-polypeptides agree in 98% of their amino acids. The N-terminus of the C-polypeptide is blocked by myristic acid; the C-chains are very similar in protein kinases I and II but are not identical. In spite of their structural and functional differences, the subunits R_I and R_{II} show significant sequence agreement. Their polypeptide chains are organized in four domains, the first of which is involved in RR dimerization, the second in reactions with the C-polypeptide, and the last two in cAMP binding. cAMP reduces the R-C affinity by about 10000-fold; this involves both cAMP-binding sites. All subunits carry phosphate residues, the significance of which is unknown [63].

Like those of the mammals, the cAMP-dependent protein kinases of the invertebrates are tetramers of two catalytic and two regulatory subunits which dissociate on activation. Also like the mammals, the insects *Drosophila melanogaster* and *Manduca sexta* and the nematode *Caenorhabditis elegans* contain two types of enzyme, which in the case of the flies are coded by different genes, and in the nematodes arise by alternate splicing of the transcript of one gene. The C-polypeptide of *Drosophila* (352 aa) shows 78% agreement with the mammalian enzyme, whereas that of *Caenorhabditis* (374 aa) shows 82%; the R-polypeptide of the nematode is 69% similar in the C-terminal region (positions 145–375) but much less so at the N-terminus [40, 74, 89, 161]. In the Protozoa, cAMP-regulated protein kinases have been detected in the malarial agents *Plasmodium berghei* and *P. chabaudi* and, after initial failures, also in the flagellate *Trypanosoma cruzi* [281].

cGMP-regulated protein serine/threonine kinases in mammals are found especially in smooth muscle, lung, intestine and heart. They are homodimers of about 155 kDa; unlike the cAMP-dependent kinases, activation by cGMP does not result in dissociation. The sequence of 670 amino acids in bovine subunits can be subdivided into six segments (A–F) or four domains (A, B/C, D/E and F): A controls dimerization; B/C are duplicated sequences, each with one cGMP-binding site; and D/E contain the ATP-binding site in D and the active centre in E. The N-terminal regulatory half shows similarity to various small cGMP-binding proteins, whereas the C-terminal half is homologous to the catalytic subunit of the cAMP-dependent kinases, the phosphorylase-b kinase, and the oncogene tyrosine kinases; the whole molecule appears to be a chimera produced by gene fusion [63, 292]. cGMP-regulated protein serine/threonine kinases also occur in arthropods; in fact, they were discovered in this group before they were recorded in mammals. Here, the catalytic and regulatory domains also lie on the same polypeptide. In *Drosophila*, four cGMP protein kinase mRNAs are produced from two genes; the encoded polypeptides agree in 55–64% of their amino acids with the enzyme from bovine lung; however, in one case, the domains responsible for dimerization and inhibition are lacking. The enzyme is also always found as a monomer in the ciliate *Paramecium*, but in the roundworm *Ascaris suum* it is a dimer [131, 180, 275]. The enzymes from the Mediterranean fruit-fly, *Ceratatis capitata*, and from the eggs of the silkworm, *Bombyx mori*, are strictly cGMP specific, whereas a protein kinase from the skin of the locust *Melanoplus sanguinipes* shows as high an affinity for cAMP as for cGMP [269]. Amongst the protozoans, cAMP-regulated protein kinases have been found in the malarial agents *Plasmodium berghei* and *P. chabaudi*. In contrast, the trypanosomes have an adenylate cyclase but no cAMP-stimulated protein kinase [48].

Three classes may be distinguished amongst the **Ca^{2+}-regulated protein kinases**; the first includes the Ca^{2+}-calmodulin-dependent enzymes, of which phosphorylase-b kinase and myosin-LC kinase are the best characterized. A second class is made up of the enzymes regulated by Ca^{2+}, phospholipid and diacylglycerol (DAG) and these

are called protein kinase C. To date, the sole representatives of the third class are two protein kinases from the ciliate *Paramecium tetraurelia* and these are stimulated by Ca^{2+} ions independently of calmodulin or lipids [92]. **Calmodulin** is a 17-kDa protein with four Ca^{2+}-binding sites. It is ubiquitous in the eukaryotes and extremely conserved in evolution; it belongs to the same family as troponin C. The phosphorylase-b kinase is a complex of 1.3 MDa, composed of 16 polypeptides according to the formula $(\alpha,\beta,\gamma,\delta)_4$. The regulatory subunits α (118–145 kDa) and β (108–128 kDA) can be phosphorylated at several positions by various protein kinases; in this way the enzyme becomes activated. The catalytic γ-subunit (44.7 kDa) is homologous to the catalytic subunit of other protein kinases; the δ-subunit is identical to calmodulin. In vitro, the enzyme phosphorylates other proteins in addition to phosphorylase-b. A multi-functional protein kinase, activated by Ca^{2+}-calmodulin, is apparently commonly found. This phosphorylates a range of very different substrates and thus participates as a regulator in many processes: glycogen synthase and tyrosine hydroxylase (metabolism), phospholamban (relaxation of heart muscle), MAP-2 (dissociation of microtubules) and synapsin (release of transmitter at the synapse) [244]. Ca^{2+}-calmodulin also indirectly affects the phosphorylation and dephosphorylation of proteins by activation of adenylate and guanylate cyclase and of protein phosphatase 2B.

The enzymes known as **protein kinase C** require Ca^{2+} and the phospholipid phosphatidylserine (PS) for their activity, and are stimulated by DAG. They were first discovered in 1979, although they are found in some mammalian tissues at levels far greater than other protein kinases. They play a central role in a system of signal transduction through the cell membrane that acts upon the membrane lipid phosphatidylinositol-4,5-biphosphate (PIP_2). There are various hormones, neurotransmitters and other extracellular signals which cause the hydrolysis of PIP_2 to inositol-1,4,5-trisphosphate (IP_3) and DAG. The IP_3 releases Ca^{2+} from intracellular stores and thus influences many cellular processes through the activation of particular protein kinases, protein phosphatases and proteases; this is caused by the increase in the intracellular calcium concentration. The protein kinase C stimulated by Ca^{2+} and DAG actually consists of a whole family of multiple enzymes; to date, seven enzymes of this type, encoded by different genes, have been isolated from rat brain [140]. An inhibitor of protein kinase C has been isolated from bovine brain, and is a zinc-binding polypeptide of 125 amino acids without homology to any known protein [207].

Ca^{2+}-PS-DAG-dependent protein kinases are apparently ubiquitous, having been detected in lower vertebrates, echinoderms, annelids, crustaceans and insects, as well as in mammals [217, 249]. Several protein kinase C genes known in *Drosophila* show specific similarities to particular mammalian genes; this multi-gene family is therefore apparently very old [242]. A protein kinase C is also responsible for the phosphorylation of nuclear proteins and increased DNA synthesis that is induced by the aggregation factor (AF) in the sponge *Geodia gigas* [115, 231]. There are many protein kinases that are not stimulated by either cAMP, cGMP or Ca^{2+}. To these belong, for example, the casein kinases I and II that are widespread in both animals and higher plants. Type I from mammalian kidney, spleen and liver is a monomer of 37 kDa; Type II, which is found in mammalian spleen and testis and also occurs in *Drosophila* and *Caenorhabditis*, is a tetramer $\alpha_2\beta_2$, the α-chain of which is homologous to the catalytic chains of other protein kinases. In bovine testis there are also tetramers of $\alpha'\beta_2$, where the α′ chain is encoded by its own gene [63, 114, 160]. Other protein kinases that are not regulated by secondary messengers include the haem-controlled repressor (HCR) involved in globin synthesis (p. 81); the pyruvate dehydrogenase kinase (p. 689) and the closely related kinase for the dehydrogenase complex of the branched keto-acids ketoleucine, ketoisoleucine and ketovaline; the rhodopsin kinase (p. 736); the glycogen synthase kinase (p. 488); the kinase for the ribosomal protein S 6 (p. 50); and the myosin HC kinase from *Acanthamoeba* (p. 339) [63].

It appears that just a few **protein phosphatases** suffice for the complicated regulation system of protein phosphorylation and dephosphorylation in the tissues of mammals. Four types of protein serine/threonine phosphatases are known, of which 1, 2 A and 2 B show significant mutual homology, whereas 2 C belongs to another family. The third family of protein phosphatases is that of the protein tyrosine phosphatases [31]. The four protein serine/threonine phosphatases may be divided into two classes according to their catalytic and regulatory properties. Protein phosphatase 1 is specific for the β-subunit in the reaction with phosphorylase-b kinase and is inhibited by two specific protein inhibitors 1 and 2. The three enzymes belonging to the second class (2 A, 2 B and 2 C) are all specific for the phosphorylase-b

kinase α-subunit and are insensitive to the two inhibitors; however, they are very different in molecular size, substrate specificity and dependence on divalent cations. The different protein phosphatases, together with their inhibitors and several protein kinases, make up an interlocking regulatory network: protein phosphatase 1 is only inhibited by inhibitor 1 when the latter is phosphorylated by a cAMP-dependent protein kinase; the dephosphorylation of inhibitor 1, which in muscle is mainly carried out by protein phosphatase 2B, leads to the inactivation of the inhibitor and hence to the activation of phosphatase 1. Some protein phosphatase 1 is present in muscle as an inactive complex with inhibitor 2. This complex dissociates, and phosphatase 1 thereby becomes activated, when inhibitor 2 is phosphorylated by the glycogen synthase kinase. Thus, the glycogen synthase kinase not only is responsible for the phosphorylation of glycogen synthase but also promotes its dephosphorylation via protein phosphatase 1. Protein phosphatase 2B, the only one that is activated by Ca^{2+}-calmodulin, is found mainly in the brain and in skeletal muscle and is identical to the Ca^{2+}-binding protein calcineurin in the brain; it possesses a high-affinity Ca^{2+}-binding site [36, 118]. Proteins isolated and characterized from the head of *Drosophila* are very similar to the protein phosphatases 1, 2A and 2B and the inhibitors 1 and 2. Here also the phosphatase 1 forms an inactive complex with the corresponding inhibitor 2. The inhibition of protein serine/threonine phosphatases by inhibitors 1 and 2 is also a common feature of organisms like baker's yeast, and the regulation of the protein phosphatases thus appears to be similar in all eukaryotes [36, 57, 201].

3.1.3 Analysis and Comparison of Protein Sequences

Whilst Sanger had to work for 10 years to determine the order of the 51 amino acids of insulin, and required several grams of the substance, the **sequence analysis** of far bigger proteins has become a routine affair since the 1960s, and can be carried out on milligram amounts. By the beginning of 1977, 1250 protein sequences were known, not including the anomalous human haemoglobins and other alleloproteins [51, 126]; the number has since increased by several-fold. Nowadays, the nucleotide sequence of structural genes or of mRNA-complementary DNA (cDNA) is far easier to determine than the amino acid sequence, the more so as the amount of substance for analysis can be amplified by cloning. Thus, amino acid sequences are now mostly determined indirectly from the gene or cDNA sequences. Whilst direct analysis of amino acid sequences is mostly carried out on proteins of known function, the analysis of DNA sequences often leads to the question of the biological function of the encoded polypeptide. Comparison with consensus sequences of particular functional regions or motifs in proteins of known function can often be helpful in such cases [3, 19].

The coincidence or similarity of protein sequences can be assessed with relatively little effort and without previous isolation using immunological and electrophoretic methods, although the room for error is here quite large. Since it was discovered in 1897 that the antigen-antibody reaction can take place as a "cross-reaction" with other (heterologous) antigens as well as with the protein used for immunization (the homologous antigen), there have been innumerable investigations of the **serological relationships** between proteins. Protein comparisons were almost always carried out like this until the invention in 1937 of paper electrophoresis. It is important to realize that reactions of proteins of different origin with a given antibody do not provide evidence of the identicalness of either the proteins or reactive sites on the molecules, but only the recognition of similar structural features (epitopes) by the antibody. If antibodies are to be used for protein identification, the highest possible specificity for a particular protein, or for a particular immunological determinant, is essential. This can be achieved by the production of monoclonal antibodies. Immunodiffusion and immunoelectrophoresis can be used to examine possible reactions between proteins of interest and antibodies, and this can be quantified, e.g. using microcomplement fixation, enzyme inhibition by antibodies, and both radio- and enzyme-immunoassays. Immunological methods are used mainly in the biochemical analysis of relationships and will be thoroughly discussed in the relevant sections.

Depending upon the system used, **electrophoretic methods**, concern charge differences (gel electrophoresis, isoelectric focusing) or molecular size differences (SDS gel electrophoresis). Two-dimensional electrophoresis makes use of two different separation criteria to achieve quite amazing separation, after which the protein spots can be detected by highly sensitive silver staining or by autoradiography. Using these methods, it

should soon be possible to record the total spectrum of cellular proteins. It is already possible to separate in 2-D electrophoregrams more than 1000 proteins from cell extracts and to compare them by computer analysis [14, 61]. The cleavage peptides obtained from the enzymic hydrolysis of proteins can also be separated two-dimensionally to give a "peptide map" or a "fingerprint". Comparisons with the peptide map of a protein of known sequence can be used to detect and localize single amino acid differences. This method is mostly used for the comparison of closely related proteins.

3.1.4 Folding of the Polypeptide Chain

During the biosynthesis of proteins, the 1-D nucleotide sequence of the mRNA is translated into a 1-D amino acid sequence; however, the biological function of the protein depends finally on a particular 3-D structure **(chain conformation)** that arises by a complicated folding process during the synthesis of the polypeptide chain. Destruction of the chain conformation (denaturation) leads to loss of biological activity [205]. The ageing of enzyme proteins, which can lead to reductions in specific activity of 40–60 %, also seems to be the result of slight alterations in chain conformation.

All the information required for the spatial structure of a protein is contained in the amino acid sequence. Evidence for this was already obtained in the 1930s with the observation that pancreas RNase denatured by urea spontaneously renatures in urea-free solution. It was initially thought that during the **folding process** many different conformations were tested until, at last, the most thermodynamically stable one was achieved. It was then realized that this would require too much time; even a small protein like the RNase would, on average, require a century to assume its specific conformation, whereas, in reality, the β-galactosidase of *E. coli*, for example, is synthesized and correctly folded within 4 minutes, despite its large size of 500 kDa. Selection apparently favours those sequences that allow the rapid assumption of a unique, functional conformation. Folding is kinetically determined and begins on the partially completed chain; spontaneous renaturation is actually observed only for a very few proteins. Initially, very small regions are apparently folded within microseconds into specific secondary structures, like α-helices or β-sheets; interactions between such regions bring about larger, stable areas (domains) in which the α and β structures take up particular positions. The final chain conformation comes about by interactions between certain binding sites on the surface of domains, or by the formation of disulphide bridges [43, 119, 142].

With multimeric proteins, there follows the association of the polypeptide chains to form one of the possible **quaternary structures** of the protein [83, 205]. About 40 % of the proteins of mammals and bacteria are monomers; most of the rest are even-numbered oligomers, mainly di- or tetramers. Less than 5 % are trimers, for example, ornithine transcarbamylase, arginase and carboxylesterase of mammalian liver, the complement components C4 and P, and the haemerythrin of the sipunculans, or pentamers like the C-reactive protein and the serum amyloid protein of vertebrates. Modification of the quaternary structure is an important mechanism for the regulation of enzyme activity. This can involve dissociation and reaggregation of the enzyme subunits, and also reversible binding to other enzymes or structural proteins [253].

The basic unit of protein structure is the **domain**, an autonomous subregion of the polypeptide chain that possesses all the characteristics of a globular protein and can often be isolated proteolytically without loss of its properties [98, 245]. It seems reasonable to relate protein domains to the exon as the structural unit of the gene. The recombination of exons (exon shuffling) can then only be imagined as an effective mechanism for the creation of new proteins when structurally and functionally autonomous protein parts are linked together in this way. In many cases, exon and domain boundaries do actually coincide, e.g. in the immunoglobulins, β-globin, lysozyme, pyruvate kinase and many membrane proteins. There are, however, numerous examples where this is not the case, e.g. carboxypeptidase, prothrombin and other serine proteases, α-amylase and carboanhydrase [5]. In such cases, the exon boundaries may in fact correspond to points on the surface of the globular molecule, so that shifts, caused by mutations in splicing signals, in the exon-intron boundaries can lead to the introduction or removal of single amino acids or short peptides without alteration of the spatial structure of the protein. [42].

The most important **secondary structures** are the right-handed α-helix and the β-sheet; occasionally, other helical structures occur such as the left-handed helix or the polyproline conformation. These structures can be deformed within certain limits: an α-helix of more than 15 amino

acids can be bent by up to 30°; β-sheets can show twisting, coiling, bending or bulging. Because the different secondary structures of a protein molecule are always tightly packed to a high density, there is a limited number of possible **basic conformations**. According to the presence and arrangement of α-helices and β-sheets, at least four classes of proteins or protein domains can be distinguished: the α-type, the β-type, the α/β-type in which α and β structures alternate, and the α+β-type in which both α and β structures are present but mostly separated [33].

In many proteins, the spatial structure has an **internal periodicity**, most clearly seen in the fibrillar proteins, the mechanical properties of which are based on such a regular structure [245]. The properties of the globular proteins are also determined by their **spatial structure** and the resulting spatial arrangement of the amino acid side-chains. Because all the information for the chain formation of a protein is in the amino acid sequence, it should be possible to predict the spatial structure from the sequence. However, of the many analytical procedures developed for this, none is entirely reliable [68]. Such predictions are particularly difficult for integrated membrane proteins; so, for example, the transmembrane domains are not always clearly delineated from the parts of the polypeptide chain lying outside of the membrane [123]. However, on the basis of the known connections between amino acid sequence and spatial structure, it is possible to presume with some certainty similarities of spatial structure from sequence comparisons of related proteins. This is important for comparative biochemistry because the direct determination of spatial structure, e.g. by X-ray analysis, is more difficult and required much more effort than amino acid sequence analysis, and has so far been achieved with only about 400 proteins. It should also be pointed out that such direct analyses provide only the mean position of the atoms in the protein crystal, whereas in native proteins the gross movement of single atoms, side-chains and domains occurs, as do folding and unfolding movements; this dynamic situation is of great importance for allosteric effects, membrane transport, enzymatic catalysis, electron transport, and the binding of large and small molecules [175].

If extensive similarities between both physicochemical and biological properties are found in the comparison of proteins of different origin, then similarities of spatial structure can be presumed. Such conclusions are in any case justified if enzymatically active **hybrid molecules** are created by combining enzyme subunits from different sources. Functional enzyme chimeras result between, for example, rabbit aldolase and the enzyme from *Drosophila* or *Ascaris*, the glyceraldehyde-3-phosphate dehydrogenase of rabbit and that of *Ascaris* yeast, and the triosephosphate isomerases of chicken and bacteria. In all these cases, the amino acid sequences are so different that the starting enzymes show no immunological cross-reactivity. If the spatial structure is sufficiently similar, even enzymes with different substrate specificities may be hybridized, for example, the creatinase of rabbit with the arginine kinase of a holothurian.

3.1.5 Multiple Forms of Proteins

In the 1950s, an increasing number of results indicated that different enzyme species could be present in the same organism, and even in the same cell, and have the same reaction and substrate specificity. Markert and Møller in 1959 suggested the name "isozyme" for such multiple enzyme forms, although linguistically "isoenzyme" is to be preferred. Multiple forms of non-enzymatic proteins are termed isoforms. The separation and identification of multiple proteins can be carried out by electrophoresis, and also by chromatography, fractionated heat inactivation, etc. The great biological and medicinal significance of this phenomenon has prompted many extensive reviews of the subject [182, 195].

The use of the term **isoenzyme** was at first independent of the nature of the molecular differences. However, in 1976 the nomenclature commission of the International Union of Biochemistry (IUB) recommended that this term be applied only when genetic differences are apparent, i.e. for groups 1–3 in their list of multiple enzyme forms (Table 3.2). It is regrettable that the same term is now applied both for the products of different loci (groups 1 and 2) and for products of different alleles of the same locus (group 3). For allelic variants, Prakash, Lewontin and Hubby suggested in 1959 the term "allozyme", which would also be better as "alleloenzyme". The allelic polymorphism of the nucleic acids and the proteins is of great importance for the theory of molecular evolution and will be extensively discussed in this connection.

Isoenzymes of group 1 are distinguished only arbitrarily from other enzymes of similar or overlapping specificity that count as separate enzyme

Table 3.2. Multiple forms of enzymes, grouped according to the recommendations of the IUPAC-IUB Commission on Biochemical Nomenclature (1976)

Groups 1–3:
genetically different enzyme forms (isoenzymes)
1. Enzyme proteins coded by different genes
2. Different heteropolymers of at least two non-covalently linked polypeptide chains
3. Enzyme proteins coded by different alleles of a gene (alleloenzymes)
Groups 4–6:
post-translationally modified enzyme forms
4.a) Enzyme forms arising by conjugation or substitution of components
4.b) Enzyme forms arising from the same polypeptide chain by hydrolytic processes
5. Different multimers of the same subunits
6. Different conformations of the same protein

species and are given their own number in the enzyme list of the IUB commission. The alkaline and acidic phosphatases, for example, are clearly different enzyme species; more questionable, on the other hand, are the lines drawn between, for example, "real" cholinesterase (EC 3.1.1.7) and "pseudocholinesterase" (EC 3.1.1.8), or between glucokinase (EC 2.7.1.2) and hexokinase (EC 2.7.1.1). Structural and functional differences between different enzyme forms are by no means always proportional. For example, the porcine cytoplasmic and mitochondrial aspartate transaminases have very similar catalytic properties but only 47% sequence identity; isoenzyme A_2 (earlier EE) of equine alcohol dehydrogenase is specific for acetaldehyde and isoenzyme B_2 (earlier SS) for steroids, whereas they differ in only 6 out of 374 amino acids. The problems of differentiating between enzyme species and isoenzymes are especially relevant in comparative biochemistry because the characteristics of enzymes of differing origin may be superimposed in a complicated manner. Phylogenetic relationships between enzyme proteins, which form the primary interest of comparative biochemistry, can in any case only be derived by the comparison of nucleotide or amino acid sequences and not by enzymatic properties.

The **maximum number of different enzyme forms** in the isoenzymes of group 2 is equal to the number of possible combinations: e.g. five homo- and heterotetramers, A_4, A_3B, A_2B_2, AB_3 and B_4, can be created from two different subunits A and B, as in the case of the much cited lactate dehydrogenase. Where, for example, due to allelic variability, more than two subunits are involved in tetramer formation, very complicated spectra of isoenzymes are the result. Not all possible combinations are so stable that they will be recorded. In numerous cases, the quaternary structure of a protein has been concluded simply from the pattern of iso- and alleloenzymes. The **sequence difference** between alleloenzymes usually only involves one or a few amino acids; in contrast, sequence differences between isoenzymes of group 1 can be as high as 50%, for example between the cytoplasmic and mitochondrial aspartate transaminases of pig heart.

The term "epizyme" has been suggested for the enzyme forms of group 4 that result from **post-translational modification**, but this was not widely accepted. The differences between the multiple enzymes of group 4 can arise by any of the previously mentioned post-translational reactions. Thus, for example, the 18 different forms of alkaline phosphatase in the human placenta differ in their content of N-acetylneuramic acid (sialic acid); the differences disappear upon treatment with neuraminidase. The isoforms of phenylalanine hydroxylase in rat kidney have different phosphate contents, etc. [182]. Hydrolytic modifications of enzyme proteins, such as the deamination of asparagine to aspartic acid, can occur; an example is the multiple aldolases in rabbit muscle, and other proteins in group 4 b. Partial proteolysis, which does not affect enzyme activity, is known, for example, for pyruvate kinase of human liver and glutaminase in rat kidneys [34, 182]. Because the enzymes responsible for the post-translational modification of proteins are under genetic control, there are also heritable differences in the spectrum of enzyme forms of group 4. So, for example, the ability of mouse liver to increase the electrophoretic mobility of mouse albumin is dependent on a dominant allele; the mobility of esterase 6 of *Drosophila melanogaster* is influenced by a gene located at another position on the same chromosome [35].

Aggregation to polymers of various molecular masses has been described for many enzymes (group 5), e.g. for the alkaline phosphatase of human placenta, the mitochondrial creatinase of bovine heart and various cholinesterases [182]. There are also innumerable cases of conformation changes due to allosteric effects; the existence of stable **conformation isomers** of enzyme proteins (group 6), on the other hand, is controversial. Indisputable evidence for this would be the demonstration that the relevant enzyme forms are interconvertible and that, at the same time, they assume characteristic properties. Up to now,

this has best been shown for enolases 2 and 3 of the roundworm *Ascaris suum* [81]. The apparent presence of multiple protein forms can be due to different artefacts. Particularly in the case of post-translationally modified proteins, forms with altered electrophoretic mobilities can occur during extraction as the result of both enzymatic and non-enzymatic processes. Because of overlapping substrate specificities, assumed enzyme-specific staining of the electrophoregram can detect other proteins, for example in the case of alcohol dehydrogenase and lactate dehydrogenase in *Drosophila melanogaster;* each of these can react with the substrate of the other.

In most cases, there are **functional differences** between multiple enzyme forms, e.g. in specific activity, substrate specificity and affinity, pH optimum, regulatory characteristics, temperature dependence and stability. These differences are generally much greater between the products of different genes, and between some post-translationally modified enzymes, than between alleloenzymes. The immunological similarities between the latter are also generally maintained, so that, for example, enzymatically inactive null variants can be detected immunologically. The existence of isoenzymes was considered from the outset to represent a fine adaptation to different functional conditions. The idea of an adaptive role is supported by the variable distribution of isoenzymes between different tissues and cell compartments, as well as by the fact that they appear in a strict order during ontogeny. On the other hand, it should be noted that gene duplication and subsequent diversification, according to our present knowledge, are important mechanisms in the evolution of enzymes; thus, some isoenzymes may only be intermediate steps in the evolutionary transformation of proteins and may themselves have no adaptive value.

There is an abundance of plausible examples of the **adaptive significance of isoenzymes**, although in many cases these are not without a rider. For example, the activity of aldolase A of mammalian muscle is five times higher with fructose-1,6-bisphosphate than with fructose-1-phosphate, whereas both substrates are used with equal efficiency by aldolase B in the liver; in this way, the muscle enzyme is adapted to glycogen degradation, and the liver enzyme to the utilization of fructose and glycogen synthesis. However, this does not explain why, in the brain, there is an aldolase C with intermediate specificity. The typical lactate dehydrogenase (LDH) isoenzyme of mammalian muscle A_4 is much less inhibited by pyruvate than is the typical heart isoenzyme B_4; if one assumes that isoenzyme A_4 is adapted to anaerobic lactate production in muscle, and isoenzyme B_4 to aerobic lactate oxidation in the heart, it is then difficult to understand why the usually aerobically active liver cells in many mammalian species contain so much a_4 [182]. The existence of different isoenzymes, e.g. of malate dehydrogenase, aspartate transaminase, malate enzyme and NADH-specific isocitrate dehydrogenase, in mitochondria and the cytoplasm is plausible in view of the different concentrations of metabolites and the opposite directions of the reactions in these different compartments. But one should also consider the fact that the mitochondrial enzymes are made on cytoplasmic ribosomes and must then be transported into the mitochondria; this requires special protein structures. It is also clear that the enzymes of the lysosomes and membrane-bound enzymes must be different from those of the cytoplasm; this has been shown, for example, for various glycosidases and phosphatases [182].

3.2 Protein Synthesis

Since the first successful attempts at the beginning of the 1950s, cell-free **protein synthesis systems** have been established from several very different eukaryotic cells. Significantly higher yields of transcription and translation products are to be obtained, however, following injection of DNA or mRNA into whole, living cells. The developing oocytes of the clawed frog, *Xenopus laevis*, are very often used for this purpose; because of their large size (a diameter of 1.2 mm) they can yield up to 100 times more product than other cells. Such in vitro or in vivo systems can also correctly translate foreign genetic material; the cytoplasmic components of protein biosynthesis are apparently very similar in all eukaryote cells. Extremely conservative evolution of the protein-synthesizing machinery is to be expected because each large change would have catastrophic consequences for the cell. Only in the prokaryotes and in eukaryote mitochondria, which are assumed to have arisen from endosymbiotic prokaryotes, are significantly different translation mechanisms to be found. In spite of the universality of protein synthesis, the translation of heterologous mRNA is often markedly reduced in contrast to that of homologous mRNA; the variable inhibitor sensitivity of the translation of different mRNAs also

points to species- and cell-specific differences in the translation apparatus.

Of particular interest is the **in vitro synthesis of the silk fibroin** which, at 350–415 kDa, is one of the largest known polypeptides. In the rabbit reticulocyte system, the synthesis of the complete chain takes 85 min. The appearance of discrete classes of shorter chains indicates that translation is interrupted at particular sites on the mRNA. In intact silk-gland cells, the synthesis of the silk fibroin molecule is completed in less than 60 min, but here too the process is discontinuous. The 320-kDA large fibroin molecule used for web construction by the spider *Nephila claviceps* is also produced discontinuously. The biological sense of this mode of synthesis is not clear [24].

3.2.1 Aminoacyl-tRNA Synthetases

To become involved in protein biosynthesis, the amino acids must first be attached to tRNA. For each amino acid there are normally available several isoacceptor tRNAs, but only one specific aminoacyl-tRNA synthetase. The synthetases bind covalently to the uridine in position 8 that is found in all cytoplasmic tRNAs. In prokaryotes, the enzyme has subunits of 50–300 kDA and these are arranged as mono-, di- or tetramers. In contrast, in mammals, the synthetases for at least nine of the standard amino acids form 18S multienzyme complexes of about 1 MDa. The synthetases of the other amino acids are found in vitro as di- or tetramers, but in living cells are probably also associated with complexes. All mammalian synthetases have subunits of 50–160 kDa [47, 82). In *Drosophila*, synthetase complexes of more than 1 MDa have also been described. The functional significance of these complexes remains a mystery.

The synthetases for glycine, alanine and serine are particularly active in fibroin synthesis and have been isolated from the silk glands of *Bombyx mori*. The glycine- and serine-specific enzymes are dimers of 160 and 124 kDa, and the alanine-specific enzyme is a monomer of 115 kDa that is so similar to the tetrameric enzyme of *E. coli*, which shows the same specificity, that it shows immunological cross-reactivity [28, 222, 285]. Using immunological methods, it can be shown that the mitochondrial and cytoplasmic phenylalanine tRNA synthetases are very closely related, even in organisms as far apart as the chicken and yeast; the two isoenzymes are presumably encoded by duplicated chromosomal genes [76]. Although there are as yet no complete sequences for animal synthetases, it is already clear that more than one protein super-family is involved. The aminoacyl-tRNA synthetases also catalyse the condensation of 2 ATP to Ap_4A, which plays an as yet unknown role in DNA synthesis, cell growth and heat-shock reactions [47].

3.2.2 Initiation, Elongation and Termination

The course of translation consists of three subprocesses, each of which requires specific protein factors; these processes are initiation, elongation and termination. Many basic studies of protein biosynthesis have made use of the cell-free, rabbit reticulocyte system. Amongst the invertebrates, the brine shrimp *Artemia salina* and the fly *Drosophila melanogaster* have been especially thoroughly investigated. Translation begins at the first AUG codon of the mRNA; one exception to this is the aldolase gene of the malarial agent *Plasmodium falciparum*, which uses the normal stop codon UAG as a start signal [79]. During **initiation** in eukaryotes, a 43 S complex first forms from the small ribosomal subunit and the initiator tRNA; binding of the mRNA then leads to a transient 48 S complex that combines with the large subunit to produce an active translation complex of 80 S. The series of protein factors involved in translation are described below using the recently agreed terminology of the IUB. The eukaryotic initiation factors (eIF) include polypeptides of 15–220 kDa and are found as monomers (eIF-1, eIF-3A, eIF-5, eIF-5A), homodimers (eIF-4B), or complicated molecules with three (eIF-2, eIF-4, eIF-4F), five (eIF-2B) or eight (eIF-3) different subunits. The phosphorylation of serine or threonine residues in certain subunits is apparently an important regulation mechanism of translation [105].

eIF-2 forms a complex with GTP and $tRNA_i^{Met}$ that then enables interaction between the mRNA AUG codon and the small subunit. After formation of the 80 S initiation complex by association of the large subunit, the GTP is hydrolysed and the eIF-2 bound to GDP is released. After exchange of GDP for GTP, eIF-2 once again takes part in the formation of an initiation complex. Haem deficiency in the rabbit reticulocyte system causes the subunit eIF-2α to be phosphorylated by a specific protein kinase, the haem-controlled suppressor (HCR), and so to lose the ability for GDP/GTP exchange. In this way, the formation of the initiation complex, and thus the

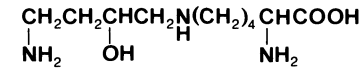


Fig. 3.2. Hypusin [N^{ε}-(4-amino-2-hydroxybutyl)lysine] is found in the initiation factor eIF-4D and arises by the donation of a butylamine residue from spermidine to a lysine residue, and subsequent hydroxylation

whole of protein synthesis, is inhibited. The eIF-2α subunit of *Artemia* is phosphorylated at the same site by HCR but remains active, the HCR is completely inactive on eIF-2α from *Drosophila* [172, 178]. The initiation factor eIF-5A (previously eIF-4D carries the unusual amino acid hypusin (Fig. 3.2) at position 50. This is produced post-translationally from a lysine residue by transfer of a butylamino group from spermidine, followed by hydroxylation. A protein of about 18 kDa with such an hypusin residue is found in all mammalian tissues and also in *Drosophila melanogaster* and baker's yeast, but is not present in *E. coli* or in the archaebacterium *Metnaococcus voltae*. Thus, this hypusin-containing peptide is ubiquitous in eukaryotes [84, 206].

Elongation, the extension of the growing polypeptide chain by one amino acid, involves three steps: the binding of the aminoacyl-tRNA at the A site of the ribosome; transfer of the peptide from the P site of the ribosome to this aminoacyl-tRNA with the formation of a peptide bond; and sliding of the ribosome to the next codon on the mRNA (translocation). In eukaryotes, three elongation factors (eEF) are involved in this process: eEF-1α, eEF-1βγ and eEF-2. eEF-1α-GTP binds the aminoacyl-tRNA and transports it to the ribosome. After delivery of the aminoacyl-tRNA and hydrolysis of the GTP, the eEF-1α-GDP leaves the ribosome. With the help of the eEF-1βγ, GDP is exchanged for GTP, thus regenerating active eEF-α-GTP [46]. eEF-2 is responsible for translocation. In unfertilized *Xenopus laevis* eggs, eEF-1α is bound to a particular type of ribonucleic acid particle. These "thesaurisomes" often occur in large numbers in the oocytes of teleosts and anurans. In *X. laevis*, there are small 7 S particles, which contain a molecule each of 5 S rRNA and the transcription factor TFIIIA, and larger tetrameric 42 S particles, whose subunits each contain one 5 S rRNA, three tRNAs and two protein molecules of 50 and 43 kDa, namely thesaurin a (42 Sp 50) and thesaurin b (42 Sp 43). Sequence analysis shows that the 42 Sp 50 is a type of eEF-1α that is only expressed in oocytes [56]. The eEF-1 α units of the mouse and man differ in only 9 of 461 positions; even the eEF-1 α sequences of *Drosophila melanogaster* and the honey bee, *Apis mellifera*, whose developmental lines diverged 250 million years ago, differ in only 49 amino acids. The eEF-2 of *Drosophila* agrees in more than 80 % of its 844-amino-acid sequence with that of the hamster [88, 289]. These data illustrate the conservative evolution of the protein synthesis apparatus. Peptide chain termination is directed by one of the stop codons (UAA, UAG, or UGA), but additional information may lie in the sequence immediately following the stop codons [27]. The three termination codons are recognized in bacteria by two termination factors of differing specificity; in mammals and insects there is only one termination factor, RF, a homodimer of 55-kDa subunits. The ribosomal subunits set free on **termination** are prevented from spontaneous reassociation by an anti-association factor, so that sufficient free 40S subunits are available for initiation [23, 27].

The **regulation of translation** appears mostly to be concerned with initiation; the initiation frequency, i.e. the number of initiation complexes per mRNA formed within a given time, depends upon the physiological status of the cell and the type of mRNA. In plasma-rich, unfertilized eggs of, for example, sea urchins, fish and amphibians, there is a large pool of untranslated mRNA; on fertilization the rate of protein synthesis increases dramatically. The inhibition of protein synthesis in the oocytes is not due to the absence of necessary components, because injected β-globin mRNA, for example, is translated. Of the possible explanations for the inhibition of translation prior to fertilization, the "masking" of mRNA by specific proteins is currently receiving most attention. The inhibitors present in the unfertilized eggs of the sea urchin *Strongylocentrotus purpuratus* inhibit translation not only in cell-free systems of the eggs themselves but also in the rabbit reticulocyte system. The 48 S complex of the small ribosomal subunit, the initiator tRNA and the mRNA is formed, but the binding of the large subunit does not take place [97]. Five translation-inhibiting proteins of 50–94 kDa have been isolated from the cytoplasmic ribonucleic acid particles of *Xenopus laevis* oocytes; globin mRNA associated with ribonucleoprotein (RNP) particles is not translated in the presence of these proteins. The regulation of translation in the brine shrimp *Artemia* is still something of a mystery. If dormant *Artemia* embryos are rehydrated, development and protein synthesis rapidly begin. There are, however, no changes in the concentrations of the initiation factors eIF-2, co-eIF-2A

and co-eIF-2B; the phosphorylation of the eIF-2α subunit seen at the beginning of development appears to have no influence on the formation of the initiation complex [295].

3.2.3 Transport of Newly Synthesized Proteins to Their Correct Destinations in the Cell

After their synthesis on free polysomes or on the ribosomes bound to the rough endoplasmic reticulum, the polypeptides must be transported to their particular destinations in the cell: secretory proteins are transported out of the cell by exocytosis; integral membrane proteins are built into the plasma membrane; and the proteins of the nucleus, mitochondria, microbodies (peroxisomes, glyoxysomes, glycosomes) and lysosomes are introduced into their respective organelles. During their translocation, proteins must normally pass at least one cellular membrane. The **signals for the distribution to different cell compartments** are contained in the amino acid sequence of the primary translation products [21, 90, 262]. Identification of these signals is made possible by genetic defects that disturb the translocation of secretory, lysosomal or membrane proteins. Proteins without an address remain in the cytoplasm; proteins linked experimentally with a translocation signal are transported into the relevant compartment [20]. During or after translocation, the polypeptides may be changed by partial proteolysis, substitution or conjugation, linked to non-protein components or built into supra-molecular structures. Proteolytic processes can bring about the removal of the N-terminal methionine or of the terminal translocation signals, or the maturation (processing) of pro-enzymes, blood-clotting factors, complement components or hormone precursors. Substitutions and conjugations create the numerous altered protein components described previously. All these processes will be dealt with in the sections on the relevant proteins. Unfortunately, there are very few data on the comparative biochemistry of translocation and post-translational modifications.

The signals and mechanisms of translocation appear, in principle, to be the same in all organisms. Hence, polypeptides encoded by foreign genetic material are correctly distributed in *Xenopus laevis* oocytes, and even in *E. coli*. After injection of the corresponding mRNA into oocytes, typical secretory proteins, like ovalbumin and other egg-white proteins, immunoglobulins or insect vitellin, are released into the medium; rabbit haemoglobin is retained in the cytoplasm; and cytochrome P-450 is introduced into membranes [147]. Firefly luciferase, synthesized in the cells of various insects and mammals, and even of higher plants and yeast, is translocated correctly to the peroxisomes [86]. The signals are so conserved in evolution that the normally cytoplasmic α-globin chain is taken up by the endoplasmic reticulum if it is coupled to the signal of the bacterial β-lactamase [20]. The fact that human genes are not only transcribed and translated in bacteria but, in many cases, also correctly processed is made use of in industry. Post-translational modification is also uniform within very wide limits. Thus, after injection of the mRNA for phaseolin into *Xenopus* oocytes, this plant storage protein is not only correctly translated but also glycosylated [173]. There are, however, species- and cell-specific mechanisms for the further processing of newly synthesized polypeptides; after the introduction of the rat prepro-insulin gene into *E. coli*, pro-insulin, and not insulin, is released into the medium [147]. Likewise, injection of the mRNA for the bee toxin prepro-melittin into *Xenopus* oocytes does not result in the release of melittin, but pro-melittin is stored for many weeks in cellular vesicles; here, in contrast to the situation in the poison glands, the C-terminal glycine is not removed and the export of the protein is blocked [149].

Secretory proteins find their way out of the cytoplasm via the lumen of the endoplasmic reticulum (ER) and the Golgi apparatus into specific transport vesicles, in which they are then transported to the cell surface. Because the vesicles of the Golgi apparatus and those of the ER arise by budding the proteins must only cross one membrane, that of the ER. According to the **signal hypothesis**, an N-terminal, non-polar signal sequence is responsible and this is recognized by a signal recognition particle (SRP) [21, 90]. Polypeptide synthesis begins on free, cytoplasmic polysomes. Through binding of an SRP to the translation complex, the elongation of the polypeptide is stopped at a length of about 70 amino acids (Fig. 3.3). The complex of SRP, ribosome and incomplete polypeptide chain binds to an SRP receptor on the ER membrane (a "docking" protein), and the SRP is split off. Now the signal sequence can interact with the signal sequence receptor and elongation is restarted. The SRP has three different tasks to accomplish (signal recognition, translation inhibition, and interac-

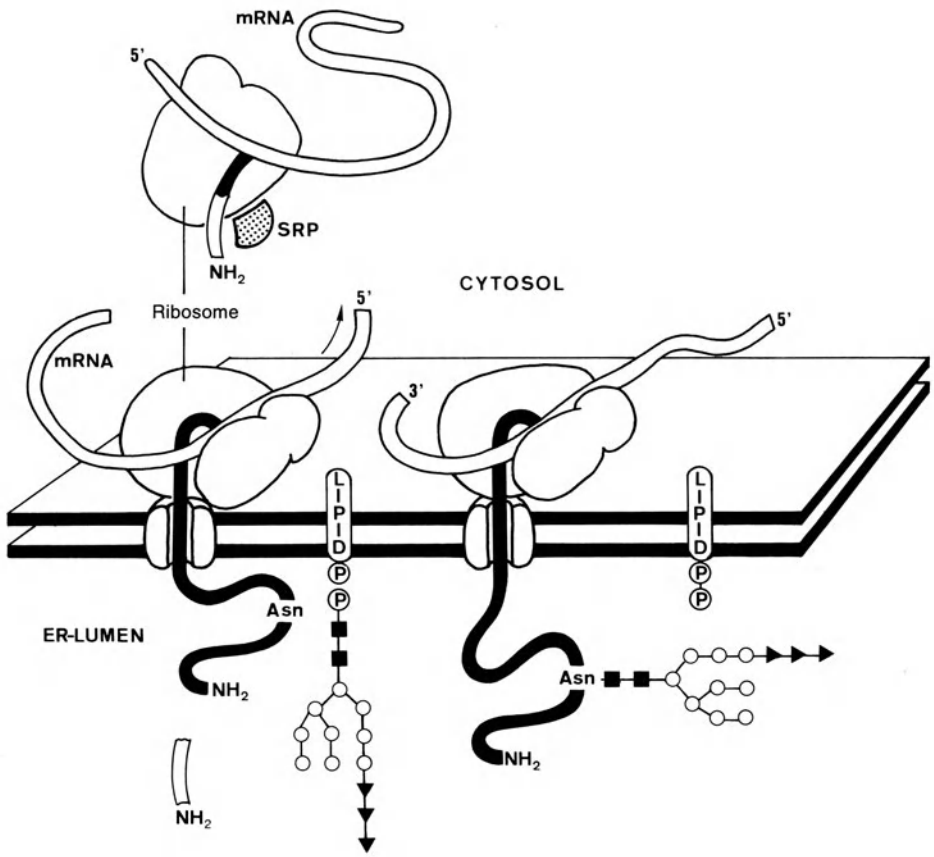


Fig. 3.3. The translocation of secreted proteins through the membrane of the endoplasmic reticulum (ER), and subsequent glycosylation. Involved in the process are: in the cytoplasm, a ribosome with the mRNA and a signal-recognition particle (SRP); and in the ER membrane, a binding complex consisting of a ribosome receptor, SRP receptor (docking protein) and the signal peptidase, and an oligosaccharide bound to a lipid (dolichol) and the glycosyltransferase. Translation begins on the cytoplasmic ribosome with the synthesis of the signal sequence, and is interrupted after approximately 70 amino acids by binding of the SRP; the translation inhibition is only removed after the ribosome becomes attached to the binding complex in the ER membrane, and the polypeptide chain extends into the ER lumen. There, the signal sequence is cleaved off and the carbohydrate chain transferred from dolichol to an asparagine residue

tion with the SRP receptor) and is therefore rather complex in structure. It consists of a 7 SL-RNA and six polypeptides of 9–72 kDa. The 7 SL-RNA has already been described (p. 23); the structures and functions of the polypeptides are only partially known. The SRP receptor is a heterodimer of 72- and 30-kDa subunits, and the signal sequence receptor is a glycoprotein of 35 kDa [215, 228].

The non-polar signal sequences of the secreted and membrane proteins are mostly 18–24 amino acids long; the shortest known signal sequence of only 13 amino acids is found in the β-crystallin of the mouse (Fig 3.4). Three regions can be distinguished in the signal sequence: the N-terminal region normally contains a positively charged, basic amino acid; the central region includes at least nine non-polar residues such as Phe, Ile, Leu, Met, Val or Trp; and polar, charged amino acids predominate in the C-terminal region. In the last and third from last positions (–1 and –3) one finds small, uncharged residues like Ala or Gly and also Ser, Cys or Thr, whereas in the second from last position (–2) there appears to be no such specificity. The "–1/–3 rule" perhaps fulfils special requirements of the signal peptidase [102]. Of the egg-white proteins secreted in the hen oviduct, ovotransferrin (conalbumin), ovomucoid and lysozyme are synthesized as pre-proteins with normal signal sequences; the signal sequence of ovalbumin, in contrast, lies within the chain and is not removed proteolytically. The interleukins IL-1 α and IL-1 β, released by activated human monocytes, also have no N-terminal signal sequence and are apparently secreted without involvement of the ER or Golgi [232].

```
                                                    -1  +1
Insulin (Myxine)              MALSPFLAAVIPLVLLLSRAPPSADT  RTTGH...
  "     (Lophius)               MAALWLQSFSLLVLLVVSWPGSQA  VAPAQ...
  "     (human)                 MALWMRLLPLLALLALWGPDPAAA  FVNQH...
Ovomucoid (chicken)           MAMAGVFVLFSFVLCGFLPDAAFG  AEVDC...
Ovotransferrin (chicken)           MKLILCTVLSLGIAAVCFA  APPKS...
Lysozyme (chicken)                  MRSLLILVLCFLPLAALG  KVFGR...
Serumalbumin (human)                MWKVTFISLLFLFSSAYS  RGVFR...
Ig H-chain (mouse)                  MKVLSLLYLLTAIPGIMS  DVQLQ...
Amylase (mouse)                        MKFFLLLSLIGFCWA  QYDPH...
β-Casein (cow)                         MKVLILACLVALALA  REQEE...
```

Fig. 3.4. Signal sequences of several secretory proteins [102]. The cleavage site lies between –1 and +1

The signal sequence of the polypeptide that extends out into the ER lumen is split off by a specific **signal peptidase** that is present in the ER membrane as an integral metalloprotein. Secretory proteins are already glycosylated during transport through the ER membrane (Fig. 3.3); however modification of the attached oligosaccharides (trimming) occurs in the Golgi cisternae. This process will be extensively described in Chapter 13. The completed glycoproteins are transported in vesicles to the cell membrane and released to the outside (exocytosis). Usually, exocytosis only occurs following an external stimulus. Transduction of the signal from the receptors on the cell membrane to the secretory vesicles may involve the activity of parfusin, a phosphoprotein of 63 kDa that is dephosphorylated on the stimulation of exocytosis. Parafusin was discovered in the ciliate *Paramecium tetraurelia* but now appears to be ubiquitous [239].

Membrane proteins have a similar fate, although, in contrast to the secretory proteins, they are not released into the ER lumen but instead remain bound to the ER membrane via the C-terminal region. If this orientation persists in the Golgi cisternae and the transport vesicles up to the time when the membrane of the transport vesicle fuses with the plasma membrane, then the N-terminus that was previously directed into the ER lumen will now point outward from the plasma membrane surface; in fact, most membrane proteins have this orientation.

Of the many hundreds of **mitochondrial proteins**, only a very few are produced in the mitochondria itself; the overwhelming majority stem from the cytoplasm. Their translocation is particularly problematic as they must be correctly distributed to four possible locations: matrix, inner membrane, intermembrane space, and outer membrane [99, 210]. Most of the cytoplasmic precursors carry an N-terminal mitochondrial target signal of 12–70 amino acids which, in contrast to the non-polar signal of the secretory and membrane proteins, is hydrophilic and has positively charged amino acids. The target signal is split off in the mitochondrial matrix by a soluble, heterodimeric metalloprotease of 55+52 kDa [204]. There are also mitochondrial proteins without a cleavable N-terminal target signal, e.g. the ADP/ATP transporter whose sequence contains at least three internal signals. Several cytoplasmic factors, including the heat-shock protein hsp70, are involved in the uptake of the precursor into the mitochondrium [132, 200]. Uptake requires energy, and insertion into the outer mitochondrial membrane occurs with the use of ATP by means of the general insertion protein (GIP); transport through the inner membrane into the matrix is dependent on the electrical potential of the latter. After cleavage of the target signal by the signal peptidase of the matrix, the matrix protein is pound to the heat-shock protein hsp60. In other proteins, following on from the N-terminal hydrophilic target signal is a less polar sequence that acts as a sorting signal and leads the protein out of the matrix back into the inner membrane or the intermembrane space [99, 165, 210].

The proteins of the different types of **microbodies** (peroxisomes, glyoxysomes, glycosomes) apparently have no cleavable signal sequence. The best examined of this type are certain glycolysis enzymes of *Trypanosoma brucei* and other Kinetoplastida; instead of being cytoplasmic these are completely or partly restricted to microbody-like cell organelles, the glycosomes (p. 515). Sequence comparisons between the glycosomal enzymes and cytoplasmic enzymes from the same or other animals should show which sequence characters lead to translocation into the glycosomes. *T. brucei* possesses one gene each for a cytoplasmic and a glycosomal isoenzyme of phosphoglycerate kinase; these agree by 95% in nucleotide sequence and by 93% in the amino acid sequence, but are only 44–46% similar to the homologous enzymes of yeast and man. The glycosomal isoenzyme differs from that in the cytoplasm by a 20-amino-acid C-terminal exten-

sion and an additional 13 positive charges [20]. The glycosomal phosphoglycerate kinase of *Crithidia fasciculata* also has a C-terminal topographical signal [264]. Compared with the homologous cytoplasmic enzyme of other eukaryotes, the glyceraldehyde-3-phosphate dehydrogenase in the glycosomes of *T. brucei* is five amino acids longer, but this does not apply to the aldolase or the triosephosphate isomerase; all three enzymes, however, have a particularly large number of positive charges, which apparently represent the translocation signal [20]. The alanine:glyoxylate aminotransferase AGT1 is located in the peroxisomes of the primates and hare-like species (lagomorphs), in the mitochondria of the carnivores, and in both organelles of the rodents. The mitochondrial target signal of the primates has been lost during evolution by mutation of the initiation codon ATG to ATA; the corresponding reverse mutation in humans leads to the lethal heritable condition of primary hyperoxaluria, because the location of the AGT1 in the mitochondria reduces its efficiency [268].

The enzyme proteins of the **lysosomes** either originate directly from the ER, from whence the lysosomes themselves come, or they are later taken up via "coated vesicles" from the cytoplasm. The markers for inclusion in lysosomes are apparently oligosaccharides with a terminal mannose-6-phosphate that is created by attachment of 1-phospho-N-acetylglucosamine to a terminal mannose residue, followed by cleavage of the N-acetylglucosamine. There are two types of mannose-6-phosphate receptors in the Golgi membrane, of which the larger (CI-MPR) is cation independent and the smaller (CD-MPR) requires divalent cations. The CI-MPR begins with an extracytoplasmic domain of 2269 amino acids made up of 15 repeats of a 147-amino-acid sequence and carrying 19 glycosylation sites. This is followed by a transmembrane segment of 23 amino acids and the C-terminal cytoplasmic domain of 163 amino acids. In the subunits of the homodimeric CD-MPR, the extracytoplasmic domain includes only 159 amino acids with five glycosylation sites, but shows significant sequence agreement (14–28 %) with the repeats of the CI-MPR. Its transmembrane segment has 25 amino acids, and the cytoplasmic domain 67 amino acids. Surprisingly, it has been found that the CI-MPR serves at the same time as a receptor for the insulin-like growth factor IGF-II [45, 144, 146, 163]. The proteins of the **cell nucleus** must possess internal non-cleavable signals as the nuclear membrane is dispersed during cell division. The signals consistently have one proline and several lysine residues; for example, in the antigen to the SV40 virus one finds the sequence -Pro-Lys-Lys-Lys-Arg-Lys-Val [2, 191, 251, 288].

3.3 Proteolysis

Proteolytic enzymes (proteases) have a variety of biological functions. In the digestion of food and the autolysis of dead cells, the amino acid constituents of all the proteins involved are similarly and completely released and thus become available again for cell metabolism. The complete cleavage of a polypeptide into its individual components requires the successive action of several proteases with varying specificities (Table 3.3). Initially, endopeptidases (proteinases) break the chain into smaller and smaller fragments by cleavage of internal bonds; several proteinases with different bond specificities are often involved. The resulting cleavage peptides are further step-

Table 3.3. Classification of the proteases [11, 189]. Only a few examples are given in each case; enzymes not of animal origin are shown within brackets. The characteristic amino acids of the active centres are given in parentheses: the position numbers refer to the first named enzyme

Exopeptidases: cleave bonds at the end of the peptide chain
Aminopeptidases: release the N-terminal amino acid
Carboxypeptidases: release the C-terminal amino acid
Carboxypeptidase A (Zn^{2+}, 255-Ile)
Carboxypeptidase B (Zn^{2+}, 255-Asp)
Dipeptidyl aminopeptidases: release the N-terminal dipeptide
Dipeptidases: cleave dipeptides
Endopeptidases or **proteinases:** cleave internal bonds of peptide chains
Serine proteinases I (102-Asp, 195-Ser, 57-Ser)
Chymotrypsin, trypsin, elastase, cathepsin G, acrosin, enteropeptidase, plasmin, kallikrein
Thrombin and other blood-clotting factors
Complement components
Serine proteinases II (32-Asp, 221-Ser, 64-His)
[Subtilisin]
Cysteine proteinases (25-Cys, 159-His, 158-Asp)
[Papain]
Calpains (calcium-dependent, papain-like proteinases)
Aspartate proteinases (33-Asp, 213-Asp)
[Penicillopepsin]
Pepsin, chymosin, cathepsins D and E, renin
Metalloproteinases I (ZN^{2+})
Collagenases
Metalloproteinases II (Zn^{2+}, 243-Glu, 231-His)
[Thermolysin]

wise degraded at both ends by amino- or carboxypeptidases, and the smallest peptides are finally cleaved by dipeptidases. As expected, a complex spectrum of digestive proteases is to be found in almost every animal species [32, 113, 128]. The proteases involved in general proteolytic processes show only limited specificity with regard to their target proteins and the amino acids that form the peptide bonds. For example, pig pepsin can cleave 1020 bonds from a total of 6910 bonds in 177 natural peptides (i.e. 14.8 %); trypsin, elastase and papain also have high cleavage potential.

The more specialized the biological function of a protease, the stronger are its substrate and bond specificities. So, for example, chymosin in the stomach of young mammals has not only general proteolytic activity but also the special property of cleaving specific bonds in casein, thereby precipitating milk proteins and allowing their digestion. Similarly, plasmin has particular tasks related to the solubilization of fibrin clots (fibrinolysis) and the inflammation process, and thus has a much narrower specificity than the closely related trypsin. Many biological processes include partial proteolysis in which only one or a few specific bonds in a certain protein are cleaved (Table 3.4). During evolution such highly specific proteases have arisen from enzymes with nonspecific proteolytic properties [189].

Intracellular proteolytic processes serve various biological functions: the release of amino acids from phagocytosed food particles by intracellular digestion, or from cellular proteins to supply energy during periods of hunger or intracellular osmoregulation; histolysis in connection with development, e.g. during metamorphosis in insects and amphibians; the destruction of superfluous, defective or abnormal proteins; general protein turnover; and partial proteolysis related to intracellular translocation of proteins, the maturation of enzyme and hormone precursors (Table 3.4), and restructuring of the cytoskeleton [18, 174, 212]. It is clear that living organisms must be protected from the uncontrolled effects of their own proteases, and there are several possible ways of achieving this:

1. Intracellular proteases can be restricted to special cell organelles like lysosomes, and thus are separated from the cytoplasm by membranes.
2. Proteases can be stored as inactive precursors (zymogens) that are activated only when required.
3. Proteases can be inactivated by binding to the animal's own protease inhibitors.
4. Individual proteins can be protected from proteases or, conversely, marked for degradation by post-translational modification.

Several different systems are available for intracellular proteolysis; however, little is known about these from a comparative biochemistry point of view. The **cathepsins of the lysosomes** are, as a rule, small glycoproteins of 20–40 kDa with rather acidic pH optima. The cathepsins B, L, H, M, N, S and T are cysteine proteinases, whereas D and E are aspartate proteinases. The lysosomal systems is not only responsible for the autolysis of the whole cell and degradation of the organelles, but also significantly involved in the turnover of cell proteins and the processing of proteins during translocation. The required specificity of proteolysis is achieved by selective uptake into the lysosomes, modulated by glycosylation of the proteins [18, 174]. Amongst the **extralysosomal proteases** are several that are ATP-dependent, e.g. cytosomal cysteine proteinases, which are stimulated by ATP without its hydrolysis, and mitochondrial serine proteinases, which require ATP hydrolysis for their activity [18]. Ca^{2+}-dependent calpain (p. 95) and the signal peptidases of the ER membrane and mitochondrial matrix are intracellular proteinases which are not involved in general proteolysis but have strict substrate or bond specificities [18].

Recognition signals for intracellular proteolysis may be contained in the amino acid sequence. In this way, for example, the signal peptidases of

Table 3.4. Examples of biological processes which include partial proteolysis [189, 245]

Protein secretion
Cleavage of the signal sequence of pre-proteins
Activation of enzyme precursors
Zymogen → active proteases
Prophenoloxidase → active phenoloxidase
Sequential activation of components of enzyme cascades
Blood clotting
Complement system
Hormone synthesis
Proinsulin → insulin (removal of the central C-peptide)
Proglucagon → glucagon (cleavage of the C-terminal octapeptide)
Angiotensinogen → angiotensin I → angiotensin II (see Fig. 8.5a, p. 302)
Kininogen → kallidin or bradykinin (see Fig. 8.5b, p. 302)
Multiple cleavage of macromolecular precursors of peptide hormones
Formation of macromolecular structures
Collagen synthesis
Formation of fibrin clots

the ER membrane and the proteases responsible for processing of hormone precursors and proenzymes recognize the pre- and pro-sequences of their substrates. In proteins with short half-lives (<2 h), one or more sequences, which do not appear in more persistent proteins, may be found that are rich in proline, glutamic acid, serine and threonine. The assumption that these are specific proteolysis signals goes under the name of the "PEST hypothesis", after the one letter symbols of these four amino acids (see Appendix) [230]. Oxidation by mixed-function oxygenases, deamidation, phosphorylation and the oxidation of SH- groups also act as signals for proteolysis. Finally, the quaternary structure of cell proteins plays a role in their susceptibility, e.g. the free globin chains in reticulocytes are more easily cleaved than is the complete tetrameric haemoglobin; certain complement factors are only activated by proteolysis when in the complexed form [212, 258].

The best-known recognition signal for extralysosomal proteolysis is conjugation with **ubiquitin**. This small polypeptide, with a length of 76 amino acids, was discovered in chromatin in 1977 as a constituent of a variant histone H2A (p. 33). As the name suggests, ubiquitin has been detected in all investigated eukaryotes and bacteria. Free ubiquitin is an α/β protein with an α-helix of 3.5 turns and a β-sheet structure of four strands. Ubiquitin is an extremely conserved protein; the sequence in mammals, fish and insects is identical (Fig. 3.5), and only differs in yeast at three positions. The most deviant ubiquitin sequence known is found in the ciliate *Tetrahymena pyriformis:* the three ubiquitins encoded in the gene pTU10 deviate from the usual animal sequence at eight or nine positions [190]. Ubiquitin is bound in an ATP-requiring reaction to the protein N-terminus or to a lysine-ε-amino group of the protein to be degraded; such protein conjugates may contain several ubiquitin molecules. The ubiquitin itself is recovered during proteolysis. From the general distribution and conservative evolution of ubiquitin, it may be concluded that ATP-ubiquitin-dependent proteolysis is a widely occurring mechanism. However, it is only really known in rabbit reticulocytes, where the system consists of a complex of ubiquitin and five to seven other polypeptides. The role of ubiquitin here is possibly not only the marking of proteins for degradation but also the repression of intracellular proteinase inhibitors [18, 174, 212]. In the reticulocytes, and in many other cells of vertebrates and insects, there are also ATP-dependent systems of proteolysis that are independent of ubiquitin [18]. Outside of its involvement in intracellular proteolysis, ubiquitin appears to have other functions; it is, for example, a heat-shock protein (p. 44) and comprises a part of the receptor that effects the uptake of lymphocytes into lymph nodes [221].

```
          10          20  T       30          40
MQIFVKTLTG KTITLEVEPS DTIENVKAKI QDKEGIPPDQ
                    T     D   S

          50          60          70      76
QRLIFAGKQL EDGRTLSDYN IQKESTLHLV LRLRGG
```

Fig. 3.5. Ubiquitin sequences. The illustrated sequence is that generally found in animals. *Above* A variant found in *Drosophila* differs in one position; and *below* the ubiquitin found in yeast differs in three positions [7]

Ubiquitin is the most conserved of all known proteins; in contrast, the ubiquitin genes vary in number and organization. Almost all eukaryotes possess several ubiquitin loci. These consist either of the coding sequence for a single ubiquitin with a C-terminal extension of 52–80 amino acids, or of several intronless, ubiquitin-coding sequences of 228 bp that are tandemly arranged without spacers. Examples of the number of ubiquitin sequences at the individual loci are in man 1, 3 and 9, in the chicken 3 and 4, and in *Drosophila* 1 and 18. Between 1 and 10 ubiquitin loci have been found in various breeding lines of the flagellates *Trypanosoma cruzi* and *Leishmania donovani*, which contain between 2 and more than 40 ubiquitin sequences [124, 143, 154, 227]. Repeated ubiquitin loci carry only one termination codon; thus, the primary translation product is a polyubiquitin that is then post-translationally reduced to single ubiquitin molecules by cleavage of the 76-Gly/1-Met bonds. In addition to the responsible proteinase, there is also a further ubiquitin-specific enzyme that cleaves ubiquitin from its conjugates [127]. In most polyubiquitins another amino acid follows behind the 76-Gly of the last ubiquitin sequence that inhibits conjugation of the precursor to other proteins; this terminal amino acid is missing in the polyubiquitin of the clawed frog *Xenopus laevis* [221].

The sequences of individual repeats become similar to each other by a special evolutionary process known as horizontal or concerted evolution (p. 123). Nevertheless, there are sequence differences between the repeats of a species which are all of the non-amino acid exchanging (synonymous) type; the number of repeats can also apparently vary between individuals of a species.

Comparisons of the repeat sequences within and between species allow conclusions to be drawn about the frequency with which sequence equalization through horizontal evolution has occurred; in the case of ubiquitin, it is on average 38 million years since the last such event in mammals, and only 11 million years in *Drosophila* [248].

Two large, proteolytic complexes are responsible for the degradation of ubiquitin-marked proteins. The 26S complex is about 1500 kDa and in vitro originates from three smaller subunits C1–C3. The well-known 650-kDa complex, termed "multicatalytic protease" because of its wide specificity, or "macropain" or "proteasome", is identical to the C3 component. This consists of subunits of 20–35 kDa, organized in a four-ringed cylinder, with each ring having up to eight subunits. The human proteasome in erythrocytes contains 13 different types of subunit, varying in both structure and specificity, although these apparently belong to the same multi-gene family [155, 202, 255]. The freshly isolated complex is stimulated by ATP, but this effect diminishes with storage. Multi-catalytic proteases are apparently ubiquitous; particles similar to mammalian proteasomes are found in archaebacteria, and they have been closely studied in lower vertebrates, insects and crustaceans [44, 59, 67, 73, 93, 185].

In the **classification of the proteases**, one can distinguish between exopeptidases, which cleave N- or C-terminal bonds, and endopeptidases, which attack internal bonds; finer characterization is based upon the structure of the active centre (Table 3.3). In the absence of information on amino acid sequences, several characteristic properties can be used to study the relationship between the proteases of the lower vertebrates and invertebrates and those of the mammals; these include, in particular, inhibition by synthetic and natural inhibitors, the specificity for certain low molecular weight, synthetic substrates, and cleavage of the insulin-B chain (Table 3.5). However, even proteases with very similar active centre structures are not necessarily closely related by evolution. Bacterial subtilisin is so different from the other animal and bacterial serine proteinases in both sequence and spatial structure that a common origin would appear to be excluded; this is clearly a case of convergent molecular evolution. Likewise, the bacterial thermolysin is not related to pancreatic carboxypeptidase, although in both cases the active centre contains an essential zinc atom [189]. Just how far the exopeptidases, as a group of proteins with similar catalytic properties, also represent a family of homologous proteins can not yet be determined in all cases due to the lack of sequence information.

Table 3.5. Typical properties of various classes and types of proteinases. The binding specificities refer to the amino acids which contribute the carboxyl group to the peptide bond

Serine proteinases: pH optimum alkaline; typical inhibitors: diisopropylfluorphosphate (DFP) and phenylmethanesulphonylfluoride (PMSF)
Chymotrypsins: specific for Tyr and Phe; typical inhibitors: tosylphenylalanine-chlormethylketone (TPCK), chymostatin and natural chymotrypsin inhibitors
Trypsins: specific for Arg and Lys; typical inhibitors: tosyllysine-chlormethylketone (TLCK), leupeptin and natural trypsin inhibitors
Elastases: specific for Ala, Leu, Gly, Val and Ile
Cysteine proteinases: pH optimum weakly acidic; typical inhibitor: p-chloromercurbenzoate (PCMB)
Cathepsin B: active against Z-Arg-Arg-NMec
Cathepsin H: active against Arg-NMec
Cathepsin L: active against azocasein and Z-Phe-Arg-NMec
Aspartate proteinases: pH optimum acidic; typical inhibitor: pepstatin
Pepsins: specific for Phe, Leu and Trp
Metalloproteinases: pH optimum alkaline; typical inhibitor: ethylenediametetraacetic acid (EDTA); EDTA inhibition relieved by Zn^{2+} or other divalent metal ions

Z-, benzoyloxycarbonyl-; -NMec, -N-methylcoumarin

3.3.1 Exopeptidases

Only the exopeptidases of the mammals have been described in any great detail [176]. The **aminopeptidases** N and A are Zn^{2+}- and Ca^{2+}-specific metalloenzymes, respectively; they are located in the microvillous membranes of the gut epithelium and kidney tubuli as symmetrical dimers with two non-polar domains [199]. The aminopeptidases involved in digestion have also been investigated in insects and several other invertebrates; like the corresponding vertebrate enzymes, they are found to be either cytoplasmic or membrane-bound in the microvillous layer of the gut cells, have a pH optimum of around 8 and are typical metalloenzymes [32, 113, 193]. The dipeptidyl aminopeptidases of the mammalian gut and kidney are also bound to the microvillous layer. These are serine enzymes that cleave dipeptides from the amino-terminus of natural and artificial

substrates. They have also been recorded in birds, amphibians, fish, insects and molluscs, but have been little investigated [95, 194]. Associated with this group of enzymes are the angiotensin-converting enzymes that release the octapeptide angiotensin II (causing an increase in blood pressure) from angiotensin I, and inactivate the nonapeptide bradykinin (which lowers the blood pressure). One form of 160–170 kDa is found in the lungs and various other tissues, and one form of 100–110 kDa, which is identical at the C-terminus with the lung enzyme, is found in the testis [64]. Various mammalian organs contain tripeptidyl hydrolases; for example, there is an enzyme in the hypophysis and ovary that sequentially cleaves tripeptides from growth hormone [139].

The **carboxypeptidases** of mammals are zinc enzymes of 34 kDa and are secreted from the pancreas as inactive precursors (procarboxypeptidases), although they also occur in the kidney and other organs. The activation reactions of the procarboxypeptidases have been recently described. Trypsin cleaves off a C-terminal activation peptide of 94 amino acids, from which a C-terminal arginine residue is then cleaved by the now active carboxypeptidase [284]. Two types of carboxypeptidase (CP), A and B, are distinguished and these differ in sequence by 53 % and have different specificities: CP-A cleaves off aromatic amino acids, whereas CP-B shows a preference for basic amino acids. The rat and other mammals possess two isoenzymes of type A (A1 and A2). Sequence comparisons show that the sole bovine carboxypeptidase A corresponds to type A1, and A2 is missing. In addition to types A and B, there are other carboxypeptidases that have special functions relating to the release of active hormones from their precursor forms; these enzymes are all homologous but show only 15–49 % sequence agreement [77, 271]. Carboxypeptidase are also widely found amongst the invertebrates as digestive enzymes and, like the vertebrate enzymes, are typical metalloenzymes [32, 113, 203]. The enzyme from the mid-gut gland of the crayfish *Astacus fluviatilis* is the only one of this type to have been sequenced so far. It has a length of 303 amino acids and a mass (without the zinc atom) of 33 899 Da; the sequence shows 44–45 % agreement with the bovine carboxypeptidases A and B. Like the bovine carboxypeptidase B, there is an aspartate residue at the primary substrate-binding site, whilst the rat and bovine carboxypeptidase A molecules have an isoleucine residue [277].

3.3.2 Serine Proteinases

Apart from in the serine proteinases, blood-clotting factors and complement components, serine is also found in the active centres of the alkaline phosphatases, esterases, lipases and haptoglobins. Just within the serine proteinases, one finds TCN (e.g. in trypsinogen) or AGY (e.g. in prothrombin) as the codon for the active serine. This leads to the hypothesis that the codons for the active serines have emerged by convergent evolution from various amino acid codons of evolutionary ancestors: e.g. from the threonine codon ACN, or from the cysteine codon TGY [22].

The mammalian pancreas secretes the zymogens of several serine proteinases with differing specificities, the chymotrypsins, trypsins and elastases (Table 3.5). Although the amino acid sequences of these enzymes differ by up to 60 %, their spatial structures agree to within 0.1 nm. The genes of these proteinases have, without doubt, all arisen from the same ancestral gene. The differences in specificity are due to only one or two amino acid exchanges in the region of the binding site: 189-Ser/Asp in chymotrypsin/trypsin; and 216-Gly/Val and 226-Gly/Thr in chymotrypsin/elastase. Similar enzymes are apparently present in all classes of vertebrates. Thus, the chymotrypsin, trypsin and elastase of the carp are all very similar to the mammalian enzymes in terms of amino acid composition, the partially known sequences and the number of disulphide bridges. The differences between the three types are notably greater than the differences found between fish and humans for just one type, suggesting that these enzyme types are already very old.

About 60 functionally different **types of serine proteinases** have so far been described in mammals and in several cases show very narrow substrate and binding specificities [11]: e.g. enteropeptidase (previously enterokinase), which converts trypsinogen into active trypsin by cleavage of a Lys/Ile bond; kallikrein (previously kininogenase), which produces the hormonally active bradykinin from kininogen by cleavage of -Lys/Arg- and -Arg/Ser- bonds (see Fig. 8.5b, p. 302); thrombin, which produces the fibrinopeptides A and B by cleavage of -Arg/Gly- bonds in the A(α) and B(β) chains of fibrinogen, and other components of the blood-clotting cascade; plasmin, which solubilizes fibrin by cleavage of -Lys/X- and -Arg/X- bonds and has a specificity, though it is somewhat narrower, for other proteins similar to trypsin; and, finally, many components of the comple-

ment system. New serine proteinases with novel functional characteristics are still being found, e.g. tonin from the submaxillary gland of the rat, or clipsin in the rat brain [152, 187]. A new serine proteinase has even been found in human pancreatric secretions and makes up 4–6 % of the total protein. It has a completely different specificity to the other pancreatic enzymes; in accord with its most prominent characteristic, it has become known as cholesterol-binding pancreas protease (CBPP) [265].

Sequence comparisons clearly indicate that the serine proteinases are all members of the same **protein super-family** [229]. In the case of prothrombin and other blood-clotting factors, the homology with the pancreatic enzymes is restricted to the C-terminal, catalytic region; the N-terminal extensions show large differences in sequence (Fig. 3.6). Haptoglobin and the α- and γ-subunits of the nerve growth factor (NGF) from the submandibular gland of the mouse (p. 304) are also serine proteinase homologues. The presently known structures of more than a dozen genes of this super-family vividly illustrate the importance of gene shuffling as a mechanism of molecular evolution (p. 122).

The super-family of the serine proteinases is clearly very old. The trypsin-, chymotrypsin- and elastase-like proteinases of many invertebrates (Table 3.6) are homologous not only to the vertebrate enzymes but also to certain serine proteinases of lower fungi and bacteria. Four proteinases are known from *Streptomyces:* one trypsin, which agrees in 34 % of its amino acids with the bovine enzyme; two proteinases A and B, which are only homologous to the mammalian enzyme in the central region but have a similar quaternary structure; and a fourth proteinase, which resembles subtilisin but belongs to another protein family. Because of their similar spatial structures, several bacterial serine proteinases are considered to be homologous to the vertebrate enzymes despite sequence differences of up to 80 % [120].

Typical of animal serine proteinases is the existence of enzymatically inactive **zymogens**, which become activated by the cleavage of an N-terminal peptide. As the bacterial enzymes have no zymogens, it is assumed that the evolution of the animal enzymes involved N-terminal extension, apparently on several independent occasions. Activation of trypsinogen by enterokinase involves the cleavage of an N-terminal hexapep-

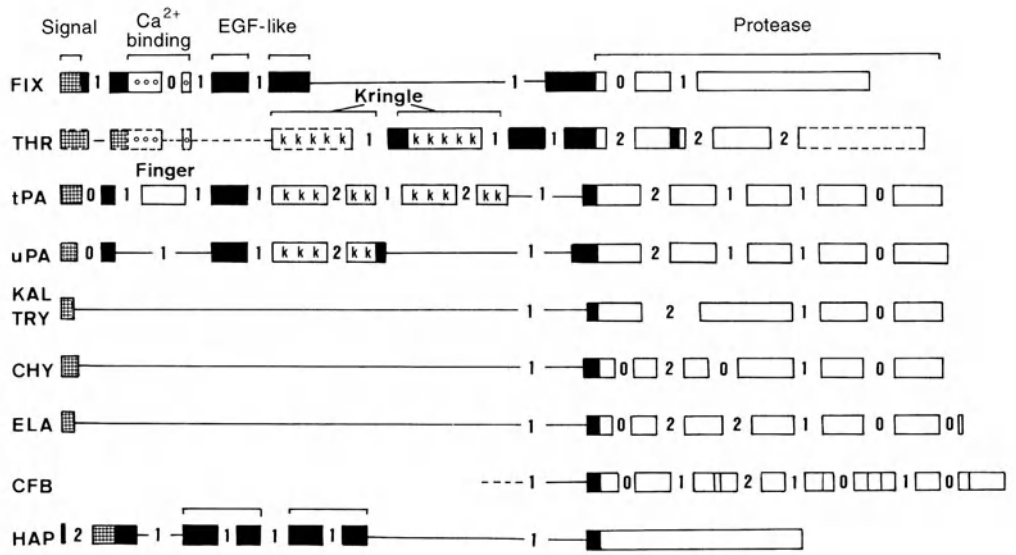


Fig. 3.6. The genes of various serine proteinases [229]. The C-terminal catalytic domains are significantly homologous; in contrast, there are large differences in both sequence and length in the N-terminal regions of different proteinases. This can be explained by the combination of exons of different origins (exon shuffling) during the evolution of the serine proteinases. Exons are *shaded* according to their homologues. *Signal* indicates the exon for the signal peptide; *Ca^{2+}binding, EGF-like* (epidermal growth factor), *kringle* and *finger* represent the exons for domains with characteristic structures. The lengths of the introns are not shown to scale, but the position of each intron in the reading frame is shown, i.e. whether it follows the first *(1)*, second *(2)* or third *(0)* nucleotide of a codon. *FIX*, human clotting factor IX; *THR*, human thrombin (the gene structure in the region indicated by the *dashed line* is not shown); *tPA*, human tissue plasminogen activator; *uPA*, pig urokinase; *KAL*, mouse kallikrein; *TRY*, rat trypsin; *CHY*, rat chymotrypsin; *ELA*, rat elastase; *CFB*, human complement factor B; and *HAP*, human haptoglobin

Table 3.6. Percentage sequence agreement between different proteases or their zymogens

	Agreement (%)	Reference
Serine proteinases		
Chymotrypsinogen A/chymotrypsinogen B (bovine)	78	[51]
Chymotrypsinogen A/trypsinogen (bovine)	43	[51]
Trypsinogen I (rat)/trypsinogen II (rat)	89	[164]
Trypsinogen I (rat)/trypsinogen (pig)	78	[164]
Trypsinogen (bovine)/*(Squalus acanthias)*	64	[51]
Elastase (pig)/chymotrypsinogen A (bovine)	40	[51]
Tonin (rat)/trypsin (bovine)	40	[152]
Prothrombin (bovine)/chymotrypsinogen A (bovine)	34	[51]
Clotting factor IX/trypsin (bovine)	44	[189]
Clotting factor X/trypsin (bovine)	41	[189]
Kallikrein/chymotrypsinogen A (bovine)	32	[51]
Kallikrein mGK-1/γ-NGF (mouse)	81	[116]
Plasminogen (human)/chymotrypsinogen A (bovine)	37	[51]
Trypsin I *(Astacus fluviatilis)*/trypsin (bovine)	44	[276]
VCP II *(Vespa crabro)*//chymotrypsin A (bovine)	37	[122]
VCP II *(V. crabro)*/VOP II *(V. orientalis)*	96	[122]
Trypsin *(Drosophila melanogaster)*//trypsin (bovine)	42	[50]
Trypsin *(D. melanogaster)*/chymotrypsin (bovine)	38	[50]
Trypsin *(D. melanogaster)*/VCP II *(V. crabro)*	35	[50]
Trypsin *(D. melanogaster)*/collagen protease *(Uca pugilator)*	31	[50]
Trypsin *(Streptomyces griseus)*/trypsin (bovine)	33	[120]
Protease A *(S. griseus)*/trypsin (bovine)	21	[120]
Cysteine proteinases		
Cathepsin B/cathepsin H (rat)	32	[270]
Papain/cathepsin H (rat)	40	[270]
Papain/cathepsin B (rat)	30	[270]
Aspartate proteinases		
Pepsinogen (bovine)/pepsinogen (pig)	81	[51]
Pepsinogen (chicken)/pepsinogen (pig)	62	[13]
Prochymosin (bovine)/pepsinogen (bovine)	60	[51]
Progastricin/pepsinogen *(Macaca fuscata)*	49	[130]
Cathepsin D/pepsinogen (human)	58	[69]
Renin/pepsinogen (human)	53	[69]
Penicillopepsin/pepsinogen (bovine)	30	[51]
Carboxypeptidases		
CP-A/CP-B (bovine)	47	[277]
CP-A (rat)/CP-A (bovine)	79	[277]
CP-B *(Astacus fluviatilis)*/CP-A (bovine)	45	[277]

tide. The human enterokinase is a glycoprotein of 300 kDa composed of three polypeptides of around 140, 102 and 54 kDa; the porcine and bovine enzymes consist of only two polypeptides and, with molecular masses of 200 and 150 kDa, are significantly smaller [158]. The activation peptides of the trypsinogens from different species are themselves homologous but differ from those of the chymotrypsinogens and proelastases (Fig. 3.7); in the components of the blood-clotting cascade, the activation peptides are actually longer than the active enzyme itself.

In mammals, there are not only numerous functionally different types of serine proteinases, but also non-allelic **isoenzymes** of individual types whose origins lie in gene duplication. The human pancreas contains two main trypsinogens, I and II, as well as a subsidiary component III, which show about 85% sequence agreement [272]; in the cow and pig one finds in each case two chymotrypsinogens (A and B) with about 20% sequence difference [11, 156]; in the rat there are two trypsinogens with 11–12% difference and two elastases with 42% difference; and in the mouse there are three trypsinogens [164, 189]. The heterogeneity of these enzymes is increased by post-translational modification. Thus, α- and then ω-trypsin are produced autoproteolytically from the initial β-trypsin. In addition to separation of the activation sequence, the

Trypsinogens:		
	FPLEDDDK	Rat I
	FPVDDDDK	Rat II, sheep, pig
	FPTDDDDK	Wild boar
	FPIDDDDK	Dog 2 and 3
	TPTDDDDK	Dog 1
	VPIDDDDK	Dromedary
	SSTDDDDK	Horse
	APDDDDK	Shark (Squalus acanthias)
	VDDDDK	Cow, goat, turkey
	DK	Man (cationic trypsin)
Chymotrypsinogens:		
	CGVPAIQPVLSGLSR	Cow A
	CGVPAIQPVLSGLAR	Cow B

Fig. 3.7. The activation peptides of several trypsinogens and chymotrypsinogens [51, 164]

activation of chymotrypsinogen and other zymogens also involves cleavage of the polypeptide chain into two fragments which are held together by a disulphide bridge; various chymotrypsinogens may be formed, depending on the site of cleavage [11].

Although the animal serine proteinases may all have a common origin, they nevertheless possess very different **physicochemical and catalytic properties**. In fish, as in many invertebrates, there are trypsins and chymotrypsins which, in contrast to most mammalian enzymes, have acidic isoelectric points, and accordingly are anionic at physiological pH values and very unstable at pH 3 [300]. The digestive enzymes of invertebrates are normally classified as chymotrypsin- or trypsin-like, according to whether their activity with synthetic, low molecular weight substrates or their inhibitor sensitivity corresponds to one or the other of the two mammalian enzyme types (Table 3.5). This classification system, however, does not suit all cases. Several proteinases (group II) from the mid-gut of the silkworm are clearly trypsin-like, whereas others (group III) are inhibited by chymostatin but not by tosylphenylalanine-chlormethylketone (TPCK), and are inactive with all tested synthetic trypsin and chymotrypsin substrates [237]. In the polychaete *Sabellaria alveolata* there are, in addition to the typical chymotrypsin- and trypsin-like proteinases, also enzymes that combine the specificity and sensitivity properties of both mammalian enzymes [208].

Many trypsin- and chymotrypsin-like proteinases of invertebrates have been characterized; understandably, these are very frequently from insects [50, 102, 122, 153, 233], but also from crustaceans [53, 87, 145, 276, 279], annelids, sipunculids and molluscs, etc. [159, 234]. Complete amino acid sequences are actually available for the enzymes from the wasp *Vespa orientalis*, the hornet *V. crabro*, the fruit-fly *Drosophila melanogaster*, the warble fly *Hypoderma lineatum*, the fiddler crab *Uca pugilator*, and the crayfish *Astacus fluviatilis* [50, 87, 122, 153, 276]. These invertebrate enzymes agree with the mammalian enzymes in only about 40 % of positions, but between themselves they show great similarity (Table 3.6). Contrary to previous assumptions, there are several cases of invertebrate zymogens, e.g. for the trypsin-like enzymes of the starfish *Derurasterias imbricata* and the clothes moth *Tineola bisselliella*, and for the cocoonase of the silkworm.

Many serine proteinases of invertebrates have acidic isoelectric points and are thus anionic at physiological pH values and unstable at low pH [87, 183, 233, 247]. Most serine proteinases have slightly alkaline **pH optima**; however, the proteinases of some dipteran and lepidopteran larvae have strongly alkaline pH optima (pH above 10), corresponding to the extreme pH values of the gut lumen [237, 247]. In contrast, proteinases which have been found in the digestive juices of the crab *Eriocheir japonica* are clearly trypsin-like in their inhibitor reactions but have weakly acidic pH optima [183]. Many invertebrate trypsins have the same pH dependence as the mammalian trypsin. This is true, for example, for the proteinases, known as hypodermin A and B, excreted by the warble fly *Hypoderma lineatum;* these aid penetration of the host's connective tissues. They have a pH optimum of 8–9 and are reversibly inhibited at pH 4.5, although they are not destroyed, as is the case for other invertebrate trypsins [153].

Several serine proteinases of invertebrates have the ability to cleave **collagen**; examples

include those of the warble fly, the fiddler crab *Uca pugilator*, the starfish *Pyenopdia helianthoides* and the turbellarian *Bipallium kewense* [87, 153]. The only vertebrate enzyme that may be referred to in this connection is a collagenase from canine pancreas [11]; the other typical vertebrate collagenases all belong to the metalloproteinases (Table 3.3) and have no digestive functions, although they are involved in developmental processes like, for example, degeneration of the tail in the tadpole. The collagen serine proteinases cleave the collagen triple helix about three-quarters of the way along the molecule in the same region as is cleaved by the metalloproteinases, although other peptide bonds are attacked. This collagen region apparently has a lower stability. Further degradation of the collagen that has been destabilized by cleavage is carried out by other proteinases.

Proteinases that create a way through the egg membrane for **spermatozoa** are found in Hydrozoa, various gastropods, the mussel *Mytilus*, sea urchins, ascidians, anurans of the *Bufo* genus, and various mammals [297]. Despite their similar biological roles, these enzymes have different binding specificities and therefore apparently arose independently during evolution. The acrosin of mammals, which is present as a zymogen (proacrosin) in the small body at the tip of the spermatozoon (acrosome), is initially activated by acrolysin and later activated autocatalytically; it allows the spermatozoon to find a way through the zona pellucida of the egg. The enzyme resembles trypsin in its amino acid sequence and specificity for -Arg/X- and -Lys/X-bonds. In the activation of pro-acrosin, a C-terminal peptide of 14 amino acids is cleaved initially. The cleavage of the peptide bond at 23-Arg then creates two different-sized subunits which are linked by a disulphide bridge. For the formation of the active enzyme, an 18-amino-acid peptide is first cleaved from the C-terminus and this is followed by a 43-amino-acid peptide that contains an unusual sequence of 23 consecutive proline residues [8]. The proteinase from the sperm of the sea urchin *Hemicentrotus pulcherrimus* corresponds rather more closely to chymotrypsin in its activity with synthetic substrates and its inhibitor sensitivity, whereas that of *Strongylocentrotus purpuratus* resembles acrosin and trypsin. In the digestion of the gelationous egg membrane, this proteinase is aided by an arylsulphatase [297]. In the spermatozoa of the ascidian *Halocynthia roretzi* there are, in fact, two very different trypsin-like proteinases; one strongly resembles the acrosin of the mammals and is so named, whilst the other (spermosin) has a much narrower specificity [241]. In the snails *Haliotis corrugata* and *H. rufescens*, the acrosome of the spermatozoa contains an amphophilic protein known as "lysin"; this produces a hole of 3 μm diameter in the egg membrane, although it possesses no enzyme activity. The lysins of the two species differ in 21 % of their 136–137 amino acids [283].

The egg membrane also presents a problem for the hatching embryo, and in the insects, echinoderms, teleosts and amphibians enzymes are involved in the process [157]. The **hatching enzyme** of *Xenopus laevis*, which would normally be classified as a serine proteinase on the strength of its inhibition by diisopropylfluorphosphate (DFP) and phenylmethanesulphonylfluoride (PMSF) (Table 3.5), is not inhibited by trypsin inhibitors, however, but by zinc ions and ethylenediaminetetraacetic acid (EDTA); its classification is thus unclear. A similar enzyme is found in *Rana chensinensis;* in contrast, the hatching enzymes of the teleosts are cysteine proteinases [282]. The best-known hatching enzyme is the cocoonase of the silkworms (Saturniidae), and this has been examined in detail in species of the genera *Bombyx* and *Antheraea*. The enzyme solubilizes only the sericin of the cocoon and not the fibroin, and it attacks other proteins with a trypsin-like specificity. This enzyme is produced as an inactive prococoonase of 28 kDa in special cells of the maxilla (galea) and is stored as a solid. The zymogen is dissolved and activated by an aqueous secretion prior to hatching. In some species, the cocoonase contains carbohydrates. The hatching enzyme of the sea urchin *Paracentrotus lividus* is a Ca^{2+}-activated protease [157].

3.3.3 Cysteine Proteinases

The best-known cysteine proteinases of vertebrates are the **lysosomal cathepsins**, which play an important role in intracellular protein degradation. Cathepsins B, H and L have been investigated most thoroughly, and cathepsins M, N, S and T to a lesser extent (Table 3.3). The cathepsins B, H and L belong to the same super-family as the plant proteinase papain and agree in up to 48 % of their sequences [226]. Cathepsin B is also apparently widely distributed in the invertebrates. Enzymes of this type are detectable, for example, in the lysosomes of protozoans, and allow the endozooic amoeba *Entamoeba histoly-*

tica to penetrate the gut wall of its host [162, 188], function as digestive enzymes in blood-sucking fleas, beetles and crustaceans [151, 184], and are involved in embryo development in various sea urchins and the brine shrimp *Artemia* as well as in insect metamorphosis [197, 290]. A proteinase, in several forms of 27–31 kDa, has been isolated from the trematode *Schistosoma mansoni*, and this corresponds to the vertebrate cathepsin B in its activity with synthetic low molecular weight substrates; it is responsible for the digestion of haemoglobin from the host's blood. *S. mansoni* possesses multiple genes for this enzyme [30, 49]. Cysteine proteinases with particular properties are found in the paragonimid *Paragonimus ohirai;* they are inhibited not only by the usual thiolprotease inhibitors, such as antipain and p-hydroxymercuriphenylsulfonate (p-HMPS), but also by α_1-antitrypsin and soybean trypsin inhibitor [298]. The recently discovered Ca^{2+}-dependent neutral proteases (CANPs) also belong to the family of cysteine proteinases. Because of their homology to the plant proteinase papain, these enzymes have been given the name **calpains** by the IUB enzyme commission. Calpains have been detected not only in the muscles and other organs of vertebrates but also in molluscs, crustaceans and insects [211]. The vertebrate enzymes are heterodimers of an 80-kDa catalytic subunit and a 28-kDa chain which is cleaved off during activation. On the basis of their calcium affinities, one can distinguish μ-calpain (type I), which requires only 1–70 μmol Ca^{2+}/l for half-maximum activity, and m-calpain, which requires 0.4–0.6 mmol Ca^{2+}/l. The breast muscle of the chicken contains a third "high m-calpain" type which needs 3.8 mmol Ca^{2+}/l [294]. The catalytic chain of m-calpain from chicken muscle has been sequenced via its cDNA. Four domains can be recognized in the sequence of 705 amino acids. The catalytic domain II (positions 81–320) has 30 % sequence agreement with other cysteine proteinases, and the regulatory domain IV resembles Ca^{2+}-binding proteins like calmodulin and troponin C; the functions of domains I and III are not yet known. The sequence agreement with proteins of other super-families suggests that calpains arose by gene fusion. The calpains apparently have no general proteolytic function but specific proteolytic activities that have not yet been described in detail. A sudden increase in intracellular Ca^{2+} concentration, as effected by the secondary messenger inositol-1,4,5,-trisphosphate, leads to the activation of this enzyme [18, 212].

3.3.4 Aspartate Proteinases

Characteristic of the aspartate proteinases is a pH optimum below 6 (acidic proteinases), the presence of two especially reactive aspartate residues, and inhibition by the pepstatin of *Streptomyces* (Table 3.5). Two classes of acidic proteinases are found in the mammalian stomach: the pepsins and gastricsins, which serve to digest proteins in the adults; and the chymosins, which precipitate casein in the stomachs of young breast-feeding animals. The chymosins show such low general proteolytic activity that undigested milk immunoglobulins can be harvested from young animals. The **gene and protein sequences** have been determined for pepsinogens and prochymosins in various mammals, the chicken and tuna fish [101, 117, 129, 130, 216, 273]. The various stomach proteinases agree in about 40 % of amino acid positions and are clearly homologous. Comparison of the three enzyme types shows that pepsinogen C (progastricsin) arose during evolution before the separation of prochymosin and pepsinogen A (Table 3.6). Not only do cathepsins D and E, renin and other intracellular aspartate proteinases of other species belong to the same super-family, but it also includes the fungal enzyme penicillopepsin, which has 24 % similarity to the stomach proteinases. The region around the active aspartate residue is particularly conserved.

The aspartate proteinases often show a high degree of **heterogeneity**. This may be due to the existence of several genes but also involves post-translational modifications such as phosphorylation, glycosylation or partial autolysis. Thus, in the pig stomach, for example, there are four different pepsins and pepsinogens. The predominant pepsin A is mainly secreted in the region of the fundus and has a pH optimum of 2; pepsins B and C are released primarily into the antro-pyloric part of the stomach, have a somewhat narrower specificity than pepsin A, and have a pH optimum of 3; pepsin D is the phosphate-free form of pepsin A. The so-called gastricsin of primates corresponds to pepsin C of the pig [267]. Multiple pepsins are known in other mammals including man and the chicken, but the homologies of the individual types in different species are uncertain. Multiple chymosins are also found; in calves there are two chymosins (A and B) with different sequences, and form C arises by autolysis. In pups of the seal *Pagophilus groenlandicus* there are four chymosins which, surprisingly, also appear in the adults [246]. In man, there are

marked individual differences in the relative proportions of the isoforms of pepsinogen A, and these are the result of differences in the number of genes [266].

The polypeptide chains in molecules of the pepsin-like proteinases form two lobes that are bound together by two antiparallel β-strands; the active aspartate residues sit opposite each other in the groove between the two lobes [121, 129]. Similar to the three-dimensional structures, the internal symmetries of the nucleotide and amino acid sequences also support the conclusion that gene duplication and fusion occurred during the evolution of these proteinases [108]. An N-terminal pro-sequence is cleaved off during the **activation** of the zymogens that form in the stomach. The pro-sequence of the pepsinogen in mammals is 48 amino acids long and in the chicken has 44 amino acids; it is normally sequentially removed in several short fragments, except in *Macaca fuscata*, where it is removed as a whole by one cleavage reaction [129]. Two activation peptides are removed from the progastricsin of *M. fuscata:* one is 25 and the other 18 amino acids long [130].

The further away from the mammals in the vertebrate hierarchy, the less comparable are the **stomach proteinases** with those of the mammals. The chicken pepsinogen is still quite similar to pig pepsin A, with 62 % identical amino acids. This is a glycoprotein that is electrophoretically heterogeneous due to a variable number of sulphate residues in the carbohydrate fraction [13]. In the frog *Rana catesbeiana*, the same four pepsinogens are produced in the oesophagal glands and in the stomach wall. Their molecular masses of 31–34 kDa are significantly lower than that of the pig pepsinogen A (43 kDa), although this omission of about 100 amino acids does not reduce either the catalytic activity or the immunological cross-reactivity to the human pepsinogen [250]. The two "gastricsinogens" of the marine pike *Merluccius* have a similar reduced size. The stomach proteinases of the cartilaginous and bony fish also differ from those of the mammals in other properties, e.g. in the presence of a chymosin-like activity, in the reduced activity with certain synthetic substrates, and in the greater resistance to alkali [80, 91].

Renin is an aspartate proteinase of extremely narrow specificity whose pH optimum, in contrast to all other enzymes of this class, extends over the neutral range (pH 5–8). The enzyme is produced in juxtaglomerular cells in the kidney and released into the blood plasma in response to certain signals. There it cleaves off angiotensin I from the dodecapeptide angiotensinogen; by the action of a "converting enzyme", the former is then converted to angiotensin II, the most effective of all blood-pressure increasing substances (see Fig. 8.5, p. 302). Whilst the renin activity of the kidneys is actually quite low, high activities are found in the submaxillary glands of males of several mouse strains. However, the glycosylated enzyme of the kidney and the carbohydrate-free enzyme of the salivary gland are encoded by different genes; the sequences of the two enzymes differ by only 3 %, with the glycosylation site, as expected, being amongst the 21 exchanged amino acids. In man there is only one renin gene, the product of which shows 69 % agreement with the kidney enzyme of the mouse. Thus, it would appear that the gene duplication which led to the formation of distinct kidney and salivary gland renins in the mouse occurred about 12 million years ago, assuming that the evolutionary lines of man and mouse separated about 80 million years ago [214].

The **cathepsins D and E** are closely related to the other aspartate proteinases (Table 3.6), but are the only members of this family to be found in the lysosomes [18]. They are involved, for example, in tail shortening during the metamorphosis of the tadpole to the frog [186]. Similar intracellular enzymes have been detected in a range of very different invertebrates, from the Protozoa upwards [138, 196, 236]. In insects, they are apparently involved in the lysis of the larval tissues during metamorphosis [138]. In the shelled molluscs they are particularly active in mantle tissues and may therefore have something to do with shell formation [196]. In the Hemiptera, on the other hand, cathepsin D appears to function extracellularly as a digestive enzyme, helped in the case of the blood-sucking forms by cathepsin C, which is otherwise normally intracellular and belongs to the cysteine proteinases [112].

3.3.5 Metalloproteinases

The best-known metalloproteinases are the vertebrate collagenases, which are inhibited by EDTA and reactivated by Zn^{2+} [71]. There are several types of such enzymes in the intercellular matrix and they have varying specificities for different matrix components. The collagenases secreted by fibroblasts and white blood cells cleave a particular bond in the triple helix of collagens type I, II, III and VII at a position about three-

quarters of the way along the molecule; the resulting cleavage products may then be further degraded by other proteases. The collagenases from human fibroblasts and neutrophils show about 57 % sequence agreement [100, 166]. The gelatinases are not active against the fibrillar collagens but are active against collagens type IV, V and VII as well as fibronectin; stromelysin predominantly degrades proteoglycans. All these enzymes, however, are homologues and show up to 50 % sequence agreement [39, 240].

The elastases of the serine enzyme group are found in all vertebrates above the Agnatha; in contrast, there is an enzyme, first discovered in 1984, that is restricted to the teleosts, cleaves typical elastase substrates, but belongs to the zinc-containing metalloenzymes [301]. In several vertebrates there are collagenolytic serine proteinases [153]; however, one of the two collagenases of the starfish *Pyenopodia helianthoides* is a metal proteinase that is, in fact, better activated by Ca^{2+} than by Zn^{2+} [4]. Four proteinases have been isolated from the digestive fluids of the spider *Argiope*, these have an alkaline pH optimum and are all inhibited by EDTA. The inhibition can be relieved by Zn^{2+}, and they are, therefore, apparently metalloproteinases. They have a particularly low molecular mass of 17 kDa. Two of the four enzymes cleave the fibroin of the web, and all the enzymes have elastase activity [137].

An enzyme discovered in 1967 in the hepatopancreas of the crayfish *Astacus fluviatilis* is unique amongst the zinc proteinases. The molecular mass of this enzyme is 22.6 kDa and there is one zinc atom per mole; the previously described molecules of only 11 kDa were artefacts caused by rapid autolysis below pH 4. The polypeptide chain of 200 amino acids shows no similarity to other protein sequences; only one short stretch (amino acids 90–99) is similar to the metal-binding site of thermolysin. The pH optimum is about 8, and the artificial heptapeptide dansyl-Pro-Lys-Arg-Ala-Pro-Trp-Val, which is cleaved at the Arg/Ala bond, is processed much better than natural protein substrates. Other proteins with immunological similarities to the *Astacus* enzyme have been detected in other crustaceans. *Astacus fluviatilis* also possesses trypsins and carboxypeptidases which, in this case, are very similar to the mammalian enzymes [260, 278].

3.4 Proteinase Inhibitors

Natural inhibitors are known for many different types of enzyme [286], but only the proteinase inhibitors appear to be ubiquitous. Although their biological role can, in general, be defined as the prevention of undesirable proteolysis, the details are often not understood. Proteinase inhibitors are usually discovered by observation of the inhibition of common proteinases like trypsin or papain, but the actual target enzyme may remain obscure. In only a minority of cases is the biological function apparent: the trypsin inhibitors prevent the premature activation of trypsinogens and other zymogens in the pancreas; the proteinase inhibitors of the blood plasma inhibit the proteolytic cascade of blood clotting and complement activation, and the release of active hormones from precursors. The plant proteinase inhibitors serve in the defence against the digestive system of herbivorous animals. With a single exception, the proteinase inhibitors are specific for one class of proteinases. Only the α_2-macroglobulins[1] (α_2M) inhibit all classes of proteinases and they also differ from all other proteinase inhibitors in their mode of action. Many inhibitors of animal serine and cysteine proteinases have already been sequenced but only one aspartate proteinase inhibitor has been sequenced; this is a pepsin inhibitor from the roundworm *Ascaris suum*. In this latter case, the sequence of 149 amino acids has no homology to other known proteins [170].

3.4.1 Serine-Proteinase Inhibitors

The presently known inhibitors of serine proteinases belong to several different **protein superfamilies** (Table 3.7) on the basis of their sequences and spatial structures, but all have the same mode of action; they thus serve as a good example of convergent protein evolution. The organization of the disulphide bridges suffices to distinguish the individual families, but the absolute definition of the phylogenetic relationships of individual inhibitors requires knowledge of the complete amino acid sequences. In addition to the examples given in Table 3.7, more than 100 other, less well-known serine-proteinase inhibitors have been described and there may possibly

1 The Greek letters and numbers refer to the localization of the inhibitors in electrophoregrams of plasma proteins.

Table 3.7. Protein super-families of the serine-proteinase inhibitors [10, 16, 41, 55, 62, 78, 106, 109, 111, 133, 134, 179, 213, 243, 280, 296]

1. Serpins
α_1-trypsin inhibitor (α_1AT): 394aa, inhibits elastase, P_1 = Met
α_1-antichymotrypsin (α_1AChy): 430aa, P_1 = Leu
Antithrombin III (AT-III): 423aa, inhibits thrombin, P_1 = Arg
α_2-antiplasmin (α_2AP): 452aa, inhibits plasmin with P_1 = 364-Arg, and chymotrypsin with P_1 = 365-Met
Nexin from glial cells: 378aa, P_1 = Arg
Contrapsin: only known from mouse, 215aa, P_1 = Lys
Ovalbumin: 385aa, P_1 = Ala
Angiotensinogen: about 450aa, no inhibitor activity, P_1 = Gly
Manduca secta haemolymph: 392aa (including the signal sequence), inhibits elastase; two similar inhibitors for chymotrypsin and trypsin

2. Bovine pancreas trypsin inhibitors (BPTI, Kunitz type; Fig. 3.8)
BPTI: apparently only in the Bovidae, in all bovine organs, 58aa
Colostrum trypsin inhibitor: 67aa, in cattle and pigs
Inter-α-trypsin inhibitor: in the blood plasma of mammals, heterodimeric or -trimeric glycoproteins of 240 kDa
Lipoprotein-associated coagulation inhibitor: 276aa, three inhibitory domains, inactivates the blood-clotting factor X_a
Proteinase inhibitors in snake toxins: 57–65aa, toxic but in some cases without inhibitory acitvity
Chelonianin from the egg albumin of an unidentified Red Sea turtle species: 110aa, with two domains; the first domain (58aa) with P_1 = Lys inhibits trypsin, the second domain (52aa) inhibits subtilisin and belongs to a new inhibitor family
Nerve growth factors from mammals and snake toxins
Helix pomatia slime: inhibitor K with 58aa, and further inhibitors with P_1 = Lys or Arg
Bombyx mori haemolymph: silkworm chymotrypsin inhibitor (SCI)-III with 63aa, and further chymotrypsin inhibitors

3. Secretory pancreas trypsin inhibitors (PSTI, Kazal type)
PTSI: 56aa, apparently present in all vertebrates in the zymogen granulae of the pancreas; inhibits trypsin but not enterokinase
"Acrosin inhibitors" from spermatozoa and seminal plasma of mammals: 57–63aa, target enzyme(s) unknown
"Many-headed" submandibular gland inhibitor in the dog: 115aa, two domains: the first (P_1 = Arg) inhibits trypsin, the second (P_1 = Met) inhibits chymotrypsin, subtilisin and elastase
Ovomucoid in avian egg albumin: 186aa, three domains with species-specific specificity
Ovoinhibitor in avian egg albumin and blood plasma: in the quail *Coturnix japonica* there are six domains, the first three of which inhibit trypsin (P_1 = Arg in all cases), and the other three inhibit chymotrypsin, subtilisin and elastase (P_1 = Tyr, Met, Met)
Testudin from turtle eggs: two domains
Bdellins from the leech *Hirudo medicinalis:* small (5–7 kDA) and large (about 20 kDa)
Elastase inhibitor from *Anemone sulcata:* 48aa

4. Hirudins
Hirudins from *H. medicinalis:* 65–66aa, inhibit thrombin

5. Ascaris trypsin inhibitors
A trypsin inhibitor (66aa) and several chymotrypsin inhibitors in *A. lumbricoides*

6. Eglins from *H. medicinalis*

7. Antistatins and ghilantens from the leeches *Haementeria officinales* and *H. ghilianii*

form further protein families. It is clear that the **nomenclature** of the serine-proteinase inhibitors is problematic; the families are named after the first-discovered inhibitor and its discoverer. Individual inhibitors are named according to both the organ in which they were first detected, even when they also occur elsewhere, and the proteinase used for the assay, even though the natural target proteinase is in most cases still unknown. Thus, for example, the ovoinhibitor of the chicken is not restricted to the egg but also appears in blood plasma; the so-called trypsin or chymotrypsin inhibitors undoubtedly also have other target enzymes.

The **action** of the serine-proteinase inhibitors involves reversible binding of the active site of the inhibitor to the active centre of the proteinase. Compared to the reaction of the proteinase with a real substrate, the affinity for the inhibitor is extremely high (low K_m) and the cleavage reaction very slow (low k_{cat}). Furthermore, the chains are held together after cleavage by a disulphide

Fig. 3.8. The N-terminal sequences of several proteinase inhibitors from the family of bovine pancreas trypsin inhibitors (Kunitz type) with the reactive bond P_1-P'_1 [136, 238]. *PTI*, bovine pancreas trypsin inhibitor; *CTI*, bovine colostrum trypsin inhibitor; *SCI-III, Bombyx mori* chymotrypsin inhibitor III; *HpTI, Helix pomatia* trypsin inhibitor; *Vr, Vipera russeli* toxin inhibitor II; *Va, Vipera ammodytes* toxin inhibitor; *Nn, Naja nivea* toxin inhibitor; *Chel*, chelonianin

bridge between the reactive sites. Because both the intact and cleaved inhibitor can bind to the proteinase, the latter is permanently inhibited. Interestingly, an inhibitor of one proteinase can be a good substrate for another: thus, for example, the bovine pancreatic trypsin inhibitor (BPTI) is rapidly cleaved at the reactive 15-Lys/16-Arg bond by a trypsin-like enzyme of the starfish *Dermasterias imbricata* [66]. The amino acids making up the reactive site are referred to as P_1 and P'_1 (Fig. 3.8); the P_1 amino acid determines the specificity of the inhibitor for trypsin (P_1 = Lys, Arg), chymotrypsin (P_1 = Phe, Leu, Met), elastase (P_1 = Ala, Ser, Met, Val) or thrombin (P_1 = Arg) [26]. Many of the inhibitors are made up of several domains, each with one reactive site; the number of domains may vary between two and six (many-headed inhibitors) (Table 3.7). The reactive sites of the proteinase inhibitors exhibit a particularly high variability in their evolution; active centres are otherwise usually highly conserved. Substitution of P_1 does not result in a loss of inhibitory activity but only a change in specificity; in general, substitution of P'_1 has no effect on either activity or specificity (only proline is not functional) [106, 150].

Proteinase inhibitors comprise about 10 % of the total protein in human blood plasma (Table 3.8). Plasma shows inhibitory activity against proteinases of all classes: in addition to the universal inhibitor α_2M and various serine-proteinase inhibitors, there are also inhibitors of cysteine proteinases and collagenolytic metalloproteinases, although these have relatively low activities. Almost all the **blood plasma inhibitors** have such a broad specificity that their biological roles are not really recognizable. Although there are some species-specific differences, all mammals have a spectrum of inhibitors similar to that of man; the rich inhibitor spectrum of birds, however, is not at all comparable with that of the mammals, and comparisons with the lower vertebrates are even more difficult to make.

The protein super-family of the **serpins** contains not only various proteinase inhibitors but also the angiotensinogens and ovalbumins, which have no proteinase inhibitory activity. All these proteins are significantly homologous. Human α_1-antitrypsin (α_1AT) shows 42 % agreement with human α_1-antichymotrypsin (α_1AChy), 28 % with AT-III, and 24 % with chicken ovalbumin. The exon/intron organization of the various serpin genes is very different; one finds eight introns (plasminogen activator inhibitor, PAI-1), seven introns (ovalbumin, PAI-2), five introns (antithrombin, AT-III) or four introns (α_1AT, α_1AChy, angiotensin) in various positions [218].

The α_1AT (or α_1 proteinase inhibitor) from human blood plasma consists of one polypeptide chain of 394 amino acids which is glycosylated at three positions. There are more than 50 human alleles [125]. α_1AT acts on all serine proteinases with variable effectiveness but its physiological substrate is the elastase released by neutrophil leukocytes. The P_1/P'_1 active centre has the sequence 358-Met/359-Ser. The rat α_1AT is 70 % similar to the human protein and 80 % similar to that of the mouse [29]. In the house mouse, *Mus musculus*, as in other mammals, α_1AT is synthe-

Table 3.8. The most important proteinase inhibitors in human blood plasma [189]

	Concentration (mg/l)	Size (kDa)
α_1-trypsin inhibitor (α_1AT)	2900	52
α_1-antichymotrypsin (α_1AChy)	500	69
Inter-α-trypsin inhibitor (IαI)	500	160
α_2-antiplasmin (α_2AP)	70	70
Antithrombin III (AT-III)	240	65
C1 inactivator	240	70
α_2-macroglobulin (α_2M)	2600	720

sized exclusively in the liver; in contrast, in the related species *M. cairoli* the concentration of α_1AT mRNA is almost as high in the kidney as in the liver. Both genes are expressed in the liver of hybrids, but only the *M. cairoli* gene is expressed in the kidney; thus, a *cis*-active element appears to be responsible for the tissue specificity. The inhibitor synthesized in the liver is released into the blood plasma, whereas that in the kidneys enters the urine [15]. Horse plasma contains a pre-α_2-elastase inhibitor that only inhibits elastase and trypsin. This is a hybrid molecule of α_1AT and a horse-specific α_2/β_1 glycoprotein [209].

In man, there is only one gene for α_1AChy, in contrast to the multi-gene family of the mouse and rat [10, 106]. The serpins involved in the regulation of blood clotting are AT-III, α_2-antiplasmin (α_2AP) and the plasminogen activator inhibitors PAI-1 and PAI-2; the lipoprotein-associated coagulation inhibitor that inhibits factor X_a belongs to the Kunitz type [296]. α_2AP possesses two overlapping P_1/P'_1 sites, of which 364-Arg/365-Met reacts with plasmin and 365-Met/366-Ser with chymotrypsin [213]. Despite their similar specificities, PAI-1 and PAI-2 have completely different structures; PAI-2, like the closely related ovalbumin, is secreted without cleavage of a signal sequence [261, 299]. Gliae cells from rat brain secrete the proteinase inhibitor nexin, which promotes neurite growth. Rat nexin agrees in 84 % of its sequence with human nexin, 32 % with AT-III and 25 % with α_1AT [254]. In addition to α_1AT in the blood plasma of the mouse, there is a further trypsin-specific inhibitor, contrapsin, which, despite the difference in specificity, is more similar to human α_1A-Chy (60 %) than to mouse α_1AT (44 %) [106].

The **family of bovine pancreas trypsin inhibitors** (BPTI family, Kunitz type) includes not only inhibitors from blood plasma but also those from snake toxins (Fig. 3.8), which will be dealt with in Chapter 9. These polypeptides are, without exception, toxic; several of them, e.g. the dendrotoxins from the black mamba, *Dendroapsis polylepsis*, have lost their proteinase inhibitory activity [60]. The chelonianin from the eggs of a Red Sea turtle species appears to be the result of gene fusion (exon shuffling); the two domains belong to separate protein super-families, one to the Kunitz family and the other to a previously unknown family [134].

In the egg albumin of birds, the ovomucoid which belongs to the **family of secretory pancreas trypsin inhibitors** (PSTI family, Kazal type) is of particular interest from a comparative biochemistry point of view. The inhibitory effect of the ovomucoids on trypsin (T), on the one hand, and on chymotrypsin, subtilisin and elastase (CSE), on the other hand, differs greatly between species. In the chicken, one T is bound per molecule of inhibitor, in the golden pheasant it is one C, in the turkey one T and one CSE, and in ducks there are two T's and one CSE. The explanation for this curious variation lies in the fact that the ovomucoid chain consists of three domains whose reactive sites are T-specific, CSE-specific or neutral, according to the nature of P_1. It is particularly clear in this case that the amino acid P_1 of the reactive centre varies considerably in its evolution. In the completely sequenced ovomucoid of the chicken, the three domains occupy positions 1–64, 65–130 and 131–186 [135]. The sequences of the third domain in 125 bird species have been compared. 45-Asn is always completely or partially glycosylated, except in the ostrich, which has a carbohydrate-free serine at this position. In the pheasants, the ovomucoid molecule carries the additional amino acids 134-Val and 135-Ser; this variation stems from inaccurate splicing of an intron/exon boundary [150]. Other proteinase inhibitors of the Kazal type also contain several domains (between one and six), each with a reactive centre (many-headed inhibitors) (Table 3.7). The two-headed inhibitors, for example, are quite well known; these were discovered in canine saliva but are also found in various cat species [223].

In the case of only a few invertebrate serine-proteinase inhibitors is there enough information to allow their classification into known or new families (Table 3.7). The proteinase inhibitors from the slime of the vineyard snail, *Helix pomatia*, are amongst the best known; the Kunitz-type inhibitor (K) has in fact been sequenced. There are three to four different inhibitors in the egg albumin glands of the investigated pulmonate snails, and the spectrum of activity of these against different proteinases varies with the species [198]. In the insects, the inhibitors from larval haemolymph of the moths *Bombyx mori* and *Manduca sexta* have received particular attention (Table 3.7). In *B. mori*, at least 15 chymotrypsin inhibitors (SCI) can be separated electrophoretically, and SCI-I, SCI-II and SCI-III have been sequenced and identified as belonging to the Kunitz type [54]. In *M. sexta*, four 47-kDa serpins with different specificities have been identified, of which the elastase inhibitor has been sequenced. In addition, the haemolymph con-

tains two Kunitz-type trypsin inhibitors [133, 219]. In the haemolymph of the horseshoe crab, *Limulus polyphemus*, a novel trypsin inhibitor has been detected and constitutes a new class of protease inhibitors [58]. The best-known invertebrate proteinase inhibitors are the hirudins, the inhibitors of blood-clotting found in the saliva of the leech *Hirudo medicinalis*. They inhibit both thrombin and trypsin and are present in more than ten isoforms, all of which are 65 amino acids long and differ in just a few positions [243]. In the anterior gut, but not in the saliva or the mid-gut of *H. medicinalis*, are two further types of proteinase inhibitors which perhaps have antibacterial functions: the bdellins belong to the Kazal group and appear as smaller (5–7 kDa) and larger (20 kDa) variants. The eglins have a length of 70 amino acids and belong to their own super-family [9]. In the saliva of the North American blood-sucking leech, *Haementaria officinalis*, is an inhibitor of the blood-clotting factor X_a; this is a protein of 119 amino acids, including 20 cysteine residues, belongs to a new super-family and has been named "antistasin". The inhibitor known as "ghilanten" from the closely related species *H. ghilinaii* is very similar [16, 62]. A highly specific X_a inhibitor has also been detected in the tick *Ornithodorus moubata;* this is a peptide of 60 amino acids of the Kunitz type [291].

The roundworm *Ascaris lumbricoides* is equipped with inhibitors against all proteinases present in the host gut lumen: in each case there are several inhibitors of low molecular weight (7–20 kDa) for pepsin, chymotrypsin and elastase, and trypsin. On the basis of their sequences, the trypsin and chymotrypsin inhibitors have been allocated to a new super-family. The horse roundworm, *Parascaris equorum*, also produces inhibitors, about which little is as yet known [170]. A strictly elastase-specific inhibitor from the sea anemone *Anemonia sulcata* has been identified as an atypical Kazal type on the basis of its sequence [280]. An inhibitor that inhibits trypsin, but not chymotrypsin or thrombin, has been isolated from the coelom fluid of the starfish *Asterias forbesi* [167].

3.4.2 Cysteine-Proteinase Inhibitors

The inhibitors of the vertebrate cysteine-, aspartate- and metalloproteinases have been much less investigated than those of the serine proteinases, and comparative biochemical data are almost completely absent. **Cystatin** from chicken egg albumin was the first representative of a super-family of cysteine-proteinase inhibitors of the same name, many more examples of which have since been found in man and other mammals. Three, or perhaps four, families of closely related proteins can be distinguished within the cystatin super-family [38, 220]. Cystatins of type 1 (stefins) are molecules of about 100 amino acids (11 kDa) without disulphide bridges. They include, for example, mammalian cystatins A and B, of which B is found in many cell types and A is restricted to the epithelium and leukocytes. The function of cystatin A is probably to inhibit the cysteine proteinases of invading parasitic Protozoa [12]. The type-2 cystatins, which include the chicken egg cystatin, consist of about 115 amino acids (13 kDa) with two disulphide bridges close to the C-terminus. The cystatin family includes not only various intra- and extracellular cystatins from mammals and birds but also an inhibitor in the poison of the puff adder, *Bitis arientans* [12, 224]. The cystatins of type 3, the kininogens, have the most complicated structure of all. Their polypeptide chain of about 355 amino acids makes up three domains, similar in structure to the type-2 cystatins but with different inhibitory activities and specificities. It appears that two gene duplications occurred during the evolution of the kininogens. The C-terminal region of the kininogen chain contains the sequence of the blood-pressure reducing peptide hormone bradykinin, which is released through the activity of the serine proteinase kallikrein (see Fig. 8.5, p. 302). The kininogens also contain additional disulphide bridges and are glycosylated [12].

Cystatins inactivate the plant cysteine proteinase papain and the lysosomal cathepsins B, H and L; calpain II, in contrast, is only inhibited by particular kininogens. The mechanism by which the cystatins inhibit enzymes is not yet completely understood [17]. A family tree can be constructed using the known sequences of cystatins from mammals and birds. Assuming that the rate of evolution of the cystatins has been constant, then the common ancestral form of the present cystatins must have existed before the eukaryotes. In fact, a cysteine-proteinase inhibitor with clear homology to the animal cystatins has been found in rice plants [1, 52]. A proteinase inhibitor specific for cathepsin L has been isolated from pig leukocytes; the sequence of 96 amino acids shows no significant homology with cystatin and apparently belongs to a new super-family [225]. In the fly *Sarcophaga peregrina*, there is a cysteine-proteinase inhibitor (sarcostatin A), of about

10 kDa, which is produced in the fat bodies and released in increasing amounts into the haemolymph during pupae development. Its function would appear to be to protect tissues from attack by proteinases, released during metamorphosis from degraded larval cells [263].

3.4.3 α_2-Macroglobulins

α_2-Macroglobulins are large glycoproteins present in the blood of vertebrates and invertebrates and in the egg albumin of birds and reptiles. The α_2M of mammalian plasma consists of four subunits connected in pairs by disulphide bridges to form dimers. The human subunits are made up of 1451 amino acids which agree at 73 % of positions with those of the rat. This super-family also includes the human pregnancy zone protein, the α_2-macroglobulin and the α_1-inhibitor 3 of the rat. Most members of the family are tetramers, but some are dimers (frog α_2M, human pregnancy zone protein) or monomers (rat α_1-inhibitor 3). Ovostatin, a proteinase inhibitor from chicken egg albumin, is similar in sequence to the blood plasma α_2M at least at the N-terminus. There is significant agreement between α_2M and partial sequences of the complement components C3 and C4 of the mouse; the latter, however, contain regions that are not present in α_2M. The evolution of the complement components possibly included a gene fusion (exon shuffling). The complement components, like α_2M, contain an internal thioester bond that is formed by reaction of a cysteine thiol group with the γ-carboxyl group of a neighbouring glutamate [256, 357].

The **inhibitory** action of α_2M is basically different from that of typical proteinase inhibitors; α_2M inhibits proteinases of all classes and the formation of the α_2M/enzyme is irreversible. Only proteolytic activity against macromolecular substrates is inhibited; the proteinases retain their activity against low molecular weight substrates, even when complexed with α_2M. The explanation for this behaviour lies in the mechanism of inhibition. According to the trap hypothesis, the proteolytic cleavage of at least one subunit brings about a conformational change in the α_2M which allows it to envelope the proteinase and prevent interaction with macromolecular substrates. The establishment of covalent bonds binds the proteinase irreversibly to the α_2M; the complex is then rapidly eliminated by uptake into macrophages or other cells [181]. The active centre of the proteinase is involved in the initial cleavage of the α_2M but not in complex formation, and therefore retains it activity against low molecular weight substrates. The α_2M chain is cleaved in a region in the last third of the molecule before the C-terminus; this contains peptide bonds that correspond to the different specificities of the individual proteinases [257]. It is characteristic of α_2M that it only inhibits that activity of the bound proteinase that is directed against protein substrates protecting it from other macromolecular proteinase inhibitors; furthermore, the activity of α_2M is itself inhibited by methylamine. Based on this property, α_2M can be detected in the blood plasma of all classes of vertebrates from the agnathans *Petromyzon* and *Myxine* to the mammals [235]. The α_2M forms of the chicken and the frog *Rana pipiens* are apparently very similar to those of the mammals in structure and mode of action [70]. In contrast, the fish plasma protein (αMh) that is homologous to α_2M is only half the size of the mammalian protein and consists of two heterodimers with 105- and 90-kDa subunits (I and II) linked by a disulphide bridge. The evolution of αMh to α_2M either involved the fusion of two genes (I and II) or, if I and II are the products of one gene and arose by post-translational modification, the cleavage site has been lost in the evolution of α_2M [259]. In the pig, dog, rat and rabbit there are two structurally different α_2Ms whose relative proportions change during development, pregnancy and acute inflammation, and which apparently have different biological functions.

α_2M-like proteinase inhibitors are also detectable in the xiphosuran *Limulus polyphemus* and in several crustaceans. Sequence comparisons between the thioester-containing regions or comparisons of the peptide cleavage patterns reveal similarities between the inhibitors of *L. polyphemus, Homarus americanus* and *Pacifastacus leniusculus* and the α_2M of mammalian plasma. Like the mammalian α_2M, the inhibitors of the above-mentioned arthropods are inactivated by methylamine; an exception to this is found in *Cancer borealis* [6, 65]. The α_2M-like protein from *Astacus* haemolymph is different to the proteinase inhibitor found in the haemocytes and cuticle; the latter inhibitor is highly specific for the proteases of the fungus *Aphanomyces astaci*, the causative organism of the crayfish pest.

References

1. Abe K. et al.: Molecular cloning of a cysteine proteinase inhibitor of rice (oryzacystatin). Homology with animal cystatins and transient expression in the ripening process of rice seeds. J. biol. Chem. 262: 16793–97 (1987)
2. Adam S. A. et al.: Identification of specific binding proteins for a nuclear location sequence. Nature 337: 276–279 (1989)
3. Aitken A.: Identification of protein consensus sequences. VHC Pubs. (Horwood), New York 1990
4. Alexander M. E. and Dresden M. H.: Collagenolytic enzymes from the starfish, Pyenopodia helianthoides. Comp. Biochem. Physiol. Pt. B 67: 505–509 (1980)
5. Argos P. and Rao J. K. M.: Relationships between exons and the predicted structure of membrane-bound proteins. Biochim. biophys. Acta 827: 283–297 (1985)
6. Armstrong P. B. et al.: Structure of α2-macroglobulin from the arthropod Limus polyphemus. J. Biol. Chem. 266: 2526–30 (1991)
7. Arribas. C., Sampedro J. and Izquierdo M.: The ubiquitin genes in Drosophila melanogaster: transcription and polymorphism. Biochim. biophys. Acta 868: 119–127 (1986)
8. Baba T. et al.:Activation and maturation mechanisms of boar acrosin zymogen based on the deduced primary structure. J. Biol. Chem. 264: 11920–27 (1989)
9. Baici A. and Seemüller U.: Kinetics of the inhibition of human leucocyte elastase by eglin from the leech Hirudo medicinalis. Biochem. J. 218: 829–833 (1984)
10. Bao J. J. et al.: Molecular evolution of serpins: Homologous structure of the human α1-antichymotrypsin and α1-antitrypsin genes. Biochemistry 26: 7755–59 (1987). Correction: Biochemistry 27: 8508 (1988)
11. Barrett A. J. and McDonald J. K.: Mammalian proteases, Vol. I: Endopeptidases. Acad Press, New York 1980
12. Barrett A. J.: The cystatins: a new class of peptidase inhibitors. Trends biochem. Sci. 12: 193–196
13. Baudys M. and Kostka V.: Covalent structure of chicken pepsinogen. Eur. J. Biochem. 136: 89–99 (1983)
14. Bauw G. et al.: Protein-electroblotting and -microsequencing strategies in generating protein data bases from two-dimensional gels. Proc. Nat. Acad. Sci. USA 86: 7701–05 (1989)
15. Berger F. G. and Baumann H.: An evolutionary switch in tissue-specific gene expression. Abundant expression of α1-antitrypsin in the kidney of a wild mouse species. J. biol. Chem. 260: 1160–65 (1985)
16. Blankenship D. T. et al.: Amino acid sequence of ghilanten: anticoagulant-antimetastatic principle of the South American leech, Haementeria ghilianii. Biochem. biophys. Res. Commun. 166: 1384–89 (1990)
17. Bode W. et al.: The 2.0 Å X-ray structure of chicken egg white cystatin and its possible mode of interaction with cysteine proteinases. Embo J. 7: 2593–99 (1988)
18. Bond J. S. and Butler P. E.: Intracellular proteases. Annual Rev. Biochem. 56: 333–364 (1987)
19. Bork P.: Recognition of functional regions in primary structures using a set of property patterns. FEBS Letters 257: 191–195 (1989)
20. Borst P.: How proteins get into microbodies (peroxisomes, glyoxysomes, glycosomes). Biochim. biophys. Acta 866: 179–203 (1986)
21. Bradshaw R. A.: Protein translocations and turnover in eukaryotic cells. Trends biochem. Sci. 14: 276–279 (1989)
22. Brenner S.: The molecular evolution of genes and proteins: a tale of two serines. Nature 334: 528–530 (1988)
23. Brown C. M. et al.: Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. Nucleic Acids Res. 18: 6339–45 (1990)
24. Candelas G. et al.: Translational pauses during a spider fibroin synthesis. Biochem. biophys. Res. Commun. 116: 1033–38 (1983)
25. Carling D. and Hardie D. G.: The substrate and sequence specificity of the AMP-activated protein kinase. Phosphorylation of glycogen synthase and phosphorylase kinase. Biochim. biophys. Acta 1012: 81–86 (1989)
26. Carrell R. and Travis J.: α1-Antitrypsin and the serpins: variation and countervariation. Trends biochem. Sci. 10: 20–24 (1985)
27. Cavener D. E. and Ray S. C.: Eukaryotic start and stop translation sites. Nucleic Acids Res. 19: 3185–92 (1991)
28. Chang P. K. and Dignam J. D.: Pimary structure of alanyl-transfer RNA synthetase and the regulation of its messenger RNA levels in Bombyx mori. J. Biol. Chem. 265: 20898–906 (1990)
29. Chao S. et al.: Molecular cloning and primary stucture of rat α1-antitrypsin. Biochemistry 29: 323–329 (1990)
30. Chappell L. C. and Dresden M. H.: Purification of cysteine proteinases from adult Schistosoma mansoni. Arch. Biochem. Biophys. 256: 560–568 (1987)
31. Charbonneau H. et al.: Human placenta protein-tyrosine-phosphatase: Amino acid sequence and relationship to a family of receptor-like proteins. Proc. Nat. Acad. Sci. USA 86: 5252–56 (1989)
32. Cheeseman M. T. and Gooding R. H.: Proteolytic enzymes from tsetse flies, Glossina morsitans and Glossina palpalis (Diptera: Glossinidae). Insect Biochem. 15: 677–680 (1985)
33. Chothia C. and Finkelstein A. V.: The classification and origins of protein folding patterns. Annual Rev. Biochem. 59: 1007–1039 (1990)
34. Clark V. M. and Curthoys P.: Cause of subunit heterogeneity in purified rat renal phosphate-dependent glutaminase. J. biol. Chem. 254: 4939–41 (1979)
35. Cochrane B. J. and Richmond R. C.: Studies of esterase-6 in Drosophila melanogaster. 1. Genetics of a posttranslational modification. Biochem. Genetics 17: 167–183 (1979)
36. Cohen P.: The structure and regulation of protein phosphatases. Annual Rev. Biochem. 58: 453–508 (1989)
37. Cohen P. T. W. et al.: Protein serine threonine phosphatases – An expanding family. FEBS Letters 268: 355–359 (1990)
38. Colella R. et al.: Chicken white cystatin. Molecular cloning, nucleotide sequence, and tissue distribution. J. Biol. Chem. 264: 17164–69 (1989)
39. Collier I. E. et al.: The structure of the human skin fibroblast collagenase gene. J. Biol. Chem. 263: 10711–13 (1988)

40. Combest W. L. and Gilbert L. I.: Particulate associated cAMP-dependent protein kinase activity in the brain of the tobacco hornworm, Manduca sexta. Insect Biochem. 19: 663–672 (1989)
41. Conlon J. M. and Thim L.: A peptide from the eel pancreas with structural similarity to human pancreatic secretory trypsin inhibitor. Eur. J. Biochem. 174: 149–153 (1988)
42. Craik C. S., Rutter W. J. and Fletterick R.: Splice junctions: association with variation in protein structure. Science 220: 1125–29 (1983)
43. Creighton T. E.: Protein folding. Biochem. J. 270: 1–16 (1990)
44. Dahlmann B. et al.: The multicatalytic proteinase (prosome) is ubiquitous from eukaryotes to archaebacteria. FEBS Letters 251: 125–131 (1989)
45. Dahms, N. M., Lobel P. and Kornfeld S.: Mannose 6-phosphate receptors and lysosomal enzyme targeting (Minireview). J. Biol. Chem. 264: 12115–18 (1989)
46. van Damme H. T. F. et al.: Elongation factor 1-beta of Artemia: Localization of functional sites and homology to elongation factor 1-delta. Biochim. biophys. Acta 1050: 241–247 (1990)
47. Dang C. V. and Dang C. V.: Multienzyme complex of aminoacyl-tRNA synthetases: an essence of being eukaryotic (Review). Biochem. J. 239: 249–255 (1986)
48. Das S. et al.: A cylic nucleotide-independent protein kinase in Leishmania donovani. Biochem. J. 240: 641–649 (1986)
49. Davis, A. H., Nanduri J. and Watson D. C.: Cloning and gene expression of Schistosoma mansoni protease. J. biol. Chem. 262: 12851–55 (1987)
50. Davis C. A. et al.: A gene family in Drosophila melanogaster coding for trypsin-like enzymes. Nucleic Acids Res. 13: 6605–19 (1985)
51. Dayhof M. O. (ed.): Atlas of protein sequence and structure, Vol. 5 and Suppl. 1–3. Nat. Biomed. Res. Foundation, Washington 1972–78
52. Delbridge M. L. and Kelly L. E.: Sequence analysis and chromosomal localization of a gene encoding a cystatin-like protein from Drosophila melanogaster. FEBS Letters 274: 141–145 (1990)
53. Dendinger J. E. and O'Connor K. L.: Purification and characterization of a trypsin-like enzyme from the midgut gland of the Atlantic blue crab, Callinectes sapidus. Comp. Biochem. Physiol. Pt. B 95: 525–530 (1990)
54. Deng L. R. et al.: Isolation and properties of two allelic chymotrypsin inhibitors from the hemolymph of the silkworm, Bombyx mori. Insect Biochem. 20: 531–536 (1990)
55. Diarra-Mehrpour M. et al.: Human plasma inter-α-trypsin inhibitor is encoded by four genes on three chromosomes. Eur. J. Biochem. 179: 147–154 (1989)
56. Djé M. K. et al.: Three genes under different developmental control encode elongation factor 1-α in Xenopus laevis. Nucleic Acids Res. 18: 3489–93 (1990)
57. Dombradi V. et al.: Cloning and chromosomal localization of Drosophila cDNA encoding the catalytic subunit of protein phosphatase 1α – High conservation between mammalian and insect sequences. Eur. J. Biochem. 183: 603–610 (1989)
58. Donovan M. A. and Laue T. M.: A novel trypsin inhibitor from the hemolymph of the horseshoe crab Limulus polyphemus. J. Biol. Chem. 266: 2121–25 (1991)
59. Driscoll J. and Goldberg A. L.: The proteasome (multicatalytic protease) is a component of the 1500-kDa proteolytic complex which degrades ubiquitin-conjugated proteins. J. Biol. Chem. 265: 4789–92 (1990)
60. Dufton M. J.: Proteinase inhibitors and dendrotoxins. Sequence classification, structural prediction and structure/activity. Eur. J. Biochem. 153: 647–654 (1987)
61. Dunbar B. S.: Two-dimensional electrophoresis and immunological techniques. Plenum, New York 1987
62. Dunwiddie C. et al.: Antistasin, a leech-derived inhibitor of factor Xa. Kinetic analysis of enzyme inhibition and identification of the reactive site. J. Biol. Chem. 264: 16695–99 (1989)
63. Edelman A. M., Blumenthal D. K. and Krebs E. G.: Protein serine/threonine kinases. Annual Rev. Biochem. 56: 567–613 (1987)
64. Ehlers M. R. W. et al.: Molecular cloning of human testicular angiotensin-converting enzyme: The testis isozyme is identical to the C-terminal half of endothelial angiotensin-converting enzyme. Proc. Nat. Acad. Sci. USA 86: 7741–45 (1989)
65. Enghild J. J. et al.: Alpha-macroglobulin from Limulus polyphemus exhibits proteinase inhibitory activity and participates in a hemolytic system. Biochemistry 29: 10070–80 (1990)
66. Estell D. A. and Laskowski M. jr.: Dermasterias imbricata Trypsin I: an enzyme which rapidly hydrolyzes the reactive site peptide bonds of protein trypsin inhibitors. Biochemistry 19: 124–131 (1980)
67. Falkenburg P. E. and Kloetzel P. M.: Identification and characterization of three different subpopulations of the Drosophila multicatalytic proteinase (proteasome). J. Biol. Chem. 264: 6660–66 (1989)
68. Fasman G. D. (ed.): Prediction of protein structure and the principles of protein conformation. Plenum, New York 1989
69. Faust P. L., Kornfeld S. and Chirgwin J. M.: Cloning and sequence analysis of cDNA for human cathepsin D. Proc. Nat. Acad. Sci. USA 82: 4910–14 (1985)
70. Feldman S. and Pizzo S. V.: Purification and characterization of frog α-macroglobulin: Receptor recognition of an amphibian glycoprotein. Biochemistry 24: 2569–75 (1985)
71. Fini M. E. et al.: A gene for rabbit synovial cell collagenase: member of a family of metalloproteinases that degrade the connective tissue matrix. Biochemistry 26: 6156–65 (1987)
72. Fischer E. H., Charbonneau H. and Tonks N. K.: Protein tyrosine phosphatases. A diverse family of intracellular and transmembrane enzymes. Science 253: 401–406 (1991)
73. Folco E. J. et al.: Multicatalytic proteinase in fish muscle. Arch. Biochem. Biophys. 267: 599–605 (1988)
74. Foster J. L., Higgins G. C. and Jackson F. R.: Cloning, sequence, and expression of the Drosophila cAMP-dependent protein kinase catalytic subunit gene. J. biol. Chem. 263: 1676–81 (1988)
75. Freedman R. B. and Hawkins H. C. (eds.): The enzymology of posttranslational modification of proteins. Acad. Press, New York 1985
76. Gabius H. J. et al.. Evolutionary aspects of accuracy of phenylalananyl-tRNA synthetase. Accuracy of fun-

gal and animal mitochondrial enzymes and their relationship to their cytoplasmic counterparts and a prokaryotic enzyme. Biochemistry 22: 5306–15 (1983)
77. Gardell S. J. et al.: A novel rat carboxypeptidase, CPA2: characterization, molecular cloning, and evolutionary implications on substrate specificity in the carboxypeptidase gene family. J. Biol. Chem. 263: 17828–36 (1988)
78. Gebhard W. et al.: Two out of the three kinds of subunits of inter-α-trypsin inhibitor are structurally related. Eur. J. Biochem. 181: 571–576 (1989)
79. Ghersa P. et al.: Initiation of translation at a UAG stop codon in the aldolase gene of Plasmodium falciparum. Embo J. 9: 1645–49 (1990)
80. Gildberg A.: Aspartic proteinases in fishes and aquatic invertebrates. Comp. Biochem. Physiol. Pt. B 91: 425–35 (1988)
81. Gockel S. F. and Lebherz H. G.: „Conformational" isoenzymes in ascarid enolase. J. biol. Chem. 256: 3877–83 (1981)
82. Godar D. E. et al.: Structural organization of the multienzyme complex of mammalian aminoacyl-tRNA synthetases. Biochemistry 27: 6921–28 (1988)
83. Goldberg M. E.: The second translation of the genetic message: protein folding and assembly. Trends biochem. Sci. 10: 388–391 (1985)
84. Gordon E. D. et al.: Eukaryotic initiation factor 4D, the hypusine-containing protein, is conserved among eukaryotes. J. biol. Chem. 262: 16585–89 (1987)
85. Gordon J. I. et al.: Protein N-myristoylation (Minireview). J. Biol. Chem. 266: 8647–50 (1991)
86. Gould S. J. et al.: Peroxisomal protein import is conserved between yeast, plants, insects and mammals. Embo J. 9: 85–90 (1990)
87. Grant G. A., Sacchettini J. C. and Welgus H. G.: A collagenolytic serine protease with trypsin-like specificity from the fiddler crab Uca pugilator. Biochemistry 22: 354–358 (1983)
88. Grinblat Y., Brown N. H. and Kafatos F. C.: Isolation and characterization of the Drosophila translational elongation factor 2 gene. Nucleic Acids Res. 17: 7303–14 (1989)
89. Gross R. E. et al.: Cloning, characterization, and expression of the gene for the catalytic subunit of cAMP-dependent protein kinase in Caenorhabditis elegans. Identification of highly conserved and unique isoforms generated by alternative splicing. J. Biol. Chem. 265: 6896–6907 (1990)
90. Grossman A.: Information transfer in biological systems. Targeting of proteins to specific organelles or to the extracellular environment (secretion). Comp. Biochem. Physiol. Pt. B 91: 389–424 (1988)
91. Guerard F. and le Gal Y.: Characterization of a chymosin-like pepsin from the dogfish Scyliorhinus canicula. Comp. Biochem. Physiol. Pt. B 88: 823–827 (1987)
92. Gundersen R. E. and Nelson D. L.: A novel Ca-dependent protein kinase from Paramecium tetraurelia. J. biol. Chem. 262: 4602–09 (1987)
93. Haass C. et al.: The Drosophila PROS-28.1 gene is a member of the proteasome gene family. Gene 90: 235–241 (1990)
94. Hamed M. B. B. and Attias J.: Isolation and partial characterization of two alkaline proteases of the greater wax moth Galleria melonella (L.). Insect Biochem. 17: 653–658 (1987)
95. Hameed, K. S. and Haard N. F.: Isolation and characterization of cathepsin D from Atlantic short finned squid Illex illecerebrosus. Comp. Biochem. Physiol. Pt. B 82: 241–246 (1985)
96. Hanks S. K., Quinn A. M. and Hunter T.: The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. Science 241: 42–52 (1988)
97. Hansen L. J., Huang W. I. and Jagus R.: Inhibitor of translational initiation in sea urchin eggs prevents mRNA utilization. J. biol. Chem. 262: 6114–20 (1987)
98. Hardie D. G. and Coggins J. R. (eds.): Multidomain proteins. Structure and funktion. Elsevier, Amsterdam 1986
99. Hartl F. U. and Neupert W.: Protein sorting to mitochondria: Evolutionary conservations of folding and assembly. Science 247: 930–938 (1990)
100. Hasty K. A. et al.: Human neutrophil collagenase. A distinct gene product with homology to other matrix metalloproteinases. J. Biol. Chem. 265: 11421–24 (1990)
101. Hayono T. et al.: Primary structure of human pepsinogen C gene. J. biol. Chem. 263: 1382–85 (1988)
102. von Heijne G.: Signal sequences. The limits of variation. J. mol. Biol. 184: 99–105 (1985)
103. Helaakoski T. et al.: Molecular cloning of the α-subunit of human prolyl 4-hydroxylase: The complete cDNA-derived amino acid sequence and evidence for alternative splicing. Proc. Nat. Acad. Sci. USA 86: 4392–96 (1989)
104. van Hemert F. J. et al.: The primary structure of elongation factor EF-1α from the brine shrimp Artemia. Embo J. 3: 1109–13 (1984)
105. Hershey J. W. B.: Protein phosphorylation controls translation rates. J. Biol. Chem. 264: 20823–26 (1989)
106. Hill R. E. and Hastie N. D.: Accelerated evolution in the reactive center regions of serine protease inhibitors. Nature 326: 96–99 (1987)
107. Hille A. et al.: Occurrence of tyrosine sulfate in proteins. A balance sheet. 1. Secretory and lysosomal proteins. Eur. J. Biochem. 188: 577–586 (1990)
108. Holm I. et al.: Evolution of aspartyl proteases by gene duplication: the mouse renin gene is organized in two homologous clusters of four exons. Embo J. 3: 557–562 (1984)
109. Holmes W. E. et al.: Primary structure of human α2-antiplasmin, a serine protease inhibitor (serpin). J. biol. Chem. 262: 1659–64 (1987)
110. Holmquist R.: Evaluation of compositional nonrandomness in proteins. J. mol. Evol. 11: 349–360 (1978)
111. Horii A. et al.: On the cDNAs for two types of rat pancreatic secretory trypsin inhibitor. Biochem. biophys. Res. Commun. 162: 151–159 (1989)
112. Houseman J. G. and Downe A. E. R.: Cathepsin D-like activity in the posterior midgut of hemipteran insects. Comp. Biochem. Physiol. Pt. B 75: 509–512 (1983)
113. Houseman J. G., Campbell F. C. and Morrison P. E.: A preliminary characterization of digestive proteases in the posterior midgut of the stable fly Stomoxys calcitrans (L.) (Diptera: Muscidae). Insect Biochem. 17: 213–218 (1987)
114. Hu E. and Rubin C. S.: Casein kinase II from Caenorhabditis elegans. Properties and developmental regulation of the enzyme; cloning and sequence ana-

lyses of cDNA and the gene for the catalytic subunit. J. Biol. Chem. 265: 5072–80 (1990)
115. Hunter T.: A thousand and one protein kinases. Cell 50: 823–829 (1987)
116. Isackson P. J., Ullrich A. and Bradshaw R. A.: Mouse 7 S nerve growth factor: Complete sequence of a cDNA coding for the α-subunit precursor and its relationship to serine proteases. Biochemistry 23: 5997–6002 (1984)
117. Ishihara T. et al.: Primary structure and transcriptional regulation of rat pepsinogen C gene. J. Biol. Chem. 264: 10193–99 (1989)
118. Ito A. et al.: The complete primary structure of calcineurin A, a calmodulin binding protein homologous with protein phosphatases 1 and 2A. Biochem. biophys. Res. Commun. 163: 1492–97 (1989)
119. Jaenicke R.: Protein folding. Local structures, domains, subunits and assemblies. Biochemistry 30: 3147–61 (1991)
120. James M. N. G., Delbaere L. T. J. and Brayer G. D.: Amino acid sequence alignment of bacterial and mammalian pancreatic serine proteases based on topological equivalences. Can. J. Biochem. 56: 396–402 (1978)
121. James M. N. G., Sielecki A. R.: Molecular structure of an aspartic proteinase zymogen, at 1.8 Å resolution. Nature 319: 33–38 (1986)
122. Jany K. D. and Haug H.: Amino acid sequence of the chymotryptic protease II from the larvae of the hornet, Vespa crabo. FEBS Letters 158: 98–102 (1983)
123. Jennings M. L.: Topography of membrane proteins. Annual Rev. Biochem. 58: 999–1027 (1989)
124. Jentsch, S., Seufert W. and Hauser H. P.: Genetic analysis of the ubiquitin system (Review). Biochim. biophys. Acta 1089: 127–139 (1991)
125. Jeppson J. O. and Laurell C. B.: The amino acid substitutions of human α1-antitrypsin M3, X and Z. FEBS Letters 231: 327–330 (1988)
126. Joernvall H., Hooeg J.-O. and Gustavsson A.-M.: Methods in protein sequence analysis. Birkhaeuser, Basel 1991
127. Jonnalagadda S. et al.: Multiple (α-NH-ubiquitin) protein endoproteases in cells. J. Biol. Chem. 264: 10637–42 (1989)
128. Jordao B. P. and Terra W. R.: Distribution, properties, and functions of midgut carboxypeptidases and dipeptidases from Musca domestica larvae. Arch. Insect Biochem. Physiol. 11: 231–244 (1989)
129. Kageyama T. and Takahashi K.: The complete amino acid sequence of monkey pepsinogen A. J. biol. Chem. 261: 4395–4405 (1986)
130. Kageyama T. and Takahashi K.: The complete amino acid sequence of monkey progastricsin. J. biol. Chem. 261: 4406–19 (1986)
131. Kalderon D. and Rubin G. M.: cGMP-dependent protein kinase genes in Drosophila. J. Biol. Chem. 264: 10738–48 (1989)
132. Kang P. J. et al.: Requirement for hsp70 in the mitochondrial matrix for translocation and folding of precursor proteins. Nature 348: 137–143 (1990)
133. Kanost M. R.: Isolation and characterization of four serine proteinase inhibitors (serpins) from hemolymph of Manduca sexta. Insect Biochem. 20: 141–147 (1990)
134. Kato I. and Tominaga N.: Trypsin-subtilisin inhibitor from Red Sea turtle eggwhite consists of two tandem domains – one Kunitz – one of a new family. Fed. Proc. Abstract of 67th Annual Meeting Nr. 3168 (1983)
135. Kato I. et al.: Chicken ovomucoid: Determination of its amino acid sequence, determination of the trypsin reactive site, and preparation of all three of its domains. Biochemistry 26: 193–201 (1987)
136. Katunuma N., Umezawa H. and Holzer H. (eds.): Proteinase inhibitors. Medical and biological aspects. Springer, Berlin 1983
137. Kavanagh E. J. and Tillinghast E. K.: The alkaline proteases of Argiope – II. Fractionation of protease activity and isolation of a silk fibroin digesting protease. Comp. Biochem. Physiol. Pt. B 74: 365–372 (1983)
138. Kawamura M., Wadano A. and Miura K.: Purification and characterization of insect cathepsin D. Insect Biochem. 17: 77–83 (1987)
139. Kenny A. J. and Ingram J.: Is there a tripeptidyl peptidase in the renal brushborder membrane? Biochem. J. 255: 373–376 (1988)
140. Kikkawa U., Kishimoto A. and Nishizuka Y.: The protein kinase C family: Heterogeneity and its implications. Annual Rev. Biochem. 58: 31–44 (1989)
141. Kikuchi Y. and Tamiya N.: Chemical taxonomy of the hinge-ligament proteins of bivalves according to their amino acid compositions. Biochem. J. 242: 505–510 (1987)
142. Kim P. S. and Baldwin R. L.: Intermediates in the folding reactions of small proteins. Annual Rev. Biochem. 59: 631–660 (1990)
143. Kirchhoff L. V. et al.: Ubiquitin genes in trypanosomatidae. J. Biol. Chem. 263: 12698–704 (1988)
144. Klier H. J., Vonfigura K. and Pohlmann R.: Isolation and analysis of the human 46-kDa mannose 6-phosphate receptor gene. Eur. J. Biochem. 197: 23–28 (1991)
145. Klimova O. A. et al.: The isolation and properties of collagenolytic proteases from crab hepatopancreas. Biochem. biophys. Res. Commun. 166: 1411–20 (1990)
146. Koster A. et al.: Molecular cloning of the mouse 46-kDa mannose 6-phosphate receptor (MPR-46). Biol. Chem. Hoppe-Seyler 372: 297–300 (1991)
147. Kreil G.: Transfer of proteins across membranes. Annual Rev. Biochem. 50: 317–348 (1981)
148. Kruh G. D. et al.: The complete coding sequence of arg defines thr. Abelson subfamily of cytoplasmic tyrosine kinases. Proc. Nat. Acad. Sci. USA 87: 5802–06 (1990)
149. Lane C. D. et al.: The sequestration, processing and retention of honey-bee promelittin made in amphibian oocytes. Eur. J. Biochem. 113: 273–281 (1981)
150. Laskowski M. et al.: Amino acid sequences of ovomucoid third domain from 25 additional species of birds. J. Protein Chem. 9: 715–726 (1990)
151. Laycock M. V. et al.: Purification and characterization of a digestive cysteine proteinase from the American lobster (Homarus americanus). Biochem. J. 263: 439–444 (1989)
152. Lazure C. et al.: The complete amino acid sequence of rat submaxillary gland tonin does contain the aspartic acid at the active site: confirmation by protein sequence analyses. Biochem. Cell Biol. 65: 321–337 (1987)
153. Lecroisey A. et al.: Complete amino acid sequence of the collagenase from the insect Hypoderma lineatum. J. biol. Chem. 262: 7546–51 (1987)

154. Lee H., Simon J. A. and Lis J. T.: Structure and expression of ubiquitin genes of Drosophila melanogaster. Mol. cell. Biol. 8: 4727–35 (1988)
155. Lee L. W. et al.: Relationships among the subunits of the high molecular weight proteinase, macropain (proteasome). Biochim. biophys. Acta 1037: 178–185 (1990)
156. Lehuerou I. et al.: Isolation and nucleotide sequence of a cDNA clone for bovine pancreatic anionic trypsinogen. Structural identity with the trypsin family. Eur. J. Biochem. 193: 767–773 (1990)
157. Lepage T. and Gache C.: Purification and characterization of the sea urchin embryo hatching enzyme. J. Biol. Chem. 264: 4787–93 (1989)
158. Light A. and Janska H.: Enterokinase (enteropeptidase): comparative aspects. Trends biochem. Sci. 14: 110–112 (1989)
159. Lin Y. M.: Characterization and peptidase specificity of lugworm (Arenicola cristata) protease C. Comp. Biochem. Physiol. Pt. B 95: 745–753 (1990)
160. Litchfield D. W. et al.: Subunit structure of casein kinase II from bovine testis. Demonstration that the α and α' subunits are distinct polypeptides. J. Biol. Chem. 265: 7638–44 (1990)
161. Lu X. et al.: Cloning, structure, and expression of the gene for a novel regulatory subunit of cDNA-dependent protein kinase in Caenorhabditis elegans. J. Biol. Chem. 265: 3293–3303 (1990)
162. Luaces A. L. and Barrett A. J.: Affinity purification and biochemical characterization of histolysin, the major cysteine proteinase of Entamoeba histolytica. Biochem. J. 250: 903–909 (1988)
163. Ma Z. M., Grubbs J. H. and Sly W. S.: Cloning, sequencing, and functional characterization of the murine 46-kDa mannose 6-phosphate receptor. J. Biol. Chem. 266: 10589–95 (1991)
164. Macdonald R. J., Stary S. J. and Swift G. H.: Two similar but nonallelic pancreatic trypsinogens. Nucleotide sequence of the cloned cDNAs. J. biol. Chem 257: 9724–32 (1982)
165. Mahlke K. et al.: Sorting pathways for mitochondrial inner membrane proteins. Eur. J. Biochem. 192: 551–555 (1990)
166. Mallya S. K. et al.: Characterization of 58-kilodalton human neutrophil collagenase. Comparison with human fibroblast collagenase. Biochemistry 29: 10628–34 (1990)
167. Marecum J. A.: A trypsin inhibitor from the coelomic fluid of the sea star Asterias forbesi. Biol. Bull. 172: 357–361 (1987)
168. van Marrewijk W. A. and Ravesetin H. J. L.: Amino acid metabolism of Astacus leptodactylus Esch. – I. Composition of the free and protein-bound amino acids in different organs of the crayfish. Comp. Biochem. Physiol. Pt. B 47: 531–542 (1974)
169. Martinage A. et al.: Primary structure of histone H2B from gonads of the starfish Asterias rubens. Identification of an N-dimethylproline residue at the aminoterminal. Eur. J. Biochem. 147: 351–359 (1985)
170. Martzen M. R. et al.: Primary structure of the major pepsin inhibitor from the intestinal parasitic nematode Ascaris suum. Biochemistry 29: 7366–72 (1990)
171. Marumo K. and Waite H.: Prolyl 4-hydroxylase in the foot of the marine mussel Mytilus edulis L.: Purification and characterization. J. exp. Zool. 244: 365–374 (1987)
172. Mateu M. G., Vicente O. and Sierra J. M.: Protein synthesis in Drosophila melanogaster embryos – Purification and characterization of polypeptide chain-initiation factor 2. Eur. J. Biochem. 162: 221–229 (1987)
173. Matthews J. A., Brown J. W. S. and Hall T. C.: Phaseolin mRNA is translated to yield glycosylated polypeptides in Xenopus oocytes. Nature 294: 175–176 (1981)
174. Mayer R. J. and Doherty F.: Intracellular protein catabolism: state of the art. FEBS Letters 198: 181–193 (1986)
175. McCammon J. A. and Harvey S. C. (eds.): Dynamics of proteins and nucleic acids. Cambridge Univ. Press, Cambridge 1987
176. McDonald J. K. and Barrett A. J.: Mammalian proteases, Vol. 2: Exopeptidases. Acad. Press, London 1986
177. McIllhinney R. A. J.: The facts of life: The importance and function of protein acylation. Trends biochem. Sci. 15: 387–391 (1990)
178. Mehta H. B. et al.: Structural studies on the eukaryotic chain initiation factor 2 from rabbit reticulocytes and brine shrimp Artemia embryos. Phosphorylation by the hemecontrolled repressor and casein kinase II. J. biol. Chem. 261: 6705–11 (1986)
179. Meloun B., Chechova D. and Jonakova V.: Homologies in the structures of bull seminal plasma acrosin inhibitors and comparison with other homologous proteinase inhibitors of the Kazal type. Hoppe-Seyler's Z. physiol. Chem. 364: 1665–70 (1983)
180. Miglietta L. A. P. and Nelson D. L.: A novel cGMP-dependent protein kinase from Paramecium. J. Biol. Chem. 263: 16096–105 (1988)
181. Moestrup S. K. and Gliemann J.: Purification of the rat hepatic α2-macroglobulin receptor as an approximately 440-kDa single chain protein. J. Biol. Chem. 264: 15574–77 (1989)
182. Moss D. W.: Isoenzymes. Chapman & Hall, London 1982
183. Muramatsu T. and Morita T.: Anionic trypsin-like enzymes from the crab Eriocheir japonicus De Haan active in more acidic media. Comp. Biochem. Physiol. Pt. B 70: 527–533 (1981)
184. Murdock L. L. et al.: Cysteine digestive proteinases in Coleoptera. Comp. Biochem. Physiol. Pt. B 87: 783–787 (1987)
185. Mykles D. L.: Purification and characterization of a multicatalytic proteinase from crustacean muscle: Comparison of latent and heat-activated forms. Arch. Biochem. Biophys. 274: 216–228 (1989)
186. Nanbu M., Kobayashi K. and Horiuchi S.: Purification and characterization of cathepsin D-like proteinase from the tadpole tail of bullfrog, Rana catesbeiana. Comp. Biochem. Physiol. Pt. B 89: 569–575 (1988)
187. Nelson R. B. and Siman R.: Clipsin, a chymotrypsin-like protease in rat brain which is irreversibly inhibited by α1-antichymotrypsin. J. Biol. Chem. 265: 3836–43 (1990)
188. Nene V. et al.: A single exon codes for the enzyme domain of a protozoan cysteine protease. J. Biol. Chem. 265: 18047–50 (1990)
189. Neurath H.: Evolution of proteolytic enzymes. Science 224: 350–357 (1984)

190. Neves A., Guerreiro P. and Rodrigues-Pousada C.: Striking changes in polyubiquitin genes of Tetrahymena pyriformis. Nucleic Acids Res. 18: 656 (1990)
191. Nigg E. A., Baeuerle P. A. and Luhrmann R.: Nuclear import-export. In search of signals and mechanisms (Meeting Review). Cell 66: 15–22 (1991)
192. Nuske J. H.: Protein methylase II in five taxa from three phyla. Comp. Biochem. Physiol. Pt. B 86: 37–47 (1987)
193. O'Donoghue G. V. and Johnson D. B.: A soluble aminopeptidase of Holothuria forskali intestinal mucosa: purification and active centre studies. Comp. Biochem. Physiol. Pt. B 85: 397–405 (1986)
194. Ogata S., Misumi Y. and Ikehara Y.: Primary structure of rat liver dipeptidyl peptidase IV deduced from its cDNA and identification of the NH2-terminal signal sequence as the membrane anchoring domain. J. Biol. Chem. 264: 3596–3601 (1989)
195. Ogita Z. I. and Markert C. L.: Isozymes. Wiley-Liss, New York 1990
196. Okada S. and Aikawa T.: Cathepsin D-like acid proteinase in the mantle of the marine mussel, Mytilus edulis. Comp. Biochem. Physiol. Pt. B 84: 333–341 (1986)
197. Okada Y. and Yokota Y.: Purification and properties of cathepsin B from sea urchin eggs. Comp. Biochem. Physiol. Pt. B 96: 381–386 (1990)
198. Okotore R. O. and Uhlenbruck G.: Proteinase-inhibitors in albumin glands of Achatina fulica. Z. Naturforsch. Sect. C 37: 142–144 (1982)
199. Olsen J. et al.: Complete amino acid sequence of human intestinal aminopeptidase N as deduced from cloned cDNA. FEBS Letters 238: 307–314 (1988)
200. Ono H. and Tuboi S.: Purification and identification of a cytosolic factor required for import of precursors of mitochondrial proteins into mitochondria. Arch. Biochem. Biophys. 280: 299–304 (1990)
201. Orgad S. et al.: The structure of protein phosphatase-2A is as highly conserved as that of protein phosphatase-1. FEBS Letters 275: 44–48 (1990)
202. Orlowski M.: The multicatalytic proteinase complex, a major extralysosomal proteolytic system. Biochemistry 29: 10289–97 (1990)
203. Osnes K. K. and Mohr V.: On the purification and characterization of exopeptidases from Antarctic krill, Euphausia superba. Comp. Biochem. Physiol. Pt. B 83: 445–458 (1986)
204. Ou W. J. et al.: Purification and characterization of a processing protease from rat liver mitochondria. Embo J. 8: 2605–12 (1989)
205. Oxender D. L. (ed.): Protein structure, folding, and design. Alan R. Liss, New York 1987
206. Park M. H.: The essential role of hypusine in eukaryotic translation initiation factor (eIF-4D). Purification of eIF-4D and its precursors and comparison of their activities. J. Biol. Chem. 264: 18531–35 (1989)
207. Pearson J. D. et al.: Amino acid sequence and characterization of a protein inhibitor of protein kinase C. J. Biol. Chem. 265: 4583–91 (1990)
208. Peaucellier G.: Purification and characterization of proteases from the polychaete annelid Sabellaria alveolata (L.). Eur. J. Biochem. 136: 435–445 (1983)
209. Pellegrini A. and von Fellenberg R.: Pre-α2-elastase inhibitor of the horse: a hybrid molecule between α1-proteinase inhibitor and α2-ß1-glycoprotein. Biochim. biophys. Acta 830: 20–24 (1985)
210. Pfanner N. and Neupert W.: The mitochondrial protein import apparatus. Annual Rev. Biochem. 59: 331–353 (1990)
211. Pinter M. and Friedrich P.: The calcium-dependent proteolytic system calpain-calpastatin in Drosophila melanogaster. Biochem. J. 253: 467–473 (1988)
212. Pontremoli S. and Melloni L.: Extralysosomal protein degradation. Annual Rev. Biochem. 55: 455–481 (1986)
213. Potempa J., Shieh B. H. and Travis J.: Alpha-2-antiplasmin: A serpin with two separate but overlapping reactive sites. Science 241: 699–700 (1988)
214. Pratt R. E., Ouelette A. J. and Dzau V. J.: Biosynthesis of renin: multiplicity of active and intermediate forms. Proc. Nat. Acad. Sci. USA 80: 6809–13 (1983)
215. Prehn S. et al.: Structure and biosynthesis of the signal-sequence receptor. Eur. J. Biochem. 188: 439–455 (1990)
216. Pungercar J. et al.: Complete primary structure of lamb preprochymosin deduced from cDNA. Nucleic Acids Res. 18: 4602 (1990)
217. Qiu T., Combest W. L. and Gilbert L. I.: Characterization of a calcium and diacylglycerol-activated and phospholipid-dependent protein kinase in the pupal brain of the tobacco hornworm, Manduca sexta. Insect Biochem. 20: 405–420 (1990)
218. Ragg H. and Preibisch G.: Structure and expression of the gene coding for the human serpin hLS2. J. Biol. Chem. 263: 12129–34 (1988)
219. Ramesh N., Sugumaran M. and Mole J. E.: Purification and characterization of two trypsin inhibitors from the hemolymph of Manduca sexta larvae. J. Biol. Chem. 263: 11523–27 (1988)
220. Rawlings N. D. and Barrett A. J.: Evolution of proteins of the cystatin superfamily. J. mol. Evol. 30: 60–71 (1990)
221. Rechsteiner M. C.: Ubiquitin. Plenum, New York 1988
222. Regan L., Dignam J. D. and Schimmel P.: A bacterial and silkworm aminoacyl-tRNA synthetase have a common epitope which maps to the catalytic domain of each. J. biol. Chem. 261: 5241–44 (1986)
223. Reisinger P. W. M. et al.: The amino-acid sequences of the double-headed proteinase inhibitors from cat, lion and dog submandibular glands. Biol. Chem. Hoppe-Seyler 368: 717–726 (1987)
224. Ritonja A. et al.: Amino acid sequence of a cystatin from venom of the African puff adder (Bitis arietans). Biochemical J. 246: 799–802 (1987)
225. Ritonja A. et al.: Primary structure of a new cysteine proteinase inhibitor from pig leucocytes. FEBS Letters 255: 211–214 (1989)
226. Ritonja T. et al.: Amino acid sequences of the human kidney cathepsins H and L. FEBS Letters 228: 341–345 (1988)
227. Rocamora N. and Agell N.: Methylation of chick UbI and UbII polyubiquitin genes and their differential expression during spermatogenesis. Biochem. J. 267: 821–829 (1990)
228. Roemisch K. et al.: Homology of 54K protein of signal-recognition particle, docking protein and two E. coli proteins with putative FTP-binding domains. Nature 340: 478–482 (1989)
229. Rogers J.: Exon shuffling and intron insertion in serine protease genes. Nature 315: 458–459 (1985)

230. Rogers S., Wells R. and Rechsteiner M.: Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234: 364–368 (1986)
231. Rottmann M. et al.: Specific phosphorylation of proteins in pore complex-laminae from the sponge Geodia cydonium by the homologous aggregation factor and phorbol ester. Role of protein kinase C in the phosphorylation of DNA topoisomerase II. Embo J. 6: 3639–44 (1987)
232. Rubarteli A. et al.: A novel secretory pathway for interleukin-1ß, a protein lacking a signal sequence. Embo J. 9: 1503–10 (1990)
233. Sakal E., Applebaum S. W. and Birk Y.: Purification and characterization of trypsins from the digestive tract of Locusta migratoria. Int. J. Peptide Prot. Res. 34: 498–505 (1989)
234. Sakanari J. A. et al.: Serine proteases from nematode and prozoan parasites. Isolation of sequence homologs using generic molecular probes. Proc. Nat. Acad. Sci. USA 86: 4863–67 (1989)
235. Sallenave J. M. and Bellot R.: Evidence of an α2-macroglobulin-like molecule in plasma of Salamandra salamandra: structural and functional similarity with human α2-macroglobulin. FEBS Letters 219: 37–39 (1987)
236. Sanchez-Chiang L. et al.: Cathepsins D from sea urchin egg Tetrapygus niger – isolation by affinity chromatography and properties. Comp. Biochem. Physiol. 85B: 81–87 (1986)
237. Sasaki T. and Suzuki Y.: Alkaline proteases in digestive juice of the silkworm, Bombyx mori. Biochim. biophys. Acta 703: 1–10 (1982)
238. Sasaki T.: Amino acid sequence of a novel Kunitz-type chymotrypsin inhibitor from hemolymph of silkworm larvae, Bombyx mori. FEBS Letters 168: 227–230 (1984)
239. Satir B. H. et al.: Species distribution of a phosphoprotein (parafusin) involved in exocytosis. Proc. Nat. Acad. Sci. USA 86: 930–932 (1989)
240. Saus J. et al.: The complete primary structure of human matrix metalloproteinase-3. Identification with stromelysin. J. Biol. Chem. 263: 6742–45 (1988)
241. Sawada H. et al.: Trypsin-like enzyme from eggs of the ascidian (protochordate) Halocynthia roretzi. Purification, properties, and physiological role. J. biol. Chem. 260: 15694–98 (1985)
242. Schaeffer E. et al.: Isolation and characterization of two new Drosophila protein kinase C genes, including one specificaly expressed in photoreceptor cells. Cell 57: 403–412 (1989)
243. Scharf M., Engels J. and Tripier D.: Primary structure of new „iso-hirudins". FEBS Letters 255: 105–110 (1989)
244. Schulman H. and Lou L. L.: Multifunctional Ca/calmodulin-dependent protein kinase: domain structure and regulation. Trends biochem. Sci. 14: 62–66 (1989)
245. Schulz G. E. and Schirmer R. H.: Principles of protein structure. Springer, Berlin 1984
246. Shamsuzzaman K. and Haard N. F.: Purification and characterization of a chymosin-like protease from the gastric mucosa of harp seal (Pagophilus groenlandicus). Can. J. Biochem. Cell Biol. 62: 699–708 (1984)
247. Sharma B. R., Martin M. M. and Shafer J. A.: Alkaline proteases from the gut fluids of detritus-feeding larvae of the crane fly, Tipula abdominalis (Say) (Diptera, Tipulidae). Insect Biochem. 14: 37–44 (1984)
248. Sharp P. M. and Li W.: Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. J. mol. Evol. 25: 58–64 (1987)
249. Shen S. S. and Ricke L. A.: Protein kinase C from sea urchin eggs. Comp. Biochem. Physiol. Pt. B 92: 251–154 (1989)
250. Shugerman R. P. et al.: A unique „mini" pepsinogen isolated from bullfrog esophageal glands. J. biol. Chem. 257: 795–798 (1982)
251. Silver P. A.: How proteins enter the nucleus (Review). Cell 64: 489–497 (1991)
252. Soderling T. R.: Protein kinases. Regulation by autoinhibitory domains. J. Biol. Chem. 265: 1823–26 (1990)
253. Somero G. N. and Hand S. C.: Protein assembly and metabolic regulation: Physiological and evolutionary perspectives. Physiol. Zool. 63: 443–471 (1990)
254. Sommer J. et al.: cDNA sequence coding for a rat glia-derived nexin and its homology to members of the serpin superfamily. Biochemistry 26: 6407–10 (1987)
255. Sorimachi H. et al.: Molecular cloning of cDNAs for two subunits of rat multicatalytic proteinase. Existence of N-terminal conserved and C-terminal diverged sequences among subunits. Eur. J. Biochem. 193: 775–781 (1990)
256. Sottrup-Jensen L.: α-Macroglobulins: shape and mechanism of proteinase complex formation (Minireview). J. Biol. Chem. 264: 11539–42 (1989)
257. Sottrup-Jensen L. et al.: The α-macroglobulin bait region. Sequence diversity and localization of cleavage sites for proteinases in five mammalian α-macroglobulins. J. Biol. Chem. 264: 15781–89 (1989)
258. Stadtman E. R.: Covalent modification reactions are marking steps in protein turnover. Biochemistry 29: 6323–31 (1990)
259. Starkey P. M. and Barrett A. J.: Evolution of α2-macroglobulin. The structure of a protein homologous with human α2-macroglobulin from plaice (Platessa platessa L.) plasma. Biochem J. 205: 105–115 (1982)
260. Stoecker W. et al.: Astacus proteinase, a zinc metalloenzyme. Biochemistry 27: 5026–32 (1988)
261. Strandberg L., Lawrence D. and Ny T.: The organization of the human plasminogen-activator-inhibitor gene. Implications on the evolution of the serine-protease inhibitor family. Eur. J. Biochem. 176: 609–616 (1988)
262. Strauss A. W., Boime I. and Kreil G. (eds.): Protein compartmentalization. Springer, Berlin 1986
263. Suzuki T. and Natori S.: Changes in the amount of sarcostatin A, a new cysteine proteinase inhibitor, during the development of adult Sarcophaga peregrina. Insect Biochem. 16: 589–595 (1986)
264. Swinkels B. W., Evers R. and Borst P.: The topogenic signal of the glycosomal (microbody) phosphoglycerate kinase of Crithidia fasciculata resides in a carboxy-terminal extension. Embo J. 7: 1159–65 (1988)
265. Sziegoleit A.: A novel proteinase from human pancreas. Biochem. J. 219: 735–742 (1984)
266. Taggart R. T. et al.: Variable numbers of pepsinogen genes are located in the centromeric region of human chromosome 11 and determine the high-frequency electrophoretic polymorphism. Proc. Nat. Acad. Sci. USA 82: 6240–44 (1985)

267. Taggart R. T. et al.: Human pepsinogen C (progastricsin). Isolation and cDNA sequence clones, localization to chromosome 6, and sequence homology with pepsinogen A. J. Biol. Chem. 264: 375–379 (1989)
268. Takada Y. et al.: Human peroxisomal L-alanine:glyoxylate aminotransferase. Evolutionary loss of a mitochondrial targeting signal by point mutation of the initiation codon. Biochem. J. 268: 517–520 (1990)
269. Takahashi S. Y.: Characterization of the guanosine 3':5'-monophosphate-dependent protein kinase from silkworm eggs and analysis of the endogenous protein substrates. J. comp. Physiol. B 155: 693–701 (1985)
270. Takio K. et al.: Homology of amino acid sequences of rat liver Cathepsins B and H with that of papain. Proc. Nat. Acad. Sci. USA 80: 3666–70 (1983)
271. Tan F. et al.: Molecular cloning and sequencing of the cDNA for human membrane-bound carboxypeptidase M. Comparison with carboxypeptidases A, B, H, and N. J. Biol. Chem. 264: 13165–70 (1989)
272. Tani T. et al.: Nucleotide sequence of the human pancreatic trypsinogen-III cDNA. Nucleic Acids Res. 18: 1631 (1990)
273. Tanji M., Kageyama T. and Takahashi K.: Tuna pepsinogens and pepsins: Purification, characterization and amino-terminal sequences. Eur. J. Biochem. 177: 251–259 (1988)
274. Taylor S. S., Buechler J. A. and Yonemoto W.: cAMP-dependent protein kinase: Framework for a diverse family of regulatory enzymes. Annual Rev. Biochem. 59: 971–1005 (1990)
275. Thalhofer H. P. and Hofer H. W.: Purification and properties of cyclic-3',5'-GMP-dependent kinase from the nematode Ascaris suum. Arch. Biochem. Biophys. 273: 535–542 (1989)
276. Titani K. et al.: Amino acid sequence of crayfish (Astacus fluviatilis) trypsin I-f. Biochemistry 22: 1459–65 (1983)
277. Titani K. et al.: Amino acid sequence of crayfish (Astacus fluviatilis) carboxypeptidase B. Biochemistry 23: 1245–50 (1984)
278. Titani K. et al.: Amino acid sequence of a unique protease from the crayfish Astacus fluviatilis. Biochemistry 26: 222–226 (1987)
279. Tsai I. H., Liu H. C. and Chuang K. L.: Properties of two chymotrypsins from the digestive gland of the prawn Penaeus monodon. FEBS Letters 203: 257–261 (1986)
280. Tschesche H., Kolkenbrock H. and Bode W.: The covalent structure of the elastase inhibitor from Anemonia sulcata – A „non-classical" Kazal-type protein. Biol. Chem. Hoppe-Seyler 368: 1297–1304 (1987)
281. Ulloa R. M. et al.: Cyclic AMP-dependent protein kinase activity in Trypanosoma cruzi. Biochem. J. 255: 319–326 (1988)
282. Urch U. A. and Hedrick J. L.: Isolation and characterization of the hatching enzyme from the amphibian, Xenopus laevis. Arch. Biochem. Biophys. 206: 424–431 (1981)
283. Vacquier V. D., Carner K. R. and Stout C. D.: Species-specific sequences of abalone lysin, the sperm protein that creates a hole in the egg envelope. Proc. Nat. Acad. Sci. USA 87: 5792–96 (1990)
284. Vendrell J., Cuchillo C. M. and Aviles S. X.: The tryptic activation pathway of monomeric procarboxypeptidase A. J. Biol. Chem. 265: 6949–53 (1990)
285. Viswanathan S. and Dignam J. D.: Seryl-tRNA synthetase from Bombyx mori. Purification and properties. J. biol. Chem. 263: 535–541 (1988)
286. Vogel R.: Natürliche Enzym-Inhibitoren. Thieme, Stuttgart 1984
287. van Waarde A.: What is the function of protein carboxyl methylation? (Review). Comp. Biochem. Physiol. 86B: 423–438 (1987)
288. Wagner P. et al.: Active transport of proteins into the nucleus (Minireview). FEBS Letters 275: 1–5 (1990)
289. Walldorf U. and Hovemann B. T.: Apis mellifera cytoplasmic elongation factor 1-alpha (EF-1-alpha) is closely related to Drosophila melanogaster EF-1-alpha. FEBS Letters 267: 245–249 (1990)
290. Warner A. H. and Shridhar V.: Purification and characterization of a cytosol protease from dormant cysts of the brine shrimp Artemia. J. biol. Chem 260: 7008–14 (1985)
291. Waxman L. et al.: Tick anticoagulant peptide (TAP) is a novel inhibitor of blood coagulation factor Xa. Science 248: 593–596 (1990)
292. Wernet W., Flockerzi V. and Hofmann F.: The cDNA of the two isoforms of bovine cGMP-dependent protein kinase. FEBS Letters 251: 191–196 (1989)
293. Wold F. and Moldave K. (eds.): Posttranslational modifications, 2 vol. set. Acad. Press, New York 1984
294. Wolfe F. H. et al.: Chicken skeletal muscle has three Ca-dependent proteinases. Biochim. biophys. Acta 998: 236–250 (1989)
295. Woodley C. L. et al.: Protein synthesis in brine shrimp embryos. Regulation of the formation of the ternary complex (Met-tRNA, eIF-2-GTP) by two purified protein factors and phosphorylation of Artemia eIF-2. Eur. J. Biochem. 117: 543–551 (1981)
296. Wun T. C. et al.: Cloning and characterization of a cDNA coding for the lipoprotein-associated coagulation inhibitor shows that it consists of three tandem Kunitz-type inhibitory domains. J. Biol. Chem. 263: 6001–04 (1988)
297. Yamada Y., Matsui T. and Aketa K.: Purification and characterization of a chymotrypsin-like enzyme from sperm of the sea urchin, Hemicentrotus pulcherrimus. Eur. J. Biochem. 122: 57–62 (1982)
298. Yamakami K.: Purification and properties of a thiol protease from lung fluke adult Paragonimus ohirai. Comp. Biochem. Physiol. 83B: 501–506 (1986)
299. Ye R. D. et al.: Structure of the gene for human plasminogen activator inhibitor-2. The nearest mammalian homologue of chicken ovalbumin. J. Biol. Chem. 264: 5495–5502 (1989)
300. Yoshinaka R. et al.: Enzymatic characterization of anionic trypsin of the catfish (Parasilurus asotus). Comp. Biochem. Physiol. Pt. B. 77: 1–6 (1984)
301. Yoshinaka R. et al.: Distribution of pancreatic elastase and metalloproteinase in vertebrates. Comp. Biochem. Physiol. Pt. B 83: 45–49 (1986)

4 Molecular Evolution

Ideas about the process and the laws of evolution are mainly based upon the comparison of characters of living organisms. Whilst classical research into evolution indirectly draws conclusions about evolutionary changes in the genotype by the observation of phenotypic differences, especially morphological characters, the molecular approach to evolution deals directly with the genome through sequence comparisons of nucleic acids and proteins. The amount of data now available at this level is enormous and increases continuously. DNA-sequence analysis, which has been available for only a few years, has already produced more information than the previous three decades of protein sequencing [377]. These days amino acid sequences are mostly determined from the coding nucleotide sequences rather than by direct analysis.

Molecular characters, such as nucleotide or amino acid sequences, and their directly correlated molecular properties are particularly useful for studies of evolution on several grounds: (1) they are unambiguous and there are no gradual transitions; (2) they are quantifiable and suitable for genetic analysis; (3) they always relate to distinct genomic sites (loci); (4) they are easily identifiable even when they are not mutated; (5) they are independent of internal and external conditions and are recognizable in heterozygotes, where there is no dominance or epistatic interaction between the genes; (6) they have an extraordinary capacity for coding information, and each

of the many positions in a sequence corresponds to a character with 20 (protein) or 4 (nucleic acid) different variants; (7) in contrast to morphological characters, they can usually be widely compared, often among all organisms; and (8) the convergence of long sequences is very unlikely and, therefore, numerical methods can be used to determine genetic relationships; derived phenotypic (e.g. morphological) characters are more uncertain of the possibility of convergence.

DNA variability is a prerequisite of evolution. According to the synthetic (neo-Darwinian) theory of evolution, a change in, or the emergence of a species is based upon the different effects of selection on genetic variation already available in the population. Alterations to the DNA as the carrier of genetic information in the germline arise because DNA replication and repair are not error-free [215]. Corresponding changes in somatic cells may have serious consequences for cell function and the fate of the whole organism – one has only to think of the genetic mechanisms of tumour formation – but are of negligible importance in evolution. The frequency of error during DNA replication and repair clearly must satisfy the need for correctly functioning gene products but, on the other hand, it should ensure sufficient variation to allow adaptation to changing conditions. The error frequency itself is quite clearly the result of selection. Gene duplication and an excess of non-coding sequences allow a relatively high variability in the DNA without endangering the supply of usable gene products. Mutations arise as genetic events in individuals. Only after the genetic change has spread through a population to the extent that it is present in many individuals, i.e. it is "fixed" in the population, does it have any importance for evolution; the genetic alterations occurring in the germline of single individuals become evolutionary events through their fixation in the population. In nucleotide substitution, for example, it is therefore necessary to distinguish clearly between "mutations" (genetic) and "substitutions" (evolutionary). Whether a newly occurring mutation becomes fixed in the population or is eliminated depends upon the effects of selection and change (drift).

Selection affects organisms rather than single genes. Its effects, however, can be detected in the whole cascade of gene function, from the nucleic acids via the proteins to the most complicated phenotypic characters (Fig. 4.1). Selection does not only affect the ability to survive; more than anything it selects for reproductive success (fit-

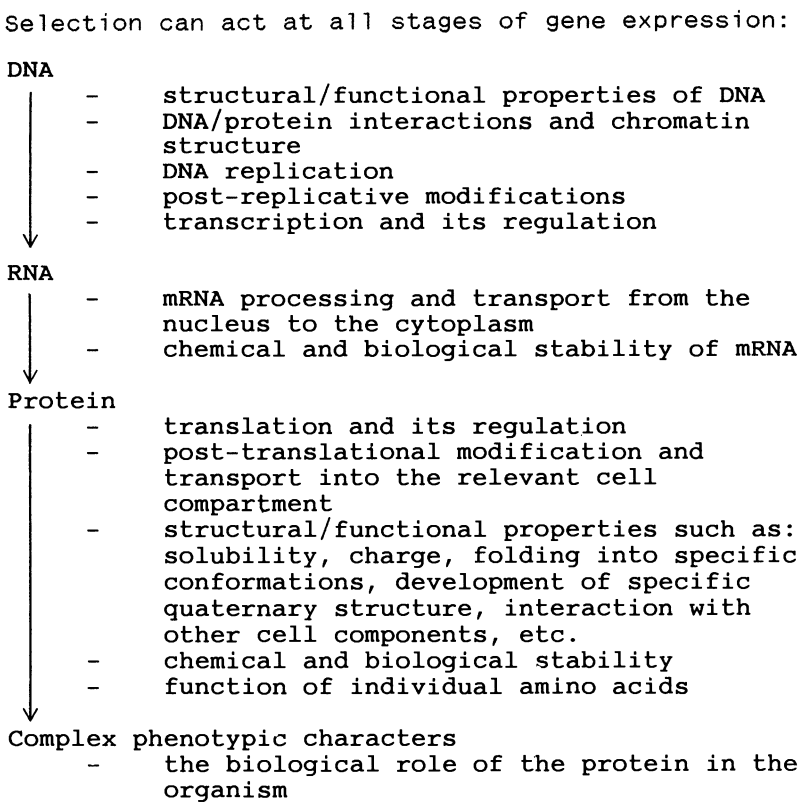


Fig. 4.1. Sites at which selection acts

ness), i.e. the ability to pass on genes to as many progeny as possible; selection doesn't always leave behind just corpses. Two types of selection can be conceptually distinguished. Negative or stabilizing selection eliminates those variants that display reduced fitness because of marked phenotypic changes. Positive selection, i.e. the fixation of fitness-increasing mutations, appears to be of rarer occurrence in nature.

Adaptive evolution is often understood to mean the "choice", made by the process of selection, of one of the many "solutions" available to the organism because of genetic variation to a "problem" posed by a change in the environment. However, the relationships between genes, organisms and the environment are much more complicated, as the following considerations show:

1. The quantitative variability of almost all characteristics of an organism arises during the unique development of that organism and results from non-trivial interactions with the environment; single genes or alleles cannot be assigned a particular selective value independent of the current environment.
2. Due to the imperfect reproducibility of the developmental process (developmental noise), various phenotypes may result from the same genotype in a given environment.
3. Internal and external factors have mutual influences; for example, in many cases, the selective pressure of the environment on a particular allele is dependent upon the genetic milieu or the gene frequency.
4. As each organism is constantly changing, the phenotype is determined not only by the genotype and the environment but also by the organism's own previous condition.
5. Organisms can actively alter the character of their environments.
6. The "environment" of most life processes is the internal milieu of the organism, and this is largely determined and maintained by regulatory processes, independent of changes in the outside world [21].

It was originally assumed that the direction and rate of evolution was determined entirely by the selection of advantageous variants. Impressed by the unexpectedly high molecular polymorphism in natural populations and the relatively constant rate of protein evolution (the molecular clock), Kimura, Ohta, King and Jukes formulated in 1968–1969 the **"neutral theory" of molecular evolution** [208, 213]. According to this theory, many perhaps most DNA alterations have such little importance for fitness that they are not subjected to any selection pressure and are only fixed by chance in the population. One could say that the organism itself does not "notice" such mutations, but they are, nevertheless, detectable by the precise tools of the molecular biologist. The lower the importance of the altered character for fitness, the higher is the probability, and thus the relative frequency, of selectively neutral alterations. Therefore, the proportion of selectively neutral alterations decreases in the gene-function cascade (Fig. 4.1) from the DNA to the complex morphological and physiological characters; the neutral theory is applicable most of all to the evolution of nucleic acids and proteins. Mutations that are neutral at a given point in evolutionary time may have a significant selective advantage or disadvantage at a later time and under different internal or external conditions. Such conditionally neutral variation represents a reservoir of genetic variability in the population for later evolutionary processes.

All in all, it may be said that research into evolution has benefited from the evaluation of molecular data by a whole series of fundamental new findings: (1) innumerable new mechanisms that give rise to genetic variation (mutation mechanisms) have been discovered; (2) the selective value of many evolutionary changes in DNA and their products is apparently so little that they are effectively "neutral" and are only spread and fixed in populations by chance (neutral theory of evolution); (3) in accordance with point 2 is the fact that the rate of evolution of molecular characters is relatively constant ("molecular clock"); (4) innovative evolution, e.g. the formation of proteins with a new function, never occurs via de novo synthesis of new DNA sequences but always through the alteration or recombination of existing sequences; and (5) the evolution of morphological characters is correlated in neither rate nor extent with the evolution of DNA and proteins.

Since the formulation of the neutral theory of evolution, the question of whether **Darwinism** has been undermined, or even completely disproved, by the new results of molecular biology has been discussed with great ideological enthusiasm. That King and Jukes chose *Non-Darwinian Evolution* as the rather unfortunate title of their first publication has undoubtedly contributed to the argument [213]. If, however, one defines Darwinian evolution as "the gradual transformation of a species over time by the forces of selection operating upon genetic variability existing in the

population" [122], it is clear that the findings of molecular biology have significantly broadened and enriched the neo-Darwinian (synthetic) theory rather than contradicted it. The neutral theory is a necessary supplement to neo-Darwinism in the molecular context and is not its antithesis. The attempts made in many publications to detect clear cases of selection at the molecular level are superfluous in as much as the neutral theory in no way denies their existence. Gene transfer between different species, and also between different eukaryotes, is now being discussed as a possible mechanism of evolution. Should it be shown that such a process is of frequent occurrence in evolution, this would make the reconstruction of phylogeny from molecular data more difficult but, looked upon as a mechanism for the emergence of new genetic variability, it fits easily into the Darwinian theory. The same argument would be valid for the recently suspected and then refuted inheritance of an acquired immunotolerance in mice; if at all applicable, this would be proof of gene transfer between immunocompetent somatic cells and germline cells but would in no way be an argument in favour of Lamarckism against Darwinism [122].

The present great interest in questions of molecular evolution is indicated by the large number of recent reviews and monographs on this theme [68, 212, 246, 369].

4.1 The Determination of Homology Between Protein DNA Sequences

Proteins or genes are considered to be homologous when they have the same origin, i.e. when they arise from the same gene. In many cases, **homology** can be assumed unconditionally, e.g. for proteins with the same function in closely related species, for immunologically cross-reactive proteins, and for proteins or genes with extensive sequence coincidence. However, to be certain about the homology between genes, it must be shown that the apparent agreement is not the result of either chance or convergence. The existence of significant agreement between long sequences due to convergent evolution can be ruled out on probability grounds. There are, in fact, many examples of proteins with similar function that show no detectable homology. For example, there is no sequence similarity between the ribonucleases of the bacteria and fungi, on the one hand, and animals, on the other hand; between phage and animal lysozymes; and between the different protein super-families of the peptide hydrolases or proteinase inhibitors. Also, the Cu, Zn superoxide dismutases from the cytoplasm of eukaryotes are not homologous to the corresponding Mn and Fe enzymes from prokaryote and eukaryote mitochondria. It is true that functionally similar amino acids are found in similar spatial structures of different proteins, but this does not produce statistically significant agreement when long stretches of sequence are considered.

The **spatial structures** of proteins are very conserved in evolution; as long as the basic function of a protein does not change, then the responsible spatial structure is maintained. It seems reasonable, therefore, to take into account spatial structure as evidence for distant relationships between proteins, perhaps even "when all traces in the sequence have been obliterated" [368]. For comparison of spatial structures, the distance between the polypeptide chains that project into each other can be calculated. A further possibility is to describe polypeptides in terms of sequences of rotational angles around the bonds $N\text{-}C^{a}$ and $C^{a}\text{-}C^{carboxyl}$ and then compare them. The probability of chance in the agreement of spatial structures can be estimated by means of relevant comparisons of simulated folded chains [193, 368, 407, 420]. It must not be forgotten, however, that convergent evolution of spatial structures to fulfil similar functions can never be totally excluded.

Two proteins may be considered as homologues if their **sequence agreement** is higher than would be possible by chance alone. In order to test this, the sequences to be compared must first be arranged opposite each other (aligned) so that they show the highest possible agreement (S). A series of appropriate alignments are then made with random sequences of the identical amino acids, and the average agreement, together with the standard deviation thereof ($S_r \pm SD_r$), is determined. If the quotient

$$A = (S - S_r)/SD_r \qquad (4.1)$$

exceeds a given value, which is arbitrarily set between 3 and 5, then the agreement shown by the original sequences is considered to be significant [87, 93].

To calculate the **similarity** (S) of two sequences, values of 1 and 0 are given to each pair of identical and non-identical amino acids, respectively, and the values for each position are added together. Instead of this "unitary matrix", which

counts only identity, other **matrices** have been suggested and in these various amino acid pairs are weighted, for example, according to differences in the coding triplets, physicochemical similarities, or the relative probability of substitutions. Of the various matrices, the most frequently used is the "mutation data matrix" (MDM-78) of Dayhoff; this assigns each of the possible 400 amino acid pairings a specific probability based on analyses of closely related proteins (p. 149). Comparison of the various methods shows that weighted matrices are only of advantage when the agreement of the compared sequences is relatively low (below 30 %). Homology determinations with a higher sensitivity may be obtained by considering the physicochemical similarities of amino acids and substitution probabilities [11, 113, 150, 349], or by reference to a consensus sequence containing all the important characteristics of a protein family [320].

To achieve optimal agreement it is often necessary to insert **"gaps"** into one or both sequences, or to shift one sequence relative to the other so that unpaired stretches (**"tails"**) are formed. In this way, however, the chances of random agreement increase. Two random sequences with the same proportions of all 20 amino acids agree, on average, by 5 %; deviations in the composition increase this value to about 6 % for the "average protein" (see Table 3.1, p. 71). If in random sequences of 100 amino acids relative shifts of up to 5 amino acids are allowed, the average agreement will have already increased to 8 %; the introduction of one chosen gap can also increase the agreement from 5 to 8.5 % [87, 93]. Tails and gaps arise in evolution through the deletion or insertion of one or more triplet codons and must be looked upon as real situations in aligning sequences. However, their influence on the statistical significance of sequence agreement must be taken into account; a particular value (a gap penalty) is usually subtracted from the similarity factor for each inserted gap [93]. The comparison of several sequences by alignment in pairs usually requires various different gaps; this avoids the method of multiple alignment based on the principle "once a gap, always a gap" [114].

The arithmetical process (algorithm) of Needleman and Wunsch [286] is most often used to find the **optimal alignment of two sequences**. Orientation of the two sequences at right angles to each other produces a field on which the similarity between an amino acid of one sequence and each of the amino acids of the opposing sequence can be plotted. The pathway across this field that gives the highest total value of agreement after subtraction of the gap penalties gives the optimal alignment (Fig. 4.2). Computer programs are now used for this manipulation and for the relevant statistical tests. There are also programs for determining internal periodicities of proteins, which can arise by duplication of gene segments, and for searching for related sequences or partial sequences in databanks. Periodicity is detected by comparing a partial sequence of a particular length with all possible partial sequences of that length in the same protein; in the search for related sequences, corresponding comparisons with all partial sequences in a databank are carried out. A further method for aligning sequences is that of the "dot matrix", in which the "dot criterion" is the number of agreements m in n consecutive amino acids [218]. Due to the continuously increasing amount of data, such sequence comparisons require increasing mathematical effort. The development of new or modified computer programs for molecular research into evolution is concerned, above all, with simplifying the arithmetic (algorithms) and saving time [29, 96]. It should then be possible, for example, to compare

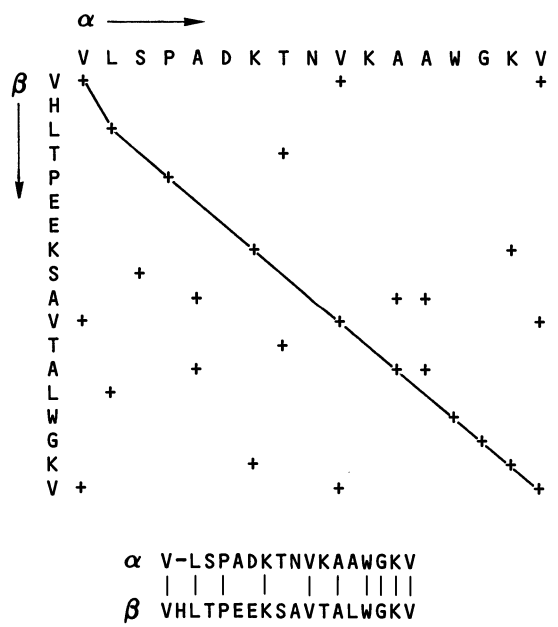


Fig. 4.2. The establishment of the optimal alignment of two sequences by the method of Needleman and Wunsch [286], illustrated here for the N-terminal sequences of the human α- and β-globin chains. A route that includes the most + points is sought from the beginning *(upper left)* to the end *(lower right)* of the sequences. Gaps in one or other sequence are seen as deviations from the diagonal. The optimal alignment of these two sequences requires a gap between positions 1 and 2 of the α-chain

Table 4.1. Pairs of proteins whose homology was tested using the method described on p. 115 [93]. In each case the length of the compared sequences are given in ( ); the percentage agreement, the number of gaps needed for optimal alignment and the quotient A [calculated from Eq. (4.1)], which describes the statistical significance, are indicated. The compared sequences may be considered homologous when A is larger than 3. To calculate the agreement, identical amino acids are given the value 1 and Cys/Cys pairs the value 2; the gap penalty (p. 115) was 2.5; "tails" (p. 115) were not considered

Compared sequences	Identity (%)	Gaps	A
Human haemoglobin: Hb-β (146)/Hb-δ (146)	93	0	54.0
Pig lactate dehydrogenase: LDH-M (333)/LDH-H (331)	75	1	75.0
Human carboanhydrase: CA-B (260)/Ca-C (259)	61	1	56.8
Bovine chymotrypsinogen-A (245)/trypsinogen (229)	46	6	22.6
Human haemoglobin: Hb-β (146)/Hb-α (141)	44	2	17.0
Human immunoglobulin: Cλ (102)/Cκ (104)	42	3	13.1
Bacterial trypsin *Streptomyces griseus* (221)/ vertebrate trypsin *Mustelus canis* (222)	38	8	16.9
Chicken lysozyme (129)/human α-lactalbumin (123)	38	3	10.7
Human fibrinogen: β-chain (461)/γ-chain (411)	33	5	31.2
Snake toxin-cardiotoxin *Bungarus* (118)/ pig prophospholipase (131)	32	5	5.0
Chicken ovalbumin (386)/human antithrombin III (423)	28	6	14.3
Carp parvalbumin (108)/bovine troponin-C (161)	27	2	6.1
Human Hb-α (141)/human myoglobin (153)	27	1	9.3
Haemoglobin: *Chironomus* (152)/*Myxine* (148)	26	3	3.7
Human cytochrome c (104)/*Euglena* cytochrome f (87)	25	3	2.9
Human apolipoprotein AI (125/245)/*Corynebacter* diphtheria toxin fragment (125)	25	1	5.2
Elephant insulin (51)/pig relaxin (48)	24	1	4.1
Human follitropin (92)/human thyrotropin-β (112)	22	2	2.9
Sheep κ-casein (171)/human fibrinogen γ-chain (179/411)	21	5	1.4
Bovine chymotrypsin-A (245)/human haptoglobin-β (245)	19	5	3.2
Human fibrinogen (N-terminal): α-chain (239)/γ-chain (239)	16	2	5.4

more than two sequences simultaneously [6, 248, 398, 406]. From time to time, completely new approaches to the comparison of protein sequences are suggested [23, 445]. With the methods described, it is possible in certain situations to recognize two amino acid sequences as homologous, although they might differ in more than 75 % of positions (Table 4.1). Large multi-domain proteins, of which there are many, present particular problems in the assessment of homology. The coding sequences of their elements ("modules") mostly have their origin in different genes that were recombined by exon shuffling. The change in function accompanying this incorporation into a new protein can lead to drastic changes in sequence which obscure its origin [321].

DNA sequences can be compared by similar methods [29]. However, significant agreement is more difficult to show statistically because the existence of only four different residues already leads to an average random-sequence agreement of 25 %. In fact, there are examples where no homology was detectable at the gene level for clearly homologous proteins, e.g. the oncogenes of the sarcoma viruses from the mouse and chicken [93]. The assessment of homology of non-coding regions is complicated by the frequent deletions and insertions that occur in association with the high incidence of substitution.

Homology determination between DNA or protein sequences is the methodological prerequisite for all investigations of the course and mechanisms of molecular evolution. Of particular importance is the demonstration that all the proteins of present-day organisms may be arranged in several hundred groups of homologous sequences, known as **protein super-families**, after the suggestion of Margaret Dayhoff. The *Atlas of Protein Sequence and Structure*, which appeared in 1978 [87], listed 248 such protein super-families. Several more have been described since then (Table 4.2), and protein sequences that do not fit into any of the known families are continuously being discovered; however it is likely that the total number of super-families is of the order of 400, or at most 1000. There are many cases of homology between prokaryote and eukaryote proteins, e.g. cytochrome, ferredoxin, glyceraldehyde-3-phosphate dehydrogenase, dihydrofolate reductase, trypsin-like enzymes and triose-

Table 4.2. Selected examples of protein super-families. The super-families listed by M. Dayhoff [87] are given in parentheses with the original numbers. The protein super-families of the serine-protease inhibitors may be found in Table 3.7 and the super-families found in mammalian blood plasma are given in Table 5.3

(1)	Cytochrome c from vertebrates, invertebrates, higher plants, lower fungi and *Tetrahymena*; various bacterial cytochromes
(4)	Cytochrome b_5 of mammals and birds; cytochrome b_2 of yeast
(6)	Ferredoxin of bacteria and higher plants; bovine adrenodoxin
(11)	Bacterial azurins and plastocyanins from blue algae and higher plants
(13)	Vertebrate and yeast alcohol dehyrogenases
(16)	Glyceraldeyhde-3-phosphate dehydrogenase from vertebrates, lobster, yeast and bacteria
(18)	Dihydrofolate reductase from mammals and bacteria
(28)	Phospholipase A_2 from vertebrates, snakes and insects
(33)	Lysozyme from mammals and bird eggs; α-lactalbumin of mammals
(39)	Trypsin, chymotrypsin, elastases, serine proteases of blood clotting, fibrinolysis and the complement cascade, haptoglobin
(43)	Pespinogen, prochymosin, penicillopepsin
(51)	Triosephosphate isomerase of vertebrates and bacteria
(68)	Prolactin, somatotropin and choriomammotropin
(73)	Glucagon, gastric inhibitor polypeptide, secretin, vasoactive intestinal peptide and pancreas hormone of mammals and birds
(75)	Proinsulins of vertebrates; insulin-like growth factors and relaxin
(76)	Gastrin and cholecystokinin-pancreozymin
(77)	Neurotoxins and cytotoxins from snake poison
(81)	Melittin of honey bees and bombinin from toads
(89)	Immunoglobulins, T cell receptors, major histocompatability complex (MHC) classes I and II, β_2-microglobulin
(90)	Globins of vertebrates, invertebrates and higher plants
(148)	Calmodulins, troponin C, alkali-soluble and regulatory light chains of myosins, parvalbumins and calcium-binding proteins from mammalian gut
(168)	Albumin, α-fetoprotein
(–)	Ceruloplasmin, clotting factors V and VIII
(–)	α_2-macroglobulin, pregnancy-associated α_2-glycoprotein, complement components C3, C4 and C5
(–)	α_{2u}-Globulin of rats, β-lactoglobulin, retinol-binding plasma protein, α_1-microglobulin, apolipoprotein D, α_1-acid glycoprotein, androgen-dependent protein of the epididymis, chicken purpurine, frog olfactory protein, insecticyanin of the tobacco hornworm *Manduca sexta*, mouse urinary proteins (MUPs), aphrodisin (α_{2u}-protein super-family)
(–)	C-reactive protein (CRP), serum amyloid protein (SAP), female-specific protein of the golden hamster (HFP) (protein super-family pentraxins)

phosphate isomerase [93]. All super-families probably came into being at the beginning of the organismic, or even at the end of the pre-organismic, phase of evolution and reflect the protein spectrum of the first primitive prokaryotes [87]. Proteins with completely novel functions also apparently arose by modifications to the 500–1000 ancestral proteins; many proteins are older than their present-day functions. Hence, a range of completely different functions may be found in one super-family (Table 4.2).

From the foregoing it is clear how much the **concept of homology** has changed during transfer to the molecular biology level [319]. With the statement that the wing of a bird and the foreleg of a mammal are homologous is meant that these two parts are similar in nature and are genealogically comparable in their structure and relations to other parts of the body, but they are not necessarily determined by transformation states of the same gene; it is exactly the latter that is understood by the homology concept in molecular biology. The presence of homologous organs, like the chorda dorsalis and neural canal, allow us to recognize family relationships between such different forms of life as ascidians, lancet fish and mice; the homology of the glyceraldehyde-3-phosphate dehydrogenases from the bacteria to man is only further evidence that all life can be traced back to a common ancestry. The homology concept of molecular biology has a stricter definition than the classical idea, but has a more limited heuristic value. However, it must be recognized that the homology of macromolecules is easier to determine than homology between morphological characters; it can be proved statistically, is less easily mistaken for convergence or parallel evolution, and is therefore especially suitable for the recognition of distant relationships.

4.2 The Mechanisms of Molecular Evolution

It is only since the possibility of comparing homologous DNA sequences in variously related individuals appeared that the wide variety of molecular

processes giving rise to genetic variability has been recognized. The different **types of mutation** and their phenotypic consequences can now be more closely studied. The genetic and evolutionary importance of point mutations, i.e. the substitution, insertion or deletion of single nucleotides, has been known for a long time. Today, however, many other genetic events are recognized that lead to changes in larger segments of DNA, e.g. duplication of genes or parts of genes and the insertion or deletion of longer or shorter DNA sequences. Gene conversion has been discovered as a mechanism which, in addition to classical crossing-over, can lead to intrachromosomal recombination. DNA sequences are transferred to the other DNA strand by inversion, and to completely different sites by transpositions; exon shuffling combines parts of different genes to produce totally new genes. The evolution of multigene families presents special problems and, above all, the question of how the often observed similarity of the multiplied sequences is maintained, despite their mutability. Of particular interest for studies of evolution is the ability of many middle repetitive sequences to be transposed to other sites in the genome, in many cases carrying with them further fragments of DNA.

The mechanisms of all these processes are as yet only incompletely understood; DNA sequence analysis shows, however, that rearrangements of the genome are very frequent events. The previous view of a more-or-less unchangeable, static genome has been replaced by a picture of a dynamic genome changing rapidly during evolution. There is now no doubt that DNA rearrangements are of particular importance for the evolution of complex phenotypic characters. Only in exceptional cases are large-scale alterations in the DNA sequence visible microscopically (chromosome mutations); in contrast, genome mutations such as polyploidization are easily detectable.

4.2.1 Nucleotide Substitution

There are $4 \cdot 3 = 12$ substitution possibilities for the four different nucleotides; in eight cases a purine is replaced by a pyrimidine or vice versa **(transversions)**, and in four cases there is a substitution of pyrimidine by pyrimidine or purine by purine **(transition)** (Fig. 4.3). The ratio transition: transversion deviates markedly from a random distribution, particularly in the early stages of molecular evolution. The proportion of transitions can be as high as 90 % for closely related sequences, but then decreases during the extended course of evolution to the theoretical value of approximately 33 %, depending upon the base composition [179]. An explanation for this phenomenon is perhaps to be found in the interactions between purines on different, neighbouring DNA strands. These interactions influence conformation parameters of the DNA double-helix; transversions lead to changes in these parameters and may, therefore, be subjected to stronger selection pressure than transitions [232].

Reference to the genetic code allows predictions of which **nucleotide substitutions** in protein-coding sequences lead to amino acid substitution, which are synonymous and which produce stop codons; synonimity results especially from most of the nucleotide substitutions in the third triplet position (Table 4.3). An average of 6.5 different amino acid substitutions per codon may result from the substitution of one nucleotide; each of the three possible nucleotide substitutions in the first or second position almost always leads to an amino acid substitution, but this happens for only a few third-position nucleotide substitutions. Thus, one-step mutations of, for example, a polypeptide like the β-globin chain with its 146 amino acids can produce $146 \cdot 6.5 = 949$ different alleles.

One might assume a priori that **synonymous mutations** would not be subject to selection and, in fact, comparisons of related sequences show synonymous substitutions to occur more frequently than would be expected from the genetic code (Table 4.3); they are mostly more frequent even than substitutions leading to amino acid exchange (Table 4.4). However, they are not

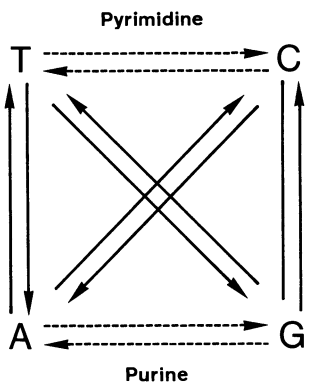


Fig. 4.3. Of the twelve possible nucleotide exchanges, four are transitions (pyrimidine ⇐---⇒ pyrimidine or purine ⇐---⇒ purine) and eight are transversions (pyrimidine ⟷ purine)

Table 4.3. The consequences of nucleotide exchange in protein-coding sequences

From the $61 \cdot 9 =$ **549** possible mutations of amino acid codons:
392 (71 %) give another amino acid codon (amino acid exchange, missense)
134 (24 %) give another codon for the same amino acid (synonymous, same-sense)
23 (4 %) give a stop codon (nonsense)

From the $3 \cdot 9 =$ **27** possible mutations of a stop codon:
23 (85 %) give an amino acid codon
4 (15 %) give a stop codon

The $61 \cdot 3 =$ **183** possible mutations of each position of an amino acid codon are:

	Amino acid exchanging	Synonymous	Nonsense
In position 1	166	8	9
In position 2	176	0	7
In position 3	50	126	7
Total	392	134	23

Of the **392** amino acid-exchanging mutations:
184 (47 %) change the polarity class[a]
128 (33 %) change the charge (i.e. charge changes account for 23 % of all possible mutations)
273 (70 %) are non-conservative, i.e. the new amino acids no longer belong to the same class (as defined by Dayhoff[b])

[a] Polarity classes [133]:
polar – Arg, His, Lys, Asn, Asp, Gln, Glu
weakly polar – Ala, Gly, Pro, Ser, Thr
non-polar – Cys, Ile, Leu, Met, Phe, Trp, Tyr, Val

[b] Amino acid classes according to Dayhoff [87]:
1. Cys
2. Hydropholic – Ala, Gly, Pro, Ser, Thr
3. Acidic and acid amides – Asn, Asp, Gln, Glu
4. Basic – Arg, His, Lys
5. Hydrophobic – Ile, Leu, Met, Val
6. Aromatic – Phe, Trp, Tyr

Table 4.4. The proportion of synonymous substitutions in various gene pairs

Genes compared	Synonymity (%)	Reference
Histones H2A, H2B and H3 *Strongylocentrotus pupuratus/Psammechinus miliaris*	92	[199]
α-Subunit gonadotropin rat/mouse	83	[136]
– their signal sequences	25	[136]
Various chorion genes *Bombyx mori*	74[a]	[195]
α-Globin rabbit/mouse	65	[199]
Preproinsulin man/rat	63	[199]
Chicken β_3-globin/β-globin rabbit, mouse, man	52[a]	[199]
Growth hormone man/rat	46	[199]
Immunoglobulin-C_x rat allotype 1b/mouse	43	[374]
Mouse β-globin minor/major	41	[199]
Human growth hormone/chorion somatomammotropin	33	[199]
Rat immunoglobulin IgC_x allotypes 1a/1b	8	[374]
mTDNA cytochrome oxidase subunit II		
– *Rattus norvegicus/R. rattus*	94	[42]
– *R. norvegicus*/bovine	78	[42]
– *R. norvegicus*/human	56	[42]

[a] Mean values.

completely neutral, they are not randomly distributed throughout the coding sequences [348], and the frequency of their occurrence is lower, for example, than the rate of evolution of pseudogenes. Above all, it is clear that not all possible synonymous codons for particular amino acids are equally used; there are codon preferences that show species-, tissue- and even gene-specific differences. The basis of this variable use of synonymous codons has not been fully explained [376], but species- and tissue-specific differences in the proportions of particular isoacceptor tRNAs are undoubtedly important [188]. In any case, it is clearly seen that even very slight varia-

tion in selection can lead to marked differences in the use of synonymous codons [244], and that molecular evolution is retarded by the existence of codon preference [210, 212].

Amongst the **mutations leading to amino acid substitution**, the stabilizing effect of selection favours conservative exchanges, i.e. those that change only minimally the physicochemical properties that are important for protein structure and function. Because the spatial structure of proteins is generally maintained during evolution, even when most of the amino acids have been exchanged, the properties of the amino acids that determine the spatial structure clearly play a significant role. These properties include hydrophobicity, charge, influence on the secondary structure of the polypeptide chains and the spatial requirements of the side-chains. On the basis of such criteria, various indices of chemical similarity have been formulated and the amino acids classified in different ways (see footnotes to Table 4.3). It has been shown in many investigations that conservative exchanges are more frequent than would be expected for random substitution [87, 210, 212]. For example, in the distribution into polarity classes of eight protein families, 88 % of substitutions in the inner regions of the molecules are conservative compared with 76 % for amino acids on the surface, i.e. there were no changes of polarity class; non-conservative exchanges would have led internally to a decrease and externally to an increase in polarity. According to the genetic code, each amino acid substitution requires one, two or three nucleotide substitutions, and each amino acid pair thus corresponds to a "minimal mutation distance" (MMD). If the genetic code were itself optimized for minimal mutational effects, then conservative exchange would be more frequent after one-step than after two- or three-step mutations; however this is not the case. Even conservative amino acid substitutions can have lethal consequences. The substitution 104-Glu to Asp in human triosephosphate isomerase makes the enzyme thermolabile; the resulting enzyme deficiency in homozygous carriers of the allele causes haemolytic anaemia and neuromuscular dysfunction [83]. The exchange of a glycine codon for a serine codon in the gart gene complex of *Drosophila melanogaster* leads to a complete breakdown of purine synthesis [168].

Nucleotide substitutions that place stop codons within coding sequences **(nonsense mutations)** are unquestionably always selectively negative. For example, β-thalassaemia, a heritable disease in humans which leads to reduced synthesis of the haemoglobin β-chain, can be caused by mutation of the 39-Gln codon CAG to the stop codon TAG [144]. The Adh^{nB} mutant of *Drosophila melanogaster* codes for a 2-kDa smaller alcohol dehydrogenase that has only 1 % of the normal activity, due mostly to enhanced degradation; the cause here is the mutation 235-TGG (Trp) to TGA (stop) [260]. Amongst the more than 100 known mutations of the unc-54 gene of the nematode *Caenorhabditis elegans* there are four mutations of the sort CAA (Gln) to TAA (stop), and one mutation of CAG (Gln) to TAG (stop) [91]. The fatal consequences of a nonsense mutation can be mitigated at least temporarily, by the presence of a nonsense-suppressor tRNA which can read through the stop codon. The **deletion** or **insertion** of one or two base pairs in a coding sequence leads to a shift in the reading frame **(frame-shift mutation)**, usually with far-reaching consequences. Thus, the δ-globin gene of the anubis ape, *Papio doguera*, is not expressed because the insertion of a nucleotide in codon 55 shifts the reading frame to the extent that a stop codon appears at position 59 [207]. The deletion or insertion of one or more complete triplets leaves the reading frame unaltered and is more often than not without negative consequences.

Nucleotide substitutions, deletions or insertions in non-coding sequences are often selectively neutral, but they may have far-reaching phenotypic consequences. For example, two Adh mutants in *Drosophila melanogaster* show a nucleotide substitution at the intron/exon border of the second intron; instead of GT ... AG, one has GT ... GG and the other has GC ... AG. Consequently the mutants have no alcohol dehydrogenase (ADH) due to deficient mRNA maturation. One case of β-thalassaemia was the result of a mutation in the TATA box; ATACAA was found instead of ATAAAA, and the transcription of the β-globin gene was drastically reduced [399]. Another heritable human disease, the persistence of the foetal $^{A}\gamma$ haemoglobin in adults, could be traced to a mutation in the upstream CCAAT box; the nucleotide substitution AACCAAT to GACCAAT changes the expression of the $^{A}\gamma$ haemoglobin gene [129]. The substitution rate in coding and non-coding DNA regions influences the base composition and, conversely, is influenced by it. The AT-rich isochores of the cold-blooded vertebrates have changed to GC-rich ones in the warm-blooded birds and mammals [26]. A comparison of 42 protein-coding genes between man and the rat demonstrated a

selection pressure against A and T in synonymous positions [411].

4.2.2 Deletion, Insertion and Duplication of DNA Sequences

Deletions, insertions and duplications arise by **mispairing of DNA strands** and this results in unequal crossing-over at meiosis or in error-prone replication or repair of DNA. Here, in contrast to nucleotide substitution, a single mutational event changes a large number of nucleotides. The paired sequences show complementary bases but belong to different DNA regions; short, repetitive sequences are most often involved. Hence, deletions and duplications are sequence-dependent mutations which occur particularly frequently in regions with several repetitive sequences ("hot spots"). For example, regions of 15 human interferon genes containing several deletions or insertions always have neighbouring normal or inverted repeats at which mismatching or loop formation could occur [137]. The gene for the LDL receptor in humans has inverted repeats in exon 13 and intron 15, and these allow the formation of a DNA loop and deletion of the loop sequence; the resulting defect in the LDL receptor is the cause of heritable hypercholesterinaemia [230].

The deletion or duplication of shorter sequences, so-called **segment mutations**, are apparently very frequent genetic events and form important genetic mechanisms of molecular evolution. The spontaneous duplication rate at the male "maroon-like" locus of *Drosophila melanogaster* was found, using genetic methods, to be $2.7 \cdot 10^{-6}$; that at the "rosy" locus was even higher at $1.7 \cdot 10^{-4}$ [372]. Many segment mutations are known for the vertebrate globin genes; however, a particularly good example is the multi-gene family of the chorion proteins of the silkworms. Because of their close relationship the evolution of this multi-gene family has not been complicated by multiple mutations [195]. Segment mutations as genetic events are just as frequent as nucleotide substitutions but become fixed less often in populations. In the chorion genes, nucleotide substitutions outweigh segment mutations in non-coding regions by 4.5:1, and in coding regions by as much as 10:1 [195]. Segment mutations in coding sequences can lead to elongation or shortening of the polypeptides and are eliminated by selection if they result in frame-shifts, disturbances in protein conformation or changes in essential protein functions [195, 455].

Table 4.5. Periodic structures in proteins [5, 87, 99]. In each case, the total length of the protein is given together with the number and length of the repeats and the proportion of the total length consisting of repeats

Protein	Total length	Repeats		Proportion
		Length	Number	(%)
Collagen α1-chain (rat)	1052	3	337	96
Lipid-binding protein A1 (human)	245	11	18	81
Keratin B2A (sheep)	171	10	13	76
Antifreeze protein (p. 207)	12–165	3	4–55	100
Tropomyosin α-chain (rabbit)	284	21/42	7	100
Plasminogen (human)	790	79	5	50
Bombinin *(Bombina sp.)*	24	4	4	67
Immunoglobulin C_μ-chain (human)	452	108	4	96
IgC_γ-chain (guinea-pig)	329	108	3	98
Serum albumin (human)	584	195[a]	3	100
Haemopexin (human)	439	45	8	>80
Ovomucoid *(Coturnix coturnix)*	186	59	3	95
Histone H3 (bovine)	135	9/13	3/2	39
Troponin C skeletal muscle (rabbit)	159	76	2	96
Protease inhibitor, salivary gland (dog)	115	54	2	94
Haptoglobin α2-chain (human)	143	59	2	83
Ceruloplasmin	564[b]	224	2	79
Parvalbumin *(Esox lucius)*	108	39	2	72
Neurophysin 2 (pig)	92	23	2	50
α-Crystallin A (bovine)	173	30	2	35
Prothrombin (bovine)	582	79	2	27

[a] The repeat itself shows periodicity.
[b] A partial sequence from a total length of 1050 amino acids.

The **duplication** of a complete gene, including the transcription signals, gives two identical genes or, in the case of repeated duplication, a series of consecutively arranged identical genes (a tandem cluster), all of which initially form the same gene product. Processes like this, which lead to the formation of multi-gene families have occurred frequently in molecular evolution. If only the coding region, or a part of it, is reproduced (an internal duplication), then the resulting protein displays an internal periodicity with two or more homologous sections; this situation is detectable by the previously described methods even when the sequence agreement has been partly eliminated by amino acid substitution. The symmetry and periodicity in the spatial structure of many proteins also points to evolution through gene multiplication, although convergence by the occurrence of similar secondary structures cannot be excluded [307]. Internal periodicities of proteins are wide spread (Table 4.5). In some cases, it has been shown that the multiply occurring protein regions themselves have internal periodicities produced by the multiplications of even shorter sequences. This observation leads to the conclusion that the original (primordial) polypeptides were short, and only by multiple duplications were the greater lengths of the present proteins reached. This is particularly clearly seen in collagens, which are largely made up of several hundred tripeptide sequences.

4.2.3 Gene Fusion and Exon Shuffling

If, between two genes, the region containing the termination and initiation signals is removed by deletion then a fusion product is formed. Good examples of **gene fusion** as a mechanism of molecular evolution are to be found in the enzymes of fatty acid synthesis. This always requires seven catalytic and one acyl-carrier function. *E. coli* and higher plants accordingly have seven monofunctional proteins; in yeast, the same functions are performed by two multi-functional proteins, and in vertebrates they are performed by just one protein. The genes for the multi-functional proteins are unquestionably the result of the fusion of the genes for the mono-functional molecules; the different orders of the domains indicate that independent fusions occurred in the fungi and the vertebrates [268]. The fusion of neighbouring genes of a gene family is by no means a rare phenomenon. For example, a series of haemoglobin anomalies have arisen by the fusion of different globin chains: Hb-Lepore from δ and β and Hb-Kenya from $^{A}\gamma$ and β; there is also a form of α-thalassaemia that involves fusion of the two α-globin genes [275].

The fact that domains with coincident structures are often found in different proteins has led to the idea of new combinations of gene segments arising by **exon shuffling**. An oft-quoted example is the nucleotide-binding "Rossmann fold" which is found, for example, in various dehydrogenases [glyceraldehyde-3-phosphate dehydrogenase (GAPDH), lactate dehydrogenase (LDH), malate dehydrogenase (MDH), and ADH], phosphoglycerate kinases and glycogen phosphorylases [276]. Despite great similarity in their spatial structure, no significant sequence agreement is found between the nucleotide-binding regions; in this case the possibility of convergent evolution can certainly not be excluded. There are, however, many clear examples of exon shuffling. Thus, a sequence of 45 amino acids, which is repeated eight times in the epidermal growth factor (EGF), is found in many other proteins with completely different functions, e.g. in the complement component C9 and urokinase, in the blood-clotting factors VII, IX and X and in thrombomodulin, in the LDL receptor and the thyroid gland peroxidase, and in the product of the notch locus of *Drosophila* and the lin-12 gene of *Caenorhabditis*. "Kringle" structures and "zinc" fingers are also widely found [95]. The best examples of exon shuffling are provided by the plasma proteases of blood clotting, fibrinolysis and the complement cascade of mammals (see Fig. 3.6, p. 91). Here, the non-catalytic regions are made up of different combinations of "kringle", "zinc finger", growth factor and Ca^{2+}-binding domains, homologues of which are found in fibronectin, EGF precursors and the LDL receptor. During the evolution of the plasma proteases, these sequences were inserted between the signal sequence and the activation peptide of a primitive trypsin-like proteinase and were subsequently partly duplicated and translocated.

4.2.4 Transposition of DNA Sequences

Systematic investigations of the **transposition of mobile elements** and the phenotypic consequences of such genetic events have been carried out, in particular, with *Drosophila* and the nematode *Caenorhabditis elegans* [74, 367]. The spread of transposition variants in a population can be described by particular mathematical models

[63]. In this way, it has become clear that this phenomenon is extremely important in evolution. The insertion or deletion of mobile elements are apparently frequent events with mutation rates of more than 10^{-5} per generation. During the "hybrid dysgenesis" that occurs in the progeny of a male *Drosophila* with transposase-producing P or I elements and a female lacking elements, the transposition rate for these elements, and also for other mobile elements, increases transiently to as high as 10^{-2} [131].

The insertion or deletion of transposable elements usually leads to extensive rearrangements in the affected DNA region, and this may be one of the main causes of sexual isolation and, thereby, the formation of new species [400]. The sequence equalization between members of a multi-gene family (horizontal evolution) can be prevented by such insertions; thus, the Alu elements found at different sites in the mammalian β-globin gene cluster have allowed the development and maintenance of adaptive differences between embryonal, foetal and adult globins [39, 365]. The insertion of a transposable element in a transcription unit or in a region flanking a gene may inhibit gene expression, or it may stimulate expression by the introduction of a promoter or a transcription-enhancing sequence (an enhancer); the enhancer may make the gene hormone-dependent. The introduction of a termination signal, e.g. in an intron, can produce an entirely new gene [131]. The phenotypic consequences of a transposition are determined by the nature of the transposed sequence and by the site of insertion. Hence, in the *Drosophila* notch locus, various mobile elements create completely different eye mutants [206].

It appears that a large proportion of spontaneous mutations of complex phenotypic characters involve the insertion of mobile elements into active regions of the genome. For example, in *Drosophila*, different eye-colour mutations are caused by the insertion of such elements into the white locus; in some cases these are reversed by their deletion [45]. The expression of the *Drosophila* gene Sgs-4 is reduced 50- to 100-fold by insertion of the mobile element "hobo" into the 5′ controlling region; this gene codes for a glue protein in the salivary gland of the larvae. The glue protein sticks the pupal membrane to the substratum surface [269]. In a "low-activity" strain of *Drosophila melanogaster*, transposition of copia or copia-like elements to particular "hot spots" brings about a drastic increase in mating success [317]. In two human patients with the blood-clotting deficiency haemophilia A, which involves the absence of factor VIII, an L1 element has been found inserted into exon 14 of the factor VIII gene on the X chromosome [203]. The insertion of Alu repeats can result in widespread rearrangements and thus to dysfunction of the genes, as seen, for example, in LDL receptors, in the β-globin cluster and in the gene for the inhibitor of the complement component C1 [39, 397].

4.2.5 The Evolution of Multi-Gene Families

Multi-gene families show special evolutionary phenomena such as homogenization of sequences, evolutionary changes in copy number and the formation of pseudogenes. In particular, a gene family organized as a coherent gene cluster represents a genetic unit, a "super gene" [21]. The sequence similarity between the genes of a family is the result of a particular evolutionary process known as concerted, coincident or **horizontal evolution**. This cannot simply be the result of selection, because there is undoubtedly no strong selection pressure if one of many identical genes mutates. The terms "concerted" and "coincident" are not appropriate in that the selection-driven, parallel evolution of all members of a family is not involved, but rather a process which, on the one hand, mutually aligns the sequences of all members and, on the other hand, allows variants to spread rapidly throughout the whole family and subsequently become fixed in the population. Horizontal evolution has been closely investigated in 18S and 28S rDNA, in 5S rDNA, in the globin, immunoglobulin, histone and heat-shock genes and in various families of repetitive DNA. It is still an open question how the sequence consistency of human mtDNA is maintained. Whilst the mtDNA of different individuals may vary by almost 0.5 %, in any one individual only one nucleotide substitution can be detected in a total of 49 kb [280].

Two mechanisms are mainly responsible for the intrachromosomal recombination and horizontal evolution in multi-gene families, namely unequal crossing-over and gene conversion (Fig. 4.4). Both involve the mismatching of DNA sequences, for which repetitive sequences are usually responsible [234]. For effective gene conversion, the homologous sequence must be at least 200 bp long [249]. In **crossing-over**, chromatid segments are exchanged during meiosis, or also during mitosis if mitotic chromosome pairing takes

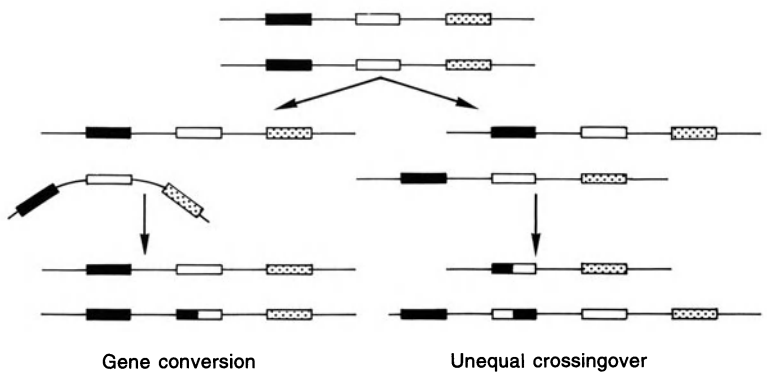


Fig. 4.4. Comparison between gene conversion and unequal crossing-over. In contrast to unequal crossing-over, the number of genes is not affected during gene conversion

place; this results in changes in the gene copy number or length. **Gene conversion** can occur during both meiosis and mitosis. It begins with the formation of a heteroduplex of DNA single strands with different sequences; this is then transformed into a homoduplex with completely complementary strands by a repair process (mismatch repair). Gene conversion is thus the replacement of one DNA segment by another; it can either equalize different sequences or create new sequence combinations. In contrast to unequal crossing-over, gene conversion is often one-sided; the number and length of the genes always remains constant [449].

The mechanisms described above are effective not only in maintaining the sequence similarity between the members of a **multi-gene family**, which arose by duplication of a common ancestral gene, but also in promoting the spread of variants through the gene family. Models of population genetics allow predictions about the rate with which mutations spread in a gene family and become fixed in the population [306, 426]. The spread of a variant in a gene cluster consisting of 200 copies requires approximately 10^3–10^4 substitutions; for example, the sequence agreement of about 85 % observed in the major histocompatability complex (MHC) families of man and the mouse would be reached with a substitution frequency of 10^{-5}–10^{-6} per generation. The distribution of a gene family between several chromosomes has only a minimal influence on its horizontal evolution, so long as the conversion rate between the genes on different chromosomes is not very low or many chromosomes are involved. Isolated members of a multi-gene family ("orphons") are, on the other hand, not subject to horizontal evolution and often show significant sequence differences to other members of the family.

Evidence for gene conversion is to be found in locally restricted sequence agreement in neighbouring DNA segments and can be detected with simple statistical tests [362]. Only a few of the many available examples can be mentioned here. The observation that the mRNAs of two human cytochromes P-450 are identical in the 5′ half but different at 36 positions in the 3′ half strongly suggests a relatively recent conversion event [13]. Particularly clear examples of gene conversion are found between the duplicated α-globin genes; thus, in both loci of the horse one finds the same polymorphism 24-Tyr/Phe [67]; in a human subject, the same mutation of 16-Lys to Glu was found at both loci [247]. During the sequencing of three γ-globin genes from the same individual, it was found that the 5′ region of the $^{A}\gamma$-gene on one chromosome had become like the neighbouring $^{G}\gamma$-gene through gene conversion and was therefore very different to the $^{A}\gamma$-gene on the other chromosome (see Fig. 4.6 a). It is possible that the $(TG)_n$ sequence in the second intron of both γ-genes was the site of the non-homologous base pairing and the starting point for the gene conversion [382, 383]. Frequent gene conversions have resulted in sequence equalization at the chorion protein locus of the silkworm *Bombyx mori*. An increase in sequence agreement from the inside to the outside was observed in each gene pair of the late Hc proteins (see Fig. 11.8, p. 393); here, the "hot spot" for gene conversion is also apparently in the spacer between the gene pairs [49, 104]. The increase in nucleotide differences towards the 3′ end of the two human α-globin genes further suggests that the starting point for gene conversion lies near the 5′ terminus [170].

The **number of copies in a gene family** (multiplicity) can change rapidly during the course of evolution. Even closely related species may show differences in multiplicity: e.g. whilst there are 24 000 5S rRNA genes in *Xenopus laevis*, in *X. borealis* there are only 9000 (see Table 2.4, p. 46). Changes in multiplicity caused by genetic events

can be very rapidly corrected. Thus, in a mutant of *Drosophila melanogaster* which possessed only 190 histone genes because of a heterozygous chromosome mutation, the normal count of 275 genes was almost fully restored in eight generations [64]. In view of the effectiveness of gene conversion, the question arises of which mechanism hinders the equalization of sequences in neighbouring genes of certain gene families, e.g. the globin genes. The answer to this problem may lie in the fact that the formation of the heteroduplex required for gene conversion is dependent upon the sequence similarity of non-coding sequences, e.g. the introns. Divergent evolution of these sequences, e.g. the frequent insertion of Alu sequences in mammalian globin gene clusters, reduces the probability of gene conversion and allows the development and maintenance of adaptive differences between the sequences [39, 365].

As long as sufficient genes of a gene family remain functional, individual genes may be inactivated by negative mutation without causing an increase in selection pressure. Non-expressed members of a gene family (pseudogenes) may also arise by the insertion into the genome of DNA copies of an RNA (processed or retropseudogenes; p. 25). In contrast to active genes, **pseudogenes** display many genetic alterations, including frame-shift and nonsense mutations, changes in splicing signals, mutations of the initiation codon, and small insertions or deletions. Pseudogenes have the highest known rates of substitution, e.g. in the case of the globin genes it is approximately 1.9 times higher than the rate of synonymous substitution. Furthermore, they approach most closely the frequently expressed idea of functionless DNA ("junk DNA") [210, 212]. However, in many pseudogenes synonymous substitutions are more frequent than amino acid exchanging substitutions, and they display the same characteristic difference in the substitution rate of non-coding regions as do active genes, albeit to a lesser degree (see Table 4.12). This might be explained by such genes becoming inactive only after a certain period of evolution [210, 212].

In contrast to the pseudogenes of other gene families, the many pseudogenes of the immunoglobulin-V_H family differ only to a minor degree from the active genes, and mostly contain only one inactivating mutation. This is undoubtedly due to continuous sequence correction by horizontal evolution. However, there are also individual pseudogenes in this family with many inactivating mutations [73]. The processes of horizontal evolution also affect the pseudogenes of multi-gene families. They can, for example, be duplicated. The known examples of this phenomenon include the bovine β-globin cluster (p. 263), the mouse urinary protein (MUP) family, and the human GAPDH genes [10, 132]. Because pseudogenes are mutated more often by deletion than by insertion, these genes become shorter during evolution [149]. Pseudogenes also represent a potentially useful reserve of genetic information; there is evidence for the formation of hybrid genes by sequence transfer from pseudo- to active genes, for example in certain allotypes of the mouse IgCγ gene and the rabbit IgCx gene [209].

4.2.6 Polyploidization

A doubling of the chromosome number polyploidization and thus the total genetic information is apparently an important mechanism of evolution [304, 307]. The event can happen in two ways: (1) the omission of meiotic reduction produces diploid gametes which fuse to give tetraploid zygotes **(autopolyploidy)**; (2) hybridization of two species results in the failure of chromosome pairing and reduction in the next meiosis **(allopolyploidy)**. Problems arise here in obligatory out-crossing organisms, and thus polyploidy occurs less frequently in animals than in flowering plants, where self-fertilization is quite common. In the animal kingdom, polyploidy is found in particular in the hermaphroditic oligochaetes, turbellarians and gastropods, and in parthenogenetic species of e.g. the Coleoptera, Lepidoptera and shrimps. Animals with the XY/XX mechanism of genotypic sex determination have special problems with polyploidy; for example, an egg with two sets of autosomes and two X-chromosomes (2AXX egg) together with a 2AXY sperm would form a 4AXXXY zygote with a disturbed XY balance; for this reason, polyploidy is quite often found in vertebrates such as fish and amphibians, which have less stringent mechanisms for sex determination. It is almost completely absent in reptiles, birds and mammals (Table 4.6). Polyploidization in the cartilaginous fish is difficult to demonstrate because of the basically high chromosome numbers; however, the existence of tetra- and octoploid species may be concluded from the bimodal distribution of DNA contents and renaturation kinetics [308].

Polyploidization produces an excess of genetic information that must be processed during evolution and there are basically two possibilities:

Table 4.6. Examples of polyploid (mostly tetraploid) vertebrates [33, 98, 146, 241, 304, 308]

Agnatha	Petromyzontidae
Chondrichthyes	Diploid: *Raja* sp. Tetraploid: *Torpedo marmorata, T. ocellata* Octaploid: *Oxynotus centrina*
	Teleostei
Catastomidae	All species (polyploidizations occurred ca. 50 million years ago)
Cyprinidae	*Cyprinus carpio, Carassius auratus, Barbus barbus* (*B. fascitatus* and *B. tetrazona* are diploid!)
Salmonidae	All species (polyploidization occurred ca. 100 million years ago)
Cobitidae	*Misgurnis fossilis*
Amphibia: Anura	
Hylidae	*Hyla versicolor* (diploid sibling species *H. chrysocelis* partly sympatric)
Leptodactylidae	
Xenopidae	Diploid: *Xenopus tropicalis* Tetraploid: *X. laevis, X. borealis, X. muelleri, X. clivii, X. fraseri, X. epitropicalis* Octaploid: *X. amieti, X. boumbaensis, X. vittei, X. vestitus, X. andrei* Dodecaploid: *X. ruwenzoriensis*
Pipidae	
Ranidae	*Pyxicephalus delalandi*[a], *Discoglossus occipitalis*[a]
Bufonidae	African *Bufo* sp.[a]
Cerato-phrydidae	Diploid: *Odontophrynus cultriceps* Tetraploid: *O. americanus* Octaploid: *Ceratophrys dorsata*

[a] Both diploid and tetraploid populations exist in these species.

1. Individual genes are silenced by mutation to non-functional null alleles or by repression of their expression.
2. The duplicated genes assume different structural and functional characters by divergent evolution and mostly become development and/or tissue-specific in their expression.

In both cases, the functional arrangement of a diploid genome is re-established, i.e. there is a **diploidization** [98]. The DNA content per cell has increased markedly by repeated duplication during the evolution of organisms; from this point of view, one can consider all eukaryotes to be diploidized polyploids. The course of diploidization can be investigated by the method of enzyme electrophoresis. Suitable for this purpose are, on the one hand, various fish families in which polyploidization occurred about 50–100 million years ago and, on the other hand, certain amphibians which, as the only vertebrate group, have both diploid and tetraploid species in the same genus, sometimes even in the same habitat (Table 4.6). In general, the evolution of genes duplicated by polyploidization happens as follows: at the beginning, the gene products are identical or very similar; later, the proteins display different electrophoretic mobilities but the duplicated loci are expressed similarly in all tissues. Loss of expression often occurs in this phase. Species- and tissue-specific expression arises during the next phase. Finally, each locus attains a particular pattern of expression which, from then on, is evolutionarily quite conservative. The most detailed investigations of diploidization have been carried out with fish from the family Catostomidae. It was found that each of the 20 duplicated loci studied was expressed in a diploid fashion in at least one of the 47 species examined. The average degree of diploidization for all species was 50 %, and for some species diploidization was as high as 70 % [116, 414]. The rate of diploidization, calculated from the time of the polyploidization and the degree of diploidization, is somewhat less than that predicted from population genetics models [430]. Evolution leading to diversification of the duplicated genes could be demonstrated in several species by comparing activities in ten different tissues. Of 115 duplicated loci, only 17 showed an activity ratio of 1:1 in all tissues, and the rest displayed tissue-specific differences; in 24 cases, one of the two loci predominated in any one tissue [116]. In addition to diploidization, the rRNA genes of several tetraploid fish present a further mechanism by which the amount of gene products can be regulated. Part of the 28S and 18S RNA molecules possess formamide-sensitive cleavage sites (p. 49). In the family Cyprinidae, in which both tetraploid and diploid species are found, the proportion of these unstable rRNAs is much higher in the tetraploids than in the diploids; in tetraploid carp, 90 % of the 28S RNA and approximately 50 % of the 18S RNA is destroyed by formamide treatment. Only the germline rRNA of this species is free of such cleavage points. Due to the higher lability of the rRNA, in vitro protein synthesis decreases more rapidly with ribosomes from the liver of tetraploid fish than of diploid fish, or from the oocytes of the carp [98].

4.2.7 Gene Transfer Between Species and "Molecular Lamarckism"

If the discovery of the dynamic state of the molecules of inheritance and the mobility of DNA sequences in the genome radically changed our ideas about the molecular mechanisms of evolution, then the **transfer of genes across species boundaries** has opened a whole new dimension in evolutionary theory. Gene transfer between prokaryotes is well known. However, there are also examples of the transfer of genetic information between prokaryotes and eukaryotes. For example, a large plasmid is responsible for the induction of crown-gall tumours in plants by *Agrobacterium tumefaciens*, and it is, in part, incorporated into the genome of the plant and expressed. This DNA includes genes for the synthesis of specific substances (opines) which are otherwise never produced in the cells of higher plants. The prokaryote *Progenitor cryptocides* secretes in culture a protein that is apparently encoded by the gene for human chorion gonadotropin. These two cases are both pathological; only systematic DNA sequence comparisons will show to what extent gene transfer between species is important as a normal mechanism of evolution of the eukaryotes. Artifical gene transfer has already been achieved in a whole series of eukaryotic cells. As the enzymes and gene vectors used in these experiments are all of natural origin, what the genetic engineer has carried out should also be possible in evolution [122].

The following cases have been offered as evidence of gene transfer during evolution.

1. The symbiotic photobacterium *Photobacter leiognathi*, regularly found in the teleost *Leiognathus splendens*, is one of a few prokaryotes known to possess a Cu, Zn superoxide dismutase. This is the typical cytosolic enzyme of the eukaryotes; only the Mn and Fe superoxide dismutases are found in other prokaryotes (p. 706).
2. *Escherichia coli* possesses two GAPDH genes, only of which is a typical prokaryotic gene; the other shows great similarity to the GAPDH genes of the eukaryotes. *E. coli* apparently took over this second gene from a eukaryote [97].
3. The repetitive DNA family 2108 is very similar in the sea urchins *Strongylocentrous purpuratus, Tripneustes gratilla* and others, although these species separated at least 200 million years ago and display large differences in single-copy DNA.
4. The histone gene h19 of the North Atlantic sea urchin *Psammechinus miliaris* differs by only 1.3 % from that of the Pacific species *Strongylocentrous miliaris*, although these species separated at least 65 million years ago. Thoughts of gene transfer are supported by the very similar sequence found in the species *Strongylocentrous droebachensis*, which colonized the North Atlantic only about 5 million years ago.
5. The fact that haemoglobin is widely found in animals but is restricted to a few groups of plants suggests gene transfer, perhaps with a virus as the vector.

At the beginning of the 1980s, there was controversy, under the heading of **"Lamarckism"**, about whether the immunotolerance conferred upon new-born mice by inoculation of spleen or bone-marrow cells was inherited by their progeny. Repetition of the experiments led, in all cases, to negative results [37]. In fact, confirmation of the results would in no way have been support for Lamarckism against Darwinism. In the view of Lamarck, the assumed process would certainly not have been a case of an acquired phenotypic character having a shaping effect on the genotype. It would have been much more a question merely of gene transfer between somatic cells and the germline, a process for which there are no precedents but which, from present-day knowledge, can certainly not be excluded [122]. The enhancement of the mutation rate in stationary-phase bacterial cultures by a potential substrate, and thus the appearance of enzymes required for its assimilation, has also been interpreted as the inheritance of an acquired character. This also conforms with neo-Darwinite ideas in that the single-stranded DNA arising during transcription is more easily mutated, and therefore substrate induction can quite possibly increase the mutation rate [85].

4.2.8 Adaptive and Innovative Protein Evolution

Even though many amino acid substitutions occurring during evolution may be selectively neutral, adaptive evolution of proteins does certainly occur. This book is full of examples of the adaptation of proteins to the specific requirements of individual animals. The evolution of single proteins relies upon the **optimization of dif-**

ferent characters according to the biological function: for cellular enzymes this is, for example, the maximal flux at the substrate concentrations occurring in the cells; for regulatory enzymes it is the high sensitivity to changing conditions; for digestive enzymes it is the constant turnover rates by substrate saturation; and for detoxifying enzymes it is the best possible protection of the organism as a result of reduced substrate specificity [48, 328]. To answer the question of how far cellular enzymes have already been "perfected" during the course of evolution, an efficiency coefficient E_f has been defined, and this reaches the value of 1 when the flux is limited by only the diffusion processes of substrate binding and product release. In these terms, there are many enzymes that may be looked upon as perfect catalysts. However, the catalytic potential of an enzyme in vivo is so dependent upon its molecular environment, for example, due to the binding to multienzyme complexes or cellular structures or to metabolic compartmentalization, that kinetic parameters defined in vitro are of doubtful value for assessing the degree of perfection reached in evolution. The interaction of enzyme molecules with other proteins is made easier by the almost completely spherical form of the enzymes; this may also be of importance in the evolution of enzymes.

The evolutionary **adaptation of proteins to changed conditions** involves the effects of positive selection. For this there is usually no requirement for new, "adaptive", genetic changes; molecular adaptation can take account of the variability already present in the population, in that previously neutral variants may acquire a positive selection value under the new conditions [210, 212]. The functional characters of protein are specified by, in particular, their spatial structure. Compared with the large number of possible amino acid sequences, there is a limited number of architectural classes. Therefore, far-reaching changes in amino acid sequence can occur during protein evolution without there being any change in spatial structure, which is very conservative. The most investigated example of this is the globin family, whose spatial structures are very similar despite the existence of up to 84 % sequence difference (see Fig. 7.4, p. 252); further examples include the serine proteases, the pepsin-like proteases and the enzyme pair, avian lysozyme/mammalian α-lactalbumin. In some cases, one finds similar spatial structure without significant sequence similarity, e.g. comparing globins with cytochrome b_5, or avian with phage lysosomes. The question remains whether these are homologous proteins, whose sequence similarities have been lost during evolution, or proteins of similar spatial structure, resulting from convergent evolution.

In the course of the evolution of organisms, **proteins with completely novel functions** have continuously appeared. A wide range of new types of protein were required when the vertebrates appeared approximately 500 million years ago and a large variety of different life forms with unique organ systems came into being. Research into molecular homology has made it possible to describe the origin and evolution of several typical vertebrate proteins. This is true, for example, for proteins of the eye lens: the exons 2 and 3 of the α-crystallin gene are related to the very ancient genes for the heat-shock proteins; β-crystallin and γ-crystallin possibly developed from Ca^{2+}-binding proteins; the δ-crystallin is homologous with arginosuccinate lyase which, because of its central role in arginine biosynthesis and purine metabolism, is ubiquitous and the ε-crystallin is simply an altered lactate dehydrogenase (p. 367).

The α-lactalbumin of mammals, the regulatory chain of the dimeric lactose synthase, is related to the lysozyme from the chicken egg. During its evolution from the carbohydrate-hydrolysing enzyme, the α-lactalbumin has retained its sugar-binding capability but lost the hydrolytic property. The super-family of the serine proteases includes not only the various protein components of blood-clotting, fibrinolysis and the complement system but also haptoglobin, which binds haemoglobin from degraded erythrocytes and transports it to the liver. Here also, the newly appearing protein has retained the ability for specific binding to polypeptides at the expense of the hydrolytic property. Caeruloplasmin, which serves to transport copper in the blood plasma, is homologous to the blood-clotting components fibrinogen, factor V and factor VII. The apparently quite old proteinase inhibitor α_2M is related not only to a pregnancy protein but also to the complement components C3, C4 and C5. There is a completely unexpected homology over a stretch of 101 amino acids between the linker protein in the proteoglycan complex of cartilage and the T-cell receptor; in this case, a component of the immune system appears to have arisen by modification of a cell-matrix protein [36, 94]. Even today, enzymes with new substrate or reaction specificities can arise; an example of this is the dichlorodiphenyltrichloroethane (DDT) dechlo-

rinase of insects. The evolution of new enzymes can be observed in bacterial experiments; for example, during the culture of β-galactosidase (lacZ)-deficient *E. coli* mutants on lactose-containing medium, a new enzyme, known as "evolved β-galactosidase" (Ebg) appears. All that is needed here are a few mutations in the gene ebg°, a weak catalytic β-galactosidase found in the wild type, and in the gene egb^R of the associated repressor [239].

The de novo synthesis of a longer DNA sequence coding for a functional protein would appear to be excluded on the grounds of probability. It is also difficult to imagine the appearance of a completely novel protein by the diversifying evolution of a duplicated gene. Gradual alteration of a DNA sequence could never really produce a functional structural gene. Frame-shift mutations and other lethal changes would accumulate in a gene that has become inactive such that as a whole it could hardly become active again. Thus, at the molecular level the evolution of novelty may come mainly by a process described by Jacob as "tinkering" [98]. By this it is meant that evolution, unlike technology, does not build new structure from scratch but achieves *novelty by the modification and recombination of what is already available*, in the same way that a father constructs a toy car for his son from the wheels of an old pushchair, an orange box and other bits and pieces. The molecular mechanisms of tinkering include, in particular, the processes of gene rearrangement, i.e. the insertion or deletion of longer sequences, duplication, transposition and fusion of gene segments (exon shuffling), and also perhaps gene transfer across species boundaries.

4.2.9 Molecular Mechanisms in the Evolution of Complex Characters

It has been shown in many investigations that there is no strong correlation between the rate and extent of molecular evolution and the evolution of morphological and other complex characters of the phenotype. The use of molecular data in the construction of species' ancestral trees gives, in many cases, branching schemes similar to those based on morphological characters, although the lengths of the branches, which indicate the extent of the changes, are usually quite different. Evolutionary changes occur, on the one hand, within a line of evolution (anagenesis) and, on the other hand, at the separation of the lines (cladogenesis or speciation). It is controversial whether the evolutionary alteration of morphological characters predominantly occurs anagenetically (gradualism) or during speciation (punctuationalism) [72, 107, 301]. The evolution of proteins is, in any case, not accelerated by the formation of species; the genetic distances determined by enzyme electrophoresis are not larger in species-rich groups of animals than in those with few species [14, 16].

The **rate of phenotypic evolution** differs quite markedly between and within different lines of evolution; one extreme is represented by species referred to as "living fossils", whose morphological characters have remained constant over many millions of years [106]. Such species include, for example, the brachiopod genus *Lingula*, which looked exactly the same in the Silurian 400 million years ago as it does today; the horseshoe crab *Limulus* (Xiphosura); or, amongst the vertebrates, the lizard *Sphenodon*, the crocodile and the opossum, all of which have hardly changed since the Jurassic. In contrast to their morphological evolution, the molecular evolution of the living fossils has in no way been retarded; e.g. the genetic variability of natural populations of *Limulus* is not less than that in other animals (see Table 4.8). So-called sibling species are hardly distinguishable morphologically, but at a molecular level they are no less different than other closely related species (see Fig. 4.18).

There are, however, also groups of species which show an especially rapid morphological evolution and whose molecular similarity is much greater than the morphological similarity. This is true, for example, of the Hawaiian *Drosophila* species *D. silvestris* and *D. heteroneura*, which are easily distinguishable morphologically but have only a small genetic distance (D) of 0.063 [370]. A similarly low molecular distance is shown by the approximately 300 species of the chequered perch (Cichlidae), which are found only in the African Malawi lake and are apparently the result of very rapid speciation [219], and the five North American species of toothed carp, belonging to the genus *Cyprinodon* [415]. In Israel, the blind mouse, *Spalax ehrenbergi*, has several forms with different chromosome numbers (2n = 52, 54 or 60) and these forms are sexually isolated due to behavioural characteristics, but they have only a very low genetic distance of 0.035 [366].

Two particularly convincing **examples for the low correlation between molecular and morphological differences** are the comparison between the anurans and the placental mammals,

and the comparison between man and the chimpanzee. The anurans, i.e. the frogs and toads, are a morphologically uniform group of 3000 species which arose about 159 million years ago and which can all be placed in the same order. The "placental mammals" with an age of about 75 million years, include about 4600 species, amongst which are such diverse forms as cats, humans, bats and whales; they are, therefore, distributed over about 16–20 orders. Although, as can be seen from the above, the organismic evolution of the two groups has progressed at very different rates, the rate of molecular evolution, as measured by, for example, the immunological distances of the serum albumins or the mtDNA sequences, appears not to be significantly different. Frog species which are similar enough to be placed in the same genus may differ at the molecular level to the same degree as the bat and the whale [123, 205, 440]. Man and the chimpanzee differ in molecular characteristics more than sibling species of other animal groups (Table 4.7); however, on the basis of their morphological differences they are placed in different families (Hominidae and Pongidae) [115, 214, 440].

Table 4.7. A molecular comparison of man and the chimpanzee [176, 214, 361]

Amino acid sequences:
Fibrinopeptide A and B; cytochrome c; and α-, β- and γ-globin chains are identical
A single difference in myoglobin and the δ-globin chain
Immunological distance:
Lysozyme = 0, carboanhydrase = 3, serum albumin = 6, serum transferrin = 8
Electrophoretic data:
For 31 intracellular and 13 extracellular proteins there are, on average, 2.4 amino acid exchanges per protein with an average length of 293 amino acids, i.e. a mean sequence difference of 0.8 %
DNA data:
12S rRNA genes in mtDNA show 3.7 % sequence difference (compared with an average of 0.36 % between the mtDNA of different human individuals)
The β-globin gene differs in the region 1–1396 in 14–19 positions (compared with differences at 9 positions between human individuals)
According to different authors, unique DNA has ΔT_m values of 0.7–1.5 °C, corresponding to an average sequence difference of 1.1 % (in comparison *Xenopus laevis* and *X. borealis* show ΔT_m = 12 °C, i.e. 12 % sequence difference)
Genetic distance:
For 44 loci, D = 0.62 (for a comparison, see Fig. 4.18; similar D values are found between sibling species of *Drosophila*; D = 1.76 for the species *Rana pipiens* and *R. corrugata*)

The **evolution of morphological characters**, right up to the appearance of completely new forms of organisms, cannot be explained simply on the basis of evolutionary changes in the amino acid sequences of proteins. Morphological characters appear to be dependent not so much upon the coding sequences of structural genes as upon the regulatory processes that determine the expression of the structural genes at a particular time during development and at a particular site in the embryo. That organic evolution mainly concerns changes in regulatory mechanisms for gene expression is shown by the formation of species hybrids. The success of hybridization between various species of the genus *Mus* correlates to the evolutionary distance as estimated from molecular data [34]. However, crosses between anuran species are possible over a 10-times larger immunological distance (of the albumins) than in the mammals [440]. Furthermore, in vitro fertilization experiments with eggs of the teleost *Micropterus salmoides floridanus* and sperms of other species showed that the incidence of non-hatching or deformed embryos was not always proportional to the genetic distance between the species [316]. The molecular basis of morphological evolution will only become accessible to research when it is known by what mechanism the molecules of the body substances are arranged into different species-specific, supramolecular structures during morphogenesis. Molecular embryology has recently made important progress, e.g. with the identification and analysis of homeotic genes, and will undoubtedly have further success. When the DNA sequences responsible for guiding morphogenesis are known in more detail, it should also become possible, by comparing different animal species, to obtain an insight into the molecular mechanisms of the evolution of morphological characters.

4.3 Protein and Nucleic Acid Polymorphism

All evolutionary processes are manifested by **changes in the gene stocks of populations**. Until the middle of the 1960s, the genetic structure of a population could be assessed only indirectly by examining phenotypes. As phenotypically recognizable spontaneous mutations are rare events, it was generally held that the individuals of a population are homozygous at almost all loci for

the "wild-type" gene, and heterozygous at only a few loci (about 0.1 %) for mutated alleles that, as a rule, would be lethal in the homozygous state. Only with the introduction in 1966 of enzyme electrophoresis and the subsequent development of other methods for analysing information-carrying macromolecules did population genetics gain direct access to the genome. A complete assessment of the genetic differences between individuals requires the comparison of gene sequences. However, the methods of DNA and protein sequencing require too much effort for comparative investigations of many individuals. Therefore, population genetics mainly makes use of the methods of protein electrophoresis and of DNA analysis by restriction endonuclease digestion. The relatively small amount of data obtained by the comparative sequencing of several individuals provides useful additional information. In 1966, the first enzyme electrophoretic investigations of *Drosophila pseudoobscura*, by Lewontin and Hubby, and of man, by Harris, revealed that the amount of genetic variability was so great that it was described as an "electrophoretic revolution". Using this and other methods, it has since been demonstrated that a marked **molecular polymorphism** exists in almost all natural populations [17, 21, 293, 294]. The characteristic allele frequency for each population remains constant over many generations, as has been shown, for example, in *Drosophila melanogaster* [61].

There are two groups of opposing theories on the origin of protein polymorphism: The "selectionistic" theories see polymorphism as the outcome of particular forms of selection (balancing selection) [17], whereas the "neutral" theories assume that a large proportion of the mutations and the resulting amino acid substitutions are selectively neutral, and only by chance (drift) do they spread through the population and become fixed. Although the controversy between "neutralism" and "selectionism" is by no means settled, the results of DNA-sequence analysis today leave no doubt that many molecular variants behave like neutral characters during evolution [210, 212].

4.3.1 Definitions and Concepts

Prakash, Lewontin and Hubby introduced in 1969 the term "allozyme" for allelic sequence variants of enzymes; the IUB nomenclature commission used the name "allelozyme"; linguistically "**alleloenzyme**" is to be preferred, or in general "**alleloproteins**". Allelic protein variants usually differ in only one or a few amino acids. Particularly large differences of seven amino acids are found between the allelic variants A and B of the haemoglobin-β chain of sheep and the types 1 and 3 alkaline phosphatases from human placenta [169]. With proteins possessing internal periodicities, allelic length variants are also possible. Strictly speaking, the terms alleloenzyme or alleloproteins should only be applied when evidence for allelic inheritance has been obtained from crossing experiments; however this condition is rarely fulfilled. It should also be remembered that the formation of certain post-translational modifications of proteins can, in some cases, be inherited as Mendelian characters (p. 79).

A protein locus is termed polymorphic when the proportion (frequency) of the most frequent allele is lower than a certain limit, which is usually given as 0.99 and sometimes as 0.95. For the quantitative description of protein polymorphism, the proportion P of polymorphic loci at all the examined loci is often given; however, P is greatly dependent upon the defined limit and the genetic constitution of the individual under investigation. Less influenced by this is the **heterozygosity** of a single locus h:

$$h = 1 - \Sigma x_i^2 \qquad (4.2)$$

where x_1 is the frequency of the allele i [115, 294]. Averaging h for all the investigated loci gives the mean heterozygosity, H. In an ideal bisexual population with random pairing, h corresponds to the proportion of animals that are heterozygous at a given locus, and H to the average proportion of heterozygous loci. In reality, the observed proportion of heterozygotes (H_{obs}) may deviate markedly from the expected value (H_{exp}). H_{obs} is heavily dependent upon mating behaviour; incest and self-fertilization lead to a reduction in H_{obs}. In haploid species, h_{obs} is zero by definition. Thus, only H_{exp} is suitable for the comparative assessment of protein polymorphism in different species, although this is not always accepted in the literature [148]. To define a protein polymorphism, the actual number of alleles n_e may be given:

$$n_e = 1/\Sigma x_1^2 \qquad (4.3)$$

The H values determined with enzyme electrophoresis differ greatly even between closely related species or different populations of a species (Table 4.8). H is seldom greater than 0.3; the

Table 4.8. Protein polymorphism in various animal species and groups as represented by the average heterozygosity H (p. 131) of at least 20 loci

Group	Species	H value	Reference
Bacteria	*Escherichia coli*	0.472	[294]
Nematoda	*Ascaris suum*	0.066	[233]
Phoronida	*Phoronopsis viridis*	0.094	[125]
Echinodermata	*Asterias forbesi*	0.041	[125]
	Nerchaster aciculosus	0.195	[125]
Mollusca	*Rumina decollata*	0.000	[331]
	Tridacna maxima	0.209	[55]
Chelicerata	*Limulus polyphemus*	0.064	[331]
Crustacea	*Homarus americanus*	0.040	[125]
	Euphausia superba	0.057	[125]
Insecta	*Gryllus integer*	0.145	[331]
	Otiorhynchus scaber	0.294	[125]
	Drosophila paulistorum	0.228	[294]
	D. melanogaster	0.194	[52]
	D. simulans	0.135	[52]
	Musca domestica	0.115	[30]
Agnatha	*Lampetra planeri*	0.076	[428]
Telostei	*Fundulus heteroclitus*	0.180	[331]
	Zoarces viviparus	0.090	[331]
	Mugil cephalus	0.071	[331]
	Gillichthys mirabilis	0.054	[125]
	Menidia menidia	0.053	[125]
	Sebastes caurinus	0.018	[125]
	Mylopharodon conocephalus	0.002	[125]
Amphibia	*Plethodon cinereus*	0.044	[125]
	Bufo viridis	0.147	[125]
Reptilia	*Anolis gingivinus*	0.100	[294]
	A. oculatus	0.050	[294]
	A. griseus	0.020	[294]
	Alligator missisipiensis	0.021	[294]
Aves	*Ardea herodias*	0.007	[154]
Mammalia	*Calomys musculinus*	0.201	[127]
	Rattus rattus	0.044	[125]
	Peromyscus maniculatus	0.128	[294]
	P. floridanus	0.062	[294]
	P. guardia	0.014	[294]
	P. interparietalis	0.000	[294]
	Myotis velifer	0.152	[125]
	Macaca fuscata	0.013	[294]
	Homo sapiens	0.143	[294]
188	Insect species[a]	0.107[a]	[151]
30	Hymenoptera species	0.036[a]	[151]
38	*Drosophila* species	0.157[a]	[341]
57	Invertebrate species	0.134[a]	[17]
31	Fish species	0.058[a]	[341]
3	Amphibian species	0.105[a]	[341]
9	Reptilian species	0.043[a]	[341]
3	Avian species	0.031[a]	[341]
25	Mammalian species	0.039[a]	[341]
68	Vertebrate species	0.060[a]	[17]

[a] Mean values.

averages for all investigated animal species are H = 0.10 and P = 0.35. The probability that two individuals carry identical alleles at z loci is given by:

$$W = (1 - H)^z \qquad (4.4)$$

With H = 0.10 and $z = 10^4$, $W = 10^{-458}$; thus, with the exception of identical (single-egg) twins, no two individuals can ever be genetically identical and therefore the genetic concept of the "wild type" is questionable.

4.3.2 Methodological Problems in the Determination of Protein Polymorphism

Enzyme electrophoresis is carried out in the following way: the soluble protein of the whole animal or a single organ is fractionated using starch-gel or polyacrylamide-gel electrophoresis; enzyme bands with a certain substrate specificity are visualized on the electrophoregram by use of a specific staining reaction (Fig. 4.5). This method only detects amino acid substitutions that lead to a change in charge. It can be easily calculated that in proteins of average composition only about one-third of all possible amino acid substitutions lead to a change in charge (Table 4.3). Thus, individual protein bands on a gel may contain several different sequence variants with the same electrophoretic mobility; in 1975, these were termed "**electromorphs**" by King and Ohta. Various other methods are used to try to detect the ("cryptic") variants which are not separated by standard electrophoresis: complicated electrophoretic and chromatographic separation techniques, differentiation by differences in substrate specificity or sensitivity to high temperature, denaturing substances or inhibitors, and identification by immunological methods or peptide patterns.

The only direct evidence for the **heterogeneity of electromorphs** comes from sequence analysis. For example, there is a polymorphism 31-Ile/Val in carboanhydrase III of human erythrocytes which is not detectable by electrophoresis [171]. The sequential polyacrylamide-gel electrophoresis introduced in 1976 attains a much greater separation by systematic variation of gel concentration and pH. This method allowed the detection of, for example, 17 out of 20 anomalous human haemoglobins, compared with detection of only 8 using the standard technique [340]. A more recent method of isoelectric focusing, using very short pH gradients, e.g. of 7.20–7.55, can even separate proteins that differ by only one neutral amino acid [433]. Pre-incubation of the gel at high temperature can be used to distinguish between alleloenzymes of different temperature sensitivity.

The extent of the cryptic protein polymorphism revealed by these refined techniques varies greatly between different loci [17], but is in many cases very impressive. The application of isoelectric focusing increased the number of known alleloenzymes of phosphoglucomutase from human erythrocytes from 3 to 10 [209]; varying

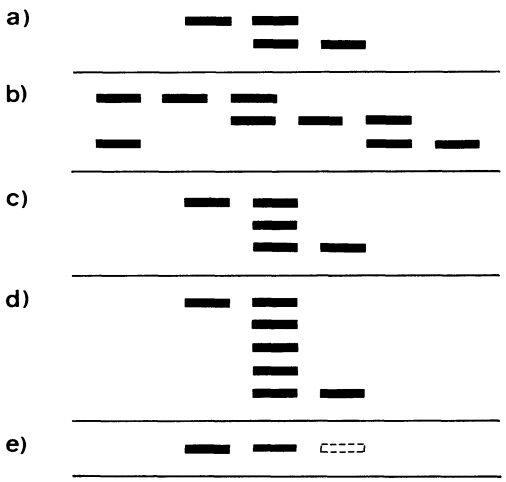


Fig. 4.5. Electrophoretic phenotype of the polymorphism of: **a** a monomeric protein, two codominant alleles; **b** a monomeric protein, three codominant alleles; **c** a dimeric protein, two codominant alleles; **d** a tetrameric protein, two codominant alleles; **e** a normal and a null allele [115]

the gel concentration in studies of 14 loci of the butterfly *Colias meadii* gave an increase from 40 to 103 [191]; sequential gel electrophoresis increased the number of distinct alleloenzymes at the esterase-5 locus of *Drosophila pseudoobscura* from 8 to 22–41 [204]. Heat inactivation experiments revealed eight variants within the two esterase-6 electromorphs of *Drosophila melanogaster*, and these could be identified as alleles by crossing experiments [69]. Use of the same method increased the number of known variants at 14 mouse loci from 27 to 53 [34]; in human erythrocytes the total number of "heat" alleles is in fact more than three times higher than the number of variants separable by electrophoresis [279].

The combination of several methods reveals extreme polymorphism at some loci. For example, using various electrophoresis methods combined with heat inactivation, 35 variants of serum esterase-1 were described in the mouse *Peromyscus maniculatus* [9]. Standard electrophoresis indicated 8 variants of xanthine dehydrogenase in *Drosophila pseudoobscura* and this estimate was increased to 27 variants by sequential electrophoresis and further to 37 using heat inactivation; thus, the apparent heterozygosity at this locus increased from 0.44 to 0.63. Similar results were obtained with *D. melanogaster* [46]. If the effective number of alleles is increased by only 20% by taking all the cryptic polymorphisms into account, the mean H values (Table 4.8) increase from 0.134 to 0.28 for the invertebrates, and from

0.06 to 0.21 for the vertebrates, i.e. there is a two- to threefold higher proportion of heterozygous loci [17].

Past investigations of protein polymorphism have almost all been carried out using the same three dozen enzymes which are easily stained and detected on gels (Table 4.9). The mean heterozygosity H, however, should really be determined from a random sample of all protein loci. For this purpose one can use the method of **2-D electrophoresis**; this was introduced in 1975 and can distinguish up to 1000 identifiable protein spots. The 2-D method almost always indicates much less polymorphism than does 1-D electrophoresis. Studies of human fibroblasts, lymphocytes and brain and kidney cells [352], of various organs of both laboratory and wild *Mus musculus* [217], and of *Drosophila* [379] gave H values of 0–0.02. Only in the case of 2-D studies of human plasma proteins was an H value (0.062) found to agree with the 1-D electrophoresis H value (0.063) from 104 human loci [352].

4.3.3 Dependence of Protein Polymorphism on Protein Type

The extent of polymorphism differs greatly between different proteins (Table 4.9). In connection with the controversy over neutralism and selectionism, there have been many investigations of which protein properties determine the extent of polymorphism. According to the neutral theories, polymorphism should increase with increasing molecular mass, because the probability of neutral amino acid substitutions increases with the number of amino acids; such a tendency was confirmed in studies involving a large amount of data [294]. Polymorphism is dependent upon the **quaternary structure;** it is larger for monomers than for di- or tetrameric enzyme proteins (Table 4.10). Neutral theories explain this relationship by restricted variability in the regions of contact. It should be noted, however, that amongst the proteins with marked polymorphism are several multimeric enzymes,

Table 4.9. The heterozygosity (H values) of various proteins: (a) considering all *Drosophila* species examined; (b) considering all invertebrate and vertebrate species examined [331]

Protein	Heterozygosity		Function type[a]
	a	b	
Xanthine dehydrogenase (XDH)	0.364	0.208	R
Phosphoglucose isomerase (PGI)	0.353	0.134	R
Esterase (EST)	0.341	0.277	U
Acetaldehyde oxidase (AO)	0.321	–	R
Amylase (AMY)	0.321	–	N
Acid phosphatase (ACPH)	0.259	0.224	U
Peptidases (PEP, LAP)	0.235	0.192	U
Adenylate kinase (ADKIN)	0.215	0.136	R
Phosphoglucomutase (PGM)	0.199	0.170	R
Glucose-6-phosphate dehydrogenase (G6PDH)	0.168	0.121	N
Alcohol dehydrogenase (ADH)	0.152	0.140	R
Isocitrate dehydrogenase (IDH)	0.127	0.082	N
Malate enzyme (ME)	0.117	0.131	R
Hexokinase (HK)	0.077	0.087	R
Aspartate transaminase (GOT)	0.075	0.057	N
Non-enzymatic proteins	0.071	0.066	
Superoxide dismutase (SOD)	0.070	0.080	U
Fumarase (FUM)	0.050	0.041	N
Malate dehydrogenase (MDH)	0.037	0.083	N
6-Phosphogluconate dehydrogenase (6PGDH)	0.037	0.039	N
α-Glyceraldehyde-3-phosphate dehydrogenase (α-GPDH)	0.013	0.039	N (insects) R (others)
Triosephosphate isomerase (TIM)	0.012	0.054	N
Relatively unspecific enzymes	0.205	0.175	U
Specific regulatory enzymes	0.210	0.161	R
Specific nonregulatory enzymes	0.086	0.073	N

[a] Function type according to Johnson: U, relatively unspecific enzyme; R, specific regulatory enzyme; N, specific nonregulatory enzyme.

Table 4.10. Dependence of the mean heterozygosity (H value) on protein quaternary structure [432]

Quaternary structure	Heterozygosity	
	Vertebrates	Invertebrates
Monomers	0.113 ± 0.050	0.186 ± 0.028
Dimers	0.040 ± 0.006	0.124 ± 0.018
Tetramers	0.015 ± 0.006	0.067 ± 0.026

e.g. esterases, xanthine dehydrogenases and aldehyde oxidases.

Less well supported is the notion of a dependence of polymorphism on **protein function**. There are several pieces of evidence that non-enzymatic proteins are less polymorphic than enzymes. On the other hand, the larval storage protein lucilin of the fly *Lucilia cuprina* is one of the most variable proteins known (p. 192). In an attempt to relate function and polymorphism, Johnson (1971) suggested grouping enzymes into three classes, whose variability decreased in the following order: (1) relatively non-specific enzymes; (2) specific regulatory enzymes; and (3) specific non-regulatory enzymes. However, there are in reality large differences in heterozygosity within each of these three classes (Table 4.9). The frequently noted high polymorphism of esterases and other enzymes of low specificity is not based on the preference of different alleles for different substrates but on the lower structural conformity required for such low specificity.

The same enzyme may show different levels of polymorphism in different species. The insect α-glyceraldehydephosphate dehydrogenase shows, in general, little variability. In only 2 out of 175 *Drosophila* species could any polymorphism be found for this enzyme by use of the standard methods, and in only 4 out of 65 species could it be found in combination with heat inactivation [223]. Furthermore, in five *Notonecta* species the mean h value of the Gdh locus was lower than 0.01. In contrast, this locus was extremely variable in 11 species of the waterbug family Gerridae, with a mean h of 0.4; it is also very polymorphic in the lepidopteran genus *Colias* [454]. The xanthine dehydrogenase locus in different *Drosophila* species has h values between 0 and 0.8 [210].

4.3.4 Differences in Protein Polymorphism Between Different Animal Groups and Habitats

The extent of polymorphism (H value) may differ greatly even between very closely related species, as is shown in Table 4.8 for the genera *Drosophila, Anolis* and *Peromyscus*. Therefore, generalizations about **protein polymorphism in different animal groups** must take into account certain conditions. Protein polymorphism determined by standard methods is markedly less in the vertebrates (means of H = 0.06 and P = 0.25) than in the invertebrates (H = 0.134 and P = 0.47) [17]; within the vertebrates, the amphibians apparently have relatively high H values. In mammals, with the exception of man (see Table 4.8), H decreases with increasing body size [446]. According to the neutral theories these differences are explained by differences in population size, and according to the selectionist theories they are explained by local variation in selection conditions. Protein polymorphism appears not to be correlated with the **type of habitat;** no consistent differences are apparent between the tropics and temperate zones, or between the sea, freshwater and terrestrial biotopes. In comparison with most other insects, the Hymenoptera display significantly less protein polymorphism (Table 4.8). In the honey bee and other bee species there was at first no indication of any polymorphic loci; however, variability has since been shown at least for alcohol dehydrogenase [262]. The limited protein polymorphism of the Hymenoptera has been associated with the haploid-parthenogenetic reproduction of the males [148]; however, high H values of 0.121–0.187 have been estimated for four parasitic species with haploid males [185, 375]. Species of solitary Hymenoptera show significantly more distinct protein polymorphism compared with the higher, social species [148].

Self-fertilization should lead to a drastic reduction in polymorphism because sexual recombination, as a mechanism for the spread of mutations in a population, is missing. The self-fertilizing terrestrial snail *Partula gibba*, which is indigenous to Tahiti and other Pacific islands, has not a single polymorphic locus; in contrast, the non-self-fertilizing species of the genus *Partula* that share the same habitat have normal polymorphisms and H values of 0.13–0.17 [192]. There are considerable differences in the allele pattern between different populations of the self-fertilizing terrestrial snail *Rumina decollata* in southern France and North Africa; however, the individual popula-

tions are uniform and thus apparently founded by single animals. The populations in various southern states of the USA, on the other hand, are completely identical and probably originate from the same population as that in southern France [294].

Single-sexed reproduction by diploid **parthenogenesis** is found in various animal groups. Here, the progeny of a single individual are genetically identical; the populations consist of one or more clones of genetically identical individuals. Because there is no segregation or recombination, new alleles arising by mutation remain heterozygous. If a population changes from bisexual to parthenogenetic reproduction, the heterozygosity at first increases, but then decreases again in the course of evolution due to the accumulation of non-functional alleles. Occasional bisexual reproduction allows new alleles into the population. In island populations of the parthenogenetic Australian snail *Thiara ballonensis*, which has no bisexual neighbours, 41 electromorphs were found for 12 enzymes, each of the 13 examined populations displaying its own uniform pattern [396]. Protein polymorphism in populations which alternate cyclically between bi- and monosexual reproduction has been studied in detail in waterfleas in the genus *Daphnia* [281, 358] and in aphids [413].

4.3.5 Dependence of Protein Polymorphism on the Size and History of a Population

According to the neutral theories, the average heterozygosity should increase with the **size of the population**. However, the electrophoretically determined H value rarely exceeds 0.30, even in very large populations. On the one hand, this may be due to the occurrence of new variants which have the same electrophoretic mobilities as already existing electromorphs and therefore remain undetected, but, on the other hand, it is also predicted by certain of the neutral theories. In any case, reduced polymorphism is expected in a population founded by a few individuals **(founder effect)** or after the sudden reduction in size of a population **(bottleneck effect);** recovery of the H value will require many thousand generations [263]. Evidence for this theory is found in several cases where the history of particular populations is well known. For example, the population of the northern elephant seal *Mirounga angustirostris* decreased in 1890 to about 20 individuals; the effective population was in fact smaller than this because only certain males can participate in mating. In 1974, investigations of the several thousand individuals of the newly expanded population showed no polymorphisms at 24 blood-protein loci [35]; in contrast, in the southern elephant seal, which has never been so greatly decimated, 5 out of 18 loci were polymorphic. The relatively limited extent of protein polymorphism in the large mammals (with the exception of man) is perhaps the result of frequent bottleneck effects.

The founder effect explains many cases of unexpectedly low polymorphism [294]. The mosquito *Aedes aegypti* first reached Asia in the nineteenth century; the Asiatic population correspondingly shows significantly lower polymorphism (H = 0.086) than the African and American populations (H = 0.129–0.155) [401]. Animal populations from small islands, lakes or ponds, or from caves are usually less polymorphic than their parent populations. Thus, the island populations of the rodent *Peromyscus polionatus* on Santa Rose island, off the coast of Florida, has an H value of 0.018 compared with that of 0.067 for the mainland population [294]. Conspicuous in this regard is the relatively high polymorphism of the laboratory rat, the 300 strains and sub-strains of which were bred during the course of the century from a few wild animals; polymorphism was found in 9 of 22 loci examined [189].

4.3.6 Quantitative Genetic Variability

The quantitative polymorphism of proteins has been relatively seldom investigated compared with structural polymorphism, because the analysis using crossing experiments and concentration measurements of single protein species requires much more experimental effort. There is, however, no doubt that heritable variation of enzyme activities or the concentration of non-enzymic proteins is widespread, and that null variants are the exception [70]. Thus, for example, in 26 out of a total of 35 examined enzymes from the laboratory mouse, heritable activity differences were found in proportions up to 1 : 8.8 [47]. At 14 loci of two natural populations of *Drosophila melanogaster*, 58 variant alleles with lower or zero activities were recorded; with one exception, these produced no reduction in viability or fertility in the heterozygous state [50]. The quantitative variability of the alcohol dehydrogenase locus of *D. melanogaster* has been thoroughly examined [387]. The frequency of mutations which change

activity was investigated on 1000 pure lines arising from one *Drosophila* pair; within only 300 generations, alcohol dehydrogenase activities were found to vary by a factor of 10–20 [283]. In various populations of *D. melanogaster*, a mean frequency of null alleles of about 0.25 % was observed at 20 autosomal loci [224]. Whilst sequence variants of proteins are often selectively neutral, quantitative variants especially when they are homozygous, do influence fitness and are thus subject to selection. Appropriate investigations have been carried out, for example, on the α-amylase locus of *D. melanogaster*. Adults and larvae with higher amylase activities ($Amy^{4,6}$) have a significantly greater life expectancy than those with lower activities (Amy^{1}) when starch is the limiting nutrient [216].

Heritable differences in enzyme activity can arise by changes in either the catalytic properties or the synthesis or degradation, and therefore the concentration, of the enzyme protein. Changes in catalytic properties are always due to amino acid substitutions. These need not necessarily influence the active centre directly but can change the physicochemical properties of the protein and thereby affect enzyme function indirectly. For example, in various *Drosophila* null variants, the ability to form active heterodimers with normal subunits has been lost. The three alleles of glucose-6-phosphate dehydrogenase in *D. melanogaster* code for enzymes of different quaternary structure: Zw^{A} and Zw^{B}, which are active dimers and tetramers, and Zw^{101}, which is monomeric and of limited activity [50, 124].

Heritable changes in the rate of enzyme synthesis can involve very different types of mutation; for example, mutations in the signals for transcription and mRNA maturation, mutations in regulatory genes, or the influence of gene expression by rearrangements or translocations. The null allele adh^{nAH52} of *D. melanogaster* has been sequenced. It contains an insertion of 8 bp and an adjacent deletion of 2 bp in intron 2; as a result, the transcription rate is reduced to one-tenth of the normal [134]. The rate of synthesis of multiple genes can be influenced by the number of active genes. A frame-shift mutation, i.e. the addition or deletion of one or two nucleotides in a coding sequence, or a nonsense mutation, i.e. the production of a stop codon within a coding sequence, prevents formation of the normal enzyme protein; resulting anomalous polypeptides are usually immediately degraded proteolytically. Whether activity variants are based upon changes in enzyme concentration or catalytic properties can mostly be determined from immunological measurements of protein concentration; however, an amino acid substitution can also change immunological reactivity and catalytic activity. Unambiguous analysis requires a knowledge of the DNA sequence.

4.3.7 DNA Polymorphism

Protein comparison reveals only a part of the genetic variation present in a natural population; synonymous nucleotide substitutions and variants in non-coding sequences can only be recognized by analysis of the DNA itself. Protein polymorphism normally involves the substitution of single nucleotides in a coding sequence; DNA polymorphism can also result from deletions, insertions or duplications of longer sequences and from other types of rearrangements. The widely found length polymorphisms of non-coding sequences are usually based on a variable number of short repetitive sequences; examples are the rDNA of *D. melanogaster* [71], the chicken vitellogenin gene [329], and the human albumin, globin and insulin genes [309]. In multi-gene families, the number of copies can vary or parts of a cluster may be duplicated or missing; examples include the rDNA of many animals, human α- and ζ-globin genes, human immunoglobulin genes and the amylase genes of several rodents, and these are described in more detail elsewhere in this book. Polymorphism can also result from the exchange of partial sequences between the members of a multi-gene family, as described for the human γ-globin gene (Fig. 4.6a).

As DNA sequencing requires considerable effort, relatively few comparative investigations have been undertaken on sequences originating from the two homologous chromosomes of cells of an individual or from different individuals of a species or population. The most extensive data on the polymorphism of DNA sequences come from the **genes for human haemoglobin** [308, 443]. Two different nucleotides were found in each of eight positions at the β-globin locus, and in one position there were in fact three differences; the combination of these characters defines at least seven haplotypes that are distributed in human populations [361]. Two types (R and T) of the human δ-globin gene found in all populations differ from each other by 16 substitutions and two deletions, which are mainly in the non-coding 5′ non-translated region; the sequences differ by almost 1 % in this region of 1.7 kb [257]. Two

variants of the human insulin gene are also known to differ by four substitutions, one each in the two introns and two in the 3′ non-coding region [419]. There are also four variant nucleotides in the human 28S rRNA [139].

Amongst the genes of the invertebrates, the Adh **locus** of *Drosophila melanogaster* has been investigated most. Two electromorphs of this enzyme (ADH^S and ADH^F), found in all natural populations, consistently display the same amino acid difference 192-Lys/Thr. In addition to the nucleotide substitution 1490-A to C, responsible for this amino acid exchange, a further 42 polymorphic nucleotide positions have been found in five Adh^S and six Adh^F genes of five populations (Fig. 4.6b); apart from this, the first intron and the 3′ terminal region have length polymorphisms at six positions with differences between 1 and 37 bp [221]. The average nucleotide difference in the Adh locus of 11 *D. melanogaster* breeding lines was found to be 0.66% [395]. The gene sequences of all ten esterase-6 alleles of *D. melanogaster* contain 52 nucleotide differences, of which 16 produce amino acid exchanges but only 3 result in electrophoretically measurable charge differences [78]. The **histone gene cluster** is also extremely polymorphic. Five repeats from four individuals of the sea urchin *Strongylocentrotus purpuratus* contained no fewer than 59 polymorphic sites in the spacers, the non-translated regions of the mRNAs, and as synonymous substitutions in coding sequences [451].

Although the complete and unambiguous assessment of polymorphism requires the direct comparison of homologous DNA sequences, other less intensive methods are available for estimating DNA polymorphism; of importance for population genetics studies is **analysis using restriction endonucleases**. The reduction of the dissociation temperature T_m of artificial DNA hybrid molecules is often used in research into molecular relationships but has only occasionally been used for studies of genetic polymorphism. For example, heteroduplices, produced from the pooled single-copy DNA of different individuals of the sea urchin *S. purpuratus* and of the starfish

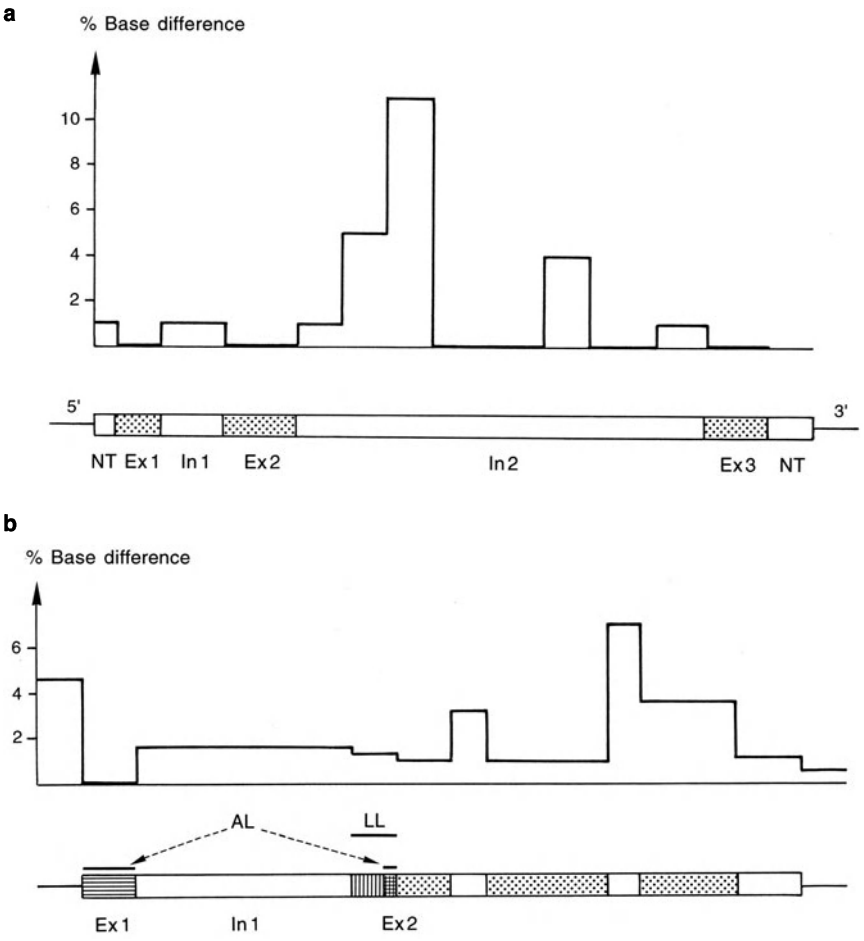


Fig. 4.6a, b. Two examples of the distribution of the sequence differences between allelic variants of the same gene between exons *(striped* or *stippled)*, introns *(blank)* and non-translated DNA sequences at the 5′ and 3′ ends of the mRNA *(blank)*. **a** $^{A}\gamma$-globin gene from the two homologous chromosomes of the same human individual [17, 382]; **b** the average difference between five alcohol dehydrogenase genes Adh^S and six Adh^F of *Drosophila melanogaster* [221]. *LL*, Larval leader sequence; *AL*, adult leader sequence

Pisaster ochraceus and compared with the reassociated DNA from one individual, showed a reduction in T_m of 4 and 5.3 °C, respectively; this corresponds to a sequence difference of 4–5 % between the individuals [151, 384]. Differences in the restriction endonuclease cleavage fragment pattern also suggest sequence differences. Such a restriction analysis can be carried out on isolated fragments of chromosomal DNA; mtDNA, which has exactly the right size of 15–19 kb and is easily obtainable in useful amount, is especially suitable for population genetics studies. A particular advantage of mtDNA is that it is purely maternally inherited and, therefore, does not undergo recombination. If, as in many species, the males are more mobile than the females, then the genetic flux in the population is dependent largely on the migrations of males; the populations are then much more clearly structured with regard to the mtDNA than the alleloproteins.

Various mathematical and statistical processes have been suggested for evaluating restriction analysis data [108, 182, 290]. However, a relationship between the polymorphism of a cleavage site and the total DNA sequence cannot be unambiguously defined because, for example, longer insertions or deletions and the rearrangement of sequences, as are frequently found in non-coding DNA regions, cannot be detected. According to Ewens [108], the mean heterozygosity per nucleotide can be calculated as follows. In n DNA sequences there may be m cleavage sites of which k are polymorphic. Given the employment of several enzymes with recognition sequences of four or six nucleotides, the mean heterozygosity per nucleotide is

$$\Theta = (k_4 + k_6 / [(8m_4 + 12m_6)\log_e n] \quad (4.5)$$

Nei [290] defined a nucleon diversity analogous to the heterozygosity [Eq. (4.2)].

$$h_n = 1 - \Sigma p^2_i \quad (4.6)$$

By nucleon is meant any defined DNA sequence for which i different cleavage patterns (nucleomorphs) are possible, each in the proportion p_i; this equation should take into account insertions and deletions.

Restriction analysis has produced evidence for **polymorphism of mtDNA** in man and various mammals, birds and reptiles (Table 4.11), as well as the invertebrates *Drosophila* and the crustacean *Panulirus argus* [270] (Table 4.11). The nucleotide substitutions in man, for example, are evenly distributed over the whole mtDNA, so that between 2 and over 40 allelic variants were detected for each of the 28 examined loci (13 proteins, 10 tRNAs, 2 rRNAs and 3 non-coding regions) in 145 individuals [434]. As well as differences in mtDNA sequences, there are also differences in mtDNA size; the relative degree of these two forms of mtDNA polymorphism differ greatly between species. No fewer than 37 size variants of mtDNA were found amongst 92 individuals of the lizard *Cnemidophorus tessaltus*, whereas the sequence difference of 0.06 % was extremely small [158]. In contrast, the mtDNA of rats and other rodents, shows considerable sequence variability but very little size difference [162]. In the insects *Drosophila melanogaster* and *Gryllus firmus*, heteroplasmy is often observed, i.e. the existence in the same individual of two mtDNA variants [158, 341]. If there is evidence of length polymorphism in the mtDNA, then not only the number but also the length of the cleavage fragments must be taken into account during restriction analysis [41, 162, 226].

Intraspecific **comparisons of chromosomal DNA sequences** by use of restriction analysis have been less often reported. It should also be noted here that the observed polymorphism can be due just as well to rearrangements of the DNA as to nucleotide substitutions. Seventeen different restriction patterns of the human β-globin locus are frequently observed, of which 14 are distributed worldwide and 3 are limited to certain populations. The sequence differences responsible are located predominantly in the non-coding flanking sequences and the introns [309]. In *D. melanogaster*, restriction analysis has been mainly concerned with the Adh locus. The aver-

Table 4.11. The percentage differences, estimated by restriction analysis, in mtDNA sequences between different individuals or populations (*) of the same species

Species	Difference (%)	Reference
Man	0.36	[158]
Chimpanzee *(Pan troglodytes)*	1.3	[117]
Dwarf chimpanzee *(Pan paniscus)*	1.0	[117]
Orang-utan *(Pongo pygmaeus)*	5.0	[117]
Gorilla *(Gorilla gorilla)*	0.55	[117]
Brown rat *(Rattus norvegicus)*	0.2–1.8	[41]
House rat *(R. rattus)*	0.2–9.6	[41]
Pocket mouse *(Geomys pinetis)*	0–0.047	[225]
Field mouse *(Microtus townsendii)**	0.1–0.9	[410]
Hamster mouse *(Peromyscus maniculatus)**	3.0–6.0	[226]
Great tit *(Parus major)*	0.19	[409]
Lizard *(Cnemidophorus tesselatus)*	0.06	[158]
Drosophila melanogaster	0.8–1.4	[158]

age sequence difference of about 0.6 %, found by use of this method, agrees with the value obtained by sequence comparison [221, 363]. Restriction analysis gave a value of 0.24 % for the variability of the heat-shock locus 87A7 of *Drosophila [40]*.

4.4 The Causes of Genetic Polymorphism

It was mainly as an explanation of the unexpectedly high protein polymorphism of natural populations that in 1968–1969, Kimura and also King and Jukes independently developed the theories of molecular evolution that became known as the "neutral theories" or "non-Darwinian evolution" [208, 213]. At the heart of these theories is the thesis that a large part of the nucleotide substitutions and resulting amino acid exchanges that occur during evolution are selectively neutral and become fixed in populations by random processes (drift). In opposition to these revolutionary ideas, various attempts were made to explain protein polymorphism as the result of some selection process. Controversy ensued between **"neutralism"** and **selectionism"**; at one time this took on an almost ideological dimension and the argument continues to this day [210, 212, 294, 297].

The arguments sometimes led to misunderstandings. Quite obviously, the proponents of the neutral theories did not claim that all mutation was selectively neutral. They did not deny the effects of selection, but held more to the idea that, under constant internal and external conditions, selection worked mainly to stabilize in that it eliminated variants deviating greatly from the norm. At the same time, the random fixation of neutral mutations brought about a high level of variability within a population and this actually then made possible adaptation to changing conditions. If the phenotypic optimum is shifted, as the result of a change in the environment, then some otherwise neutral variant may become advantageous; this spreads through the population because of positive selection and becomes "fixed". According to this concept, it is not necessary to assign special importance to specific "adaptive" mutations, although some may arise [210, 212].

4.4.1 Neutral Theories of Molecular Evolution

A significant advantage of the "neutral mutation – random drift" theory, proposed by Kimura in 1968, is the possibility to make quantitative statements about the relationship between the neutral mutation rate v_o and the effective population size N_e, on the one hand, and the rate of evolution and the extent of polymorphism on the other hand [210, 212]. The effective population N_e refers to an ideal population with random mating; for various reasons N_e in real populations is usually considerably smaller than the total number of available individuals. In addition, it is assumed that the population size will significantly change several times during the period required for the development of the polymorphism. N_e is, therefore, considered to have an abstract size in population genetics models, and the real value can be only roughly estimated [294]. A population of N_e individuals with a neutral mutation rate of v_o will accumulate $2N_ev_o$ neutral mutations per generation, and will have a fixation probability of $1/2N_e$. Thus, the fixation rate of neutral mutations is given by

$$k = 2N_ev_o/2N_e = v_o \qquad (4.7)$$

This is therefore independent of the population size and identical to the neutral mutation rate: a neutral mutation will become fixed in the population every $1/v_o$ generations. Definite fixation consequently requires about $4N_e$ generations and, in contrast, elimination an average of $2 \cdot \log_e(2N_e)$ generations (Fig. 4.7).

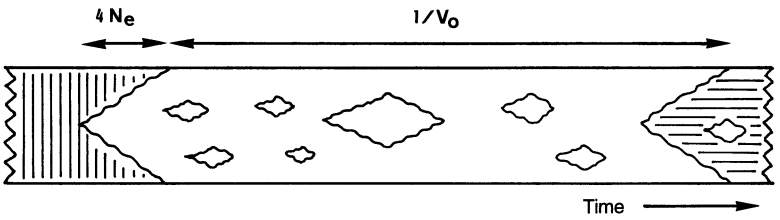


Fig. 4.7. The fate of neutral alleles in a population with an effective size of N_e and a rate of neutral mutation v_o. About 4 N_e generations intervene between the occurrence and the fixation of a neutral mutation; a new mutation is fixed every $1/v_o$ generations; the majority of neutral mutations are not fixed but are eliminated soon after their occurrence [209]

Only a certain proportion p_n of all mutation is selectively neutral. P_n may be estimated from the number and frequency of rare alleles, i.e. those whose frequency is lower than 0.01. Using this method, p_n for various enzymes and other proteins of both vertebrates and invertebrates has been estimated as 0.14 ± 0.06 [211]. The value of p_n can be used to calculate the rate of occurrence and fixation of neutral mutations from the total mutation rate v_T,

$$v_o = k = v_T p_n \qquad (4.8)$$

A neutral mutation is one that has no significant effect on the fitness of an organism under a given set of selection conditions. In contrast to the classical concept, according to which the strength of positive selection determines the **rate of evolution**, the neutral theory maintains that it is the characters without any fitness value, all mutations of which are neutral, that show the maximal rate of evolution and the highest polymorphism. The more important a molecular character is for the maintenance of structure and function of a nucleic acid or protein, the greater is the restriction of its molecular evolution. As discussed in section 4.6, this concept has been confirmed many times. The highest rates of evolution are found amongst the pseudogenes that are no longer expressed; non-coding DNA sequences change quicker and are more polymorphic than coding sequences (see Table 4.13, p. 162), and synonymous substitutions predominate over amino acid exchanges in coding sequences. The more complex the interaction of individual proteins with other cell components and the more important their function for the whole organism, the lower is their rate of evolution (see Table 4.12; p. 161); thus, the amino acids that are most frequently exchanged are those of lesser importance for the structural and functional properties of the protein [440].

The following numerical example illustrates the process of formation and fixation or elimination of neutral mutations. Consider a gene of approximately 1000 nucleotides that codes for a protein of about 330 amino acids; the effective population size is $N_e = 10^5$, and the neutral mutation rate $v_o = 5 \cdot 10^{-9}$ per nucleotide and generation, corresponding to the value observed for pseudogenes. Thus, one neutral mutation occurs in this particular gene per generation and, in most cases, will be randomly eliminated in the course of about 24 generations. The probability that the mutation will be fixed in the population is approximately 1:200 000. It requires about 400 000 generations for the mutation to spread throughout the whole population, and $1/v_o =$ 200 million generations before the mutated nucleotide is finally replaced by another. Many speciation processes will, of course, occur within this period. Therefore, neutral mutations are usually very old, in fact mostly older than the species in which they are found. The nucleotide position (the mutation site) is polymorphic only during the 400 000 generations required for fixation; during the following 200 million generations it will be monomorphic. The probability of polymorphy is correspondingly larger when the numerous nucleotides of the whole gene are considered [210, 212]. The time required for the fixation of a useful mutation is comparatively short; for a mutation with a selective advantage s (fitness = 1 + s), the time is given by $(2/s)\log_e(2N_e)$; with $s = +0.01$ and $N_e = 10^5$, the time required is 2900 generations. Thus, advantageous mutations in the phase of their fixation contribute very little to the polymorphism of the population [294].

If we make the assumption that, in reality, each mutation produces a new allele (the infinite allele model), then the **mean heterozygosity** is given by

$$H = 4N_e v_o/(1 + 4N_e v_o) \qquad (4.9a)$$

Taking into account the limited resolution of electrophoretic investigations (stepwise allele model), then

$$H = 1 - (1 + 8N_e v_o)^{-1/2} \qquad (4.9b)$$

Where $N_e v_o < 1$, the two models give very similar results, where $N_e v_o$ is large, the value of H from Eq. (4.9b) will be smaller [210, 212]. Extremely low, almost undetectable heterozygosity results when the population is very small ($N_e < 1000$), as is often the case, for example, with large mammals [210, 212, 294]. According to Eq. (4.9a), in very large populations H should approach 1.0, whereas in fact the mean heterozygosity rarely exceeds 0.3. The relatively low value of $H = 0.177$ in the tropical species *Drosophila willistoni*, which includes thousands of millions of individuals, was used by Ayala as an argument against the neutral theory of polymorphism. However, it is also probably the case in this species that the effective population size is much smaller than the total number of individuals, and the polymorphism is heavily reduced by the frequent elimination of local populations [210, 212, 294]. Polymorphism can also be reduced by highly selected substitutions [201].

During its further development, the neutral theory diverged from the classification of mutations into three groups (negative, neutral and positive) to take into account all intermediate stages between essentially neutral mutations and those that cause distinct phenotypic effects and are, therefore, subject to selection [210, 212]. The fixation probability of a mutation depends upon the selection coefficient s and the population size N_e. Very weak negative or positive mutations for which $[N_e s] < 1$, i.e. where less than one individual will be selected per generation, behave like neutral mutations. Significantly positive mutations have a higher fixation probability and significantly negative mutations a lower fixation probability. Even a mutation with a high selective advantage of 1 % ($s = +0.01$) has only a 2 % chance of fixation; in 98 % of cases it will be lost by chance. A weakly negative mutation which behaves neutrally in a small population will be subject to selection in a large population [210, 212].

Based on the contribution of single amino acids to the biochemical and biological properties of protein molecules, Zuckerkandl has proposed a further **theory of protein evolution**. He distinguishes between functions of individual amino acids, e.g. in the active centre of an enzyme, and general functions connected to solubility, charge density, isoelectric point and polarity of the protein. Each amino acid has several general functions, and each general function involves several amino acids. The exchange of any one amino acid may change a general function either positively or negatively, i.e. on balance such changes may be selectively neutral. As one of the general functions deviates more and more from the optimum by such gradual changes, the selective pressure increases. Thus, the protein fluctuates about the optimal value for any general function [458, 459]. New insights into the importance of specific individual amino acids for protein stability support this concept [2].

4.4.2 Selection Theories of Polymorphism

For the neutral theories, polymorphism is an obvious concomitant of molecular evolution (Fig. 4.7). Selectionism, on the other hand, attempts to explain stable polymorphism, which does not just represent a short transitional phase between different adaptive conditions; it does this by assuming special **mechanisms of "balancing" selection**. Three of the most frequently quoted mechanisms are [135, 294, 300]:

1. Over-dominance between two alleles, when the heterozygotes are superior to both homozygotes.
2. A selection which is dependent upon the allele frequency or the population density, or which differs in consecutive generations.
3. Diversifying selection, i.e. adaptation to the spatial or temporal variability of environmental parameters (niche theory of protein polymorphism).

Heterozygotic advantage is not infrequently observed in crossing experiments, but may be confused with the effect of a neutral allele coupled to a directionally or balanced selected gene (hitch-hiking effect and associated over-dominance). Perfect cases of over-dominance are otherwise quite rare. The classical case is human sickle-cell haemoglobin, which increases resistance to malaria in the heterozygotic state and, as a consequence, reaches frequencies of 20 % in some African populations. The increased resistance to malaria of heterozygotes is also thought to be responsible for locally increased frequencies of β-thalassaemia, a heritable anomaly in which the β-globin chain shows reduced synthesis. One argument brought against over-dominance as the explanation for all polymorphism is the high protein polymorphism of natural populations of haploid organisms like *E. coli, Neurospora* sp. and certain moss species, in which no heterozygotes are possible [294, 448].

Frequency-dependent selection as a cause of protein polymorphism has been convincingly demonstrated, for example, for the amylase of *Drosophila melanogaster* larvae. Homozygotes for the null allele Amy^{null} produce no active enzyme, and the higher the frequency of normal Amy^1 homozygotes, the higher is the chance of survival of the null mutants on starch-containing substrates. Amy^1 animals apparently release glucose and/or amylose into the substratum [156]. It is not known exactly which selective advantage protects the Amy^{null} allele from complete elimination. Frequency-dependent, and thus polymorphism-producing, selection could also be involved in interactions, for example with predators, parasites or competitors, but there are no really convincing examples of this [460].

Most important amongst the **niche theories** is the environmental grain hypothesis, which is based on the concept of the habitat of each population as a mosaic of regions with different properties. In the case of a "coarse-grained species", each individual is specifically adapted to

one of these niches; the population as a whole is highly polymorphic. In contrast, each individual of a "fine-grained species" can survive in each niche. In the latter case, there is little polymorphism as the same universal alleles are found in all individuals. Similar terms, but with variation of environmental parameters in time rather than in space, are used in the "environment amplitude variation" hypothesis to explain protein polymorphism. As these hypotheses were later contradicted by the discovery of high polymorphism amongst deep-sea creatures and tropical invertebrates living in constant, homogeneous biotopes, Valentine and Ayala developed, in 1971, the "resource predictability variation" hypothesis: With a continuous, high food supply, specialization is advantageous because it reduces intraspecific competition; this results in higher polymorphism. During periods of food shortage, all individuals must be able to make use of any form of nutrient and they are, therefore, genetically similar [17, 294, 439]. Because it is never known exactly what selective value the time- or space-variable parameters have at any one moment, clear confirmation or rejection of the niche theories is hardly possible.

4.4.3 The Controversy Between Neutralism and Selectionism

Neutralism and selectionism are not really alternatives; the question is no longer whether (quasi-)neutral genetic alterations exist, but rather what proportion they represent of the total genetic variability. An irrefutable argument for the predominance of neutral alterations in molecular evolution is the fact that the lower their importance, the faster molecular characters evolve, and the more they vary. If one argues that characters of minor importance even react to weaker selection pressure [439], then with the assumption of very weak selection effects the selectionist and neutral theories are reconciled. The exact proportion of neutral genetic variation will in fact never be determined. Neutrality, i.e. the absence of selection, as a negative property cannot be conclusively proven. The particular advantage of the neutral theories is that they allow relatively simple, quantitative models of molecular evolution and molecular polymorphism to be established.

The following arguments are mainly brought in favour of **selection as the cause of protein polymorphism:**

1. The relative frequency of the alleles in natural populations does not agree with the predictions of the neutral theories.
2. Protein polymorphism (in the terms of the niche theory) is correlated with variation in time and space of the environment.
3. Selection can be shown directly for single gene loci:
   a) Alleloenzymes have different properties in vitro.
   b) Organisms with different alleloenzymes have different physiological properties and are subject to selection in the laboratory or the field.
   c) Allele frequencies are correlated with the geographical location or a site-dependent environmental factor.
   d) Reproductively isolated populations (e.g. different species) in the same biotope have matching allele frequencies.

Many studies have used **statistical methods** to examine whether the allele frequencies in natural populations agree with the predictions of the neutralist theories [183, 254, 294], and a negative result has often been interpreted as evidence for the correctness of the selectionistic explanation of protein polymorphism. There are, however, many basic problems with such studies. Most of the equations are only valid for equilibrium; however, this can be disturbed by brief changes in population size and may not be immediately re-established. Phenomena such as coupling and epistatic interaction between genes are not taken into account. Finally, the models are always concerned with on the one hand, the relationship between deterministic forces, like the mutation rate v, the selection coefficient s or the migration rate m, and on the other hand, probability, which is proportional to $1/N_e$. Terms, such as $N_e v$, $N_e s$ or $N_e m$, are derived in which neither the very high value of N_e nor the very low values of v, s or m can be sufficiently accurately determined; almost any model can be "proven" with appropriate plausible assumptions for these parameters [237].

According to the **niche theories**, protein polymorphism should be correlated with the variability of living conditions [126]. Generalists, i.e. animal species with large distribution areas, broad ecological niches and large populations, should have a higher mean heterozygosity than specialists; this trend is confirmed by the analysis of several large collections of data [21]. The low heterozygosity of the mammals (Table 4.8) is also indicated by the niche theories; mobile mammals

must be able to survive under very different living conditions. However, a whole series of results, particularly on animals from deep-sea and tropical biotopes with constant and homogeneous environments, contradict the niche theories. Distinct protein polymorphism with H = 0.075–0.189 (with a mean of 0.15) was found in ten deep-sea animal groups (one species each of gastropods, brachiopods and crustaceans, and seven echinoderms [80]. The giant mussel *Tridacna maxima*, which lives in the very constant conditions of the Great Barrier Reef, has an H value for 37 loci of 0.209 [55]. There is also no evidence that periodic changes in environmental parameters have a significant influence on protein polymorphism; investigations over several years, of *Drosophila pseudoobscura* and *D. persimilis* revealed only a few cases of synchronous changes in polymorphism at several loci [282]. No correlation was found between the polymorphism at 21–30 loci of 13 marine teleosts and the extent of the seasonal changes in water temperature, although the fluctuations ranged from about 1 °C (for antarctic species) to 20 °C [389].

In vitro differences between alleloenzymes in biologically important characters are often taken as evidence for balancing selection as a cause of protein polymorphism. However, it is always necessary to ask how far such in-vitro-determined differences are important for the fitness of the animal. Alleloenzymes show particularly high variability in their temperature stability; this is actually used as a distinguishing feature. For example, *Notropis lutrensis* possesses like other teleosts, two soluble malate dehydrogenases (sMDH) A and B, of which sMDH-B shows higher polymorphism with three alleles (F, M and S). The enzymes of homozygotes have different temperature optima: 20 °C for FF, 25 °C for MM and 30 °C for SS; in each case incubation was for 45 minutes. The temperature of the habitat varies between 11 °C and 29 °C [346]. Variable temperature dependency of alleloenzymes is also found in the lactate dehydrogenase (LDH) and glucose phosphate isomerase of fish [22, 81]. Corresponding results have been obtained with *Drosophila*, e.g. for xanthine dehydrogenases and esterases-5 of *D. pseudoobscura* [109, 436], α-glycerophosphate dehydrogenases of *D. virilis* and *D. melanogaster* [285], and the superoxide dismutases of *D. melanogaster* [229]. The pH of body fluids also depends upon the temperature; an increase of 1 °C reduces the pH by 0.015–0.020 pH units. Thus, the variable pH-dependence of the LDH-B alleloenzymes of the teleost *Fundulus heteroclitus* many be considered as an adaptation to temperature [330].

Many cases of variation in the catalytic properties of alleloenzymes have also been established, for example of substrate affinity or specificity. The alleloenzymes of the larval esterase 4 of *D. mojawensis* show particularly high variability. The K_m for 1-naphthylacetate of the esterase 4-100α is ten times lower than that of the 4-86β (28:284 μmol/l), and the k_{cat} is more than twice as high (832:367 min^{-1}), giving an overall 23-fold increase in catalytic efficiency, k_{cat}/K_m, of the 4-100α alleloenzyme (29.3:1.3) [324]. Differences in the catalytic properties of alleloenzymes have also been recorded for the superoxide dismutase and alcohol dehydrogenase of *D. melanogaster* [167, 229] and the lactate dehydrogenase of various teleosts [315]. Various attempts have been made to find evidence for a heterozygote advantage in proteins by in vitro measurements. The bactericidal ovotransferrin of bird eggs is found in several allelic variants in the pigeon *Columba livia* and the pheasant *Phasianus colchicus;* the protein of the heterozygotes was more inhibitory against yeasts in vitro than was that of the two homozygotes. The rate of hatching was also increased in the heterozygotes [253]. In contrast, in vitro hybridization of the allelic subunits of esterase-5 of *D. pseudoobscura* did not increase either specific activity or temperature resistance [24].

Compared with the numerous results showing differences between alleloenzymes in vitro, there is little clear evidence for **different functions in vivo**. In *D. melanogaster*, the different rates of direct glucose oxidation depend mainly on the allelic variants of 6-phosphogluconate dehydrogenase (6PGDH). The 6PGDH alleles with lower activity or null alleles of this enzyme result in reduced viability. However, animals in which the activity of glucose-6-phosphate dehydrogenase (G6PDH) is also reduced are completely viable; this situation is probably due to the inhibition by accumulated 6-phosphogluconate of the enzymes of glycolysis and, in particular, of hexosephosphate isomerase [100]. The significance of the heritable differences in α-amylase activity in *D. melanogaster* has already been mentioned (p. 137). The selection conditions for alleles of α-glycerophosphate dehydrogenase of *Drosophila* are quite complicated. Although the equilibrium concentrations of glycolytic intermediary products are the same in FF and SS homozygotes of *D. mercatorum* during both rest and flight [66], SS flies of *D. melanogaster* have a 2–4 % higher

flight performance. Breeding at 15 °C selects for the S allele and at 30 °C selects for the F allele [18]. Systematic investigations of the influence of the alleloenzyme spectrum on mechanical flight performance in *D. melanogaster* showed no clear effects of any of the 15 enzymes of energy-converting metabolism tested [227]. In *Mytilus edulis*, the rate at which the intracellular concentration of free amino acids in the mid-gut gland is osmotically increased after transfer to 120 % sea water is significantly higher in animals with leucine aminopeptidase (Lap^{94}) than in animals with any other allele [173]. In erythrocytes of the teleost *Fundulus heteroclitus*, found on the Atlantic coast of North America, there are two allelic variants of the lactate B subunit, whose kinetic properties involve different temperature sensitivities. Thus, Ldh-B^bB^b animals are superior to Ldh-B^aB^a animals in swimming performance and developmental rate at 10 °C but not at 25 °C. The frequency of Ldh-b^b alleles increases from south to north, reaching 100 % in the cold waters off Maine [92].

It should be possible to demonstrate direct selection for allelic enzyme variants that show differences in vivo, and the prime example here is that of the **alcohol dehydrogenase** (ADH) of *Drosophila melanogaster* [166]. This is a homodimer coded by a single gene, for which at least eight alleles can be distinguished by electrophoresis and heat inactivation. Predominant in natural populations are the variants S (slow) and F (fast), which differ in only one amino acid; the rarer variant UF (ultrafast) differs from F and S by two to three amino acids (Fig. 4.8). Thermostable "fast" variants have been isolated from laboratories in three continents and are known as fast-71k in Europe, fast-resistant in America and fast-Chateau Douglas in Australia. They apparently arose from the F allele; in any case, the fast-Chateau Douglas has the typical amino acid 192-Thr in addition to the substitution 214-Pro to Ser [75]. The variant 71k has been shown to have a higher thermal stability and a deviant substrate specificity; it has activity also with sarcosine (N-methylglycine) and dihydroorotate, which appear as intermediates in choline metabolism and pyrimidine biosynthesis, respectively [105, 167].

The electrophoregrams of ADH show three to five sub-bands for each alleloenzyme, and these arise by reversible formation of abortive NAD^+-carbonyl complexes. The specific activity depends upon the number of such complexes in the dimer; ADH-1 with two complexes is inactive, ADH-3 with one complex is half-active, and ADH-5 without any complexes is fully active [441, 442]. The degradation rates for propane-2-ol and ethanol in homozygous larvae decreases in the allelic order Adh-71k > Adh-F > Adh-S; however, this has more to do with genetically regulated differences in the amount of enzyme than with the specific activity [166, 228]. Post-translational modification also plays a role in the regulation of enzyme activity. Thus, in FF animals the effect of the secondary alcohol propane-2-ol in the medium is to increase complex formation and reduce the specific enzyme activity, although, at the same time, enzyme degradation is inhibited and therefore the equilibrium enzyme concentration is increased; the overall effect is a reduction in enzyme activity and the formation of highly toxic ketones [8]. FF larvae survive better than SS animals on alcohol-containing substrates; the SS animals are only superior if the primary oxidation product of the alcohol is particularly toxic, e.g. in the case of 1-pentene-3-ol. Breeding on an ethanol-containing substrate selects the Adh-71k allele before the Adh-F and the Adh-S alleles [165].

Position	F	S	UF
8	Asn	Asn	Ala
45	Ala	Ala	Asp
192	Thr	Lys	Thr

Fig. 4.8. The allelic variants of alcohol dehydrogenase in *Drosophila melanogaster* differ at three positions [441]

ADH-SS in vitro is more heat-stable than the FF enzyme. SS and FS animals have a significantly higher temperature resistance than do FF animals in the presence of high alcohol concentrations. As evidence for a difference in temperature adaptability of the two ADH alleloenzymes, it is often pointed out that in the eastern USA the frequency of S increases north to south from 0.50 to 0.90 [378]. On the other hand, a comparison of the ADH of different *Drosophila* species indicates that their temperature stability is not correlated with the habitat, i.e. it is not especially higher in tropical species than in those from more temperate climates [1]. In general, selection of the Adh locus appears to be very complicated. Whilst at high alcohol concentrations in the laboratory, the F allele is always selected, in natural populations, e.g. animals from wine cellars compared with those in the surroundings, there is often no difference in the F frequency; of course, the possibility of gene exchange between the subpopulations must also be considered. *Drosophila melanogaster* appears to be specifically adapted to ethanol-containing substrates, whereas many

other *Drosophila* species show less alcohol tolerance [166]. Very similar ADH electromorphs are found in the flour moth, *Ephestia kühniella* and in *D. melanogaster*, although the former never encounters alcohols in its natural habitat [231]. In experiments with the kelp fly, *Coelopa frigida*, the Adh allele frequency was not affected by 0.2 % alcohol in the medium, perhaps because here, unlike in *Drosophila*, ADH has no detoxifying function [86].

Correlation of the allele frequency with the geographical site or ecological factors is often given as evidence for the effect of selection. The directed alteration of a gene frequency over a geographical distance is known as a **cline**, according to the concept derived by Huxley in 1939. Many environmental parameters also show directed site dependency such that, for example, a north/south cline can be looked upon as an adaptation to temperature. However, which of the many site-dependent environmental parameters is really effective in the selection of a given locus can, at best, be presumed. Neutral alleles also form frequency gradients as they disperse from the site of the mutation. There are many examples of clines, and particularly of north/south clines. In the sea anemone *Metridium senile*, found on the east coast of North America, the frequency of the alleloenzymes PGI-F of glucose phosphate isomerase declines from almost 100 % in the south to only 50 % in the north. The explanation offered for this is that, based on the temperature dependency of its kinetic properties, PGI-F is better adapted than PGI-S to higher temperatures in the surroundings. As in many such conclusions, the question remains open as to exactly which selective advantage maintains the relatively high frequency of the PGI-S allele in the northern population [178]. In the mussel *Guekensia demissa*, the frequency of a cryptic phosphoglucomutase allele, which was identified by its higher temperature sensitivity, is correlated with the average water temperature [145].

There are many particularly appropriate examples in the insects. The cline of the alcohol dehydrogenase of *Drosophila melanogaster* in the eastern USA has already been mentioned (p. 145); a similar cline also exists in Australia but is correlated more with rainfall than with temperature. The 6-phosphogluconate dehydrogenase of the mosquito *Culex pipiens* is a dimeric enzyme for which two subunit alleles, F and S, are known. The frequency of S declines from 30 % in the north to below 10 % in the south. This may be partly explained by the fact that the temperature stability increases in the order SS < FS < FF; however, the exact selective advantage of S is again not clear. Further enzymes of this species also show north/south clines [337, 338]. In various fish species, north/south clines have been found for lactate dehydrogenase and other enzymes [330, 351]. Selection is indicated as the cause of polymorphism when parallel clines appear in closely related species existing in close proximity (sympatric), or the same cline is seen on different continents for widely dispersed species. Thus, the sympatric cricket species *Gryllus veletis* and *G. pennsylvanicus*, which are phenotypically very similar but sexually isolated, show the same type of site dependency in the frequency of glucose phosphate isomerase alleles [159]. A similar dependency on geographical range is found for a whole series of polymorphic enzymes in populations of *Drosophila melanogaster* from all continents [7]. If a protein polymorphism allows adaptation to climatic factors, it should show regular seasonal changes; however, no annual periodicity was detected, for example, in 12 enzyme loci in *D. melanogaster* [61].

In several cases, **a correlation of gene frequency with climatic or ecological factors** has been directly demonstrated. Two populations of the barnacle *Balanus amphitrite* live in the inflowing and outflowing cooling-water canals of a power station in Haifa, with a temperature difference of 9–12 °C. The animals in the cooler canal are three times larger and have a four-fold higher population density than those in the warmer canal. Differences in allele frequency were found at 8 out of 12 loci, and these may be looked upon as adaptation to the different conditions [299]. In *Drosophila simulans*, sequential electrophoresis distinguishes 27 alleles of esterase 6; the pattern relates more to the available fruit than to the location and thus represents ecological races [3]. The electrophoregrams of various enzymes of aphid species reveal host-plant-specific differences [393]. Amongst different North American *Drosophila* species, whose larvae develop on cacti, *D. longicornis*, which is restricted to *Opuntia*, shows only half as much polymorphism at four enzyme loci as the species *D. mojavensis* and *D. arizonensis*, which live on various cactus genera [345]. The mosquito species *Anopheles aquasalis*, whose larvae require salt water for their development, have a significantly lower polymorphism (H = 0.081) than five related species with H = 0.17–0.27 [284].

Coincidental electrophoretic patterns in sexually isolated populations are commonly observed

and taken as evidence for selection at the relevant locus. However, this conclusion is not at all certain. Firstly, identical electrophoretic mobility of proteins does not necessarily mean structural identity. Secondly, alterations in the frequency of neutral alleles occur so slowly that similarities between populations are maintained for a long time, e.g. up to 2 million years in *Drosophila*, after their separation, even in the absence of selection [240]. Finally, all gene exchange between the compared populations must be excluded as, independent of the population size, the migration of just a few individuals per generation prevents the occurrence of large differences. It could be shown for *D. pseudoobscura* that intensive gene exchange occurred even between populations 15 km apart. Thus, the experimentally disturbed frequency pattern of esterase 5 in one oasis population was restored already after 1 year; marked flies travelled up to 10 km in 1 day [196].

Although conclusions cannot be drawn with any certainty from the coincidence of electrophoretic patterns of isolated populations, different allele frequencies in closely related, sympatric, but sexually isolated, populations are a convincing argument against selection as the cause of polymorphism at any given locus. The nicest example of this is found in the salmon *Oncorhynchus garbuscha*. This shows such a rigid biennial reproductive rhythm that two sexually isolated populations exist side-by-side in many Alaskan rivers; these are known as "even-year" and "odd-year". These genetically separated populations living in the same habitat show large differences in allele frequencies of the malate dehydrogenases Mdh-A and Mdh-B and in α-glycerophosphate dehydrogenase [12]. Completely different heterozygosity values (H = 0.04–0.27) and variable frequency patterns were found at six enzyme loci in 12 sympatric species of *Drosophila* [355]. Furthermore, it remains a mystery why, in human populations, the frequency of rare alleles at 21 erythrocyte protein loci varies between the English (0.07 %) and native inhabitants of Australia (1.1 %) [387].

4.5 Methods and Problems in the Molecular Approach to Evolutionary Relationships

Homologous DNA regions in different organisms are always found to contain numerous substitutions and rearrangements which can be used in the analysis of the genealogical relationships of individuals and species. Man and the chimpanzee, which are considered to be closely related species, differ in about 2 % of their DNA sequences, i.e. in approximately 60 million nucleotides, although, as it happens, the majority of the sequence differences have no phenotypic effects. Even individual humans differ in up to 5 million nucleotides [38]. Each human gamete has, on average, about 20 nucleotide substitutions [288]. The comparison of molecular characters need not always make use of the costly techniques of DNA or protein sequencing. There is a whole series of molecular properties which are correlated to the DNA or protein sequences and are easily compared in a large number of individuals or species; they often involve large parts of the genome. The following contains a description of these various methods for comparing molecular characters and their use in constructing phylogenetic trees [175].

The reconstruction of phylogeny is closely associated with the **classification of the organisms;** this latter is the task of a special biological discipline known as systematics or taxonomy [198, 347]. Basically, organisms may be grouped according to chosen characters and criteria, e.g. their usefulness or destructiveness towards man, their habitats or any other character appropriate to their identification. However, as a general frame of reference, a preferred system stems from the organisms themselves, i.e. it can be considered a natural system. The three most important schools of systematics strive for such a system in quite different ways.

1. "Phenetic systematics" organizes organisms according to the degree of their similarity. A particular advantage of this method is the possibility for quantification. The results of a phenetic analysis can be presented in the form of a branching scheme (a cladogram), which may be considered to be a picture of the most likely phylogenetic relationships. The founders of such "numerical taxonomy" were Sneath and Sokal [111, 385].
2. The development of "phylogenetic systematics", mainly by Willi Hennig, is based entirely on the phylogenetic relationships of the species as depicted in a cladogram with dichotomous branching; the principle of this method is the identification of the sibling species or group of a particular species that share the most recent common ancestor [347].

3. "Evolutionary systematics" is also based on phylogenetic relationships but takes into account phenetic similarity. The important proponents of this system are Simpson and Mayr [198].

A large problem in the analysis of phylogenetic relationships is the fact that similarities in a particular character may not simply be due to a common origin (homology), but rather to the development of similar characters in two independent evolutionary lines **(parallel evolution)** or to the adaptation of characters to similar functional requirements **(convergence)**. In the face of these difficulties, the phenetic school makes the assumption that the analysis of a large enough number of characters reduces the influence of parallel evolution or convergence on the results. The other two schools attempt to recognize the existence of parallel evolution and convergence, in particular by examining the agreement between cladograms constructed using different characters (congruence criteria) [198, 385].

The basic unit of biological systematics is the **species**. The definition of this term already presents large problems. Comprehensible in real terms for bisexual organisms is the population, whose members form a reproductive community and can freely exchange genes. In this connection, a species may be defined biologically as a group of individuals that actually or potentially forms a reproductive community and which is reproductively isolated from other such groups. As this definition applies only to organisms with bisexual reproduction, and as, furthermore, a "potential" reproductive community cannot be determined, a phenetic species concept is normally applied in practice, according to which the members of one species are distinguished from another species by particular (diagnostic) characters. Groups of species are brought together in higher taxa as genera, families, orders, classes, phyla and kingdoms; further classification may be into sub-orders or superfamilies. The assignment of a taxon to one of these categories is based on different rules in the different schools of systematics: according to the estimated degree of similarity in the phenetic school; based on the phylogenetic age in the phylogenetic school; and taking into account the number of species and the heterogeneity in the evolutionary school. Whatever the case, taxa of the same rank in different parts of the system are not directly comparable [198, 385].

In accordance with the international **rules of nomenclature**, each organism is known by a Latin name consisting of at least two parts, the genus and the species; where necessary, the name of the subspecies is also added. The genus name is always written with an initial capital letter, whereas the species or subspecies name starts with a lower-case letter. In order to refer completely and unambiguously to a species, the name of the first describer should be given. Unfortunately, biochemical texts frequently break the rules of biological nomenclature and the organisms in question are often very carelessly named. Trivial names, such as "frog", "toadfish" or "locust", are often applied to very different species and in no way clearly define the species actually under examination. Because considerable biochemical differences can exist between various species, biochemical results lacking exact reference to the organisms involved are quite worthless.

4.5.1 The Evolutionary Distance Between Amino Acid or Nucleotide Sequences

The simplest measurement of the difference between two amino acid sequences is the number of varying amino acids. If, however, by **evolutionary distance** is meant the number of nucleotide substitutions that have become fixed during the diversifying evolution of two structural genes, then it must be taken into account that one, two or three substitutions may be required for one amino acid exchange, depending on the degree of difference in the codons. The minimum sum of substitutions, calculated from the genetic code, to explain the transformation of one amino acid sequence into another was described by Fitch in 1967 as the "minimum mutation distance" (MMD) [120]. The MMD, however, gives only an incomplete estimate of the genetic distance between the structural genes. It takes no account of synonymous substitutions that do not lead to amino acid exchange, and cannot deal with multiple, parallel and back mutations (Fig. 4.9).

In the last two decades numerous methods have been developed for the determination of evolutionary distances between proteins, and the relative suitability of these has been the subject of vigorous debate. Conclusions drawn from differences in two amino acid sequences about the corresponding structural genes are questionable, because information is clearly lost during translation; the information content decreases from $3 \cdot \log_2 4 = 6$ bits for a triplet to $\log_2 20 = 4.32$ bits for one amino acid. On the other hand, it is not only the selection pressure on the protein that is

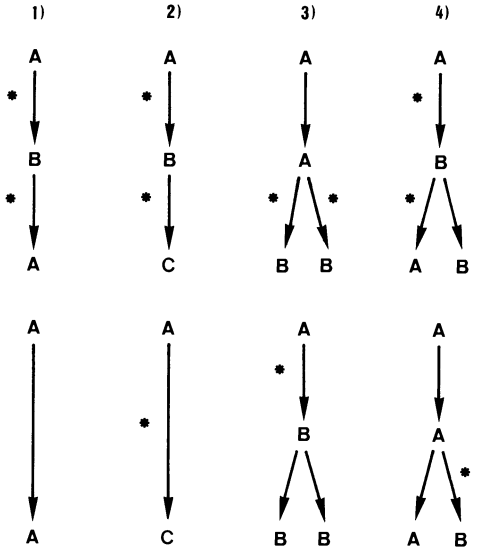


Fig. 4.9. Faulty reconstruction of the evolutionary relationship between different characters *(A, B* and *C);* this can result from back mutation *(1* and *4)*, multiple mutation *(2)* or parallel mutation *(3)*. The true process is illustrated in the *upper part* of the diagram and the mistaken reconstruction is shown in the *lower part*. The *asterisks* each represent a mutation

effective in the evolution of a structural gene but also the selection for particular structural/functional properties of the DNA and mRNA. Synonymous nucleotide substitutions can, of course, only be recognized by investigation of the DNA or mRNA. The number of multiple and back mutations can be calculated on the basis of a mathematical model of evolution, the accuracy of which remains questionable because it cannot take into account the effect of selection and other non-stochastic factors. Conclusions drawn about genes from protein sequences are decreasing in importance in the wake of advances in nucleic acid sequencing. The problem of how to include multiple and back mutations, however, also applies to nucleotide sequences. To take account of multiple and back mutations in estimates of genetic distance using amino acid sequences, Zuckerkandl and Pauling suggested, in 1965, the use of a **Poisson correction**. Accordingly, the average number of exchanges per amino acid between two sequences is given by

$$K_{aa} = -\log_e[1-(d_{aa}/n_{aa})] = -\log_e(1-p_d) \quad (4.10)$$

and the rate of protein evolution is given by

$$k_{aa} = K_{aa}/(2T) \quad (4.11)$$

In these equations, p_d is the proportion of varying amino acids, calculated from the number of amino acid differences (d_{aa}) and the length of the sequences (n_{aa}) and T is the separation time of the evolution lines of the two sequences. The equation gives usable results for up to approximately 40 % sequence difference ($p_d \leq 0.40$) [210, 212]. For the determination of genetic difference in the case of larger differences, Margaret Dayhoff derived a matrix using 1572 amino acid exchanges in closely related proteins; this gives the relative probabilities of individual amino acid exchanges ("mutation data matrix" MDM78). With the help of this matrix it is possible to calculate a genetic distance for each sequence pair of the amino acid differences, and for this purpose the **distance measurement PAM ("point accepted mutations"**, defined as substitutions per 100 codons) is introduced [87]. The procedure has been recently improved [435]. Numerically similar values are obtained with the following equation [210]:

$$K_{aa} = -\log_e(1-p_d-\tfrac{1}{5}p_d^2) \quad (4.12)$$

A new matrix (EMPAR, exchange matrix derived from parameters) similar to the MDM78, has now been suggested and makes use of certain physicochemical parameters of the amino acids, instead of observations on amino acid exchanges. As evolution prefers conservative amino acid exchanges in which the physicochemical properties are only slightly altered, the EMPAR and the MDM78 matrices assign individual amino acid exchanges similar probabilities [342]. Since 1972, Holmquist and Jukes have been working on the development of a **measurement of distance (REH, "random evolutionary hits")** from a stochastic model of evolution; this makes allowance for all substitutions, including synonymous, multiple and back substitutions. This model assumes that at any point during evolution only some of the codons vary, and these "varions" change after each substitution [141]. The **maximum parsimony method**, developed by Goodman and Moore from 1972 to 1976, starts with a phylogenetic tree of minimal total length ("maximum parsimony tree"). The lower the "density" of any part of this tree, i.e. the lower the amount of sequence information to be evaluated, the more the frequency of multiple and back mutations will be underestimated. This is taken into account by the calculation of an **"augmented distance"** (AD) [141].

The AD method always makes use of a directly determined **nucleotide sequence**, or one derived from an amino acid sequence. The REH method in a modified form can also use nucleic acid data [141]. In addition to the possibility of multiple

and back mutations, the comparison of DNA sequences presents further problems due to: (1) differences in base composition; (2) variable frequencies of the 12 possible substitutions; (3) unequal distribution of the substitutions between the codons of a gene; and (4) unequal distribution between the three codon positions.

To calculate the **number of substitutions** per nucleotide (K) from the proportion of variant nucleotides (λ) in two compared DNA sequences, taking into account multiple mutations, an equation, which was suggested by Jukes and Cantor in 1969, is now frequently used:

$$K = -\tfrac{3}{4}\log_e(1-\tfrac{3}{4}\lambda) \qquad (4.13)$$

Equation (4.13) assumes equal frequencies for all possible substitutions. Based on the observation that the ratio of transition to transversion is much higher than that expected by chance, especially in closely related sequences, Kimura (1980) developed an equation which includes separate terms for nucleotide differences of the transition type (P) and the transversion type (Q) [210]:

$$K = -\tfrac{1}{2}\log_e[(1-2P-Q)\sqrt{(1-2Q)}] \qquad (4.14)$$

Since then, other methods have been suggested that allow not only two but four or more different substitution probabilities to be considered [31], and even include the influence of deletions and insertions on the nucleotide differences [238, 402]. As many gaps are required in non-coding sequences to obtain the optimal alignment, because of the high frequency of deletions and insertions, the evolutionary distance calculated from the above equations therefore depends very much upon the evaluation of such gaps [210]. From the sequence differences and the separation time of the compared sequences, it is possible to calculate the **substitution rate per** nucleotide and year [210]:

$$k_{nuc} = K/(2T) \qquad (4.15)$$

There are also methods for calculating the rate of evolution which, unlike Eq. (4.15), do not assume that the two sequences are equally distant from the common ancestral sequence [31]. It is desirable to distinguish between synonymous and amino acid-exchanging substitutions in the determination of sequence differences and substitution rates for coding DNA sequences. Whilst rates of evolution of different proteins calculated from amino acid exchanges are highly variable, the rate of synonymous nucleotide substitution is remarkably constant, and thus particularly suitable for specifying evolutionary distance. As synonymous substitutions almost exclusively affect the third codon position (Table 4.3), K and k_{nuc} are often separately calculated for each of the three codon positions.

The **number of synonymous substitutions and those that alter amino acids** in two compared DNA sequences can usually be determined by direct observation. If, however, the compared codons differ in more than one nucleotide, then more than one evolutionary route is possible. For two or three nucleotide differences there are two or six routes, respectively, with different probabilities due to variability in the order with which synonymous and amino acid exchanging substitutions must occur. This is taken into account in several methods for determining the distance between DNA sequences [210, 243, 296]. Polymorphism can lead to substantial errors in the determination of sequence differences and substitution rates, and methods have also been developed to deal with this situation [403]. There are special methods that use the detection of clusters of associated nucleotide exchanges as evidence for conversion events [394]. Amongst the further sequence comparison methods published very recently is one which uses an unsimilarity measure, the calculation of which does not require optimal alignment of the sequences [32]. A new form of multi-variance analysis, known as "correspondence analysis", has become available and is used, for example, for analysis of the 5S rRNA sequences in numerous prokaryotes and eukaryotes [258]. There are, of course, computer programs for all these methods, and these are constantly updated [29, 96, 202, 323].

4.5.2 Determination of Evolutionary Distance from the Amino Acid Composition of Proteins

Each amino acid exchange leads to an alteration in amino acid composition. Consequently conclusions may be drawn about the extent of sequence differences between closely related proteins by considering differences in their amino acids spectra. However, the probability of reciprocal alterations increases concomitantly with an increase in the number of amino acid exchanges and can lead to an underestimate of sequence differences. Of the various equations suggested for determining evolutionary distance from the amino acid spectrum, that of Cornish-Bowden is the simplest [79, 344]:

$$S\Delta n = \tfrac{1}{2}\Sigma(n_{iA}-n_{iB})^2 \qquad (4.16)$$

where n_{iA} and n_{iB} indicate the number of amino acids of type i in two proteins A and B with the same length N. For technical reasons, Asp/Asn and Glu/Gln are not distinguished and therefore only 18 amino acids are considered. $S\Delta n$ is an estimate of the number of positions at which the compared sequences disagree. Values for $S\Delta n/N$ of up to 0.42, i.e. an estimated sequence difference of up to 42 %, can be taken as clear evidence of homology; in several thousand comparisons of proteins with known sequences, only one exception was found [79]. The $S\Delta n$ value for protein pairs of differing length can also be calculated using quite a complicated equation [344]. According to Ghiretti-Magaldi, the variable probabilities for the origin of an amino acid i, dependent upon the number of synonymous codons, should be taken into account thus:

$$D_c = \{\Sigma([x_{iA}-x_{iB}]/n_i)^2\}^{1/2} \qquad (4.17)$$

In this equation, x_{iA} and x_{iB} indicate the proportions of the amino acids i in A and B; n_i is the number of triplets coding for i. The values obtained in this way for 15 homologous proteins were found to be strongly correlated with the sequence differences [133].

4.5.3 Immunological Distance Between Proteins

The most widely used immunological method for the quantitative determination of protein similarity is that of **microcomplement fixation** (MC'F). MC'F exploits the competition between two different complement reactions: on the one hand, the irreversible binding to antigen-antibody complexes and, on the other hand, the ability to lyse sheep erythrocytes previously treated with an anti-sheep antiserum (haemolysin). To carry out MC'F, the relevant antiserum is allowed to react with the homologous and heterologous proteins. In immunology, the homologue is the antigen protein used to produce the required antiserum; the heterologue is the other protein on which this antiserum is tested. Of course, in the terminology of research into molecular relationships, the two proteins under investigation would be homologues. Depending upon the strength of the antigen-antibody reaction, a larger or smaller proportion of the introduced complement will be bound to the antigen-antibody complex. The excess of complement lyses the sensitized sheep erythrocytes and the released haemoglobin is measured photometrically (Fig. 4.10). Allowing for the dependence of the reaction on the antigen concentration, the calculation is made of how much higher the antiserum concentration must be in order to obtain the same degree of complement

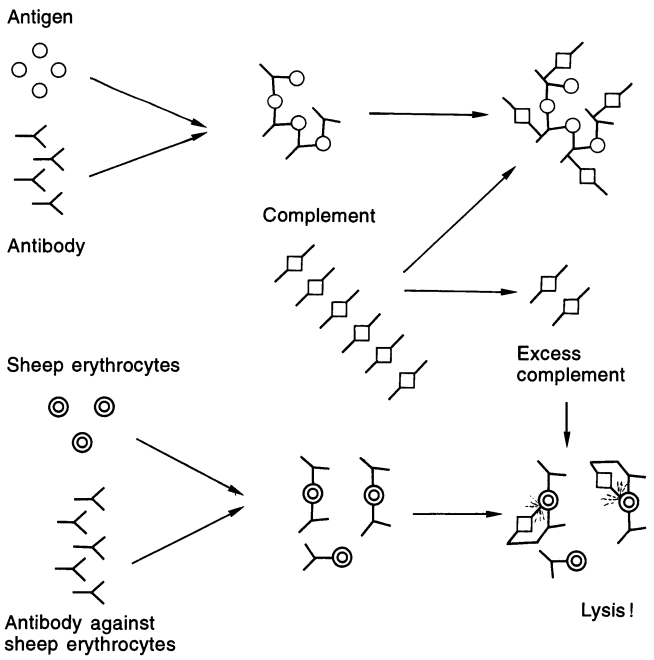


Fig. 4.10. The method of microcomplement fixation [115]. See text for further description

binding with the "heterologous" as with the "homologous" protein. The resulting factor is termed the "index of dissimilarity" (ID), and the value 100 · log ID represents the immunological distance [115]. The distances determined in this way are independent of the antiserum used. A radio-immunological method has recently been developed which gives results similar to those from the MC'F, but is about 1000-times more sensitive and requires only a few nanograms of protein [251].

Theoretically, the **immunological distance** is proportional to the percentage difference in the sequences [267] and this has been confirmed for many proteins. Thus, an immunological distance of 100 corresponds to a sequence difference of 20 % in bacterial azurines, tryptophan synthetases, and avian egg lysozymes, and of about 14 % in pancreas RNases [334]. However, the error may be as large as 60 % with small distances and the accuracy of immunologically determined evolutionary distances should not be overestimated [289]. On the other hand, MC'F is particularly suitable for phylogenetic studies in which as many species as possible must be compared. The method has been used in investigations of several hundred vertebrate species, in particular for comparisons of serum albumin, which is easy to obtain, contains a lot of information in its 580 amino acids, and shows a relatively high rate of evolution. Despite the problems referred to above, these data are still of great importance today to the discussion of central questions in molecular evolution research: the concept of the molecular clock, which is sometimes even termed the albumin clock; the problem of the missing correlation between molecular evolution and the evolution of complex phenotypic characters; and finally, in the solution of individual phylogenetic problems and the construction of phylogenetic trees. Other immunological methods, e.g. immunoprecipitation, immunodiffusion and immunoelectrophoresis, as well as radio- and enzyme-immunological procedures, have mainly been used qualitatively, e.g. as evidence for the homology of proteins, and only rarely for quantitative determination of evolutionary distance. The enzyme-linked-immunosorption-assay (ELISA) should give accurate results for sequence differences of up to 60 %, whereas MC'F is only useful for differences of up to 40 % [177].

4.5.4 Genetic Distance Given by Electrophoretic Data

Comparisons of amino acid or nucleotide sequences, amino acid composition and immunological properties give information of the evolutionary distance of only single genes. In contrast, electrophoretic data on a large number of loci can be used, with a few reservations, to derive the average evolutionary distance of whole genomes. The valid objections here are the same as those for the determination of genetic variability from electrophoregrams:

1. Standard electrophoresis detects only amino acid-exchanging substitutions, and only those which result in a change in charge; equal electrophoretic mobility cannot be taken unconditionally as evidence for sequence identity. In addition, in the determination of evolutionary distance between distantly related organisms, the probability of opposite charge change, which cannot be recognized, increases with increasing distance. Thus, electrophoretically determined distances become more unreliable the further apart the compared taxa. It is therefore recommended that the more efficient method of 2-D electrophoresis be used for the determination of genetic similarity between different groups of animals [278, 457].
2. The commonly investigated proteins (Table 4.9) should not be looked upon as representative of the whole genome.

There exists a whole series of distance measures derived from electrophoretic data. By far the most frequently used is the **"standard genetic distance" D of Nei** [292]. D corresponds to the mean number of electrophoretically recognizable codon differences per locus between two animal groups, X and Y, minus the mean codon difference within the groups. D is equivalent to the negative natural logarithm of the standard gene identity I:

$$D = -\log_e I \tag{4.18a}$$

$$I = J_{xy}/(J_x J_y)^{1/2} \tag{4.18b}$$

$$\text{therein } J_x = (1/r)\,\Sigma_j^r \Sigma_i x_{ij}^2 \tag{4.18c}$$

$$J_y = (1/r)\,\Sigma_j^r \Sigma_i y_{ij}^2 \tag{4.18d}$$

$$J_{xy} = (1/r)\,\Sigma_j^r \Sigma_i x_{ij} y_{ij}, \tag{4.18e}$$

where r is the number of the compard loci, and x_{ij} and y_{ij} are the frequency of the allele i at locus j in the animal groups X and Y. Thus, J_x and J_y are the probabilities, averaged for all the loci, that two

given genes from groups X and Y, respectively, are identical; J_{xy} refers then to the comparison of a gene from group X with one from group Y. I can have values between 0 and 1; I = 1 if identical alleles occur with the same frequency at all loci of X and Y, and I = 0 if this occurs at no locus [292]. The estimation of genetic distance using electrophoretic data is not free of problems:

1. D depends upon the choice of the loci to be compared, and increases as the heterozygosity of the chosen loci increases; thus, highly polymorphic loci also show large interspecific differences [62, 381].
2. Even otherwise insignificant gene exchange between two populations X and Y hinders the development of a large genetic distance. Assuming that per generation m_x individuals migrate from X to Y and m_y migrate in the opposite direction, then, according to the neutral theory, the genetic identity is given by

$$I = (m_x + m_y)/(m_x + m_y + v), \quad (4.19)$$

where v is the number of mutations arising per generation at the relevant locus [292].

The **measurement of distance D_R** of Roger (1972) is also relatively frequently applied; in contrast to Nei's D value, D_R conforms to the so-called triangle condition that is important for the construction of phylogenetic trees, i.e. the estimated three distance values between three animal groups form a triangle:

$$D_R = (1/r)\ \Sigma_j^r\ [½\,\Sigma_i\ (x_{ij} - y_{ij})^2]^{½}. \quad (4.20)$$

D_R can have values between 0 and 1 and is usually somewhat smaller than D [292].

The evolutionary distance values obtained from electrophoretic data agree fairly well with distances determined immunologically or by DNA hybridization, but are inevitably subject to error because different loci usually have different rates of substitution [175]. D and I values obtained from 1-D electrophoresis are mainly suitable for intraspecific comparisons or those between closely related species; 2-D electrophoresis is more suitable for the comparison of distantly related taxa [391].

4.5.5 Comparison of DNA Sequences from the Thermostability of Heteroduplices

DNA double helices dissociate on heating into single strands which reassociate on cooling; the value of the **dissociation or melting temperature (T_m)** depends upon the strength of the interaction of the complementary strands. A heteroduplex made up of DNA single strands from different species has a lower T_m than fully complementary DNA; the difference ΔT_m is a measure of the sequence difference. Until recently, it was accepted that a ΔT_m of 1 °C corresponded to a sequence difference of about 1 %; however, new measurements suggest 1.7 % [53]. Reassociation of total eukaryotic DNA only really involves repetitive DNA which, because of its limited information content and rapid rate of evolution, is unsuitable for the analysis of relationships. Therefore, since 1970, fragmented single-copy DNA has been used for such measurements.

The estimation of DNA sequence differences by these methods is associated with several difficulties and sources of error [44]:

1. An error arises in ΔT_m determination because the latter is dependent upon the base composition; however, this effect can be mitigated by careful choice of the buffer.
2. T_m is dependent upon the length of the DNA fragments; this can be corrected for mathematically.
3. Distinct DNA polymorphism, e.g. in the echinoderms, leads to the situation whereby reassociated DNA from two or more individuals already shows a 2–4 °C lower T_m compared with duplex DNA from one individual [151].
4. Reassociation is never complete; this can also be corrected for mathematically. However, in some *Drosophila* species, one-third of the DNA takes no part in duplex formation, whereas the remaining two-thirds shows great sequence similarity; in some starfish, only one-third of the DNA reacts at all. In such cases, and generally for high ΔT_m values, the accuracy of this form of sequence comparison should not be overestimated.

The genetic distances derived from DNA hybridization are usually significantly larger than those obtained by other methods. Thus, in the case of four *Drosophila* species from three neighbouring Hawaiian islands, enzyme-electrophoresis and ΔT_m determinations gave the same picture of relationships during colonization of the islands (Fig. 4.11), but the electrophoretically determined genetic distances were significantly smaller, perhaps as the result of selection at the investigated loci [186]. Amongst the various methods for the analysis of relationships by investigations of the DNA, the sequencing of individual genes stands at one extreme and T_m

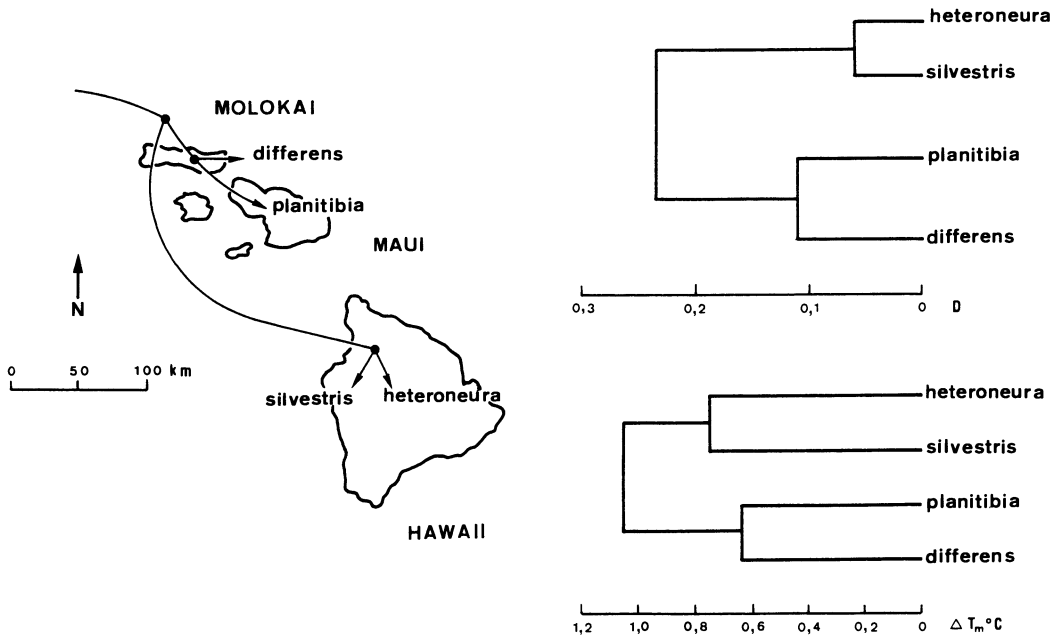


Fig. 4.11. The spread of the four Hawaiian *Drosophila* species and their relationships [186]. The genetic distance derived from enzyme electrophoresis, and data from the melting point depression ΔT_m of DNA heteroduplices allow the same species family tree to be constructed; this agrees with the supposed migration routes. However, the length of the branches of the phylogenetic trees constructed by the two methods do not agree

measurements stand at the other. In between, there are methods such as the analysis of limited DNA regions with the help of restriction endonucleases, or the relatively new procedure of Southern analysis of total chromosomal DNA, i.e. the digestion of the DNA by restriction enzymes into defined fragments, their electrophoretic separation, transfer to a solid phase and, finally, hybridization with DNA probes [250].

4.5.6 DNA Restriction Analysis

Restriction endonucleases cut double-stranded DNA at certain recognition sequences of 4–6 bp in length; many enzymes of different specificity are commerically available. The DNA fragments can be separated by gel electrophoresis according to their lengths, and conclusions can be drawn from the restriction patterns about the extent of the differences between compared DNAs. Restriction analysis of different individuals of the same species allows the determination of DNA polymorphism, and comparisons of two species allows the calculation of genetic distance. The method can be applied to parts of the chromosomal DNA, but for the analysis of relationships it is most often applied to mtDNA, which has on optimal size of 15–19 kb and is easily isolated. If the mtDNA shows a significantly higher rate of evolution than the chromosomal DNA, as in the vertebrates, then it is particularly suitable for analysing close relationships.

The mathematical **evaluation of restriction patterns** is based entirely on the fundamental investigations of Upholt (1977), and the equations of Nei and Li are frequently used [291]. According to these, the mean fraction of identical recognition sequences for two compared mtDNAs, X and Y, is given by

$$S = 2n_{xy}/(n_x + n_y), \qquad (4.21)$$

and from this, the mean number of substitutions per nucleotide as

$$\delta = -(1/r)\log_e S, \qquad (4.22)$$

where n_x, n_y and n_{xy} are the number of restriction sites in X, Y or both mtDNAs, and r is the length of the recognition sequence. Nei has produced a less time-consuming method especially for use in population genetics studies [298]. For larger sequence differences (where $\delta > 0.20$), complicated equations must be applied in order to correct for back mutations. Errors that can arise in these calculations stem from variability in the proportions of different nucleotides, differences in the frequency of the various types of substitution, and the non-random distribution of substitutions

in the DNA. From the wealth of discussion and large number of suggestions for solutions in the literature [29], it is clear that the numerical accuracy of genetic distance values obtained by mtDNA restriction analysis should not be overestimated; with very large genetic distances ($\delta > 0.40$) the method is completely unusable. The use of mtDNA in the analysis of relationships presents some special problems. Thus, the considerable polymorphism of mtDNA has a disturbing effect. By the chance crossing-in of a foreign female, the mtDNA can be replaced by that of another species, and because mtDNA is passed on maternally this situation then persists in the population (pp. 166 and 168).

DNA relationship analysis with the potential of extremely high resolution has now become possible through the introduction of DNA fingerprinting. This method uses as a probe a certain type of repetitive DNA that is widely dispersed in the genome, the so-called mini-satellite. Total DNA of suitable cells (e.g. leukocytes) is cleaved by restriction enzymes into defined fragments, the fragments are separated electrophoretically and hybridized on a Southern blot with a radioactively labelled mini-satellite. The resulting pattern is specific for each individual; the similarity increases with increasing relationship, being highest between parents, children and siblings. A greater degree of separation, as is required, for example, for parenthood analysis or forensic purposes, can be obtained by 2-D electrophoresis [418]. Highly variable mini-satellites are apparently widespread and the method is, therefore, increasingly used for population genetics investigations on vertebrates of all classes [155, 255, 343, 416].

4.5.7 Construction of Phylogenetic Trees from Molecular Data

The evolutionary relationship between homologous genes or proteins can be represented by a phylogenetic tree. This can be looked upon as a genealogical tree of the molecules but also of the species from which the molecules were obtained. A **molecular phylogenetic tree** describes which nucleotide substitutions or amino acid exchanges have occurred in the course of evolution. This, on the one hand, provides data from which the laws of molecular evolution may be derived and, on the other hand, can be viewed as a natural experiment that allows conclusions to be drawn about the functions of the individual components of the macromolecules. A **species phylogenetic tree** presents special possibilities because macromolecules can be compared more widely than morphological characters, in fact, often right across the whole animal kingdom or across all organisms.

A particular form of molecular relationship arises through gene duplication, which is a common event in molecular evolution. The genes resulting from duplication are, of course, homologous but may become quite different during the course of further evolution. It is therefore necessary in comparing homologous genes or gene products to distinguish between **paralogous genes**, which belong to two different evolutionary lines arising by gene duplication, and **orthologous genes**, which are of the same evolutionary line. Whereas a molecular phylogenetic tree may include both paralogous and orthologous molecules, only orthologous molecules may be used in the construction of a species phylogenetic tree. A confusing situation may arise when only one of two gene lines, arising by duplication, is retained in a particular species. The best-known example of this is the lysozyme of avian eggs; this is very different in the goose and the chicken despite the close relationship of these species. This puzzle was only solved when the enzymes of the goose and the chicken type were discovered together in the black swan, *Cygnus atratus*, presenting evidence for a duplication of the lysozyme gene (p. 504); thus, the lysozymes from eggs of the goose and chicken are paralogous.

Qualitative variation in individual characters may be used for the **construction of molecular phylogenetic trees**, as in classical phylogenetics, or quantitative data, as in numerical taxonomy. **Numerical methods** are preferred in studies of molecular relationships for two reasons: (1) only phylogenetic trees constructed with the help of quantitative methods can be tested statistically; and (2) according to prevalent ideas, the influence of selection can be largely neglected in phylogenetic tree construction using molecular characters, and the rate of evolution may be taken to be constant over long periods (molecular clock hypothesis). Thus, the expression: "the more similar, the more closely related", which is easily refutable for morphological characters, is actually more appropriate for the molecular analysis of relationships. Many of the molecular data are quantitative from the outset, e.g. immunological distance, amino acid composition or ΔT_m values. Whenever essentially qualitative data, e.g. from sequencing, restriction analysis or electrophoresis, are converted to a quantitative measure of

distance, information tends to be lost and events such as parallel or back mutations are difficult both to recognize and to include in the calculations.

The very successful **method of Willi Hennig**, which is based on morphological characters and results in dichotomously branched phylogenetic trees, involves the search for the sibling species or groups sharing the most recent, common ancestor. This requires evidence of the joint possession of derived (apomorphous) characters (synapomorphy). The occurrence of an apomorphous character in only one taxon (autapomorphy) or the joint possession of plesiomorphous characters, i.e. those also found in other taxa (symplesiomorphy), are not relevant to the analysis of relationships. Hennig's method has, up to now, only rarely been applied using molecular characters and then with only modest success, e.g. with electrophoretic data [322] and data on the disulphide bridges of transferrins [437]; it would be interesting to apply the method to sequence data.

With the aid of **numerical methods**, surviving forms can be arranged in a network (Wagner network) according to their degree of similarity. The endpoints constitute the surviving forms, and the internal intersections the ancestral forms (Fig. 4.12). A dichotomously branched network of N surviving forms has N-2 ancestral forms and 2N-3 connecting lines. Such a network can be converted to a phylogenetic tree so that an original form or root is defined on one of the connecting lines (Fig. 4.12); however, this requires further information or assumptions that are not present in the surviving sequences:

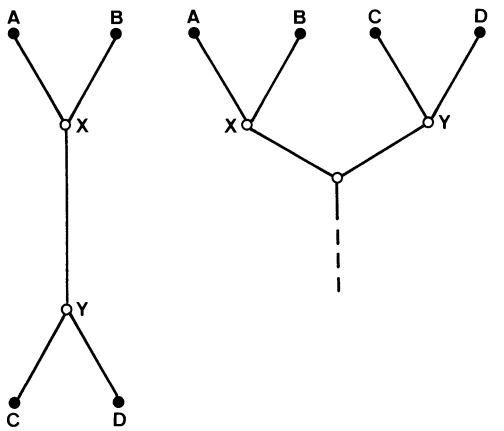


Fig. 4.12. The network of four present-day species *(A–D)* and two ancestral forms *(X* and *Y)*, together with the phylogenetic tree derived by defining a root along the line connecting X and Y

1. The direction of the evolutionary change is known; this is generally not the case for molecular characters.
2. If the molecular clock hypothesis is correct, all routes from the root to the surviving forms are equal in length.
3. If the surviving forms can be reasonably placed in two groups (e.g. animal/plant or globin α-/non-α-chain), the root lies between the two groups.

The **number of possible networks** for N surviving forms is given by

$$\Pi_{n=3}^{N}\,(2n–5) = 1 \cdot 3 \cdot 5 \cdot 7 \ldots (2N - 5). \quad (4.23)$$

This means that for 10 surviving forms there are 2 027 025 different networks, for 20 forms the number is $2.2 \cdot 10^{20}$, and for 40 forms it is $4.5 \cdot 10^{53}$ [325, 424]. From this enormous number of possible networks, it is necessary to find that which is correct or at least optimal; of course, computers are used here, but an increasing number of species can quickly tax the potential of even the most powerful machine. The evaluation of the alternative networks involve several different criteria, of which the most frequently used are goodness of fit, the parsimony principle, and the compatibility principle. In **goodness of fit** [120] the distance between the surviving forms (input) is compared with the total length of the connecting lines in the network (output); an effort is made to minimize the difference between input and output. In the **parsimony principle** (economy principle) an attempt is made to minimize the total length of the network. This principle, which was suggested by Camin and Sokal in 1965 [385], is not based on the assumption that evolution always follows the shortest path; it is much more a question of the methodological principle of considering only that which is necessary for evaluation of the data and accepting only events for which there is evidence. The parsimony criterion is advantageous, for example, for sequence data from which distances in genealogical trees can only be under- and not overestimated; in input/output comparison, on the other hand, is to be preferred if distance data could include errors in both directions, i.e. with immunology or electrophoresis data and ΔT_m values [335]. The **compatibility principle** can also be used for the optimization of phylogenetic trees. Character sets which lead to different phylogenetic trees are incompatible. This can be demonstrated by means of a table of paired events (Fig. 4.13). The compatibility principle serves particularly for the recogni-

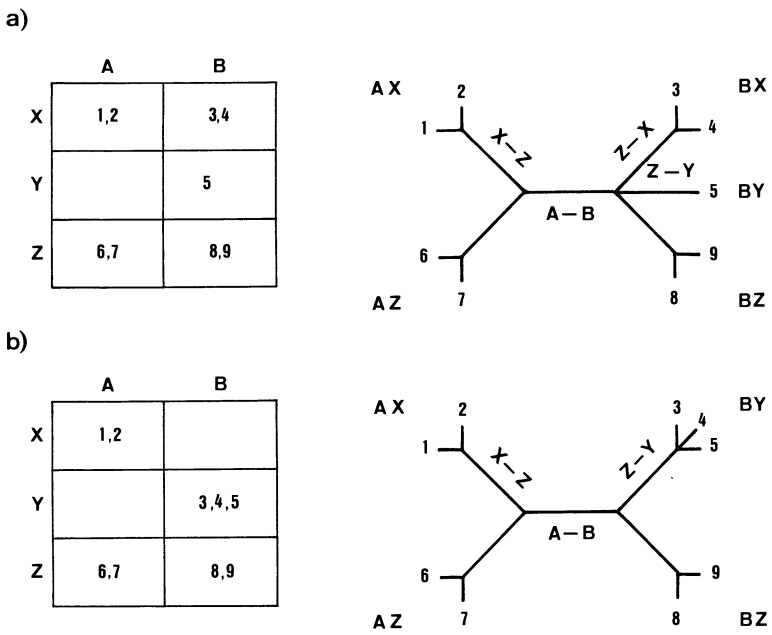


Fig. 4.13 a, b. A test of compatibility [153]. The sets of characters *A* and *B* and *X–Z* occurring in species *1–9* could be **a** incompatible or **b** compatible. To test compatibility, the characters are entered pair-wise into a table *(left)*. The are incompatible when connection of all the entries results in a right-angle **(a)**, and compatible whenever that is not the case **(b)**. The *right* side of the figure shows possible networks and character changes for the nine species. A change of character A to B, or the converse, is shown as *A–B*. In the case of incompatibility of the character sets, at least one change in a character (here *X–Z*) must be assumed to occur twice

tion of parallel substitutions. Pair-wise comparisons with the help of a computer program are used to determine for which species incompatibility appears to be most prevalent. A parallel substitution is then assumed for these species, e.g. that character A was replaced by A′; the procedure is repeated with character A′ and so on until all the parallel substitutions have been localized [153]. In simple cases, the characters found to be compatible can be used directly in the construction of a phylogenetic tree.

The problems of genealogical tree construction from molecular data are especially well illustrated by the following descriptions of classical procedures; these in effect form the starting point for the development of the present-day, more effective computer methods [112, 163, 164]. The **ancestral sequence method** of Eck and Dayhoff (1966) not only connects the surviving sequences to a network but also derives the ancestral sequence. In both steps, the minimization of the network is attempted by means of the following procedure: The three possible networks are drawn for the first four sequences. Here, the surviving sequences form the end-points and the ancestral sequences the internal intersections. To derive the ancestral sequence, amino acids are sought which are present in more than one of the branches arising at the intersection; such amino acids are assumed to be at the intersection (Fig. 4.14). Of the possible networks, the one that requires the least number of alterations (e.g. amino acid exchanges) is chosen. The fifth sequence is now introduced at all possible positions and the shortest network is again chosen; this is continued through to the last sequence. The resulting network is then rearranged in that each branch is removed and transferred to another position (branch exchange or shuffling). Each still shorter network that is found is then the subject of further shuffling. Considerable calculation is required and, furthermore, simulation experiments have shown that with sequences differing by more than 50 %, the construction of phylogenetic trees by use of matrix methods gives better results [87].

Matrix methods are based on a distance matrix. The first large molecular phylogenetic tree was constructed, using a matrix method developed by Fitch and Margoliash in 1967, from 21 cytochrome c sequences [120]; the same method was in fact later used to calculate a better tree from the same data [386]. The Fitch-Margoliash method is still in use today (Fig. 4.15). The "unweighted pair group" method (UPGM), which was invented by Sokal in 1958 as a generally applicable method for numerical taxonomy, is the simplest of the matrix methods. In contrast to the Fitch method, the first step here always involves the combination of the two most similar species (Fig. 4.16). The construction of trees from a Wagner network of minimal total length was described by Farris in 1970 and is also a frequently used method (Fig. 4.17). The subsequent

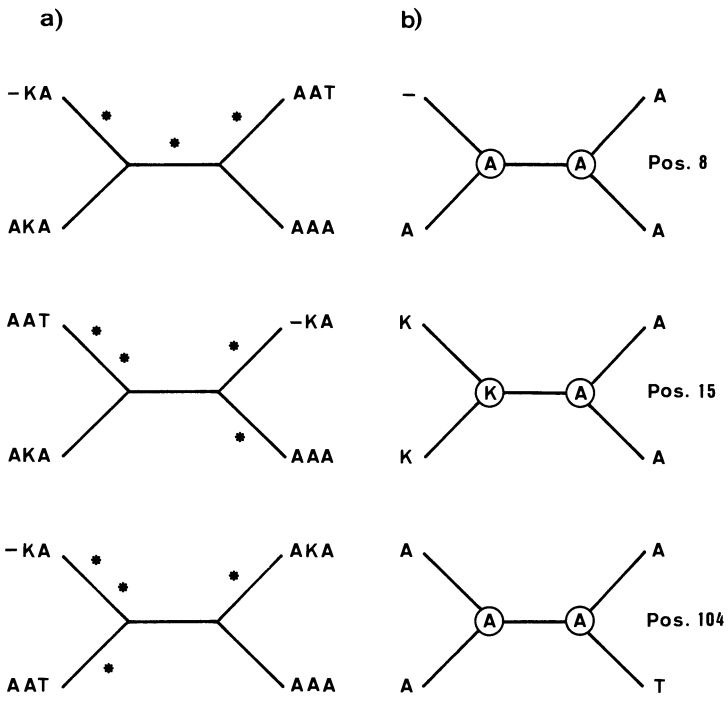


Fig. 4.14a, b. The construction of a phylogenetic tree by the ancestral sequence method of Eck and Dayhoff (1966). The example makes use of amino acid positions 8, 15 and 104 of the cytochrome c of green plants *(AAA)*, fungi *(AAT)*, insects *(AKA)* and vertebrates *(-KA)*. **a** The three possible networks between the four groups. The *uppermost* network requires three and the other two networks four amino acid exchanges (marked with*). **b** Amino acids in the individual positions of the ancestral sequence of the upper network in **a**; any amino acid which occurs in more than one of the lines originating with the ancestral sequence is assumed to belong to that sequence; thus, the ancestral sequence of vertebrates and insects is AKA and that of fungi and higher plants is AAA

development of these mathematical procedures has been aimed at deriving as much phylogenetic information as possible from the molecular data in the shortest possible time [29, 222, 357]. The problem of finding the shortest possible network from a given set of data can only really be solved empirically. There is a whole series of calculation methods (algorithms) which include this search process but none can guarantee that the automatically produced, final phylogenetic tree is optimal in the sense of the criteria described above [357].

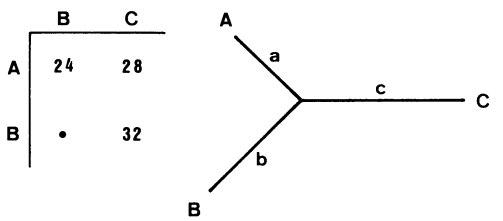


Fig. 4.15. The construction of a phylogenetic tree according to Fitch [120]. From the distances between the sequences *A, B* and *C* (given in the matrix), the length of the branches *a, b* and *c* of the corresponding network, assuming a + b = A/B = 24, a + c = A/C = 28, and b + c = B/C = 32, are a = 10, b = 14 and c = 18. When more than three species are involved, individual sequences such as A and B are compared, in all possible combinations, with the combined, remaining sequences C. Branch lengths are calculated from the mean distances from A and B to all the sequences included in C. The two sequences A and B with the lowest distance are united as (A, B) and the process is repeated. Therefore, all sequences are inserted into the network one after the other in order of decreasing similarity

The **parsimony method**, which was developed in Goodman's group and has been used for the construction of numerous genealogical trees, attempts to construct the tree with the smallest total length and to define the ancestral sequence [141]. Nucleotide sequences are used here, either directly determined or derived from amino acid sequences using the genetic code. A tree is first constructed by the Fitch or the Farris method and is then rearranged by branch exchange until the total number of substitutions no longer decreases. The process is repeated with trees begun in other ways to find those of even smaller total length. Quite often the phylogenetic trees produced in this way contradict in some places the confirmed ideas of relationships between some species. Assumptions are made about gene duplication or changes in gene expression in order to correct for these deficiencies. The tree of minimal total length is that obtained by branch exchange, which assumes the least number of nucleotide substitutions, gene duplications and changes in expression. In the less dense areas of the tree, the length of the branches is underestimated because of the greater probability of multiple substitutions. To correct for this error, an

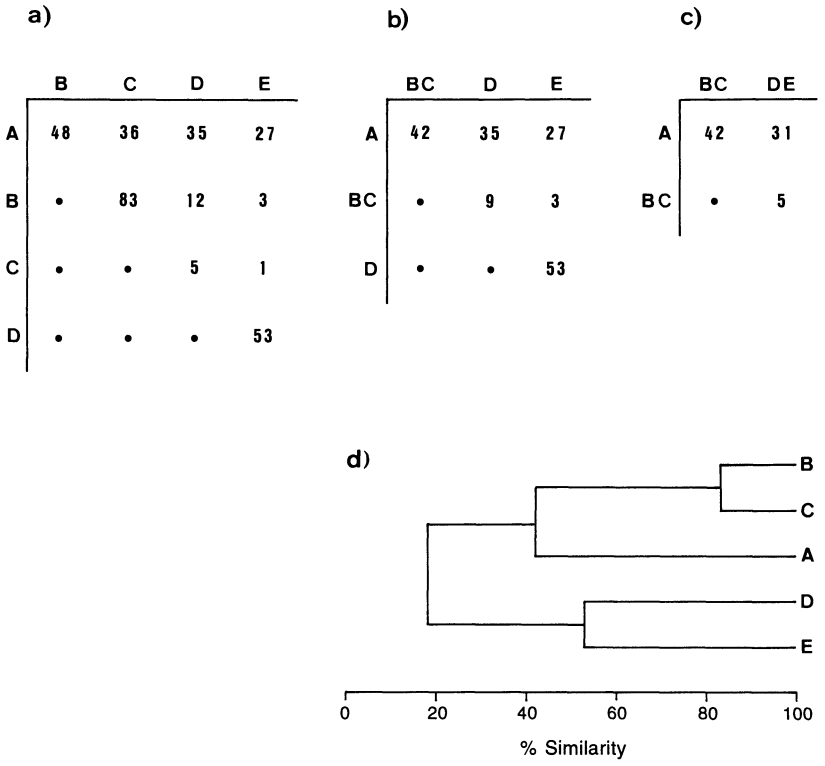


Fig. 4.16a–d. The construction of a phylogenetic tree according to the unweighted pair group method (UPGM) [115]. **a** The method starts with an identity matrix of all species. **b** In the first step, the two most similar species are combined and a new matrix is calculated; thus, the identity value of *(BC)* to another species X is given by (BC)/X = ½(B/X + C/X), e.g. (BC)/A = ½(48 + 36) = 42. **c** In the second step, the next most similar species *D* and *E* are combined and the matrix is calculated. In the third step, A is united with BC; this gives a similarity value (ABC)/(DE) = 18. **d** The dendrogram illustrates the relationships schematically

"augmented distance" is calculated by extrapolating information from denser regions of the tree [141]. Neither the theoretical nor the practical problems of phylogenetic tree construction from molecular data have yet been satisfactorily solved, and new methods are continuously being suggested [405, 412].

The critical **evaluation of molecular phylogenetic trees** requires a test of how far the tree constructed by these methods is correct, i.e. how far

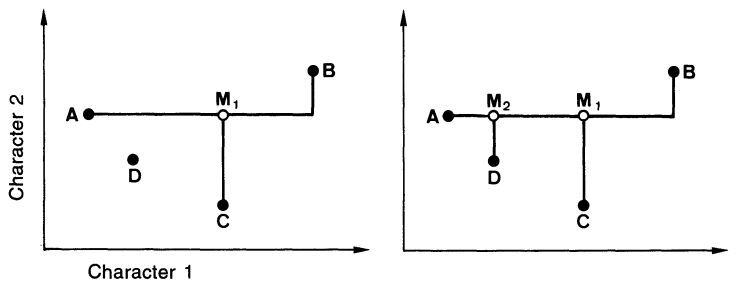


Fig. 4.17. The construction of a phylogenetic tree with the help of a Wagner network [385]. The distance d(J,K) between two species J and K is calculated from the extent of expression X of all the considered characters $J(X_{ij})$ and $K(X_{ik})$ according to

$$d(J,K) = \Sigma_i X_{ij} - X_{ik}. \quad (4.24)$$

The distances between all possible pairs of species are determined in this way. The two species A and B with the greatest distance are chosen. For each of the other species J, the distance to the nearest point M on the connecting line AB is calculated according to

$$d(J,M) = \tfrac{1}{2}[d(A,J) + d(B,J) - d(A,B)]. \quad (4.25)$$

The species C with the largest distance d(C,M) is chosen; the ancestral form M_1 corresponds to the median value of the character expression in A, B and C. Next, species D with the second-largest distance d(D,M) is taken; the distances from D to the connecting lines AM_1, M_1B and M_1C are calculated using Eq. (4.25); M_2 lies on the line with the shortest distance from D. The process is continued until all species are connected in the network

it represents the true phylogenetic relationships between the genes and the corresponding species. Of critical importance to the discussion of the molecular clock hypothesis is the reliability of the conclusions about the evolutionary distance between ancestral forms and their successors that can be drawn from the lengths of tree connections. However, because the real course of the phylogenetic history is unknown, the usefulness of the methods used clearly cannot be judged from the actual relationships. There is also no justification for any assumption that particular characters of the living organism are a priori superior for the construction of phylogenetic history such that the trees based on them could be used as a standard for the evaluation of other methods. Consequently, there is no reason at the outset to assign trees constructed by classical methods, using morphological characters, any special degree of correctness. Each phylogenetic tree represents a hypothesis about the course of evolution, and the plausibility of this hypothesis must be judged against alternative hypotheses. The best possible results from attempts to reconstruct the history of life on earth can actually only be expected when all available data have been evaluated, be they palaeontological, biogeographical, morphological or molecular [318].

One way out of the dilemma that the real course of evolution is unknown and cannot be used to judge the accuracy of the various reconstruction methods is offered by the application of the above methods to sets of data obtained in precisely known ways from **simulated evolution**. Against such a procedure it can be argued that simulation of evolution must involve certain basic assumptions which possibly deviate from the laws operating in nature. Investigations of simulated sequences have so far shown that, depending upon the character of the data, different methods produce the best results, but also that these results, even under favourable conditions, may often be a false branching scheme and involve large errors in distances [112, 358, 390]. On the basis of such investigations, particularly heavy criticism has been levelled against Goodman's parsimony method [180, 190].

In the ideal case, the same phylogenetic tree should always emerge for any given species, irrespective of the method used. In actual fact, quite different trees are often produced when different methods are applied to the same data or the same method is used for different sets of molecular data from the same species [326]. With many methods, several phylogenetic trees of the same length (parsimony) are produced from exactly the same set of data; the introduction of more modern forms changes the branching scheme for the previously considered species, and the family relationships between closely related forms become particularly uncertain. There are various reasons for these deficiencies [153]:

1. The information contained in the available data is insufficient. For example, at least 22 characters are required to define a dichotomously branched family tree of 25 species, and singular (only available for one species) or incompatible characters are not appropriate.
2. A lack of clarity, caused by polymorphism, particularly affects the determination of genetic distance between closely related species [314].
3. The evolutionary rate plays a significant role. Where it is low, there are usually too few evaluable differences; where it is too high, evolutionary events may be superimposed and therefore not distinguishable. A variable rate of evolution complicates the mathematical process.
4. The largest problem is that of parallel substitution which, contrary to previous notions, seems to be very frequent. Tests using the method shown in Fig. 4.13 have shown that sets of molecular data may contain up to 50 % parallel substitutions [153]. This only applies, however, to the evaluation of the individual positions in a sequence; the fear that "almost or completely identical sequences" may be found in distantly related species [350] will probably not be realized.

4.6 The Rate of Molecular Evolution

4.6.1 The Rate of Protein Evolution

The rate of evolution of proteins can be expressed in several ways: (1) as the rate of amino acid exchange per amino acid per year; (2) as the time required for the development of 1 % sequence difference between two evolving lines (unit evolutionary period, UEP); and (3) as "accepted point mutations" (PAM) per 100 amino acids per 100 million years [87, 115, 210, 212]. Individual proteins evolve at very different rates (Table 4.12), with the extremes differing by more than two orders of magnitude. The neutral theory explains these differences, in that only certain

Table 4.12. The rate of evolution of different proteins [440]

Protein	UEP	Exchange rate
Histone H4	400	0.013
Histone H3	330	0.015
Histones H2A and H2B	60	0.08
Glutamate dehydrogenase	55	0.09
Glucagon	43	0.12
Collagen-α1	36	0.14
Corticotropin	24	0.21
Crystallin-αA	22	0.23
Glyceraldehyde-P-dehydrogenase	20	0.25
Triosephosphate isomerase	19	0.26
Lactate dehydrogenase H4	19	0.26
Cytochrome C	15	0.33
Insulin	14	0.36
Lactate dehydrogenase M4	13	0.38
Thyrotropin β-chain	9	0.56
Lipotropin β-chain	8	0.6
Histone H1	8	0.6
Lutropin α-chain	7	0.7
Myoglobin	6	0.8
Trypsinogen	6	0.8
Prolactin	5	1.0
Parvalbumin	5	1.0
Carboanhydrase B	4	1.3
Growth hormone	4	1.3
Haemoglobin α-chain	3.7	1.4
Haemoglobin β-chain	3.3	1.5
Albumin	3	1.7
Lutropin β-chain	3	1.7
Lysozyme	2.5	2.0
Ribonuclease	2.3	2.2
α-Lactalbumin	2.3	2.2
Carboanhydrase C	2.1	2.4
Insulin C peptide	1.9	2.6
Immunoglobulin-C_γ, -C_λ	1.7	2.9
Fibrinopeptide A	1.7	2.9
x-Casein	1.4	3.6
Fibrinopeptide B	1.1	4.5
Immunoglobulin-C_x	0.9	5.6
Short snake neurotoxins	0.8	6.0

UEP (unit evolutionary period), the time (in millions of years) in which a 1 % sequence difference occurs between two evolutionary lines. Exchange rate, the number of exchanges per amino acid per year multiplied by 10^9.

amino acids may be exchanged in a given protein without unduly changing the properties of the protein, and the organism places various demands on each individual protein. For example, the rate of evolution of the C peptide is higher than that of the A and B chains of insulin because the former functions only to bring the future A and B chains together in the correct orientation to each other during folding of the pro-insulin chain. The relatively rapid evolution of serum albumin relates to the fact that this protein, as a whole, is dispensable; analbuminemia in humans causes no clinical symptoms [210, 212, 440]. The lowest rate of evolution is shown by the histones, which are involved in complex interactions with other chromatin proteins and DNA. Surprisingly, amongst the most conserved of proteins is thymidilate synthase, which is required for dTMP synthesis in rapidly dividing cells; the enzymes from mouse and *E. coli* agree in not less than 55 % of their amino acids [327].

The rate of evolution may differ not only between proteins but also between different parts of a protein. So, for example, amino acid exchanges in the globins are 10-fold more frequent on the surface of the molecule than in the region of the haem pockets. Various models of evolution have been derived from the differing substitution rates encountered for individual codons, according to which only a fraction of all codons are available for amino acid-exchanging substitutions, which itself changes at each stage in evolution; this somewhat variably defined fraction has been referred to as "covarions" by Fitch in 1970 and as "varions" by Holmquist and Jukes in 1972. Four classes can be distinguished amongst the 114 codons of cytochrome c: 25 non-variable, 53 variable with an average of 3.7 substitutions, 33 hypervariable with an average of 11.1 substitutions, and the 3 most variable codons, 44, 58 and 100, each with 25–26 substitutions [121]. Differences in the rate of evolution of individual proteins may therefore be explained by their differing contents of covarions: 10 % in cytochrome c, 30 % in the α-globin chain, and almost 100 % in the fibrinopeptides [210].

The relationship between function and evolutionary variability of individual amino acids is especially clear in cases where the exchange of one amino acid is compensated by the exchange of another. For example, there is a stabilizing salt bridge in the myoglobin molecule and in most mammals this consists of a glutamic acid at position 27 and a lysine at position 118. In whale myoglobin, however, the pair 27-Asp ... 118-Arg is found; the greater length of the aspartic acid is compensated exactly by the smaller size of the arginine. The sterically unfavourable combination 27-Asp ... 118-Lys occurs in only 2 out of 43 mammals (the galago and the kangaroo), and the combination 27-Glu ... 118-Arg is never found. In vertebrate myoglobin, the amino acid in position 45 is always basic, whereas that in position 60 is always acidic; in the myoglobin of the snail *Aplysia*, the reverse pairing 45-Asp ... 60-Lys is found [58].

4.6.2 The Rate of Evolution of the Nucleic Acids

The rate of nucleic acid evolution is normally given as the mean number of substitutions per nucleotide per year; for coding sequences, a distinction is often made between the rates of amino acid-exchanging and of synonymous substitutions. According to the neutral theory of evolution, there is a relationship between the fixation rate and the selective value of mutations, and this can be described mathematically [Eq. (4.8), p. 141]. Thus, the substitution rate increases with the proportion p_n of mutations which behave as selectively neutral. When all mutations have no influence on fitness ($p_n = 1$), the rate of evolution reaches a maximum which is equal to the mutation rate. The rate of amino acid-exchanging substitutions differs between different genes by more than two orders of magnitude, as has already been shown by comparison of the proteins. The rate of synonymous substitution is approximately the same for all genes and is always significantly higher than that of amino acid-exchanging substitution; most recorded values are around $5 \cdot 10^{-9}$ per nucleotide per year, with extremes of $1.3 \cdot 10^{-9}$ and $7.4 \cdot 10^{-9}$. The variability in the rate of synonymous substitution is at least partly due to codon preference; the more the use of the various synonymous codons deviates from random, the lower is the substitution rate [38, 210, 212, 277]. If the rate remains constant, the number of synonymous substitutions that can be detected on comparing two sequences will eventually reach a saturation value; for the actin genes this occurred after 30–40 million years. Synonymous substitutions may be used only for estimations of phylogenetic relationships and the evolution rate before the point of saturation [4].

The frequency of mutation is actually quite similar in all parts of the genome, but the probability of fixation of a mutation depends upon its consequences; thus, the changes (substitutions) that become fixed during evolution are not stochastically distributed throughout the genome. In coding sequences, there are large differences in substitution rate between the three codon positions. For example, the distribution of substitutions between the three positions in the globin genes of man, the mouse and the rabbit is 24:21:55, showing a higher proportion of synonymous mutations in the third position [210, 212]. In the non-coding sequences of genes there are also characteristic differences in substitution rate in different regions which are related to function; similar tendencies are found in different genes (Table 4.13 and Fig. 4.5). The substitution rate in the introns and in the 5′ section of non-coding 3′ regions [3′ non-translated (NT) regions] corresponds approximately to the rate of synonymous substitution. In contrast, the substitution rate in 5′-NT regions and the 3′ sections of 3′-NT regions has only about half that value; these regions include important signals for transcription and mRNA maturation. Conservative evolution of 3′-NT regions with relatively low substitution rates has been detected, for example, in the genes of isotypic actins and tubulins of various species and in the genes for human and mouse epithelial growth factors; it is much more pronounced here than in the globin genes [450]. The observation that the substitution rate is lower in smaller than in larger introns is due to differences in the degree of restriction. Sequence comparisons between homologous genes of closely related species of *Drosophila* have revealed large differences in the substitution rate between different introns or neighbouring sections of non-coding gene regions [101, 259]. Substitutions in pseudogenes should be completely selection neutral ($p_n = 1$); the substitution rate of $12.6 \cdot 10^{-9}$ per nucleotide per year calculated for the globin pseudogene should correspond to the mutation rate. The rate of synonymous substitutions in globin genes is 1.7- to 1.9-fold lower, suggesting that synonymous mutations are also subject to selection pressure [210, 212]. Comparison of the Adh loci of *Drosophila pseudoobscura* and *D. mauritiana* also shows that not only amino acid-exchanging but also synonymous substitutions

Table 4.13. The percentage sequence differences between gene regions of various gene pairs [209]

Regions	1	2	3	4
Coding				
amino acid exchanging	8	19	11	15
synonymous	48	51	49	30
Introns				
small	40	31	–	–
large	49	53	–	–
5′ non-coding	32	23	24	6
3′ non-coding				
5′ section	44	48	50	21
3′ section	19	23	22	22

1, prepro-insulin genes man/rat; 2, β-globin genes rabbit/mouse; 3, α-globin genes rabbit/mouse; 4, α-globin gene mouse/ψα-pseudogene mouse

are unevenly distributed between the exons [364]. The chorion genes of Hawaiian *Drosophila* species show extremely high rates of evolution. The amino acid exchange rate of $21 \cdot 10^{-9}$ per site per year is the highest yet recorded (Table 4.12). The rate of amino acid-exchanging substitutions is twice that of synonymous substitutions [261].

Extremely high rates of synonymous substitution have been observed in vertebrate **mtDNAs**. The mtDNAs of various mammals, with synonymous substitution rates of $35 \cdot 10^{-9}$ to $94 \cdot 10^{-9}$, have a rate of evolution 10–20 times greater than that of chromosomal DNA; similar results have been obtained with the mtDNAs of different *Xenopus* species [57, 360]. There is, as yet, no explanation for these differences in rates of synonymous substitution, except that in the vertebrates mtDNA possibly has a higher mutation rate than chromosomal DNA. The results from vertebrates, however, may not be used as a basis for generalization. In various *Drosophila* species, the rate of mtDNA evolution is at most only twofold higher than that of chromosomal DNA [332, 360, 373]; in some sea urchins, the mtDNA substitution rate is just as high as in the vertebrates but that of single-copy DNA is also high, i.e. about five times the rate of the mammals [422]. The mtDNA of higher plants evolves very slowly, the substitution rate being 100-times lower than that of vertebrate mtDNA and three- to fourfold lower than chloroplast DNA [312].

4.6.3 Is There a Molecular Clock?

Based on the random character of mutation, Zuckerkandl and Pauling expressed, at the beginning of the 1960s, the idea that the sequence of amino acid exchanges in a line of evolution must be so regular that it could be used as a molecular clock. The sequence differences between homologous polypeptides would then allow conclusions to be drawn about the time that has elapsed since their separation. This concept has subsequently been applied to evolutionary distance derived from data other than molecular data and has proved to be very profitable. On the occasion of the 25th anniversary of the idea, the *Journal of Molecular Evolution* devoted the whole of a special issue to the molecular clock [200, 236]. In order to apply the principle of the clock, the relevant molecular data are used to construct a phylogenetic tree, the branch lengths of which reflect the evolutionary distance. Then, with the help of the fossil record, the time from at least one branch fork to the present is determined and related to the evolutionary distance. The molecular clock is thus calibrated, and all evolutionary distances in that particular tree can be converted to units of time. This process includes two obvious and significant **sources of error**. One is to be found in the separation time used for the calibration. Although the age of fossils can now be determined with great accuracy, it is usually the case that no available fossil corresponds exactly to any branch point in the tree being investigated. The second source of error involves the problems, which have already been mentioned, of the estimation of evolutionary distance and the construction of family trees from molecular data. Clearly, only orthologous and not paralogous sequences may be used for age determination. Statistical methods may be used to estimate the range of error for the branch points of the tree [295]. In view of the variable evolution rates determined within the mammalian phylogenetic tree, the question has been discussed of whether the number of generations, rather than absolute time, should be used as the time scale of evolution; in general, however, molecular evolution is proportional to time [102, 440, 447].

The largest **amount of data**, related to the molecular clock, that has so far been analysed consists of the thousands of immunological distances determined by microcomplement fixation of serum albumins. The calibration of the "albumin clock" gives 1.7–1.9 units of immunological distance per million years [440]. Other sets of molecular data obtained by other methods have shown constant rates of evolution corresponding to a molecular clock, e.g. immunological data from turtles, crocodiles and other vertebrates, and hybrid DNA melting points (ΔT_m values) from rodents [43, 88, 303]. Convincing arguments for the functioning of the molecular clock may be found in evolutionary events that are clearly defined in terms of time, as, for example, the separation of sibling species between the Pacific and Atlantic sides of the isthmus of Panama. According to Nei, the genetic distance, calculated from electrophoretic data, of such fish species, had a value of 0.234; this would indicate a separation time of about 3.5 million years, which fits quite well to the geologically determined age of the isthmus. The mean immunological distances for the serum albumins of 103 and 114 for marsupials and tree frogs, respectively, from South America and Australia, given 1.7 units per million years, suggest a separation time of 70 million

years; this also agrees with the geological data on the separation of the two continents.

The relative constancy of the rate of evolution, together with the great extent of genetic polymorphism, was one of the main arguments in the development of the neutral theory of evolution. In connection with **the controversy between neutralism and selectionism**, the concept of a molecular clock was, therefore, questioned from the very beginning. Criticism was also levelled at certain results obtained using the molecular clock approach, in particular those related to the evolution of man and the anthropoids (p. 169). As arguments against the claimed constancy in the rates of molecular evolution, examples have been quoted where varying rates were observed in different evolutionary lines or at different moments in the evolution of a line. Up to fivefold differences in the rates of synonymous substitution were cited for different animal groups, with the lowest value of about $1.3 \cdot 10^{-9}$ for the primates and birds, and the highest value of about $6.6 \cdot 10^{-9}$ for the rodents, sea urchins and *Drosophila* [38, 59, 373]. The rates of synonymous substitution can vary by as much as twofold within a species [373].

There are also many results showing variable rates of evolution within an evolutionary line. For example, the molecular phylogenetic trees constructed by Goodman and colleagues by use of the maximum parsimony method, consistently show strongly variable rates of amino acid exchange, with an extreme difference of 1:32 [141, 142]. Kimura, however, challenged these results, quoting deficiencies in the maximum parsimony method and errors in the choice of separation time used for calibration of the phylogenetic trees [210, 212]. Goodman also found that individual proteins were not usable measures of time, but considered that using the combined data from several proteins, "the hypothesized protein clock does not perform too badly" [141]. A comparison of the globin genes of sheep and goats showed that the rates of amino acid-exchanging substitutions increased after each gene duplication; this increase is especially distinct in functionally less critical parts of the sequence, and is therefore related more to the reduced effect of eliminating selection than to an increase in positive selection pressure [242]. An increased rate of evolution due to positive selection is shown by the ruminant lysozymes, which have assumed the function of a digestive enzyme [194]. A variation in the substitution rate may therefore be explained by changes in the fraction of nucleotide positions that are variable without negative consequences during different periods of evolution [313]. The rate of substitution is dependent upon the base composition of the DNA region (isochore) and is lower in GC-rich regions [444]. Thus, constant rates of substitution are only to be expected so long as the gene remains in a stable state, i.e. the base composition of its surroundings does not change. Non-stable genes show significantly higher rates of substitution [356].

Because the molecular clock is a **probabilistic** and not a **metronomic clock**, there is already theoretically a minimum standard deviation of size $\sqrt{M}$, where M is the number of evolutionary events used for measuring time, e.g. the number of amino acid exchanges or nucleotide substitutions. In reality, the standard deviation of the molecular clock is found to be two- to fourfold larger [210, 212, 440]. This increased variability is also compatible with the neutral theories [404]. First, the rare, significantly selection-positive substitutions will be fixed at an increased rate. Second, in the course of, and as the result of, the evolution of a protein, the proportion p_n of the neutral evolutionary changes can vary. Finally, the further developed neutral theory assumes that whether a certain substitution behaves neutrally is dependent upon the population size N_e (p. 140); changes in the population size, which of course occur frequently during evolution, will therefore lead to changes in the rate of evolution [210, 212]. Because, according to the neutral theory, the greater and more variable the influence of selection, the greater will be the variation in the rate of molecular evolution, synonymous substitutions confer a particular reliability on a molecular clock; however, due to their high rates they may be used only in the most recent phases of evolution.

4.7 Some Results of Molecular Research into Evolutionary Relationships

Biochemical characters vary in their suitability for the different tasks of biological systematics. They are not basically superior to other characters for species diagnosis, i.e. for determining whether an animal belongs to a particular species or subspecies, but are often the last resort if, for example, insufficient morphological characters are available. It is mostly only possible with the help of molecular characters to recognize whether

morphologically indistinguishable groups of animals living in the same biotope are sexually isolated, i.e. represent "true" species (sibling species), or whether gene flow exists between given species (hybrid species). Enzyme electrophoresis allows the process of species formation (speciation) to be followed directly. The initial expectation that molecular analysis of relationships would bring a breakthrough in systematics and, for example, the phylogenetic relationships between the animal phyla would be clearly defined, has not proved to be the case. Nevertheless, compared with morphological characters, proteins and nucleic acids possess certain advantages for the analysis of relationships and the reconstruction of phylogeny (p. 111). Furthermore, the comparison of molecular characters has provided insights into the early phases of evolution; these are amongst the most fantastic results of molecular biology but can only be listed, and not dealt with in depth in a book on the comparative biochemistry of animals. They are as follows:

1. The mitochondria and plastids of the eukaryotes have their origins in symbiotic prokaryotes.
2. In addition to the two classical organism kingdoms of the (Eu-)bacteria and the Eukaryotes, there is at least one more kingdom, that of the Archaebacteria.
3. The unicellular organisms possess a heterogeneity that had been unimaginable before the era of molecular biology and which cannot be described by the traditional concepts of fungus, plant or animal.

In the following, a few examples will serve to illustrate the different possibilities for the molecular analysis of evolutionary relationships.

4.7.1 Molecular Phylogenetic Trees

Amino acid sequences have been more-or-less systematically collected since the 1960s and used for the construction of phylogenetic trees; particularly extensive comparative material is available for the globins, cytochrome c, α-crystallin, immunoglobulins, pancreas RNases and fibrinopeptides. The knowledge of phylogenetic development gained in this way will be presented in each case by the individual proteins. Phylogenetic considerations must, of course, be restricted to the evolution within single protein super-families; how these came into being in the early phases of life on earth is at present beyond our knowledge. The analysis of DNA sequences, which began in the 1980s and has constantly increased in intensity, has already made available, in terms of the total length of analysed sequences, more extensive material than protein sequencing did; however, comparative data on homologous sequences from a large number of species is required for the analysis of relationships and, as yet, is only available for the globin genes as well as for the rRNAs and the tRNAs and their genes.

4.7.2 Species Systematics

For **species diagnosis**, i.e. the assignment of single individuals to a certain species or a taxon below the species level, biochemical characters can be applied just as well as morphological or other characters. Here also, macromolecular materials appear particularly suitable because of their high information content, but low molecular weight substances are also used, although more frequently in botanical than in zoological systematics. The substance used need not necessarily be identified chemically, e.g. unidentified spots on a chromatogram can also have taxonomic value so long as they are only "diagnostic", i.e. exclusively present in the members of a particular species and preferably also in all individuals. Because, in general, the analysis of biochemical characters requires a greater technical effort, they are usually only taken into account for species diagnosis when suitable morphological features are unavailable. This applies in particular to animal groups, like amoebae, trypanosomes, sporozoans or nematodes, that are relatively poor in characters, to eggs and early developmental stages, and, finally, to isolated tissues whose origin is of interest, for example, in the fields of foodstuffs or crime. Numerous such examples are presented in the book by Ferguson [115], and recent investigations have been carried out, for example, on the protozoans and the muscle tissue of fish [27, 432].

Even in animals that are rich in characters, species may be so similar to each other that they are indistinguishable using morphological features alone **(sibling species)**. In such cases, the method of enzyme electrophoresis has been used with particular success, e.g. to subdivide several groups of sibling species of *Drosophila* [25, 90], and for the separation of the brachyuran *Uca speciosa* and *U. spinicarpa* [359], the nematodes *Caenorhabditis elegans* and *C. briggsae* [51], the sea

urchins *Echinostreptus aciculatus* and *E. molaris* [265], the cephalopods *Eledone massyae* and *E. gaucha* [235], the New Zealand species of mackerel *Trachurus declivis* and *T. novaezelandiae* [128], and the field-mouse species *Microtus arvalis* and *M. subarvalis* [453]. The ciliates of the genus *Tetrahymena* can be divided into a complex of more than 30 sibling species on the basis of their rRNA sequences [336]. Species diagnosis in the disease-transmitting mosquitoes of the genus *Anopheles*, which are morphologically difficult to distinguish, has made use of electrophoretic methods as well as gas chromatographic analysis of the cuticular hydrocarbons [56].

Of particular interest are cases in which apparently true species were first recognized after electrophoretic analysis as being pairs or groups of sibling species. For example, the marine polychaete *Capitella capitata*, which as an indicator of water pollution is of practical importance, turned out to be really a complex of at least six sibling species [147]. The Hawaiian teleost *Albula virgata* could, by electrophoretic analysis, be separated into two completely isolated populations, for which only later diagnostic morphological differences were found [371]. The leopard frog *Rana pipiens*, used in numerous experimental investigations, is apparently a complex of at least eight sibling species which mostly live in separate habitats (allopatric) and, in addition to possessing electrophoretic differences, are also distinguishable by their mating calls [174]. Enzyme electrophoresis can also help to clarify the taxonomic situation when morphologically quite different individuals are mistakenly assigned to different species. An example here is provided by the ant species *Formica rufa, F. polyctena* and *F. pratensis*, which all have the same pattern of 68 protein bands and 18 esterase bands. The morphological variability in this case is so great that even animals from the same nest appear to belong to different species [408].

A further interesting biological-taxonomic application of biochemical characters is the **identification of hybrid species**. Because, in contrast to most morphological characters, molecular features are inherited co-dominantly, i.e. both alleles are always recognizable, hybrids can be clearly recognized by electrophoretic investigation. Crosses between true species are rarely found in the field, even when they are easy to achieve in the laboratory. Thus, at least 159 different hybrid species of *Drosophila* have so far been bred in the laboratory, although only eight cases are known in nature [152]. In most cases hybrid species are sterile or, at least, show reduced fertility. In spite of this, naturally occurring hybrids have been detected for a whole series of species pairs using electrophoretic methods. It turns out that on average 0.4%, and locally up to 5.9%, of the "salmon" caught in Irish rivers and lakes are hybrids between the salmon species *Salmo salar* and the trout *S. trutta* [82]. Hybrids occur especially in the overlap zone between neighbouring habitats of two species as, for example, in the case of the toads *Bufo boreas* and *B. punctatus* in California, and *Bombina bombina* and *B. variegata* in Poland, and the marine turtles *Eretmochelys imbricata* and *Caretta caretta* on the coast of Brazil [77, 110]. Hybrid species are more frequent in fish and amphibians than in the reptiles, birds and mammals. This is at least in part related to the different rates of morphological and molecular evolution (p. 129). Naturally occurring hybrid species are also known in the invertebrates, e.g. between the ants *Solenopsis geminata* and *S. xyloni* in Texas, or between the mussels *Mytilus edulis* and *M. galloprovincialis* on the English coast [184, 380]. Molecular methods may also allow the correction of false conclusions, which are based on mistakes in the interpretation of morphological data, relating to hybrid species. For example, it was previously claimed that up to 29% of the populations of the darter (perch) *Etheostoma spectabile* and *E. caeruleum*, living in the same habitat (sympatric), were hybrid species. Tests with enzyme electrophoresis, however, gave no indication of hybrids but showed instead that single individuals of *E. caeruleum* could be easily confused visually with *E. spectabile* [271].

By far the most interesting example of naturally occurring hybrid species is that of the European **aquatic** or **green frogs**. In the course of the evolution of these species, special genetic mechanisms arose by which the normal problems of chromosome pairing during meiosis in hybrids were overcome. The central European aquatic frog *Rana esculenta*, which was originally described by Linnaeus and has since been used in innumerable physiological investigations, is really a hybrid between the species *R. ridibunda* and *R. lessonae;* this has been shown in various investigations since 1968, above all by the electrophoretic examination of serum albumins and various enzymes and also by restriction analysis of mtDNA. In southern France, *R. ridibunda* hybridized with *R. perezi*, and in Italy and the Balkans it hybridized with *lessonae*-like species; hybrids between *R. nigromaculata* and *R. brevipoda* have been found in Japan [392]. The central

European populations consist mostly of *R. lessonae* and *R. esculenta*, but both parental species never occur together. The hybrid species known as *R. esculenta* shows increased vigour compared with the parental species, e.g. it has a greater resistance to oxygen shortage at lower temperatures. Selfing *R. esculenta* results in progeny with defective meiosis, reduced fertility and increased lethality; in contrast, crosses of *R. esculenta* x *R. lessonae* give normal *R. esculenta*. Furthermore, electrophoretic investigations of oocytes show that most of the *lessonae* genome is eliminated during gametogenesis in *R. esculenta* and only the *ridibunda* genome is passed on to the progeny. The same is true for the populations of the green frog in southern France and Italy; however, the mechanism by which the non-*ridibunda* genome is eliminated has not been fully explained. Only about 7 % of hybrids are found in the Balkan populations and these are apparently hardly fertile; in this case the mechanisms of species hybridization appear to be not so far advanced [392]. The analysis of mtDNA provides more detailed information about the pairing behaviour of green frog populations. As a rule, a male pairs with a larger female; the average body size increases in the order *R. lessonae* < *R. esculenta* < *R. ridibunda*. Pairing of an *R. lessonae* male with an *R. esculenta* female results in an *R. esculenta* individual with *ridibunda* mtDNA; the rarer pairing of a large *R. lessonae* female with a small *R. esculenta* male produces *R. esculenta* with *lessonae* mtDNA. There are two types of mtDNA in *R. ridibunda:* the A form shows 8 % difference to *lessonae* mtDNA and is apparently typical for *ridibunda;* the B form is only 0.3 % different to *lessonae* mtDNA and probably found its way into the *R. ridibunda* population via occasional crosses between *R. esculenta* females and *R. ridibunda males* [392].

Molecular data also provide answers to the interesting question of the extent of the genetic changes that occur during **species formation (speciation)**. In this respect, electrophoretic methods have in particular been used to determine the genetic distance between taxa of different rank that represent successive stages in speciation. Especially instructive data are available for flies of the American *Drosophila willistoni* group and for mice from the genus *Peromyscus* (Table 4.14). The rule derived here, that the distance increases with increasing taxonomic rank, has since been confirmed by numerous investigations, especially with all classes of vertebrates [15, 16]; however, an overview of all the available data shows a marked overlap of the distances between the different taxonomic ranks (Fig. 4.18). In the vertebrates, the lowest species distances are found amongst the birds and the largest amongst the amphibians (Table 4.15). This results from the determination of taxon rank, e.g. the classification into different subspecies, species, etc., mainly on the basis of morphological differences

Table 4.14. The genetic distances, according to Nei, between taxa of different levels in the *Drosophila willistoni* group [16] and the rodent genus *Peromyscus* [456]

	Drosophila	Peromyscus
Populations	0.031 ± 0.007	0.03 ± 0.014
Subspecies	0.230 ± 0.016	0.052 ± 0.014
Semi-species	0.226 ± 0.033	0.178
Sibling species	1.056 ± 0.068	0.334 ± 0.043
Easily distinguishable species	1.056 ± 0.068	0.334 ± 0.043

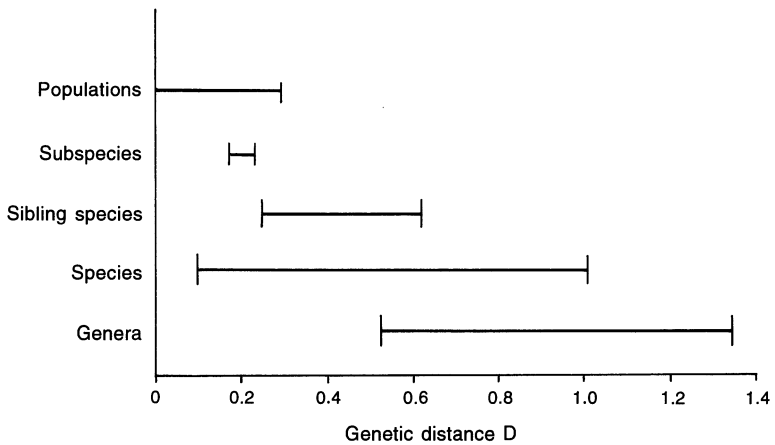


Fig. 4.18. The average genetic distances, according to Nei, between taxa of different rank [115]

Table 4.15. The average genetic distances, according to Nei, between species of a genus and species of different genera in the classes of vertebrates, weighted according to the number of pair-wise comparisons [15]

Class	Genetic distance between	
	species	genera
Osteichthyes	0.60	0.76
Amphibia	1.75	–
Reptilia	0.67	–
Aves	0.08	0.24
Mammalia	0.41	1.10

which are not strongly correlated to the magnitude of molecular differences (p. 129).

The increase in evolutionary distance during the course of species formation is also clearly shown by restriction analysis of mtDNA: in North American rodents, the sequence differences between populations of the species *Geomys pinetis* amount to 0–4.7 %, between the sibling species *Peromyscus polionotus* and *P. maniculatus* to 15 %, and between the morphologically quite distinct species of the genus *Peromyscus* to consistently more than 20 % [225]. The mtDNA differences in local races of the house mouse, *Mus musculus*, are at most 0.4 %, and between subspecies are up to 7.1 % [452]. The relationship described above allows testing of the taxonomic rank in critical cases. For example, in mussel taxonomy it was controversial whether two species groups should be designated as the independent genera of *Musculium* and *Sphaerium;* the D values between the two groups of 1.10–2.11 are much higher than those within the groups and support the classification as genera [181]. In contrast, the butterflies *Brenthis daphne* and *B. ino*, which exist in spatially separate populations on different forage plants, have such a low genetic distance, with D = 0.048, that doubts about their status as species are quite justified [264].

With the aid of molecular data it is also possible to reconstruct, in some appropriate cases, the evolutionary events in the **history of individual species**. Because of its purely maternal inheritance the mtDNA provides hints about the ancestors of parthenogenetic species, e.g. in the fish genus *Menidia* or in the lizard genus *Cnemidophorus* [89, 103]. Restriction analysis of mtDNA has also shown that the inbred line of mice *Mus musculus*, now widely distributed throughout many laboratories, is derived from a single female. Fifteen different mtDNA types were found in 20 examples of this species captured in the wild; in contrast, BALB/c, C57BL/6 and seven further established inbred lines all have the same mtDNA, whereas new lines bred from wild animals have variant mtDNAs. It is possible that with time different wild males were crossed-in so that the gene pool of the chromosomal DNA could, nevertheless, be very extensive [118]. In another example, comparative investigations of the mtDNA of different races of the domestic pig provide evidence for the introduction of various European as well as Asiatic wild-pig females [429]. The brown rat, *Rattus norvegicus*, is derived from the house rat, *R. rattus*, which, however, occurs worldwide in several subspecies. The highest similarity of mtDNA (an mtDNA difference of 2.3 % and the same chromosome number of 2n = 44) is found between the brown rat and the Asiatic house rat, *R. rattus flaviceps*, whereas other subspecies of *R. rattus* show 8–9 % mtDNA difference to the brown rat and have variant chromosome numbers [161]. Molecular data can also give information on the history of human populations. Protein and DNA polymorphisms show that the first phylogenetic branching was the separation of African and non-African populations, and that the non-African populations then split into two clusters, from one of which there arose the Caucasian, East Asian, Arctic and native American populations, and from the other arose the populations of Southeast Asia, the Pacific islands, New Guinea and Australasia. Linguistic data are in good agreement with this scheme [60].

4.7.3 Molecular Taxonomy Above the Species Level

Whilst only sequence data are suitable for the construction of molecular phylogenetic trees, all the other types of data mentioned above can be used for **species phylogenetic trees**. As the reliability of the tree should increase with the information content of the data, species phylogenetic trees are often constructed using the combined data from several proteins. Such trees have, of course, also been constructed for invertebrates, especially for the more than 1000 species of the genus *Drosophila* [28]. However, the vertebrates, and especially humans, are particularly suitable for demonstrating the possibilities and limitations of molecular taxonomy.

Related to humans **(hominoids)** are the chimpanzee *(Pan troglodytes)*, the dwarf chimpanzee *(P. paniscus)*, the gorilla *(Gorilla gorilla)*, the

orang-utan *(Pongo pygmaeus)* and the seven species of gibbons *(Hylobates sp.)*. In order to clarify the family relationships between man and the various anthropoids, various sets of molecular data have been applied, in particular DNA sequences from the η-globin pseudogenes, ε-immunoglobulin pseudogenes, various non-coding DNA regions and mtDNAs, ΔT_m values, restriction data, amino acid sequences, and immunological and electrophoretic data. All the molecular family trees first separate the gibbons and the orang-utan from the evolutionary line. The branching scheme between man, chimpanzee and gorilla, however, remains controversial; the difference in time between the splitting off of the two anthropoids is apparently so short that the molecular data do not allow a statistically certain conclusion to be drawn about the sequence of events. The majority of authors assume a closer relationship between man and the chimpanzee, with the branch point between them given as 4–6 million years ago, depending upon which data are used; the dwarf chimpanzee arose later. This conclusion has now been accepted by the palaeontologists, who initially proposed a much greater phylogenetic age for man [54, 140, 143, 160, 180, 354, 417, 438]. Man and the chimpanzee together present a particularly instructive example of the fact that the rate of evolution of complex phenotypic characters is not strongly correlated with the substitution rate for DNA, and that the evolution of complex features cannot simply be explained by gradual, randomly distributed changes in DNA sequences. It only became possible to solve the puzzle of the taxonomic position of the **giant panda or bamboo bear** *(Ailuropoda melanoleuca)* with the use of molecular data; this had not been possible with morphological and palaeontological data alone. Does the giant panda belong to the family of bears (Ursidae) or the racoons (Procyonidae), or should it be assigned to a unique family (Ailuropodidae)? What exactly is its relationship to the lesser panda *(Ailurus fulgens)*? DNA hybridization data, electrophoretically determined D values, and the immunological distance of the albumins and transferrins clearly placed the giant panda close to the brown bears and the lesser panda with the raccoons. However, it is not yet decided whether the giant panda should be classified as a separate family or as a subfamily of the Ursidae [302]. From the sequences of α-crystallin and α- and β-globin, it was concluded that the Procyonidae is a sibling group of the Mustelidae (martens) and should not, as has often been attempted, be placed close to the Ursidae and Canidae (canines) [197]. The phylogenetic origin of the South African aardvark *Orycteropus afer* was previously a complete puzzle; however, a family tree constructed from data on the lens protein α-crystallin indicated as close relatives the manatee (Sirenia), hyrax (Hyracoidae) and elephant (Proboscoidae) [195]. From the immunological data on albumin for 34 of the 37 existing cat species (Felidae), a family tree may be constructed that coincides with the karyotype data and the fossil record: about 12 million years ago the small South American species, such as the ocelot, first separated; 8–10 million years ago the relatives of the domestic cat separated; 4–6 million years ago the cheetah, serval and puma diverged; and the lines to the lynx and the *Panthera* species of the lion and tiger separated only about 2 million years ago [76]. Restriction analysis of the mtDNA of the seven species of the Equidae points to there having been a common ancestor about 3.9 million years ago [130].

Immunological investigations on various egg and serum proteins have shown that the southern and central American curassow (Cracidae) are not more closely related to the chicken (Galliformes) than to the duck (Anseriformes) and should not be placed in the Galliformes family; these methods have also indicated that the **penguin**, whose evolution was hitherto a complete mystery, is apparently quite closely related to the grebes (Gaviiformes and Podicipediformes), the petrel (Procellariiformes), and the bawler (Ciconiiformes) [333]. The urodelan family **Plethodontidae**, with 23 genera and more than 200 species, is widely distributed in the New World; only two species from this group, *Hydromantes italicus* and *H. genei* are found in southern Europe. Enzyme electrophoresis and the immunological distances of the serum albumins show that these two species migrated from North America in the Oligocene and further developed in isolation [425]. Immunological albumin comparisons carried out on numerous species of the **tree frog (Hylidae)** have assisted in the clarification of a whole series of phylogenetic-taxonomic questions. In this way, it was shown that the morphological similarity of *Hyla regilla* to other species of the *Hyla* genus is a case of convergence, that *Anotheca spinosa* belongs to the Hylinae rather than to the pouched frogs (Amphignathodontinae), and finally that the branch frog *Phyllomedusa* is not more closely related to *Hyla* than to *Bufo* and should, therefore, not be counted amongst the Hylidae but placed in its own family [266]. The approximately

200 species of speckled perch (Cichlidae), which are found only in the 1-million-year-old Victoria Lake in Africa, are monophyletic according to their mtDNA sequences, i.e. they originate from a single common ancestral species [273]. Molecular data have also helped to solve the problem of the roots of the terrestrial vertebrates. Impressed by the discovery of *Latimeria chalumnae* as the surviving representative of the lobe-finned fish (Crossopterygii), it was thought likely that here was the origin of the tetrapods. However, the fact that the mtDNA of *Xenopus* is more similar to that of the three lungfish species than is that of *Latimeria* speaks more for the Dipnoi as the forerunners of the tetrapods [272].

4.7.4 Molecular Data and the Large-Scale Classification of Organisms

Molecular data provide evidence that all present living organisms have a common origin, but they do not allow unambiguous conclusions to be drawn about the genealogical relationships of the various groups of organisms. Protein phylogenetic trees may, in many cases, represent authentic phylogenetic relationships between higher eukaryotes; however, several authors doubt that these may be expanded to include the prokaryotes. It may well be the case that evolving amino acid sequences come up against the limits of difference that are set by structural-functional restrictions and no longer represent evolutionary distance [274]. The many attempts to evaluate the nucleotide sequences of the 5S rRNAs in a broad taxonomic frame are of questionable worth, because such small rRNAs contain too little information [157]. More suitable are the rRNAs of the small ribosomal subunit (SS-rRNAs), the sequences of which are now known completely or in part for about 400 organisms [138, 421]. Phylogenetic trees constructed from these data allow the three lines of the Archebacteria, Eubacteria and Eukaryotes to be distinguished and indicate the path from the (various) Eubacteria to the mitochondria and chloroplasts [311].

However, the attempts to use these data in the construction of a clear branching scheme that leads to the different groups of the eukaryotes [65, 119] are not convincing. The **evolution of the eukaryotes** appears at present not to be a tree whose branches arise from the stem at clearly defined intervals but a bush in which the origins of the individual branches from the common roots cannot be defined. The classical division of the organisms overvalued the morphological diversity of the multicellular organisms and undervalued the molecular diversity of the unicellular eukaryotes and the prokaryotes. This is illustrated by the finding that three of the five algal species assigned to the genus *Chlorella* are more similar to other algae than to *Chlorella vulgaris* in their SS-rRNA sequences [187]. Results continue to accumulate which indicate that various groups of unicellular eukaryotes had already split off from the common evolutionary line before the separation of the fungi, plants and animals [19, 20, 388, 423]. There are particularly convincing arguments for a special position for the Ciliates: They have a deviant genetic code; they possess the most unusual cytochrome c of all eukaryotes; their histone H4 differs by more than 20 % from that of all other eukaryotes, the H4 of which differs without exception by less than 10 %; and the other histones, the 5.8s rRNA and the tRNAs of the Ciliates all differ markedly from those of other eukaryotes.

4.8 Palaeobiochemistry

Quite understandably, the examination of **fossils** plays only a minor role in molecular research into animal evolution. The macromolecular biosubstances retained in fossils consist mainly of insoluble scleroproteins which can be identified by hydrolysis and amino acid analysis. Such substances include collagen in fossil bones and teeth and also in invertebrates (brachiopods) since the Palaeozoic, ceratin in turtle shells, conchiolins in mussel shells, and other proteins in the remains of extinct graptoliths, in the calcareous bodies of the foraminifers and in the silicic acid skeletons of the radiolarians. The egg shells of dinosaurs contain proteins with a high glycine content and which are similar to the calcium-binding proteins of bird eggs [220]. The amino acids of fossil proteins are, however, subject to change through spontaneous chemical reactions; thus, L-aspartate is converted to the D-isomer at the rate of 0.1–0.25 % per year; D-allo-leucine is formed from L-leucine at the much slower rate of 0.001 % per thousand years [84]. In spite of these age-related chemical changes, proteins extracted from 60- to 80-million-year-old mussel shells still show immunological cross-reactivity with the corresponding proteins of present-day mussels [431]. Unfortunately, little phylogenetically useful information

may be expected from the collagens, which are made up of repetive structures that are low in information and coded by many genes, and the conchiolins, which are a mixture of proteins of poorly defined structure [310]. The frozen specimen of **mammoth** *(Mammuthus primigneius)*, found in Siberia in 1977–1979, presented a unique opportunity for biochemical studies on an animal species which became extinct thousands of years ago. Particularly interesting was the completely preserved baby mammoth "Dima", whose age was determined as 40 000 years using ^{14}C dating. However, the planned DNA analysis turned out to be unavoidably difficult. Fossil DNAs disintegrate through ageing processes into fragments of a few hundred base pairs; become heavily altered chemically by loss of bases, oxidation of pyrimidines and the growth of cross-links; and are often contaminated with foreign DNA. Only the polymerase chain reaction (PCR), which facilitates sequencing starting from minute amounts of DNA, was able to provide molecular evidence for the close relationship of the mammoth to the present elephant species [310]. It was much easier to examine museum materials from more recently extinct species. Steller's sea cow *(Hydrodamalis gigas)* was discovered in 1741 on the coast of the Komandor Islands off Kamchatka and was already extinct by 1768. Immunological comparisons of albumins extracted from bone with those from the four living sirenians show that *Hydrodamalis* separated 4–8 million years ago from its nearest relative, *Dugong dugong* of the Indo-Pacific, whilst the three manatee species of the genus *Trichechus* split off 17–20 million years ago [339]. The **quagga** was once the most abundant zebra species in Africa; the last specimen died in captivity in 1883. Until recently, it was not clear whether the quagga should be considered a separate zebra species or the southern type of *Equus burchelli*. Proteins extracted from quagga skin in fact showed much more similarity to plasma proteins of *E. burchelli* than to the other two zebra species, *E. zebra* and *E. grevyi* [252]. Whereas partial sequences of quagga and *E. burchelli* mtDNA are identical, those of the horse *E. caballus* contained 11 synonymous substitutions [172].

References

1. Alahiotis S. N.: Adaptation of Drosophila enzymes to temperature. 4. Natural selection at the alcohol dehydrogenase locus. Genetica 59: 81–87 (1982)
2. Alber T.: Mutational effects on protein stability. Annual Rev. Biochem. 58: 765–798 (1989)
3. de Albuquerque C. M. R. and Napp M.: Genetic variability at the esterase-6 locus in natural populations of Drosophila simulans in relation to environmental heterogeneity. Genetics 98: 399–407 (1981)
4. Alonso S. et al.: Comparison of three actin-coding sequences in the monuse: evolutionary relationships between the actin genes of warm-blooded vertebrates. J. mol. Evol. 23: 11–22 (1986)
5. Altruda F. et al.: The primary structure of human hemopexin deduced from cDNA sequence: evidence for internal repeating homology. Nucleic Acids Res. 13: 3841–59 (1985)
6. Altschul S. F. and Lipman D. J.: Protein database searches for multiple alignments. Proc. Nat. Acad. Sci. USA 87: 5509–13 (1990)
7. Anderson P. R. and Oakeshott J. G.: Parallel geographical patterns of allozyme variation in two sibling Drosophila species. Nature 308: 729–730 (1984)
8. Anderson S. M. and McDonald J. F.: Effect of environmental alcohol on in vivo properties of Drosophila alcohol dehydrogenase. Biochem. Genetics 19: 421–430 (1981)
9. Aquadro C. F. and Avise J. C.: An assessment of „hidden" heterogeneity within electromorphs at three enzyme loci in deer mice. Genetics 102: 269–283 (1982)
10. Arcari P., Martinelli R. and Salvatore F.: Human glyceraldehyde-3-phosphate dehydrogenase pdeudogenes: Molecular evolution and possible mechanisms for amplification. Biochem. Genet. 27: 439–450 (1989)
11. Argos P.: Sensitive procedure to compare amino acid sequences. J. mol. Biol. 193: 385–396 (1987)
12. Aspinwall N.: Genetic analysis of North American populations of the pink salmon, Oncorhynchnus gorbnuscha, possible evidence for the neutral mutation-random drift hypothesis. Evolution 28: 295–305 (1974)
13. Atchison M. and Adesnik M.: Gene conversion in a cytochrom P-450 gene family. Proc. Nat. Acad. Sci. USA 83: 2300–04 (1986)
14. Avise J. C.: Is evolution gradual or rectangular? Evidence from living fishes. Proc. Nat. Acad. Sci. USA 74: 5083–87 (1977)
15. Avise J. C. and Aquadro C. F.: A comparative summary of genetic distances in the vertebrates. In: Evolutionary Biology Vol. 15, pp. 151–185. Plenum, New York 1982
16. Ayala F. J.: Genetic differentiation during the speziation process. In: Evolutionary Biology Vol. 8, pp. 1–78. Plenum, New York 1975
17. Ayala F. J.: Genetic polymorphism: from electrophoresis to DNA sequences. Experientia 39: 813–823 (1983)
18. Barnes P. T. and Laurie-Ahlberg C. C.: Genetiv variability of flight muscle metabolism in Drosophila melanogaster. III. Effects of Gpdh allozymes and environmental temperature on power output. Genetics 112: 267–294 (1986)
19. Baroin A. et al.: Partial phylogeny of the unicellular eukaryotes based on rapid sequencing of a portion of 28S ribosomal RNA. Proc. Nat. Acad. Sci. USA 85: 3474–78 (1988)

20. Baverstock P. R. et al.: SrRNA evolution and phylogenetic relationships of the genus Naegleria (Protista: Rhizopoda). Mol. Biol. Evol. 6: 243–257 (1989)
21. Bendall D. S. (ed.): Evolution from molecules to man. Cambridge Univ. Press, Cambridge 1983
22. van Beneden R. J. and Powers D. A.: Structural and functional differentiation of two clinally distributed glucosephosphate isomerase allelic isozymes from the teleost Fundulus heteroclitus. Mol. Biol. Evol. 6: 155–170 (1989)
23. Benson D. C.: Fourier methods for biosequence analysis. Nucleic Acids Res. 18: 6305–10 (1990)
24. Berger D.: Esterases od Drosophila. II. Biochemical studies of esterase-5 in D. pseudoobscura. Genetics 78: 1157–72 (1974)
25. Berlocher S. H.: Insect molecular systematics. Annual Rev. Entomol. 29: 403–433 (1984)
26. Bernardi G. et al.: Compositional patterns in vertebrate genomes: Conservation and change in evolution. J. mol. Evol. 28: 7–18 (1989)
27. Beverley M., Ismach R. B. and McMahon Pratt D.: Evolution of the genus Leishmania as revealed by comparisons of nuclear DNA restriction fragment patterns. Proc. Nat. Acad. Sci. USA 84: 484–488 (1987)
28. Beverley S. M. and Wilson A. C.: Ancient origin for Hawaiian Drosophilinae inferred from protein comparisons. Proc. Nat. Acad. Sci. USA 82: 4753–57 (1985)
29. Bishop M. J. and Rawlings C. J. (eds.): Nucleic acids and protein sequence analysis. IRL Press, Oxford 1987
30. Black IV W. C. and Krafsur E. S.: Electrophoretic analysis of genetic variability in the house fly (Musca domestica L.). Biochem. Genetics 23: 193–203 (1985)
31. Blaisdell B. E.: A method of estimating from two aligned present day DNA sequences their ancestral composition and subsequent rates of substitution, possibly different in the two lineages, corrected for multiple and parallel substitutions at the same site. J. mol. Evol. 22: 69–81 (1985)
32. Blaisdell B. E.: Effectiveness of measures requiring and not requiring prior sequence alignment for estimating the dissimilarity of natural sequences. J. mol. Evol. 29: 526–537 (1989)
33. Bogart J. P. and Tandy M.: Polyploid amphibians: Three more diploid-tetraploid cryptic species of frogs. Science 193: 334–335 (1976)
34. Bonhomme F. et al.: Biochemical diversity and evolution in the genus Mus. Biochem. Genetics 22: 275–303 (1984)
35. Bonnell M. L. and Selander R. K.: Elephant seals: genetic variation and near extinction. Science 184: 908–909 (1974)
36. Bonnet F. et al.: An unexpected sequence homology between link proteins of the proteoglycan complex and immunoglobulin-like proteins. Biochim. biophys. Acta 873: 152–155 (1986)
37. Brent L. et al.: Further studies on supposed lamarckian inheritance of immunological tolerance. Nature 295: 242–244 (1982)
38. Britten R. J.: Rates of DNA sequence evolution differ between taxonomic groups. Science 231: 1393–98 (1986)
39. Britten R. J. et al.: Sources and evolution of human Alu repeated sequences. Proc. Nat. Acad. Sci. USA 85: 4770–74 (1988)
40. Brown A. J. L.: Variation of the 87A heat shock locus in Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 80: 5350–54 (1983)
41. Brown G. G. and Simpson M. V.: Intra- and interspecific variation of the mitochondrial genome in Rattus norwegicus and Rattus rattus: Restriction enzyme analysis of variant mitochondrial DNA molecules and their evolutionary relationships. Genetics 97: 125–143 (1981)
42. Brown G. G. and Simpson M. V.: Novel features of animal mtDNA evolution as shown by sequences of 2 rat cytochrome oxidase subunit-II genes. Proc. Nat. Acad. Sci. USA 79: 3246–50 (1982)
43. Brownell E.: DNA/DNA hybridisation studies of muroid rodents: Symmetry and rates of molecular evolution. Evolution 37: 1034–51 (1983)
44. Brunk C. F. and Olson E. C. (eds.): The C.S.E.O.L. conference on DNA-DNA hybridisation and evolution, Lake Arrowhead, California, May 11–14, 1989. J. mol. Evol. 30: 101–297 (1990)
45. Bryan G., Garza D. and Hartl D.: Insertion and excision of the transposable element mariner in Drosophila. Genetics 125: 103–114 (1990)
46. Buchanan B. A. and Johnson D. L. E.: Hidden electrophoretic variation at the xanthine dehydrogenase locus in a natural population of Drosophila melanogaster. Genetics 104: 301–315 (1983)
47. Bulfield G., Moore E. A. and Kacser H.: Genetic variation in activity of the enzymes of glycolysis and gluconeogenesis between inbred strains of mice. Genetics 89: 551–561 (1978)
48. Burbaum J. J. et al.: Evolutionary optimization of the catalytic effectiveness of an enzyme. Biochemistry 28: 9293–9305 (1989)
49. Burke W. D. and Eickbush T. H.: The silkworm late chorion locus. I. Variation within two paired multigene families. J. mol. Biol. 190: 343–356 (1986)
50. Burkhart B. D. et al.: Characterization of allozyme null and low activity alleles from two natural populations of Drosophila melanogaster. Genetics 107: 295–306 (1984)
51. Butler M. H. et al.: Molecular relationships between closely related strains and species of nematodes. J. mol. Evol. 18: 18–23 (1981)
52. Cabrera V. M. et al.: Electrophoretic variability in natural populations of Drosophila melanogaster and Drosophila simulans. Genetica 59: 191–201 (1982)
53. Caccone A., Desalle R. and Powell J. R.: Calibration of the change in thermal stability of DNA duplexes and degree of base pair mismatch. J. mol. Evol. 27: 212–216 (1988)
54. Caccone A. and Powell J. R.: DNA divergence among hominids. Evolution 43: 925–942 (1989)
55. Campbell C. A., Valentine J. W. and Ayala F. J.: High genetic variability in a population of Tridacna maxima from the Great Barrier Reef. Mar. Biol. 33: 341–345 (1975)
56. Carlson D. A. and Service M. W.: Identification of mosquitoes of Anopheles gambiae species complex A and B by analysis of cuticular components. Science 207: 1089–91 (1980)
57. Carr S. M., Brothers A. J. and Wilson A. C.: Evolutionary inferences from restriction maps of mitochondrial DNA from 9 taxa of Xenopus frogs. Evolution 41: 176–188 (1987)

58. Castillo O., Lehmann H. and Diamond R.: Selection at molecular level in mammalian myoglobins. FEBS Letters 96: 12–14 (1978)
59. Catzeflis F. M. et al.: DNA-DNA hybridisation evidence of the rapid rate of muroid rodent DNA evolution. Mol. Biol. Evol. 4: 242–253 (1987)
60. Cavalli-Sforza L. L. et al.: Reconstruction of human evolution: Bringing together genetic, archaeological, and linguistic data. Proc. Nat. Acad. Sci. USA 85: 6002–06 (1988)
61. Cavener D. R. and Clegg M. T.: Temporal stability of allozyme frequencies in a natural population of Drosophila melanogaster. Genetics 98: 613–623 (1981)
62. Chakraborty R. et al.: Heterozygosity and genetic distance of proteins. Nature 304: 755–756 (1983)
63. Charlesworth B. and Langley C. H.: The evolution of self-regulated transposition of transposable elements. Genetics 112: 359–383 (1986)
64. Chernyshev A. I.: Further study of histone structural gene multiplication in Drosophila melanogaster. Mol. gen. Genetics 185: 176–180 (1982)
65. Christen R. et al.: An analysis of the origin of metazoa, using comparisons of partial sequences of the 28S RNA, reveals an early emergence of the triploblasts. Embo J. 10: 499–503 (1991)
66. Clark R. L. et al.: Studies of enzyme polymorphism in the Kamuela population of Drosophila mercatorum. III. Effects of variation at the αGPD locus and subflight stress on the energy charge and glycolytic intermediate concentrations. Genetics 104: 661–675 (1983)
67. Clegg J. B.: Gene conversions in the horse α-globin gene complex. Mol. Biol. Evol. 4: 492–503 (1987)
68. Clegg M. T. and Obrien S. J. (eds.): Molecular evolution. Wiley & Liss, New York 1990
69. Cochrane B. J. and Richmond R. C.: Studies of esterase-6 in Drosophila melanogaster. II. The genetics and frequency distribution of naturally occurring variants studied by electrophoretic and heat stability criteria. Genetics 93: 461–478 (1979)
70. Cockerham C. C. and Tachida H.: Evolution and maintenance of quantitative genetic variation by mutations. Proc. Nat. Acad. Sci. USA 84: 6205–09 (1987)
71. Coen E., Strachan T. and Dover G.: Dynamics of concerted evolution of ribosomal DNA and histone gene families in the melanogaster species subgroup of Drosophila. J. mol. Biol. 158: 17–35 (1982)
72. Cohen A. S. et al.: Speciation in molluscs from Turkana basin. Nature 304: 659–663 (1983)
73. Cohen J. B. and Givol D.: Conservation and divergence of immunoglobulin VH pseudogenes. Embo J. 2: 1795–1800 (1983)
74. Cohn W. E. and Moldave K. (eds.): Transposable elements in mutagenesis and regulation of gene expression. Acad. Press, San Diego 1989
75. Collet C.: Recent origin for a thermostable alcohol dehydrogenase allele of Drosophila melanogaster. J. mol. Evol. 27: 142–146 (1988)
76. Collier G. E. and Obrien S. J.: A molecular phylogeny of the felidae – Immunological distance. Evolution 39: 473–487 (1985)
77. Conceicao M. B. et al.: Electrophoretic characterization of a hybrid between Eretmochelys imbricata and Caretta caretta (Cheloniidae). Comp. Biochem. Physiol. Pt. B 97: 275–278 (1990)
78. Cooke P. H. and Oakeshott J. G.: Amino acid polymorphisms for esterase-6 in Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 86: 1426–30 (1989)
79. Cornish-Bowden A.: Interpretation of amino acid compositions. Trends biochem. Sci. 6: 217–219 (1981)
80. Costa R. and Bisol P. M.: Genetic variability in deep-sea organisms. Biol. Bull. 155: 125–133 (1978)
81. Crawford D. L. and Powers D. A.: Molecular basis of evolutionary adaptation at the lactate dehydrogenase-B locus in the fish Fundulus heteroclitus. Proc. Nat. Acad. Sci. USA 86: 9365–69 (1989)
82. Crozier W. W.: Electrophoretic identification and comparative examination of naturally occurring F1-hybrids between brown trout (Salmo trutta L.) and Atlantic salmon (S. salar L.). Comp. Biochem. Physiol. Pt. B 78: 785–790 (1984)
83. Daar I. O. et al.: Human triose-phosphate isomerase deficiency: A single amino acid substitution results in a thermolabile enzyme. Proc. Nat. Acad. Sci. USA 83: 7903–07 (1986)
84. Davies K. H.: Amino acid analysis of Pleistocene marine molluscs from the Gower peninsula. Nature 302: 137–139 (1983)
85. Davis B. D.: Transcriptional bias: A non-Lamarckian mechanism for substrate-induced mutations. Proc. Nat. Acad. Sci. USA 86: 5005–5009 (1989)
86. Day T. H. and Buckley P. A.: Alcohol dehydrogenase polymorphism in the seaweed fly, Coelopa frigida. Biochem. Genetics 18: 727–742 (1980)
87. Dayhoff M. O. (ed.): Atlas of protein sequence and structure, Vol. 5 and Suppl. 1–3. Nat. Biomed. Res. Foundation, Washington 1972–78
88. Densmore L. D. and Owen R. D.: Molecular systematics of the order Crocodilia. Amer. Zool. 29: 831–841 (1989)
89. Densmore L. D., Wright J. W. and Browns W. M.: Mitochondrial-DNA analyses and the origin and relative age of parthenogenetic lizards (genus Cnemidophorus). 2. C. neomexicanus and the C. tesselatus complex. Evolution 43: 943–957 (1989)
90. Desalle R., Giddings E. V. and Kaneshiro K. Y.: Mitochondrial DNA variability in natural populations of Hawaiian Drosophila. II. Genetic and phylogenetic relationships of natural populations of D. silvestris and D. heteroneura. Heredity 56: 87–96 (1986)
91. Dibb N. J. et al.: Sequence analysis of mutations that affect the synthesis, assembly and enzymatic activity of the unc-54 myosin heavy chain of Caenorhabditis elegans. J. mol. Biol. 183: 543–551 (1985)
92. Dimichele L. and Powers D. A.: Physiological basis for swimming endurance differences between LDH-B genotypes of Fundulus heteroclitus. Science 216: 1014–16 (1982)
93. Doolittle R. F.: Similar amino acid sequences: Chance or common ancestry? Science 214: 149–159 (1981)
94. Doolittle R. F.: The genealogy of some recently evolved vertebrate proteins (Review). Trends biochem. Sci. 10: 233–237 (1985)
95. Doolittle R. F.: Similar amino acid sequences revisited. Trends biochem. Sci. 14: 244–245 (1989)
96. Doolittle R. F. (ed.): Molecular evolution: Computer analysis of protein and nucleic acid sequences. Acad. Press, San Diego 1990

97. Doolittle R. F. et al.: A natural occurring horizontal gene transfer from a eukaryote to a prokaryote. J. mol. Evol. 31: 383–388 (1990)
98. Dover G. A. and Flavell R. B. (eds.): Genome evolution. Acad. Press, New York 1982
99. Dwulet F. E. and Putnam F. W.: Internal duplication and evolution of human ceruloplasmin. Proc. Nat. Acad. Sci. USA 78: 2805–09 (1981)
100. Eanes W. F., Katona L. and Longtine M.: Comparison of in vitro and in vivo activities associated with the G6PD allozyme polymorphism in Drosophila melanogaster. Genetics 125: 845–853 (1990)
101. Easteal S. and Oakeshott J. G.: Estimating divergence times of Drosophila species from DNA sequence comparisons. Mol. Biol. Evol. 2: 87–91 (1985)
102. Easteal S.: Generation time and the rate of molecular evolution. Mol. Biol. Evol. 2: 450–453 (1985)
103. Echelle A. A. et al.: Mitochondrial-DNA diversity and the origin of the Menidia-clarkhubbsi complex of unisexual fishes (Atherinidae). Evolution 43: 984–993 (1989)
104. Eickbush T. H. and Burke W. D.: The silkmoth late chorion locus. II. Gradients of gene conversion in two paired multigene families. J. mol. Biol. 190: 357–366 (1986)
105. Eisses K. T. et al.: Analysis of the gene encoding the multifunctional alcohol dehydrogenase allozyme ADH-71k of Drosophila melanogaster. Mol. Biol. Evol. 7: 459–469 (1990)
106. Eldredge N. and Stanley S. M. (eds.): Livings fossils. Springer, Berlin 1984
107. Eldredge N.: Time frames: The rethinking of Darwinian evolution and the theory of punctuated equilibria. Simon & Schuster, New York 1985
108. Ewens W. J., Spielman R. S. and Harris H.: Estimation of genetic variation at the DNA level from restriction endonuclease data. Proc. Nat. Acad. Sci. USA 78: 3748–50 (1981)
109. Farmer J. L. and Carter M. W.: Properties og partially purified allozymes of esterase-5 of Drosophila pseudoobscura. Comp. Biochem. Physiol. Pt. B 93: 451–458 (1989)
110. Feder J. H.: Natural hybridization and genetic divergence between toads Bufo borealis and Bufo punctatus. Evolution 33: 1089–97 (1979)
111. Felsenstein J. (ed.): Numerical taxonomy. Springer, Berlin 1983
112. Felsenstein J.: Phylogenies from molecular sequences: Inference and reliability. Annual Rev. Genet. 22: 521–565 (1988)
113. Feng D. F., Johnson M. S. and Doolittle R. F.: Aligning amino acid sequences: Comparison of commonly used methods. J. mol. Evol. 21: 112–125 (1985)
114. Feng D. F. and Doolittle R. F.: Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. mol. Evol. 25: 351–352 (1987)
115. Ferguson A.: Biochemical systematics and evolution. Blackie, Glasgow 1980
116. Ferris S. D. and Whitt G. S.: Evolution of the differential regulation of duplicate genes after polyploidization. J. mol. Evol. 12: 267–317 (1979)
117. Ferris S. D. et al.: Extensive polymorphism in the mitochondrial DNA of apes. Proc. Nat. Acad. Sci. USA 78: 6319–23 (1981)
118. Ferris S. D., Sage R. D. and Wilson A. C.: Evidence from mtDNA sequences that common laboratory strains of inbred mice are descended from a single female. Nature 295: 163–165 (1982)
119. Field K. G. et al.: Molecular phylogeny of the animal kingdom. Science 239: 748–753 (1988)
120. Fitch W. M. and Margoliash E.: Construction of phylogenetic trees. Science 155: 279–284 (1967)
121. Fitch W. M.: The molecular evolution of cytochrome c in eukaryotes. J. mol. Evol. 8: 13–40 (1976)
122. Fitch W. M.: The challenges to darwinism since the last centennial and the impact of molecular studies. Evolution 36: 1133–43 (1982)
123. Formas J. R., Vera M. I. and Lacrampe S.: Allozymic and morphological differentiation in the South American frog genus Eupsophus. Comp. Biochem. Physiol. Pt. B 75: 475–478 (1983)
124. Freeth A. L., Gibson J. B. and de Couet H. G.: The partial characterization of alcohol dehydrogenase null alleles from natural populations of Drosophila melanogaster. Biochem. Genetics 24: 957–972 (1986)
125. Fuerst P. A., Chakraborty R. and Nei M.: Statistical studies on protein polymorphism in natural populations. I. Distribution of single locus heterozygosity. Genetics 86: 455–483 (1977)
126. Garciadorado A.: The effect of niche preference on polymorphism protection in a heterogenous environment. Evolution 40: 936–945 (1986)
127. Gardenal C. N., Sabattini M. S. and Blanco A.: Enzyme polymorphism in a population of Calomys musculinus (Rodentia, Cricetidae). Biochem. Genetics 18: 563–575 (1980)
128. Gauldie R. W., Purnell D. and Robertson D. A.: Some biochemical similarities and differences between two jack mackerel species, Trachurus declivis and T. novaezelandiae. Comp. Biochem. Physiol. Pt. B 58: 389–391 (1977)
129. Gelinas R. et al.: G to A substitution in the distal CCAAT box of the A-gamma-globin gene in Greek hereditary persistence of fetal hemoglobin. Nature 313: 323–325 (1985)
130. Goeorge jr. M and Ryder O. A.: Mitochondrial DNA evolution in the genus Equus. Mol. Biol. Evol. 3: 535–546 (1986)
131. Georgiev G. P.: Mobile genetic elements in animal cells and their biological significance. Eur. J. Biochem. 145: 203–220 (1984)
132. Ghazal P., Clark A. J. and Bishop J. O.: Evolutionary amplification of a pseudogene. Proc. Nat. Acad. Sci. USA 82: 4182–85 (1985)
133. Ghiretti-Magaldi A. and Tamino G.: Evolutionary studies on haemocyanin. In: Bannister J. V. (ed.): Structure and function of haemocyanin, pp. 271–278. Springer, Berlin 1977
134. Gibson J. B. and Wilks A. V.: Molecular structure of a naturally occurring alcohol dehydrogenase null activity alelle in Drosophila melanogaster. Biochem. Genet. 27: 679–688 (1989)
135. Gillespie J. H.: Could natural selection account for molecular evolution and polymorphism? Genome 31: 311–315 (1989)
136. Godine J. E., Chin W. W. and Habener J. F.: Subunit of rat pituitary glycoprotein hormones. Trimary structure of the precursor determined from the nucleotide sequence of cloned DNAs. J. biol. Chem. 257: 8368–71 (1982)
137. Golding G. B. and Glickman B. W.: Sequence-directed mutagenesis: Evidence from a pylogenetic

history of human α-interferon genes. Proc. Nat. Acad. Sci. USA 82: 8577–81 (1985)
138. Gomez E. et al.: What is a genus in the trypanosomatidae family. Phylogenetic analysis of two small rRNA sequences. Mol. Biol. Evol. 8: 254–259 (1991)
139. Gonzales I. L. et al.: Variation among human 28 S ribosomal RNA genes. Proc. Nat. Acad. Sci. USA 82: 7666–70 (1985)
140. Gonzalesz I. L. et al.: Ribosomal RNA gene sequences and hominoid phylogeny. Mol. Biol. Evol. 7: 203–219 (1990)
141. Goodman E. (ed.): Macromolecular sequences in systematics and evolutionary biology. Plenum, New York 1982
142. Goodman M., Weiss M. L. and Czelusiak J.: Molecular evolution above the species level: Branching pattern, rates, and mechanisms. Syst. Zool. 31: 376–399 (1983)
143. Goodman M. et al.: Molecular phylogeny of the family of apes and humans. Genome 31: 316–335 (1989)
144. Gorski J., Fiori M. and Mach B. A.: A new nonsense mutation as the molecular basis for β° thalassemia. J. mol. Biol. 154: 537–540 (1982)
145. Gosling E.: Hidden genetic variability in two populations of a marine mussel. Nature 279: 713–715 (1979)
146. Graf J. D. and Fischberg M.: Albumin evolution in polyploid species of the genus Xenopus. Biochem. Genetics 24: 821–837 (1986)
147. Grassle J. P. and Grassle J. F.: Sibling species in the marine pollution indicator Capitella (Polychaeta). Science 192: 567–569 (1976)
148. Graur D.: Gene diversity in Hymenoptera. Evolution 39: 190–199 (1985)
149. Graur D., Shuali Y. and Li W. H.: Deletions in processed pseudogenes accumulate faster in rodents than in humans. J. mol. Evol. 28: 279–285 (1989)
150. Gribskov M., Maclachlan A. D. and Eisenberg D.: Profile analysis: detection of distantly related proteins. Proc. Nat. Acad. Sci. USA 84: 4355–58 (1987)
151. Grula J. W. et al.: Sea urchin DNA sequence variation and reduced interspecies differences of the less variable sequences. Evolution 63: 665–676 (1982)
152. Gupta J. P., Dwivedi Y. N. and Singh B. K.: Natural hybridization in Drosophila. Experientia 36: 290 (1980)
153. Gutfreund H. (ed.): Biochemical evolution. Cambridge Univ. Press, Cambridge 1981
154. Guttman S. I., Grau G. A. and Karlin A. A.: Genetic variation in Lake Erie great blue herons (Ardea herodias). Comp. Biochem. Physiol. Pt. B 66: 167–169 (1980)
155. Gyllensten U. B. et al.: Nucleotide sequence and genomic organization of bird minisatellites. Nucleic Acids Res. 17: 2203–14 (1989)
156. Haj-Ahmad Y. and Hickley D. A.: A molecular explanation of frequency-dependent selection in Drosophila. Nature 299: 350–352 (1982)
157. Halanych K. M.: 5S ribosomal RNA sequences inappropriate for phylogenetic reconstruction. Mol. Biol. Evol. 8: 249–253 (1991)
158. Hale L. R. and Singh R. S.: Extensive variation and heteroplasmy in size of mitochondrial DNA among geographical populations of Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 83: 8813–17 (1986)
159. Harrison R. G.: Parallel variation at an enzyme locus in sibling species of field crickets. Nature 266: 168–170 (1977)
160. Hasegawa M. et al.: Mitochondrial DNA evolution in primates. Transition rate has been extremely low in the lemur. J. mol. Evol. 31: 113–121 (1990)
161. Hayashi J. I. et al.: Evolutionary aspects of variant types of rat mitochondrial DNAs. Biochim. biophys. Acta 564: 202–211 (1979)
162. Hayashi J. I., Moriwaki K. and Yosida T. H.: Polymorphisms of mitochondrial DNAs in Norway rats (Rattus norwegicus) – Cleavage site variations and length polymorphism of restriction fragments. Mol. gen. Genetics 184: 337–341 (1981)
163. Hein J.: A new method that simultaneously aligns and reconstructs ancestral sequences for any numer of homologous sequences, when the phylogeny is given. Mol. Biol. Evol. 6: 649–668 (1989)
164. Hein J.: A tree reconstruction method that is economical in the number of pairwise comparisons used. Mol. Biol. Evol. 6: 669–684 (1989)
165. Heinstra P. W. H. et al.: Alcohol dehydrogenase of Drosophila melanogaster: metabolic differences mediated through cryptic allozymes. Heredity 57: 23–29 (1986)
166. Heinstra P. W. H., Scharloo W. and Thörig G. E. W.: Physiological significance of the alcohol dehydrogenase polymorphism in larvae of Drosophila. Genetics 117: 75–84 (1987)
167. Heinstra P. W. H. et al.: The metabolism of ethanol-derived acetaldehyde by alcohol dehydrogenase (EC 1.1.1.1) and aldehyde dehydrogenase (EC 1.2.1.3) in Drosophila melanogaster larvae. Biochem. J. 259: 791–797 (1989)
168. Henikoff S. et al.: Two Drosophila melanogaster mutations block successive steps of de novo purine synthesis. Proc. Nat. Acad. Sci. USA 83: 3919–23 (1986)
169. Henthorn P. S. et al.: Products of two common alleles at the locus for human placental alkaline phosphatase differ by seven amino acids. Proc. Nat. Acad. Sci. USA 83: 5597–5601 (1986)
170. Hess J. F., Schmid C. W. and Shen C. K. J.: A gradient of sequence divergence in the human adult α-globin duplication units. Science 226: 67–70 (1984)
171. Hewett-Emmett D., Welty R. J. and Tashian R. E.: A widespread silent polymorphism of human carbonic anhydrase III (31 Ile – Val) – implications for evolutionary genetics. Genetics 105: 409–420 (1983)
172. Higuchi R. G. et al.: Mitochondrial DNA of the extinct quagga: Relatedness and extent of postmortem change. J. mol. Evol. 25: 283–287 (1987)
173. Hilbish T. J., Deaton L. E. and Koehn R. K.: Effect of an allozyme polymorphism on regulation of cell volume. Nature 298: 688–689 (1982)
174. Hillis D. M., Frost J. S. and Wright D. A.: Phylogeny and biogeography of the Rana pipiens complex: a biochemical evaluation. Syst. Zool. 32: 132–143 (1983)
175. Hillis D. M. and Moritz C.: Molecular systematics. Sinauer, Sunderland 1990
176. Hixson J. E. and Brown W. M.: A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications. Mol. Biol. Evol. 3: 1–18 (1986)

177. Hoffmann D. and van Regenmortel M. H. V.: Detection of distant antigenic relationships between insect and bird lysozymes by ELISA. J. mol. Evol. 21: 14–18 (1984)
178. Hoffmann R. J.: Temperature modulation of the kinetics of phosphoglucose isomerase genetic variants from the sea anemone Metridium senile. J. exp. Zool. 227: 361–370 (1983)
179. Holmquist R.: Transitions and transversions in evolutionary descent: an approach to understanding. J. mol. Evol. 19: 134–144 (1983)
180. Holmquist R., Miyamoto M. M. and Goodman M.: Higher-primate phylogeny: Why can't we decide? Mol. Biol. Evol. 5: 201–216 (1988)
181. Hornbach D. J., McLeod M. J. and Guttman S. I.: On the validity of the genus Musculium (Bivalvia: Sphaeriidae). Can. J. Zool. 58: 1703–07 (1980)
182. Hudson R. R.: Estimating genetic variability with restriction endonucleases. Genetics 100: 711–719 (1982)
183. Hudson R. R., Kreitman M. and Aguadé M.: A test of neutral molecular evolution based on nucleotide data. Genetics 116: 153–159 (1987)
184. Hung A. C. F. and Vinson S. B.: Interspecific hybridization and caste specificity of protein in fire ant. Science 196: 1458–60 (1977)
185. Hung A. C. F., Hedlund R. C. and Day W. H.: High level of genetic heterozygosity in the hyperparasitic wasp, Mesochoris nigripes. Experientia 42: 1050–51 (1986)
186. Hunt J. A. and Carson H. L.: Evolutionary relationships of four species of Hawaiian Drosophila as measured by DNA reassociation. Genetics 104: 353–364 (1983)
187. Huss V. A. R. and Sogin M. L.: Phylogenetic position of some Chlorella species within the chlorococcales based upon complete small-subunit ribosomal RNA sequences. J. mol. Evol. 31: 432–442 (1990)
188. Ikemura T.: Codon usage and tRNA content in unicellular and multicellular organisms (Review). Mol. Biol. Evol. 2: 13–34 (1985)
189. Jimenez-Marin D. and Dessauer H. C.: Protein phenotype variation in laboratory populations of Rattus norwegicus. Comp. Biochem. Physiol. Pt. B 46: 487–492 (1973)
190. Jin L. and Nei M.: Limitations of the evolutionary parsimony method of phylogenetic analysis. Mol. Biol. Evol. 7: 82–102 (1990)
191. Johnson G. B.: Characterization of electrophoretically cryptic variation in alpine butterfly Colias meadii. Biochem. Genetics 15: 665–693 (1977)
192. Johnson M. S., Clarke B. and Murray J.: Genetic variation and reproductive isolation in Portula. Evolution 31: 116–126 (1977)
193. Johnson M. S., Sutcliffe M. J. and Blundell T. E.: Molecular anatomy: Phyletic relationships derived from three-dimensional structures of proteins. J. mol. Evol. 30: 43–59 (1990)
194. Jollès J. et al.: Amino acid sequences of stomach and nonstomach lysozymes of ruminants. J. mol. Evol. 30: 370–382 (1990)
195. Jones C. W. and Kafatos F. C.: Accepted mutations in a gene family: Evolutionary diversification of duplicated DNA. J. mol. Evol. 19: 87–103 (1982)
196. Jones J. S. et al.: Gene flow and the geographical distribution of a molecular polymorphism in Drosophila pseudoobscura. Genetics 98: 157–178 (1981)
197. de Jong W.: Protein sequence evidence for a monophyly of the carnivore families Procyonidae and Mustelidae. Mol. Biol. Evol. 3: 276–281 (1986)
198. Joysey K. A. and Friday A. E. (eds.): Problems of phylogenetic reconstruction. Acad. Press, New York 1982
199. Jukes T.H.: Silent nucleotide substitutions and the molecular evolutionary clock. Science 210: 973–978 (1980)
200. Jukes T. H. (ed.): Special issue – Molecular evolutionary clock. J. mol. Evol. 26: 1–164 (1987)
201. Kaplan N. L., Hudson R. R. and Langley C. H.: The „hitchhiking effect" revisited. Genetics 123: 887–899 (1989)
202. Karlin S. et al.: Efficient algorithms for molecular sequence analysis. Proc. Nat. Acad. Sci. USA 85: 841–845 (1988)
203. Kazazian jr. H. H. et al.: Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332: 164–166 (1988)
204. Keith T. P.: Frequency distribution of esterase-5 alleles in two populations of Drosophila pseudoobscura. Genetics 105: 135–155 (1983)
205. Kessler L. G. and Avise J. C.: A comparative description of mitochondrial DNA differentiation in selected avian and other vertebrate genera. Mol. Biol. Evol. 2: 109–125 (1985)
206. Kidd S. and Young M. W.: Transposon-dependent mutant phenotypes at the Notch locus of Drosophila. Nature 323: 89–91 (1986)
207. Kimura A. and Takagi Y.: A frameshift addition causes silencing in the δ-globin gene in an old world monkey, an anubis (Papio doguera). Nucleic Acids Res. 11: 2541–50 (1983)
208. Kimura M.: Evolutionary rate at the molecular level. Nature 217: 624–626 (1968)
209. Kimura M. (ed.): Molecular evolution, protein polymorphism and the neutral theory. Springer, Berlin 1982
210. Kimura M.: The neutral theory of molecular evolution. Cambridge Univ. Press, Cambridge 1983
211. Kimura M.: Rare variant alleles in the light of the neutral theory. Mol. Biol. Evol. 1: 84–93 (1983)
212. Kimura M. and Nakahata N. (eds.): New aspects of the genetics of molecular evolution. Springer, Berlin 1991
213. King J. L. and Jukes T. H.: Non-Darwinian evolution. Science 164: 788–798 (1968)
214. King M. and Wilson A. C.: Evolution at two levels in humans and chimpanzees. Science 188: 107–116 (1975)
215. Kirkwood T. B. L. and Rosenberger R. F. (eds.): Accuracy in molecular processes – Its control and relevance to living systems. Chapman & Hall, London 1986
216. Klarenberg A. J. et al.: Genetic and dietary regulation of tissue-specific expression patterns of α-amylase in larvae of Drosophila melanogaster. Comp. Biochem. Physiol. Pt. B 89: 143–146 (1988)
217. Klose J.: Genetic variability of soluble proteins studied by two-dimensional electrophoresis on different inbred mouse strains and on different mouse organs. J. mol. Evol. 18: 315–328 (1982)
218. Konkel D. A., Maizel J. V. and Leder P.: The evolution and sequence comparison of two recently

diverged mouse chromosome beta-globin genes. Cell 18: 865–873 (1979)
219. Kornfield I. L.: Evidence for rapid speciation in African cichlid fishes. Experientia 34: 335–336 (1978)
220. Krampitz G. et al.: Calcium-binding peptide in Dinosaur egg shells. Naturwissenschaften 64: 583 (1977)
221. Kreitman M. E. and Aguadé M.: Excess polymorphism at the Adh locus in Drosophila melanogaster. Genetics 114: 93–110 (1986)
222. Lake J. A.: A rate-independent technique for analysis of nucleic acid sequences: Evolutionary parsimony. Mol. Biol. Evol. 4: 167–191 (1987)
223. Lakovaara S. and Keränen L.: Variation at the α-Gpdh locus of drosophilids. Hereditas 92: 251–258 (1980)
224. Langley C. H. et al.: Null allele frequencies at allozyme loci in natural populations of Drosophila melanogaster. Genetics 99: 151–156 (1981)
225. Lansman R. A. et al.: The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. J. mol. Evol. 17: 214–226 (1981)
226. Lansman R. A. et al.: Extensive genetic variation in mitochondrial DNAs among geographic populations of the deer mouse, Peromyscus maniculatus. Evolution 37: 1–16 (1983)
227. Laurie-Ahlberg C. C. et al.: Genetic variability of flight muscle metabolism in Drosophila melanogaster. II. Relationship between power output and enzyme activity levels. Genetics 111: 845–868 (1985)
228. Laurie C. C. and Stam L. F.: Quantitative analysis of RNA produced by slow and fast alleles of Adh in Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 85: 5161–65 (1988)
229. Lee Y. M., Misra H. P. and Ayala F. J.: Superoxide dismutase in Drosophila melanogaster – Biochemical and structural characterization of allozyme variants. Proc. Nat. Acad. Sci. USA 78: 7052–55 (1981)
230. Lehrman M. A. et al.: Exon-alu recombination deletes 5 kilobases from the low density lipoprotein receptor gene, producing a null phenotype in familian hypercholesterolemia. Proc. Nat. Acad. Sci. USA 83: 3679–83 (1986)
231. Leibenguth F. and Kutz U.: Alcohol dehydrogenase polymorphism in the flour moth, Ephestia kuehniella. Comp. Biochem. Physiol. Pt. B 82: 455–459 (1985)
232. Lennon G. G. and Nussinov R.: Homonyms, synonyms and mutations of the sequence/structure vocabulary. J. mol. Biol. 175: 425–430 (1984)
233. Leslie J. F., Cain G. D. and Meffe G. K.: Enzyme polymorphism in Ascaris suum (Nematoda). J. Parasitol. 68: 576–587 (1982)
234. Levinson G. and Gutman G. A.: Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. 4: 203–221 (1987)
235. Levy J. A., Haimovici M. and Conceicao M.: Genetic evidences for two species to the genus Eledone (Cephalopoda: Octopodidae) in South Brazil. Comp. Biochem. Physiol. Pt. B 90: 275–277 (1988)
236. Lewin R.: Molecular clocks turn a quarter century. Science 239: 561–563 (1988)
237. Lewontin R. C.: The genetic basis of evolutionary change. Columbia Univ. Press, New York 1974
238. Lewontin R. C.: Inferring the number of evolutionary events from DNA coding sequence differences. Mol. Biol. Evol. 6: 15–32 (1989)
239. Li B. F. L. et al.: The catalytic consequences of experimental evolution. Transitionstate structure during catalysis by the evolved β-galactosidases of Escherichia coli (ebg enzymes) changed by a single mutational event. Biochem. J. 260: 109–114 (1989)
240. Li W. H. and Nei M.: Persistence of common alleles in two related populations or species. Genetics 86: 901–914 (1977)
241. Li W. H.: Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. Genetics 95: 237–258 (1980)
242. Li W. H. and Gojobori T.: Rapid evolution of goat and sheep globin genes following gene duplication. Mol. Biol. Evol. 1: 94–108 (1983)
243. Li W. H., Wu C. I. and Luo C. C.: A new method for estimating synonymous and non-synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. 2: 150–174 (1985)
244. Li W. H.: Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. mol. Evol. 24: 337–345 (1987)
245. Li W. H. et al.: Molecular phylogeny of Rodentia, Lagomorpha, Artiodactyla, and Carnivora and molecular clocks. Proc. Nat. Acad. Sci. USA 87: 6703–07 (1990)
246. Li W. H. and Graur D.: Fundamentals of molecular evolution. Sinauer Assoc., Sunderland 1991
247. Liebhaber S. A. et al.: Hemoglobin I mutation encoded at both α-globin loci on the same chromosome: concerted evolution in the human genome. Science 226: 1449–51 (1984)
248. Lipman D. J., Altschul S. F. and Kececioglu J. D.: A tool for multiple sequence alignment. Proc. Nat. Acad. Sci. USA 86: 4412–15 (1989)
249. Liskay R. M., Letsou A. and Stachelek J. L.: Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics 115: 161–167 (1987)
250. Loukas M., Delidakis C. and Kafatos F. C.: Genomic blot hybridization as a tool of phylogenetic analysis: Evolutionary divergence in the genus Drosophila. J. mol. Evol. 24: 174–188 (1986)
251. Lowenstein J. M., Sarich V. M. and Richardson B. J.: Albumin systematics of the extinct mammoth and Tasmanian wolf. Nature 291: 409–411 (1981)
252. Lowenstein J. M. and Ryder O. A.: Immunological systematics of the extinct quagga (Equidae). Experientia 41: 1192–93 (1985)
253. Lucotte G. and Kaminski M.: Molecular heterosis at the conalbumin locus in the ring-necked pheasant (Phasianus colchicus). Theor. appl. Genetics 48: 251–253 (1976)
254. Lynch M. and Hill W. G.: Phenotypic evolution by neutral mutation. Evolution 40: 915–935 (1986)
255. Lanch M.: The similarity index and DNA fingerprinting. Mol. Biol. Evol. 7: 478–484 (1990)
256. MacIntyre R. J. (ed.): Molecular evolutionary genetics. Plenum, New York 1985
257. Maeda N., Bliska J. B. and Smithies O.: Recombination and balanced chromosome polymorphism suggested by DNA sequences 5' to the human δ-globin gene complex. Proc. Nat. Acad Sci USA 80: 5012–16 (1983)
258. Manella C. A., Frank J. and Delihas N.: Interrelatedness of 5 S RNA sequences investigated by correspondence analysis. J. mol. Evol. 24: 228–235 (1987)

259. Martin C. H. and Meyerowitz E. M.: Characterization of the boundaries between adjacent rapidly and slowly evolving genomic regions in Drosophila. Proc. Nat. Acad. Sci. USA 83: 8654–58 (1986)
260. Martin P. F. et al.: UGA nonsense mutation in the alcohol dehydrogenase gene of Drosophila melanogaster. J. mol. Biol. 184: 221–229 (1985)
261. Martinez-Cruzado J. C.: Evolution of the autosomal chorion cluster in Drosophila. IV. The Hawaiian Drosophila: Rapid protein evolution and constancy in the rate of DNA divergence. J. mol. Evol. 31: 402–423 (1990)
262. Martins E., Mestriner M. A. and Contel E. P. B.: Alcohol dehydrogenase polymorphism in Apis mellifera. Biochem. Genetics 15: 357–366 (1977)
263. Maruyama T. and Fuerst P. A.: Population bottlenecks and nonequilibrium models in population genetics. III. Genic homozygosity in populations which experience periodic bottlenecks. Genetics 111: 691–703 (1985)
264. Matsuoka N., Chiba Y. and Saitoh K.: Allozymic similarity in two species of the genus Brenthis (Lepidoptera: Nymphalidae). Comp. Biochem. Physiol. Pt. B 74: 385–387 (1983)
265. Matsuoka N. and Suzuki H.: Electrophoretic study on the taxonomic relationship of the two morphologically very similar sea-urchins, Echinostrephus aciculatus and E. molaris. Comp. Biochem. Physiol. Pt. B 88: 637–641 (1987)
266. Maxson L. R.: Immunological detection of convergent evolution in the frog Anotheca spinosa. Syst. Zool. 26: 72–76 (1977)
267. Maxson R. D. and Maxson L. R.: Micro-complement fixation: A quantitative estimator of protein evolution. Mol. Biol. Evol. 3: 375–388 (1986)
268. McCarthy A. D., Goldring J. P. D. and Hardie D. G.: Evidence that the multifunctional polypeptides of vertebrate and fungal fatty acid synthases have arisen by independent gene fusion events. FEBS Letters 162: 300–304 (1983)
269. McGinnis W., Shermoen A. W. and Beckendorf S. K.: A transposable element inserted just 5' to a Drosophila glue protein gene alters gene expression and chromatin structure. Cell 34: 75–84 (1983)
270. McLean M., Okubo C. K. and Tracey M. L.: mtDNA heterogeneity in Panulirus argus. Experientia 39: 536–538 (1983)
271. McLeod M. J., Wynes D. L. and Guttman S. I.: Lack of biochemical evidence for hybridization between two species of darters. Comp. Biochem. Physiol. Pt. B 67: 323–325 (1980)
272. Meyer A. and Wilson A. C.: Origin of tetrapods inferred from their mitochondrial DNA affiliation to lungfish. J. mol. Evol. 31: 359–364 (1990)
273. Meyer A. et al.: Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature 347: 550–553 (1990)
274. Meyer T. E., Cusanovich M. A. and Kamen M. D.: Evidence against use of bacterial amino acid sequence data for construction of all-inclusive phylogenetic trees. Proc. Nat. Acad. Sci. USA 83: 217–220 (1986)
275. Michelson A. M. and Orkin S. A.: Boundaries of gene conversion within the duplicated human α-globin genes. Concerted evolution by segmental recombination. J. biol. Chem. 258: 15245–54 (1983)
276. Michelson A. M. et al.: Structure of human phosphoglycerate kinase gene and the intron-mediated evolution and dispersal of the nucleotide-binding domain. Proc. Nat. Acad. Sci. USA 82: 6965–69 (1985)
277. Minghetti P. P., Law S. W. and Dugaiczyk A.: The rate of molecular evolution of α-fetoprotein approaches that of pseudogenes. Mol. Biol. Evol. 2: 347–358 (1985)
278. Miyazaki J. I., Sekiguchi K. and Hirabayashi T.: Application of an improved method of two-dimensional electrophoresis to the systematic study of horseshoe crabs. Biol. Bull. 172: 212–224 (1987)
279. Mohrenweiser, H. W. and Neel J. V.: Frequency of thermostability variants – Estimation of total rare variant frequency in human populations. Proc. Nat. Acad. Sci. USA 78: 5729–33 (1981)
280. Monnat R. J. and Loeb L. A.: Nucleotide sequence preservation of human mitochondrial DNA. Proc. Nat. Acad. Sci. USA 82: 2895–99 (1985)
281. Mort M. A. and Wolf H. G.: Enzyme variability in large-lake Daphnia populations. Heredity 55: 27–36 (1985)
282. Mueller L. D., Barr L. G. and Ayala F. J.: Natural selection vs. random drift: evidence from temporal variation in allele frequencies in nature. Genetics 111: 517–554 (1985)
283. Mukai T., Harada K. and Yoshimaru H.: Spontaneous mutations modifying the activity of alcohol dehydrogenase (ADH) in Drosophila melanogaster. Genetics 106: 73–84 (1984)
284. Narang S.: Genetic variability in natural populations, evidence in support of the selectionist view. Experientia 36: 50–51 (1980)
285. Narise S.: Purification and biochemical properties of allelic forms of cytoplasmic glycerol-3-phosphate dehydrogenase from Drosophila viridis. Biochim. biophys. Acta 615: 289–298 (1980)
286. Needleman S. B. and Wunsch C. D.: A general method applicable to the search for similarities in the amino acid sequence of two proteins. J. mol. Biol. 48: 443–453 (1970)
287. Neel J. V. and Rothman E.: Is there a difference among human populations in the rate with which mutation produces electrophoretic variants? Proc. Nat. Acad. Sci. USA 78: 3108–12 (1981)
288. Neel J. V. et al.: The rate with which spontaneous mutation alters the electrophoretic mobility of polypeptides. Proc. Nat. Acad. Sci. USA 83: 389–393 (1986)
289. Nei M.: Standard error of immunological dating of evolutionary time. J. mol. Evol. 9: 203–211 (1977)
290. Nei M. and Tajima F.: DNA polymorphism detectable by restriction endonucleases. Genetics 97: 145–163 (1981)
291. Nei M. and Tajima F.: Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105: 207–217 (1983)
292. Nei M., Tajima F. and Tateno Y.: Accuracy of estimated phylogenetic trees from molecular data. II. Frequency data. J. mol. Evol. 19: 153–170 (1983)
293. Nei M. and Koehn R. K. (eds.): Evolution of genes and proteins. Sinauer Assoc., Sunderland 1983
294. Nei M. and Graur D.: Extent of protein polymorphism and the neutral mutation theory. In : Evolutionary Biology Vol. 17, pp. 73–118. Plenum, New York 1984

295. Nei M., Stephens J. C. and Saitou, N.: Methods for computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. Mol. Biol. Evol. 2: 66–85 (1985)
296. Nei M. and Gojobori T.: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3: 418–426 (1986)
297. Nei M.: Molecular evolutionary genetics. Columbia Univ. Press, New York 1987
298. Nei M. and Miller J. C.: A simple method for estimating average number of nucleotide substitutions within and between populations from restruction data. Genetics 125: 873–879 (1990)
299. Nevo E., Shimony T. and Libni M.: Thermal selection of allozyme polymorphisms in barnacles. Nature 267: 699–701 (1977)
300. Nevo E.: Genetic diversity in nature: Patterns and theory. In: Evolutionary biology Vol. 23, pp. 217–246. Plenum, New York 1988
301. Newman C. M., Cohen J. E. and Kipnis C.: Neo-Darwinian evolution implies punctuated equilibria. Nature 315: 400–401 (1985)
302. O'Brien S. J. et al.: A molecular solution to the riddle of the giant panda's phylogeny. Nature 317: 140–144 (1985)
303. Ohnishi K. I.: Constant rate of evolution in the antigenicity of glucose 6-phosphate dehydrogenase revealed by the enzyme inhibition method. Comp. Biochem. Physiol. Pt. B 80: 217–222 (1985)
304. Ohno S.: Evolution by gene duplication. Springer, Berlin 1970
305. Ohta T. and Aoki K. (eds.): Population genetics and molecular evolution. Springer, Tokyo 1985
306. Ohta T.: Simulating evolution by gene duplication. Genetics 115: 207–213 (1987)
307. Ohta T.: Role of gene duplication in evolution. Genome 31: 304–310 (1989)
308. Olmo E. et al.: Repetitive DNA and polyploidy in selachians. Comp. Biochem. Physiol. Pt. B 73: 739–745 (1982)
309. Orkin S. H. and Karazian jr. H. H.: The mutation and polymorphism of the human β-globin gene and its surrounding DNA. Annual Rev. Genetics 18: 131–171 (1984)
310. Paabo S., Higuchi R. G. and Wilson A. C.: Ancient DNA and the polymerase chain reaction. The emerging field of molecular archaeology. J. Biol. Chem. 264: 9709–12 (1989)
311. Pace N. R., Olsen G. J. and Woese C. R.: Ribosomal RNA phylogeny and the primary lines of evolutionary descent. Cell 45: 325–326 (1986)
312. Palmer J. D. and Herbon L. A.: Plant mitochondrial DNA evolves rapidly in structure, but slowly in sequence. J. mol. Evol. 28: 87–97 (1989)
313. Palumbi S. R.: Rates of molecular evolution and the fraction of nucleotide positions free to vary. J. mol. Evol. 29: 180–187 (1989)
314. Pamilo P. and Nei M.: Relationships between gene trees and species trees. Mol. Biol. Evol. 5: 568–583 (1988)
315. Panepucci L., Schwantes M. L. and Schwantes A. R.: Biochemical and physiological properties of the lactate dehydrogenase allozymes of the Brazilian teleost, Leporinus friderici, Anostomidae. Comp. Biochem. Physiol. Pt. B 87: 199–206 (1986)
316. Parker H. R., Philipp D. P. and Whitt G. S.: Gene regulatory divergence among species estimated by altered developmental patterns in interspecific hybrids. Mol. Biol. Evol. 2: 217–250 (1985)
317. Pasyukova E. G. et al.: Concerted transpositions of mobile genetic elements coupled with fitness changes in Drosophila melanogaster. Mol. Biol. Evol. 3: 299–312 (1986)
318. Patterson C. (ed.): Molecules and morphology in evolution: conflict or compromise? Cambridge Univ. Press, New York 1987
319. Patterson C.: Homology in classical and molecular biology. Mol. Biol. Evol. 5: 603–625 (1988)
320. Patthy L.: Detecting homology of distantly related proteins with consensus sequences. J. mol. Biol. 198: 567–577 (1987)
321. Patthy L.: Detecting distant homologies of mosaic proteins. J. mol. Biol. 202: 689–696 (1988)
322. Patton J. C. and Avise J. C.: An empirical evaluation of qualitative hennigian analyses of protein electrophoretic data. J. mol. Evol. 19: 244–254 (1983)
323. Pearson W. R. and Lipman D. J.: Improved tools for biological sequence comparison. Proc. Nat. Acad. Sci. USA 85: 2444–48 (1988)
324. Pen J. et al.: Differences in specificity and catalytic efficiency between allozymes of esterase-4 from Drosophila mojavensis. Mol. Biol. Evol. 3: 366–373 (1986)
325. Penny D., Foulds L. R. and Hendy M. D.: Testing the theory of evolution by comparing phylogenetic trees constructed from five different protein sequences. Nature 297: 197–200 (1982)
326. Penny D. and Hendy M. D.: Estimating the realiability of evolutionary trees. Mol. Biol. Evol. 3: 403–417 (1986)
327. Perryman S. M. et al.: Sequence of a cDNA for mouse thymidilate synthase reveals striking similarity with the prokaryotic enzyme. Mol. Biol. Evol. 3: 313–321 (1986)
328. Pettersson G.: Effect of evolution on the kinetic properties of enzymes. Eur. J. Biochem. 184: 561–566 (1989)
329. Philipsen J. N. J. et al.: Characterization of a polymorphism in the 3' part of the chicken vitellogenin. J. mol. Evol. 28: 185–190 (1989)
330. Place A. R. and Powers D. A.: Kinetic characterization of the lactate dehydrogenase (LDH-B4) allozymes of Fundulus heteroclitus. J. biol. Chem. 259: 1309–18 (1984)
331. Powell J. R.: Protein variation in natural populations of animals. In: Evolutionary Biology Vol. 8, pp. 79–119. Plenum, New York 1975
332. Powell J. R. et al.: Rates of nucleotide substitution in Drosophila mitochondrial DNA and nuclear DNA are similar. Proc. Nat. Acad. Sci. USA 83: 90–90–93 (1986)
333. Prager E. M. and Wilson A. C.: Congruence of phylogenies derived from different proteins. A molecular analysis of the phylogenetic position of cracid birds. J. mol. Evol. 9: 45–57 (1976)
334. Prager E. M., Welling G. W. and Wilson A. C.: Comparison of various immunological methods for distinguishing among mammalian pancreatic ribonucleases of known amino acid sequence. J. mol. Evol. 10: 293–307 (1978)
335. Prager E. M. and Wilson A. C.: Construction of phylogenetic trees for proteins and nucleic acids: Empir-

ical evaluation of alternative matrix methods. J. mol. Evol. 11: 129–142 (1978)
336. Preparata R. M. et al.: Ciliate evolution: The ribosomal phylogenies of the tetrahyminae. J. mol. Evol. 28: 427–441 (1989)
337. Pryor S. C. et al.: Biochemical genetics of the Culex pipiens complex – I. 6-Phosphogluconate dehydrogenase. Comp. Biochem. Physiol. Pt. B 65: 663–668 (1980)
338. Pryor S. C. et al.: Biochemical genetics of the Culex pipiens complex – II. Hexokinase. Comp. Biochem. Physiol. Pt. B 67: 705–710 (1980)
339. Rainey W. E. et al.: Sirenian molecular systematics – including the extinct Steller's sea cow (Hydrodamalis gigas). Naturwissenschaften 71: 586–588 (1984)
340. Ramshaw J., Coyne J. A. and Lewontin R. C.: The sensitivity of gel electrophoresis as a detector of genetic variation. Genetics 93: 1019–37 (1979)
341. Rand D. M. and Harrison R. G.: Molecular population genetics of mtDNA size variation in crickets. Genetics 121: 551–569 (1989)
342. Rao J. K. M.: New scoring matrix for amino acid residue exchanges based on residue characteristic physical parameters. Int. J. Peptide Protein Res. 29: 276–281 (1987)
343. Reeve H. K. et al.: DNA „fingerprinting" reveals high levels of inbreeding in colonies of the eusocial naked mole-rat. Proc. Nat. Acad. Sci. USA 87: 2496–2500 (1990)
344. Reisner A. H. and Westwood N. H.: Assessment of similarities of pairs and groups of proteins using transformed amino-acid residues data. J. mol. Evol. 18: 240–250 (1982)
345. Richardson R. H., Smouse P. E. and Richardson M. E.: Patterns of molecular variation – II. Associations of electrophoretic mobility and larval substrate within species of the Drosophila mulleri complex. Genetics 85: 141–154 (1977)
346. Richmond M. C. and Zimmerman E. G.: Effect of temperature on activity of allozymic forms of supernatant malate dehydrogenase in the red shiner, Notropis lutrensis. Comp. Biochem. Physiol. Pt. B 61: 415–419 (1978)
347. Ridley M.: Evolution and classification. The reformation of cladism. Longman, London 1986
348. Riley M. A.: Nucleotide sequence of the Xdh region in Drosophila pseudoobscura and an analysis of the evolution of synonymous codons. Mol. Biol. Evol. 6: 33–52 (1989)
349. Risler J. L. et al.: Amino acid substitutions in structurally related proteins. A pattern recognition approach determination of a new and efficient scoring matrix. J. mol. Biol. 204: 1019–29 (1988)
350. Romero-Herrera A. E. et al.: The use of amino acid sequence analysis in assessing evolution. Biochimie 61: 767–779 (1979)
351. Ropson I. J., Brown D. C. and Powers D. A.: Biochemical genetics of Fundulus heteroclitus (L.). 6. Geographical variation in the gene frequencies of 15 loci. Evolution 44: 16–26 (1990)
352. Rosenblum B. B., Neel J. V. and Hanash S. M.: Two-dimensional electrophoresis of plasma polypeptides reveals „high" heterozygosity indices. Proc. Nat. Acad. Sci USA 80: 5002–06 (1983)
353. Ruvinsky A. O. et al.: Factors organizing and maintaining polymorphism in a cylic parthenogenetic species: Daphnia pulex. Heredity 57: 15–22 (1986)
354. Ruvolo M. et al.: Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. Proc. Nat. Acad. Sci. USA 88: 1570–74 (1991)
355. Sabath M. D.: Enzyme variability in 12 sympatric drosophilid species (genera: Chymomyza, Leucophenga, Scaphomyza and Drosophila). Amer. Midland Naturalist 94: 144–153 (1975)
356. Saccone C., Pesole G. and Preparata G.: DNA microenvironments and the molecular clock. J. mol. Evol. 29: 407–411 (1989)
357. Saitou N. and Nei M.: The neighbour-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4: 406–425 (1987)
358. Saitou N. and Imanishi T.: Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbour-joining methods of phylogenetic tree construction in obtaining the correct tree. Mol. Biol. Evol. 6: 514–525 (1989)
359. Salmon M. et al.: Behavioral and biochemical evidence for species distinctiveness in the fiddler crabs, Uca speciosa and U. spinicarpa. Evolution 33: 182–191 (1979)
360. Satta Y., Ishiwa H. and Chigusa S. I.: Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species. Mol. Biol. Evol. 4: 638–650 (1987)
361. Savatier P. et al.: Nucleotide sequence of the beta-globin genes in gorilla and macaque: The origin of nucleotide polymorphism in human. J. mol. Evol. 24: 309–318 (1987)
362. Sawyer S.: Statistical tests for detecting gene conversion. Mol. Biol. Evol. 6: 526–538 (1989)
363. Schaefer S. W., Aquadro C. F. and Anderson W. W.: Restriction-map variation in the alcohol dehydrogenase region of Drosophila pseudoobscura. Mol. Biol. Evol. 4: 254–265 (1987)
364. Schaefer S. W. and Aquadro C. F.: Nucleotide sequence of the Adh gene region of Drosophila pseudoobscura: evolutionary change and evidence for an ancient gene duplication. Genetics 117: 61–73 (1987)
365. Schimenti J. C. and Duncan C. H.: Ruminant globin gene structures suggest an evolutionary role for Alu-repeats. Nucleic Acids Res. 12: 1641–55 (1984)
366. Schueller C. et al.: The amino-acid sequence of pancreatic ribonuclease from the mole rat Spalax ehrenbergi, chromosomal species 2n=60. Biol. Chem. Hoppe-Seyler 370: 583–589 (1989)
367. Schuking R. F. and Plasterk R. H. A.: TcA, the putative transposase of C. elegans Tc1 transposon, has an N-terminal DNA binding domain. Nucleic Acids Res. 18: 895–900 (1990)
368. Schulz G. E.: Protein-Differenzierung: Entwicklung neuartiger Proteine im Laufe der Evolution. Angew. Chemie 93: 143–151 (1981)
369. Selander K., Clark A. G. and Whittam T. S. (eds.): Evolution at the molecular level. Sinauer Assoc., Sunderland 1991
370. Sene F. M. and Carson H. L.: Genetic variation in Hawaian Drosophila. IV. Allozymic similarity between D. silvestris and D. heteroneura from the island of Hawaii. Genetics 86: 187–198 (1977)
371. Shaklee J. B. and Tamaru C. S.: Biochemical and morphological evolution of Hawaiian bonefishes (Albula). Syst. Zool. 30: 125–146 (1981)
372. Shapira S. K. and Finnerty V. G.: The use of genetic complementation in the study of eukaryotic macro-

molecular evolution: Rate of spontaneous gene duplication at two loci of Drosophila melanogaster. J. mol. Evol. 23: 159–167 (1986)
373. Sharp P. M. and Li W. H.: On the rate of DNA sequence evolution in Drosophila. J. mol. Evol. 28: 398–402 (1989)
374. Sheppard H. W. and Gutman G. A.: Allelic forms of rat kappa chain genes. Evidence for strong selection at the level of nucleotide sequence. Proc. Nat. Acad. Sci. USA 78: 7064–68 (1981)
375. Sheppard W. S. and Heydon S. L.: High levels of genetic variability in 3 male-haploid species (Hymenoptera, Argidae, Tenthredinidae). Evolution 40: 1350–53 (1986)
376. Shields D. C. et al.: „Silent" sites in Drosophila genes are not neutral. Evidence of selection among synonymous codons. Mol. Biol. Evol. 5: 704–716 (1988)
377. Sillince J. A. A. and Sillince M.: Molecular databases for protein sequences and structure studies. An Introduction. Springer, Berlin 1991
378. Simmons G. M. et al.: Molecular analysis of the alleles of alcohol dehydrogenase along a cline in Drosophila melanogaster. 1. Maine, North Carolina, and Florida. Evolution 43: 393–409 (1989)
379. Singh R. S.: Population genetics and evolution of species related to Drosophila melanogaster. Annual Rev. Genet. 23:425–453 (1989)
380. Skibinski D. O. F., Ahmad M. and Beardmore J. A.: Genetic evidence for naturally occurring hybrids between Mytilus edulis and Mytilus galloprovincialis. Evolution 32: 354–364 (1978)
381. Skibinski D. O. F. and Ward R. D.: Correlation between heterozygosity and evolutionary rate of proteins. Nature 298: 490–492 (1982)
382. Slightom J. L., Blechl A. E. and Smithies O.: Human fetal Gγ- and Aγ-globin genes: Complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. Cell 21: 627–638 (1980)
383. Slightom J. L. et al.: Chimpanzee fetal Gγ- and Aγ-globin gene nucleotide sequences provide further evidence of gene conversions in hominoid evolution. Mol. Biol. Evol. 2: 370–389 (1985)
384. Smith M. J. et al.: Single copy DNA homology in sea stars. J. mol. Evol. 18: 92–101 (1982)
385. Sneath P. H. A. and Sokal R. R.: Numerical taxonomy. Freeman, San Francisco 1973
386. de Soete G.: On the construction of „optimal" phylogenetic trees. Z. Naturforsch. Sect. C 38: 156–158 (1983)
387. Sofer W. and Martin P. F.: Analysis of alcohol dehydrogenase gene expression in Drosophila. Annual Rev. Genet. 21: 203–225 (1987)
388. Sogin M. L. et al.: Phylogenetic meaning of the kingdom concept: An unusual ribosome from Giardia lamblia. Schience 243: 75–77 (1989)
389. Somero G. N. and Soulé M.: Genetic variation in marine fishes as a test of the niche-variation hypothesis. Nature 249: 670–672 (1974)
390. Sourdis J. and Nei M.: Relative efficiencies of the maximum parsimony and distance-matrix methods in obtaining the correct phylogenetic tree. Mol. Biol. Evol. 5: 298–311 (1988)
391. Spicer G. S.: Molecular evolution among some Drosophila species groups as indicated by two-dimensional electrophoresis. J. mol. Evol. 27: 250–260 (1988)
392. Spolsky C. and Uzzell T.: Evolutionary history of the hybridogenetic hybrid frog Rana esculenta as deduced from mtDNA analyses. Mol. Biol. Evol. 3: 44–56 (1986)
393. Steiner W. W. M. et al.: Electrophoretic comparison of aphid species: detecting differences based on taxonomic status and host plant. Comp. Biochem. Physiol. Pt. B 81: 295–299 (1985)
394. Stephens J. C.: Statistical methods of DNA sequence analysis. Detection of intragenic recombination or gene conversion. Mol. Biol. Evol. 2: 539–556 (1985)
395. Stephens J. C. and Nei M.: Phylogenetic analysis of polymorphic DNA sequences at the Adh locus in Drosophila melanogaster and its sibling species. J. mol. Evol. 22: 289–300 (1985)
396. Stoddart J. A.: The accumulation of genetic variation in a parthenogenetic snail. Evolution 37: 546–554 (1983)
397. Stoppa-Lyonett D. et al.: Clusters of intragenic Alu repeats predispose the human C1 inhibitor locus to deleterious rearrangements. Proc. Nat. Acad. Sci. USA 87: 1551–55 (1990)
398. Subbiah S. and Harrison S. C.: Method for multiple sequence alignment with gaps. J. mol. Biol. 209: 539–548 (1989)
399. Surrey S. et al.: Functional analysis of a β-globin gene containing a TATA box mutation from a Kurdish jew with β-thalassemia. J. biol. Chem. 260: 6507–10 (1985)
400. Syvanen M.: The evolutionary implications of mobile genetic elements. Annual Rev. Genet. 18: 271–293 (1984)
401. Tabachnik W. J. and Powell J. R.: A world-wide survey of genetic variation in the yellow fever mosquito, Aedes aegypti. Genetic Res. 34: 215–229 (1979)
402. Tajima F. and Nei M.: Estimation of evolutionary distance between nucleotide sequences. Mol. Biol. Evol. 1: 269–285 (1984)
403. Takahata N. and Nei M.: Gene genealogy and variance of interpopulational differences. Genetics 110: 325–344 (1985)
404. Takahata N.: On the overdispersed molecular clock. Genetics 116: 169–179 (1987)
405. Tateno Y.: A method for molecular phylogeny construction by direct use of nucleotide sequence data. J. mol. Evol. 30: 85–93 (1990)
406. Taylor W. R.: A flexible method to align large numbers of biological sequences. J. mol. Evol. 28: 161–169 (1989)
407. Taylor W. R. and Orengo C. A.: Protein structure alignment. J. mol. Biol. 208: 1–22 (1989)
408. Tegelström H., Nilsson G. and Wyoni P. I.: Lack of species differences in isoelectric focused proteins in the Formica rufa group (Hymenoptera, Formicidae). Hereditas 98: 161–165 (1983)
409. Tegelström H.: Genetic variability in mitochondrial DNA in a regional population of the great tit (Parus major). Biochem. Genetics 25: 95–110 (1987)
410. Thomas W. K. and Beckenbach A. T.: Mitochondrial DNA restriction site variation in the Townsend's vole, Microtus townsendii. Can. J. Zool. 64: 2750–56 (1986)
411. Ticher A. and Graur D.: Nucleic acid composition, codon usage, and the rate of synonymous substitu-

tions in protein-coding genes. J. mol. Evol. 28: 286–298 (1989)
412. Toha J., Soto M. A. and Chinga H.: Algorithm for construction of phylogenetic trees. Z. Naturforsch. Sect. C 44: 312–316 (1989)
413. Tomiuk J. and Wöhrmann K.: Enzyme variability in populations of aphids. Theor. appl. Genetics 57: 125–127 (1980)
414. Tsoi S. C. M., Lee S. C. and Chao W. C.: Duplicate gene expression and diploidization in an Asian tetraploid catostomid, Myxocyprinus asiaticus (Cypriniformes, Catostomodae). Comp. Biochem. Physiol. Pt. B 93: 27–32 (1989)
415. Turner B. J.: Genetic divergence of Death Valley pupfish species: biochemical versus morphological evidence. Evolution 28: 281–294 (1974)
416. Turner B. J. et al.: Genetic variation in clonal vertebrates detected by simple-sequence DNA fingerprinting. Proc. Nat. Acad. Sci. USA 87: 5653–57 (1990)
417. Ueda S. et al.: Nucleotide sequences of immunoglobulin-epsilon pseudogenes in man and apes and their phylogenetic relationships. J. mol. Biol. 205: 85–90 (1989)
418. Uitterlinden A. G. et al.: Two-dimensional DNA fingerprinting of human individuals. Proc. Nat. Acad. Sci. USA 86: 2742–46 (1989)
419. Ullrich A. et al.: Genetic variation in the human insulin gene. Science 209: 612–615 (1980)
420. Usha R. and Murthy M. R. N.: Protein structural homology: a metric approach. Int. J. Peptide Protein Res. 28: 364–369 (1986)
421. Vanherck M. et al.: The 18S ribosomal RNA sequence of the sea anemone Anemonia sulcata and its evolutionary position among other eukaryotes. FEBS Letters 269: 445–449 (1990)
422. Vawter L. and Brown W. M.: Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science 234: 194–196 (1986)
423. Vossbrinck C. R. et al.: Ribosomal RNA sequence suggests Microsporidia are extremely ancient eukaryotes. Nature 326: 411–414 (1987)
424. Wagner H. J.: The minimum number of mutations in an evolutionary network. J. theor. Biol. 91: 621–636 (1981)
425. Wake D. B., Maxson L. R. and Wurst G. Z.: Genetic differentiation, albumin evolution and their biogeographic implications in plethodontid salamanders in California and southern Europe. Evolution 32: 529–539 (1978)
426. Walsh J. B.: Interaction of selection and biased gene conversion in a mutigene family. Proc. Nat. Acad. Sci. USA 82: 153–157 (1985)
427. Ward R. D.: Relationship between enzyme heterozygosity and quaternary structure. Biochem. Genetics 15: 123–135 (1977)
428. Ward R. D., McAndrew B. J. and Wallis G. P.: Enzyme variation in the brook lamprey, Lampetra planeri (Bloch), a member of the vertebrate group Agnatha. Genetica 55: 67–73 (1981)
429. Watanabe T. et al.: Pig mitochondrial DNA: Polymorphism, restriction map orientation, and sequence data. Biochem Genetics 24: 385–396 (1986)
430. Watterson G. A.: On the time for gene silencing at duplicate loci. Genetics 105: 745–766 (1983)
431. Westbroeck P. et al.: Fossil macromolecules from cephalopod shells: characterization, immunochemical response and diagenesis. Palaeobiology 5: 151–167 (1979)
432. Whitmore D. H.: Identification of sunfish species by muscle protein isoelectric focusing. Comp. Biochem. Physiol. Pt. B 84: 177–180 (1986)
433. Whitney III J. B. et al.: Detection of neutral amino acid substitutions in proteins. Proc. Nat. Acad. Sci. USA 82: 7646–50 (1985)
434. Whittam T. S. et al.: Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. Proc. Nat. Acad. Sci. USA 83: 9611–15 (1986)
435. Wilbur W. J.: On the PAM matrix model of protein evolution. Mol. Biol. Evol. 2: 434–447 (1985)
436. Wilcox D. R. and Prakash S.: Variation in biochemical properties of allozymes of xanthine dehydrogenase in Drosophila pseudoobscura. Genetics 96: 927–938 (1980)
437. Williams J.: The evolution of transferrin. Trends biochem. Sci. 82: 394–397 (1982)
438. Williams S. A. and Goodman M.: A statistical test that supports a human/chimpanzee clade based on noncoding DNA sequence data. Mol. Biol. Evol. 6: 325–330 (1989)
439. Wills C.: Genetic variability. Clarendon Press, Oxford 1981
440. Wilson A. C., Carlson S. S. and White T. J.: Biochemical evolution. Annual Rev. Biochem. 46: 573–639 (1977)
441. Winberg J. O., Thatcher D. R. and McKinley-McKee J. S.: Drosophila melanogaster alcohol dehydrogenase: An electrophoretic study of the Adh-S, Adh-F, and Adh-UF alloenzymes. Biochem. Genetics 21: 63–80 (1983)
442. Winberg J. O. and McKinley-McKee J. S.: Drosophila melanogaster alcohol dehydrogenase. Biochemical properties of the NAD+-plus-acetone-induced isoenzyme conversion. Biochem. J. 251: 223–227 (1988)
443. Winter W. P. (ed.): Hemoglobin variants in human populations. CRC Press, Boca Raton USA 1986
444. Wolfe K. H., Sharp P. M. and Li W. H.: Mutation rates differ among regions of the mammalian genome. Nature 337: 283–285 (1989)
445. Wooley K. J. and Athalye M.: A use for principal coordinate analysis in the comparison of protein sequences. Biochem. biophys. Res. Commun. 140: 808–813 (1986)
446. Wooten M. C. and Smith M. H.: Large mammals are genetically less variable. Evolution 39: 210–212 (1985)
447. Wu C. I. and Li W. H.: Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Nat. Acad. Sci. USA 82: 1741–45 (1985)
448. Wyatt R., Odrzykoski I. J. and Stoneburner A.: High levels of genetic variability in the haploid moss Plagiomnium ciliare. Evolution 43: 1086–96 (1989)
449. Wysocki L. J. and Gefter M. L.: Gene conversion and the generation of antibody diversity. Annual Rev. Biochem. 58: 509–531 (1989)
450. Yafee D. et al.: Highly conserved sequences in the 3' untranslated region of mRNAs coding for homologous proteins in distantly related species. Nucleic Acids Res. 13: 3723–37 (1985)
451. Yager L. N., Kaumeyer J. F. and Weinberg E. S.: Evolving sea urchin histone genes – Nucleotide polymorphisms in the H4 gene and spacers of Strongylo-

centrotus purpuratus. J. mol. Evol. 20: 215–226 (1984)
452. Yonekawa H. et al.: Evolutionary relationships among five subspecies of Mus musculus based on restricting enzyme cleavage patterns of mitochondrial DNA. Genetics 98: 801–816 (1981)
453. Zakijan S. M. et al.: An estimation of the degree of the genetic divergence of sibling species Microtus arvalis and Microtus subarvalis (Rodentia) based on electrophoretic analysis. Biochem. Genetics 22: 1081–91 (1984)
454. Zera A. J.: Extensive variation at the α-glycerophosphate dehydrogenase locus in species of waterstriders (Gerridae: Hemiptera). Biochem. Genetics 19: 797–812 (1981)
455. Zharikh A. A., Solovyon V. V. and Kolchanov N. A.: Conformational changes in the globin family during evolution – I. Analysis of the evolutionary role of insertions and deletions. J. mol. Evol. 21: 42–53 (1984)
456. Zimmerman E. G., Kilpatrick C. W. and Hart B. J.: The genetics of speciation in the rodent genus Peromyscus. Evolution 32: 565–579 (1978)
457. Zimny-Arndt U. and Klose J.: Qualitative and quantitative variability in different classes of proteins: comparison of mouse and rat. J. mol. Evol. 24: 260–271 (1987)
458. Zuckerkandl E.: Evolutionary processes and evolutionary noise at the molecular level – I. Functional density in proteins. J. mol. Evol. 7: 167–183 (1976)
459. Zuckerkandl E.: Evolutionary processes and evolutionary noise at the molecular level – II. A selectionist model for random fixations in proteins. J. mol. Evol. 7: 269–311 (1976)
460. Special Issue: Frequency-dependent selection. Phil. Trans. Roy. Soc. London B 319: 459ff (1988)

5 Plasma Proteins, Yolk Proteins and Metal-Binding Proteins

5.1 The Variety of Plasma Proteins

The extracellular fluid of the metazoans is not only a transport vehicle but also, for the majority of the body's cells, their growth environment. Proteins play an important role here, providing colloid-osmotic pressure and acting as buffers. The most important parameter in this respect is their concentration which, depending upon the species, the developmental stage and the physiological conditions, can vary from less than 1 to more than 200 mg/ml (Table 5.1). In addition to these general functions, individual plasma proteins have various specific roles, e.g. in the transport of substances, in defence reactions, in blood clotting or in the solution of clots. At least in the case of the more highly developed animals, the plasma proteins may be looked upon as a well-defined extracellular system with certain general functions and regulation mechanisms; in all animals, they exist as a mixture of proteins of very different structures and functions.

Table 5.1. Protein concentrations (mg/ml) in the blood plasma or haemolymph of various animals (after Prosser 1961)

Group	Species	Concentration (mg/ml)
Annelida	*Pheretima* sp.	0.13
Mollusca	*Mytilus edulis*	1.5
	Anodonta sp.	3.0
	Sepia officinalis	109
Crustacea	*Nephrops norvegicus*	42
	Maja verrucosa	53
	Callinectes sapidus	20–231
Insecta	*Samia cynthia* pupae	43
	Bombyx mori	60
	Gastrophilus sp. larvae	108
Vertrebrales	*Scomber scombrus*	35
	55 different reptile species	29–78
	Rattus norvegicus	59
	Canis vulgaris	58
	Homo sapiens	75

5.1.1 Plasma Proteins of Vertebrates

The analysis of vertebrate plasma proteins is based upon investigations of man. The **number of known plasma proteins** has steadily increased in parallel with the resolution and sensitivity of the detection methods. In the previous century, a distinction was already made between albumin, which remains in solution even after substantial

dilution of the blood plasma, and the globulins, which are insoluble under these conditions. The first electrophoretic investigations carried out in the 1940s produced the well-known separation at pH 8.6 of the fastest, anodic albumin fraction and the slower globulin fractions α_1, α_2, β and γ. Fifteen years later, 20 components could be distinguished by gel electrophoresis and as many as 30–35 by immunoelectrophoresis. Today, detailed knowledge is available on more than 100 different proteins from 14 super-families in human blood plasma (Table 5.2), not including either the protein hormones or the enzymes originating in body cells, and without taking into account the genetic variability of the immunoglobulins and other plasma proteins. More than 95 % of the total plasma protein is accounted for by just eight components (Table 5.3).

Electrophoresis also reveals complex patterns of the plasma proteins in other vertebrates, although in cartilaginous fish the plasma proteins are at much lower concentrations and show less variety than in the higher vertebrates [25]. The further the animal is from the mammals, the less directly comparable are individual bands with those of human plasma proteins. For example, in the agnathans the fastest fraction is not an albumin but rather a mixture of glycoproteins that is, if anything, more comparable to human α-globulin; many other fish, amphibians and reptiles also present similar problems of protein identification [73]. With the exception of the albumins, all plasma proteins are glycoproteins. The carbohydrate portion increases the solubility of the plasma proteins due to the presence of many polar groups and the negative charge of the sialic acid; cleavage of the sialic acid serves as an indicator of degradation.

Table 5.2. Protein super-families in the blood plasma of mammals [63, 147]

1. Albumin, α-fetoprotein, vitamin D-binding protein
2. Immunoglobulins, β_2-microglobulin
3. Fibrinogen chains: α, β and γ
4. Serine proteases of blood clotting, fibrinolysis, the complement system, including haptoglobin
5. α_2-Macroglobulin, pregnancy zone protein (α_2-glycoprotein), complement factors C3, C4 and C5
6. β_2-Glycoprotein, complement factors B (non-enzymatic region) and H, β-chain of the blood clotting factor XIII
7. Retinol-binding protein, α_1-microglobin, α_{2u}-globulin (rodents), apolipoprotein apoD
8. Antithrombin III, α_1-antitrypsin, α_1-antichymotrypsin, angiotensinogen
9. Caeruloplasmin, clotting factors V and VIII
10. Lipid-binding proteins: A, B, C, etc.
11. Kininogens, β_1-microglobulin, acute-phase proteins, cysteine-proteinase inhibitors
12. Transthyretin (pre-albumin), glucagon, glycentin
13. β-Thromboglobulin, platelet factor 4
14. Serum amyloid protein (SAP) components, C-reactive protein (CRP)
15. Transferrins

Table 5.3. The major plasma proteins of man and other mammals [63]. These eight proteins and protein classes consistently make up more than 95 % of the total plasma protein

Component	Concentration (mg/ml)
Albumin	45
Immunoglobulins	15
Lipoproteins	10
Haptoglobins	6
Fibrinogen	3.5
Transferrin	3.0
α_1-Antitrypsin	2.9
α_2-Macroglobulin	2.6

The **classification of the plasma proteins according to function** in man, and presumably all other vertebrates, recognizes the groups of the multi-enzyme systems (complement, blood clotting, fibrinolysis, kinin production), antibodies, lipoproteins, transport proteins and protease inhibitors, as well as numerous α- and β-glycoproteins and trace components of unknown function. The antifreeze proteins of fish living in cold seas are a specific adaptation to an extreme environmental condition. The **concentration of the individual proteins** in humans varies across seven orders of magnitude, from 50 mg/ml for albumin to 5 ng/ml for IgE. In fact, this latter protein was only discovered because its concentration is 50 000 times higher than normal in patients with multiple myeloma. Essentially the same proteins are to be found in the extravasal space as in the blood; a few of the plasma proteins, e.g. fibrinogen, are almost entirely restricted to the blood itself. The concentration of proteins in the tissue and body-cavity fluids is 5–103 times lower than in the blood; however, as the extravasal space has a much greater volume than the blood plasma, it contains an equivalent amount of total plasma protein. In addition to the plasma proteins, the extracellular space contains other specific proteins, e.g. the coelomic fluid-specific protein (CFSP) in the body cavity of the salmon *Oncor-*

hynchus keta, a dimeric protein with 16-kDa subunits that is synthesized in the coelomic epithelium and the mesovarium. The function of CFSP is not known; immunologically similar proteins are found in all *Oncorhynchus* species but not in other salmon [160].

All **plasma proteins** are already present **in the embryo**, e.g. in humans from the 6th week, but in relative amounts that are different to those found in the adult. In pig foetuses, however, about one-third of the plasma proteins are not cross-reactive with the antisera against adult plasma [137]. Thus, there are specific foetal proteins, of which α-fetoprotein and fetuin have received particular attention. The synthesis of the majority of plasma proteins occurs in the liver; in embryos, the yolk-sac is also always involved. Whilst in humans, plasma protein synthesis in the yolk-sac ceases at about 12 weeks, it continues in rats and chickens up to the moment of birth or hatching, respectively. In contrast to the human and chicken yolk-sacs, that of the rat produces no albumin; investigations in other species have unfortunately not been carried out. Many vertebrates produce specific plasma proteins when in certain physiological states. For example, the body reacts to infection, inflammation or tissue damage by producing acute-phase proteins. Also relevant here are the mammalian pregnancy proteins, and the precursors of the yolk proteins that appear during the development of yolked eggs.

5.1.2 Plasma Proteins of Invertebrates

By means of high-resolution electrophoretic methods, more than 100 different polypeptides with molecular masses of 15–200 kDa have been demonstrated in the haemolymph of the mussel *Mytilus edulis* [245]. In insect haemolymph, 10–30 protein fractions can be electrophoretically separated and grouped, according to function, into the vitellogenins, lipoproteins, hormone-binding proteins, storage proteins, defence proteins and enzymes. The haemolymph proteins show species-specific changes during each developmental stage. In the haemolymph of older larvae, however, there are usually only a few protein types, in particular the larval haemolymph proteins (LHPs) and the lipid-transporting lipophorins. Their concentration is low (10–20 mg/ml) at the beginning of each developmental stage but reaches high values at each ecdysis, e.g. in Lepidoptera 60–100 mg/ml and in bluebottle larvae as much as 200 mg/ml [87, 217]. Synthesis of the haemolymph proteins occurs mainly in the fat bodies; it has been shown, however, in tracer experiments with caterpillars of the butterfly *Calpodes ethlius* that the mid-gut and epidermis cells are involved in the synthesis of arylphorins and other haemolymph proteins [197, 198].

In animal species with haemoglobin or haemocyanin freely dissolved in the plasma, these two substances make up the largest fraction of the plasma proteins. Other non-respiratory proteins are found in many chelicerates and decapod crustaceans but, as yet, little is known about either their structure or function. There is no real basis for the assumption that they are mainly copper-free "apohaemocyanins". The coelomic fluid of the annelids and the haemolymph of the molluscs [3] also contain up to 20 proteins that are detectable by gel electrophoresis (30–60 are seen on 2-D electrophoregrams) but about which little is known.

5.2 Serum Albumin and α-Fetoprotein

Albumin is the most easily identifiable plasma protein of the higher vertebrates due to its solubility at low ionic strength, low molecular mass, high charge, particular binding characteristics and paucity of carbohydrates. At a concentration of 35–50 mg/ml, it makes up about 60% of human plasma protein; it also forms the largest fraction in the lower vertebrates. Because of its relatively high concentration and low molecular mass, albumin is the main factor determining the colloid-osmotic pressure of the plasma. In the tadpoles of *Rana catesbeiana*, its concentration is less than 1 mg/ml and only increases markedly during metamorphosis. Serum albumins have also been isolated from non-mammalian species; they appear to correspond broadly in their molecular masses and electrophoretic behaviour to those of the mammals and man [30, 57, 92, 154]. In the agnathans and cartilaginous fish no plasma protein has so far been found which corresponds to albumin in its physicochemical and functional properties, such as the ability to bind fatty acids [73]. The so-called pre-albumins of many vertebrates appear to have nothing in common with real albumins, except a high electrophoretic mobility.

Complete **sequences** have been obtained, either directly or via cDNA, for human, bovine, sheep, pig, rat and mouse albumins and the two albumins of the clawed frog *Xenopus laevis*. The albumins consist of a single polypeptide chain of

581–584 amino acids corresponding to a molecular mass of about 66 kDa. The long, strung-out molecule has dimensions of 4 × 14 nm and is subdivided into three domains which, despite similarities in structure, appear to have differences in function [36, 92]. The significant sequence homology between the three domains indicates that albumin arose by the triplication of a polypeptide of about 190 amino acids. Of physiological importance is the binding capacity of albumin for organic anions like gall pigments, steroid and thyroid hormones, fatty acids and drugs. A high-affinity fatty-acid-binding site is found in domain III; evolutionary rudiments in the form of low-affinity fatty-acid-binding sites are detectable in domains I and II. The binding site for bilirubin lies on domain II and that for indole on domain I. There is a marked species-specifity in the binding properties of the albumins. For example, bilirubin binds much more strongly to human and chicken albumin than to bovine or rabbit albumin; frog albumin, in contrast to the others, cannot bind tryptophan. Albumin, like almost all other plasma proteins, is synthesized in the liver. The prepro-albumin formed initially has an N-terminus which bears a signal sequence of 24 amino acids as well as a propeptide of 5–6 amino acids; pro-albumin has been isolated from the liver of various mammals and birds.

Serum albumin belongs to the proteins that show a high rate of evolution (see Table 4.12, p. 161). This fact, and the ease with which albumin can be isolated, makes it especially suitable for **immunological studies of relationships** by the method of microcomplement fixation (p. 151). Such investigations have now been carried out on many hundreds of vertebrate species. Positive immunological cross-reactions between serum albumins, corresponding to sequence differences of not more than 40 %, are only obtained between vertebrates of the same class; one exception is the positive reaction between the alligator and the chicken. Thus, an albumin comparison is suitable, above all, for clarifying relationships at the level of the family or order. **Multiple albumins** are known from the polyploid species of the genus *Xenopus* (see Table 4.6, p. 126): tetraploid species have two, octaploid species two or three and the dodecaploid *X. ruwezoriensis* three albumins, which apparently can differ significantly in size and sequence. The peptide cleavage patterns are species specific; thus, it is seen that the tetraploid species *X. andrei* possesses two *fraseri* and one *vestitus* albumin and is actually an allopolyploid that arose by hybridization of these latter two species [90]. Genetic variants of serum albumins are known from man and several vertebrates. There are human and rat "analbumin" individuals whose blood plasma, however, still contains 5–25 μg albumin/ml. Although this is a 1000 times less than the normal concentration, it should not be concluded that serum albumin is dispensable. The heritable analbumin in one rat line is the result of a failure to transcribe exon H because of a 7-bp deletion in the preceding intron [228].

The predominant plasma protein of all mammalian embryos is **α-fetoprotein**, which on electrophoregrams, appears in the region of the α_1-globulins, but belongs to the same super-family as the serum albumins. In humans, the maximal concentration of ca. 3 mg/ml is reached in the 13th week of embryo development; in adults it is found at only ng/ml concentrations, except in cases of liver regeneration or liver tumours when higher values are recorded. The α-fetoprotein is apparently not just a passive plasma protein but has various immunological and growth-regulating functions [170]. The amino acid sequences, obtained from cDNA or gene sequences, of human, rat and mouse fetoproteins agree to about 35 % with those of the albumins. The rate of evolution of the α-fetoproteins is even higher than that of the albumins and is comparable with the rates for the pseudogenes [86]. In contrast to the non-glycosylated albumins, α-fetoprotein carries at least two oligosaccharide chains. Serum albumin and α-fetoprotein are each encoded by single genes which lie close to each other on the same chromosome. Both genes consist of 15 exons; exons 3–14 are subdivided into three similar sets of four exons, each of which makes up a domain. In mammalian foetuses, the α-fetoprotein is made in the yolk–sac, the liver and the gut wall. This tissue-specific expression is controlled in the mouse by three enhancer sequences in the 5′ flanking gene region [183]. Artificial "minigenes" with enhancer I are expressed only in the gut cells, those with enhancer II in the liver and yolk-sac, and those with enhancer III in the yolk-sac and the gut [94]. Small amounts of mature albumin and α-fetoprotein mRNAs are also detectable in rat kidney and pancreas [184].

In the neighbourhood of these two genes is a further gene of similar structure which encodes the very important **vitamin D-binding protein** (DBP) of the blood plasma; this is also known as group specific component (Gc). Serum albumin, α-fetoprotein and DBP belong to the same super-family; they have about 20 % identical amino

acids and the same periodicity of structure with 15 exons and 3 mutually homologous protein domains. It has been calculated from sequence comparisons that the separation of DBP occurred 560–600 million years ago, and the separation of albumin and α-fetoprotein about 280 million years ago, i.e. shortly after the separation of the reptiles from the amphibians. Accordingly, one should find larval α-fetoprotein in the reptiles but not in the fish or amphibians; up to now this hypothesis has only been tested in *Xenopus* tadpoles, where α-fetoprotein is, in fact, not present [92]. A further protein, **fetuin**, is found in the foetal plasma of the ungulates (cattle, sheep, goats, pigs) at concentrations up to 5 mg/ml but it is only present in trace amounts in adult animals. Immunological analysis has demonstrated the presence of fetuin in all mammals tested. In sheep and bovine embryos, fetuin is also found at high concentrations in the cerebral cortex, and it is detectable immunologically in trace amounts in the brains of human embryos. In the latter case, significantly greater amounts are found of the α_2HS glycoprotein, which has 70 % amino acids in common with fetuin [66].

5.3 Plasma Proteins with Special Binding and Transport Functions

Plasma proteins are involved in many ways in the transport activities of the blood. Relatively non-specific binding of organic anions to serum albumin occurs in vertebrate blood, and, in addition, there are specific plasma proteins binding iron (transferrin), copper (caeruloplasmin), haemoglobin (haptoglobin), haem (haemopexin), thyroxin (thyroxin-binding protein), steroid hormone (transcortin), vitamin B (transcobalamin) and lipids (lipoproteins). Corresponding transport proteins are also present in the invertebrates, although these have been less thoroughly investigated, e.g. the juvenile hormone and ecdysone-binding proteins of the insects, and the lipoproteins of various arthropods.

5.3.1 Transferrin

Free Fe^{3+} ions cannot exceed a concentration of ca. 10^{-17} mol/l in neutral solution without forming insoluble ferric oxide. In vertebrates, which exhibit a very intensive iron metabolism in connection with haemoglobin synthesis, iron is transported between the sites of resorption, storage, use and elimination by specific transport proteins. The most important are the **serum transferrins;** these are glycoproteins of 65–85 kDa which are red in an iron-saturated condition (λ_{max} = 470 nm). Under normal physiological conditions they are, for example in humans, up to only 30 % saturated, and thus only 3–5 mg of iron, which represents 1/1000th of the total body iron, is found dissolved in the blood plasma. In female birds, more than 50 % of the serum iron is bound to vitellogenin. Each transferrin molecule can bind two iron atoms. This process is unique in so far as one bicarbonate is bound per iron. The affinity of transferrin for iron is extraordinarily high and varies very little between different mammalian species [271].

The transferrins also have antimicrobial activity, denying microorganisms essential iron by virtue of their high affinity. The uptake of iron, e.g. into haemoglobin-synthesizing reticulocytes, occurs via receptor-mediated endocytosis at coated pits. The primary structure of human **transferrin receptors** has been determined from the mRNA sequence. It is a dimeric transmembrane protein with two polypeptides of 90 kDa and three N-bound oligosaccharide chains, as well as phosphoric acid and fatty acid residues. Of the 760 amino acids in each chain, the first 62 form the cytoplasmic region, the following 26 the transmembrane part, and the remaining 672 make up the extracellular domain [254]. The binding of transferrin to the receptor is not species specific, but it is, to a limited extent, group specific. Thus, the receptors of higher mammals (placentals) bind transferrin of other Placentalia but not that of pouched animals (marsupials), birds or amphibians [148].

There are three types of vertebrate transferrin: in addition to serum transferrin (siderophilin), which is found in all vertebrates, there is the **ovotransferrin** (conalbumin) of bird eggs, and **lactotransferrin** (lactoferrin), which occurs in the milk of all mammals and also in tears and leukocytes. Ovo- and lactotransferrin are probably mostly antimicrobial in function. The three transferrins comprise a protein super-family. The ca. 700-amino-acid-long polypeptide chain of human lactotransferrin shows about 59 % agreement with human serum transferrin and about 49 % with chicken ovotransferrin. The serum transferrin and ovotransferrin of the chicken differ only in the carbohydrate components and not in the amino acid sequence. Whilst human and bovine milk transferrins differ markedly from those of the

blood plasma, in the rabbit they are identical, apart from differences in the sugar residues [119, 201, 271].

The polypeptide chains of all transferrins are subdivided into two domains which agree in 37–40 % of their sequences and each carry one binding site. These binding sites consist of three tyrosine and two histidine residues as well as an HCO_3^- ion bound electrostatically to an arginine residue. The internal periodicity of the **amino acid sequence** indicates that all transferrins arose by the duplication of a short ancestral sequence; this is recognizable from the spatial structure and from the gene sequence [11, 119]. The sequences of the two domains in human serum transferrin and in chicken ovotransferrin have, in each case, only about 40 % identical positions and thus differ to a greater extent than the homologous domains of both proteins, which share 50 % identity. Thus, the duplication of the ancestral transferrin appears to have occurred before the separation of the different types. The evolutionary advantage of the gene duplication may stem from the failure of the kidneys to retain the ancestral form of transferrin because of its low molecular weight [271]. Comparative investigations of numerous fish, birds and mammals show similar molecular masses of the serum transferrin polypeptide chains, ranging from 61 to 87 kDa [24]. Transferrin belongs to the proteins with a high **rate of evolution;** the sequence agreement between human and pig transferrin is 70 %, and between human and *Xenopus laevis* is only 46 % [15, 176]. The ovotransferrins and serum transferrins of all vertebrate classes show unusually high variability. This is due in part to differences in the carbohydrate fraction (sialic acid content) and in part to allelic variation in the amino acid sequence [47]. The **polymorphism** of the transferrins has often been exploited for studies of population genetics.

Although the problem of Fe^{3+} transport also exists for the invertebrates, little is known about iron transport proteins. They have been detected, for example, in the ascidian *Pyura stolonifera*, the spider *Dugesiella hentzi* and the pocket crab *Cancer magister.* A haemolymph protein of 280 kDa, isolated from the horseshoe crab *Limulus polyphemus*, consists of ten subunits each with two iron-binding sites. In the butterfly *Manduca sexta* an 80-kDa monomeric haemolymph protein containing one iron atom is present together with an iron-rich protein of 490 kDa; this latter protein has subunits of 24 and 26 kDa and is very similar structurally and functionally to ferritin [16, 111, 258].

5.3.2 Haptoglobin and Haemopexin

Haptoglobins (Hp) are α_2-glycoproteins which bind in a ratio of 1:1 to the haemoglobin released from disintegrated erythrocytes. The whole haptoglobin-haemoglobin complex is taken up and metabolized by cells of the reticulo-endothelial system; the haptoglobin molecules are not recycled. Haptoglobins are widely distributed in mammals and birds but have not yet been definitely identified in amphibians. In the urodelan *Taricha granulosa*, the haemoglobin in haemolysed blood binds to albumin. Whilst a dramatic increase in α_2-glycoproteins, including haptoglobin, can be induced in higher vertebrates by, for example, the injection of turpentine, this acute-phase reaction is absent from *Taricha* [77].

The haptoglobin molecule is a tetramer with two types of subunit bound by disulphide bridges: smaller α-chains of 9 or 17 kDa and larger, glycosylated β-chains of 40 kDa. Canine Hp lacks the disulphide bridges between the dimers and the rabbit Hp molecule has no disulphide bridges; therefore, in SDS or urea the canine Hp splits into dimers and the rabbit Hp into single chains [136]. There are three **allelic variants** of the human α-chain, one of which is increased in length from 83 to 142 amino acids [152]. In the neighbourhood of the human hp gene there is a related gene, hpr (hp-related), which contains a retroviral element in the first intron. Anthropoids and Old World apes have three hp genes (hp, hpr, hpp) but New World apes have only one. There is also a retroviral element in the first intron of the chimpanzee hpr gene and a further one in the spacer between hpr and hpp. The triplication of the hp gene apparently occurred after the separation of the apes of the Old World and the New World; in humans, a gene was lost during unequal crossing-over between hpr and hpp [162].

The amino acid sequence of the human haptoglobin β-chain has ca. 30 % agreement with different serine proteases (see Fig. 3.6, p. 91). There is 10–20 % difference between the N-terminal sequence of the human haptoglobin β-chain and those of various other mammals. Immunologically, the haptoglobins of different Artiodactyla show strong mutual cross-reactivity but only weak interaction with human Hp; thus, haptoglobin shows a medium **rate of evolution** similar to that of haemoglobin. Chicken haptoglobin is very different, both structurally and functionally, from that of mammals; full sequence analysis is needed in order to determine whether it belongs to the same protein super-family: it is a 68-kDa mono-

meric protein with 11–12 disulphide bridges in the chain. It binds haemoglobin from bird and reptile blood but not from human blood; conversely, however, mammalian haptoglobin binds chicken haemoglobin [60]

The **haemopexins** are β-glycoproteins that bind free haem and transport it to the liver cells, where it is degraded to gall pigments and the iron is bound to ferritin; the haemopexin returns to the bloodstream. Human haemopexin has a native molecular mass of about 600 kDa with 20 % carbohydrate; the polypeptide chain of 439 amino acids contains at least eight repeats of ca. 45 amino acids [4]. Haemopexin appears to exist in all classes of vertebrates and apparently has a relatively low rate of evolution. Antisera against human haemopexin react with the haemopexins of all placental mammals but not with those of the marsupials or other vertebrate classes. As in the case of haptoglobin, chicken haemopexin differs markedly from that of man and the mammals: it migrates in the α_1 instead of the β fraction during electrophoresis and contains different carbohydrate components [194].

5.3.3 Caeruloplasmin and Pre-Albumins

Human **caeruloplasmin** is a blue α_2-glycoprotein of 132 kDa which carries over 90 % of the copper transported in the blood plasma. There are six or eight copper-binding sites with a high affinity and a further ten with a lower affinity. In addition, the caeruloplasmin molecule has the enzyme activities of a ferroxidase, aminoxidase and superoxide dismutase. It is synthesized in the liver and increases in concentration as an acute-phase protein in response to inflammation. The polypeptide chain of human caeruloplasmin consists of 1046 amino acids and is subdivided into three domains. There is 93 % agreement with the protein from the rat. A similar domain structure is found in the clotting factors V and VIII, which belong to the same protein super-family [75, 133]. Caeruloplasmins appear to be generally distributed throughout the vertebrates and are also identifiable in the cartilaginous fish on the basis of their copper content and aminoxidase activity [25, 34, 105].

Transthyretin or **pre-albumin**, which runs in front of albumin on electrophoregrams of vertebrate plasma, has been identified in man and other mammals as a protein with a double transport function: it binds thyroxin and also the retinol-binding protein. The protein is a homotetramer with 14-kDa subunits of 127 amino acids that form a central canal containing the thyroxin-binding site. The amino acid sequences from man, sheep, rabbit, rat and mouse have been determined either directly or via cDNA; they show 80 % agreement and also have significant homology to human thyroxin-binding-protein, a further, functionally more important transport protein for the thyroid hormone [114, 246, 260]. Transthyretin is made in adult and foetal livers and in the yolk-sac of early embryos. Large amounts of transthyretin mRNA are also found in the chorioid plexus on the inner surface of the neural canal; thus, transthyretin is apparently important for the transport of the thyroid hormone from the blood to the brain and the spinal cord. Transthyretin made in the retina is possibly involved in the transport of retinal [81, 158].

5.4 Acute-Phase Proteins

One response to bacterial infection or certain irritants by man and many mammals is a drastic increase in the concentration of several plasma proteins (acute-phase reaction). In the human acute phase, 100- to 1000-fold increases in concentration occur for the C-reactive protein (CRP) and the serum amyloid protein (SAP), and four- to ten-fold increases, for example, in the case of the acid α_1-glycoprotein or orosomucoid, α_1-antitrypsin, α_2-macroglobulin, haptoglobin, haemopexin, fibrinogen and several complement factors. An acute-phase reaction is also observed in many mammals but both the spectrum of proteins and the increases in their concentration vary with the species [50]. The most important regulator of the acute-phase reaction is interleukin-6.

The **C-reactive protein** (CRP) gets its name from its ability to precipitate the C-polysaccharide of *Streptococcus pneumoniae* in the presence of Ca^{2+}. It has a very unusual molecular structure in which five identical subunits of 22.5 kDa are associated non-covalently into disk-shaped pentamers. The evolution of CRP is relatively slow; the 205 amino acids of the mouse and rabbit CRPs agree 70–80 % with human CRP [270]. A CRP induced by pathogenic bacteria has also been found in the trout *Salmo gairdneri* [178]. The protein super-family (**pentraxine**), of which CRP is a member, also includes the **serum amyloid protein** (SAP), which is constructed of two pentameric disks, and the SAP-related, female-specific protein (hamster female protein,

HFP) of the Syrian golden hamster [206]. The known mammalian pentraxins share 69 invariant amino acids [270]. CRP and SAP can be easily identified on the basis of their binding specificities: CRP shows specific affinity for phosphorylcholine and SAP for agarose. Both proteins are found in a wide range of very different vertebrates, including the cartilaginous and the bony fish; the CRP of the dogfish *Mustelus canis* is even cross-reactive with that of the rabbit [210]. It is noteworthy that CRP-like proteins have also been detected in an invertebrate, the horseshoe crab *Limulus polyphemus*. The CRP in this case has a concentration of 1–5 mg/ml and is a constitutive, major component of the haemolymph; in contrast to the vertebrate protein, it consists of non-identical, 24-kDa subunits in a 2×6 arrangement. Three CRP genes from *Limulus* have been sequenced. The encoded polypeptides of 218 amino acids differ from each other by up to 10 % but agree to about 25 % with human CRP, and even more so in two conserved regions. *Limulus* possesses more than three CRP genes compared with the one gene of man [191, 192].

5.5 Larval Haemolymph Proteins of Insects

A type of haemolymph protein was discovered in 1969 in the fly *Calliphora vicina* (identical to *C. erythrocephala*); this protein appears only in older larvae and the pupa, and disappears again during metamorphosis to the imago. Subsequently, similar proteins were demonstrated in other dipterans, various lepidopterans, the bug *Rhodnius prolixus* and the cockroach *Blatta orientalis* [121, 128, 129, 207]. Their function is apparently that of amino acid storage during the periods of ecdysis and metamorphosis when there is intensive metabolism but no nutrient intake. Initially, these proteins were named according to their source, e.g. calliphorin, lucilin, manducin etc.; today, they are generally known as larval haemolymph proteins (LHPs) or, in view of their high aromatic amino acid content, as arylphorins.

LHPs are synthesized in the fat bodies of older larvae and secreted into the haemolymph. In the Lepidoptera, the arylphorin genes are also expressed in other tissues, although at lower levels than in the fat bodies [121, 165]. At the end of the feeding phase, LHPs make up 60–80 % of the total haemolymph proteins, i.e. up to 6 % of larval fresh weight. During the migration phase or shortly before pupation, the LHPs are taken up again by the fat bodies and stored in 1.5- to 3.0-μm-large granulae. The **recovery** of LHPs **into the fat bodies** has been studied in *Sarcophaga peregrina* and *Musca domestica*. There is a characteristic sequence in the house fly with a maximum at larval migration. The LHPs from *Drosophila melanogaster* and *Rhynchosciara americana* can also be taken up by the fat bodies of *Musca*, although that of *R. americana* in fact shows no cross-reactivity with the *Musca* LHP. The responsible membrane receptor therefore appears to have no great specificity [121, 155]. LHPs disappear completely during the histolysis that occurs in the pupa, but little is known in detail about their degradation. It has been shown in *Calliphora vicina* that the ubiquitin which is present is not involved in the intracellular cleavage of calliphorin [146].

The **biological functions** of the LHPs are not yet completely clear. It can be assumed that LHPs provide amino acids for the synthesis of structural and nutritional substances during metamorphosis, including the aromatic amino acids for the sclerotization of the cuticle. However, arylphorins are also found intact in the sclerotized cuticle of the fly *Calliphora vicina* and the butterfly *Manduca sexta;* the LHPs of *Drosphila, Calliphora* and *Manduca* are polymerized in vitro under oxidizing conditions by sclerotizing substances such as N-acetyldopamine and N-β-alanyldopamine. Thus, the arylphorins appear to be structural components of the sclerotized cuticle. The LHPs also probably have other functions, e.g. as transport proteins for ecdysteron. A strain of *Drosophila melanogaster* which cannot make arylphorin because of a genetic defect showed a 15-fold reduction in fertility during a 2-year observation period but there were no negative effects on survival [121].

The LHPs have a very unusual **amino acid spectrum** with a total of 17–26 % phenylalanine and tyrosine. Small amounts of carbohydrates could be detected in all examined LHPs and in some cases there was also a lipid component [121, 145]. The structure of the N-glycosyl chain has, as yet, only been analysed in the butterfly *M. sexta*, where it turns out to be the typical asparagine-bound oligosaccharide $Man_9GlcNAc_2$ (see Fig. 13.17, p. 491) [216, 224]. The LHPs are hexamers of 450–500 kDa with subunits of 72–83 kDa. Multiple LHPs are found in several species and are referred to, for example, as LHP-I, II, etc. In these cases, several types of subunit have been detected which combine to form differ-

ent hexamers. Only one type of arylphorin monomer is found in the lepidopterans *Calpodes ethlius, Heliothis zea, Hyalophora cecropia* and *Papilio polyxenes* and in the honey bee *Apis mellifera;* two occur in the lepidopterans *Bombyx mori, Galleria melonella* and *Manduca sexta* and in the cockroach *Blatta orientalis;* and three occur in the dipterans *Drosophila melanogaster, Ceratitis capitata* and *Musca domestica*. In *Sarcophaga peregrina, Lucilia styga* and *Calliphora vicina* there are several other components in addition to the main subunit. The genes for the LHP subunits form a multi-gene family that in the fly *Calliphora vicina*, for example, includes 20 members. The heterogeneity of the LHPs is increased even further by genetic variability, in some cases to a quite unusual degree. Thus, no less than 90 LHP phenotypes were found in 180 flies of the species *Lucilia cyprina* from an Australian population, the genetic analysis of which indicated at least 12 homologous genes. The random combination of isoforms and allelic variants should result in a host of different LHPs [121]. LHP cDNAs or genes from the dipterans *S. peregrina* and *D. melanogaster* and from the lepidopterans *B. mori, M. sexta* and *Heliothis virescens* have been completely or partially sequenced [79, 121, 140, 187, 272]. The two LHP subunits α and β from *M. sexta* agree by 68 % in their 686–687 amino acids. Surprisingly, the *Manduca* LHP is significantly homologous to the respiratory blood pigment haemocyanin from the crab *Panulirus interruptus*. The highest sequence similarity is found in the region of the polypeptide chain which, in both proteins, forms the contact between the subunits of the hexamers [272]. Stage-specific expression of arylphorin genes is regulated in the fly *Sarcophaga peregrina* by a protein which only appears in 46-hour-old larvae and binds specifically to the sequence ACCACAACA in position –255 to –247 [130].

Apart from the arylphorins, a second type of larval haemolymph protein is found in the lepidopterans; these contain 4–8 % methionine and are therefore known as "methionine-rich storage proteins". They are synthesized only in the final larval stage and reach higher concentrations in female than in male animals. *Spodoptera litura* has two proteins of this type, whereas *Bombyx mori* and *Hyalophora cecropia* have only one. The methionine-rich proteins are significantly homologous to the arylphorins and have the same hexameric structure [121]. In *B. mori* the methionine-rich protein SP-1 agrees in 30 % of its sequence with the arylphorin SP-2 [79]. *Drosophila melanogaster* possesses a larval haemolymph protein, LSP-2, which is rich in aromatic amino acids but shows no immunological cross-reactivity to the arylphorins. Animals with null alleles for LSP-2 mostly die during development, and in any case remain sterile. Similar proteins are found in other dipterans [121]. In addition to the typical arylphorin and methionine-rich proteins, there is a further haemolymph protein, found in the pupae of *Hyalophora cecropia*, which is rich in histidine and contains bound riboflavin and copper. A further hexameric haemolymph protein of ca. 500 kDa, which is present in the larvae of the migratory locust *Locusta migratoria* but is missing in the adult, fits into none of the known classes of haemolymph proteins in holometabolic insects. There are, in all probability, further classes of as yet unknown larval haemolymph proteins in other insect groups [121].

5.6 Plasma Lipoproteins

In order to be transported between the sites of resorption, de novo synthesis, conversion, and use, hydrophobic lipids must be solubilized by binding to proteins of the blood plasma. Consequently, **transport lipoproteins** are to be expected in many animals but have so far only been examined in any detail in vertebrates and insects. Both groups of animals contain spherical lipid-protein particles which can, in each case, be resolved by electron microscopy but which have very different molecular architectures. In vertebrates, there is a nucleus of non-polar lipids surrounded by a coat of proteins and polar lipids (phospholipids and cholesterol); in insects, on the other hand, the nucleus of the lipoprotein particles consists mainly of proteins and is encased predominantly by relatively polar lipids (diacylglycerol). Specific lipoproteins, **vitellogenins**, are found in the blood plasma of animals that produce yolked eggs, and are also incorporated into the yolk material in more or less modified forms. Three methods are usually used to examine lipoproteins: electrophoresis and lipid-specific staining give information about the availability and variety of the lipoproteins, and density gradient centrifugation or fractionated flotation in media of differing density by the help of ultracentrifugation serves for their isolation. This takes advantage of the reduction in density of particles which occurs with increasing lipid content. In the investigation of human lipoproteins, the **density**

Table 5.4. Classes of human lipoproteins [55]

	Chylomicrons	VLDL	LDL	HDL
Density (g/cm^3)	0.950	0.950–1.006	1.019–1.063	1.063–1.210
Particle diameter (mm)	80–500	30–90	20–25	8–12
Major apoproteins	apoC-I–III apoB apoA-I apoA-IV	apoB apoE	apoB	apoA-I apoA-II
Protein content (%)	1–2	8–10	25	50
Lipid composition (%):				
Triacylglycerol	86–95	45–65	4– 8	2– 7
Phospholipids	2– 7	15–20	18–24	26–32
Cholesterol (esters)	3–10	20–30	51–58	18–25

VLDL, Very low-density lipoproteins; LDL, low-density lipoproteins; HDL, high-density lipoproteins; VHDL, very high-density lipoproteins (density greater than 1.210 g/cm^3)

classes are chosen such that each includes the lipid complexes of particular apolipoproteins (Table 5.4). The application of the same density classes to other animals is questionable as the classes defined in man quite possibly do not represent a uniform apolipoprotein spectrum in other animals.

5.6.1 Plasma Lipoproteins of Vertebrates

Human plasma lipoproteins have been intensively studied because of their clinical importance in atherosclerosis and coronary disease; the investigations of rat and various primates, therefore, have been primarily undertaken with the aim of finding a model for the human situation [2, 46, 103, 223]. In fact, it appears that the lipoproteins of all vertebrates correspond in their basic structure to the model developed for man: a nucleus of neutral lipids surrounded by a layer of proteins and polar lipids (phospholipids, cholesterol). The specific density is mainly determined by the ratio nucleus: coat (Table 5.4). The chylomicrons are actually lipid droplets with a thin covering of proteins and phospholipids; in the other lipoprotein particles; the outer envelope constitutes at least 35–40 % of the mass [39]. Both the total concentration of the plasma lipoproteins and the relative **proportions of the individual density classes** in vertebrates vary considerably. Among the domesticated mammal species one already finds very different lipoprotein profiles [88, 108], and even greater deviation is seen amongst the amphibians and fish [8, 39]. In particular, the teleosts may have lipoprotein concentrations that are more than 10-times higher than those found in humans. However, in making such comparisons, one must take into account the dependence of the values on nutritional status and other factors.

In spite of large quantitative differences, the **spectrum of lipids** in the individual lipoprotein classes of all vertebrates is essentially the same. Triacylglycerols predominate in the chylomicrons and very-low-density lipoproteins (VLDL), whereas sterols and sterol esters predominate in low-density (LDL) and high-density lipoproteins (HDL). In certain fish, a portion of the triacylglycerols is replaced by other neutral fats; thus, up to 19 % hydrocarbon is found in the VLDL and LDL of sharks and sturgeons, and in the LDL of other shark species there is up to 24 % monoalkyl-diacylglycerol [39]. Each lipoprotein particle contains between 10 and over 100 protein molecules (apolipoproteins) with molecular masses between 12 and over 500 kDa. Ten different types of **apolipoprotein** are found in humans. The main protein component of the chylomicrons, VLDL and LDL is apoB, of which there are two forms of different size: apoB-100, which with a length of 4536 amino acids and a mass of 512 kDa belongs to the largest known polypeptides, and apoB-48, the sequence of which is identical to the 2151 N-terminal amino acids of apoB-100. There is, however, only one apoB gene and its primary transcript is processed to apoB-100 mRNA. RNA editing results in a C/U substitution in one part of the transcript such that the codon 2153 CAA(Gln) becomes the stop codon UAA. The C-terminal amino acid 2152-Met of the apoB-48 is cleaved post-translationally, and the mature apoB-48 thus has 2151-Met as the C terminus. Contrary to previous results, both mRNAs are found in the liver and small intestine of humans

as well as rats [252]. A smaller apoB form is missing in the chicken [250]. The human apolipoproteins A-I, A-II, A-IV, C-I, C-II, C-III, D and E have lengths between 57 (apoC-I) and 377 amino acids (apoA-I and apoA-IV). Just recently, the rather curious apolipoprotein J was discovered in a particular subclass of human HDL. This is synthesized as a 427-amino-acid polypeptide which is cleaved post-translationally at 205-Arg/206-Ser; the cleavage fragments apoJα (205 amino acids) and apoJβ (222 amino acids) are held together by disulphide bridges. In contrast to all other apolipoproteins, apoJ is synthesized in almost all cell types, the highest concentrations being found in brain, ovary, testis and liver [236]. Several isoforms are known for many human apolipoproteins, e.g. allelic sequence variants (apoE) or the products of post-translational glycosylation (apoA-II, apoB, apoC-III, apoE), deamidation (apoA-I), acylation or phosphorylation [68]. As lipoprotein particles may contain hexoses, hexosamines and sialic acids up to a total concentration of 9 %, they may be considered to be glycolipoproteins [39].

The different types of lipoprotein particles have different **transport functions**. In most vertebrates, fats reabsorbed in the small intestine are transported away by chylomicrons, which are formed in the cells of the intestine and are introduced into the blood via the lymph. In birds, however, the reabsorbed lipids are included into VLDL and directly enter the blood in the portal vein. The chylomicrons are rapidly degraded in the blood. **Lipoprotein lipases**, located in particular on the surface of capillary endothelial cells, release fatty acids and monoacylglycerols, which are transported to the neighbouring cells with the help of serum albumin. The remains of the chylomicrons, which retain only the proteins apoB and apoE and are rich in cholesterol esters, are taken up into liver cells. Fats synthesized in the liver are secreted into the blood as VLDL; the HDL is also formed in the liver. In contrast, the fatty acids released from fat tissues are bound directly to serum albumin. The VLDL particles formed in the liver serve primarily to transport triacylglycerols or fatty acids to the peripheral tissues. Due to the action of lipoprotein lipases, the VLDL particles lose triacylglycerols and are converted to particles of intermediate density (IDL). Non-esterified cholesterol of the VLDL is transferred to HDL, esterified by **lecithin: cholesterol acyltransferase (LCAT)** and returned to the IDL. With the removal of further triacylglycerols and all apoproteins except apoB, the IDL is converted to LDL, which transports cholesterol to the peripheral tissues for utilization and to the liver for degradation. ApoB and apoE are mainly responsible for the binding to LDL receptors [147]. At least two other proteins are encoded by the same multi-gene family as the LDL receptor: the LDL receptor-related protein (LRP) and glycoprotein 330. LRP appears to be the receptor for the remnants produced by chylomicron lipolysis. In the chicken, there are receptors on the oocyte membrane for the yolk protein precursor VLDL and vitellogenin, and a receptor on the body cells which is mainly involved in cholesterol metabolism. Interestingly, only the oocyte receptor is similar to the human LDL receptor [240]. The liver cell receptors of the carp *Cyprinus carpio* bind human and carp LDL and, conversely, carp LDL also binds to human LDL receptors [71].

Comparative investigations of all classes of vertebrates have revealed large variation in LCAT activity, e.g. compared with the value for man there are lower values in the rat, sheep, cow and dog, and higher values in the rabbit and rainbow trout *Salmo gairdneri;* the activities are correlated to the cholesterol content of VLDL [91]. The activity of lipoprotein lipase also varies by up to two orders of magnitude in the fat tissue, lung and muscle of various mammals; there is no apparent correlation with the nature of the lipid metabolism in each species [53]. Apolipoproteins are also exchanged between lipoprotein particles. This is of particular importance as the apolipoproteins have regulatory effects on enzyme activities: apoC-II stimulates lipoprotein lipase; this effect is inhibited by apoC-III; apoA-I, and also apparently apoC-I, stimulate LCAT [151]. All in all, the lipoproteins circulating in the blood plasma may be considered as highly dynamic structures. The few comparative studies clearly show that the mode of action of the lipoproteins is very similar amongst all vertebrates [213, 231].

Although no significant sequence similarity is found between the different **apolipoprotein types**, with the exception of apoD, they nevertheless appear to belong to the same protein superfamily. If, in assessing similarity, one takes into account similar physicochemical properties of the amino acids, periodic structures of 22 amino acids can be detected in all members of the family. These apparently form α-helices which on one side carry polar side-chains and on the other side non-polar side-chains (amphiphilic α-helices); they can therefore bind both the polar groups of phospholipids and the hydrocarbon chains of the fatty acids. ApoB-100 also contains numerous

22-amino-acid repeats which build amphiphilic α-helices. ApoD differs from the other small apolipoproteins in many significant properties. The 169-amino-acid apoD chain has an α-helix portion of less than 5 % and no periodicity. The highest concentration of apoD mRNA is found in the medulla of the adrenal body, and it is high in the kidney, pancreas and small intestine, but low in the liver. The apoD gene has a very different exon-intron structure compared with that of the genes of the other apolipoproteins [147]. Apparently, apoD does not belong to the protein superfamily of the apolipoproteins but, together with the retinol-binding protein, to that of the α_{2u}-globulins (Table 5.2). The **genes of human apolipoproteins** are dispersed on different chromosomes: apoA-II, apoB-100/48 and apoD are found as single copies on chromosomes 1, 2 and 3, respectively; apoA-I, apoC-III and apoA-IV form a gene cluster on chromosome 11, in the same order as is found in the chicken; and apoE, apoC-I and apoC-II form a cluster on chromosome 9. With the exception of apoD, the genes of all the small apolipoproteins have a similar structure with one intron each in the 5′-NT (non-translated) region, the signal region and the region coding for the mature protein; the first intron is missing in the gene for apoA-IV. The apoD gene contains no intron in the signal region but three introns in the remaining coding region. The gene of apoB-100, with a total length of 43 kb, is unexpectedly short when one considers that it includes 4564 codons in its 29 exons, including the 27 signal-coding triplets and the stop-codons; in fact, exons 26 and 29, with 7572 and 1906 bp, respectively, are exceptionally long [147].

The **protein or gene sequences** are known for all human apolipoproteins and for many of those from the chicken, rat, mouse, guinea-pig, rabbit and dog [56, 147, 161 280]. Sequence comparisons indicate relatively high rates of evolution. For example, human and rat apoA-IV differ at 38 % of positions, and rabbit apoA-I differs by 20 % from that of the dog, by 32 % from human and by 40 % from rat [123, 279]. The rate of evolution of the apolipoproteins A-II, C-II and C-III is approximately three times higher than that of β-globin (see Table 4.12), and that of A-I and A-IV is about twice as high; the apoE rate is also higher than that of β-globin, but that of apoB is markedly lower. Evolution apparently began with an apolipoprotein similar to apoC-I and resulted consecutively in the forms C-II, C-III, A-II, E, A-IV and A-I [147]. Many of the apolipoproteins found in lower vertebrates correspond to those of man in their electrophoretic mobility but should not necessarily be assumed to be identical. Up to now, only two apolipoproteins of lower vertebrates have been sequenced (via cDNA), namely those from the HDL of the sea lamprey *Petromyzon marinus;* they show no significant sequence similarity to any apolipoprotein class of mammals. Including the signal peptide, pro-peptide and mature protein, LAL1 consists of 21 + 8 + 76 amino acids, and LAL2 of 23 + 168 amino acids [202]. A close relationship between the lipoproteins of the lower and higher vertebrates is, however, indicated by immunological cross-reactivity between the VLDL and LDL of trout, chicken and guinea-pig.

5.6.2 Plasma Lipoproteins of Insects and Other Invertebrates

Lipid transport in insects is fundamentally different from that in vertebrates. In the latter, the triacylglycerols form the non-polar nucleus of lipoprotein particles and their fatty acids can be released only by radical reshaping of the particle. In insects, fatty acids are transported as strongly polar diacylglycerols (DAGs) arranged close to the surface of the lipoprotein particles, from whence they can be easily removed. Thus, in contrast to the vertebrates, the lipid transport particles of insects are reusable transport proteins. In the migratory locust *Locusta migratoria*, the half-lives of apolipoproteins and DAGs is 5–6 days and 2–3 h, respectively [101]. The lipid-transporting proteins of insects were originally termed "diacylglycerol-carrying proteins"; after it became known that they were also involved in lipid resorption in the intestine and in the transport of sterols, hydrocarbons and carotinoids, Chino in 1981 suggested the name **lipophorins**.

The lipophorin systems of representatives from various different insect orders have been examined in detail [72, 89, 106, 168, 229, 247, 261]. They appear under the electron microscope as globular particles with a diameter of 13–16 nm, corresponding to molecular masses of 500–700 kDa [221]. Their **lipid content** is normally 40–50 % but varies with the metabolic state. According to the increase in density with decreasing lipid content, the lipophorins are placed into the classes LDLp, HDLp and VHDLp [20]. Diacylglycerols and phospholopids are the predominant lipids; triacylglycerols are only found in trace amounts. Nuclear magnetic resonance (NMR)

spectroscopic investigations have shown that the phospholipids in the lipophorin particles of *Locusta* are located on the surface [124]. Sterols are always present but never sterol esters. The hydrocarbon content varies markedly between species, reaching 28 % in *Periplaneta* but only 1.4 % in *Philosamia*. The hydrocarbon composition corresponds approximately to that of the cuticular lipids [125]. Dependent on the type of nutrition, the lipophorins may contain more or less carotinoids, and are accordingly more or less intensively yellow. Juvenile hormones are also bound to the lipophorins [121].

Comparative investigations of a total of seven insect orders show that each lipophorin particle contains at least two **apolipoprotein molecules:** an apoLp-I of 230–250 kDa and an apoLp-II of 70–85 kDa [20, 121, 215]. Antisera against the apoLp-II of *Manduca* show cross-reactivity with the apoLp-II of all other species, but apoLp-I, on the other hand, shows no immunological similarity with other apoLp-I [215]. Differences in sensitivity to protease attack indicate that apoLp-I is more exposed than apoLp-II at the particle surface [62]. An additional, small apoLp-III of 17–20 kDa is found in the lipid-rich lipophorins of the Lepidoptera, Orthoptera and Hemiptera [220]. The apoLp-III gene of *Manduca sexta* consists of four exons which code for a signal sequence of 15 amino acids, a pro-sequence of 5 amino acids, which is cleaved off during maturation, and a mature protein of 165 amino acids. The amino acid sequence and the spatial structure of apoLp-III has also been determined for the migratory locust *Locusta migratoria*. Although there is only 29 % sequence similarity between these two proteins, the *Manduca* apoLp-III combines with the *Locusta* lipophorins to give a complex which is active on isolated fat bodies [49, 107, 121]. All lipophorins contain mannose and glucosamine [121, 215] but the N-bound carbohydrate chains have been examined in detail only in *Locusta*. In this case, apoLp-I carries six different oligosaccharides with the general formulae of $Man_{5-9}GlcNAc_2$ and $Glc_1Man_9GlcNAc_2$; the smaller apoLp-II has only three oligosaccharides [181]. The associated apoLp-III is a glycoprotein in the Orthoptera but carbohydrate-free in the Lepidoptera and Hemiptera [98].

The **cyclic alterations in the lipophorins** during transport of lipids from the fat bodies to the flight muscles (lipid shuttle) have been particularly well investigated in the butterflies *Manduca sexta* and *Acherontia atropos* [104, 121, 218, 247]. The apolipoproteins are formed in the fat bodies. Isolated fat bodies secrete a lipoprotein, for example in *Manduca*, that differs significantly from that found in the haemolymph: the density is 1.24–1.28 compared with at the highest 1.15 g/cm^3, and the ratio phospholipid:diacylglycerol is 8.3 instead of at the highest 0.9. Only in animals on a fat-free diet can such a protein be found in the haemolymph. Normally, the lipoprotein formed in the fat body, consisting mainly of apolipoproteins and phospholipids, is immediately loaded with diacylglycerols [203]. The haemolymph of the resting animal contains 9- to 16-nm-large HDLp particles that are poor in diacylglycerols and are known as A^{yellow} due to their strongly visible carotinoid coloration. Owing to the effect of the adipokinetic hormone (AKH), secreted by the corpora cardiaca at the beginning of flight, diacylglycerol is increasingly released from the fat bodies into the haemolymph and loaded onto the lipophorin particles. Their density is consequently reduced, and the continuously expanding particles are stabilized by the incorporation of apoLp-III (protein C_2), which is available in large amounts in the haemolymph. In actively flying animals, one finds mainly lipid-rich LDLp particles of 20–50 nm which contain as many as 16 apoLp-III and are known as lipophorin A^+ [48].

Following contact of the LDLp (lipophorin A^+) with the flight muscle, the diacylglycerols are hydrolysed by a lipoprotein lipase and apoLp-III is released; the low-lipid HDLp (lipophorin A^{yellow}) that remains can then be reloaded with diacylglycerols. Flight muscle lipoprotein lipase has been isolated from the migratory locust and found to be very different from the mammalian enzyme (p. 662). The characteristic changes in the lipophorin system that occur in the course of development from the larva to the pupa and eventually to the adult animal have been examined in detail in *Manduca sexta* and *Bombyx mori* [169]. The haemolymph of *M. sexta* contains specific lipid transfer particles (LTP) which catalyse transfer of lipids between lipoproteins of different density. The LTP is a VHDL particle of $1.4 \cdot 10^6$ Da, spherical with a long tail, and made up from the glycoproteins apoLTP-I (320 kDa), apoLTP-II (85 kDa) and apoLTP-III (55 kDa) with 14 % lipid. The LTP catalyses lipid transfer from a donor LDLp to an acceptor LDLp, resulting in an HDLp and a VLDLp; the LTP, however, can also use other lipoproteins as substrate, including even some from vertebrates [219, 221]. The fat-body cells of *Manduca* carry on the surface a receptor of approximately 120 kDa which binds lipoproteins more strongly the more dia-

cylglycerol they contain. This dependence of receptor affinity on lipid content facilitates lipid transport from the mid-gut to the fat bodies in the feeding larvae [259].

In addition to the lipophorins, further **species- and stage-specific lipoproteins** are found in insect haemolymph; at least seven different types can be distinguished electrophoretically [39, 121]. Vitellogenins play a dominant role in mature females. A VHDL found in the larva of the honey bee contains only 10% lipid and 2.6% carbohydrate; this is a homodimer with 160-kDa subunits and has no similarity to any other known haemolymph protein [224, 233]. In the last larval stage of the butterfly *Heliothis zea*, the haemolymph takes on a green colour; this is caused by the appearance of a homotetrameric protein of 560 kDa which contains 8.4% lipid and biliverdin as the colour component. The protein, for which there are no known parallels, is not found in young larvae, pupae or adults [97, 224].

Very little is known about the **lipoproteins in the haemolymph of other arthropods**. All the examined chilopods, diplopods and arachnids (scorpions, araneae, solifugae) have two lipophorin-like lipoproteins in the haemolymph and these are made up of two types of subunit with molecular masses of 220–250 and 80–90 kDa. A completely different type of haemolymph lipoprotein with polypeptides of about 100 kDa is found in the crustaceans. Surprisingly, the lipoprotein of the velvet mite *Dinothrombium pandorae* is more similar to that of the crustaceans than to that of other arachnids [99, 141]. Lipoproteins have been detected occasionally in the body fluids of molluscs and other invertebrates by specific staining reactions on electrophoregrams but have not been investigated further.

5.7 Vitellogenins and Yolk Proteins

The quantity of yolk in the eggs of most non-mammalian species and many invertebrates is so great that it cannot all be synthesized in the oocyte alone. In these cases, lipoproteins are synthesized as precursors of yolk proteins outside of the ovary, transported in the blood to the oocyte and there incorporated into the yolk. These lipoproteins are known as vitellogenins (VGs); their formation is mostly under hormonal control. The proteins are significantly altered between their synthesis and their incorporation into the yolk structure of the oocyte; the nature of this post-translational modification varies greatly between different groups of animals. Vertebrate VGs remain almost unchanged from their synthesis in the liver until they are taken up into the oocyte; they are then converted by proteolytic cleavage, phosphorylation and other processes into two fundamentally different types of yolk protein (**lipovitellogenins and phosvitins**).

In many insects, the primary translation products of the VG genes are already greatly modified by partial proteolysis, glycosylation, phosphorylation and other **post-translational processes** in the fat body before they enter the haemolymph; only minor changes occur after uptake into the oocytes. The comparative biochemistry of the VGs is a particularly attractive subject because their evolution is determined by so many structural/functional requirements: secretion out of VG-producing cells, transport in the blood plasma or haemolymph with protection from hydrolysis, specific recognition and uptake into oocytes, chemical changes in the oocytes, and incorporation into the yolk structure, e.g. the crystalline yolk platelets of the lower vertebrates. Their investigation, however, is hindered by the fact that they are usually large and complicated glycolipophosphoprotein complexes. The VGs are apparently quite old proteins; at least in the case of the chicken, the clawed frog *Xenopus laevis*, the sea urchin *Strongylocentrotus purpuratus* and the nematode *Caenorhabditis elegans* the similarity in VG gene structure and sequence indicates a common origin [185, 234]. There is significant homology between partial VG sequences of the chicken, frog or *Caenorhabditis* and the human apoB-100, as well as between *Drosophila* VG and human lipoprotein lipase. This suggests that apolipoproteins and the lipase evolved from a vitellogenin-like precursor [13]. Use of the yolk proteins during embryogenesis of vertebrates and arthropods involves lysosomal enzymes that are under the control of specific protease inhibitors [70, 163].

5.7.1 Vitellogenins and Yolk Proteins of Vertebrates

The process of **yolk formation** has been very conservative throughout the evolution of the vertebrates: in all oviparous amphibians, reptiles and birds dimeric VGs of 400–600 kDa are synthesized under oestrogen control in the liver, secreted into the blood and taken up by the oocytes, where they undergo proteolytic cleavage and are

used in the production of two types of yolk protein, the vitellins (VTs) and phosvitins (PVs). In addition to the PVs, there are often other smaller phosphoproteins, the phosvettes (PVT); PVs and PVTs are probably alternative cleavage products of the same VG region. The monotreme mammals, which produce large, abundantly yolked eggs, have unfortunately not yet been investigated. However, the formation of yolk proteins has been particularly well-described for the chicken and the clawed frog *Xenopus laevis*. General questions relating to cell biology and biochemistry, e.g. about the glycosylation and phosphorylation of proteins, have been studied using these model systems. A comparison of the gene sequences of VG-II from chicken and VG-A2 of *Xenopus* illustrates the homology of the vertebrate VGs. The two genes are very alike in structure with 35 often quite similar exons; in other regions the bird and frog genes are very different, in particular in the PV region, where numerous nucleotide exchanges and segment mutations are recognizable [186].

The **VGs of the domestic chicken** make up a family of related proteins encoded by different genes. They are homodimers with subunits between 170 and 190 kDa whose molecular mass is increased by about 10 kDa through glycosylation and phosphorylation. The main components, VG-I and VG-II, are very similar in their amino acid composition and each contains 116 phosphate residues. The gene of the minor component VG-III shows major differences to the VG-II gene in the region coding for the phosphoproteins PV and PVT: the agreement in amino acid sequence between VG-II and VG-III in this region is only 23 % compared with 40 % for the whole protein; exon 23 is reduced in length by more than half and the number of serine codons, and therefore phosphate-binding sites, is drastically reduced. As a result, VG-III carries only 44 phosphate residues [33]. Three VGs are also distinguishable in the quail *Coturnix coturnix*. The VG receptor on the surface of chicken oocytes has been solubilized and characterized in detail [241].

Centrifugation of diluted chicken-egg yolk leads to the sedimentation of granulae which make up approximately 20 % of the dry weight. They correspond to the yolk platelets of the lower vertebrates but are not crystalline. The major components of these granulae are two pairs of VTs and PVs derived from the two major VGs. The VTs are tetramers of two each of the subunits VT-I (125 kDa) and VT-II (30 kDa); the PVs have molecular masses of 34 kDa (with 104 phosphate residues) and 28 kDa (with 85 phosphate residues). In the supernatant there is a lipoprotein fraction with a density of 0.95 g/cm^3; this constitutes 70 % of the yolk dry weight and shows immunological similarity to the plasma VLDL. The lipid requirement for yolk formation in birds can apparently not be satisfied by VGs alone; a considerable part of the yolk lipid originates in the VLDL of the blood plasma. However, only a small apoprotein of the plasma VLDL, and not apoB, is found unchanged in the yolk as apoVT-I. It is likely that fragments of apoB are present in the yolk, in particular in apoVT-IV and apoVT-VI, which together have approximately the same size as apoB. There are also proteins present in the yolk which do not come from the plasma, e.g. apoVT-III and apoVT-V. The apoVT-I proteins of many birds have been sequenced and they show a relatively high rate of evolution, differing between the chicken and the turkey, for example, in 12 of the 82 amino acids [118]. Apart from the VLDL fraction, the yolk supernatant also contains the so-called livetins, a mixture of plasma proteins (serum albumin, transferrins, γ-globulins, etc.) which makes up the remaining 10 % of the yolk. Thus, uptake into the oocyte is apparently not specifically restricted to VG and VLDL.

Little is known about the VGs and the yolk proteins of the **reptiles**. The VG induced by 17β-estradiol in the snake *Thamnophis sirtalis* does not have the homodimer structure of other vertebrate VGs but is constructed from a VG2 chain of 124 kDa bound by a disulphide bridge to a VG1 chain of 149 kDa. In the turtle *Chrysemys picta*, on the other hand, there is a normal homodimeric VG with subunits of 210–220 kDa [18]. The predominant yolk protein of the lizard *Anolis pulchellus*, like the VT-I of chicken and *Xenopus*, has a molecular mass of 110–120 kDa [172].

The liver of sexually mature females of the **clawed frog** *Xenopus* continuously produces VGs, whereas in the male VG is induced by oestrogens. The VGs have molecular masses of 460 kDa and contain 12 % lipid, 1.3 % protein-bound phosphorous, 1 % carbohydrate and two identical polypeptides of about 200 kDa. *X. laevis* possesses four active VG genes coding for two VG pairs, A1/A2 and B1/B2. The amino acid sequences vary at about 20 % of positions between the pairs and at about 5 % within each pair. The genes for A1-A2-B2 are clustered, but B1 is located separately. Taking into account that *X. laevis* is a tetraploid species, the evolution of the VG genes

possibly occurred as follows: the A/B duplication happened about 150 million years ago, and was followed by the duplication A1/A2; the resulting A1-A2-B species took over a further B gene by allopolyploidization with another species [225]. The tetraploid *X. borealis* also has four VG genes of which A1–B1 are clustered but A2 and B2 are not. In contrast, in the diploid *X. tropicalis*, only one A and one B gene are found together with a weakly expressed A* gene [12].

In accordance with the multiplicity of genes, three different VGs are detectable in the blood of oestrogen-induced *X. laevis* females. The yolk proteins are present as crystalline structures in the so-called yolk platelets. These mainly consist of the two protein types vitellin (VT, 200 kDa) and phosvitin (PV, 35 kDa) as well as lower amounts of the phosvettes PVT1 (19 kDa) and PVT2 (13 kDa). Vitellin contains 20 % lipid and 0.5 % protein-bound phosphate and is made up of two subunits, VT1 (116 kDa) and VT2 (32 kDa). Both VT1 and VT2 separate electrophoretically into three differently sized fractions, which correspond in fact to the three VGs. PV is unique in that it contains more than 50 % serine residues, the majority of which are phosphorylated, and therefore 9.5 % protein-bound phosphorus. Dephosphorylated PV can also be separated into two fractions of different size. In the 1807-amino-acid-long chain of VG-A2, the VT1 sequence lies at the N-terminal end, the VT2 sequence lies close to the C-terminal end and the PV sequence is in the middle region; it is not yet clear what happens to the C-terminal 20 kDa encoded by exons 30–35 [264]. The VGs pass into the oocytes by a specific endocytosis process which shows a 30- to 50-fold preference for VG over other proteins. The VG receptor has a molecular mass of 115 kDa and is immunologically similar to the chicken VG receptor. Each receptor can bind the VG of the other species [242].

Vitellogenesis of other amphibian species has also been examined in some detail. A VG of 427 kDa is present in the tetraploid frog *Odontophrynus americanus;* this VG consists of two phosphoglycopeptides VGT1 (208 kDa) and VGT2 (204 kDa). The yolk contains two vitellins and three phosphoproteins: VT1 is made up of two glycopeptides of 105 and 93 kDa, and VT2 consists of three different glycopeptides of 32, 30 and 28 kDa. The phosphoproteins are phosphovitin (37 kDa) and the two phosvettes PVT1 (28 kDa) and PVT2 (26 kDa). The VGs from *O. americanus* are very similar to those of *X. laevis*, but result in different yolk proteins, probably because of differences in proteolytic cleavage caused by deviations in the amino acid sequence [273]. Very little is known about the yolk proteins of the urodelans and their ancestors. Four VGs, which can be electrophoretically defined in the blood plasma of *Pleurodeles waltii*, give at least two different polypeptides of 210 and 180 kDa on denaturation [37].

The quite variable results so far obtained in investigations of **yolk formation in fish** do not allow firm conclusions to be drawn. The uptake of female-specific plasma proteins into the oocytes of the agnathan *Eptatretus* has been demonstrated immunologically [281]. However, the possibility of hormone-regulated VG formation was not investigated in this case; corresponding experiments with cartilaginous fish were entirely negative and only with teleosts were the results always positive. The molecular sizes of the yolk proteins and the plasma precursors and apoproteins in the teleosts show great variety. VGs of 550–600 kDa were found in the rainbow trout *Salmo gairdneri*, the flounder *Pleuronectes flesus* and the winter flounder *Pseudopleuronectes americanus;* there appear to be several VGs of 380 kDa in the goldfish *Carassius auratus*. The VG of the Japanese eel *Anguilla japonica* is reported to be a 350-kDa homotetramer with subunits of 85 kDa. The VG of *Ameiurus nebulosus* has a mass of 145 kDa and the corresponding mRNA is also shorter than that of the chicken and *Xenopus* [182, 243].

The yolk proteins are just as heterogeneous; no less than nine VG-derived polypeptides have been found in the egg yolk of the antarctic teleost *Caenocephalus aceratus* and these have sizes between 13 and 172 kDa and differing phosphate contents [232]. Denaturation of the VTs of the rainbow trout and the goldfish yield subunits of 90–140 kDa and 15–25 kDa, whilst the PV subunits have molecular masses of 7.6–15 kDa [208]. Partial proteolysis of teleost VGs seems to produce at first larger and then smaller phosphoproteins that are analogous to the phosvettes of *Xenopus*. The cleavage products apparently suffer partial dephosporylation; in any case, the content of protein-bound phosphate in PVs varies between 0.0012 % in the halibut *Hippoglossus hippoglossus* and 0.73 % in the dogfish *Scyliorhinus canicula* [251, 263]. The yolk platelets of the Agnatha, of the archaic fish genera *Amia, Lepisosteus, Polypterus* and *Latimeria*, and of the freshwater teleosts are crystalline like those of the amphibians; the yolk crystals of the Agnatha are monoclinal whereas the others are of the orthorhombic type [139].

5.7.2 Vitellogenins and Yolk Proteins of Insects

The VG of the haemolymph and the major components of the yolk proteins of all insects so far examined are essentially identical in terms of molecular size and both electrophoretic and immunological properties; in contrast to the vertebrates, the VGs of insects are not significantly altered after their uptake into the oocytes. Many authors, therefore, also use the term **vitellogenins** to include the yolk proteins. However, there are often differences in solubility and lipid content between the VGs of the haemolymph and the corresponding yolk proteins and, in several cases, proteolytic processing or aggregation to higher polymers has also been described [27, 121, 204]. It is therefore justifiable to distinguish between the **vitellins** (VTs) of the yolk protein and the VGs of the haemolymph [69, 121].

The VGs are lipoglycoproteins with about 7–16 % lipid and 1–14 % carbohydrate. In many insects (e.g. *Hyalophora, Apis*), VGs are the predominant haemolymph proteins found during vitellogenesis; in others (e.g. *Periplaneta, Leucophaea, Drosophila*), the proportion hardly exceeds 1 %. This is not the result of quantitative differences in synthesis; in tracer experiments with labelled leucine, more than 80 % was incorporated into VGs in *Leucophaea* as well as in *Apis* and *Locusta*. It appears more likely that the oocyte transport system is much more efficient in *Drosophila* and the cockroaches [69]. In contrast to the lipophorins, the VGs and the VTs contain more phospholipid than diacylglycerol; they also contain sterols but never hydrocarbons. The carbohydrate fractions of the VGs are mainly the classical asparagine-bound chains $Man_9GlcNAc_2$ which, in this case, can be later modified. Following inhibition of N-glycosylation by tunicamycin, the fat bodies of the cockroach *Blattella germanicus* accumulate the primary translation product of over 200 kDa. The release of VGs into haemolymph is therefore only possible after glycosylation, as has been recorded for other haemolymph proteins of the wax moth *Galleria melonella*. Conversely, *Xenopus* liver is able to secrete non-glycosylated VG [121, 135, 196]. The VGs and VTs of many insects contain phosphorylated serine residues or sulphated tyrosine residues [10, 45, 116, 249].

The **synthesis of VG** takes place predominantly, or entirely, in the fat bodies. Before release into the haemolymph, the precursor formed in the fat bodies is converted into VG by glycosylation, phosphorylation, addition of lipid and, in many cases, partial proteolysis. In most insects, VG synthesis in the fat bodies is regulated by the juvenile hormone; synthesis declines after removal of the corpora allata and is restored by injection of juvenile hormone or analogues thereof. VG synthesis can also be induced in male animals by hormone injection. In several dipterans, in addition to, or instead of, juvenile hormone, 20-hydroxyecdysone is responsible for the control of VG synthesis [29, 35, 69, 84, 121, 135, 278]. VG is also normally found in the haemolymph of males of several insect species *(Hyalophora, Tenebrio, Oncopeltus, Rhodnius)*. VG is the predominant haemolymph protein in the honey bee *Apis mellifera*, not only in the egg-laying queens but also in freshly hatched queens and in workers; it is missing only in the drones. In the bumble bee *Bombus terrestris*, the workers produce VG only when they lack a queen [69]. Uptake into oocytes proceeds via receptor-mediated endocytosis in coated pits; it was in the oocytes of *Aedes aegypti* that this widely distributed mechanism was first found in 1964. Uptake is highly specific; isolated ovaries of *Hyalophora cecropia* take up only the own-species type of VG from a mixture of *Hyalophora* and *Blattella* VGs. Small amounts of haemolymph proteins other than VGs are also found in the oocytes, but only VG is enriched by 30-fold or more. The oocyte receptor of *Locusta migratoria* has been solubilized and the binding constant for VG was found to be $4.2 \cdot 10^{-8}$ mol/l [211].

Insects may be subdivided into three classes according to the molecular architecture of the yolk proteins and their precursors, and to the in vitro translation products of the corresponding mRNAs [29, 95, 121, 135]. In **group 1**, the VGs consist of two polypeptide classes, H (100–180 kDa) and L (43–86 kDa). As a rule, they have a molecular mass of 440–560 kDa and contain two each of these polypeptides according to the formula H_2L_2; however, the VG of *Manduca sexta* is only 260 kDa and has the formula HL. During synthesis in the fat body, a precursor of 200–260 kDa is formed and this is cleaved into two unequal fragments prior to release into the haemolymph. This partial proteolysis is apparently pre-programmed in the sequence of the precursor; this also occurs if the fat-body mRNA of *Locusta* is translated in *Xenopus* oocytes [135]. Group 1 includes the mayflies, cockroaches [31, 45, 244, 274], locusts, earworms, butterflies [19, 127], beetles [199], bug *Rhodnius prolixus* [159] and bristletail *Thermobia domestica* [214].

Unexpectedly, the VG of the cockroach *Nauphoeta cinerea* is a homodimer of 244-kDa polypeptides [115].

The VGs of **group 2** are monomers or dimers of 170- to 190-kDa subunits, the precursors of which are already of approximately equal size in the fat body. To this group belong the Hymenoptera and the primitive dipterans, the Nematocera (mosquitoes). One can imagine that in the evolution of this group the VG genes lost the exons for the L-polypeptide [269]. Deviating somewhat from this picture is the 65-kDa subunit, arising from the same precursor, which was recently found in addition to the 200-kDa subunit in the VG of the mosquito *Aedes aegypti* [61].

Group 3 includes only the higher dipterans, the Brachycera (flies). The native VGs are of about 200 kDa and are made up of polypeptides of about 50 kDa corresponding to the primary translation products. In this case, evolution of the VG genes resulted in the loss of the exons for the H subunit. The VGs of about ten species of flies have been investigated in detail [21, 22, 41, 96, 135, 144, 209]. One peculiarity of this group is that in addition to the continuous synthesis of VG in the fat body there is also stage-specific synthesis in the ovary, probably in the follicle epithelial cells [22, 282]. This is possibly related in some way to the already-mentioned characteristic hormonal control of VG synthesis. According to experiments with *Drosophila* cell cultures, juvenile hormone affects only synthesis in the ovary, and 20-hydroxyecdysone only synthesis in the fat body [166]. In the fly *Stomoxys calcitrans*, no female-specific proteins are ever found in the haemolymph or yolk proteins in the fat body; VG synthesis is restricted to the ovary [41, 109]. Minor VG synthesis has also been observed in isolated ovaries of the potato beetle *Leptinotsara decemlineata* [199].

Most insects contain **multiple VGs and VTs** that are recognizable by their varying molecular size, for example in the cockroach *Leucophaea maderae* [31], or by differences in immunological or electrophoretic properties. This is probably a case of products of non-allelic loci. *Locusta migratoria* has two VG genes, most dipterans have three, and *Aedes aegypti* has four [29, 35, 84, 121]. There are three VG polypeptides in *Drosophila* which agree in 43 % of their 420–442 amino acids, and are encoded by three homologous genes on the X chromosome. Of these, YP-1 and YP-2 contain only one intron and are separated by 1.2 kb; YP-3 lies about 1000 kb away and has two introns. The three genes are undoubtedly the result of a double duplication [83]. The genes are expressed in a sex-, tissue- and stage-specific manner, and the recently isolated yolk-specific factor I (YPF1), which binds to a specific site on the YP-1 gene, is probably involved in this regulation [167].

Proteins originating in the VGs account for 60–90 % of the total soluble egg protein; other plasma proteins are found in minor amounts in the egg [109]. Amongst these subsidiary components of the yolk proteins are, for example, the microvitellin of *Manduca sexta* and the paravitellin of *Hyalophora cecropia*. Microvitellin is a protein of 26 kDa without carbohydrate, lipid or phosphate and which is produced in the fat body 17 days before hatching of the adult animal; it is transported to the ovary by the haemolymph and taken up unchanged into the oocytes. The sequence of 232 amino acids derived from the cDNA has no similarity to any other known protein [265]. In the eggs of *Bombyx mori*, 40 % of the yolk protein consists of vitellin, 35 % of the group of 30-kDa proteins [174] and 25 % of the egg-specific protein (ESP). Mature ESP is a glycophosphoprotein of 225 kDa consisting of two subunits of 72 kDa and one of 64 kDa. The two types of subunit arise by different post-translational processing from a primary translation product of 558 amino acids [116, 222]. Only after the onset of embryogenesis is a trypsin-like proteinase of 30.5 kDa produced which cleaves certain bonds in the ESP in a highly specific way [117]. It has been shown in *Manduca* that lipophorin is also taken up into the egg from the haemolymph [126, 134]. In many cockroaches, locusts, beetles and butterflies, the females reabsorb a part of the spermatophore material before the remaining spermatophore is ejected. The males of *Melanoplus sanguinipes* pass on an average of seven spermatophores during copulation, although one alone contains more than enough sperm; it could be shown immunologically that male proteins find their way unchanged into the oocytes [78].

5.7.3 Vitellogenins and Yolk Proteins of Crustaceans and Other Invertebrates

The egg yolk of crustaceans contains one or more lipoproteins (lipovitellins) of 310–600 kDa. The lipovitellin of the small brine shrimp *Artemia salina* is a lipoglycoprotein of 600 kDa with polypeptides of 190 and 68 kDa, 3.3 % carbohydrate and 8.6 % lipid; it is coloured by the presence of

the carotinoid canthaxanthin [38]. Other crustacean lipovitellins have a higher lipid content and 2–11 different apolipoproteins ranging in molecular mass between 45 and 190 kDa [38, 205, 257]. Tracer experiments with isolated organs of the prawn *Penaeus japonicus* showed that immunologically identifiable VGs were synthesized by the ovary but not by the hepato-pancreas. VG synthesis also occurs in the ovary in other decapods like *Palaemon* and *Pachygrapusus*, but occurs in the fat tissues of amphipods and isopods [248, 257] and of the tick *Ornithodorus moubata* [44].

The vitellin of the **polychaete** *Nereis virens* is a lipoglycoprotein of about 420 kDa which is coloured green by bilin. It arises in free coelom cells (eliocytes) and is taken up into the oocytes (which are also found free in the coelom fluid) by receptor-mediated endocytosis; the egg-maturation process lasts for about 1 year [74]. It has also been shown by tracer experiments in the related species *Perinereis cultrifera* that VGs secreted by coelomocytes are taken up by egg cells. The VG has a molecular mass of 530 kDa and contains two 176-kDa polypeptides. The protein taken up by the egg cells is progressively reduced in size from 530 to 390 kDa; a series of five vitellins (VT-1 to VT-5) of decreasing molecular mass is found in young oocytes but older oocytes contain only VT-5, which has a molecular mass of 390 kDa and about 16 % lipid [9].

In the **nematode** *Caenorhabditis elegans* the yolk contains two different lipoprotein complexes, of which the A-complex contains three different polypeptides of 170, 115 and 88 kDa (yp170A, yp115 and yp88) and the B-complex has only one type (yp170B). These polypeptides arise from at least three different precursors produced in the gut cells of the hermaphroditic animal. Two of these (yp170A and yp170B) are taken up unchanged into the oocytes, whilst the third is first cleaved into two fragments (yp115 and yp88). *Caenorhabditis elegans* possesses six VG genes, the coding sequences of which agree 85 % with the five homologous genes of the closely related species *C. briggsae* [230, 283]. The eggs of the **sea urchin** *Hemicentrotus pulcherrimus* contain lipoprotein particles that are 29–48 nm in diameter with 55–72 % lipid (predominantly triacylglycerols) and 8–13 % carbohydrate. Denaturation of the particles always produces the same four polypeptides [59]. VG synthesis in the sea urchin *Stronglyocentrotus purpuratus* is quite unique in several respects. There is only one VG gene and this is expressed in the gut wall and gonads of both males and females; it has similarities to the VG genes of the vertebrates not only in its relatively large size of 19 kb but also in the extreme shortness of the signal sequence and in various sequence elements of the 5′-NT region [234].

5.8 Blood Clotting

In the animal kingdom there are three mechanisms for reducing blood losses from damaged blood vessels:

1. Contraction of the edges of the wound and the vessels.
2. Blockage of the wound by aggregates of blood cells.
3. Clotting of the blood fluid.

Here we will consider only the third process, which has been observed not only in the vertebrates but also in the blood of arthropods and certain molluscs (oysters) and in the coelom fluids of sipunculids, brachiopods, echinoids and holothurians. In the case of the invertebrates, biochemical data are available only for the arthropod groups Xiphosura, Crustacea and Insecta.

5.8.1 Blood Clotting in Vertebrates

The formation of a solid blood clot in all vertebrates involves the polymerization and precipitation of fibrin; this occurs by partial proteolysis of the fibrinogen in the blood plasma. The protease responsible is thrombin, which itself is formed by proteolysis from its precursor prothrombin. These two reactions are the last steps in a **cascade of proteolytic processes** which each produces an active clotting factor from a precursor (zymogen) (Fig. 5.1). The cascade has an amplifying effect and also increases the number of possible regulation points. The activation of the zymogens is accelerated when they are complexed with protein cofactors, Ca^{2+} ions and phospholipids of cell membranes. Most clotting factors are serine proteases which, like trypsin, attack arginyl bonds; only the factors Va and VIIIa have no enzyme activity, but they increase the effectiveness of proteases Xa and IXa. The zymogens of the clotting cascade are homologous in their C-terminal region (250 amino acids) with the catalytic domain of the pancreas proteases; their N-terminal sequences, however, are much longer

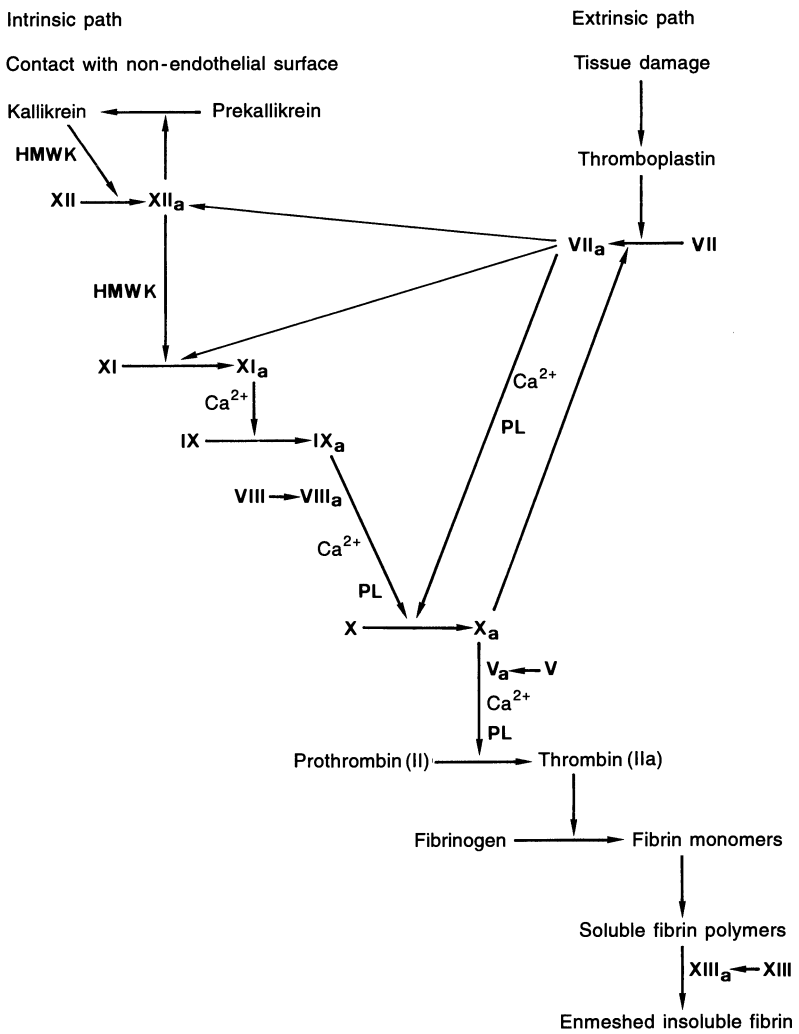


Fig. 5.1. The human clotting cascade [264]. For further explanation, see text. *HMKW*, high molecular weight kininogen; *PL*, phospholipid

(150–582 amino acids). Consequently, their molecular masses of 45–160 kDa are considerably larger than the approximately 25-kDa pancreas zymogens and enzymes (see Fig. 3.6, p. 91). During the proteolytic activation of all zymogens of the blood-clotting cascade, the activating peptide, or a large part thereof, remains bound to the active enzyme via a disulphide bridge. The only exception is during the **activation of prothrombin** by the factor Xa. The prothrombin chain is initially cleaved into two approximately equal parts by the hydrolysis of an Arg/Thr bond; the N-terminal half remains bound to the clotting complex whilst the C-terminal, catalytic half is released and can diffuse to its various sites of activity. Due to the cleavage of an Arg/Ile bond in the catalytic half, active thrombin consists of two chains, A and B, held together by a disulphide bridge. In human, but not in bovine, thrombin the N-terminal tridecapeptide of the A chain is removed by hydrolysis of an Arg/Thr bond. Active thrombin is responsible not only for the cleavage of fibrinogen but also for the activation of factors V, VIII and XIII and the clotting-inhibitor protein C [82, 284].

The **scheme** in Fig. 5.1 shows the known distribution of factors in the **intrinsic clotting pathway**, which also takes place in isolated blood, and the **extrinsic pathway**, which is induced by contact with the edges of the wound; the diagram is incomplete in so far as it does not show the interaction between the two pathways. In addition to plasma proteins, the clotting process also involves various factors released from the aggregated and degraded thrombocytes, nucleated blood cells which in mammals degenerate into non-nucleated blood platelets. The binding of the blood platelets to the damaged vessel walls is mediated by a large glycoprotein, the von Willebrand factor (vWF). The mature vWF of 2050 amino acids arises by

proteolysis, glycosylation and sulphation of a precursor of 2813 amino acids and is constructed from two to five repeats of each of four different types of domain. The vWF mosaic gene is undoubtedly the result of evolutionary processes such as exon shuffling and the duplication of fragments from very different genes. Parts of the sequence are in fact homologous to the vitellogenin of the nematode *Caenorhabditis elegans* [14, 153, 255]. In the intact organism, the clotting cascade is more often induced by the tissue factor TF (thromboplastin), which activates plasma factor VII, than by the platelet components. TF is a membrane-bound glycoprotein with 263 amino acids and shows no significant homology to serine proteases [238]. Factor VII is a serine protease with 406 amino acids and upon activation is cleaved into two disulphide-bridge-linked chains of 156 and 254 amino acids [195]. Contact with various negatively charged materials activates only the intrinsic pathway in isolated blood. This involves the factors XII and XI, prekallikrein and the high molecular weight kininogen (HMWK). Prekallikrein is found in human blood plasma as a complex with HMWK and is converted by factor XIIa to active kallikrein, which itself consists of two chains bound by a disulphide bridge. Kallikrein releases kinin from kininogen and activates plasminogen, factor IX and the surface-bound factor XII. A further action of kallikrein is the production of the pharmacologically active nonapeptide bradykinin from HMWK; the remaining product is a molecule, made up of two chains, which further promotes clotting [277]. In addition to HMWK, which has a molecular mass of 110 kDa, there are also smaller kininogens (LMWK) corresponding to the N-terminus of HMWK. Rats possess a K gene for HMWK and LMWK and also two homologous T genes, which differ markedly from the K gene due to the insertion of an Alu element and a frame-shift mutation in the HMWK region. Both HMWK and LMWK function as cysteine-proteinase inhibitors [132].

The structure and evolution of the individual **components of the human clotting cascade** will be only briefly referred to here as unfortunately there are no comparative biochemical data. Factor XI is a homodimer that, when complexed with HMWK and phospholipid, is cleaved by the action of factor XIIa into two light and two heavy chains; in the presence of Ca^{2+} it can then activate factor IX. Factor XI agrees in 58% of its sequence with human prekallikrein [80, 284]. Factor IX belongs, together with factors VII and X, prothrombin and the clotting-inhibitory proteins C and S, to the vitamin K-dependent proteins, which are all homologous [262]. In all these proteins, the first 10–12 glutamine residues of the chain are converted to γ-carboxylglutamic acid (Gla) residues by the action of a vitamin K (phyllochinon)-dependent carboxylase (see Fig. 3.1, p. 72). These Gla residues bind Ca^{2+} and thereby allow the interaction of the clotting factors with phospholipids. The primary translation products of the relevant genes always consist of a signal peptide, a propeptide which is perhaps involved in Gla formation, the Gla region, two domains homologous to the epidermal growth factor (EGF), a connecting region, and the catalytic domain that is homologous to the usual serine proteases; the vitamin K-dependent proteins are model examples of evolution by exon shuffling [262] (see Fig. 3.6, p. 91). Factor X, as found in the plasma, consists of a light chain (15 kDa) and a heavier chain (40 kDa) bound together by a disulphide bridge. The light chain includes 12 γ-carboxylglutamic acid residues, and the heavy chain contains the active centre. In man, the non-enzymatic factors V and VIII are very large proteins which were first sequenced via the cDNA. The 2224 and 2332 amino acids, respectively, of these two factors are subdivided into three A regions of 330 amino acids, a B region of 980 amino acids and two C regions of 220 amino acids in the order A1-A2-B-A3-C1-C2. The two factors show about 40% sequence similarity; their A regions share 35–40% similarity with the copper protein caeruloplasmin [120].

It may be assumed from the results of physiological investigations that the blood-clotting process progresses similarly in all vertebrates, although with some quantitative differences. However, comparative biochemical data are only available for fibrinogen. Human **fibrinogen** is a glycoprotein of 340 kDa with about 4% carbohydrate; it is symmetrically constructed from an Aα chain of 610 amino acids, a Bβ chain of 461 amino acids and a γ chain of 411 amino acids, according to the formula $(A\alpha, B\beta, \gamma)_2$. Thrombin cleaves with high specificity an arginine-glycine bond in the Aα and Bβ chains, thereby setting free the short-chained fibrinopeptides A and B and producing the soluble fibrin monomers $(\alpha, \beta, \gamma)_2$. The polymerization and precipitation of the monomers to the insoluble fibrin clots is a very complicated process involving all three chains [175, 284], in the course of which the factor XIIIa (glutaminylpeptide-γ-glutamyltransferase) forms isopeptide bonds between glutamine-γ-carbamoyl and lysine-ε-amino groups (Fig. 5.2).

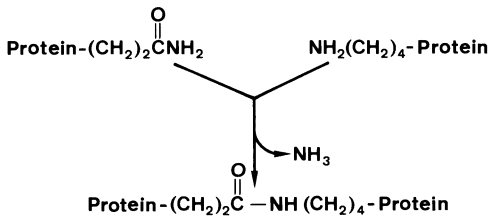


Fig. 5.2. Factor XIIIa (glutaminylpeptide-γ-glutamyltransferase) links fibrin chains by isopeptide bonds between glutamine-γ-carbamoyl groups and lysine-ε-amino groups [264]

Factor XIII/XIIIa belongs to a group of enzymes ("transglutaminases") which are widely distributed in the vertebrate body and are involved in many different biological processes [112].

The three human fibrinogen chains are clearly homologous; the genes are closely linked on chromosome 4 in a gene cluster with the structure $\vec{\gamma}$–$\vec{\alpha}$–$\overleftarrow{\beta}$, which would appear to have arisen from a common ancestral gene by two duplications and one inversion [122]. The Bβ chain agrees in about 30 % of positions with the γ chain, and both together have only 10 % similarity with the Aα chain, although this increases to 66 % if the central region alone is considered. Human fibrinogen is heterogeneous; there are γ variants with different C-terminal regions resulting from alternative splicing, and all three chain types show allelic polymorphisms and post-translational variants, e.g. with a varying sialic acid content [284].

Sequence comparisons between the fibrinogen chains of different vertebrates indicate greater species-specific differences in the Aα chains than in the Bβ and γ chains. Thus, the human Bβ and γ chains have about 80 % sequence similarity to the bovine and rat chains, and still about 50 % similarity with those of the sea lamprey *Petromyzon marinus*, one of the most primitive living vertebrates; on the other hand, large parts of the Aα chain of the lamprey are very different to that of the mammals [173, 266, 268, 284]. Nevertheless, mammalian thrombin can coagulate lamprey fibrinogen, although it only cleaves off fibrinopeptide B, one of the longest known fibrinopeptides with 36 amino acids. In addition, lamprey thrombin releases fibrinopeptide A, the shortest of all fibrinopeptides with only six amino acids. The sequences of fibrinopeptides have been determined for numerous mammalian species but for only one species each of the birds, reptiles and amphibians; at least 55 A sequences and 46 B sequences are known so far and have been used in the molecular analysis of familial relationships. As the Bβ and γ chains of the same vertebrate species agree by only about 33 %, whereas the corresponding chains of humans and lampreys agree by up to 50 %, the gene duplication must have occurred before the appearance of the vertebrates, and fibrinogen-like sequences should occur in invertebrates. In fact, coding sequences have been found by the polymerase chain reaction (PCR) in the DNA of the holothurian *Parastichopus parvimensis*, whose putative products are homologous to the carboxy-terminal two-thirds of both fibrinogen chains [276].

Fibrin clots are solubilized by the protease **plasmin**, which is formed from plasminogen. This polypeptide of 790 amino acids contains five of the so-called Kringel domains found in thrombin, and a catalytic region, which is homologous to other serine proteases. The bovine and porcine plasminogens agree by 78–83 % with that of man [157]. Activation involves cleavage of the bond at 560-Arg to give a plasmin molecule consisting of two chains linked by a disulphide bridge. Activation can occur (1) intrinsically by factor XIIa, kallikrein and HMWK, (2) by urokinase, or (3) by tissue plasminogen activator (see Fig. 3.6, p. 91). The urokinase-like chicken plasminogen activator, which was recently sequenced via the gene, agrees in only 43 % of its 434 amino acids with that of man, but shows the same domain structure [143]. The plasmin set free from blood clots is inactivated in the plasma by α_2-antiplasmin, which belongs to the serpine family of protease inhibitors (p. 98).

A **system of clotting inhibitor factors** prevents the clotting of blood in intact vessels. The most important regulator of blood clotting is, in fact, the C-protein which inactivates factors Va and VIIIa and stimulates fibrinolysis. The Ca-protein, resulting from the activation of the C-protein, is a disulphide-linked dimer of 420 amino acids made up of a light chain with γ-carboxylglutamic acid residues and a heavy chain with the catalytic centre of a serine protease. The formation of the Ca-protein from its C-protein precursor is catalysed by thrombin, especially when this is bound to thrombomodulin, which is an integral membrane protein of the endothelial cells in the vessel walls [102]. A necessary cofactor of the Ca protein is the S-protein which, like the Ca-protein, is a vitamin K-dependent protein and contains four domains similar to the EGF but is not a serine protease [150]. A series of proteins from the annexin family have inhibitory effects upon both the extrinsic and intrinsic clotting pathways by binding to the phospholipids involved. There are

at least six such proteins in humans which have been given the names, for example, lipocortin, placental anticoagulant protein (PAP) or vascular anticoagulant (VAC) [100]. Finally, reference must be made to the clotting-regulating components of the thrombin inhibitor AT–III, which has about the same size (of 60 kDa) in all vertebrates from the mammals to the amphibians; the human protein is 423 amino acids long. The clotting inhibitory effect of the polysaccharide heparin is based on its binding to AT–III, although the affinity of different heparin fractions varies by ten-fold [23, 284]. Various types of clotting inhibitor are found in the blood-sucking Hirudinae. The best known is hirudin from *Hirudo medicinalis;* this is a thrombin inhibitor of 65 amino acids (p. 101). Antistasin from *Haementeria officinalis* (with 119 amino acids) and a similar factor from *H. ghilianii* have factor Xa-inhibiting activity, and hementin from *H. ghilianii* is fibrinolytic. Decorsin (39 amino acids) from the North American leech *Macrobdella decora* inhibits aggregation of the thrombocytes [227].

5.8.2 Blood Clotting in Arthropods

The **clotting capacity of arthropod blood** varies considerably according to the species, the developmental stage and the physiological state: it ranges from the complete absence of plasma clotting to the formation of blood clots whose strength exceeds those of mammals. For a long time there was controversy about whether the proteins analogous to fibrinogen (coagulogens) originated in blood cells or were always to be found dissolved in the plasma; however, there was no doubt that the clotting enzyme was of cellular origin. It would now appear that the arthropods demonstrate all possibilities for the origin of coagulogen: in the xiphosurans they arise only from blood cells, in the crustaceans only from the haemolymph, whilst in the insects both cellular and extracellular proteins appear to be involved.

A haemocyte lysate from the **xiphosuran** *Limulus polyphemus* is converted into a gel upon contact with small amounts of bacterial endotoxins (lipopolysaccharide, LPS). The sensitivity of this test is so high that it is used in pharmacology for endotoxin detection [267]. The clotting system includes the coagulogen and three serine proteases. LPS activates factor C; the activated factor C converts factor B into its active form, which in turn promotes the conversion of pro-clotting enzyme to clotting enzyme [180, 256]. The coagulogens from the blood cells of the three species *Limulus polyphemus* (North and Central America), *Tachypleus tridentatus* and *Carcinoscorpius rotundicauda* (Southeast Asia) have been sequenced directly and that of *Limulus* has been sequenced via the cDNA. They are polypeptides of 20 kDa with 175 amino acids, including 16 cysteine residues, and in the case of *Limulus* and *Tachypleus* show 69 % similarity. The clotting enzyme cleaves the arginyl bonds at 18-Arg and 46-Arg; of the three chains that are thereby created, A (1–18) and B (47–175) form the coagulate, whilst C (19–46) is removed [43, 239]. Factor C is a dimeric glycoprotein made up of a heavy (80 kDa) chain and a light (43 kDa) chain. During activation by LPS, the light chain is split into an A-chain of 8.5 kDa and a B-chain of 34 kDa; the latter contains the catalytic centre. This arises by cleavage of the bond 72-Phe/73-Ile, which constitutes a difference between the xiphosuran system and the serine proteases of blood clotting, fibrinolysis and the complement cascade in mammals, where activation is by cleavage of an Arg-Ile/Val bond. There is apparently no other proteolytic enzyme involved in the activation reaction of the xiphosurans, but the reaction mechanism requires further investigation. The A-chain has some similarity in a part of its 72-amino-acid sequence to several mammalian complement components [180, 256]. The clotting enzyme is a serine protease of 150 kDa and is present in the blood cells as a zymogen. The haemocytes of *Limulus* and *Tachypleus* contain an inhibitor of LPS-dependent clotting. This anti-LPS factor is a polypeptide of 102 amino acids and, surprisingly, has significant homology (22 %) to rabbit α-lactalbumin [1].

According to Tait (1911), the **crustaceans** can be subdivided into three groups based upon their clotting type. In group A (e.g. *Cancer* and *Maja*) there is only cell aggregation; in group B (e.g. *Macropipus, Carcinus, Galathea, Homarus*) plasma clotting follows cell aggregation; and in group C (e.g. *Astacus, Panulirus*) the solidification of the plasma spreads out from "explosive" blood cells. However, coagulogens can be purified from the haemolymph of representatives of all eight genera mentioned, and clot with a haemocyte extract from *Macropipus puber.* Thus, the variation between the three groups concerns only differences in the concentration of the coagulogens [85]. The coagulogens of these eight genera appear to be structurally quite similar as they all react with the same clotting enzyme. The coagulogen of the spiny lobster *Panulirus interruptus* is

a homodimer with 200-kDa subunits, and as such has a completely different structure to the vertebrate fibrinogen with its three chains. The amino-terminal sequence is homologous to vertebrate and invertebrate vitellogenin. The clotting reaction is not a proteolytic process but involves the transglutaminase-catalysed formation of pseudopeptide bonding between glutamate and lysine residues (Fig. 5.2) [64].

Amongst the **insects** there is a wide spectrum of blood clotting phenotypes, but in every case specific types of blood cell (coagulocytes) are involved. In some species an aggregate of coagulocyte is formed and surrounded by a network of granular fibrils; in other species single coagulocytes appear to send out thread-like processes; and in some cases, there is no indication of any plasma clotting [26]. The proteins involved have been investigated in a few cockroach and locust species and in most detail in the cockroach *Leucophaea maderae* and the migratory locust *Locusta migratoria*. Two proteins are active here: a haemocyte coagulant (HC) secreted from the blood cells forms an insoluble clot by interacting with the plasma coagulant (PC) already present in the haemolymph. In reducing medium, HC alone produces a soluble gel which is not solidified by later addition of PC. The plasma coagulant of both species mentioned above is identical to the lipid-transporting protein lipophorin [17, 26].

5.9 Antifreeze Proteins

The blood plasma of teleosts in the polar oceans and in the coastal waters, which are cold in winter, of the northern temperate zone contains large quantities of special proteins that depress the freezing point of the plasma by several degrees. Because seawater freezes at about –1.9 °C and the blood plasma of most marine teleosts freezes at approximately –0.8 °C, depending on the solute concentration, large areas of cold seas would be inaccessible to teleosts without this remarkable adaptive mechanism. Surprisingly, there are several types of such antifreeze proteins with the same function but very different structures [58].

The first type to be discovered was the **"antifreeze glycoproteins" (AFGPs)**, which were isolated from two species of the antarctic Nototheniidae, *Pagothenia* (formerly *Trematomus borchgrevinki* and *Dissostichus mawsoni*, and from five Gadidae of the Northern Hemisphere, *Eleginus gracilis, Boreogadus siada, Gadus ogac, G. morrhua* and *Microgadus tomcod;* these were characterized in detail [58]. They consist of a series of (mostly) eight polypeptides varying in size between 2.4 and 34 kDa, of which the larger types (> 10 kDa) consist only of alanine, threonine, galactose and N-acetyl-galactosamine, and are made up of repeated glycosylated tripeptides (Fig. 5.3). In contrast, the smaller peptides also contain proline, and in *Eleginus* and *Microgadus* also arginine. Like the larger proteins, their sequences of 14–17 amino acids are organized as tripeptides, but with proline replacing alanine in one or two positions. The small polypeptides alone have almost no antifreeze activity but they amplify the effectiveness of the larger proteins by two- to eight-fold. Similarly composed glycopeptides have been detected in almost 30 other teleost species from the Ross Sea [32, 58, 110].

The second type of antifreeze protein has no carbohydrate. These **"antifreeze polypeptides" (AFPs)** may be subdivided into three basically different classes [58]. **Class I** includes alanine-rich polypeptides of 3–5 kDa with a secondary structure of amphiphilic α-helices. Their distribution appears to be restricted to several flounder species of the subfamily Pleuronectinae, for example *Pseudopleuronectes americanus, Limanda ferruginea*, and *Liopsetta putnami*, and to sculpins of the genus *Myoxocephalus*.

Antifreeze proteins of **class II** have so far been detected in only one species of the Cottidae, the Arctic Sea raven *Hemitripterus americanus*. These are larger proteins (14 kDa) with 8 % cysteine and they contain many reverse turns and five disulphide bridges. The proteins assigned to **class III** are of intermediate size (6–7 kDa) and have no particularly characteristic amino acid composition or secondary structure. Class III proteins were first discovered in the antarctic Zoarcidae (eelpouts), e.g. *Macrozoarces americanus, Rhigophila dearboni* and *Austrolycicthys brachycephalus*. Interestingly, the sequence of the antifreeze protein of the arctic eelpout *Lycodes polaris* agrees 78–84 % with proteins from the antarctic species [42]. DNA coding for class III proteins has also been found in four related fami-

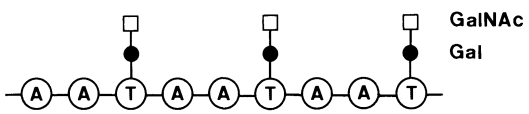


Fig. 5.3. The antifreeze glycoproteins (AFGPs) from teleosts are constructed of alanine and glycosylated threonine repeats

lies of the suborder Zoarcoidei: Stichaeidae (pricklebacks), Cryptacanthodidae (wrymouths), Pholididae (gunnels) and Anarhichadidae (wolf fish). The variety and distribution of the three AFP classes suggests that they arose independently and relatively recently: classes I and II in the Arctic and class III in the Antarctic. The effective selection pressure was probably the appearance of ice in surface waters. With the resulting unavoidable contact with ice, it became necessary for the surface-water fish to cool their body fluids below the freezing point of –0.8 °C; this was not a requirement for the deep-sea fish. The fact that the AFPs of all species show marked heterogeneity suggests that they are always encoded by multi-gene families. In fact, the sea raven *Hemitripterus americanus* has 12–15 AFP genes, the winter flounder *Pseudopleuronectes americanus* 30–40 genes, and different populations of the ocean pout *Macrozoarces americanus* between 30–40 and about 150 genes. The AFPs are only produced during the cold season; this seasonal regulation of AFP biosynthesis in the winter flounder involves a reduction during the winter in the formation or release of a specific inhibitor by the hypophysis [58].

Like AFPs, AFGPs have a **specific effect** on the freezing point but not on the melting point (thermal hysteresis). The large differences between the AFGPs and AFPs and within the AFPs prompts the question whether the effect of all these proteins involves the same mechanism; the answer is still uncertain [58]. The biosynthesis of both AFGPs and AFPs takes place in the liver. It is interesting to ask why the relatively small molecules of the antifreeze proteins are not lost in the urine. The antarctic Nototheniidae have aglomerulous kidneys in which there is no filtration. *Pseudopleuronectes* and *Rhigophila* do have glomeruli but filtration of the AFPs is apparently prevented [67].

Antifreeze proteins which give rise to thermal hysteresis have been detected in a series of terrestrial **arthropods:** beetle species from six different families, a cockroach, a bug, a scorpion fly and even a spider [65, 275]. However, only a few of these have been characterized in any detail. Seven antifreeze proteins of 9–17 kDa have been isolated from the larvae of the flour beetle *Tenebrio molitor*, and the concentration of these proteins increases to 10 mg/ml during cold acclimatization. The proteins vary greatly in amino acid composition, containing only a little alanine but large numbers of polar amino acids; one such protein contains 28 % cysteine. The antifreeze proteins from the caterpillars of *Choristoneura fumifrana* are of 13–16 kDa and also contain little alanine and up to 6 % cysteine [65]. In many insect species frost resistance is achieved not by the prevention of haemolymph freezing by special proteins or other substances but by the converse mechanism of ice nucleation: specific haemolymph proteins (ice nucleators) lead to rapid ice formation in the extracellular fluid under freezing conditions, thereby preventing lethal intracellular ice formation. There has as yet been little biochemical characterization of the ice nucleators. In the hornet *Vespula maculata* there is a protein with 20 % glutamate residues, and in the crane-fly *Tipula trivittata* several specific proteins and lipoproteins are involved [190].

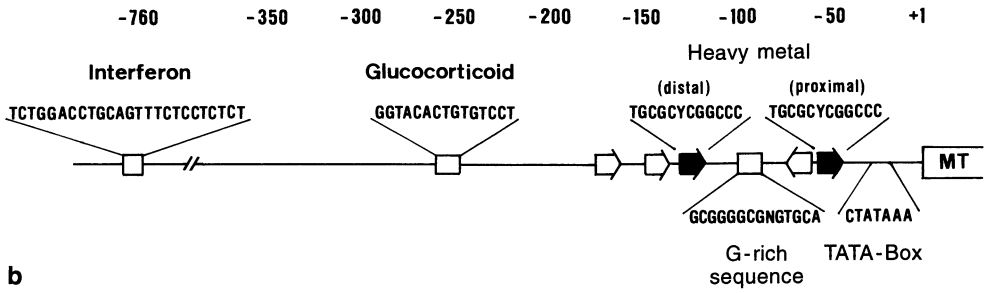
5.10 Metallothioneins

The metallothioneins (MTs) were discovered in 1957 in the liver and kidneys of mammals exposed to cadmium; they have subsequently been detected not only in the liver, kidneys and other organs of all vertebrate classes but also in a variety of invertebrates (annelids, crustaceans, insects, molluscs and echinoderms), in the ciliate *Tetrahymena* and in higher plants, lower fungi and yeast. They are therefore probably ubiquitous in the eukaryotes. In contrast, the metal-binding proteins of the prokaryotes do not appear to be related to the mammalian MTs. The MTs are small polypeptides of less than 10 kDa which contain 23–33 % cysteine residues and have the capacity for complex formation with 4–12 metal ions. The cysteine residues all take part in metal binding and form no intra- or intermolecular disulphide bridges. The MTs lack the aromatic amino acids phenylalanine and tyrosine and therefore, unlike most proteins, they show no absorption at 280 nm [93].

Mammalian MTs consist of 61–62 amino acids, including 20 cysteines, 6–8 lysines and 7–10 serines but no histidine; the N-terminus is N-acetylated. The polypeptide chain is divided into two domains, each with a cluster of cysteines which are able to bind four cadmium, four zinc or five to six copper ions; cluster A in the C-terminal α-domain contains 11 cysteines and cluster B in the N-terminal β-domain 9 cysteines (Fig. 5.4a). Each divalent metal ion is bound tetrahedrally with four cysteine residues. The affinity of the MTs for different metals varies considerably, being 1000-fold higher for cadmium than for zinc, and a further 100-fold higher for copper

```
Ech  MPDVKcVccTEGKEcAcFGQDccKGDTccGIcNAAcKcANGcKcGSGcScTEGNcAc
     |<---------- β ---------->|<---------- α ---------->|
Hum  MDPNcScAAGDScTcAGScKcKE-cKcTScKKSccSccPVGcAKcAQGcIcKGASDK-cSccA
Scy   PGPc-c--NDKcVcKEG-GcKEGcQcTScRcSPcEKcSSGc-KcANKEEcSKTcSKAcSccPT
Dm-n                MPc-PcGSGcKcASQATKGScNcGSDc-KcGGDKKSAcGcSE
Dm-o               MVcK-GcGTNcQcSAQKcGDNcAcNKDc-QcVcKNGPKDQccSNK
```

a



b

Fig. 5.4a, b. Metallothioneins. **a** Amino acid sequences. *Ech*, Strongylocentrotus purpuratus [188]; *Hum*, human MT II [142] with the domains β and α; *Scy, Scylla serrata* MT-1 [142]; *Dm-n, Drosophila melanogaster* Mtn [171]; *Dm-o, Drosophila melanogaster* Mto [171]. All sequences except Ech are aligned to give maximum agreement [142, 171]; cysteine residues are marked *c*. **b** Control sequences of mammalian MT genes (primates, rodents) [93]

than for cadmium. Affinities also vary between different MTs and between the two cysteine clusters of any one MT. The metal content of isolated MTs depends upon the species, tissue and previous history of the individual. It seems reasonable to assign the MTs a protective function. In fact, treatment with cadmium or another heavy metal leads to MT induction. Of the many mono- and divalent metals which can bind to MTs in vitro only copper and zinc are trace elements essential for metabolism; however, these too are toxic at higher concentrations. Just how far MTs are involved in normal zinc and copper metabolism is not clear. Although the transfer of zinc from mammalian MT to carboanhydrase, aldolase and alkaline phosphatase has been shown experimentally, tissue culture cells without functional MTs show normal growth and differentiation processes which depend upon the incorporation of zinc and copper [93].

All vertebrates possess two **MT classes**, MTI and MTII, which differ, for example in the horse, in 7 out of 61 amino acids, and which, for example in the rat, also show functional differences in the reconstitution of apo-carboanhydrase. Isoforms are present in several species and are described as MTI_A, MTI_B, etc. More than a dozen mammalian MTs have been sequenced either directly or via the DNA. The results of sequence analysis suggest a relatively slow **evolution of the metallothioneins;** thus, human, sheep and bovine MTIIs have more than 87 % amino acids in common [93]. Avian metallothioneins with 63 amino acids are slightly longer than those of mammals and agree in about 68 % of positions. It is interesting to note that the MT sequences of the duck species *Anas platyrhynchos* and *Cairina moschata* are identical with that of the chicken [149]. The MTs of the amphibians and fish are also very similar to those of the mammals [40]. In the mouse, there are only two MT genes (MTI and MTII), whereas rat MTI is a multi-gene family including one active gene and three retropseudogenes [5, 93]. The sheep has at least 9 MT genes and man has 12; only 6 of the latter (MTI_A, MTI_B, MTI_E, MTI_F, MTI_G and $MTII_A$) are functional [76, 200]. The transcription of MT genes is stimulated by heavy metals and also by glucocorticoids and interferon, and the regulatory sequences lying in front of the MT gene are very similar in the primates and rodents (Fig. 5.4b). A 108-kDa protein isolated from the mouse binds to the controlling element of the MTI gene [226]. In mammalian cell cultures it has been shown that cadmium treatment not only influences gene transcription but also causes amplification of the MTI and MTII genes [93].

The MTs of the crab *Scylla serrata*, the fly *Drosophila melanogaster* and the fungus *Neurospora* are apparently homologous to the mammalian MTs. The cysteine residues have identical locations, despite the fact that the length of the polypeptide chain varies between 57–58 residues in *Scylla* and 25 residues in *Neurospora*. However,

because the number and order of the cysteines in such metal-binding proteins are largely determined by functional considerations, convergent evolution cannot be excluded, and the apparent homology between the proteins of *Drosophila* and *Neurospora*, for example, and those of the vertebrates is difficult to confirm. The attempt to find a common central region in all these proteins has also failed to demonstrate unambiguous relationships [188]. The first **invertebrate metallothioneins** to be characterized closely were the isoforms MT-1 and MT-2 from the hepato-pancreas of the crab *Scylla serrata*. The chains of 58 and 57 amino acids contain 18 cysteines in two clusters, each of which binds three metal ions; the two forms agree 83 % overall, and MT-2 of *Scylla* is 47 % similar to human MTII (Fig. 5.4a). In contrast to the mammals, the *Scylla* MTs do not have an N-acetyl group blocking the N-terminus [142]. *Drosophila melanogaster* has two MT genes which code for very different sequences. The MT encoded by the gene Mtn includes 10 cysteine residues amongst its 40 amino acids, and the Mto product has 12 cysteines among 43 amino acids. The genes have only 11 amino acids in common, of which 8 are cysteines (Fig. 5.4a). The Mto gene is expressed from early in embryogenesis until the third larval stage, and Mtn is expressed from late embryogenesis (12–15 hours) until the adult stage [235]. Individuals with MT gene duplications have been found in natural *Drosophila* populations from several continents; these flies produce more MT mRNA and have increased cadmium and copper tolerance [138, 156].

During adaptation to high cadmium concentrations, the nematode *Caenorhabditis elegans* produces two cysteine-rich, MT-like proteins of 62 and 74 amino acids; however, their sequences have no significant homology to the mammalian MTs [113, 237]. A cysteine-rich protein with just as little homology to mammalian MTs has also been sequenced via the cDNA from the sea urchin *Strongylocentrotus purpuratus* (Fig. 5.4) [188, 189]. Finally, there are small, cadmium-binding glycoproteins in the hepato-pancreas of the sea-snail *Buccinum tenuissimum* and the terrestrial snails *Helix pomatia, Arianta arbustorium* and *Cepaea hortensis* [54, 212].

5.11 Ferritins

Under the conditions predominating inside organisms, iron exists almost entirely in the trivalent form; Fe^{3+} ions, however, have such strong hydrolytic properties that they precipitate as insoluble iron (III) hydroxide at concentrations above 10^{-17} mol/l. To retain essential iron in a soluble form it is bound to specific proteins. Iron transport in vertebrates is carried out by **transferrins** found in blood serum (serotransferrin), in milk and other secretions (lactotransferrin), and in egg albumen (ovotransferrin or conalbumin). These proteins were referred to in Section 5.3.1. The storage forms of iron are the **ferritins** and the structurally less well-defined haemosiderin. About 0.54 mmol (30 mg) iron is set free daily in human macrophages following the degradation of erythrocytes, and it is stored initially as ferritin. Iron which is slowly remobilized from the ferritin is bound to the apotransferrin of blood plasma, transported to the erythroblasts, and the cycle repeated. Some of the ferritins found in vertebrate cells have special functions, e.g. in the degradation of erythrocytes in the macrophages or as iron reserves in the liver, whereas others are simply involved in iron metabolism in the cell. The ferritin concentration can correspondingly vary by orders of magnitude. The ferritins are apparently ubiquitous, having been detected in many invertebrates, in higher plants and fungi (phyto- and mycoferritins), and in bacteria [52, 253].

In mammals, ferritins are found at especially high concentrations in the spleen, liver and bone marrow; the first to be investigated in detail was that from horse spleen. The iron-free apoferritins are hollow spheres of about 480 kDa, with an external diameter of 12 nm and walls 2 nm thick; they are made up of 24 subunits of about 20 kDa. The inner space consists of eight pockets for iron uptake. Loading of the apoferritins occurs when Fe^{2+} ions pass through channels in the protein membrane, are oxidized to Fe^{3+} and form the crystalline nucleus of iron (III) hydroxide; the simultaneous storage of phosphate ions leads to disturbance of the crystal structure. Each ferritin molecule can incorporate up to 4500 iron atoms, corresponding to an iron content of 30 %; normally, the iron content lies between 10 and 20 %, according to the tissue, with higher values, for example, in the spleen and liver, and lower values in heart muscle. The remobilization of the ferritin iron is the result of reduction or chelate formation [52, 253].

The **ferritin subunit** is a polypeptide of 172–183 amino acids, the chain of which is subdivided into two bundles of α-helices with a long connecting segment. Human mRNAs for two types of

subunit have been isolated; type H (182 amino acids, 21.1 kDa) predominates in the heart, and type L (175 amino acids, 19.8 kDa) predominates in the spleen and the liver. The two polypeptide chains differ at 45 % of their positions. The human genome contains 8L and 12H genes, amongst which, however, there are many pseudogenes. On the basis of these findings, the heterogeneity of mammalian ferritins was initially thought to be due to the presence in each organ of a population of hybrid molecules which vary in their average molecular mass according to the proportions of two different-sized subunits: e.g. in horse spleen 460 kDa with 10 % H, in liver 480 kDa with 40 % H, and in the heart 515 kDa with 85 %H. It was subsequently shown that the heterogeneity of the ferritin subunits is greater than that expected from the H/L hypothesis. Thus, in porcine spleen there are two different H chains, and in the tadpole of *Rana catesbeiana* there is, in addition to the H chain (175 amino acids) and L chains, a further chain which differs from H in 16 % of its 175 amino acids [51, 177]. SDS electrophoresis clearly shows that the subunit related to the human L chain does not always have a molecular mass lower than that of the H-like subunit [253].

The only **ferritin genes** so far characterized in any detail are those of man and the rat; they contain three introns in the coding sequence. The H and L mRNAs in the cytosol associate only with polyribosomes and become active in translation when the iron level in the cytoplasm rises; this regulation may be the function of an upstream 28-bp sequence, conserved in both man and the rat, which can fold back on itself [179]. The ferritins show a markedly slow **evolution:** the human and rat H chains have 95 % identical amino acids and the L chains 85 % ; the H chains of the mouse and chicken are 89 % identical. The H chain of frog larvae is 67 % similar to the human H chain, compared with 61 % similarity to its own L chain [177, 253]. Thus, the gene duplications leading to the separation of the different chain types occurred early in vertebrate evolution.

The iron nuclei of ferritins have diameters of 7–8 mm and are only visible in the electron microscope. In contrast, the iron concrescence of a **haemosiderin** is visible in the light microscope, but only with electron microscopy is it revealed as bunches of particles which are, in fact, very similar to ferritin nuclei. These particle groups are the residue of iron-laden secondary lysosomes (siderosomes) which have the same X-ray pattern as ferritin and immunological similarity to ferritin amongst their few proteins. Haemosiderins are apparently formed from ferritins through destruction of the protein shell by lysosomal enzymes [6].

The **ferritins of invertebrates** appear in all cases to be very similar to those of the mammals, but only in a few cases have they been investigated in detail. Ferritin-like particles from many insects have been viewed in the electron microscope but only the iron-binding haemolymph proteins of the lepidopterans have been characterized. *Manduca sexta* contains a 490-kDa protein consisting of 24- and 26-kDa subunits, and in *Calpodes ethlius* there is a glycoprotein of 600 kDa with 24- and 31-kDa subunits [111, 193]. The ferritins of molluscs have been investigated frequently, e.g. those involved in haemoglobin biosynthesis, or identified as components of radula teeth, or yolk proteins in snail eggs [164]. The apoferritin of the chiton (Polyplacophora) *Clavarizona hirtosa* has a molecular mass of 530 kDa, that of the mussel *Corbicula sandai* 503 kDa, and that of the earthworm *Octolasium complanatum* 460 kDa. Two different types of subunit have been found in many, if not all, invertebrates, e.g. in *Clavarizona* (28 and 25.5 kDa), in *Octolasium* (20 and 19.7 kDa), and in the egg yolk of the snail *Lymnaea stagnalis* (24 and 19 kDa); however, comparison to the H and L types of the mammals is not reliable without sequence data [7, 131]. Of the invertebrate ferritins mentioned above, only that of *L. stagnalis* shows immunological cross-reactivity to the ferritin of equine spleen [131]. Ferritin has been detected in the eggs and the hepato-pancreas of various gastropods; this suggests that, like a typical yolk protein, ferritin is produced in the hepato-pancreas and transported in the blood to the oocytes. However, the subunits of egg ferritin in *L. stagnalis* are 24 kDa, whereas those in the hepato-pancreas are 19 kDa and have a different (cyanogen bromide) cleavage peptide pattern; therefore, at least in this snail species, the metabolic relationship between the ferritins of the two organs is not at all clear [28].

References

1. Aketagawa J. et al.: Primary structure of Limulus anticoagulant anti-lipopolysaccharide factor. J. biol. Chem. 261: 7357–65 (1986)
2. Albers J. J. and Segrest J. P. (eds.): Plasma lipoproteins. Acad. Press, New York 1986
3. Allen W. V. and Conley H.: Transport of lipids in the blood of the Pacific oyster, Crassostrea gigas (Thunberg): Comp. Biochem. Physiol. Pt. B 71: 201–207 (1982)

4. Altruda F. et al.: The primary structure of human hemopexin deduced from cDNA sequence: evidence for internal repeating homology. Nucleic Acids Res. 13: 3841–59 (1985)
5. Andersen R. D. et al.: Rat metallothionein-I structural gene and three pseudogenes, one of which contains 5'-regulatory sequences. Mol. Cell Biol. 6: 302–314 (1986)
6. Andrews S. C., Teffry A. and Harrison P. M.: Siderosomal ferritin. The missing link between ferritin and haemosiderin? Biochem. J. 245: 439–446 (1987)
7. Arosio P. et al.: Properties of ferritin from the earthworm Octolasium complanatum. Biochim. biophys. Acta 774: 264–269 (1984)
8. Babin P.J. and Vernier J. M.: Plasma lipoproteins in fish (Review). J. Lipid Res. 30: 467–489 (1989)
9. Baert J. L.: A 15000-Mr protein proteolytically derived from vitellogenin within oocyte of Perinereis cultrifera (polychaete annelid). Eur. J. Biochem. 177: 625–630 (1988)
10. Baeuerle P. A., Lottspeich F. and Huttner W. B.: Purification of yolk protein 2 of Drosophila melanogaster and identification of its site of tyrosine sulfation. J. Biol. Chem. 263: 14925–29 (1988)
11. Bailey S. et al.: Molecular structure of serum transferrin at 3.3-Å resolution. Biochemistry 27: 5804–12 (1988)
12. Baker B. S., Steven J. and Tata J. R.: Vitellogenin genes and their products in closely and distantly related species of Xenopus. Comp. Biochem. Physiol. Pt. B 82: 497–505 (1985)
13. Baker M. E.: Is vitellogenin an ancestor of apolipoprotein B-100 of human low density lipoprotein and human lipoprotein lipase? Biochem. J. 255: 1057–60 (1988)
14. Baker M. E.: Invertebrate vitellogenin is homologous to human von Willebrand factor. Biochem. J. 256: 1059–61 (1988)
15. Baldwin G. S. and Weinstock J.: Nucleotide sequence of porcine liver transferrin. Nucleic Acids Res. 16: 8720 (1988)
16. Bartfeld N. S. and Law J. W.: Isolation and molecular cloning of transferrin from the tobacco hornworm, Maduca sexta. Sequence similarity to the vertebrate transferrins. J. Biol. Chem. 265: 21684–91 (1990)
17. Barwig B.: Isolation and characterization of plasma coagulogen (PC) of the cockroach Leucophaea maderae (Blattaria). J. comp. Physiol. B 155: 135–143 (1985)
18. Bast R. E. and Gibson A. R.: Characterization of reptilian vitellogenin: Subunit composition and molecular weights of vitellogenin from the colubrid snake Thamnophis sirtalis (L.). Comp. Biochem. Physiol. Pt. B 80: 409–418 (1985)
19. Bean D. W., Shirk P. D. and Brookes V. J.: Characterization of yolk proteins from the eggs of the Indian meal moth, Plodia interpunctella. Insect Biochem. 18: 199–210 (1988)
20. Beenakkers A. M. T., Chino H. and Law J. H.: Lipophorin nomenclature. Insect Biochem. 18: 1–2 (1988)
21. de Bianchi A. G. and Pereira S. D.: Comparative structural studies of vitellogenin and vitellin of Rhynchosciara americana. Comp. Biochem. Physiol. Pt. B 80: 895–900 (1985)
22. de Bianchi A. G. et al.: Vitellogenin and vitellin of Musca domestica. Quantifications and synthesis by fat bodies and ovaries. Insect Biochem. 15: 77–84 (1985)
23. Bikfalvi A. and Beress L.: Natural proteinase inhibitors: blood coagulation inhibition and evolutionary relationships. Comp. Biochem. Physiol. Pt. B 87: 435–441 (1987)
24. Bobák P., Stratil A. and Valenta M.: A comparison of molecular weights of transferrins from various vertebrates. Comp. Biochem. Physiol. Pt. B 79: 113–117 (1984)
25. Bodine A. B., Luer C. A. and Gangjee S.: Determination of ceruloplasmin and other copper transport ligands in the blood sera of the nurse shark and clearnose shark. Comp. Biochem. Physiol. Pt. B 77: 779–783 (1984)
26. Bohn H.: Hemolymph clotting in insects. In: Brehélin M. (ed.): Immunity in invertebrates, pp. 188–207. Springer, Berlin 1986
27. Borovsky D. and Whitney P. L.: Biosynthesis, purification, and characterization of Aedes aegypti vitellin and vitellogenin. Arch. Insect Biochem. Physiol. 4: 81–99 (1987)
28. Bottke W. and Crichton R. R.: Vitellogenic ferritin of Lymnaea stagnalis L. (Mollusca, Gastropoda) differs in structure from soma cell type ferritin. Comp. Biochem. Physiol. Pt. B 77: 57–61 (1984)
29. Bownes M.: Expression of the genes coding for vitellogenin (yolk protein). Annual Rev. Entomol. 31: 507–531 (1986)
30. Bradley G., Naudé R. J. and Oelofsen W.: The isolation and characterization of serum albumin from the ostrich (Struthio camelus). Comp. Biochem. Physiol. Pt. B 82: 829–835 (1985)
31. Brookes V. J.: The polypeptide structure of vitellogenin and vitellin from the cockroach, Leucophaea maderae. Arch. Insect Biochem. Physiol. 3: 577–591 (1986)
32. Burcham T. S. et al.: Purification and primary sequences of the major arginine-containing antifreeze glycopeptides from the fish Eleginus gracilis. J. biol. Chem. 261: 6384–89 (1986)
33. Byrne B. M. et al.: Rudimentary phosvitin domain in a minor chicken vitellogenin gene. Biochemistry 28: 2572–77 (1989)
34. Calabrese L., Carbonaro M. and Musci G.: Chicken ceruloplasmin. Evidence in support of a trinuclear cluster involving type 2 and 3 copper centers. J. Biol. Chem. 263: 6480–83 (1988)
35. Cardoen J., Huybrechts R. and de Loof A.: Yolk polypeptide processing in the fat body of Sarcophaga bullata. Arch. Insect Biochem. Physiol. Suppl. 1: 87–96 (1986)
36. Carter D. C. et al.: Three-dimensional structure of human serum albumin. Science 244: 1195–98 (1989)
37. Cayrol C. and Deparis P.: Identification of the vitellogenin proteins in the newt Pleurodeles waltlii (urodele amphibian). Comp. Biochem. Physiol. Pt. B 83: 135–142 (1986)
38. de Chaffoy de Corcelles D. and Kondo M.: Lipovitellin from the crustacean Artemia salina. Biochemical analysis of lipovitellin complex from the yolk granules. J. biol. Chem. 255: 6727–33 (1980)
39. Chapman M. J.: Animal lipoproteins: chemistry, structure, and comparative aspects (Review). J. Lipid Res. 21: 789–853 (1980)
40. Chan K. M. et al.: Molecular cloning of metallothionein cDNA and analysis of metallothionein gene expression in winter flounder tissues. Can. J. Zool. 67: 2520–27 (1989)

41. Chen A. C. et al.: Vitellogenesis in the stable fly, Stomoxys calcitrans. Comp. Biochem. Physiol. Pt. B 88: 897–903 (1987)
42. Cheng C. H. C., and de Vries A. L.: Structures of antifreeze peptides from the anarxtic eel pout, Australycicthys brachycephalus. Biochim. biophys. Acta 997: 55–64 (1989)
43. Cheng S. M. et al.: Characterization of a complementary deoxyribonucleic acid for the coagulogen of Limulus polyphemus. Biochim. biophys. Acta 868: 1–8 (1986)
44. Chinzei Y. and Yano I.: Fat body is the site of vitellogenin synthesis in the soft tick, Ornithodoros moubata. J. comp. Physiol. B 155: 671–678 (1985)
45. della Choppa G. and Engelmann F.: The vitellogenin of Leucophaea maderae. Synthesis as a large phosphorylated precursor. Insect Biochem. 17: 401–415 (1987)
46. Clark S. B., Tercyak A. M. and Glander K. E.: Plasma lipoproteins of free-ranging howling monkeys (Alouatta palliata). Comp. Biochem. Physiol. Pt. B 88: 729–735 (1987)
47. Coddeville B. et al.: Primary structure of horse serotransferrin glycans. Demonstration that heterogeneity is related to the number of glycans and to the presence of N-acetylneuraminic acid and N-acetyl-4-O-acetylneuraminic acid. Eur. J. Biochem. 186: 583–590 (1989)
48. Cole K. D. et al.: Primary structure and comparative sequence analysis of an insect apolipoprotein. Apolipoprotein-III from Manduca sexta. J. biol. Chem. 262: 11794–11800 (1987)
49. Cole K. D., Smith A. F. and Wells M. A.: The structure of the apolipoprotein-III gene from Manduca sexta. Insect Biochem. 20: 381–388 (1990)
50. Cole T. and Schreiber G.: Synthesis of thiostatins (major acute phase α1-proteins) in different strains of Rattus norvegicus. Comp. Biochem. Physiol. Pt. B 93: 813–818 (1989)
51. Collawn J. F. jr. et al.: Isolation and partial amino acid sequence of three subunit species of porcine spleen ferritin: Evidence of multiple H subunits. Arch Biochem. Biophys. 259: 105–113 (1987)
52. Crichton R. R. and Charloteaux-Wauters M.: Iron transport and storage. Eur. J. Biochem. 164: 485–506 (1987)
53. Cryer A. and Jones H. M.: The distribution of lipoprotein lipase (clearing factor lipase) activity in the adiposal, muscular and lung tissues of ten animal species. Comp. Biochem. Physiol. Pt. B 63: 501–505 (1979)
54. Dallinger R., Berger B. and Bauer-Hilty A.: Purification of cadmium-binding proteins from related species of terrestrial helicidae (gastropoda, mollusca). A comparative study. Mol. cell. Biochem. 85: 135–145 (1989)
55. Danielsson H. and Sjovall J. (eds.): Sterols and bile acids. New comprehensive Biochemistry Vol. 12. Elsevier, Amsterdam 1985
56. Datta S. et al.: Structure and expression of dog apolipoprotein C-II and C-III mRNAs. Implications for the evolution and functional constraints of apolipoprotein structure. J. biol. Chem. 262: 10588–93 (1987)
57. Davidson W. S. et al.: Palmitate-binding, serum albumin-like proteins in salmonids. FEBS Letters 233: 299–302 (1988)
58. Davies P. L. and Hew C. L.: Biochemistry of fish antifreeze proteins. Faseb J. 4: 2460–68 (1990)
59. Deguchi K. et al.: Water-soluble lipoproteins from yolk granules in sea urchin eggs. II. Chemical composition. J. Biochem. Tokyo 85: 1519–25 (1979)
60. Delers F. et al.: Characterization of a chicken hemoglobin-binding protein: a novel haptoglobin. Comp. Biochem. Physiol. Pt. B 74: 745–748 (1983)
61. Dhadialla T. S. and Raikhel A. S.: Biosynthesis of mosquito vitellogenin. J. Biol. Chem. 265: 9924–33 (1990)
62. Dillwith J. W., Lenz C. J. and Chippendale G. M.: Isolation and characterization of lipophorin from the hemolymph of diapausing larvae of the south-western corn borer, Diatraea grandiosella. J. comp. Physiol. B 156: 783–789 (1986)
63. Doolittle R. F.: The evolution of the vertebrate plasma proteins. Biol. Bull. 172: 269–283 (1987)
64. Doolittle R. F. and Riley M.: The amino-terminal sequence of lobster fibrinogen reveals common ancestry with vitellogenins. Biochem. biophys. Res. Commun. 167: 16–19 (1990)
65. Duman J. and Horwath K.: The role of hemolymph proteins in the cold tolerance of insects. Annual Rev. Physiol. 45: 261–270 (1983)
66. Dziegielewska K. et al.: The complete cDNA and amino acid sequence of bovine fetuin. Its homology with α2-HS glycoprotein and relation to other members of the cystatin superfamily. J. Biol. Chem. 265: 4354–57 (1989)
67. Eastman J. T. et al.: Renal conservation of antifreeze peptide in Antarctic eelpout, Rhigophila dearborni. Nature 282: 217–218 (1979)
68. von Eckardstein A. et al.: Structural analysis of human apolipoprotein A-I variants. Amino acid substitutions are nonrandomly distributed throughout the apolipoprotein A-I primary structure. J. Biol. Chem. 265: 8610–17 (1990)
69. Engelmann F.: Insect vitellogenin: Identification, biosynthesis, and role in vitellogenesis. Adv. Insect Physiol. 14: 49–108 (1979)
70. Fagotto F.: Yolk degradation in tick eggs. Arch. Insect Biochem. Physiol. 14: 217–252 (1990)
71. Fainaru M. et al.: Interactions between human and carp (Cyprinus carpio) low density lipoproteins (LDL) and LDL receptors. Comp. Biochem. Physiol. Pt. B 91: 331–338 (1988)
72. Fernando-Warnakulasuriya G. J. P. and Wells M. A.: Isolation and characterization of lipophorin from Drosophila melanogaster larvae. Arch. Insect Biochem. Physiol. 8: 243–248 (1988)
73. Filosa M. F., Sargent P. A. and Youson J. H.: An electrophoretic and immunoelectrophoretic study of serum proteins during the life cycle of the lamprey, Petromyzon marinus L. Comp. Biochem. Physiol. Pt. B 83: 143–149 (1986)
74. Fischer A. et al.: In vitro studies on the origin of yolk protein and its uptake by oocytes in the annelid, Nereis virens (Polychaeta). (In German). Verh. Deutsche Zool. Ges. 77: 145 (1984)
75. Fleming R. E. and Gitlin J. D.: Primary structure of rat ceruloplasmin and analysis of tissue-specific gene expression during development. J. Biol. Chem. 265: 7701–07 (1990)
76. Foster R. et al.: Structure and expression of the human metallothionein-IG gene. Differential promoter activity of two linked metallothionein-I genes in

response to heavy metals. J. Biol. Chem. 263: 11528–35 (1988)
77. Francis R. T. jr. and Becker R. R.: Two hemoglobin-binding proteins identified in the plasma of the amphibian Taricha granulosa. Comp. Biochem. Physiol. Pt. B 77: 341–347 (1984)
78. Friedel T. and Gillott C.: Contribution of male-produced proteins to vitellogenesis in Melanoplus sanguinipes. J. Insect Physiol. 23: 145–151 (1977)
79. Fujii T. et al.: Structure of the gene for the arylphorin-type storage protein SP2 of Bombyx mori. J. Biol. Chem. 264: 11020–25 (1989)
80. Fujikawa K. et al.: Amino acid sequence of human factor XI, a blood coagulation factor with four tandem repeats that are highly homologous with plasma prekallikrein. Biochemistry 25: 2417–24 (1986)
81. Fung W. P. et al.: Structure and expression of the rat transthyretin (prealbumin) gene. J. biol. Chem. 263: 480–488 (1988)
82. Furie B. and Furie B. C.: The molecular basis of blood coagulation. Cell 53: 505–518 (1988)
83. Garabedian M. J. et al.: The nucleotide sequence of the gene coding for Drosophila melanogaster yolk protein 3. Gene 55: 1–8 (1987)
84. Gemmill R. M. et al.: Isolation of mosquito vitellogenin genes and induction of expression by 20-hydroxyecdysone. Insect Biochem. 16: 761–774 (1986)
85. Ghidalia W. et al.: Coagulation in decapod crustacea. Comparative study of the clotting process in species from groups A, B and C. J. comp. Physiol. B 142: 473–478 (1981)
86. Gibbs P. E. M. et al.: Structure, polymorphism, and novel repeated DNA elements revealed by a complete sequence of the human α-fetoprotein gene. Biochemistry 26: 1332–43 (1987)
87. Glinski Z., Jarosz J. and Wernicki A.: Serological characterization of soluble proteins in greater wax moth larval blood. Comp. Biochem. Physiol. Pt. B 84: 131–135 (1986)
88. le Goff D. et al.: Lipid and apolipoprotein distribution as a function of density in equine plasma. Comp. Biochem. Physiol. Pt. B 93: 371–377 (1989)
89. Gondim K. C. et al.: Lipophorin from Rhodnius prolixus: purification and partial characterization. Insect Biochem. 19: 153–161 (1989)
90. Graf J. D. and Fischberg M.: Albumin evolution in polyploid species of the genus Xenopus. Biochem. Genetics 24: 821–837 (1986)
91. Ha Y. C. and Barter P. J.: Differences in plasma cholesteryl ester transfer activity in sixteen vertebrate species. Comp. Biochem. Physiol. Pt. B 71: 265–269 (1982)
92. Haefliger D. N. et al.: Amphibian albumins as members of the albumin, alpha-fetoprotein, vitamin-D-binding protein multigene family. J. mol. Evol. 29: 344–354 (1989)
93. Hamer D. H.: Metallothionein. Annual Rev. Biochem. 55: 913–951 (1986)
94. Hammer R. E. et al.: Diversity of alpha-fetoprotein gene expression in mice is generated by a combination of separate enhancer elements. Science 235: 53–58 (1987)
95. Harnish D. G. and White B. N.: Insect vitellins: Identification, purification, and characterization from eight orders. J. exp. Zool. 220: 1–10 (1982)
96. Hatzopoulos P. and Kambysellis M. P.: Comparative biochemical and immunological analysis of the three vitellogenins from Drosophila grimshawi. Comp. Biochem. Physiol. Pt. B 89: 557–564 (1988)
97. Haunerland N. H. and Bowers W. S.: A larval-specific lipoprotein: Purification and characterization of a blue chromoprotein from Heliothis zea. Biochem. biophys. Res. Commun. 134: 580–586 (1986)
98. Haunerland N. H. et al.: Lipophorin from the grasshopper, Gastrimargus africanus. Isolation and properties of apolipophorin III. Insect Biochem. 16: 797–802 (1986)
99. Haunerland N. H. and Bowers W. S.: Comparative studies on arthropod lipoproteins. Comp. Biochem. Physiol. Pt. B 92: 137–141 (1989)
100. Hauptmann R. et al.: Vascular anticoagulant β: a novel human Ca/phospholipid-binding protein that inhibits coagulation and phospholipase-A2 activity. Its molecular cloning, expression and comparison with VAC-α. Eur. J. Biochem. 185: 63–71 (1989)
101. Hayakawa Y.: Characterization of lipophorin receptor in locust flight muscles. Biochim. biophys. Acta 919: 58–63 (1987)
102. Hayashi T. et al.: Further localization of binding sites for thrombin and protein-C in human thrombomodulin. J. Biol. Chem. 265: 20156–59 (1990)
103. Herbert P. N. et al.: Homologues of the human C and A apolipoproteins in the Macaca fascicularis (Cynomolgus) monkey. Biochemistry 26: 1457–63 (1987)
104. van Heusden M. C. et al.: The recycling of protein components of the flight-specific lipophorin in Locusta migratoria. Insect Biochem. 17: 771–776 (1987)
105. Hilewicz-Grabska M. et al.: Purification and partial characterization of goose ceruloplasmin. Arch. Biochem. Biophys. 260: 18–27 (1988)
106. Hirayama Y. and Chino H.: Lipid transfer particle in locust hemolymph: purification and characterization. J. Lipid Res. 31: 793–799 (1990)
107. Holden H. M. et al.: Crystallization and preliminary analysis of crystals of apolipophorin III isolated from Locusta migratoria. J. Biol. Chem. 263: 3960–62 (1988)
108. Hollanders B. et al.: Comparison of the lipoprotein profiles obtained from rat, bovine, horse, dog, rabbit and pig serum by a new two-step ultracentrifugal gradient procedure. Comp. Biochem. Physiol. Pt. B 84: 83–89 (1986)
109. Houseman J. G. and Morrison P. E.: Absence of female-specific protein in the hemolymph of stable fly Stomoxys calcitrans (L.) (Diptera: Muscidae). Arch. Insect Biochem. Physiol. 3: 205–213 (1986)
110. Hsiao K. C. et al.: An antifreeze glycopeptide gene from the Antarctic cod Notothenia coriceps neglecta encodes a polyprotein of high peptide number. Proc. Nat. Acad. Sci. USA 87: 9265–69 (1990)
111. Huebers H. A. et al.: Iron binding proteins and their roles in the tobacco hornworm, Manduca sexta (L.). J. comp. Physiol. B 158: 291–300 (1988)
112. Ikura K. et al.: Amino acid sequence of Guinea pig liver transglutaminase from its cDNA sequence. Biochemistry 27: 2898–2905 (1988)
113. Imagawa M. et al.: Characterization of metallothionein cDNAs induced by cadmium in the nematode Caenorhabditis elegans. Biochem. J. 268: 237–240 (1990)

114. Imamura S. et al.: Molecular cloning and primary structure of rat thyroxine-binding globulin. Biochemistry 30: 5406–11 (1991)
115. Imboden H. et al.: Characterization of the native vitellogenin and vitellin of the cockroach, Nauphoeta cinerea, and comparison with other species. Insect Biochem. 17: 353–365 (1987)
116. Inagaki S and Yamashita O.: Complete amino acid sequence of Bombyx egg-specific protein deduced from cDNA clone. Arch. Insect Biochem. Physiol. 10: 131–139 (1989)
117. Indrasith L. S., Sasaki T. and Yamashita O.: A unique protease responsible for selective degradation of a yolk protein in Bombyx mori. Purification, characterization, and cleavage profile. J. biol. Chem. 263: 1045–51 (1988)
118. Inglis A. S., Atrike P. M. and Burley R. W.: Two low-molecular-weight apoproteins (apovitellinins I and II) from a piloprotein of goose's egg yolk: a comparison with related species. Aust. J. biol. Sci. 35: 263–269 (1982)
119. Jeltsch J. M. et al.: Sequence of the chicken ovotransferrin gene. Nucleic Acids Res. 15: 7643–45 (1987)
120. Jenny R. J. et al.: Complete cDNA and derived amino acid sequence of human factor V. Proc. Nat. Acad. Sci. USA 84: 4846–50 (1987)
121. Kanost M. R. et al.: Insect haemolymph proteins. Adv. Insect Physiol. 22: 299–396 (1990)
122. Kant J. A. et al.: Evolution and organization of the fibrinogen locus on chromosome 4: gene duplication accompanied by transposition and inversion. Proc. Nat. Acad. Sci. USA 82: 1244–48 (1985)
123. Karathanasis S. K., Yunis Y. and Zannis V. I.: Structure, evolution, and tissue-specific synthesis of human apolipoprotein A-IV. Biochemistry 25: 3962–70 (1986)
124. Katagiri C.: Structure of lipophorin in insect blood: location of phospholipid. Biochim. biophys. Acta 834: 139–143 (1985)
125. Katase H. and Chino H.: Transport of hydrocarbons by the lipophorin of insect hemolymph. Biochim. biophys. Acta 710: 341–348 (1982)
126. Kawooya J. K. and Law J. H.: Role of lipophorin in lipid transport to the insect egg. J. Biol. Chem. 263: 8748–53 (1988)
127. Kim H. R., Ko Y. G. and Mayer R. T.: Purification, characterization, and synthesis of vitellin from the cabbage butterfly Pieris rapae L. Arch. Insect Biochem. Physiol. 9: 67–79 (1988)
128. Kim H. R., Kang C. S. and Mayer R. T.: Storage proteins of the fall webworm, Hyphantria cunea Drury. Arch. Insect Biochem. Physiol. 10: 115–130 (1989)
129. Kim H. R., Seo S. J. and Mayer R. T.: Properties, synthesis, and accumulation of storage proteins in Pieris rapae L. Arch. Insect Biochem. Physiol. 10: 215–228 (1989)
130. Kim J. W. et al.: Stage-specific detection of a DNA-binding protein for the storage protein gene of Sarcophaga peregrina. Biochim. biophys. Acta 1008: 79–84 (1989)
131. Kim K. S., Webb J. and Macey D. J.: Properties and role of ferritin in the hemolymph of the chiton Clavarizona hirtosa. Biochim. biophys. Acta 884: 387–394 (1986)
132. Kitagawa H. et al.: Differing expression patterns and evolution of the rat kininogen gene family. J. biol. Chem. 262: 2190–98 (1987)
133. Koschinsky M. L. et al.: Complete cDNA sequence of human preceruloplasmin. Proc. Nat. Acad. Sci. USA 83: 5086–90 (1986)
134. Kulakosky P. C. and Telfer W. H.: Lipophorin as a yolk precursor in Hyalophora cecropia: Uptake kinetics and competition with vitellogenin. Arch. Insect Biochem. Physiol. 14: 269–285 (1990)
135. Kunkel J. G. and Nordin J. H.: Yolk proteins. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology Vol. 1, pp. 83–111. Pergamon Press, Oxford 1985
136. Kurosky A., Hay R. E. and Bowman B. H.: Canine haptoglobin: a unique haptoglobin subunit arrangement. Comp. Biochem. Physiol. Pt. B 62: 339–344 (1979)
137. Lampreave F. and Piñeiro A.: Characterization of an new alpha-glycoprotein as the major serum component in later fetal and newborn pigs. Comp. Biochem. Physiol. Pt. B 72: 215–219 (1982)
138. Lange B. W., Langley C. H. and Stephan W.: Molecular evolution of Drosophila metallothionein genes. Genetics 126: 921–932 (1990)
139. Lange R. H., Grodzinski Z. and Kilarski W.: Yolk-platelet crystals in three ancient bony fishes. Cell Tissue Res. 222: 159–165 (1982)
140. Leclerc R. F. and Miller S. G.: Identification and molecular analysis of storage proteins from Heliothis virescens. Arch. Insect Biochem. Physiol. 14: 131–150 (1990)
141. Lee R. F. and Puppione D. L.: Lipoproteins I and II from the hemolymph of the blue crab Callinectes sapidus: Lipoprotein II associated with vitellogenesis. J. exp. Zool. 248: 278–289 (1988)
142. Lerch K., Ammer D. and Olafson R. W.: Crab metallothionein. Primary structures of metallothioneins 1 and 2. J. biol. Chem. 257: 2420–26 (1982)
143. Leslie N. D. et al.: The chicken urokinase-type plasminogen activator gene. J. Biol. Chem. 265: 1339–44 (1990)
144. Levedakou E. N. and Sekeris C. E.: Isolation and characterization of vitellin from the fruitfly, Dacus oleae. Arch. Insect Biochem. Physiol. 4: 297–311 (1987)
145. Levenbook L.: Insect storage proteins. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology Vol. 10, pp. 307–346. Pergamon Press, Oxford 1985
146. Levenbook L., Bauer A. C. and Chou J. Y.: Ubiquitin in the blowfly Calliphora vicina. Partial characterization, titers during the life cycle and relation to calliphorin breakdown. Insect Biochem. 16: 509–515 (1986)
147. Li W. et al.: The apolipoprotein multigene family: biosynthesis, structure, structure-function relationships, and evolution (Review). J. Lipid Res. 29: 245–271 (1988)
148. Lim B. C., McArdle H. J. and Morgan E. H.: Transferrin-receptor interaction and iron uptake by reticulocytes of vertebrate animals – a comparative study. J. comp. Physiol. B 157: 363–371 (1987)
149. Lin L. Y. and Huang P. C.: Complete homology in metallothionein from two genera of ducks and their hybrids. Biochem. biophys. Res. Commun. 168: 182–187 (1990)
150. Lundwall A. et al.: Isolation and sequence of the cDNA for human protein S, a regulator of blood coa-

gulation. Proc. Nat. Acad. Sci. USA 83: 6716–20 (1986)
151. Luo C. C. et al.: Structure and evolution of the apolipoprotein multigene family. J. mol. Biol. 187: 325–340 (1986)
152. Maeda N.: Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair. The haptoglobin-related gene contains a retrovirus-like element. J. biol. Chem. 260: 6698–6709 (1985)
153. Mancuso D. J. et al.: Structure of the gene for human von Willebrand factor. J. Biol. Chem. 264: 19514–27 (1989)
154. Mao S. H., Guo Y. W. and Yin F. Y.: Physical properties of some snake plasma albumins. Comp. Biochem. Physiol. Pt. B 82: 655–658 (1985)
155. Marinotti O. and de Bianchi A. G.: Uptake of storage protein by Musca domestica fat body. J. Insect Physiol. 32: 819–825 (1986)
156. Maroni G. et al.: Metallothionein gene duplications and metal tolerance in natural populations of Drosophila melanogaster. Genetics 117: 739–744 (1987)
157. Marti T., Schaller J. and Rickli E. E.: Determination of the complete amino-acid sequence of porcine miniplasminogen. Eur. J. Biochem. 149: 279–285 (1985)
158. Martone R. L. et al.: Thransthyretin is synthesized in the mammalian eye. Biochem. biophys. Res. Commun. 151: 905–912 (1988)
159. Masuda H. and Oliveira P. L.: Characterization of vitellin and vitellogenin from Rhodnius prolixus. Identification of phosphorylated compounds in the molecule. Insect Biochem. 15: 543–550 (1985)
160. Matsubara T., Hara A. and Takano K.: Immunochemical identification and purification of coelomic fluid-specific protein in chum salmon (Oncorhynchus keta). Comp. Biochem. Physiol. 81B: 309–314 (1985)
161. Matsushima T., Getz G. S. and Meredith S. C.: Primary structure of guinea pig apolipoprotein E. Nucleic Acids Res. 18: 202 (1990)
162. McEvoy S. M. and Maeda N.: Complex events in the evolution of the haptoglobin gene cluster in primates. J. Biol. Chem. 263: 15740–47 (1988)
163. Medina M., Leon P. and Vallejo C. G.: Drosophila cathepsin B-like proteinase: A suggested role in yolk degradation. Arch. Biochem. Biophys. 263: 355–363 (1988)
164. Miksys S. and Saleuddin A. S. M.: Ferritin as an exogenously derived yolk protein in Helisoma duryi (Mollusca: Pulmonata). Can. J. Zool. 64: 2678–82 (1986)
165. Miller S. G.: Synthesis and transport of storage proteins by testes in Heliothis virescens. Arch. Insect Biochem. Physiol. 14: 151–170 (1990)
166. Mintzas A. C. and Kambysellis M. P.: The yolk proteins of Drosophila melanogaster: isolation and characterization. Insect Biochem. 12: 25–33 (1982)
167. Mitsis P. G. and Wensink P. C.: Purification and properties of yolk protein factor I, a sequence-specific DNA-binding protein from Drosophila melanogaster. J. Biol. Chem. 164: 5195–5202 (1989)
168. Miura K. and Shimizu I.: Identification and properties of lipophorin of the silkworm Bombyx mori. Comp. Biochem. Physiol. Pt. B 89: 95–103 (1988)
169. Miura K. and Shimizu I.: Changes in properties in lipophorin of the silkworm, Bombyx mori, with ontogeny. Comp. Biochem. Physiol. Pt. B 92: 197–204 (19989)
170. Mizejewski G. J. and Jacobson H. I. (eds.): Biological activities of alpha-1-fetoprotein. CRC Press, Boca Raton USA 1987
171. Mokdad R., Debec A. and Wegnez M.: Metallothionein genes in Drosophila melanogaster constitute a dual system. Proc. Nat. Acad. Sci. USA 84: 2658–62 (1987)
172. de Morales M. H., Vallés A. M. and Baerga-Santini C.: Studies of the egg proteins of tropical lizards: Purification and partial characterization of yolk proteins of Anolis pulchellus. Comp. Biochem. Physiol. Pt. B 87: 125–136 (1987)
173. Morgan J. G., Holbrook N. J. and Crabtree G. R.: Nucleotide sequence of the gamma chain of rat fibrinogen: conserved intronic sequences. Nucleic Acids Res. 15: 2774–76 (1987)
174. Mori S., Izumi S. and Tomino S.: Structures and organization of major plasma protein genes of the silkworm, Bombyx mori. J. mol. Biol. 218: 7–12 (1991)
175. Mosesson M. W. et al. (eds.): Fibrinogen 3. Biochemistry, biological functions, gene regulation and expression. Elsevier, New York 1988
176. Moskaitis J. E., Pastori R. L. and Schoenberg D. R.: The nucleotide sequence of Xenopus laevis transferrin messenger RNA. Nucleic Acids Res. 18: 6135 (1990)
177. Moskaitis J. E., Pastori R. L. and Schoenberg D. R.: Sequence of Xenopus laevis ferritin mRNA. Nucleic Acids Res. 18: 2184 (1990)
178. Murai T. et al.: Isolation and characterization of rainbow trout C-reactive protein. Develop. comp. Immunol. 14: 49–58 (1990)
179. Murray M. T., White K. and Munro H. N.: Conservation of ferritin heavy subunit gene structure: Implications for the regulation of ferritin gene expression. Proc. Nat. Acad. Sci. USA 84: 7438–42 (1987)
180. Muta T. et al.: Limulus factor-C. An endotoxin-sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like, and lectin-like domains. J. Biol. Chem. 266: 6554–61 (1991)
181. Nagao E., Takahashi N. and Chino H.: Asparagine-linked oligosaccharides of locust lipophorin. Insect Biochem. 17: 531–538 (1987)
182. Nagler J. J. and Idler D. R.: Ovarian uptake of vitellogenin and another very high density lipoprotein in winter flounder (Pseudopleuronectes americanus) and their relationship with yolk proteins. Biochem. Cell Biol. 68: 330–335 (1990)
183. Nahon J. L.: The regulation of albumin and α-fetoprotein gene expression in mammals. Biochimie 69: 445–459 (1987)
184. Nahon J. L. et al.: Albumin and α-fetoprotein gene expression in various nonhepatic tissues. J. Biol. Chem. 263: 11436–42 (1988)
185. Nardelli D. et al.: Vertebrate and nematode genes coding for yolk proteins are derived from a common ancestor. Biochemistry 26: 6397–6402 (1987)
186. Nardelli D. et al.: Comparison of the organization and fine structure of a chicken and a Xenopus laevis vitellogenin gene. J. biol. Chem. 262: 15377–85 (1987)
187. Naumann U. and Scheller K.: Complete cDNA and gene sequence of the developmental regulated arylphorin of Calliphora vicina and its homology to insect hemolymph proteins and arthropod hemocyanins.

Biochem. biophys. Res. Commun. 177: 963–972 (1991)
188. Nemer M. et al.: Sea urchin metallothionein sequence: key to an evolutionary diversity. Proc. Nat. Acad. Sci. USA 82: 4992–94 (1985)
189. Nemer M. et al.: Structure, spatial and temporal expression of two sea urchion metallothionein genes, SpMTB1 and SpMTA. J. Biol. Chem. 266: 6586–93 (1991)
190. Neven L. G. et al.: Purification and characterization of an insect hemolymph lipoprotein ice nucleator: evidence for the importance of phosphatidylinositol and apolipoprotein in the ice nucleator activity. J. comp. Physiol. B 159: 71–82 (1989)
191. Nguyen N. Y. et al.: Isolation and characterization of Limulus C-reactive protein genes. J. biol. Chem. 261: 10450–55 (1986)
192. Nguyen N. Y. et al.: The amino acid sequence of Limulus C-reactive protein. Evidence of polymorphism. J. biol. Chem. 261: 10456–65 (1986)
193. Nichol H. and Locke M.: The characterization of ferritin in an insect. Insect Biochem. 19: 587–588 (1989)
194. Nikkila H., Gitlin J. D. and Mullereberhard U.: Rat hemopexin: molecular cloning, primary structural characterization, and analysis of gene expression. Biochemistry 30: 823–829 (1991)
195. O'Hara P. et al.: Nucleotide sequence of the gene coding for human factor VII, a vitamin K-dependent protein participating in blood coagulation. Proc. Nat. Acad. Sci. USA 84: 5158–62 (1987)
196. Ostr E. O. et al.: Studies on the carbohydrate moiety of vitellogenin from the tobacco hornworm, Manduca sexta. Insect Biochem. 16: 471–478 (1986)
197. Palli S. R. and Locke M.: The synthesis of hemolymph proteins by the larval midgut of an insect Calpodes ethlius (Lepidoptera: Hesperidae). Insect Biochem. 17: 561–572 (1987)
198. Palli S. R. and Locke M.: The synthesis of hemolymph proteins by the larval epidermis of an insect Calpodes ethlius (Lepidoptera: Hesperidae). Insect Biochem. 17: 711–722 (1987)
199. Peferoen M. and de Loof A.: Synthesis of vitellogenic and non-vitellogenic yolk proteins by the fat body and the ovary of Leptinotarsa decemlineata. Comp. Biochem. Physiol. 83B: 251–254 (1986)
200. Peterson M. G., Hannan F. and Mercer J. F. B.: The sheep metallothionein gene family: Structure, sequence and evolutionary relationship of five linked genes. Eur. J. Biochem. 174: 417–424 (1988)
201. Pierce A. et al.: Molecular cloning and sequence analysis of bovine lactotransferrin. Eur. J. Biochem. 196: 177–184 (1991)
202. Pontes M. et al.: cDNA sequences of two apolipoproteins from lamprey. Biochemistry 26: 1611–17 (1987)
203. Prasad S. V. et al.: Lipoprotein biosynthesis in the larvae of the tobacco hornworm, Manduca sexta. J. biol. Chem. 17174–76 (1986)
204. Purcell J. P., Kunkel J. G. and Nordin J. H.: Yolk hydrolase activities associated with polypeptide and oligosaccharide processing of Blatella germanica vitellin. Arch. Insect Biochem. Physiol. 8: 39–58 (1988)
205. Quinitio E. T. et al.: Identification and characterization of vitellin in a hermaphrodite shrimp, Pandalus kessleri. Comp. Biochem. Physiol. Pt. B 94: 445–451 (1989)
206. Ray B. K. and Ray A.: Complementary DNA cloning and nucleotide sequence of rabbit serum amyloid α protein. Biochem. biophys. Res. Commun. 178: 68–72 (1991)
207. Rehn K. G. and Rolim A. L. R.: Purification and properties of a storage protein from the hemolymph of Rhodnius prolixus. Insect Biochem. 20: 195–201 (1990)
208. Riazi A., Fremont L. and Gozzelino M. T.: Characterization of egg yolk proteins from rainbow trout Salmo gairdneri (Rich.). Comp. Biochem. Physiol. Pt. B 89: 399–407 (1988)
209. Rina M. D. and Mintzas A. C.: Two vitellins-vitellogenins of the Mediterranean fruit fly Ceratitis capitata: a comparative biochemical and immunological study. Comp. Biochem. Physiol. Pt. B 86: 801–808 (1987)
210. Robey F. A., Tanaka T. and Liu T. Y.: Isolation and characterization of two major serum proteins from the dogfish, Mustelus canis, C-reactive protein and amyloid P component. J. biol. Chem. 258: 3889–94 (1983)
211. Roehrkasten A. et al.: Isolation of the vitellogenin-binding protein from locust ovaries. Arch. Insect Biochem. Physiol. 10: 141–149 (1989)
212. Roesijadi G. et al.: Structure and time-dependent behavior of acetylated and non-acetylated forms of a molluscan metallothionein. Biochim. biophys. Acta 1074: 230–236 (1991)
213. Rogie A. and Skinner E. R.: The roles of the intestine and liver in the biosynthesis of plasma lipoproteins in the rainbow trout, Salmo gairdneri Richardson. Comp. Biochem. Physiol. Pt. B 81: 285–289 (1985)
214. Rousset A. and Bitsch C.: Synthesis and maturation of the major yolk protein by the ovaries of the firebrat, Thermobia domestica (Insecta, Thysanura). Comp. Biochem. Physiol. Pt. B 94: 599–605 (1989)
215. Ryan R. O., Schmidt J. O. and Law J. H.: Chemical and immunological properties of lipophorins from seven insect orders. Arch. Insect Biochem. Physiol. 1: 375–383 (1984)
216. Ryan R. O. et al.: Arylphorin from Manduca sexta: carbohydrate structure and immunological studies. Arch. Biochem. Biophys. 243: 115–124 (1985)
217. Ryan R. O. et al.: Major hemolymph proteins from larvae of the black swallowtail butterfly, Papilio polyxenes. Arch. Insect Biochem. Physiol. 3: 539–550 (1986)
218. Ryan R. O. et al.: Lipoprotein interconversions in an insect, Manduca sexta. Evidence for a lipid transfer factor in the hemolymph. J. biol. Chem. 261: 563–568 (1986)
219. Ryan R. O. et al.: Manduca sexta lipid transfer particle acts upon a lipoprotein to catalyze lipid and apoprotein disproportion. J. Biol. Chem. 265: 546–552 (1990)
220. Ryan R. O. et al.: Characterization of apolipophorin III from Barytettix psolus and Melanoplus differentialis. Insect Biochem. 20: 127–133 (1990)
221. Ryan R. O., Howe A. and Scraba D. G.: Studies on the morphology and structure of the plasma lipid transfer particle from the tobacco hornworm, Manduca sexta. J. Lipid Res. 31: 871–879 (1990)
222. Sato Y. and Yamashita O.: Post-translational processing in the synthesis of egg-specific protein in the silkworm, Bombyx mori. Insect Biochem. 19: 293–300 (1989)

223. Scanu A. M. and Spector A. A. (eds.): Biochemistry and biology of plasma proteins. M. Dekker, New York 1986
224. Scheller K.: 3. Workshop on larval serum proteins of insects. Biol. Chem. Hoppe-Seyler 368: 571–578 (1987)
225. Schubiger J. L. and Wahli W.: Linkage arrangement in the vitellinogenin gene family of Xenopus laevis as revealed by gene segregation analysis. Nucleic Acids Res. 14: 8723–34 (1986)
226. Séguin C. and Prévost J.: Detection of a nuclear protein that interacts with a metal regulatory element of the mouse metallothionein 1 gene. Nucleic Acids Res. 16: 10547–60 (1988)
227. Seymour J. L. et al.: Decorsin. A potent glycoprotein IIb-IIIa antagonist and platelet aggregation inhibitor from the leech Macrobdella decora. J. Biol. Chem. 265: 10143–47 (1990)
228. Shalaby F. and Shafritz D. A.: Exon skipping during splicing of albumin mRNA precursors in Nagase analbuminemic rats. Proc. Nat. Acad. Sci. USA 87: 2652–56 (1990)
229. Shapiro J. P.: Isolation and fluorescence studies on a lipophorin from the weevil Diaprepes abbreviatus. Arch. Insect Biochem. Physiol. 7: 119–131 (1988)
230. Sharrock W. J. et al.: Two distinct yolk lipoprotein complexes from Caenorhabditis elegans. J. Biol. Chem. 265: 14422–31 (1990)
231. Sheridan M. A., Friedlander J. K. I. and Allen W. V.: Chylomicra in the serum of postprandial steelhead trout (Salmo gairdneri). Comp. Biochem. Physiol. Pt. B 81: 281–284 (1985)
232. Shigeura H. T. and Haschemeyer A. E. V.: Purification of egg-yolk proteins from the Antarctic fish Caenocephalus aceratus. Comp. Biochem. Physiol. Pt. B 80: 935–939 (1985)
233. Shipman B. A. et al.: Purification and properties of a very high density lipoprotein from the hemolymph of the honeybee Apis mellifera. Biochemistry 26: 1885–89 (1987)
234. Shyu A. B., Blumenthal T. and Raff R. A.: A single gene encoding vitellogenin in the sea urchin Strongylocentrotus purpuratus – Sequence at the 5' end. Nucleic Acids Res. 15: 10405–17 (1987)
235. Silar P. et al.: Metallothionein Mto gene of Drosophila melanogaster: Structure and regulation. J. mol. Biol. 215: 217–224 (1990)
236. de Silva H. V. et al.: Apolipoprotein J: Structure and tissue distribution. Biochemistry 29: 5380–89 (1990)
237. Slice L. W., Freedman J. H. and Rubin C. S.: Purification, characterization and cDNA cloning of a novel metallothionein-like, cadmium-binding protein from Caenorhabditis elegans. J. Biol. Chem. 265: 256–263 (1990)
238. Spicer E. K. et al.: Isolation of cDNA clones coding for human tissue factor: Primary structure of the protein and cDNA. Proc. Nat. Acad. Sci. USA 84: 5148–52 (1987)
239. Srimal S. et al.: The complete amino acid sequence of coagulogen isolated from southeast Asian horseshoe crab, Carcinoscorpius rotundicauda. J. Biochem. Tokyo 98: 305–318 (1985)
240. Steyrer E., Barber D. L. and Schneider W. J.: Evolution of lipoprotein receptors. The chicken oocyte receptor for very low density lipoprotein and vitellogenin binds the mammalian ligand apolipoprotein. J. Biol. Chem. 265: 19575–81 (1990)
241. Stifani S., George R. and Schneider W. J.: Solubilization and characterization of the chicken oocyte vitellogenin receptor. Biochem. J. 250: 467–475 (1988)
242. Stifani S., Nimpf J. and Schneider W. C.: Vitellogenesis in Xenopus laevis and chicken: cognate ligands end oocyte receptors. The binding site for vitellogenin is located on lipovitellin I. J. Biol. Chem. 265: 882–888 (1990)
243. Stifani S. et al.: Regulation of oogenesis: the piscine receptor for vitellogenin. Biochim. biophys. Acta 1045: 271–279 (1990)
244. Storella J. R., Wojchowski D. M. and Kunkel J. G.: Structure and embryonic degradation of two native vitellins in the cockroach, Periplaneta americana. Insect Biochem. 15: 259–275 (1985)
245. Suhr-Jessen P. and Rasmussen L. P. D.: Biochemical and immunological characterization of Mytilus edulis plasma polypeptides. Comp. Biochem. Physiol. Pt. B 91: 45–49 (1988)
246. Sundelin J. et al.: The primary structure of rabbit and rat prealbumin and a comparison with the tertiary structure of human prealbumin. J. biol. Chem. 260: 6481–87 (1985)
247. Surholt B. et al.: Lipophorin conversions during flight of the death's-head hawkmoth Acherontia atropos. Insect Biochem. 18: 117–126 (1988)
248. Suzuki S., Yamasaki K. and Katakura Y.: Vitellogenin synthesis by fat body and ovary in the terrestrial isopod, Armadillidium vulgare. Gen. comp. Endocrinol. 74: 120–126 (1989)
249. Takahashi S. Y.: Post-translational processing, by phosphorylation, of vitellogenin and vitellin from the silkworm, Bombyx mori. Comp. Biochem. Physiol. Pt. B 87: 255–266 (1987)
250. Tarugi P. et al.: Absence of apolipoprotein B-48 in the chick, Gallus domestucus. J. Lipid Res. 31: 417–427 (1990)
251. Taza I. et al.: Isolation and characterization of the major and the minor phosvitin from eight species (three genera) of salmonid fishes. Comp. Biochem. Physiol. Pt. B 89: 475–482 (1988)
252. Tennyson G. E. et al.: Expression of apolipoprotein B mRNAs encoding higher- and lower-molecular weight isoproteins in rat liver and intestine. Proc. Nat. Acad. Sci. USA 86: 500–504 (1989)
253. Theil E. C.: Ferritin: Structure, gene regulation, and cellular function in animals, plants, and microorganisms. Annual Rev. Biochem. 56: 289–315 (1987)
254. Thorstensen K. and Romslo I.: The role of transferrin in the mechanism of cellular iron uptake (Review Article). Biochem. J. 271: 1–10 (1990)
255. Titani K. and Walsh K. A.: Human von Willebrand factor: The molecular glue of platelets. Trends biochem. Sci. 13: 94–97 (1988)
256. Tokunaga F. et al.: Lipopolysaccharide-sensitive serine protease zymogen (factor C) of horseshoe crab hemocytes – Identification and alignment of proteolytic fragments produced during the activation show that it is a novel type of serine protease. Eur. J. Biochem. 167: 405–416 (1987)
257. Tom M., Goren M. and Ovadia M.: Purification and partial characterization of vitellin from the ovaries of Parapenaeus longirostris (Crustacea, Decapoda). Comp. Biochem. Physiol. Pt. B 87: 17–23 (1987)

258. Topham R. et al.: Isolation, purification and characterization of an iron-binding protein from the horseshoe crab (Limulus polyphemus). Biochem. J. 252: 151–157 (1988)
259. Tsuchida K. and Wells M. A.: Isolation and characterization of a lipoprotein receptor from the fat body of an insect, Manduca sexta. J. Biol. Chem. 265: 5761–67 (1990)
260. Tu, G. et al.: The nucleotide sequence of transthyretin cDNA isolated from a sheep choroid plexus cDNA library. Nucleic Acids Res. 17: 6384 (1989)
261. Vanderhorst D. J.: Lipid transport function of lipoproteins in flying insects. Biochim. biophys. Acta 1047: 195–211 (1990)
262. Vermeer C.: Gamma-carboxyglutamate-containing proteins and the vitamin K-dependent carboxylase (Review Article). Biochem. J. 266: 625–636 (1990)
263. Wallace R. A. and Begovac P. C.: Phosvitines in Fundulus oocytes and eggs. Preliminary chromatographic and electrophoretic analyses together with biological considerations. J. biol. Chem. 260: 11268–74 (1985)
264. Wallace R. A., Hoch K. L. and Carnevali O.: Placement of small lipovitellin subunits within the vitellogenin precursor in Xenopus laevis. J. mol. Biol. 213: 407–409 (1990)
265. Wang X. Y., Cole K. D. and Law J. H.: cDNA cloning and deduced amino acid sequence of microvitellogenin, a female-specific hemolymph and egg protein from the tobacco hornworm, Manduca sexta. J. Biol. Chem. 263: 8851–55 (1988)
266. Wang Y. Z. et al.: Complete sequence of the lamprey fibrinogen α-chain. Biochemistry 28: 9801–06 (1989)
267. Watson F. W., Levin J. and Nivitsky T. J. (eds.): Detection of bacterial andotoxins with the Limulus amebocyte lysate test. Alan R. Liss, New York 1987
268. Weissbach L. and Grieninger G.: Bipartite mRNA for chicken alpha-fibrinogen potentially encodes an amino acid sequence homologous to beta- and gamma-fibrinogens. Proc. Nat. Acad. Sci. USA 87: 5198–5202 (1990)
269. Wheeler D. E. and Kawooya J. K.: Purification and characterization of honey bee vitellogenin. Arch. Insect Biochem. Physiol. 14: 253–267 (1990)
270. Whitehead A. S. et al.: Mouse C-reactive protein. Generation of cDNA clones, structural analysis, and induction of mRNA during inflammation. Biochem. J. 266: 283–290 (1990)
271. Williams J.: The evolution of transferrin. Trends biochem. Sci. 7: 394–397 (1982)
272. Willott E., Wang X. Y. and Wells M. A.: cDNA and gene sequence of Manduca sexta arylphorin, an aromatic acid-rich larval serum protein. Homology to arthropod hemocyanins. J. Biol. Chem. 264: 19052–59 (1989)
273. Winter C. E. et al.: Yolk proteins and their plasmatic precursor in the tetraploid Odontophrynus americanus (Amphibia, Anura). Comp. Biochem. Physiol. Pt. B 82: 515–524 (1985)
274. Wojchowski D. M. and Kunkel J. G.: Purification of two distinct oocyte vitellins and identification of their corresponding vitellogenins in fat body and hemolymph of Blaberus discoidalis. Insect Biochem. 17: 189–198 (1987)
275. Wu D. W. et al.: Purification and characterization of antifreeze proteins from larvae of the beetle Dendroides canadiensis. J. comp. Physiol. B 161: 271–278 (1991)
276. Xu X. and Doolittle R. F.: Presence of a vertebrate fibrinogen-like sequence in an echinoderm. Proc. Nat. Acad. Sci. USA 87: 2097–2101 (1990)
277. Yamamoto T.: Characterization of guinea-pig high-molecular-weight kininogen as multi-functional molecule. Biochim. biophys. Acta 914: 259–274 (1987)
278. Yan Y. L., Kunert C. J. and Postlethwait J. H.: Sequence homologies among the three yolk polypeptide (Yp) genes in Drosophila melanogaster. Nucleic Acids Res. 15: 67–85 (1987)
279. Yang C. et al.: The primary structure of apolipoprotein A-I from rabbit high-density lipoprotein. Eur. J. Biochem. 160: 427–431 (1986)
280. Yang C. et al.: The complete amino acid sequence of proapolipoprotein A-I of chicken high density lipoproteins. FEBS Letters 224: 261–266 (1987)
281. Yu J. Y. L. et al.: Sexual patterns of protein metabolism in liver and plasma of hagfish, Eptatretus stouti, with special reference to vitellogenesis. Comp. Biochem. Physiol. Pt. B 65: 111–117 (1980)
282. Zongza V. and Dimitriadis G. J.: Vitellogenesis in the insect Dacus oleae: Isolation and characterization of yolk protein mRNA. Insect Biochem. 18: 651–660 (1988)
283. Zucker-Aprison E. and Blumenthal T.: Potential regulatory elements of nematode vitellogenin genes revealed by interspecies sequence comparison. J. mol. Evol. 28: 487–496 (1989)
284. Zwaal R. F. A. and Hemker H. C. (eds.): Blood coagulation. In: New comprehensive Biochemistry, Vol. 13. Elsevier, Amsterdam 1986

6 Immunoproteins

All cells are able to communicate with the environment, i.e. receive signals from their surroundings and broadcast their (immunological) identity to the outside. The registration of signals is via receptors on the surface; for example, there are receptors for hormones or neurotransmitters and also for the recognition of other cells as "self" or "non-self". Foreign cells or substances induce specific defence reactions **(immune responses)**. In both the vertebrates and invertebrates, immune responses may involve certain blood cells (cell-mediated immunity) or soluble proteins of the plasma (humoral immunity). In accordance with the general aims of this book, the emphasis of the account that follows is placed on the comparative biochemistry and evolution of the proteins involved in the immune response. However, their functions cannot be understood without some basic knowledge of cell biology.

Of the various **cell types** in the blood and other extracellular fluids, the B and T lymphocytes and the macrophages are the most important for the responses of man and other mammals to molecular structures (antigens) recognized as foreign. Soluble macromolecular antigens in the blood can induce an immune response only after they are taken up by macrophages or other antigen-presenting cells (APCs) and subsequently "presented" to the B lymphocytes or certain T lymphocytes (T helper cells). The antigen is proteolytically cleaved into fragments in the APCs prior to presentation. Low molecular weight substances can function as antigens only when they occur as haptens bound to a "self" protein. Each B lymphocyte carries on its surface immunoglobulin molecules of one of the many existing structural variants. These function as receptors for the corresponding (complementary) antigen. There is a special type of receptor on T lymphocytes. T helper cells can recognize an antigen only when it is presented on the surface of an APC together with an Ia antigen, which is the product of a class II MHC (major histocompatibility complex) gene. In collaboration with the APCs and the T helper cells, the B cells specific for the antigen are induced to proliferate; fully differentiated plasma cells are produced and these synthesize and secrete large quantities of antibodies. The soluble **immunoglobulins** produced by the plasma cells bind to the complementary antigens on the surface of cells which are then destroyed (lysed) by the complement system. Soluble antigens associate with the complementary immunoglobulins to form large molecular complexes, which bind to specific receptors on neutrophil blood cells and natural killer (NK) cells and are thereby removed from the circulatory system [159].

Cell-mediated immunity requires the activation of the cytotoxic T cells and phagocytic macrophages. The extent of the response is determined by T helper and suppressor cells. The cytotoxic T lymphocytes only destroy cells which carry the self MHC-class-I antigen in addition to the foreign antigen, e.g. virus-infected cells of the

body. The restriction of antigen recognition by cytotoxic T cells and T helper cells via MHC is a specific task of the T cell receptors; the biological role of this process lies in the safer distinction between self and non-self.

Only recently has it become clear that the immunoglobulins, T cell receptors, MHC antigens, cell-adhesion molecules (CAM) and several other proteins are all members of one and the same super-family which, besides its involvement in the immune system, has apparently played a central role in the **evolution of cell–cell recognition** in vertebrates. In addition to the immunoglobulins, further plasma proteins involved in the immune response are the components of the complement system, the lymphokines and other immuno-modulator proteins, the antiviral interferons, acute-phase proteins and, especially in lower vertebrates, lectins, agglutinins and lysins with low specificity.

Compared with the defence mechanisms of the vertebrates, those of the **invertebrates** are far less complex. In the latter, phagocytosis by specialized blood cells is widely found as a cellular mechanism. In addition, the extracellular fluids of many invertebrates and lower vertebrates contain relatively unspecific defence proteins (lectins, agglutinins, lysins), the concentrations of which may, in some cases, be increased in response to infection. Finally, many of the low molecular weight defence substances of the invertebrates (which will be dealt with in Chap. 19 as products of secondary metabolism) have primarily bactericidal or fungicidal functions.

The **molecular structure of the cell surface**, which is responsible for the external identification of the cell and its contact interactions with other cells, has mainly been investigated in the vertebrates. However, in some special instances attention has been drawn to the corresponding membrane proteins of invertebrates. For example, the surface proteins of the sponges have been studied in some detail because of the unique property of these most primitive, multicellular animals to reaggregate spontaneously a complete, viable organism from isolated single cells. The surface proteins of the parasitic trypanosomes and malarial agents are of medical interest because as antigens they are responsible for the immune response of their hosts; at the same time they have attracted the attention of molecular biologists.

6.1 The Immunoglobulin Super-Family

The characteristic structural element of this super-family is the antibody fold or **Ig homology unit**, a polypeptide chain of about 100 amino acids which is folded into two layers of antiparallel β-sheets and is stabilized by a central disulphide bridge (Fig. 6.1). The Ig super-family consists of a structurally and functionally very heterogeneous group of surface proteins which are involved in widely different processes of cell–cell interaction (Fig. 6.2). The genes of the heavy and light immunoglobulin chains, of the T cell receptors and of the MHC classes I and II form multigene families. However, the Ig super-family also includes β_2-microglobulin (β_2m), which is associated with class I MHC molecules; various accessory T cell proteins that, amongst other functions, are involved in the recognition of MHC antigens (CD8, CD4, CD3 complex); the Thy-1 antigen from thymocytes and lymphocytes; cell-adhesion molecules such as N-CAM; various growth-factor receptors with tyrosine kinase activity (platelet-derived growth factor receptor, PDGF R; colony-stimulating factor 1 receptor, CSF-1 R); the pregnancy-specific β_1-glycoprotein (PSβG); an opioid-binding protein; and many others [79, 162, 202]. Most members of the Ig super-family are known only from the vertebrates. Thy-1, however, is also found in invertebrate cells and Ig homology units have been

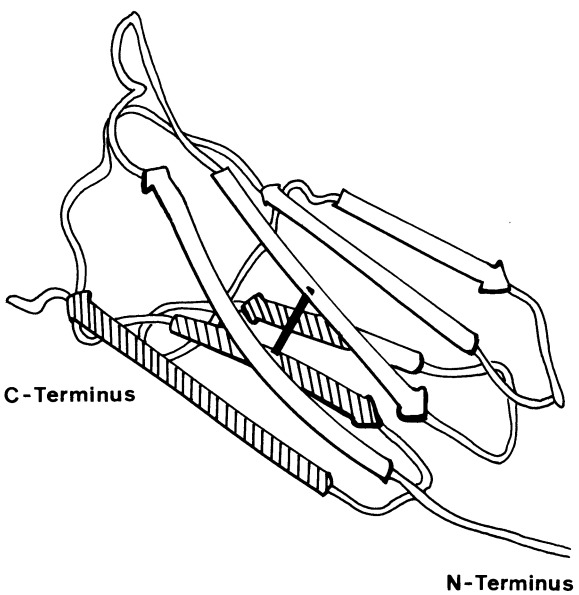


Fig. 6.1. The structural units of the immunoglobulin superfamily (immunoglobulin fold), illustrated here by the C domain of the human λ-chain, consist of two layers of antiparallel β structures. The *black rod* symbolizes a disulphide bridge

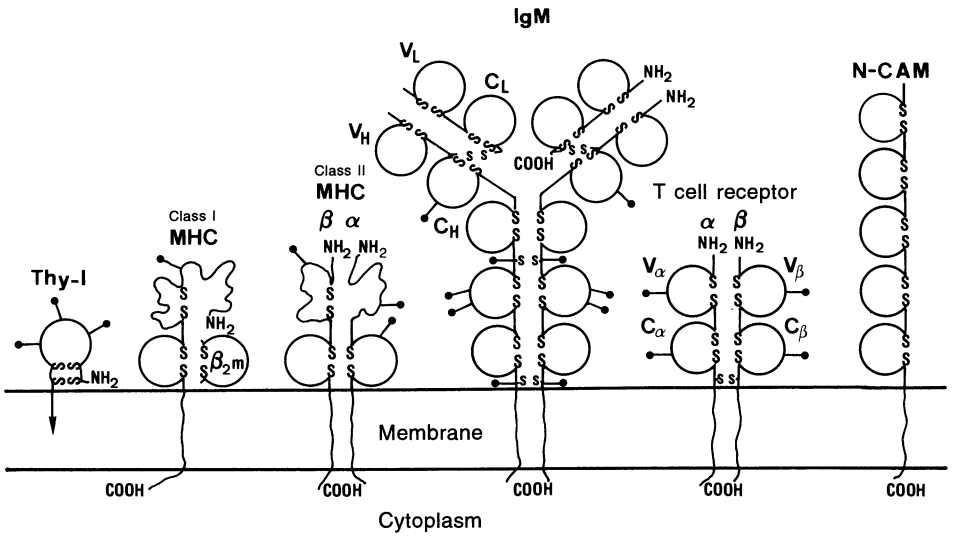


Fig. 6.2. Different membrane-bound proteins from the immunoglobulin super-family [74, 79]. *Thy 1*, Thy-1 glycoprotein of the thymocytes; *MHC class I*, MHC (major histocompatibility complex) antigen class I; *MHC class II*, MHC antigen class II; *IgM*, immunoglobulin M; *N-CAM*, neuronal cell-adhesion molecule. The *drumsticks* represent carbohydrate chains

detected in various cell-adhesion proteins from insects.

Extensive material is already available which allows **sequence comparisons within the Ig superfamily;** thus, 77 different Ig-C domains, more than 800 Ig-V domains, various β_2-microglobulins, domains of MHC classes I and II, Thy-1, poly-Ig receptors and T cell receptors have been compared with each other [8]. These sequences can be divided into four groups:

1. The Ig-C domains are similar to β_2m, the α3 domain of class I MHC, and the α2 and β2 domains of MHC class II.
2. Ig-V is similar to Thy-1.
3. α2 of MHC class I and β1 of MHC class II have distant similarity to the immunoglobulins.
4. The α1 domains of MHC classes I and II show no sequence similarity to the immunoglobulins.

In general, the agreement between domains of the same type in different species is greater than between different types in the same species. For example, the sequence similarity between the mouse and man for Ig-Cγ1 is 64 % and for Ig-Cγ2 is even 73 %, whereas Cγ1 and Cγ2 themselves agree by only 24–25 %. Comparison of the human and mouse α2 and β2 domains of MHC class II also show 82 % agreement, whereas between the two types themselves there is only 30 % similarity. Thus, each type is optimized for its specific function and contains a large number of strongly conserved positions. There is unfortunately very little corresponding information for non-mammals.

6.2 Immunoglobulins

6.2.1 Basic Structure of Immunoglobulins

Man and other mammals possess five **classes of immunoglobulins** with different structures and functions: IgM, IgG, IgA, IgD and IgE (Table 6.1). IgG makes up the greater part of the blood plasma γ-globulins and is usually taken as the prototype of the immunoglobulins. The introduction of a new antigen results initially in the production of IgM; the switch to IgG occurs after about 1–2 weeks, and renewed antigen injection mainly causes the synthesis of further IgG. IgA is the only immunoglobulin found in secretions such as saliva or milk. The functions of IgD and IgE, found at much lower concentrations, have not been completely determined.

The **basic structural unit** of the immunoglobulins consists of two light (L) chains and two heavy (H) chains coupled by a disulphide bridge (Fig. 6.2). There are five types of H chain (γ, μ, α, δ and ge), corresponding to the five Ig classes, and two types of L chain (κ and λ), which occur in all five classes. All H chains contain carbohydra-

Table 6.1. Human immunoglobulins [23]

	IgG				IgM	IgA			IgD	IgE
	IgG1	IgG2	IgG3	IgG4		IgA1	IgA2	sigA		
Serum concentration (mg/ml)	5	3	0.5	0.5	1.5	3	0.5	0.05	0.03	$5 \cdot 10^{-5}$
Quaternary structure	mono	mono	mono	mono	pent	mono	mono	di	mono	mono
Size (kDa)	105	150	160	150	970	160	160	415	175	190
H chain										
– designation	γ^1	γ^2	γ^3	γ^4	μ	α^1	α^2	α^1/α^2	δ	ε
– size (kDa) incl. carbohydrate	51	51	56	51	72	57	57	–	63	72
– number of domains	4	4	4	4	5	4	4	–	4	5
– amino acids in hinge region	15	12	62	12	–	20	7	–	64	–

mono, monomer; pent, pentamer; di, dimer

tes of different amount and composition. Each chain consists of a variable (V) region, where the specific interaction with the antigen occurs, and a constant (C) region, which is responsible for effector functions such as binding to the cell surface or complement activation. The chains are subdivided into **domains** of about 110 amino acids, corresponding in their spatial structure to the Ig homology units mentioned above. The subdivision of the polypeptide chain into domains corresponds to the exons of the Ig gene (Figs. 6.3 and 6.4a). The light chains contain a variable and a constant domain (V_L and C_L). In the heavy chains, the V region is equivalent to one domain; the C regions from μ and ε consist of four domains, CH1 to CH4; in the evolution of the other H chains, the original CH2 domain was reduced to the hinge (H-) region (Table 6.1).

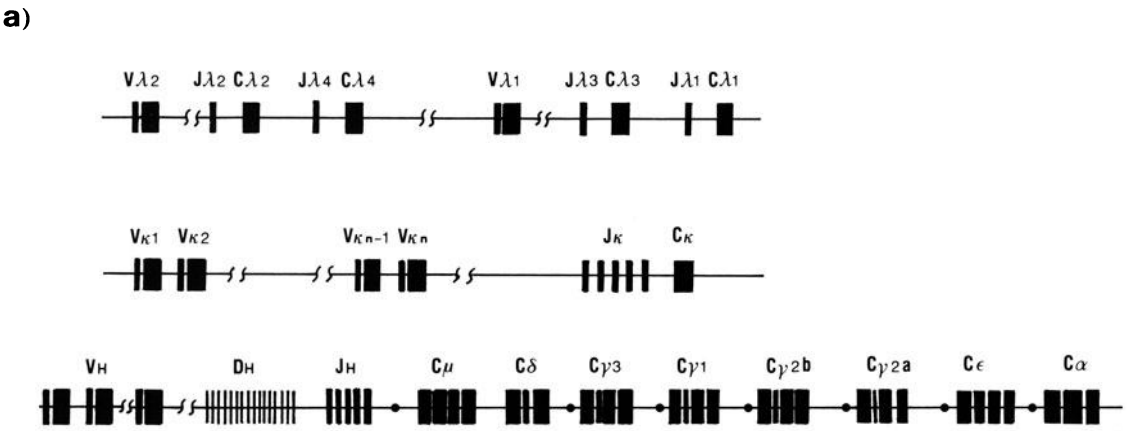


Fig. 6.3a, b. Immunoglobulin genes. **a** Genes of the mouse [23, 74]. *Black rectangles* represent exons and *dots* indicate the S regions. **b** The order of the genes in the IgH chain of the mouse, rat and man [19]. The boundaries of the segments formed by gene duplication in the human gene cluster are marked by *vertical lines*

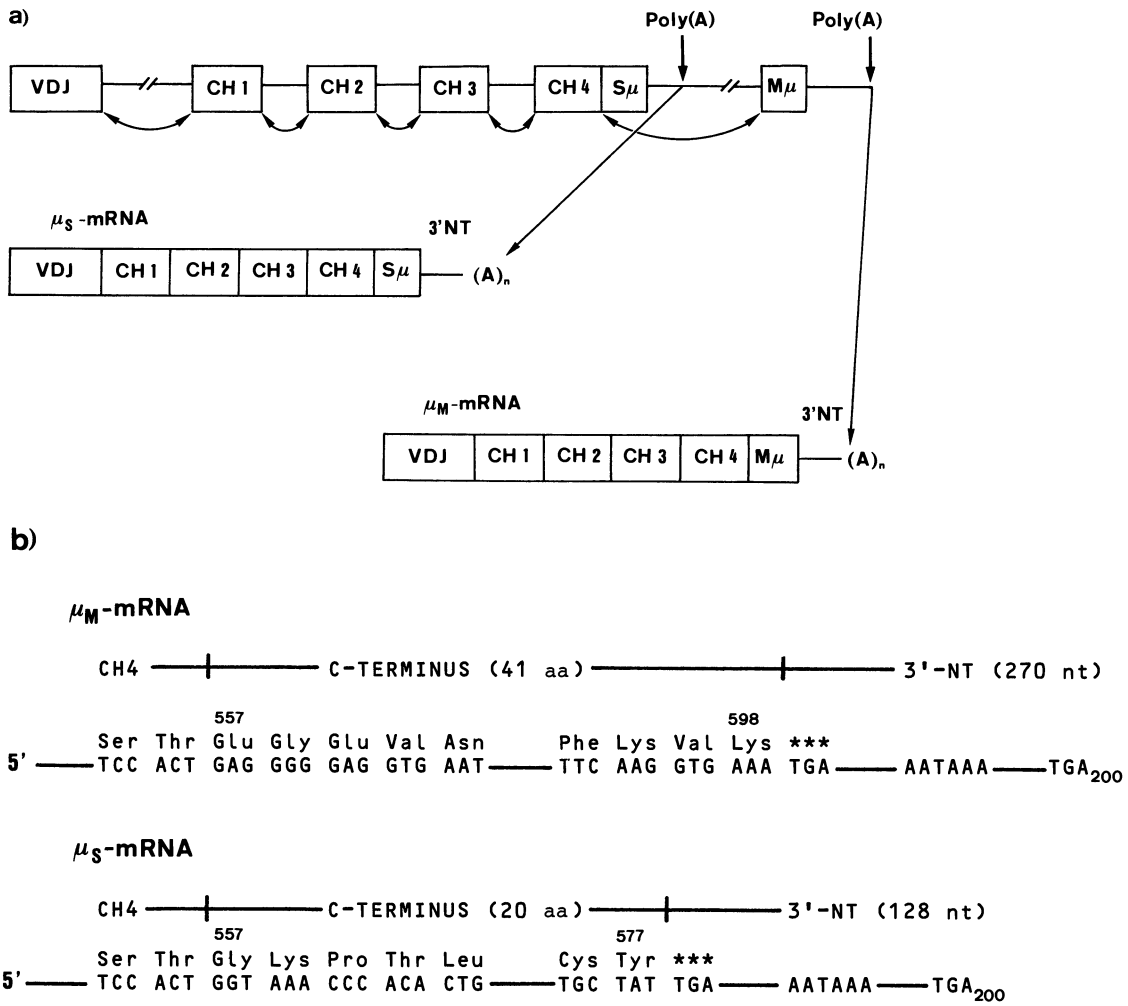


Fig. 6.4a, b. The mRNAs of the heavy chain of membrane-bound IgM (μM) and the secreted IgM (μS) arise from the same primary transcript. **a** The splicing scheme. **b** The nucleotide sequence and corresponding amino acid sequence in the C-terminal region of human μM and μS [194]

The IgG molecule corresponds to one of the basic structural units described above; the IgM in solution in the plasma consists of five units together with a 15-kDa component (JC); the soluble sIgA is made up of two types of unit, JC and a 60- to 70-kDa secretory component (SC), either singly or in pairs (Table 6.1). The plant protease papain cleaves the Ig molecule in the H region to give two Fab fragments, each consisting of the light chain together with the V region and CH1 domain of the heavy chain, and one Fc fragment, which consists of the other two CH domains. Because Fab fragments still have the capacity to bind antigen, they are a frequently used tool of immunology.

6.2.2 The Variety of Immunoglobulins

Each basic structural unit carries two **antigen-binding sites** which complement particular molecular structures of the antigen. These antigen structures, known as antigen determinants or epitopes, involve just a few (5–7) amino acids. The antigen-binding site consists of 10–12 amino acids in the V region of the light and heavy chains and these are brought together by folding of the L and H chains. The amino acids of the antigen-binding site are distributed over three particularly variable regions of the V domain sequence known as complementarity-determining regions (CDRs) or hypervariable regions. The **variety of immunoglobulins** in the mammals is unique; it has been estimated that each individual human or mouse can produce millions of antibodies with different

structures and antigen specificity. The hypervariability of the CDRs is related to their extremely high rates of evolution. Substitution rates of $3.14 \cdot 10^{-9}$, $3.72 \cdot 10^{-9}$ and $7.47 \cdot 10^{-9}$ per nucleotide per year have been calculated for the three codon positions of human and murine CDRs; thus in this case there is no effective selection pressure [58]. The sections of the V region outside of the hypervariable CDRs are much less variable and evolve so slowly that common molecular structures may be detected immunologically even when the V_H chains of agnathans or cartilaginous fish and the mammals are compared [153]. One of the V_H genes of the clawed frog *Xenopus laevis* is identical in 61 % of its nucleotides to the V-T15 gene of the mouse [199]. The different H and L chains can be freely combined; thus, a million different immunoglobulins would result from a thousand each of H and L variants. The actual total number of V sequences in mammals is unknown but could well be of the above order; there is probably considerable fluctuation according to chain type and species.

The large variety of immunoglobulins has its origin in the construction of the coding sequences of the V regions from two or three gene segments during lymphocyte development (somatic rearrangement): the heavy chains arise from V_H (variable) with 95–100 amino acid codons, D (diversity) with 13–17 codons, and J_H (joining) with just a few codons; several (in the case of D and J_H) or many (V_H) variants of these segment types exist. The V regions of the light chains are also made up from V_L and J_L segments. The **gene rearrangement** involves the recognition of recombination signals which are conservative nonamer and heptamer sequences flanking each V, J and D segment on both sides. The sequences lying in front of the J_L segments and behind the V_L segments are complementary and can recombine. The same is true for D/V_H and J_H/D. The variety of the H and L genes is further increased by variability in the regions joining the gene segments and by somatic point mutations (nucleotide substitutions) which may occur in the genes themselves [10, 74].

The following terms are used to describe the diversity of antibodies: the variants present in all individuals of a species, e.g. the Ig classes and subclasses, are known as **isotypes. Allotypes** are allelic variants which differ particularly in the C region of the heavy chain. **Idiotypes** show differences in the structure of the antigen-binding site, i.e. in the hypervariable segment of the V region. The general number of gene segments available for combination is known for only a few species (Fig. 6.3): the λ locus of the mouse contains two V_λ segments, each of which is coupled to two J_λ and two C_λ segments, whereas the rat λ locus has only one V_λ and two C_λ genes [177]. The human λ locus is considerably more complicated. The κ locus in man, as in the mouse, consists of many V_κ, five J_κ and one C_κ; the murine H locus is made up of about 1000 V_H, 15 D, 5 J_H and 8 C_H, whereas that of man has at least 200 V_H, more than 20 D, 6 J_H and 9 functional C_H [25]. Multiple C genes of the same type define different Ig subclasses (isotypes), e.g. the 4 C_γ genes of the mouse and man, the two human C_α genes (Fig. 6.3b and Table 6.1), or the 13 C_α genes of the rabbit [22].

Each B lymphocyte expresses only one H and one L gene at any one time **(allelic exclusion);** this is of biological advantage in that simultaneous expression of further loci would lead to a reduction in the density on the cell surface of receptors with a particular specificity. The V sequence combination arising during lymphocyte development is linked to a C sequence to give a complete H or L gene. In the case of an H gene, C_μ initially lies nearest to J_H; the lymphocyte thus produces an IgM which is specific for a particular antigen. Stimulation by this antigen causes the B lymphocytes to proliferate and to differentiate into Ig-secreting plasma cells. The secretory form of Ig is now produced instead of the membrane-bound form, and this requires a gene rearrangement in the region of the H chain C-terminus (Fig. 6.4). The first cells to be formed still produce IgM; the specific V_H sequence is later translocated to one of the C_γ or C_α sequences (class switch) and IgG or IgA is formed. The **class switch** involves recombination in the area of the S regions which lies in front of all C_H genes except C_δ (Fig. 6.3a). The S regions are built up of repetitive 49-bp elements and, for example in the mouse, are all very similar. Hybridization experiments have shown comparable sequences not only in humans and the clawed frog *Xenopus laevis* but also in *Drosophila* and sea urchins, suggesting that they originally had other functions. Rearrangements, gene switching and allelic exclusion are all molecular biological specialities of the immune system [74].

6.2.3 Comparative Biochemistry of Immunoglobulins

Most information about the immunoglobulins and their genes comes from studies of man and

the laboratory mouse, but additional investigations have been carried out on other mammals. All mammals appear to possess the five Ig classes described, and IgG predominates in the blood plasma even in the monotremes. The proportions of the two L-chain types $\varkappa$ and λ, which in man are about 60:40, vary in the mammals over a very wide range, from 95 % $\varkappa$ in the mouse to more than 95 % λ in the horse. Many mammals have Ig subclasses and Ig isotypes that suggest the existence of multiple C_H or C_L genes; for example, subclasses of IgG are detectable even in the marsupials [110].

In birds there are three, or perhaps four, **Ig classes**, and in the remaining vertebrates there are at the most two, which fall into two groups, high molecular weight (HMW) and low molecular weight (LMW), according to their molecular size. They are only partially comparable with the Ig classes of the mammals. The HMWs found in all non-mammals have sedimentation coefficients of 16–19S and molecular masses of 720–920 kDa and are very similar to IgM. The HMWs contain H chains of 70–72 kDa and L chains of 23–25 kDa and are therefore polymers of the basic Ig structural unit. Their structures can be described by the formula $(L_2H_2)_n$, where n may have a value of 2, 4 or 5. Different degrees of polymerization are found even in closely related species. Thus, the ray *Dasyatis centrura* has an HMW with n = 2, but in *D. americana* n = 5. HMWs with n = 4, corresponding to about 720 kDa, are widely found in the bony fish, e.g. in *Lepisosteus, Polyodon*, and many teleosts. The teleost *Archosargus probatocephalus* has two HMWs with n = 2 and n = 4 [138, 154]. The chains of some immunoglobulins, e.g. the HMW of the sea lamprey *Petromyzon marinus* and the LMW of the frog *Rana catesbeiana*, can be separated without previous reduction and are therefore apparently not linked by disulphide bridges.

The agnathans, cartilaginous fish and bony fish usually contain only the HMW form. Some fish, however, also possess LMW. In the case of many shark species, these LMWs contain H chains of the same size and immunological character as those of the HMW and are therefore, perhaps, HMW monomers [154]. The ray *Raja kenojei* and the other skates and rays may be considered as the lowest vertebrates to have two well-defined Ig classes. *R. kenojei* and other Rajidae possess pentameric HMWs with chains of 70 and 23 kDa as well as dimeric LMWs with chains of 45–50 kDa and 23 kDa [67]. The sharks *Carcharhinus plumbeus* and *Heterodontus francisci* have several C_L gene segments; the encoded amino acid sequences are more similar (40 %) to the λ chains of mammals than to the $\varkappa$ chains [160, 165]. Well-characterized immunoglobulins of low molecular mass are found in the anurans; the axolotl, a representative of the urodelans, has only HMW. The LMWs of the anurans, reptiles and birds are more similar to the mammalian IgA than to IgG; they are often referred to as IgY [133]. In the chicken, but also in the anuran *Xenopus laevis*, there are three different C_H gene segments [3, 133].

The **Ig genes** of all vertebrates from the agnathan and cartilaginous fish onwards are apparently formed by combination of V, (D), J and C segments. In various teleosts and the clawed frog *Xenopus laevis*, the variable Ig regions are encoded, as in the mammals, by tandemly arranged multiple V_H, D and J gene segments [4, 163]. Both membrane-bound and secretory IgMs have been identified in the teleost *Ictalurus punctatus;* however, μ_M is markedly smaller than μ_S as a result of deviant splicing [198]. In Elasmobranchii, such as the horned shark *Heterodontos francisci* or the ray *Raja erinacea*, the Ig segments are arranged in a manner very different from that in mammals, anurans and teleosts. In this case, the Ig locus consists of many tandemly arranged clusters with the structure V_H-D_1-D_2-J_H-C_H. The free combination of single gene segments found in the mammalian system is thus excluded and each variable segment V_H-D_1-D_2-J_H is apparently assigned to a certain C_H. The sequences of the numerous C_H segments are very varied [67, 90]. If the birds are also taken into account, then two quite different strategies for creating antibody diversity are recognizable amongst the vertebrates. In the Elasmobranchii, as in the mammals, there is a large repertoire of V genes available; at the λ locus of the chicken, however, there is only one single functional V_λ gene in addition to a J_λ gene, but this is diversified by gene conversion with about 25 available V_λ pseudogenes. In addition to this, the variety of L chains in the chicken, as in the mammals, is increased by somatic mutation and imprecise V_L/J_L linkage. As a result of this process about 10^6 different antibodies are formed; this is only one order of magnitude less than in mammals [133]. The **variety of immunoglobulins** in the lower vertebrates is much less than that found in the mammals.

Following immunization with a low molecular weight hapten like dinitrophenol, mice produce about 500 different antibodies which are specific for this antigen but are electrophoretically distinguishable (spectrotypes); the pattern varies

from individual to individual within an inbred line. In the anurans, one finds at the most 40 spectrotypes and in carp there are only 23 spectrotypes. Different individuals of the shark *Heterodontus francisci* produce essentially identical antibodies in response to the same hapten [147]. The significantly lower Ig variety, e.g. in the clawed frog *Xenopus*, is particularly puzzling because here the gene organization and the number of V_H, D and J_H segments is similar to that of the mammals [163].

6.2.4 Evolution of the Immunoglobulins

The direct determination of the **amino acid sequences of the immunoglobulins** is hampered by their diversity. Earlier, use was made of the production of immunoglobulins with a uniform V sequence by particular clonal tumours (myeloma proteins); the resulting excess of L chains appears in the urine (Bence-Jones proteins) [179]. Large quantities of uniform (monoclonal) antibodies can also be obtained by experimental selection of lymph cell lines. More recently, the indirect determination of amino acid sequences by DNA analysis has made it possible to obtain a wealth of immunoglobulin sequence data. However, the sequences obtained are almost exclusively from man, the mouse and a few other mammals; sequence data for lower vertebrates, which would be essential for the analysis of relationships and for the construction of genealogical trees, are as yet in short supply [101].

The multiple segments of the Ig genes arose by **duplications**. This can be demonstrated, for example, with the C_H genes (Fig. 6.3b). Whereas in the mouse the four C_γ genes form a tandem cluster, in man it appears that the cluster C_γ-C_γ-C_ε-C_α was duplicated as a whole [19]. The duplication of Ig genes in the germline is apparently a rare evolutionary event; the frequency of $5 \cdot 10^{-7}$ per gene per year for the Ig genes is about 100-fold less than for the rDNA. This reduced rate of horizontal evolution helps maintain the genetic variability of the Ig segments but, on the other hand, it allows the formation of pseudogenes [58]. Almost all the known Ig pseudogenes show relatively few defects when compared with other pseudogenes. Thus, they probably arose quite recently from active genes and have been constantly corrected by horizontal evolution; they can, therefore, be easily reactivated or used in part for the construction of new gene sequences [74]. Gene duplications or deletions are apparently very frequent somatic mutations; 1–3 % of human individuals are said to be heterozygous for the deletion of single C_H genes. A consideration of the degree of similarity indicates different subgroups within the multiple V sequences. The V_H subgroup V_HIII, initially defined in man and the mouse, was easy to sequence because of its free N-terminus and it has been especially well characterized. V_HIII-like sequences are already to be found in the shark *Ginglymostoma cirratum* and are apparently quite old. Such multi-gene subgroups can either expand or contract during evolution and may differ in extent, as for example between man and the mouse.

The gradient of similarity between the V and C domains of the different H and L chains suggests the following model for **immunoglobulin evolution**. Duplication of an ancestral Ig gene led at first to V and C and the joining system with the J segment. This was followed by duplication of the C gene to give the C_L and C_H genes. The original C_H region with four domains then arose by internal duplication. From this emerged the different C_H types, of which C_γ, C_δ, C_α lost all but the H (hinge) section of the CH2 domain. In agreement with this scheme, for example, is the greater similarity between the homologous domains of μ, γ^1 and γ^{2b} than between the different C domains of μ in the mouse. Most of the H chains carry terminal extensions of unknown biological function which are not entirely homologous to each other. The highest substitution rate is shown by the hypervariable sections of the V region. It is also clear from sequence comparisons of homologous C domains of different mammalian species that the rate of evolution of the immunoglobulin C regions is consistently high, despite large differences between the various chain types and individual C domains (see Table 4.12, p. 161). The rate of evolution of λ is higher than that of $\varkappa$; in most of the H chains investigated, the rate of evolution decreases towards the C-terminal domains. The evolution of the H region (hinge) appears to be particularly rapid. The immunoglobulins (HMW and LMW) of many anuran species show immunological cross-reactivity with those of *Xenopus laevis*, despite the fact that some of these species separated more than 150 million years ago. In this case, the evolution of at least some parts of the immunoglobulins appears to be progressing more slowly than in the mammals.

6.3 T Cell Receptors

T cells recognize antigens only when they are presented on other cells together with an MHC antigen (MHC restriction); cytotoxic T cells are specific for the class I MHC products which are found on most cells of the body, and T helper cells are specific for the class II MHC products which are produced by the APCs. Specific glycoproteins present on the surface of the T cell are apparently responsible for this **MHC restriction:** CD4 for the MHC class II-specific T cells, and the heterodimeric CD8 for the class I-specific cells; both belong to the immunoglobulin super-family [127]. The antigen-specific receptor of the T cell is a heterodimer of the subunits α and β. A minor proportion of the cells express receptor dimers from two other chains, γ and δ. A third class of T cell receptor is found in the chicken [32]. The αβ or γδ receptor dimers are associated with the CD3 complex, which is composed of four or five different types of subunit according to the formula γδεεξξ or γδεεξη. The ε subunit is present twice; ξ is present either as a homodimer or as a heterodimer with the homologous η chain [16]. With the exception of ξ and η, all polypeptides of the τ cell receptor complex belong to the Ig superfamily [196]. The α and β subunits are both about 40–50 kDa and have an N-terminal, variable V domain and a C-terminal, relatively conserved C domain; each domain corresponds to one Ig homology unit (Fig. 6.2). The variable V domains of the α and β chains together make up the antigen-binding site.

The four receptor subunits α–δ in man and in the mouse are encoded by gene families which, like the immunoglobulins, include a large number of tandemly arranged V, J and C gene segments; the β and δ families also contain D segments. Furthermore, as for the immunoglobulins, the complete gene arises by somatic rearrangements of V-(D)-J-C. The gene segments are flanked by nonamer and heptamer sequences which are almost completely identical with those of the immunoglobulins [38]. The Vβ domain consists of about 109 amino acids. The murine Vβ genes can be divided into several subfamilies of one to three members, and there are also many subfamilies of human Vβ genes. The sequence similarity of members of the same subfamily in different species, e.g. man and the mouse, is often greater than that between the subfamilies within the same species, i.e. the subfamilies are apparently very old [74, 128, 186]. The two murine Cβ genes are each made up of four exons; the first codes for the extracellular, Ig-homologous domain of 125 amino acids, and the others for the hinge sequence with 6, the transmembrane region with 36, and the cytoplasmic tail with 6 amino acids. The two Cβ sequences differ in only 5 of their 173 amino acids and therefore arose quite recently by gene duplication [74]. Two Cβ genes are found also in man and several rabbit varieties; other rabbits possess three Cβ genes [104]. The only T cell receptor subunit of a non-mammal to be sequenced (in this case via the cDNA) is the β chain of the chicken. Although there is only about 31 % agreement in the amino acid sequence, a whole series of conserved structural features are recognizable [188].

Thy-1 also belongs to the immunoglobulin super-family and is involved with the differentiation of thymocytes to mature T lymphocytes. Many more Thy-1 molecules are present on mouse and rat than on canine thymocytes; they are not present on human thymocytes or on mature rat T lymphocytes, but in all species they are found on neuronal brain cells, fibroblasts and some other cell types. In the mouse and rat, the molecule is a glycoprotein with a polypeptide chain of 111 amino acids, the molecular mass of which is increased from 12.5 to 17.5 kDa in the brain and to 18.7 kDa in the thymus by the addition of three N-linked carbohydrate chains. The Thy-1 C-terminus in the rat brain is attached to the cell membrane via glycosyl-phosphatidylinositol. This method of anchoring the membrane proteins is found widely in unicellular eukaryotes but is very rare in the vertebrates [73]. Thy-1 has the typical three-dimensional structure of an Ig homology unit and agrees by 36 % with the amino acid sequence of the murine V_L chain [109]. Proteins resembling Thy-1 have also been found, for example, in lower chordates (ascidians), insects, annelids and molluscs [17, 36, 109, 158].

6.4 MHC Antigens and β_2-Microglobulins

The proteins encoded by the major histocompatibility complex (MHC) may be divided into three **classes** according to their structure and function. MHC antigens of classes I and II are integral membrane proteins which are responsible not only for transplant rejection but also for the general differentiation between "self" and "non-self" by the immune system. As described above, the T lymphocytes recognize the complementary for-

eign antigen only when accompanied by an MHC antigen; however, the mechanism of this MHC restriction has not yet been defined. The class III MHC genes encode components of the complement system and will be discussed in Section 6.6.

The best-characterized MHCs are the murine H-2 complex and the human HLA complex [50, 87]. The H-2 complex on mouse chromosome 7 extends over 4000 kb. The classical transplantation antigens of **class I** are encoded in loci K, D and L. All the genes of this group so far sequenced contain eight exons corresponding to the subdivisions of the polypeptide; exon 1 encodes the signal peptide, exons 2–4 encode the three extracellular domains α1, α2 and α3, exon 5 encodes the transmembrane region and exons 6–8 code for the small cytoplasmic domain. The complete polypeptide of 45 kDa, anchored in the cell membrane, is bound non-covalently via domain α3 to the 12-kDa large $\beta_2$microglobulin (β_2m). The domain α3 and the β_2m each correspond to one Ig homology unit (Fig. 6.2). Class I also includes the Qa antigen, which is found predominantly on B and T lymphocytes, and the Tla antigen, which is found on the thymocytes; these two molecules are much more polymorphic than the transplantation antigen and are not involved in the MHC-restricted recognition by T lymphocytes. In all there are 30–40 class I genes in the mouse, most of which can be allotted to the Qa and Tla regions. Many of these are pseudogenes. The number of class I genes varies between inbred lines of laboratory mice, probably because of frequent unequal crossing-over [15, 50]. Of the 65 MHC genes detected in the blind mouse *Spalax ehrenbergi* only three are active [193]. Further class I genes, other than the transplantation antigen in the H-2 complex, are apparently absent from some rodent species and are clearly lacking in other mammals. For example, DNA sequences corresponding to the Tla gene of the laboratory mouse are detected in the rat and the golden hamster but not in the guinea-pig [152]. The human HLA complex consists of more than 20 class I genes, many of them pseudogenes. The functional transplantation antigens of man have been designated A, B and C.

The Ia antigens of **class II** are expressed primarily on B lymphocytes and macrophages [77, 173]. These are heterodimers of two membrane-anchored, non-covalently bound chains, a 35-kDa α chain and a 29-kDa β chain, each of which has two extracellular domains. The domains α2 and β2 have the structure of Ig homology units (Fig. 6.2). As in the case of class I genes, the exon boundaries correspond to domain boundaries in the proteins. The class II region of human MHC consists of the subregions HLA-DP, -DZ/DO, -DQ and -DR. Each subregion contains α and β genes, except DZ, which contains only α, and DO, which has only β; in all regions there are both functional genes and pseudogenes [87]. The mouse also possesses loci for all four subclasses, although fewer than in man. Only DP and DQ genes have been detected in the blind mouse *Spalax ehrenbergi* and none of the subclasses DZ/DO and DR [77].

Like the transplantation antigens of class I, the class II antigens also show extreme **polymorphism**. Between 50 and 100 alleles of both classes may be present in a single mouse population. The alleles can differ by up to 50 amino acids, mainly localized in the outer domains, i.e. α1 and α2 of the transplantation antigens, and α1 and β1 of the class II antigens. The mechanism leading to this extreme polymorphism is not completely clear. Sequence comparisons of man, the chimpanzee and the mouse show that many of the MHC alleles are very old; this would speak against an increase in mutation rate. On the other hand, gene conversion has certainly played a role in the creation of MHC polymorphism, in addition to overdominant selection due to a heterozygote advantage [47, 77, 95, 99]. MHC-antigen polymorphism is important for the survival of the species, in so far as the combination of foreign antigen and MHC antigen may correspond by chance to a particular "self" structure; in such a case there will be no immune response and the receptor repertoire of the T lymphocyte will have a "blind spot". The high polymorphism ensures that this danger does not affect the whole population [78, 87]. In the tamarin *Sanguinus oedipus* (a primate) there are only a few class I MHC alleles and this may well be the cause of the observed high infectivity of this species [195].

Little can be said at present about the **evolution of the MHC** because of the absence of detailed investigations of the lower vertebrates. Nevertheless, it can be taken as certain that all vertebrates possess an MHC. The molecularly well-characterized MHC complex of the chicken (B complex) has significantly smaller introns, and is therefore smaller than both human and murine MHC; it also differs by having much closer linkage of the class I and II genes with each other and with other genes [63]. The surface of the leukocytes of the frog *Xenopus laevis* bears both class I antigen molecules, made up of a polymorphic chain of about 45 kDa and a β_2m-like

chain of 13 kDa, and class II molecules, consisting of two non-covalently bound glycosylated chains of 32–34 kDa. Class II genes, but not those of class I, are expressed in the tadpole. Gene sequences for both classes of MHC antigen have been identified in the carp *Cyprinus carpio* [49, 69, 110]. There is even some evidence for the existence of an MHC in a tunicate, the ascidian *Botryllus* sp. [49].

β_2m is found both free in the body fluids and associated with the MHC antigen on the cell surface. It is a non-glycosylated protein of 12–13 kDa which shows up to 40 % difference in pairwise comparisons of its sequence of 99–100 amino acids between humans and various other mammals [61]. β_2m is ubiquitous in the vertebrates but has rarely been examined in the lower vertebrates [36, 61]. β_2m-like proteins have also been detected immunologically on haemocytes of the earthworm *Lumbricus terrestris*, on tissue culture cells of *Drosophila melanogaster*, and on other invertebrate cells [17]. MHC molecules of class I in the membranes of vertebrate cells occur mostly bound to β_2m, but can also serve as the anchor for other proteins. For example, the tetrameric insulin receptor in human cell cultures has been shown to be associated with class I molecules. Thus, the range of functions for MHC molecules has expanded in unexpected directions [43].

6.5 Cytokines and Interferons

The complex interactions between B lymphocytes, T cells and macrophages involve the activity of regulatory polypeptides which have become known as "lymphokines" or "interleukins". It soon became apparent that the same, or similar, polypeptides function not only as regulators of the immune system but also as hormone-like growth and differentiation factors. They are involved, for example, in the control of haematopoiesis and the conversion of the pluripotent stem cells of the bone marrow into blood cells, such as erythrocytes, granulocytes, monocytes and lymphocytes, and into certain tissue cells, like macrophages, mast cells and dendritic cells. They control inflammation processes in that they react chemotactically with neutrophils or macrophages, and they stimulate hepatocytes to produce acute-phase proteins. In view of these multiple activities, the general term "cytokines" was introduced. Cells stimulated by cytokines in turn produce further mediators; a cytokine network serves to connect functionally lymphocytes, macrophages and haematopoietic cells. About 50 cytokines are presently known. They include, for example, the interleukins, of which eight have been described (IL-1 to IL-8), although some authors count 11 [56, 134], interferon IFN-γ, growth factors such as epidermal growth factor (EGF) and PDGF, colony-stimulating factors (CSFs), tumour necrosis factors and erythropoietin [6, 39, 114]. The cytokinins function by binding to specific receptors on the cell surface; these may be grouped into families on the basis of structural similarity. The IL-1 and PDGF receptors belong to the Ig super-family. The receptors for Il-2, -3, -4, -6 and -7, various CSFs and erythropoietin, as well as growth hormone and prolactin all possess corresponding modules of 200 amino acids with four cysteine residues in the N-terminal half and the sequence WSXWS close to the C-terminus. The binding domains of the receptors for interferons α–γ and the receptor for the clotting factor VII (known as tissue factor) have the same structure, a sequence of 210 amino acids with characteristic cysteine pairs at both the C- and the N-terminii. The latter two receptor families show similarities to the Ig homology units in their tertiary structure, despite having different amino acid sequences [7].

Interleukin-1 (IL-1) was discovered as a product of activated macrophages, but is also formed by other cells such as fibroblasts, keratinocytes and B cells. It is actually a whole family of proteins which are involved in various reactions to injury and infection: thymocyte proliferation, B cell activation, multiplication of the neutrophilic granulocytes, synthesis of acute-phase proteins in the liver, and triggering of fever [6]. Il-1 activities have been detected in all classes of vertebrates including fish; they are usually bound to proteins of 12–20 kDa but sometimes to larger proteins of 35–70 kDa. IL-1-like proteins have also been detected in the starfish *Asterias forbesi* and various ascidians [9]. There are two different proteins of this family in man and other mammals (IL-1α and IL-1β); they have identical activities and bind to the same receptors but agree by only 26 % in their sequences. Both arise from precursors of about 270 amino acids by cleavage of the N-terminal 112–116 amino acids. The cDNA or gene sequences are known for several species; the common gene structure of seven exons indicates that IL-1α and -1β, despite the poor similarity of their amino acid sequences, arose by duplication from a common ancestral gene [6].

Interleukin-2 (IL-2) is secreted by T helper cells in cooperation with foreign antigens and IL-1; it induces the proliferation of T cells and their production of γ-interferon. IL-2 has so far not been unambiguously detected in birds and lower vertebrates and shows large variation in immunological and functional characters amongst the mammals. Human and bovine pre-IL-2, including the 20 residues of the signal peptide, have lengths of 153 and 155 amino acids, respectively; in the rat, there are 169 amino acids, including a unique sequence of 12 consecutive glutamines. There is one N-linked carbohydrate chain in the bovine IL-2 and two in the human and rat forms. The sequences of IL-2 agree in these three species in 60–65 % of positions [31]. **Interleukin-3** (IL-3) is mainly produced by antigen-stimulated T cells and stimulates the formation of all the blood cell types originating in the stem cells of the bone marrow. Rat IL-3 differs from the murine form in 46 % of its amino acids and has only a limited effect on murine bone marrow cells. Surprisingly, the sequence agreement in the flanking NT regions (90 %) and the four introns (80 %) is greater than in the exons (76 %) of these species [35]. There are hardly any comparative data available for the other interleukins. The α_1-**microglobulin** inhibits antigen stimulation of lymphocytes. It is a glycoprotein which in various primates has a length of 198 amino acids and a molecular mass of 31–32 kDa; in other mammals (mouse, rat, guinea-pig, horse) and in the chicken, the molecular weight is only 24–26 kDa [1, 190].

Virus infection of almost all vertebrate cells induces the production and secretion of **interferons** (IFN). These bind to high-affinity receptors of other cells and enhance resistance to the virus by an as yet unknown mechanism. The interferons have immunoregulatory functions in addition to their antiviral role: they stimulate the cytotoxic activity of T lymphocytes, macrophages and the so-called natural killer cells, which lyse foreign and transformed self cells but do not belong to the T- or B-lymphocyte categories. All interferons activate the expression of class I MHC genes and the β_2-microglobulins; one particular type of interferon (IFN-γ) also stimulates the expression of class II MHC genes [136].

Three families of IFNs can be distinguished in the mammals: the leukocyte IFN or IFN-α, the fibroblast IFN or IFN-β, and the immune IFN or IFN-γ. IFN-α and IFN-β are very similar and are therefore classified as type I; IFN-γ is classified as type II. **IFN**-α has a molecular weight of 16–27 kDa and in man usually consists of 165–166 amino acids. In addition, there is a further type (IFN-ω, -α_L or -αII) in man and some other mammals with a length of 172 amino acids. There are 23 IFN-α genes on human chromosome 9, of which 14 can be expressed. Only one of the 14 active genes encodes an IFN with an asparagine that can be glycosylated; the other IFNs are, in all cases, O-glycosylated. The IFN-α molecules encoded by the different genes differ in about 20 % of positions. In addition, the heterogeneity of IFN-α is increased by post-translational cleavage of C-terminal amino acids by partial proteolysis. IFN-α multi-gene families are also found in the mouse, rat and cow. Pairwise sequence comparisons between these species show agreement of about 60 % [136].

The **IFN-β gene** is also located on human chromosome 9. The encoded sequence of 166 amino acids shows only about 33 % agreement with IFN-α2. Nevertheless, IFN-α and -β compete for the same receptors on target cells. Surprisingly, the activity remains unchanged when the 62-Glu found in all known human, murine and bovine IFN-α and -β is exchanged for lysine [191]. Whilst humans, like most mammals, have only one IFN-β gene, the lion and the rabbit have two genes, and cattle, horses and pigs have many. There are IFN-β-like sequences on other human chromosomes. However, the so-called IFN-β2 gene is not significantly homologous to IFN-β and its product has little antiviral activity; apparently, it does not belong to the IFN-β family. Both IFN-β and IFN-α genes are always without introns. In contrast, the gene for **IFN**-γ on chromosome 12 has three introns. The sequence of 146 amino acids shows no significant homology to either IFN-α or -β, and IFN-γ binds to its own specific receptor. The carbohydrate-free chain has a molecular mass of 17.1 kDa, and the native molecular mass is 20 or 25 kDa, depending on whether only the 28-Asn or also the 100-Asn is glycosylated [21].

The three IFN families are found in all mammals. In other vertebrates, proteins similar to IFN-α and -β but not to IFN-γ have been detected. The sequence similarity between IFN-α and IFN-β indicates that the gene duplication that separated the two families occurred about 300 million years ago, i.e. before the separation of the reptiles, birds and mammals. Accordingly, the fish and amphibians should possess only one IFN family [136].

6.6 The Complement System

With its 20 or so proteins, the complement system plays an effective part in the defence against microorganisms by causing partial destruction of the cell membrane and, thereby, cell lysis. The **activation of the complement system** can occur in two reaction cascades, the "classical" and the "alternative", in each step of which a zymogen is proteolytically converted to an active protease. The enzymatically active C-terminal fragment is designated by addition of the suffix "b" to the complement component number, and the cleaved N-terminal partial peptide by addition of the suffix "a". The specificity of complement action is maintained through the triggering of the activation cascade, either via the classical pathway by antigen-bound immunoglobulin or via the alternative pathway by the lipopolysaccharides of the bacterial cell wall. The classical pathway results in amplification of the immune reaction, and the alternative pathway can be initiated before the immune reaction. The complement system is remarkable in that its lytic activity requires the association of five different proteins [98, 120].

Three phases can be distinguished in the **classical pathway** and these occur at different sites on the membrane. In the first phase, a recognition complex is formed on antigen-bound antibodies by the accumulation and proteolytic activation of the components C1q, C1r and C1s. Only certain immunoglobulin classes or subclasses are able to bind complement: in man these are IgG1, IgG3 and IgM; in the mouse IgG2a and IgM; in the guinea-pig IgG2; and in ruminants IgG1 [98, 120]. The C1 complex contains one molecule of the hexameric C1q and two molecules each of C1r and C1s. The C1 complex is activated by interaction with the Fc regions of several Ig molecules; in the process a bond is cleaved at about one-third of the length in each of the zymogens C1r and C1s. The smaller N-terminal cleavage fragment remains bound to the enzymatically active C-terminal fragment by a disulphide bridge [103]. The second phase is initiated when the activated C1s produces the fragments C4b and C2b from C4 and C2; these become enzymatically active (C3 convertase) after binding to a neighbouring binding site on the membrane. The cleavage fragment C3b, produced from C3, binds to the C4b,2b complex and modifies its activity to that of a C5 convertase. The fragment C5b, produced from C5, accumulates at a third membrane site and becomes associated with one molecule each of C6, C7 and C8. Binding of several (up to 16) C9 molecules results in the formation of a channel structure which perforates the membrane [65].

In the **alternative pathway**, the C5 convertase is formed via a feedback loop involving bound C3b, factor B, factor D and free C3. C3b, bound to lipopolysaccharide of the bacterial cell wall or another activator, forms an inactive C3b,B complex with factor B. Cleavage of B by factor D results in the complex C3b,Bb, which as an active C3 convertase produces more C3b. Finally, a further C3b binds to the C3b,Bb complex; the resulting C3b,Bb,C3b complex is the C5 convertase of the alternative pathway. This produces C5b and thereby allows the formation of the membrane-attacking complex [98, 120]. Because of its C3 convertase activity, the C3b,Bb complex represents a positive feedback system which is regulated by factors H, I and P (properdin). Factor H binds C3b, displacing the Bb subunit and making C3b accessible for proteolytic inactivation by factor I. C1Inh and the C4b-binding protein function as inhibitors of the classical pathway; the former inhibits interaction of C1 with the immune complexes. The decay-accelerating factor (DAF), an inhibitor in both pathways, is a membrane protein found on almost all cells of the body, and together with factor I it proteolytically inactivates C3b and C4b [70, 148].

The **proteins of the complement system** are well known in man and several rodents (Table 6.2). Many have been sequenced via their cDNA, and therefore conclusions can be drawn about their relationships to each other and to other proteins. The subunits of the hexameric C1q molecule are made up in each case of three proline-rich polypeptide chains; the subunits together form a stem with a collagen-like structure and a head, which interacts with the Fc region of immunoglobulins. C1r, C1s, C2, C6, factor B and factor D are serine proteases. C1r and C1s are very similar in their domain structures and agree in about 40 % of their amino acids. The domains I and III are apparently the products of a gene duplication; domain II corresponds to the sequence of the EGF; domains IV and V resemble a repeated sequence of about 60 amino acids which has also been detected in other complement components and other proteins, e.g. C2, factor H and its homologues CR1, C4b-binding protein and DAF, factor B, the β_2-glycoprotein I and the IL-2 receptor [148]. The C-terminal catalytic domains of C1r and C1s are typical serine proteases [103, 189]. The sequences of C2 (732

Table 6.2. The components of the human complement system. Given is the concentration in the blood plasma in µg/ml, the molecular mass in kDa and the number of polypeptide chains per molecule

Classical pathway				Alternative pathway			
	µg/ml	kDa	Chains		µg/ml	kDa	Chains
C1q	75	410	6 × 3	C/3	1500	180	2
C1r	34	190	2	B	200	93	1
C1s	30	87	1	D	1	24	1
C2	25	115	1	P	20	220	3
C3	1500	180	2	H	470	150	1
C4	450	210	3	I	34	88	2
C5	75	190	2				
C6	60	128	1				
C7	60	121	1				
C8	80	163	3				
C9	58	79	1				

amino acids) and factor B agree by 39 %; they have similarities with trypsin (17 %) and other serine proteases in the C-terminal catalytic half (see Fig. 3.6, p. 91), and at the N-terminus they contain three of the 60-amino-acid tandem repeats mentioned previously [11]. Factor H is made up entirely from 20 such tandem repeats [148]. As in the case of the other serine proteases, the evolution of these complement components involved exon shuffling. C6 agrees in more than 30 % of its 913 amino acids with C7 and shows the same distribution of all 56 cysteine residues [65].

C3, C4 and C5 belong to the same protein super-family as the protease inhibitor α_2M [129]. The comparison of partial sequences of C3 from the clawed frog *Xenopus laevis* and various mammals showed not less than 49 % agreement [60]. C8 is a glycoprotein of 150 kDa which consists of three polypeptide chains. The α-chain (64 kDa) and the γ-chain (25 kDa) form a disulphide-linked dimer with which the β-chain (64 kDa) interacts non-covalently. The amino acid sequence of C8β agrees 33 % with C8α and 26 % with C9; all three are members of the same protein super-family [64, 76]. In contrast, the C8γ chain belongs to the α_{2u} protein super-family together with 10 further functionally very different plasma proteins (see Table 4.2, p. 117) [80]. C2, factor B and C4 are encoded by MHC and form class III of the MHC products. All three are highly polymorphic: 13 alleles are known for the human C4 locus C4A and 22 alleles are known for C4B [126].

As expected, the immunological similarity to human complement components gradually decreases in the order: anthropoids, Old World apes, New World apes; the prosimians show no cross-reactivity. **Complement systems functionally similar to that of the mammals** exist in all vertebrates, at least from the cartilaginous fish onwards [91, 111]. On the other hand, the supposed evidence for a complement system in invertebrates should be treated with some scepticism; not every haemolytic molecular species or cascade corresponds to the complement system. However, there exist, for example in the coelom fluid of the sea urchin *Strongylocentrotus droebachiensis*, both lytic activity against antibody-treated rabbit erythrocytes and a phagocytosis-stimulating (opsonic) effect on self coelom cells and mouse macrophages which can be inhibited by specific inhibitors of the complement system [12].

In agnathans, such as *Myxine glutinosa* and *Lampetra japonica*, there are, at most, single components of the alternative pathway. On the other hand, six complement components (C1n, C2n, C3n, C4n, C8n and C9n, where n stands for "nurse shark") isolated from *Ginglymostoma serratum* produce 8-nm holes in sheep erythrocytes; this is similar to the effect of guinea-pig complement. The holes made by human complement are somewhat larger (10–11 nm). The shark components are comparable with those of the mammals, although component C4n also possibly represents mammalian components C5, C6 and C7; the mammalian system thus appears to be a refinement that arose during evolution. Only a few of the shark components may be combined with the mammalian components and this always results in reduced effectiveness; in contrast, the complement components of different mammals can usually substitute for each other without loss of activity [171]. Both activation pathways for the complement system have been detected in several teleosts, such as the tuna *Thunnus alalunga* and the rainbow trout *Salmo gairdneri*, and in amphibians; the alternative pathway appears to be the only possibility in other teleosts and the chicken [125].

6.7 Lectins

Lectins are proteins or glycoproteins that bind specifically to carbohydrates. They are found in all groups of organisms from the viruses to the mammals [166, 167, 200]. A lectin molecule typically consists of several subunits, each with one

sugar-binding site, and it can, therefore, link cells together (agglutination) or precipitate glycoconjugates. The **detection of lectins** is usually based upon the agglutination of vertebrate erythrocytes. Many lectins require Ca^{2+} for the **binding reaction**. Specificity is defined by examining which free or bound sugar competitively inhibits haemagglutination. The specificity is most frequently for L-fucose, N-acetyl-D-galactose, D-galactose, N-acetyl-D-glucose, D-mannose or N-acetyl-D-neuraminic acid (sialic acid). Neuraminic acid-specific lectins are also found in animal groups which contain no sialic acid, and in this case they probably have a bactericidal function [17, 108, 166, 167]. Several invertebrate lectins are specific for 2-keto-3-deoxyoctonate (KDO; Fig. 6.5), which occurs frequently as a sugar component of the bacterial cell surface. In almost all cases, glycosidic, non-reducing sugars are bound more strongly than free sugar molecules; the binding affinity depends also upon the type of glycosidic bond and the extent of the carbohydrate chain branching. Thus, the lectin from the haemolymph of the oyster *Crassostrea gigas* has a 16000-fold greater affinity for sialoglycoprotein from the bovine salivary gland than for free sialic acid. The haemolymph lectin from the cricket *Teleogryllus commodus* has a binding constant of $1.8 \cdot 10^{-7}$ mol/l for the sialo-protein fetuin from calf plasma and 10^{-2} mol/l for free sialic acid. Because the glycoconjugates on the erythrocyte surface show species- and blood group-specific differences, lectins with strong sugar specificity cannot agglutinate all erythrocytes. This is illustrated by achatinin H from the haemolymph of the pulmonate snail *Achatina fulica;* this is highly specific for 9-O-acetylneuraminic acid and can, therefore, agglutinate the erythrocytes of the rabbit, rat and guinea-pig but not those of the horse, which mainly have 4-O-acetylneuraminic acid, or those of man, sheep, goat and chicken, on which only N-acetyl- or N-glycolylneuraminic acid are found [106]. A similar high specificity is shown by the haemolymph lectin from the rock-dwelling crab *Cancer antennarius;* this lectin binds to 4-O- and 9-O- acetylated sialic acid residues [143].

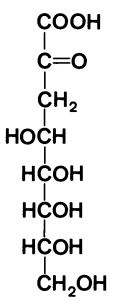


Fig. 6.5. 2-Keto-3-deoxyoctanoate is found in glycoconjugates on the surface of many bacteria

Their high specificity for certain sugar sequences has led to various uses of the lectins as **biochemical reagents**, e.g. for the identification and localization of glycoconjugates on cell surfaces, for the taxonomy of bacteria and protozoa, and for the isolation of glycoproteins, polysaccharides and particular cell populations by affinity chromatography with bound lectins. Largely because of these practical considerations, lectins have been sought, isolated and characterized from many plant and animal species [17, 166, 167, 200]. There have been many investigations in the invertebrates, in particular of the sponges [84], snails and mussels [107, 116, 121, 183], crustaceans [28, 122], insects [30, 81, 86, 94, 135, 141, 146] and ascidians [66, 132, 161, 184].

The **structure of animal lectins** is known rather poorly from the point of view of comparative biochemistry. They are in any case a very heterogeneous group of proteins. The molecular masses themselves present a very non-uniform picture: the values for native proteins vary between about 35 kDa and more than 1 million Da, and the subunits between 12 and more than 50 kDa [12, 200]. A good many lectins have carbohydrate contents of 10–30 %, e.g. the lectins from the eggs of the fish *Coregonus lavaretus* and the clawed frog *Xenopus laevis*, or the agglutinin from the haemolymph of the pulmonate snail *Achatina fulica* [116]. Others have little or no carbohydrate. Lectins exist intracellularly, as integral membrane proteins, or freely dissolved, in either extracellular fluids or mucous secretions [53, 166, 167, 200]. Animal lectins may be subdivided into three classes based on both structural and functional criteria. The C-type requires Ca^{2+} ions, has a carbohydrate-binding domain of about 130 amino acids, and is characterized by 18 conserved positions and two intra-chain disulphide bridges; a typical example is the asialoglycoprotein receptor. The S-type requires no Ca^{2+} in vitro and contains a domain approximately 100 amino acids long with 39 conserved positions which are completely different from those of the C-type; the cysteine residues in this case have free SH groups. Typical examples of the S-type are the β-galactose-specific lectins. Lectins which lack the characteristics of the other two types are designated N-type; these include proteins whose principal functions are other than carbohydrate binding, e.g. the mannose phos-

phate receptor which is responsible for transport of proteins into the lysosomes [42, 100, 175]. The best-characterized membrane-bound forms are galactose-, mannose- and acetylgalactosamine-specific lectins on the hepatocytes of birds and mammals. Membrane-bound lectins have also been described in invertebrates, e.g. on the haemocytes of the oyster *Crassostrea virginica* and other mussels [17]. Detailed concepts of the molecular structure are available for only a few lectins. The galactose-specific, soluble lectins of fish, amphibians, birds and mammals are mostly dimers with subunits of 13–17 kDa; the eggs of fish and amphibians are especially lectin-rich, with more than 1 % of the total soluble protein falling into this category. These **vertebrate lectins** form a significantly homologous group; for example, the lectins of the electric eel *Electrophorus electricus* and chicken skin agree at 39 % of comparable amino acid positions. The lectins have no similarities to other proteins and apparently constitute their own protein super-family [27, 34, 48, 131, 174]. The 111-amino-acid sequence of the sialic acid-specific lectin from the egg of the American bullfrog *Rana catesbeiana* shows no similarity to other vertebrate lectins or to any other protein [191].

The **limulin** molecule from the haemolymph of the xiphosuran *Limulus polyphemus* appears as a hexagonal structure in electron micrographs; each of the six 67-kDa subunits consists of three 22-kDa polypeptide chains whose sequence of 163 amino acids shows no similarity with the known lectins of vertebrates or plants. Only a very few invertebrate lectins have been either fully or partially sequenced. The α-chain of the lectin from the fly *Sarcophaga peregrina* and the lectins from the barnacle *Megabalanus rosea*, the sea urchin *Anthocidaris crassispina* and the ascidian *Polyandrocarpa misakiensis* are all C-type; they have about 20–30 % sequence similarity amongst themselves and with vertebrate C-type lectins [89, 122, 130, 183, 200]. The lectin from the albumen gland of the edible snail *Helix pomatia* is widely used as a specific reagent; it is constructed from three non-covalently bound dimers with internal disulphide bridges. That of the fly *Sarcophaga peregrina* is constructed from six chains according to the formula $\alpha_4\beta_2$ [89]. In many invertebrates, several variants of a lectin (isolectins) may appear together; thus, upon isoelectric focusing the albumen-gland lectin of the edible snail separates into 12 fractions of varying amino acid composition. It was a surprising finding that the lactose-specific haemolymph lectin of the cephalopod *Octopus vulgaris* is copper-containing and both electrophoretically and immunologically similar to the subunits of the respiratory pigment haemocyanin, although it has no oxygen-binding capacity [151]. The echinoidin of the coelom fluid from the sea urchin *Anthocidaris crassispina* has a native molecular weight of about 300 kDa and consists of 147-amino-acid subunits with three internal disulphide bridges and a serine-bound oligosaccharide. The C-terminal sequence shows significant homology to the mannose-binding protein of rat liver and other mammalian lectins, but also has a certain resemblance to the central sequence of the *Sarcophaga* lectin [55].

The **biological role of animal lectins** is largely undetermined; in all probability it is as varied as their structures. The first animal lectin discovered, the limulin of *Limulus polyphemus*, was considered to be a defence substance. Various observations suggest a defence function for the invertebrate lectins; the occurrence of lectins with marked specificity for 2-keto-3-desoxyoctonate (KDO; Fig. 6.5), which is found only on bacterial cells, can only be understood in this context. Various insect lectins agglutinate parasitic flagellates; the lectin of the blood-sucking bug *Rhodnius prolixus* is specific for the epimastigotes of *Trypanosoma cruzi*, living in insects, but not for the trypomastigotes in vertebrate blood. The phagocytosis-stimulating (opsonic) effect of lectins has often been discussed but has seldom been demonstrated with pure lectins and species-specific phagocytes, e.g. in the lobster *Homarus americanus*, the caterpillar of the butterfly *Spodoptera exigua*, the oyster *Crassostrea virginica*, the edible mussel *Mytilus edulis* and the snail *Lymnaea stagnalis* [17]. The carcinoscorpin of the Indian horseshoe crab *Carcinoscorpius rotundicauda* agglutinates its own amoebocytes and could, therefore, possibly influence their defence function [176]. The lectin bound to granulae in the egg cortex of *Xenopus* is supposedly involved in the formation of the fertilization membrane between the vitellin membrane and the egg jelly [124]. A role in tissue differentiation may be assumed for several vertebrate lectins whose concentration changes in a tissue-specific manner during ontogeny, e.g. the chicken CLL-I, which makes up no less than 0.1 % of the total protein in 16-day embryonal breast muscle but is entirely absent from adult muscle. The membrane-bound lectins of mammalian and bird liver cells are responsible for the binding, pinocytosis and further degradation of circulating gly-

coproteins which, as the result of partial hydrolysis, have lost their protective sialic acid residues [42, 166, 167, 200].

6.8 Humoral Defence in Invertebrates

The invertebrates have no system equivalent to the complex immune system of the vertebrates. Nevertheless, the interior of the invertebrate body is just as sterile as that of vertebrates, although the danger of microbial contamination is just as great. They must also have mechanisms for distinguishing self and non-self, although we know next to nothing about these. In the case of invertebrate animals, one can also differentiate between **cellular and humoral defence mechanisms**. The most important cellular mechanisms are phagocytosis and the encapsulation of foreign cells and organisms. The humoral defence components of the invertebrates encompass the cell-agglutinating agglutinins (lectins), the cell-wall-perforating lysins [24, 150], and lysozyme, which has the capacity to dissolve the cell walls of many bacteria. Humoral defence mechanisms are particularly well developed in the annelids, arthropods, molluscs and echinoderms. Of the three groups of defence substances, lectins have already been discussed, and the lysozymes will be dealt with in Chapter 13. Lysins have been detected in many representatives of the animal groups mentioned but have been little characterized biochemically. Detailed studies of humoral defence have been carried out for only a few insects and annelids [17, 44, 110].

Amongst the insects, the giant diapause pupa of the moth *Hyalophora cecropia* has received particular attention. With a body weight of up to 10 g, this provides ca. 1–2 ml of haemolymph. Treatment with living, non-pathogenic, or killed pathogenic bacteria strongly enhances the bactericidal activity of the haemolymph over several days. The expression of immunoprotein genes in the fat bodies is increased, whilst that of the other genes remains at basal levels. This response has allowed the isolation from the haemolymph of no fewer than 15 proteins which are involved in humoral defence; the cDNA or gene sequences have been determined for several of them. These antibacterial proteins are distributed amongst three families: the small cecropins (3–5 kDa), the larger attacins (20 kDa), and lysozyme. The whole system becomes especially important at the end of the pupal resting phase when histolysis begins and bacteria from the gut cavity may enter the haemolymph. The induction of antibacterial proteins occurs in the Lepidoptera, Diptera and Hymenoptera, i.e. in all highly developed insect orders. Amongst the proteins induced by bacteria are several which have no direct antibacterial effect, e.g. the 48-kDa protein P4, which is the major protein induced in *H. cecropia* [86, 180].

The **cecropins** are polypeptide chains of 35–37 amino acids with a basic N-terminus and a non-polar C-terminal region. The C-terminal amino acid carries an amide group formed from the terminal glycine of the primary translation product. The prepro-cecropins possess a signal sequence of 22 amino acids and a pro-sequence of 2–4 amino acids, both of which are absent from the mature protein. Three cecropins (A, B and D) are known from *H. cecropia* to kill and partially lyse both Gram-negative and Gram-positive bacteria. Comparisons of protein and gene sequences show that the antibacterial proteins of many Lepidoptera and Diptera are homologues of cecropin, even though only five amino acids are conserved. Some of these proteins received various other names from their discoverers: "lepidopterans A and B" in the lepidopterans *Manduca sexta, Antheraea pernyi* and *Bombyx mori;* and "sarcotoxins IA, IB and IC" in the dipterans *Sarcophaga peregrina* and *Drosophila melanogaster.* There was even a recent report of a cecropin in porcine gut tissue, indicating that these peptides are widely distributed in the animal kingdom [62, 85, 86, 96, 117].

H. cecropia possesses two forms of **attacin**, one basic and one acidic, encoded by different genes; they agree in 79 % of their 1844 amino acids. Like the cecropins, the attacins are also synthesized as a prepro-protein. They are specific for Gram-negative bacteria. A homologous protein has been detected in *M. sexta* via cDNA cloning. The attacins belong to the same family as the sarcotoxins IIA, IIB and IIC from *S. peregrina;* the sarcotoxin sequence of 270 amino acids is 20 % similar to that of the attacins [86, 181]. A third family of antibacterial insect proteins includes the three **diptericins** from the fly *Phormia terranovae*. The predominant form is a polypeptide of 82 amino acids with an amidated C-terminus, but it has no sequence similarity to either the cecropins or attacins. The subsidiary forms are highly homologous. The diptericins are also specific for Gram-negative bacteria [41, 86, 197]. Sequence similarity to the diptericins is shown by four antibacterial proteins which appear in the haemolymph of the honeybee *Apis mellifera* after the injection of

living bacteria: these are the three apidaecins Ia, Ib and II, which differ from each other in only 2 of their 18 amino acids, and the larger abaecin with 34 amino acids [29]. "Sapecin" from tissue cultures of the fly *S. peregrina*, and "phormicin" from the haemolymph of *P. terranovae* are specific for Gram-positive bacteria. In view of the similarities in their 40-amino-acid sequence and the positions of three disulphide bridges to the antibacterial "defensins" from neutrophil blood cells and mammalian macrophages, these proteins are better termed **insect defensins** [41, 86, 97, 112]. The same family includes the antibacterial protein royalisin (51 amino acids) from the royal jelly of the honeybee [52].

Very little is known about the antibacterial proteins of **other invertebrates**. Of the annelids, most attention has been paid to the oligochaete *Eisenia foetida*, whose coelom fluid has an unusually high protein concentration of 4–6 mg/l. Several protein components are normally present which not only lyse (haemolysins) or agglutinate (haemagglutinins) the erythrocytes of various vertebrates but also have bacteriostatic effects. The haemolysins, which are also bactericidal, have molecular masses of 40–45 kDa and show pronounced genetic polymorphism; they are produced in the chloragocytes. The haemagglutinins (lectins) have sizes between 11.5 and 40 kDa, with the larger components also having haemolytic properties. In the case of the earthworm *Lumbricus terrestris*, the coelom fluid presents a spectrum of multiple haemolysins and agglutinins after immunization with mammalian erythrocytes or inoculation with bacteria [88, 150, 178]. Lysozymes have also been detected in the coelom fluid of various annelids [17, 71, 88]. The haemocytes of the Japanese horseshoe crab *Tachypleus tridentatus* produce a basic peptide of 17 amino acids known as tachyplesin; this has an antibacterial action against both Gram-positive and Gram-negative bacteria. Similarly to the anti-lipopolysaccharide (LPS) factor from *T. tridentatus* and the North American xiphosuran *Limulus polyphemus* (102 amino acids; see p. 206), tachyplesin also inhibits haemolymph clotting caused by bacterial LPS [169].

6.9 Immune-Reactive Proteins of the Cell Surface

Animal cells are able to recognize and bind to other cells of the same type or to certain other cells of their own bodies, as has been clearly shown in cell aggregation experiments. Adhesive interactions between the same or different cells play a vital role in the processes of morphogenesis and tissue development [5, 123]. Cell-adhesion molecules are involved in the interaction between egg and sperm in fertilization. The proteins on the surface of parasitic protozoans have been intensively investigated for medical reasons. In intracellular species of parasites they are responsible for attachment to the host cell. For protozoans parasitizing vertebrates, it is the surface proteins which are the target for immune defence, and several protozoan species have evolved mechanisms to avoid the defence reactions of their host.

6.9.1 Cell-Adhesion Molecules of Vertebrates

The proteins involved in the linking of vertebrate cells to each other and to the extracellular matrix are divided into three groups according to function: cell-adhesion molecules (CAMs), cell-junction proteins (CJMs) and matrix- or substrate-adhesion molecules (SAMs). The SAMs include not only the receptor molecules of the cell surface but also the components of the intracellular matrix with which they interact, e.g. collagens, laminin, fibronectin and glycosamineglycans; these are discussed elsewhere in this book. The amino acid sequences derived from cDNA clones are known for many of these adhesion molecules. They often contain sequence motifs that are repeated several times and are known from other proteins, e.g. the Ig homology units of about 100 amino acids, the equally long type-III fibrinogen domains, EGF-like domains of about 40 amino acids, or calcium-binding regions with numerous Asp, Asn, Thr or Ser residues. The tripeptide sequence Arg-Gly-Asp (RGD sequence) is widely found in proteins of the extracellular matrix and in the corresponding receptors [5]. Electron micrographs reveal very complex three-dimensional structures for many of the adhesion molecules, with globular regions, stiff rods and flexible threads; it is assumed that the different structural elements have different functions [5].

The functional category of the **CAMs** includes proteins from two different super-families: the Ig super-family and the cadherin family. The best-known CAM of the **IG super-family** is N-CAM (N for neuronal), which already appears on neurons early in embryo development and is later to be found on other cell types. N-CAM is homophilic, i.e. the N-CAM of one cell binds to

the N-CAM of another cell. The N-terminal extracellular portion of N-CAM contains five Ig homology units and two segments corresponding to type III fibrinogen repeats (Fig. 6.2). There is a total of seven glycosylation sites on the third to fifth Ig domains; the carbohydrate chains of the fifth domain carry α(2–8)-linked polysialic acids which make up 20 % of the N-CAM molecular mass in embryonic brain but only 9 % in the adult brain. The homophilic binding of the N-CAM molecules is hindered by mutual repulsion of the negatively charged sialic acids; consequently, binding is stronger in the adult animal. Three N-CAMs of different size are found in chicken and mouse brain with deglycosylated molecular masses of 180 kDa (ld form), 140 kDa (sd form) and 120 kDa (ssd form). All three contain the same N-terminal extracellular sequence of about 862 amino acids. In the ld form, the transmembrane segment is followed by a large cytoplasmic domain which is 261 amino acids smaller in the sd form. There is no transmembrane or cytoplasmic domain in the ssd form; in this case, the extracellular chain is bound to a short non-polar sequence which is anchored to the membrane via phosphatidylinositol. The three forms arise by alternative splicing of a primary transcript from a gene with 19 exons. It has recently been found that the mRNA of mouse brain also has two sites in the extracellular region in which 3, 18 or 30 nucleotides can be introduced by alternative splicing. At least 27 alternatively spliced forms of the N-CAM mRNA are expressed during rat heart development [144]. The N-CAM of *Xenopus laevis* brain, which has been sequenced via the cDNA, corresponds approximately to the ld form of the higher vertebrates, and much of the sequence is in common [93]. Further CAMs from the Ig super-family have been described in vertebrates, e.g. the myelin-associated glycoprotein (MAG) with five Ig units and the neuronal L1 with six. Surprisingly, adhesion molecules from the Ig super-family have been found in *Drosophila melanogaster:* amalgam with three Ig units, fascilin II with five, and neuroglian with six [14].

The **cadherins** require Ca^{2+} ions for their adhesive activity; the Ca^{2+}-binding sites lie on the extracellular domain but have not yet been identified. Cadherins are known from mammals, birds and amphibians and are probably present in all vertebrates; similar adhesion proteins have also been found in *Drosophila*. The cadherins are integral membrane proteins of 120–130 kDa with 723–748 amino acids and only one transmembrane segment. The sequence, which is partly made up of repeats of about 110 amino acids, agrees by at least 40 % in the bird and mammal species examined. The first members of this family were discovered in various tissues of the mouse and chicken, and received corresponding names. In spite of the multiplicity of names, all vertebrate cadherins may be categorized into three types: the E-cadherin found in epithelial tissues is apparently identical to uvomorulin, L-CAM and cell CAM; the neuronal N-cadherin is identical to A-CAM and NcalCAM; and the final class includes the placental P-cadherin. The three classes show overlapping tissue distributions. However, they separated from each other early in vertebrate evolution; this may be deduced from the fact that mouse N-cadherin is 92 % similar to that of the chicken, but only 49 and 43 % similar to mouse E- and P-cadherin, respectively [185].

The special binding structures between neighbouring cells (cell junctions) include specific **junction proteins**. The desmosomes on which the intermediary filaments are anchored contain desmoplakin I (250 kDa), and the different types of intermediate junctions which are bound to microfilaments contain cell-specific proteins such as E-cadherin, vinculin and α-actinin. Plakoglobin (82 kDa), which is found in both types of junction, has 744 amino acids and sequence similarities to other known proteins [51]. "Gap junctions" form channel-like connections between cells and these allow the exchange of ions, secondary messengers and metabolites; furthermore, their low electrical resistance facilitates the transmission of action potentials. Each cell-cell channel is composed of six identical molecules of the protein connexin. A whole series of connexins have been isolated from mammals, the chicken and *Xenopus laevis;* they have chain lengths between 283 and more than 400 amino acids and molecular masses of 21–45 kDa. The connexins of the frog and rat show 32–41 % sequence similarity, the resemblance being greatest in the four transmembrane segments and the extracellular domain [13, 45].

Of the **SAMs**, only the cellular receptors will be considered here; their corresponding matrix components are the subjects of Chapters 11 and 13. The majority of the receptors for matrix components belong to the family of **integrins**, which are heterodimeric membrane proteins with an α- and a β-chain of 700–800 amino acids, including 56 invariant cysteine residues. Until recently, only three types of β-chain were known (β1, β2 and β3), and these combine with about ten types of α subunit to form three integrin subfamilies: the

VLA family with the β1 subunit includes various receptors for collagens, laminin and fibronectin; the LEUCAM family with the β2 subunit is made up of different leukocyte adhesion molecules; and the cytoadhesin family with β3 includes, amongst others, the vitronectin receptor. Sequence comparisons indicate the great age of the three β-chains. The β1 subunits of man, the mouse, chicken and *Xenopus* agree 82–86 %, but human β1, β2 and β3 are only 40–48 % similar [170]. With the recent discovery of a further three β subunits, a still larger variety is now recognizable in the integrin super-family [142, 168]. The most unusual of these newly discovered β-chains is the β4 subunit, which was sequenced via cDNA clones from epithelial cells. With 1778 amino acids, it is more than twice as long as the other β-chains; the extracellular domain (683 amino acids) lacks eight of the otherwise conserved 56 cysteine residues. The cytoplasmic domain (1072 amino acids) of β4 is much longer than that of the other β-chains (50 amino acids), shows no homology to the latter and, furthermore, contains four type-III fibronectin repeats [72, 182]. In addition to the integrins, further types of matrix receptor proteins are the collagen-binding protein anchorin CII [46] and the membrane-bound proteoglycan syndecan; this is approximately 300 amino acids long and has an extracellular domain that binds glycosamineglycans [105].

6.9.2 Cell-Binding Proteins of Invertebrates

Only very recently have **adhesion molecules from insects** been described in any detail. Quite surprisingly, amongst them are members of the Ig super-family: amalgam, fasciclin II and neuroglian. Amalgam is the product of a gene in the antennipedia complex of *Drosophila melanogaster;* during embryo development it is expressed on the surface of various mesodermal and neural cells. In addition to the signal sequence and a short hydrophobic C-terminal region, the encoded sequence of 333 amino acids includes three Ig homology units [164]. Fasciclin II and neuroglian belong to the four membrane-bound glycoproteins found on the growing axons and glial cells of various insects. Fasciclin II is composed of an extracellular domain of 742 amino acids with five Ig-like domains and two type-III fibronectin repeats, a transmembrane segment of 25 amino acids and a cytoplasmic portion of 108 amino acids [68]. Neuroglian contains six Ig homology units and five type-III fibronectin repeats and has a great similarity to L1 of mammals [14]. On growing axon membranes, the glycoproteins known as fasciclin I and III are not homologous to fasciclin II or to any other known protein. Fasciclin I is a homophilic adhesion molecule made up of four homologous domains of about 150 amino acids and anchored in the membrane via glycosyl-phosphatidyl-inositol. Probably as the result of hydrolytic cleavage of the anchor, large amounts of free soluble fasciclin I are found in *Drosophila* embryos [75]. The sequence of 488 fasciclin-III amino acids was determined from the cDNA, and has no similarity to the other two fasciclins [172]. The chaoptin located in the rhabdomers of the *Drosophila* eye contains within its sequence 41 copies of a leucine-rich repeat, which is known in numerous proteins from those of yeast to those of man, and is possibly involved in specific protein–protein or protein–membrane interactions [92]. Integrins are also known from insects, for example in *Drosophila*, where two integrins with the same β subunits are expressed in various very different tissues and are responsible for, amongst other things, the close juxtaposition of the dorsal and ventral wing epithelia [18].

Fertilization in the sea urchin, and cell aggregation in sponges are two examples of cell–cell interactions in invertebrates that have been well characterized at the molecular level. During the first step in the fertilization of sea urchins, an important role is played by a small body (acrosome) lying at the tip of the sperm. Interaction between fucose sulphate in the egg jelly and receptors in the sperm membrane leads to ejection of the acrosome granula from the sperm and the formation of the acrosome process; this connects the sperm with the vitellin membrane of the egg and makes possible fusion of the plasma membranes of the egg and sperm cells. The main component of the acrosome granula and the acrosome process is a carbohydrate-free protein **bindin**, which binds to a receptor protein in the vitellin membrane. The bindins show great species-specific diversity; thus, the acrosome reaction represents the first barrier to species intercrossing. The precursor (pro-bindin), encoded by a cDNA from *Stronglyocentrotus lividus*, includes an unusually large pro-peptide of 245 amino acids and the mature bindin of 236 amino acids. The N-terminal 73 amino acids of the bindin from the related species *S. franciscanus* differ in at least 45 % of positions [54].

In the **Porifera**, one can observe a fascinating example of biological self-organization: if the

cells of a sponge are separated mechanically or chemically, completely functional small sponges form in the cell suspension within a matter of days. Because of the great mobility of the sponge cells and the dynamic changeability of the structures they form, even this primitive multicellular organism requires highly developed mechanisms for cell–cell recognition; these may be effectively investigated by such reaggregation experiments.

Two **aggregation systems** can be distinguished in most sponge species: a non-species-specific primary system and a species-specific secondary system, which is lacking only in the calcareous sponges (Calcarea). Both require Ca^{2+} and therefore treatment with ethylenediaminetetraacetic acid (EDTA) leads to cell separation. In the **primary process**, aggregates of about 400–500 μm in diameter are formed; this also applies to mixtures of cells from different species, e.g. *Cliona celata* + *Geodia cydonium*. The responsible factor here is a primary aggregation factor (pAF) on the cell surface and this has been purified from *G. cydonium*. It is a glycoprotein with 50 % carbohydrate and consists of three polypeptides of 16.5, 15.5 and 13.5 kDa [118, 119].

In the **secondary process**, for example in *G. cydonium*, large aggregates with a diameter of more than 5 mm are formed, and can differentiate into small functional sponges. The secondary aggregation factors (sAF) from *Microciona prolifera* and *G. cydonium* are large glycoproteins of complex structure. The aggregation factor from *M. prolifera* is a proteoglycan of $2 \cdot 10^7$ Da containing 60–70 % carbohydrate [115]. The aggregation factor from *Geodia* is a bivalent molecule with the capacity to link the aggregation receptors (AR) of different cells; in addition, it has both glucuronyl transferase and galactosyl transferase activity. sAF-AR binding initiates a process of signal transduction across the cell membrane, in the course of which particular nuclear proteins are phosphorylated with the involvement of inositol trisphosphate and diacylglycerol as secondary messengers [156]. The aggregation factor of *G. cydonium* is associated with a 32-kDa protein, calpactin, which binds Ca^{2+} ions in cooperation with a membrane-bound phospholipid. The sequence of this protein, derived from its cDNA, agrees 80 % with that of vertebrate calpactin II [149]. The membrane-bound AR from *M. prolifera* has been purified and characterized; it is a glycoprotein of 15–18 kDa with 81 % carbohydrate. Binding of sAF and AR is via a lysine residue of the aggregation factor and the terminal glucoronic acid of the receptor. Cleavage at the glucuronyl residue by a membrane-bound glucuronidase results in cell separation; conversely, glucuronyl transfer by the sAF restores the aggregation capacity. In parallel, there is a further regulatory mechanism in which an anti-aggregation receptor on the cell surface interacts with the sAF-AR complex and cleaves off a terminal galactose residue. The anti-aggregation receptor of *Geodia* is a glycolipoprotein of about 180 kDa. It can be inactivated by a galactose-specific lectin, and it may be that the cell's own galactosidase and the galactosyl transferase activity of the sAF are involved in inactivation and reactivation. The species specificity of cell aggregation is determined largely by the specific structure of the aggregation factor and its receptor. In addition, transplant rejection may occur through the action of an "inhibiting aggregation factor", a glycoprotein of 26 kDa with 60 % carbohydrate [118].

6.9.3 The Variable Surface Glycoproteins of the Trypanosomes

The cell surface of the pathogenic African *Trypanosoma* species that infest mammals, e.g. *T. brucei* and *T. congolense*, which cause the nagana disease of cattle, or *T. equiperdum*, which is the causative agent of horse murrain, is completely covered by a 12- to 15-nm-thick layer of identical glycoprotein molecules which constitute 7–10 % of the total cell protein. The variability of these surface proteins is unique; one single clone, i.e. the progeny of a single cell, of *T. equiperdum* was found to possess more than 100 different variant proteins with no common immunological determinants. Alterations in these **"variant-specific surface glycoproteins"** (VSGs) at a frequency of about 10^{-6} per cell division enable the parasite to avoid the immune defences of the host. The VSGs are made up of polypeptide chains with a length of 450–500 amino acids and these chains form a compact aggregate of homodimers on the cell surface. The outwardly directed N-terminal domains have several variable antigen determinants along a length of about 400 amino acids. The C-terminal domain carries several oligosaccharides, giving a carbohydrate content of 7–17 % [139]. Anchorage of VSGs in the membrane is brought about by binding of the C-terminus to a complex glycosyl-phosphatidyl-inositol via a covalent bond between the carboxyl group and an ethanolamine residue. The precursor of this structure in *T. brucei* is the free glycolipid ethanolamine-

phosphate-6Manα(1-2)Manα(1-6)Manα(1-4)Glc Nα(1-6)-phosphatidylinositol. After its incorporation, this structure is modified by the attachment of several galactose residues to the mannose next to the glucosamine [113]. The most closely investigated of the VSG variants is MITat 1.4 of *T. brucei:* the complete amino acid sequence, details of the secondary structure, the glycolipid anchor and the structure of the Asn-bound carbohydrate are all known [201].

T. equiperdum possesses at least 100 and *T. brucei* more than 1000 closely linked **VSG genes**, which together make up about 10 % of the DNA content of $8.6 \cdot 10^7$ bp. As the chromosomes of the trypanosomes never appear in a condensed form in the cell cycle, the ploidy level cannot be directly determined; however, genetic observations suggest that *T. brucei* is diploid. Nevertheless, there is only one copy of each VSG gene in the genome; the evolution of the VSG genes is apparently so rapid that homologous positions on the chromosomes are always occupied by different genes. The mechanisms for maintaining the VSG repertoire are mainly those of gene duplication and conversion [139].

The VSG genes are expressed only when they are at one of the **expression sites** at the ends of the chromosomes (telomeres). The VSG genes lying internally on the chromosome can be activated by the transposition of the silent basic copy (BC) to one of the telomeric expression sites (duplicative transposition) to give an "expression-linked copy" (ELC). The transposed DNA segment begins 1–2 kb in front of the gene and ends just behind it. Some ELCs are hybrids of different BCs [155]. *T. brucei* has about 200 chromosomes whose telomeres carry VSG genes; only one of the expression sites is active at any one time, but the selection mechanism is unknown. Silent VSG genes located close to expression sites can be activated in various ways; the inactive gene can be transposed duplicatively to an expression site or can exchange places with an active gene. ELCs that become inactive are not retransposed to internal chromosomes sites but are eliminated [102, 139]. The activation of VSGs after infection follows a certain course whose regulation is not known but whose biological advantage is obvious. The repeated activation of any one gene would undermine the successful competition with the host immune system. In the tsetse fly, one particular set of 10–20 VSG genes is activated. The VSGs produced during the early states of infection in vertebrates arise from genes which are already located at the telomeres [102].

The **transcription unit** of a VSG gene has a length of 60 kb and is polycistronic. In front of the VSG genes at the expression site (ES) are more than seven ES-associated genes (ESAGs) together with a more than 10-kb-long segment of 70-bp repeats. The importance of the ESAGs is unclear; one was found to code for a membrane-bound protein of unknown function [2, 139]. Quite surprisingly, the transcription of the VSG genes, and of the procyclic acidic repetitive protein (PARP) genes, described below, from insect trypanosomes, is not sensitive to α-amanitin. All other protein-coding trypanosome genes are transcribed by a normal, amanitin-sensitive polymerase II. In this case, there is either an amanitin-insensitive form in addition to the normal polymerase II, or the normal polymerase II can be made insensitive by means of some additional factor [59, 157]. As for all mRNAs of the Kinetoplastida, the VSG mRNAs contain a 5′-terminal mini-exon of 35 nt that is introduced by trans-splicing (p. 39). The **primary translation** product (pre-VSG) has not only a relatively long, 25- to 33-bp N-terminal signal sequence but also a C-terminal extension of 17–23 non-polar amino acids. Immediately after translation, but whilst still in the endoplasmic reticulum, this is exchanged for the phosphoglycolipid which anchors the VSG to the cell membrane. A phospholipase C can release the VSG, in which case sn-1,2-dimyristylglycerol remains behind in the membrane. This enzyme, which may play a role in the exchange of VSG variants, has been isolated; it has a molecular weight of 39 kDa and is specific for VSGs [20]. The procyclic stages of *T. brucei* from the intermediate host *Glossina* sp. or from cultures at 27 °C have no VSGs on their surface but instead have PARPs. This is a family of invariable proteins composed mainly of Glu-Pro repeats with a single N-glycosylation site [33].

6.9.4 Surface Proteins of Other Protozoa

The South American *T. cruzi*, the agent causing Chagas' disease, has no VSG system. In this case, both the epimastigotes from the insect host and the trypomastigotes from mammalian blood have different glycoproteins on their surfaces, but these do not show the variability typical of VSGs [192]. The penetration into cells by parasitic forms in vertebrate blood is inhibited by antibodies against the surface protein. The gene for this protein has been sequenced: the coding sequence contains the nonapeptide DKKESGDSE in five

complete copies and two with slight changes; it is noteworthy that the surface protein of another cell parasite, the circumsporozoite protein (CSP) of the malarial agent, also has a repetitive structure [137]. The forms of the genus *Leishmania* that live in insects (promastigotes) all possess the surface protein P63, which is anchored in the membrane via the same phosphoglycolipid as occurs in the VSGs from *Trypanosoma*. The proteins from five different species of *Leishmania* have the same peptide pattern, i.e. P63 is extremely conserved in evolution. There is immunological evidence for P63 in the vertebrate forms of *Leishmania* (amastigotes), but it is not located on the cell surface, probably due to the absence of the anchoring lipid [82].

In the malarial agents of the genus *Plasmodium*, it is the proteins on the surface of the infectious sporozoites, the CSPs, which are probably important for the interaction with erythrocyte membranes. There is only one CSP gene, but the evolutionary alterations of the CSPs are so rapid as to radically hinder serum therapy for malaria; each inoculation results in the selection of insensitive variants. The CSP genes from various *Plasmodium* species have been sequenced. The primary translation product is a polypeptide of 350–370 amino acids and carrys a signal sequence at the N-terminus and a sequence for membrane anchorage at the C-terminus. The central region is made up of repeats of 9–11 amino acids, of which both the number and sequence vary between different lines in a species [37]. This variable region is flanked by two regions which are similar in all species. Region II apparently includes the erythrocyte attachment site with the critical sequence – VTCG – [145]. Whilst the surface of the sporozoites has only CSPs, the merozoite surface bears many immunodominant proteins of which only one species has been sequenced [40]. Immunogenic surface proteins with repetitive sequences are characteristic of parasitic protozoans; apart from *Trypanosoma cruzi, Leishmania major* and various *Plasmodium* species, such proteins have been found on the merozoites of *Eimeria acervulina*, a gut parasite of the chicken [83].

The surface of various *Paramecium* species is covered by a layer of large glycoprotein molecules which together make up about 3 % of the total cell protein. Because *Paramecium* can be immobilized and killed by antisera raised against their surface glycoproteins, the proteins have, rather unfortunately, been named immobilization-antigens (i-ags). They consist of a family of related proteins of about 300 kDa which are anchored in the membrane via glycosylinositol phospholipid [26]. The i-ags are specific to each *Paramecium* strain, after which they are named; for example, the 11 i-ags from strain 51 of *P. tetraurelia* are designated 51A to 51K. Each i-ag is encoded by a single gene but only one such gene is expressed in a strain at any one time. Environmental conditions, e.g. temperature or pH, have a strong influence on gene expression. Several i-ag genes have been sequenced and found to contain no introns. So, for example, the 156G gene of *P. primaurelia* consists of an uninterrupted coding sequence of 8145 bp; the polypeptide of 2715 amino acids has a marked periodic structure. As in almost all ciliate genes, TAA and TAG code for glutamine, and only TGA codes for "stop" [57, 140].

References

1. Åkerström B.: Immunological analysis of α1-microglobulin in different mammalian and chicken serum. α1-Microglobulin is 5–8 kilodaltons larger in primates. J. biol. Chem. 260: 4839–44 (1985)
2. Alexandre S. et al.: Putative genes of a variant-specific antigen gene transcription unit in Trypanosoma brucei. Mol. cell. Biol. 8: 2367–78 (1988)
3. Amemiya C. T., Haire R. N. and Litman G. W.: Nucleotide sequence of a cDNA encoding a third distinct Xenopus immunoglobulin heavy chain. Nucleic Acids Res. 17: 5389 (1989)
4. Amemiya C. T. and Litman G. W.: Complete nucleotide sequence of an immunoglobulin heavy-chain gene and analysis of immunoglobulin gene organization in a primitive teleost species. Proc. Nat. Acad. Sci. USA 87: 811–815 (1990)
5. Anderson H.: Adhesion molecules and and animal development (Review). Experientia 46: 2–13 (1990)
6. Arai K. I. et al.: Cytokines: Coordinators of immune and inflammatory responses. Annual Rev. Biochem. 59: 783–836 (1990)
7. Bazan J. F.: Structural design and molecular evolution of a cytokine receptor superfamily. Proc. Nat. Acad. Sci. USA 87: 6934–38 (1990)
8. Beale D.: A comparison of the amino acid sequences of the extracellular domains of the immunoglobulin superfamily. Possible correlations between conservancy and conformation (Review). Comp. Biochem. Physiol. Pt. B 80: 181–194 (1985)
9. Beck G. et al.: Characterization of interleukin-1 activity in tunicates. Comp. Biochem. Physiol. Pt. B 92: 93–98 (1989)
10. Becker R. S. and Knight K. L.: Somatic diversification of immunoglobulin heavy chain VDJ genes. Evidence for somatic conversion in rabbits. Cell 63: 987–997 (1990)
11. Bentley D. R.: Primary structure of human complement component C2. Homology to two unrelated protein families. Biochem. J. 239: 339–345 (1986)

12. Bertheussen K.: Complement-like activity in sea urchin coelomic fluid. Develop. comp. Immunol. 7: 21–31 (1983)
13. Beyer E. C.: Molecular cloning and developmental expression of two chick embryo gap junction proteins. J. Biol. Chem. 265: 14439–43 (1990)
14. Bieber A. J. et al.: Drosophila neuroglian: A member of the immunoglobulin superfamily with extensive homoloy to the vertebrate neural adhesion molecule L1. Cell 59: 447–460 (1989)
15. Bjorkman P. J. and Parham P.: Structure, function, and diversity of class I major histocompatibility complex molecules. Annual Rev. Biochem. 59: 253–288 (1990)
16. Blumberg R. S. et al.: Structure of the T-cell antigen receptor. Evidence for two CD3-epsilon subunits in T-cell receptor-CD3 complex. Proc. Nat. Acad. Sci. USA 87: 7220–24 (1990)
17. Brehelin M. (ed.): Immunity in invertebrates. Cells, molecules, and defense reactions. Springer, Berlin 1986
18. Brower D. L. and Jaffe S. M.: Requirement for integrins during Drosophila wing development. Nature 342: 285–287 (1989)
19. Brüggemann M. et al.: Immunoglobulin heavy chain locus of the rat: Striking homology to mouse antibody genes. Proc. Nat. Acad. Sci. USA 83: 6075–79 (1986)
20. Bülow R. and Overath P.: Purification and characterization of the membrane-form variant surface glycoprotein hydrolase of Trypanosoma brucei. J. biol. Chem. 261: 11918–23 (1986)
21. Bulleid, N. J. et al.: Source of heterogeneity in secreted interferon-gamma. A study on products of translation in vitro. Biochem. J. 268: 777–781 (1990)
22. Burnett R. C. et al.: The IgA heavy-chain gene family in rabbit. Cloning and sequence analysis of 13 Cα genes. Embo J. 8: 4041–47 (1989)
23. Calabi F. and Neuberger M. S.: Molecular genetics of immunoglobulin. New comprehensive Biochemistry Vol. 17. Elsevier, Amsterdam 1987
24. Canicatti C.: Hemolysins. Pore-forming proteins in invertebrates. Experientia 46: 239–244 (1990)
25. Capra J. D. and Tucker P. W.: Human immunoglobulin heavy chain genes. J. Biol. Chem. 264: 12745–48 (1989)
26. Cardoso de Almeida M. and Deregnaucourt C.: The membrane-anchor of Paramecium temperature-specific surface antigens is glycosylinositol phospholipid. Biochem. biophys. Res. Commun. 147: 1219–25 (1987)
27. Caron M., Joubert R. and Bladier D.: Purification and characterization of a β-galactoside-binding soluble lectin from rat and bovine brain. Biochim. biophys. Acta 925: 290–296 (1987)
28. Cassels F. J., Marchalonis J. J. and Vasta C. R.: Heterogenous humoral and hemocyte-associated lectins with N-acylaminosugar specificities from the blue crab, Callinectes sapidus Rathburn. Comp. Biochem. Physiol. Pt. B 85: 23–30 (1986)
29. Casteels P.: Isolation and characterization of abaecin, a major antibacterial response peptide in the honeybee (Apis mellifera). Eur. J. Biochem. 187: 381–386 (1990)
30. Castro V. M., Boman H. G. and Hammarström S.: Isolation and characterization of a group of isolectins with galactose/N-acetylgalactosamine specificity from hemolymph of the giant silk moth Hyalophora cecropia. Insect Biochem. 17: 513–523 (1987)
31. Ceretti D. P. et al.: Cloning, sequence and expression of bovine interleukin 2. Proc. Nat. Acad. Sci. USA 83: 3223–27 (1986)
32. Chen C. H. et al.: TCR3: a third T-cell receptor in the chicken. Proc. Nat. Acad. Sci. USA 86: 2351–55 (1989)
33. Clayton C. E. and Mowatt M. R.: The procyclic acidic repetitive proteins of Trypanosoma brucei. Purification and post-translational modification. J. Biol. Chem. 264: 15088–93 (1989)
34. Clerch L. B. et al.: Sequence of a full-length cDNA for rat lung β-galactoside-binding protein: Primary and secondary structure of the lectin. Biochemistry 27: 692–699 (1988)
35. Cohen D. R., Happel A. J. and Young I. G.: Cloning and expression of the rat interleukin-3 gene. Nucleic Acids Res. 14: 3641–58 (1986)
36. Cooper E. L., Langlet C. and Bierne J. (eds.): Developmental and comparative immunology. Alan R. Liss, New York 1987
37. de la Cruz V. F. et al.: Evolution of the immunodominant domain of the circumsporozoite protein gene from Plasmodium vivax. J. biol. Chem. 262: 6464–67 (1987)
38. Davis M. M.: T cell receptor gene diversity and selection. Annual Rev. Biochem. 59: 475–496 (1990)
39. Dawson M.: Lymphokines and interleukins. Open Univ. Press, Ballmoor 1990
40. Delportillo H. A. et al.: Primary structure of the merozoite surface antigen-1 of Plasmodium vivax reveals sequences conserved between different Plasmodium species. Proc. Nat. Acad. Sci. USA 88: 4030–34 (1991)
41. Dimarcq J. et al.: Insect immunity: expression of the two major inducible antibacterial peptides, defensin and diptericin, in Phosmia terranovae. Embo J. 9: 2507–15 (1990)
42. Drickamer K. et al.: Two distinct classes of carbyhydrate-recognition domains in animal lectins. J. Biol. Chem. 263: 9557–60 (1988)
43. Due C., Simonsen M. and Olsson L.: The major histocompatibility complex class I heavy chain as a structural subunit of the human cell membrane insulin receptor: Implications for the range of biological functions of histocompatibility antigens. Proc. Nat. Acad. Sci. USA 83: 6007–11 (1986)
44. Dunn P. E.: Biochemical aspects of insect immunology. Annual Rev. Entomol. 31: 321–339 (1986)
45. Ebihara L. et al.: Cloning and expression of a Xenopus embryonic gap junction protein. Science 243: 1194–95 (1989)
46. Fernandez M. P. et aal.: The structure of anchorin CII, a collagen-binding protein isolated from chondrocyte membrane. J. Biol. Chem. 263: 5921–25 (1988)
47. Figueroa F., Guenther E. and Klein J.: MHC polymorphism pre-dating speciation. Nature 335: 265–267 (1988)
48. Fink de Cabutti N. E. et al.: Purification and some characteristics of a β-galactoside binding soluble lectin from amphibian ovary. FEBS Letters 223: 330–334 (1987)
49. Flajnik M. F. et al.: Evolution of the major histocompatibility complex. Molecular cloning of a major histocompatibility complex class-I from the amphibian

Xenopus. Proc. Nat. Acad. Sci. USA 88: 537–541 (1991)
50. Flavel R. A. et al.: Molecular biology of the H-2 histocompatibility complex. Science 233: 437–443 (1986)
51. Franke W. W. et al.: Molecular cloning and amino acid sequence of human plakoglobin, the common junctional plaque protein. Proc. Nat. Acad. Sci. USA 86: 4027–31 (1989)
52. Fujiwara S. et al.: A potent antibacterial protein in royal jelly. Purification and determination of the primary structure of royalisin. J. Biol. Chem. 265: 11333–37 (1990)
53. Gabius H. J.: Detection and function of mammalian lectins. With emphasis on membrane lectins (Review). Biochim. biophys. Acta 1071: 1–18 (1991)
54. Gao B. et al.: Sequence of mRNA coding for bindin, a species-specific sea urchin sperm protein required for fertilization. Proc. Nat. Acad. Sci. USA 83: 8634–38 (1986)
55. Giga Y., Ikai A. and Takahashi K.: The complete amino acid sequence of echinoidin, a lectin from the coelomic fluid of the sea urchin Acanthocidaris crassispina. Homologies with mammalian and insect lectins. J. biol. Chem. 262: 6197–6203 (1987)
56. Gimbrone M. A. jr., Obin M. S. and Brock A. F.: Endothelial interleukin-8: A novel inhibitor of leukocyte-endothelial interactions. Science 246: 1601–03 (1989)
57. Godiska R.: Structure and sequence of the H surface protein gene of Paramecium and comparison with related genes. Mol. gen. Genetics 208: 529–536 (1987)
58. Gojobori T. and Nei M.: Concerted evolution of the immunoglobulin VH gene family. Mol. Biol. Evol. 1: 195–212 (1984)
59. Grondal E. J. M. et al.: Characterization of the RNA polymerases of Trypanosoma brucei: trypanosomal RNAs are composed of transcripts derived from both RNA polymerases II and III. Embo J. 8: 3383–89 (1989)
60. Grossberger D. et al.: Conservation of structural and functional domains in complement component C3 of Xenopus and mammals. Proc. Nat. Acad. Sci. USA 86: 1323–27 (1989)
61. Groves M. L., Greenberg R. and Farrell H. M. jr.: Isolation and properties of goat β-2 microglobulin. Comp. Biochem. Physiol. 81B: 621–627 (1985)
62. Gudmundsson G. H. et al.: The cecropin locus: Cloning and expression of a gene cluster encoding 3 antibacterial peptides in Hyalophora cecropia. J. Biol. Chem. 266: 11510–17 (1991)
63. Guillemot F. et al.: The major histocompatibility complex in the chicken. Trends Genet. 5: 300–304 (1989)
64. Haefliger J. A. et al.: Complementary DNA cloning of complement C8ß and its sequence homology to C9. Biochemistry 26: 3551–56 (1987)
65. Haefliger J. A. et al.: Complete primary structure and functional characterization of the sixth component of the human complement system. Identification of the C5b-binding domain in complement 6. J. Biol. Chem. 264: 18041–51 (1989)
66. Harada-Azumi K., Yokosawa H. and Ishii S. I.: N-Acetyl-galactosamin-specific lectin, a novel lectin in the haemolymph of the ascidian Halocynthia roretzi: isolation, characterization and comparison with galactose-specific lectin. Comp. Biochem. Physiol. Pt. B 88: 375–381 (1987)
67. Harding F. A. et al.: 2 distinct immunoglobulin heavy chain isotypes in a primitive, cartilaginous fish, Raja erinacea. Nucleic Acids Res. 18: 6369–76 (1990)
68. Harrelson A. L. and Goodman C. S.: Growth cone guidance in insects. Fasciclin II is a member of the immunoglobulin family. Science 242: 700–708 (1988)
69. Hashimoto K., Nakansihi T. and Kurosawa Y.: Isolation of carp genes encoding major histocompatibility complex antigens. Proc. Nat. Acad. Sci. USA 87: 6863–67 (1990)
70. Hillarp A. and Dahlbaeck B.: Novel subunit in C4b-binding protein required for protein S binding. J. Biol. Chem. 263: 12759–64 (1988)
71. Hirigoyenberry F., Lasalle F. and Lassegues M.: Antibacterial activity of Eisenia fetida andrei coelomic fluid. Transcription and translation regulation of lysozyme and protein evidenced after bacterial infection. Comp. Biochem. Physiol. Pt. B 95: 71–75 (1990)
72. Hogervorst F. et al.: Cloning and sequence analysis of beta-4 cDNA: an integrin subunit that contains a unique 118 kd cytoplasmic domain. Embo J. 9: 765–770 (1990)
73. Homans S. W. et al.: Complete structure of the glycosyl phosphatidylinositol membrane anchor of rat brain Thy-1 glycoprotein. Nature 333: 269–272 (1988)
74. Honjo T. and Habu S.: Origin of immune diversity: genetic variation and selection. Annual Rev. Biochem. 54: 803–830 (1985)
75. Hortsch M. and Goodman C. S.: Drosophila fasciclin I, a neural cell adhesion molecule, has a phosphatidylinositol lipid membrane anchor that is developmentally regulated. J. Biol. Chem. 265: 15104–09 (1990)
76. Howard O. M. Z., Rao A. G. and Sodetz J. M.: Complementary DNA and derived amino acid sequence of the β-subunit of human complement protein C8: Identification of a close structural and ancestral relationship to the α-subunit and C9. Biochemistry 26: 3565–70 (1987)
77. Hughes A. L. and Nei M.: Evolutionary relationships of class-II major-histocompatibility-complex genes in mammals. Mol. Biol. Evol. 7: 491–514 (1990)
78. Hughes A. L., Ota T. and Nei M.: Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class-I major-histocompatibility-complex molecules. Mol. Biol. Evol. 7: 515–524 (1990)
79. Hunkapiller T. and Hood L.: The growing immunoglobulin gene superfamily. Nature 323: 15–16 (1986)
80. Hunt L. T., Elzanowski A. and Barker W. C.: The homology of complement factor C8 gamma-chain and α1-microglobulin. Biochem. biophys. Res. Commun. 149: 282–288 (1987)
81. Ingram G. A. and Molyneux D. H.: Lectins (haemagglutinins) in the haemolymph of Glossina fuscipes fuscipes: isolation, partial characterization, selected physico-chemical properties and carbohydrate-binding specificities. Insect Biochem. 20: 13–27 (1990)
82. Jaehnig F. and Etges R.: Secondary structure of the promastigote surface of Leishmania. FEBS Letters 241: 79–82 (1988)
83. Jenkins M. C.: A cDNA encoding a merozoite surface protein of the protozoan Eimeria acervulina con-

tains tandem-repeated sequences. Nucleic Acids Res. 16: 9863 (1988)
84. Kamiya H. et al.: Isolation and characterization of hemagglutinins from the sponge Dysidea herbacea. Experientia 41: 1201–02 (1985)
85. Kanai A. and Natori S.: Cloning of gene cluster for sarcotoxin-I, antibacterial proteins of Sarcophaga peregrina. FEBS Letters 258: 199–202 (1989)
86. Kanost M. R. et al.: Insect hemolymph proteins. Adv. Insect Physiol. 22: 299–396 (1990)
87. Kappes D. and Strominger J. L.: Human class II major histocompatibility complex genes and proteins. Annual Rev. Biochem. 57: 991–1028 (1988)
88. Kauschke E. and Mohrig W.: Comparative analysis of hemolytic and hemagglutinating activities in the coelomic fluid of Eisenia foetida and Lumbricus terrestris (Annelida, Lumbricidae). Develop. comp. Immunol. 11: 331–341 (1987)
89. Kobayashi A. et al.: Cloning and in vitro transcription of the Sarcophaga lectin gene. Biochim. biophys. Acta 1009: 244–250 (1989)
90. Kokubu F. et al.: Extensive families of constant region genes in a phylogentically primitive vertebrate indicate an additional level of immunoglobulin complexity. Proc. Nat. Acad. Sci. USA 84: 5868–72 (1987)
91. Koppenheffer T. L.: Serum complement systems of ectothermic vertebrates. Develop. comp. Immunol. 11: 279–286 (1987)
92. Krantz D. E. and Zipursky S. L.: Drosophila chaoptin, a member of the leucine-rich repeat family, is a photoreceptor cell-specific adhesion molecule. Embo J. 9: 1969–77 (1990)
93. Krieg P. A., Sakaguchi D. S. and Kintner C. R.: Primary structure and developmental expression of a large cytoplasmic domain form of Xenopus laevis neural cell adhesion molecule (NCAM). Nucleic Acids Res. 17: 10321–35 (1989)
94. Kubo T., Kawasaki K. and Natori S.: Sucrose-binding lectin in regenerating cockroach (Periplaneta americana) legs: its purification from adult hemolymph. Insect Biochem. 20: 585–591 (1990)
95. Kuhner M. et al.: Gene conversion in the evolution of both the H-2 and Qa class-I genes of the murine major histocompatibility complex. Genetics 126: 1115–26 (1990)
96. Kylsten P., Samakovlis C. and Hultmark D.: The cecropin locus in Drosophila: a compact gene cluster involved in response to infection. Embo J. 9: 217–224 (1990)
97. Lambert J. et al.: Insect immunity: Isolation from immune blood of the dipteran Phormia terranovae of two insect antibacterial peptides with sequence homology to rabbit lung macrophage bactericidal peptides. Proc. Nat. Acad. Sci. USA 86: 262–266 (1989)
98. Law S. K. A. and Reid K. B. M.: Complement. Oxford Univ. Press, New York 1988
99. Lawlor D. A. et al.: HLA-A and B polymorphisms predate the divergence of humans and chimpanzees. Nature 335: 268–271 (1988)
100. Lee Y. M. et al.: Primary structure of the bovine conglutinin, a member of the C-type animal lectin family. J. Biol. Chem. 266: 2715–23 (1991)
101. Litman G. W. et al.: Complete nucleotide sequences of three VH genes in Caiman, a phylogenetically ancient reptile: evolutionary diversification in coding segments and variation in the structure and organization of recombinant elements. Proc. Nat. Acad. Sci. USA 82: 844–848 (1985)
102. Longacre S. and Eisen H.: Expression of whole and hybrid genes in Trypanosoma equiperdum antigenic variation. Embo J. 5: 1057–63 (1986)
103. Mackinnon C. M. et al.: Molecular cloning of cDNA for human complement factor C1s – The complete amino acid sequence. Eur. J. Biochem. 169: 547–553 (1987)
104. Mage R. G. et al.: Evolution of genes for allelic and isotypic forms of immunoglobulin kappa chains and of genes for T-cell receptor beta chains in rabbits. J. mol. Evol. 25: 292–299 (1987)
105. Mali M. et al.: Sequence of human syndecan indicates a novel gene family of inegral membrane proteoglycans. J. Biol. Chem. 265: 6884–89 (1990)
106. Mandal C. and Basu S.: A unique specificity of a sialic acid binding lectin, achatinin, from the hemolymph of Achatina fulica snail. Biochem. biophys. Res. Commun. 148: 795–801 (1987)
107. Mandal C., Basu S. and Mandal C.: Physicochemical studies on achatinin H, a novel sialic acid-binding lectin. Biochem. J. 257: 65–71 (1989)
108. Mandal C. and Mandal C.: Sialic acid-binding lectins. Experientia 46: 433–441 (1990)
109. Mansour M. H., Negm H. I. and Cooper E. L.: Thy-1 evolution. Develop. comp. Immunol. 11: 3–15 (1987)
110. Marchalonis J. J. and Reinisch C. L. (eds.): Defense molecules. Wiley-Liss, New York 1990
111. Matsuyama H. et al.: Characterization of the alternate complement pathway of carp. Develop. comp. Immunol. 12: 403–408 (1988)
112. Matsuyama K. and Natori S.: Molecular cloning of cDNA for sapecin and unique expression of the sapecin gene during the development of Sarcophaga peregrina. J. Biol. Chem. 263: 17117–21 (1988)
113. Mayor S. et al.: Glycolipid precursors for the membrane anchor of Trypanosoma brucei variant surface glycoproteins. I. Glycan structure of the phosphatidylinositol-specific phospholipase C sensitive and resistant glycolipids. J. Biol. Chem. 265: 6164–73 (1990)
114. Meager A.: Cytokines. Prentice-Hall, New York 1991
115. Misevic G. N. and Burger M. M.: The species-specific cell-binding site of the aggregation factor from the sponge Microciona prolifera is a highly repetitive novel glycan containing glucuronic acid, fucose, and mannose. J. Biol. Chem. 265: 20577–84 (1990)
116. Mitra D., Sakkar M. and Allen A. K.: Further characterization of the cold agglutinin from the snail Achatina fulica. Biochem. J. 242: 331–338 (1987)
117. Morishima I. et al.: Isolation and structure of the cecropins, inducible antibacterial peptides, from the silkworm, Bombyx mori. Comp. Biochem. Physiol. Pt. B 95: 551–554 (1990)
118. Mueller W. E. G.: Cell membranes in sponges. Int. Rev. Cytol. 77: 129–181 (1982)
119. Mueller W. E. G. et al.: Identification and isolation of the primary aggregation factor from the cell membrane of the sponge Geodia cydonium. Mol. Cell Biochem. 67: 55–64 (1985)
120. Mueller-Eberhard H. J.: Molecular orgaization and function of the complement system. Annual Rev. Biochem. 57: 321–347 (1988)
121. Mullainadhan P. and Renwrantz L.: Comparative analysis of agglutinins from hemolymph and albumin

gland of Helix pomatia. J. comp. Physiol. B 159: 443–452 (1989)
122. Muramoto K. and Kamiya H.: The amino-acid sequence of multiple lectins of the acorn barnacle Megabalanus rosa and its homology with animal lectins. Biochim. biophys. Acta 1039: 42–51 (1990)
123. Muramatsu T.: Cell surface and differentiation. Chapman and Hall, London 1990
124. Nishihara T. et al.: Isolation and characterization of a lectin from the cortical granules of Xenopus laevis eggs. Biochemistry 25: 6013–20 (1986)
125. Nonaka M. et al.: Identification and characterization of a variant of the third component of complement (C3) in rainbow trout (Salmo gairdneri) serum. J. biol. Chem. 260: 809–815 (1985)
126. Nonaka M. et al.: Complete nucleotide and derived amino acid sequences of the fourth component of mouse complement (C4). Evolutionary aspects. J. biol. Chem. 260: 10936–43 (1985)
127. Norment A. M. and Littman D. R.: A second subunit of CD8 is expressed in human T cells. Embo J. 7: 3433–39 (1988)
128. Novotny J. et al.: Secondary, tertiary, and quaternary structure of T-cell-specific immunoglobulin-like polypeptide chains. Proc. Nat. Acad. Sci. USA 83: 742–746 (1986)
129. Ogata R. T., Rosa P. A. and Zepf N. E.: Sequence of the gene for murine complement component C4. J. Biol. Chem. 264: 16565–72 (1989)
130. Ozeki Y. et al.: Amino acid sequence and molecular characterization of a D-galactoside-specific lectin purified from sea urchin (Acanthocidaris crassispina) eggs. Biochemistry 30: 2391–94 (1991)
131. Paroutod P. et al.: Extensive amino acid sequence homologies between animal lectins. Proc. Nat. Acad. Sci. USA 84: 6345–48 (1987)
132. Parinello N. and Arizza V.: Sugar-specific cellular lectins of Phallusia mamillata hemocytes: Purification, characterization and evidence for cell surface localization. Develop. comp. Immunol. 13: 113–121 (1989)
133. Parvari R. et al.: Somatic diversification of chicken immunoglobulin light chains by point mutations. Proc. Nat. Acad. Sci. USA 87: 3072–76 (1990)
134. Paul S. R. et al.: Molecular cloning of a cDNA encoding interleukin 11, a stromal cell-derived lymphopoietic and hematopoietic cytokine. Proc. Nat. Acad. Sci. USA 87: 7512–16 (1990)
135. Pendland J. C., Heath M. A. and Boucias D. G.: Function of a galactose-binding lectin from Spodoptera exigua larval hemolymph: opsonization of blastopores from entomogenous hyphomycetes. J. Insect Physiol. 34: 533–540 (1988)
136. Pestka S. et al.: Interferons and their actions. Annual Rev. Biochem. 56: 727–777 (1987)
137. Peterson D. S., Wrightsman R. A. and Manning J. E.: Cloning a major surface-antigen gene of Trypanosoma cruzi and identification of a nonapeptide repeat. Nature 322: 566–568 (1986)
138. Phillips J. O. and Ourth D. D.: Isolation and molecular weight determination of two immunoglobulin heavy chains in the channel catfish, Ictalurus punctatus. Comp. Biochem. Physiol. 85B: 49–54 (1986)
139. van der Ploeg L. H. T.: Control of variant surface antigen switching in Trypanosomes (Minireview). Cell 51: 159–161 (1987)
140. Prat A. et al.: Nucleotide sequence of the Paramecium primaurelia G surface protein. A huge protein with a highly periodic structure. J. mol. Biol. 189: 47–60 (1986)
141. Qu X. M. et al.: Purification of a lectin from the hemolymph of chinese oak silk moth (Antherea pernyi) pupae. J. Biochem Tokyo 101: 545–552 (1987)
142. Ramaswamy H. and Hemler M. E.: Cloning, primary structure and properties of a novel human integrin β subunit. Embo J. 9: 1561–68 (1990)
143. Ravindranath M. H. et al.: Purification and characterization of an O-acetyl-sialic acid-specific lectin from a marine crab, Cancer antennarius. J. biol. Chem. 260: 8850–56 (1985)
144. Reyes A. A., Small S. J. and Akeson R.: At least 27 alternatively spliced forms of the neural cell adhesion molecule messenger RNA are expressed during rat heart development. Mol. cell. Biol. 11: 1654–61 (1991)
145. Rich K. A. et al.: Cell-adhesive motif in region II of malarial circumsporozoite protein. Science 249: 1574–77 (1990)
146. Richards E. H., Ratcliffe N. A. and Renwrantz L.: Isolation and characterization of a serum lectin from the giant stick insect Extatosoma tiaratum. Insect Biochem. 18: 691–700 (1988)
147. Richter R. F. and Ambrosius H.: The immune response of carp against dextran: an analysis of the idiotype heterogeneity. Develop. comp. Immunol. 12: 761–772 (1988)
148. Ripoche J. et al.: The complete amino acid sequence of human complement factor H. Biochem. J. 249: 593–602 (1988)
149. Robitzki A. et al.: cDNA structure and expression of calpactin, a peptide involved in Ca-dependent cell aggregation in sponges. Biochem. J. 271: 415–420 (1990)
150. Roch P., Canicatti C. and Valembrois P.: Interactions between earthworm hemolysins and sheep red blood cell membranes. Biochim. biophys. Acta 983: 193–198 (1989)
151. Rögener W., Renwrantz L. and Uhlenbruck G.: Comparison of hemolymph lectin from Octopus vulgaris with hemocyanin. Comp. Biochem. Physiol. Pt. B 85: 119–123 (1986)
152. Rogers J. H.: Mouse histocompatibility-related genes are not conserved in other mammals. Embo J. 4: 749–753 (1985)
153. Rosenshein I. L. et al.: Phylogenetic conservation of heavy chain determinants of vertebrates and protochordates. Develop. comp. Immunol. 9: 783–795 (1985)
154. Rosenshein I. L. and Marchalonis J. J.: The immunoglobulins of carcharhine sharks: a comparison of serological and biochemical properties. Comp. Biochem. Physiol. Pt. B 86: 737–747 (1987)
155. Roth C. et al.: Active late-appearing variable surface antigen genes in Trypanosoma equiperdum are constructed entirely from pseudogenes. Proc. Nat. Acad. Sci. USA 86: 9375–79 (1989)
156. Rottmann M. et al.: Specific phosphorylation of proteins in pore complex-laminae from the sponge Geodia cydonium by the homologous aggregation factor and phorbol ester. Role of protein kinase C in the phosphorylation of DNA topoisomerase II. Embo J. 6: 3939–44 (1987)

157. Rudenko G. et al.: Alpha-amanitin resistant transcription of protein coding genes in insect and bloodstream form Trypanosoma brucei. Embo J. 8: 4259–63 (1989)
158. Saad A. H. and Cooper E.: Evidence for a Thy-1-like molecule expressed on earthworm leucocytes. Zool. Sci. 7: 217–222 (1990)
159. Scallon B. J. et al.: A human immunoglobulin G receptor exists in both polypeptide-anchored and phosphatidylinositol-glycan-anchored forms. Proc. Nat. Acad. Sci. USA 86: 5079–83 (1989)
160. Schluter S. F. et al.: Evolution and immunoglobulin light chains: cDNA clones specifying sandbar shark constant regions. Proc. Nat. Acad. Sci. USA 86: 9961–65 (1989)
161. Schluter S. F. and Ey P. L.: Purification of three lectins from the haemolymph of the ascidian Botrylloides leachii. Comp. Biochem. Physiol. Pt. B 93: 145–146 (1989)
162. Schofield P. R. et al.: Molecular characterization of a new immunoglobulin superfamily protein with potential roles in opoid binding and cell contact. Embo J. 8: 489–495 (1989)
163. Schwager J. et al.: Genetic basis of the antibody repertoire in Xenopus: analysis of the VH diversity. Embo J. 8: 2989–3001 (1989)
164. Seeger M. A., Haffley L. and Kaufman T. C.: Characterization of amalgam: a member of the immunoglobulin superfamily from Drosophila. Cell 55: 589–600 (1988)
165. Shamblott M. J. and Litman G. W.: Complete nucleotide sequence of primitive vertebrate immunoglobulin light chain genes. Proc. Nat. Acad. Sci. USA 86: 4684–88 (1989)
166. Sharon N. and Lis H.: Lectins as cell recognition molecules. Science 246: 227–234 (1989)
167. Sharon N. and Lis H.: Lectins. Chapman and Hall, London 1989
168. Sheppard D. et al.: Complete amino acid sequence of a novel integrin β subunit (ß6) identified in epithelial cells using the polymerase chain reaction. J. Biol. Chem. 265: 11502–07 (1990)
169. Shigenaga T. et al.: Antimicrobial tachyplesin peptide precursor: cDNA cloning and cellular localization in the horseshoe crab (Tachypleus tridentatus). J. Biol. Chem. 265: 21350–54 (1990)
170. de Simone D. and Hynes R. O.: Xenopus laevis integrins. Structural conservation and evolutionary divergence of integrin β subunits. J. Biol. Chem. 263: 5333–40 (1988)
171. Smith S. H. and Jensen J. A.: The second component (C2n) of the nurse shark complement system: Purification, physico-chemical characterization and functional comparison with guinea pig C4. Develop. comp. Immunol. 10: 191–206 (1986)
172. Snow P. M., Bieber A. J. and Goodman C. S.: Fasciclin III: A novel homophilic adhesion molecule in Drosophila. Cell 59: 313–323 (1989)
173. Solheim B. G., Moller E. and Ferrone S. (eds.): HLA II antigens. A comprehensive review of structure and function. Springer, Berlin 1986
174. Southan C. et al.: Amino acid sequence of β-galactoside-binding bovine heart lectin: member of a novel class of vertebrate proteins. FEBS Letters 214: 301–304 (1987)
175. Spiess M.: The asialoglycoprotein receptor: A model for endocytic transport receptors. Biochemistry 29: 10009–18 (1990)
176. Srimal S. et al.: A new haemagglutinin from the amoebocytes of the horseshoe crab Carcinoscorpius rotundicauda – Purification and role in cellular aggregation. Biochem. J. 230: 321–327 (1985)
177. Steen M. L., Hellman L. and Pettersson U.: The immunoglobulin lambda locus in rat consists of two C-lambda genes and a single V-lambda gene. Gene 55: 75–84 (1987)
178. Stein E. A., Younai S. and Cooper E. L.: Bacterial agglutinins of the earthworm, Lumbricus terrestris. Comp. Biochem. Physiol. Pt. B 84: 409–415 (1986)
179. Stevens F. J., Solomon A. and Schiffer M.: Bence Jones proteins. A powerful tool for the fundamental study of protein chemistry and pathophysiology. Biochemistry 30: 6803–05 (1991)
180. Sun S. C. et al.: Hemolin: An insect-immun protein belonging to the immunoglobulin superfamily. Science 250: 1729–32 (1990)
181. Sun S. C. et al.: Structure and expression of the attacin genes in Hyalophora cecropia. Eur. J. Biochem. 196: 247–254 (1991)
182. Suzuki S. and Naitoh Y.: Amino acid sequence of a novel integrin β subunit and primary expression of the mRNA in epithelial cells. Embo J. 9: 757–763 (1990)
183. Suzuki T. and Mori K.: A galactose-specific lectin from the hemolymph of the pearl oyster, Pinctada fucata martensii. Comp. Biochem. Physiol. Pt. B 92: 455–462 (1989)
184. Suzuki T. et al.: A calcium-dependent galactose-binding lectin from the tunicate Polyandrocarpa misakiensis. Isolation, characterization, and amino acid sequence. 265: 1274–81 (1990)
185. Takeichi M.: Cadherins: A molecular family important in selective cell-cell adhesion. Annual Rev. Biochem. 59: 237–252 (1990)
186. Tillinghast J. P., Behlke M. A. and Loh D. Y.: Structure and diversity of the human T-cell receptor β-chain variable region genes. Science 233: 879–883 (1986)
187. Titani K. et al.: Amino acid sequence of sialic acid binding lectin from frog (Rana catesbeiana) eggs. Biochemistry 26: 2189–94 (1987)
188. Tjoelker L. et al.: Evolutionary conservation of antigen recognition: The chicken T-cell receptor β chain. Proc. Nat. Acad. Sci. USA 87: 7856–60 (1990)
189. Tosi M. et al.: Complete cDNA sequence of human complement C1s and close physical linkage of the homologous genes C1s and C1r. Biochemistry 26: 8516–24 (1989)
190. Traboni C and Cortese R.: Sequence of a full length cDNA coding for human protein HC (α1-microglobulin). Nucleic Acids Res. 14: 6340 (1986)
191. Valenzuela D., Weber H. and Weissmann C.: Is seqence conservation in interferons due to selection for functional proteins? Nature 313: 698–700 (1985)
192. Villalta F., Lima M. F. and Zhou L.: Purification of Trypanosoma cruzi surface proteins involved in adhesion to host cells. Biochem. biophys. Res. Commun. 172: 925–931 (1990)
193. Vincek V. et al.: Evolutionary expansion of MHC class I loci in the mole-rat, Spalax ehrenbergi. Mol. Biol. Evol. 4: 483–491 (1987)

194. Wall R.: Immunoglobulin RNA processing. Trends biochem. Sci. 5: 325–327 (1980)
195. Watkins D. I., Hodi F. S. and Letvin N. L.: A primate species with limited major histocompatibility complex class I polymorphism. Proc. Nat. Acad. Sci. USA 85: 7714–18 (1988)
196. Weissman A. M. et al.: Molecular cloning and chromosomal localization of the human T-cell receptor zeta chain: Distinction from the molecular CD3 complex. Proc. Nat. Acad. Sci. USA 85: 9709–13 (1988)
197. Wicker C. et al.: Insect immunity: Characterization of a Drosophila cDNA encoding a novel member of the diptericin family of immune peptides. J. Biol. Chem. 265: 22493–98 (1990)
198. Wilson M. E. et al.: The immunoglobulin M heavy chain constant region gene of the channel catfish, Ictalurus punctatus: an unusual mRNA splice pattern produces the membrane form of the molecule. Nucleic Acids Res. 18: 5227–33 (1990)
199. Yamawaki-Kataoka Y. and Honjo T.: Nucleotide sequences of variable region segments of the immunoglobulin heavy chain of Xenopus laevis. Nucleic Acids Res. 15: 5888 (1987)
200. Yoshizaki N.: Functions and properties of animal lectins. Zool. Sci. 7: 581–591 (1990)
201. Zamze S. E. et al.: Characterisation of the asparagine-linked oligosaccharides from Trypanosoma brucei type-I variant surface glycoproteins. Eur. J. Biochem. 187: 657–663 (1990)
202. Zheng Q. X. et al.: Characterization of cDNAs of the human pregnancy-specific β1-glycoprotein family, a new subfamily of the immunoglobulin gene superfamily. Biochemistry 29: 2845–52 (1990)

7 Respiratory Pigments

In the animal kingdom there are four types of O_2-binding (respiratory) pigment with different structures but very similar functional properties. They have characteristic colours in their oxygenated states and the absorption spectra of the pigments with bound O_2 or another ligand, such as CO or CN^-, are used for purposes of identification (Table 7.1). The structures of the **binding sites** vary (Fig. 7.1): the prosthetic group of the globins is protohaem, i.e. Fe(II)-protoporphyring (Fig. 7.2), which can bind one ligand. Chlorocruorin is also a haemoprotein but with a haem component (spirographis haem) which differs from protohaem in one substituent (Fig. 7.2). In the copper protein haemocyanin and the iron protein haemerythrin, the binding site in each case contains two metal atoms (Fig. 7.1). Chlorocruorins and haemocyanins are always found dissolved in the blood plasma; haemerythrins occur only intracellularly, and haemoglobins are both intra- and extracellular (Table 7.2). The intracellular respiratory pigments consistently have molecular masses under 100 kDA and only one to eight O_2-binding sites per molecule. Most of the extracellular blood pigments have far larger molecular masses of up to several million kDA and often more than 100 O_2-binding sites; in this way, the colloid osmotic effects in the blood plasma are reduced. There are, however, some exceptions to this rule, e.g. the extracellular haemoglobins of chironomid larvae are only 16–32 kDA.

Haemocyanins and haemerythrins are found in relatively few groups of animals, and chlorocruorin is in fact restricted to just a few polychaete families. In contrast, the globins are found not only in almost all vertebrates and representatives of many invertebrate lines (Table 7.2), but also in higher plants and even bacteria. Various invertebrate species have both myoglobin in the muscles and chlorocruorin or haemocyanin in the blood. Despite far-reaching differences in sequence, the **globins of animals, plants and bacteria** may be considered homologous, with a com-

Table 7.1. Animal respiratory pigments: changes in the colour and the absorption spectra upon deoxygenation

	Colour	Absorption maxima
Haemoglobin		
Oxy-	Bright red	α = 574–581 nm; β = 538–545 nm
Deoxy-	Crimson	Broad maximum of 556–565 nm
Chlorocruorin		
Oxy-	Green	α = 604–605 nm; β = 558–560 nm
Desoxy-	Green	About 525 nm
Haemocyanin		
Oxy-	Blue	Molluscs 570–580 nm Crustaceans 557–559 nm
Deoxy-	Colourless	–
Haemerythrin		
Oxy-	Violet	330 nm; further peak at about 500 nm
Desoxy-	Colourless	–

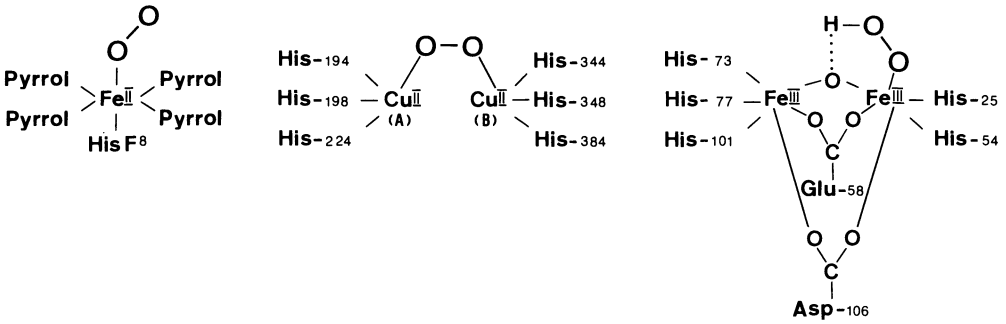


Fig. 7.1. O_2-binding sites of respiratory pigments in the oxygenated state, illustrated by the basic structure of *(left)* haemoglobin, *(middle)* the haemocyanin subunit Pint-a of the spiny lobster *Panulirus interruptus*, and *(right)* the haemerythrin of the echiurid *Themiste dyscritum* [108]

mon origin in a cytochrome-b-like haemoprotein. The globins of legume root nodules (leghaemoglobins) have been known for a long time; the soybean globin has 22 amino acids in common with the α-chain of human haemoglobin. More recently, globins have been detected and characterized in various non-legumes: in the root nodules of *Parasponia* (Ulmacea) and *Casuarina* (Casuarunacea) harbouring nitrogen-fixing bacteria, and in the nodule-free roots of *Trema*, a genus related to *Casuarina*. The haemoglobins of legumes and non-legumes have about 40–50 % amino acids in common [15, 90]. The sporadic occurrence of globins in higher plants has led to speculation of a possible gene transfer between animals and plant. In the meantime, however, a dimeric haemoprotein has been isolated from *Vitreoscilla*, a thread-forming bacterium of the *Beggiatoa* group; this haemoprotein agrees in 34 of its 146 amino acids (24 %) with leghaemoglobin II from the yellow lupin, and apparently serves as an O_2 store during transient oxygen shortage. Globin has also been found in the free-living bacteria of the genus *Rhizobium* [181]. Chlorocruorins and haemoglobins are so similar in their structure as to be assumed homologous; the haemocyanins and haemerythrins represent two separate groups of homologous proteins. The similarities in the functional characteristics of all respiratory pigments are consequently the result of convergent evolution in three different protein super-families [108, 112 174].

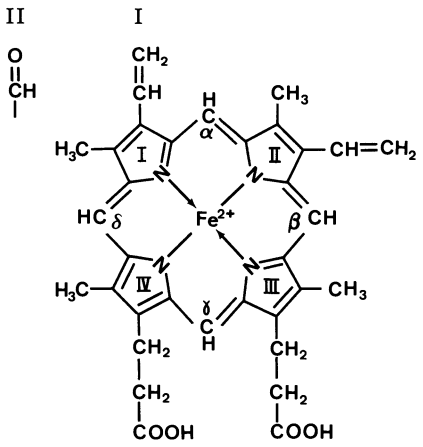


Fig. 7.2. The prosthetic groups of the haemoglobins (protohaem) and the chlorocruorins (spirographis haem) differ only in the substituents at position 2 of the porphyrin ring. Protohaem has a vinyl residue *(I)* and spirographis haem a formyl residue *(II)*

The primary **biological task of the respiratory pigments** is one of transport. The O_2 capacity of the blood depends upon the pigment concentration and varies, e.g. in the vertebrates, between 1 (agnathans) and 36 vol % (in the seal *Cystophora*). The capacity of invertebrate blood is generally in the range 1–5 vol % and seldom higher than 10 vol % (larvae of non-biting midges *Chironomus*, the annelids *Arenicola* and *Megascolex*, and the pogonophore *Riftia*). Respiratory pigments can serve as O_2 stores during periods of transient oxygen deficiency. Myoglobin accelerates O_2 diffusion in muscle. Such pigments, especially those dissolved in the blood, may have non-respiratory functions. Because they are often the predominant extracellular protein, they may be the most important factor in the buffering capacity and colloid-osmotic pressure. Finally, they represent a potential nutritional reserve which is important, for example, during metamorphosis of the chironomids or ecdysis of the crustaceans.

Despite differences in the atomic structure of the O_2-binding sites, the process of O_2 exchange is similar for all the respiratory pigments. A complete description of the functional properties of a respiratory pigment is contained in its **O_2-binding curve**; an approximate measure of O_2 affinity is

Table 7.2. The distribution of respiratory pigments in the animal kingdom

Haemoglobins (intra- or extracellular)	
Bacteria	In *Vitreoscilla* and *Rhizobium*
Flowering plants	In the Leguminosae, Ulmaceae and Casuarinaceae, in root nodules and nodule-free roots
Ciliates	In *Paramecium* and *Tetrahymena*
Plathelminthes	In the parenchyme and muscles of the pharynx of turbellarians and trematodes
Nemerteans	Soluble in the blood and in nerve cells
Nemathelminthes:	
Nematoda	Soluble in the body cavity fluid and in various body cells
Gastrotricha	In special cells of *Neodysis*
Molluscs	
Polyplacophores	In muscles
Gastropods	Soluble in the haemolymph and in muscles
Bivalves	Soluble in blood, blood cells and muscle
Annelids	Soluble in the blood or in coelom fluid, in cells of the coelom fluid, in muscles and in nerve cells
Echiurids	In cells of the coelom fluid, in various body cells and in the eggs
Crustacea	Soluble in the haemolymph of Copepoda, Ostracoda, Phyllopoda and Anostraca; also in muscle, fat and egg cells of *Daphnia* and other species
Insects	Soluble in the haemolymph of chironomid larvae; in the tracheal cells of several diptera and hemiptera; and in gland cells and eggs of several other insects
Tentaculates	In the blood cells of Phoronida
Echinoderms	In cells of the coelom fluid in Holothurioidea and Ophiuroidea
Pogonophores	Dissolved in the blood and the coelom fluid of *Riftia pachyptila* and *Lamellibrachia* sp.
Vertebrates	In blood cells, and skeletal and heart muscle
Chlorocruorins (exclusively extracellular)	
Polychaetes	Only in the families Sabellidae, Serpulidae, Chlorhaemidae and Ampharetidae
Haemocyanins (exclusively extracellular)	
Molluscs	In polyplacophores, gastropods, cephalopods and several bivalves
Chelicerates	In Xiphosura, Scorpiones, Uropygi, Amblypygi and Aranea
Crustacea	In the isopods and decapods
Chilopods	In *Scutigera coleoptrata*
Haemerythrins (exclusively intracellular)	
Annelids	In blood cells of the polychaete *Magelona*
Priapulids	In cells of the coelom fluid
Sipunculids	In cells of the coelom fluid, blood cells and muscles
Brachiopods	In cells of the coelom fluid

given by the p50, i.e. the oxygen partial pressure at which the O_2-binding sites are 50 % occupied. The p 50 values, even of respiratory pigments with the same prosthetic group, vary greatly, e.g. from 2 Pa for the haemoglobin from the body cavity of the roundworm *Ascaris* to more than 100 000 Pa for certain fish haemoglobins at low pH; thus, the protein structure has a large influence on the O_2 affinity of the prosthetic group. The O_2-binding curve of respiratory pigments with just one binding site per molecule is always a rectangular hyperbola, and that of pigments with more than one site is often S-shaped (sigmoid), due to positive homotropic interaction between the binding sites **(cooperativity)**. The steepness of a sigmoid binding curve in the region of oxygen release into the tissue contributes to the maintenance of constant tissue O_2 (Fig. 7.3). The properties of many respiratory pigments are influenced by allosteric modulators like H^+, CO_2, inorganic ions or organophosphates (heterotropic interactions).

7.1 The Haemoglobins and Myoglobins of Vertebrates

7.1.1 The Structure of Haemoglobin

Almost all vertebrates possess red blood cells, which in the case of the mammals lack nuclei. Only antarctic fish of the family Chaenichthyidae have neither erythrocytes nor haemoglobin. The concentration of haemoglobin in human **erythrocytes** is close to the limit of solubility at 330 mg/ml. In the anomalous Hb C, the amino acid substitution βA3-Glu to Lys somewhat lowers the solubility; Hb C in the erythrocytes of Hb C homozygotes can crystallize out [66]. The

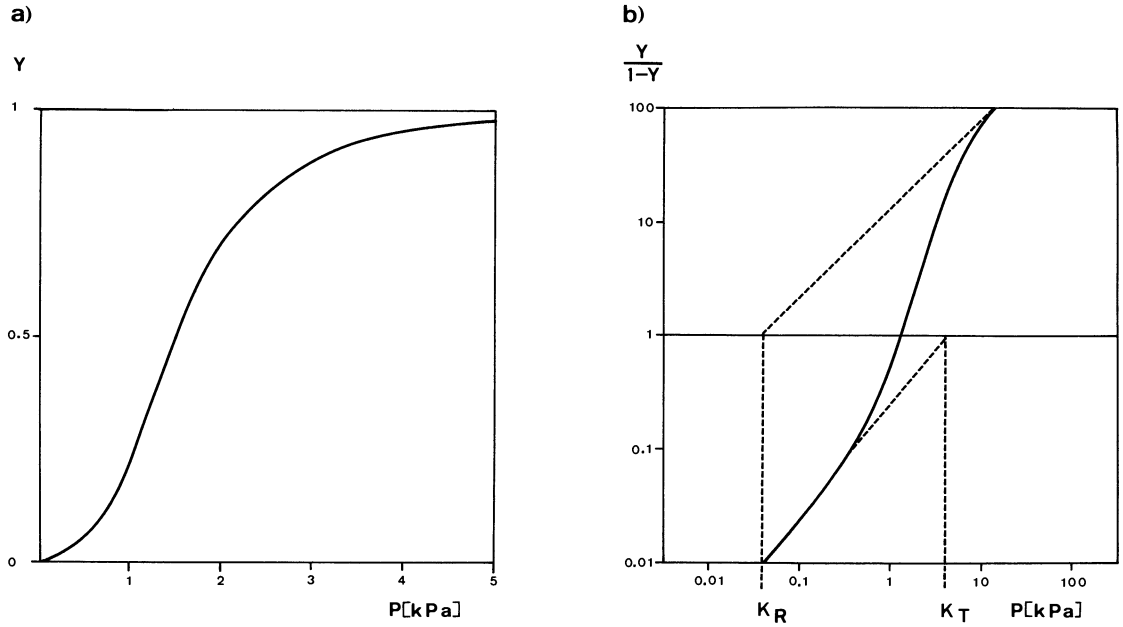


Fig. 7.3a, b. The O_2-binding curves of respiratory pigments represented in different ways. **a** The proportion of pigment loaded with O_2 (Y) as a function of the O_2 partial pressure (P). **b** The Hill plot of log (Y)/(1-Y) as a function of log P. The Hill constant n is the slope of the curve and is usually given for Y/(1-Y) = 1 or log (Y)/(1-Y) = 0, i.e. at 50 % saturation; K_R and K_T are the dissociation constants for the R and T forms

haemoglobin content of the blood in mammals is 100–220, in birds 70–190, in reptiles 50–100, in amphibians 20–70, and in fish 30–150 g/l; given that 1 g haemoglobin can bind 1.34 ml O_2, the corresponding oxygen capacity of the blood can be calculated. The longevity of the erythrocytes is 700–1400 days in amphibians, 600–800 days in reptiles and 120 days in man; the life-span is shorter in birds than in mammals. These data give an indication of the rate of haemoglobin synthesis required for the replacement of erythrocytes, viz. about 6 g per day in man.

The haemoglobins of almost all vertebrates are tetramers of two different subunits of 141–147 amino acids according to the general formula $\alpha_2\beta_2$; the haemoglobins of the agnathans are unique in being monomers in the oxygenated state and di- or tetramers when deoxygenated. The **amino acid sequences** are known for several hundred haemoglobin chains of vertebrates from all classes. In spite of extensive differences in sequence, the polypeptides have similar **secondary and tertiary structures** (Fig. 7.4). The β-chain consists of eight helical sections A–H; helix D is missing in the α-chain (Fig. 7.5) [87]. The intervening, non-helical regions are designated AB, BC, etc. The short N-terminal segment in front of helix A is termed NA and the corresponding C-

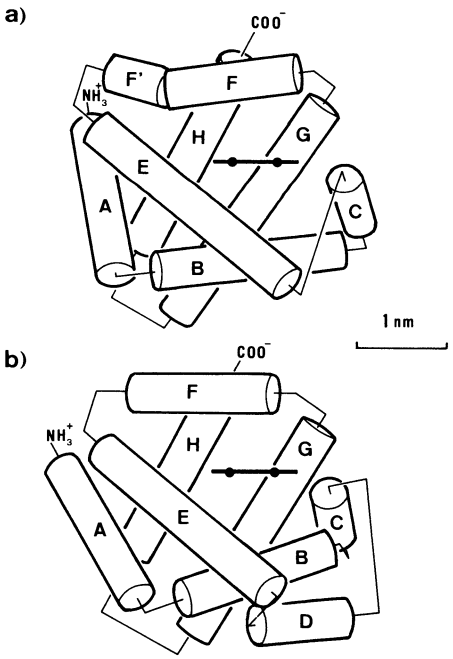


Fig. 7.4a, b. The spatial structures of **a** the human haemoglobin α-chain, and **b** a haemoglobin chain from the larvae of the non-biting midge *Chironomus thummi*. The *cylinders* indicate the actual relative lengths and positions of the α helices, whereas in the interhelical sections the form of the chain is shown only schematically. The *horizontal lines with two points* represent the haem groups [101]

Fig. 7.5. The amino acid sequences of the α- and β-chains of human haemoglobin. *A–H* indicate the α helices

terminus is HC. The amino acids are numbered according to the segment; this has the advantage over consecutive numbering of the whole chain that the key positions for particular functions have the same designation, despite individual insertions or deletions. Thus, the haem-binding proximal histidine in the human haemoglobin chain (Fig. 7.5) corresponds to positions α97-His and β102-His but because in both cases it is located at the eight position of helix F, it can be designated simply F8-His. There are, of course, very many monographs on the structure and function of haemoglobin, and in particular of the human variety [28].

Aggregation of the haemoglobin tetramers to form higher polymers is found in many birds and amphibians as well as in several species of elasmobranchs, teleosts and reptiles. In many cases, this is clearly due to the formation of disulphide bridges between the tetramers. In fact, many of the haemoglobins with a tendency to aggregate are rich in cysteine; the record is held by the marine teleost *Lophius americanus* with 16 reactive cysteine-SH groups per tetramer [18]. Disulphide bridges forming spontaneously in intact erythrocytes are always immediately reduced again. Several human and murine anomalous haemoglobins have a high cysteine content and a tendency towards polymerization. The most investigated example is that of haemoglobins B and C from the American bullfrog *Rana catesbeiana*; on deoxygenation these form an aggregate BC_2 with a drastically reduced O_2 affinity. In this case, the β-chains of both haemoglobins are identical. The cysteine residues in positions β 122 and β 129 are excluded, on geometrical grounds, from the formation of disulphide bridges between the two chains; therefore the aggregation apparently involves interaction between the different α-chains of the two haemoglobins [62, 139].

If the binding forces between the haemoglobin subunits are weakened by high concentrations of neutral salts, propylurea or guanidine hydrochloride, the tetramers dissociate to αβ dimers or to monomers. Because the amino acids involved in the interactions between the subunits show large species-specific differences, the ease with which different vertebrate haemoglobins can be brought to **dissociation** also shows large variation. The haemoglobins of fish are especially resistant; in contrast, the haemoglobins of the turtle *Phrynops hilarii* dissociate simply on dilution of the haemolysate [136].

7.1.2 Ligand Binding and Cooperativity

The blood of almost all vertebrates shows the typical sigmoid O_2-binding curve of cooperative haemoglobins. The **mechanisms of homotropic interaction** between subunits or binding sites (cooperativity) can be described in abstract mathematical models using thermodynamic or kinetic data. For vertebrate haemoglobin in particular, the underlying molecular processes are also known in some detail down to the atomic level. Cooperativity is a general phenomenon observed for many proteins with several binding sites, and is especially relevant in the case of allosterically regulated enzymes. The basic theories of cooperativity were developed for vertebrate haemoglobin, in particular human Hb A. It was because of the importance of these theories for understanding allosteric enzyme modulation that Monod jokingly suggested naming haemoglobin an "honorary enzyme".

The **cooperativity theories** [13] start with the simple case of the binding of a ligand A to a monomeric protein E:

$$E + A \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} EA. \quad (7.1)$$

According to the mass effect theory, the binding constant is

$$K_a = \frac{k_{on}}{k_{off}} \frac{[EA]}{[E] \cdot [A]} \text{ l/mol.} \quad (7.2)$$

Instead of the binding constant K_a, one can give the inverse, i.e. the dissociation constant $K_d = 1/K_a$ in mol/l. The first mathematical model for the binding of A to a protein with n binding sites was derived by Hill in 1909:

$$E + nA \rightleftharpoons EA_n \quad (7.3)$$

$$K_a = \frac{[EA_n]}{[E] \cdot [A]^n}. \quad (7.4)$$

Where Y is the completely occupied proportion of the protein E, then

$$\frac{Y}{1-Y} = [A]^n \cdot K_a \quad (7.5a)$$

$$\log \frac{Y}{1-Y} = n \log [A] + \log K_a. \quad (7.5b)$$

According to the Hill equation (7.5b), log [Y/1–Y] plotted against log [A] (a Hill plot) gives a straight line whose slope n (the Hill constant) corresponds to the number of binding sites. However, this is not the case for either the tetrameric haemoglobins or other respiratory pigments (Fig. 7.3); more usually, the slope n gradually decreases up to extreme values of log [A] and never reaches a value on the curve equal to the number of subunits. The Hill constant is usually read at 50 % saturation, i.e. at 0 on the ordinate; for the tetrameric human Hb A the value obtained is 3.0. Equations (7.5a) and (7.5b) assume that all the A's are bound simultaneously, i.e. that only E or EA_n exist. However, binding is apparently also simultaneous if the affinity of E is very low but is increased by the binding of an A. All the following A's will bind rapidly, and cooperativity results. Thus, the Hill constant, n, does not correspond to the number of binding sites but is more a measure of the degree of interaction between the subunits [13].

The **model of Adair (1925)** assigns a different K_a value to each binding step. In this case, the O_2-binding curves may be described perfectly but the individual K_a values cannot be directly measured. The sequential **model of Koshland (1958)** assumes that each subunit can exist in two states, T and R, with different affinities. Somewhat simpler is the model proposed by Monod, Wyman and Changeux in 1965 **(MWC model)**; this assumes two states, T and R, of the whole molecule with spontaneous transition between the two. The ratio of the concentrations of the two forms in the absence of the ligand A is given as L, and the ratio of their binding constants as c:

$$L = \frac{[T_0]}{R_0]} \quad (7.6)$$

$$c = \frac{K_R}{K_T}. \quad (7.7)$$

Assuming $\alpha = [A]/K_R$, then

$$Y = \frac{Lc\alpha(1 + c\alpha)^{n-1} + \alpha(1 + \alpha)^{n-1}}{L(1 + c\alpha)^n + (1 + \alpha)^n}. \quad (7.8)$$

A sigmoid binding curve then results only if $n > 1$, $L > 1$ and $c \ll 1$ [13].

Instead of using oxygen, investigations of the molecular mechanisms of ligand binding have often made use of carbon monoxide, for which vertebrate haemoglobins have a 60- to 550-fold higher affinity. The CO complex can be dissociated by a light flash and this method therefore allows measurements of the uptake kinetics in the millisecond range. The **alterations on haemoglobin structure after ligand binding** are known down to the atomic level, thanks largely to the pioneering efforts of Perutz and coworkers [131, 132]. The haemoglobin tetramer consists of two $\alpha\beta$ dimers, the chains of which are held together by so-called $\alpha^1\beta^1$ contacts (packing contacts). On the binding of a ligand, the $\alpha^1\beta^1$ dimer rotates 15° relative to the $\alpha^2\beta^2$ dimer, and shifts 0.1 nm; responsible for this are the so-called $\alpha^1\beta^2$ contacts (sliding contacts) between the two dimers. Dimers and homotetramers are not cooperative. In deoxygenated haemoglobin, the distance between the β-chains is increased so that anions, such as organophosphates or chloride ions, can bind to the positively charged side-chains. Substitutions in the $\alpha^1\beta^2$ contact region can reduce cooperativity [131, 132].

In terrestrial vertebrates, the oxygenation of haemoglobin is, as a rule, sufficiently accurately described by the MWC model. Deoxygenated haemoglobin is almost completely in the T state and oxygenated haemoglobin in the R state; in marine turtles, HbO_2 has been found in the T state [136]. The value of the **Hill constant** for the haemoglobins of mammals is always in the region $n = 3.0$; this is also true for most reptiles and amphibians, which show values under 2.0 only at low pH. Hill constants over 4.0, as are found in all bird haemoglobins, the haemoglobins of *Rana catesbeiana* and other frogs and several reptiles,

are related to tetrameric aggregation [139]. The haemoglobin of the semi-aquatic snake *Liophis miliaris* dissociates to dimers on oxygenation. Sequence comparisons with human Hb A show five amino acid differences in the $\alpha^1\beta^2$ contact region, including the exchanges βG3Glu → Val and βCD2Glu → Thr. It is probably this loss of two negative charges that is responsible for the lower stability of binding between the dimers [45, 118]. In fish, there are haemoglobins whose properties are not in total agreement with the MWC mode. They show Hill constant values as low as 1.0, particularly in acid conditions, and this indicates a marked inhibition of the T → R transition [21]. Hill constants under 1.0 are also possible if the O_2 affinities of th α- and β-chains differ greatly [139]. According to the MWC model, heterotropic ligands should change only the L values; K_T and K_R should remain constant. However, in the case of the haemoglobins of several fish species and mammals, including man, organophosphates not only increase the L value but also reduce K_T; K_R remains unchanged and the T → R transition is hindered. This effect can be explained by a three-state model, according to which there exists, in addition to T and R, a third state, S, in which the low affinity state T is stabilized by the organophosphate [75]. More recent cooperativity models are based upon the MWC model but take into account the thermodynamic equivalence of the salt bridges between the dimers, as postulated by Perutz, and allow a quantitative description of homotropic and heterotropic interactions in vertebrate haemoglobin [98].

The only vertebrate haemoglobins which are not built on the tetramer model described above are from the **agnathans**, the lampreys (Petromyzontoidae) and the hagfish (Myxinoidae). The marine lamprey *Petromyzon marinus* has a principal haemoglobin (Hb II) and two electrophoretically distinct subsidiary components (Hb I and Hb III), all of which are monomeric in the oxygenated state but can aggregate to dimers after oxygenation at weakly acidic pH:

$$G_2 + 2O_2 \leftrightarrow 2GO_2.$$

This **change in the aggregation state** with oxygenation or deoxygenation may be considered an extreme case of conformational change as the result of homotropic interaction; that is why these haemoglobins also show cooperativity. The reason why they cannot form tetramers can be found in their amino acid sequence. All three haemoglobins lack nine amino acids in the GH region, and in several positions important for $\alpha^1\beta^1$ contacts there are more strongly polar amino acids to be found than in the other vertebrates; on the other hand, the $\alpha^1\beta^2$ contacts are more similar to those of the mammals. Thus, these haemoglobin chains can be linked by contacts which correspond to the $\alpha^1\beta^2$ contacts of the tetrameric vertebrate haemoglobins, but the $\alpha^1\beta^1$ contacts necessary for tetramer formation cannot be produced. The haemoglobins of the marine lamprey *(P. marinus)* and the river lamprey *(Lampetra fluviatilis)* differ at only three to five positions. In contrast, the haemoglobin sequences of the common hagfish *Myxine glutinosa* and the Petromyzontidae differ by 87–90 amino acids. The haemoglobin of *Eptatretus stoutii* (Myxinoidae) appears to exist as monomers in the presence or absence of oxygen [65, 67]; on the other hand, the deoxyhaemoglobins of the related species *E. burgeri* and *E. cirrhatus* have a tendency to aggregate at low pH and high protein concentration [20].

The $\alpha^1\beta^2$ contacts in the haemoglobins of the Petromyzontidae may be looked upon as the first **evolutionary step** from the monomeric globins, resembling myoglobin, to the tetrameric vertebrate haemoglobins. This first step already allowed cooperative and allosteric interactions. In the following evolutionary step, two chains arose after a gene duplication; these could form heterodimers and- tetramers, and allowed finer grading of the homotropic and heterotropic interactions. The "discovery" and refinement of cooperativity and the heterotropic modulation of vertebrate haemoglobins should be considered in the light of the increase in metabolism, which required high and constant oxygen partial pressures in the tissues as independent as possible from O_2 use and CO_2 production. Further gene duplications resulted in the appearance of many α- and β-like chains, from which a whole spectrum of haemoglobins with varying properties could be produced. Finer adaptations to various internal and external conditions consequently became possible. The course of haemoglobin evolution can be illustrated by a molecular genealogical tree (Fig. 7.6).

7.1.3 Heterotropic Interactions

Vertebrate haemoglobins show an extraordinary variety of functional characteristics, such as their intrinsic O_2 affinity and their sensitivity to allosteric modulation, which allow adaption to very different internal and external conditions. **Allos-**

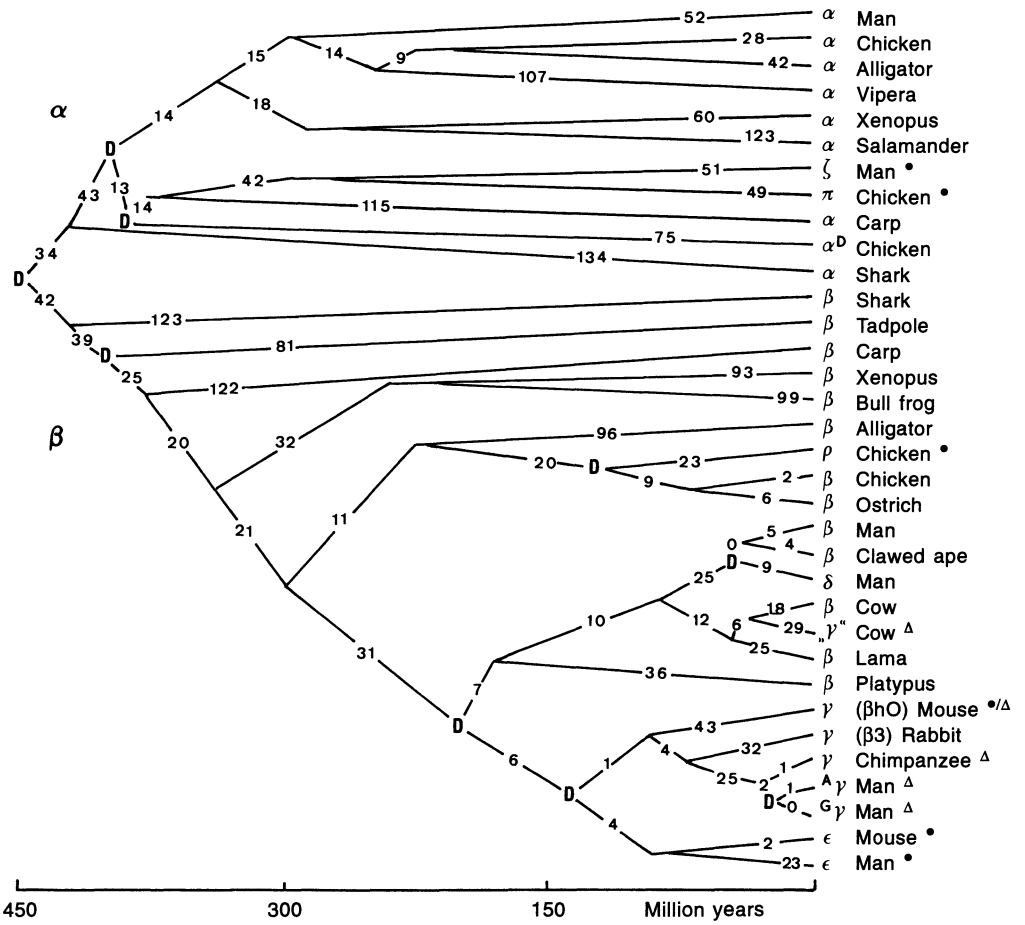


Fig. 7.6. A genealogical tree of the vertebrate globins [53]. *Open triangles* denote foetal globin chains and *filled circles* embryonal globin chains. The *numbers* indicate the number of substitutions calculated by the maximum parsimony method; gene duplications are shown by *D*

teric modulators, like H^+, inorganic ions, organophosphates and CO_2, reduce O_2 affinity because they bind more strongly to the T form and have a stabilizing effect by creating electrostatic interactions (salt bridges) between the subunits. The allosteric effectors interact with particular amino acid residues (Table 7.3); thus, in the evolution of haemoglobins, substitution of one or a few of these key positions changed the respiratory properties of the haemoglobin, whilst exchange of other amino acids is without functional effect and therefore more or less selectively neutral [131, 132, 183].

The O_2 affinity of haemoglobin in terrestrial vertebrates is diminished by H^+ ions under physiological conditions. This effect, which facilitates oxygen off-loading in tissues and oxygen uptake in the lungs, is known as the (alkaline) **Bohr effect**. Below pH 6.5, however, there is an increase in O_2 affinity by H^+ ions: this is the "acidic" or "inverse" Bohr effect. In both cases, the Hill constant remains unchanged. A quantitative measure of the change in O_2 affinity by protons is the quotient $\varphi = \Delta \log p50/\Delta pH$, which is negative for the alkaline and positive for the acidic Bohr effect. For most vertebrates, φ under physiological conditions has a value between 0 and –1.0; extreme values up to –2.0 are encountered with some fish haemoglobins. Here, the acidic Bohr effect is missing and the O_2 affinity declines markedly with decreasing pH; air-saturated haemoglobin releases more than half the bound oxygen on reduction of the pH from 8 to 6. The haemoglobin is not completely loaded with O_2 even at partial pressures of 10 mPa (100 bar) **(Root effect)**. At the same time, n declines to 1.0, or even lower, because the T state is so stabile that the haemoglobin tetramer is not transformed to the R state by the binding of the first O_2 molecule. Values of n lower than 1.0 are only possible where the α- and β-chains have different O_2 affinities, as for example in the carp and

Table 7.3. The amino acids responsible for heterotropic interactions in various vertebrates [131]. The latest results showing no Root effect in *Xenopus* haemoglobin [26] are taken into account

	Hum	FIV	FI	Xen	R.ad	R.la	Lep
β Chain							
NA2	His	Asp	Glu	Gly	–	His	His
EF6	Lys	Lys	Leu	Lys	Lys	Lys	Lys
F6	Glu	Val	Glu	Lys	Glu	Glu	His
F9	Cys	Ser	Ala	Ser	Ser	Ala	Ser
FG1	Asp	Glu	Asn	Glu	Gly	Asn	Glu
H21	His	Arg	Ser	Lys	Lys	His	Arg
HC1	Lys	Gln	Arg	Gly	Gly	Ser	Glu
HC3	His	His	Phe	His	His	His	His
α Chain							
NA1	Val	aSer	aSer	Leu	?	aSer	mArg
C3	Thr	Gln	Gln	Lys	?	Gln	Gly
Bohr effect	Norm	High	None	Norm	?	Inv	Norm
Root effect	None	Pres	None	None	None	None	None

Hum, human; FIV and FI, rainbow trout *Salmo irideus (S. gairdneri)* Hb IV and HB I, respectively; Xen, *Xenopus laevis* (adult), R.ad and R.la, frog *Rana catesbeiana* adult and larval, respectively; Lep, lungfish *Lepidosiren paradoxus; aSer, N-acetyl-serine; mArg, methylarginine; Norm, normal; Inv, inverse; Pres, present*

the tuna fish *Thunnus thynnus* [21, 131]. The stability of the T state at low pH makes Root-effect haemoglobins extremely interesting for studies of the R–T transition [64]. The biological importance of the Root effect was initially thought to lie in the achievement of higher gas pressure in the swim bladder by counter-current multiplication in the rete mirabile of the gas gland, a mechanism described in all fish physiology textbooks. However, Root-effect haemoglobins are also found in fish lacking a functional gas gland. In this case, they are involved in the O_2 supply to the eye by a multiplication effect in the network of blood vessels in the choroid (rete chorioidale) [21, 38, 131].

The **salt bridges** responsible for the Bohr effect in human Hb A are depicted in Fig. 7.7; similar bonds are found in all Bohr-effect haemoglobins [139]. The pK values of the amino acids involved increase on deoxygenation and protons become bound. Thus, O_2 binding and H^+ dissociation are linked: on the one hand, O_2 binding is pH dependent (Bohr and Root effects) and, on the other hand, protons are released by oxygenation (Haldane effect) and bound by deoxygenation. This promotes HCO_3^- formation in the tissues and CO_2 release in the lungs. According to Perutz, the Root effect arose due to the exchange of βF9-Cys for a serine, which allows two further H-bridges to be formed with βHC3-His, thereby stabilizing the T state, increasing L and reducing K [131]. Apart from βF9-Ser, Root-effect haemoglobins usually also contain βFG1-Glu, βH21-Arg and βHC3-His. The connection between the amino acid sequence and the occurrence of the Root effect is, however, not absolute. The haemoglobins of the frogs *Xenopus laevis*, *Rana esculenta* and *R. catesbeiana*, the lungfish *Lepidosiren paradoxus* and the electric eel *Electrophorus electricus* all contain βF9-Ser but exhibit no Root effect [69]. In the genus *Rana* this is explainable by the occurrence of a glycine or asparagine instead of a glutamine at position βFG1 (Table 7.3); as a result, βHC3-His forms a weak salt bridge to βF6-Glu instead of forming the stronger link to βFG1-

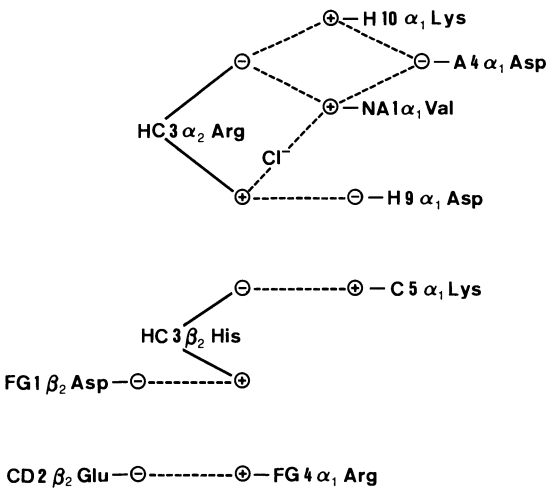


Fig. 7.7. Salt bridges in human deoxyhaemoglobin [131]. *(–)* denotes a carboxyl group, and *(+)* an amino group, a guanidine group or the imidazole ring. Amino acid residues are numbered according to the chain regions (see p. 253)

Glu [131, 132]. *Xenopus* haemoglobin contains both βFG1-Glu and βF9-Ser and, despite previous findings to the contrary, still shows no Root effect [26]. The haemoglobin from *Lepidosiren* has the same amino acid composition and also exhibits no Root effect. Thus, although it is possible to assign certain amino acids to key positions in the haemoglobin molecule [131, 132], the functionally important properties of the haemoglobin are nevertheless determined by the structure of the whole molecule [21].

The goldfish possesses a single haemoglobin showing a marked Root effect. The three principle haemoglobins of the carp are also **Root-effect haemoglobins**. During rapid swimming, the reduction in blood pH in the gill may be so great that Root-effect haemoglobins can no longer be saturated; in such cases, haemoglobins with and without the Root effect are found together. The best-known example of this is the rainbow trout *Salmo irideus (S. gairdneri)*: Hb IV (about 65 % of the total) is a typical Root-effect haemoglobin, but Hb I and II, in contrast, show practically no Bohr effect; the properties of Hb III (about 3 % of the total) are not known. In Hb I, all the polar amino acids involved in the Bohr effect and in binding organophosphates have been replaced by neutral residues: βEF6-Lys with leucine, βF9-Ser with alanine, βFG1-Glu with asparagine, βH21-Arg with serine, and βHC3-His with phenylalanine (Table 7.3). The sequence difference between Hb IV and Hb I of the trout is no less than 41 %, and between Hb IV and human Hb A it only 50 %. This clearly illustrates the extreme specialization of the Root-effect haemoglobins [21, 131, 133]. Haemoglobin systems with mixed properties, which provide adaptation to different living conditions, are also found in adult amphibians. Whilst in the aquatic Surinam toad *Pipa carvalhoi* both principal haemoglobins show minor Bohr effects ($\varphi = -0.05$) and strong interaction with organophosphates, the related semi-aquatic species *Leptodactylus labrynthicus* has the "aquatic" Hb II and a further haemoglobin, Hb III, which, like the haemoglobins of the terrestrial anurans, has a normal Bohr effect and a low ATP sensitivity [107].

Carbon dioxide reduces the oxygen affinity mainly by the protons it generates. A specific **CO_2 effect** involves the formation of carbamino groups with the terminal $-NH_2$ of the α- and β-chains and a consequent reduction in O_2 affinity. Quantitatively, this effect plays only a minor role compared with that of the H^+ ions. Organophosphates compete with CO_2 for the terminal-NH_2 of the β-chains and thereby further reduce the CO_2 effect. Blocking of the α-N-termini, as found in the haemoglobins of many fish and amphibians and of several mammals (Table 7.3), also prevents carbamino binding [183].

Certain amino acids of the β-chains form positively charged fields after protonation and these promote the T → R transition by mutual repulsion. In the mammals, this involves the N-terminus, βNA2-His, βEF6-Lys and βH21-His. Inorganic **anions** neutralize these charges and thereby reduce the O_2 affinity of most haemoglobins, although to a somewhat lesser degree than do organophosphates. Because of its relatively high concentration, the Cl^- ion is of particular physiological importance [24, 49]. The haemoglobin of the crocodile *Caimand crocodylus* is only slightly sensitive to H^+, CO_2, Cl^- and organophosphates, and this clearly increases its diving capacity; conversely, HCO_3^- has a reducing effect on O_2 affinity, as in other crocodile species, and this facilitates the efficient use of the oxygen bound to the haemoglobin. The HCO_3^- effect requires the three amino acid substitutions βNA1-Val to Ser, βNA2-His to Pro and βHC1-Lys to Glu; reduction of the other heterotropic interactions comes from the exchanges βH21-His to Ala and αH14-Ser to Ala [131, 183]. **Urea**, which in human Hb A increases the O_2 affinity by dissociating the tetramers, has no effect in several shark species, even at concentrations as high as 5.0 mol/l. The haemoglobins of *Squalus acanthias* and *Cephaloscyllium isabella* are, however, sensitive to urea: in contrast to various enzymes, the urea effect on haemoglobin is not compensated by trimethylaminoxide [4].

Organophosphates like 2,3-bisphosphoglycerate (DPG), inositol phosphates (IP_6, IP_5, etc.), ATP or GTP induce salt bridges between the β-chains and thereby stabilize the Z state. In mammals DPG is bound to βNA1-Val, βNA2-His, βEF6-Lys and βH21-His, and in birds IP_5 additionally binds to βH13-Arg and βH17-His. The teleosts have the amino acids βNA2-Glu/Asp and βH21-Arg instead of the histidine found in the equivalent positions in man; this represents a steric adaptation to the binding of nucleoside triphosphates. ATP binds to five amino acids (β^2NA1-Val, β^1NA2-Glu, β^1H21-Arg, β^{1+2}EF6-Lys); GTP additionally binds to β^1NA1-Val and thereby has a twofold higher modulating effect [183]. Curiously, the two haemoglobins of the rhinoceros *Rhinoceros unicornis* contain not only the DPG-binding sites typical for mammals, βNA1-Val, βEF6-Lys and βH21-His, but also in position βNA2 either glutamic acid (Hb A) or aspartic acid

(Hb B), as found in the NTP-specific haemoglobins of the teleosts. Thus, rhinoceros haemoglobins are modulated by both ATP and DPG [2]. The organophosphates reduce the intrinsically high O_2 affinities of most mammalian and avian haemoglobins to physiologically meaningful values, e.g. the p50 for human haemoglobin is shifted from 1500 to 3200 Pa. They also effect an increase of the Bohr effect. Some groups of mammals have haemoglobins with intrinsically low O_2 affinities and these are insensitive to organophosphates. This is true, above all, for the super-family of the cat species (Feloidea) with the families Felidae, Viverridae and Hyaenidae; the super family of the bovine species (Bovoidea) with the families Cervidae, Antelocapridae and Bovidae; and the lemurs as the only primates. The basic amino acid βNA2-His, important for DPG binding, is replaced here by a neutral residue such as leucine, methionine or phenylalanine, or, in the case of the ruminants, it has been completely lost [131, 132]. In some vertebrate haemoglobins, e.g. Hb B of the domestic cat or the haemoglobins of *Crocodilus niloticus* and *Alligator mississipiensis*, the sensitivity to organophosphates is reduced by blockage of the α-chain N-terminus by an acetyl group [1, 131, 132].

Changes in the effect of DPG in mammals often have to do with **adaption** to unfavourable conditions for O_2 transport. Llamas and guanacos are adapted to the lower pO_2 at high altitudes not only by having a lower concentration of DPG in their erythrocytes but also the DPG sensitivity of their haemoglobin is markedly reduced by the exchange of βNA2-His for asparagine. It is more difficult to understand why the same amino acid substitution is also found in the haemoglobin of elephants and the armadillo *Dasypus novemcinctus*; in the latter case, it might be an adaptation to a burrowing habit. A reduced organophosphate sensitivity of its haemoglobin also helps the goose *Anser indicus* cope with the low pO_2 of its habitat. This species breeds in Central Asian mountains at altitudes of about 4000 m. The intrinsic O_2 affinity of its haemoglobin is little different to that of the domestic goose *Anser anser* but it is reduced much less by IP_5 [19]. The effect of phosphate increases with the number of negative charges, e.g. in the haemoglobin of the ostrich in the order GTP, ATP < IP_4 < IP_5 < IP_6. GTP has a much stronger effect than ATP in many fish. This is related to the fact that ATP has the capacity to form only five H-bridges compared with six formed with GTP [131, 132]. Nucleoside triphosphates affect not only the T → R transition, e.g. in the haemoglobin of the tench *Tinca tinca*, but also the O_2 affinities in the T and R states. This does not fit the MWC model. Whereas heterotropic interactions affect and stabilize the T state of haemoglobins in higher vertebrates, nucleoside triphosphates also affect the R state, in particular the binding constant K_R [183].

The concentration and composition of the **organophosphates in the erythrocytes** vary greatly with the species and, in addition, can be regulated during adaptation to large fluctuations in the internal or external environment. The ATP content of enucleate mammalian erythrocytes is consistently lower (less than 1 μmol/ml) than that of the nucleated erythrocytes of other vertebrates (up to 20 μmol/ml). The erythrocytes of some teleosts and amphibian larvae contain about as much GTP as ATP, and therefore assume a rather special position amongst living cells. DPG is found in all cells as a cofactor of phosphoglycerate mutase, but occurs only at high concentrations in the erythrocytes of amphibians and mammals. Inositol tetracis-, pentacis- and hexacisphosphate (IP_4, IP_5 and IP_6) are found in high concentrations not only in avian erythrocytes but also in those of several fish and reptile species [8]. ATP normally predominates in **fish** erythrocytes; more GTP than ATP is found in *Mustelus*, *Cyprinus*, *Carassius*, *Anguilla*, *Lepomis* and *Esox*; DPG predominates in *Entosphenus*, IP_5 in *Arapaima* and *Squalus*, and IP_2 in *Lepidosiren* [8, 169]. In waters low in O_2, the concentration of nucleoside triphosphates is reduced in most fish and the O_2 affinity of the haemoglobin is therefore increased. In some species (e.g. *Pleuronectes*, *Fundulus*, *Salmo*) it is the ATP which is regulated and in others (e.g. *Cyprinus* and *Anquilla*) it is the GTP [183].

Amongst the **amphibians** and **reptiles**, the adults of *Ambystoma*, *Rana* and *Bufo* have both DPG and ATP in their erythrocytes. Embryonal DPG is found in turtles and crocodiles; adult crocodiles have only ATP and adult turtles also have IP_5. *Iguana iguana* and various snakes apparently have IP_6 [8]. In the viviparous species *Sphenomorphus quoyii* and *Agkistrodon piscivorus* (Reptilia), *Typhlonectes compressicauda* (Amphibia: Gymnophiona) and *Squalus suckleyi* (Elasmobranchii), O_2 transport from the mother to the embryo is facilitated by the presence of organophosphates at a lower concentration in the embryonal than in the maternal erythrocytes [183]. The predominant organophosphate of **avian erythrocytes** was previously thought to be IP_6, but in 1969 it was identified definitely as IP_5. Other inositol phosphates are present but play a subsidiary role; the ostrich *Struthio camelus* is

alone in having more IP_4 than IP_5 in its erythrocytes. Avian IP_5 is 1,3,4,5,6-IP_5 and the IP_4 is always 1,4,5,6-IP_4, in contrast to the mammalian isomer 1,3,4,5-IP_4 [119]. DPG is present in embryos of the duck, chicken, pigeon and various other birds, and is gradually replaced by IP_5. The reduction in DPG and ATP in the chicken is related to the maximum in blood O_2 affinity before hatching; this increases the efficiency of O_2 transport through the egg membrane [183].

In most **mammalian erythrocytes**, DPG is at such a high concentration (4–12 μmol/ml erythrocytes) that the haemoglobin is unsaturated and its O_2 affinity is reduced two- to threefold. In the Feloidea and Bovoidea, however, not only is the haemoglobin organophosphate insensitive but the DPG concentration in the erythrocytes is unusually low. During hibernation, e.g. of the hedgehog, the ground squirrel *Spermophilus mexicanus*, the marmot and the hamster, the O_2 affinity of the blood is increased by a reduction in erythrocyte DPG concentration [27]. The O_2 supply to mammalian foetuses is maintained in various ways. Primates and ruminants produce a special **foetal haemoglobin**. In the case of the ruminants, this has a high intrinsic O_2 affinity and a reduced DPG sensitivity; furthermore, the amino acid NA1 is missing in the non-α chains and βNA2-His is replaced by methionine. The foetal haemoglobins of man and other primates do not have a higher O_2 affinity but do have a lower sensitivity to DPG; in human haemoglobin F, this is achieved by the presence of serine instead of histidine in position γH21. However, in most mammals, e.g. the dog, horse, pig, guinea-pig, rabbit and rat there is no special foetal haemoglobin but simply a lower DPG concentration in foetal erythrocytes [27].

7.1.4 Temperature Effects and Adaptations

The oxygenation of haemoglobin is usually an exothermic process; higher temperatures therefore reduce O_2 affinity. The **enthalpy value** (ΔH) of oxygenation is extraordinarily variable in the vertebrates: in the fish it is 0 to –70 kJ/mol O_2, in the reptiles –30 to –50 kJ/mol O_2, and in mammals –40 to –60 kJ/mol O_2. This is related to the fact that O_2 binding produces heat but the simultaneous T → R transition absorbs heat in proportion to the number of broken H bonds. The difference between the two processes is measurable in terms of the ΔH value. Organophosphates reduce the numerical value of ΔH; because the binding of organophosphates is exothermic, their removal during oxygenation requires heat. Thus, for human haemoglobin ΔH is –50 kJ/mol O_2 in the absence of DPG and only –30 kJ/mol O_2 in its presence [75, 183].

Cold-blooded vertebrates show various ways of preventing a **reduction in O_2 affinity with increasing temperature**. The first possibility is for several or all haemoglobins to have ΔH values close to zero. This is true, for example, for the haemoglobin of *Iguana iguana* and Hb I of the rainbow trout, compared with ΔH = –45 kJ/mol O_2 for the Root-effect haemoglobin Hb IV [135]. A corresponding solution is found in fish with a greater than normal swimming capacity: the arterial blood, which is cooled in the gills, subsequently passes through a venous heat-exchanger, where its temperature is increased by 3–5 °C (in the porbeagle *Lamna ditrotus*), or even by 15 °C in the tuna *Thunnus thynnus*, before it reaches the underlying red musculature. The unavoidable O_2 loss by the venous blood is limited by the absence of a temperature effect on the haemoglobin *(Lamna)* or by an inverse temperature effect *(Thunnus)*, i.e. the O_2 affinity remains constant or increases with an increase in temperature [131, 132]. The variable temperature sensitivity of fish haemoglobin apparently does not correlate with the temperature variability of the environment: haemoglobins with ΔH values of –60 kJ/mol O_2 are found in antarctic species of the genus *Trematomus* and in species like *Fundulus heteroclitus* which inhabit regions with extremes of temperature [135]. The heterothermic vertebrates have a second possibility to reduce the influence of a temperature change on their haemoglobin: they can reduce the concentration of organophosphates in the erythrocytes at increasing temperatures. This mechanism is found in, for example, the turtle *Malacochersus tornieri*, the frog *Rana esculenta* and the fish *Ictalurus nebulosus*.

7.1.5 The Gene Families of Vertebrate Haemoglobins

The possession of **multiple haemoglobins** is the rule rather than the exception amongst the vertebrates. These are mostly the products of single genes which are expressed only at certain developmental stages or under particular physiological conditions. Allelic variants and post-translationally altered forms are, however also widely distributed amongst the haemoglobins. This variety can be particularly well illustrated in

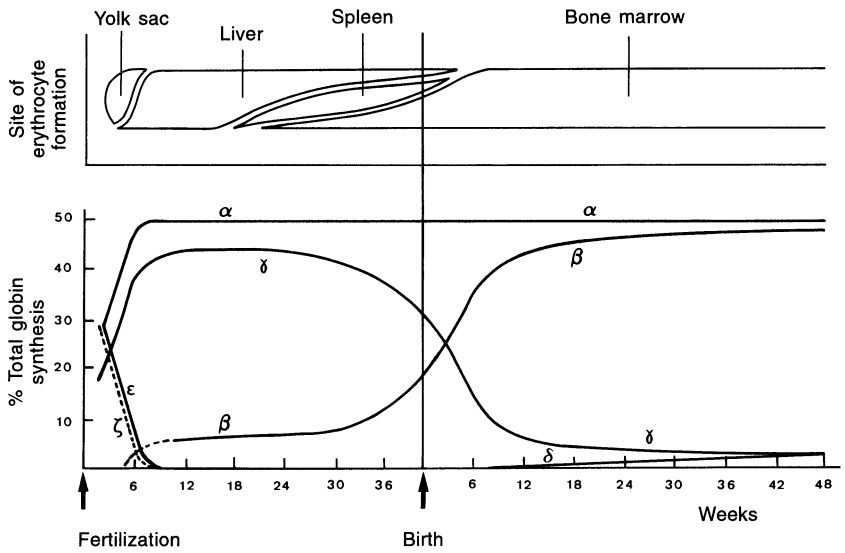


Fig. 7.8. Embryonal, foetal and adult haemoglobins are formed consecutively during human development [186]. Shown are the changes with time in the proportions of the different globin chains, and the organs in which the erythrocytes are formed and haemoglobins synthesized

man (Fig. 7.8 and Table 7.4). The erythrocytes of the adult contain about 94% Hb A ($\alpha_2\beta_2$), 2% Hb A_2 ($\alpha_2\delta_2$) and about 4–6% Hb A_1, which is formed post-translationally from Hb A by N-terminal glycosylation and contains mainly glucose. The glycosylation first occurs in the blood and depends upon the blood-sugar level; the proportion of Hb A_1 in diabetics can be as high as 2%. Human embryos contain Hb Gower 1 ($\zeta_2\varepsilon_2$), Gower 2 ($\alpha_2\varepsilon_2$) and Hb Portland ($\zeta_2\gamma_2$). Little is known about the functional properties of these embryonic haemoglobins; however, it is known that the ζ-chain is N-terminal acetylated. From about the 10th week, the embryonal haemoglobins are replaced by the foetal Hb F ($\alpha_2\gamma_2$), which then disappears at birth. There are two different γ-chains in Hb F, one of which carries glycine ($^{G}\gamma$) at position γH14 and the other alanine ($^{A}\gamma$). Hb F, like Hb A, can be glycosylated in the circulatory system; about 10–15% of the γ-chains are N-terminal acetylated. The amino acid sequences of all eight human globin chains (Mb, α, β, $^{G}\gamma$, $^{A}\gamma$, δ, ε and ζ) are known [66].

The **human globin genes** are distributed on three chromosomes: chromosome 11 bears the gene cluster for β-like chains, chromosome 16 bears the α-globin cluster, and chromosome 22 has the myoglobin genes [66]. The β-gene cluster includes the genes for the five non-α chains and a pseudogene (Table 7.4); the so-called ψβ2 lying in front of the ε gene shows no similarity to the globin genes. The number of genes in the α cluster is higher than was first thought. Just 65 bp behind the poly(A) signal of ψζ is the start of a newly discovered pseudogene, ψα2, which contains long inserts in exon 1 and intron 1 [59]. Recently, an apparently functional gene, Θ1, was discovered behind α1 in the orang-utan, and this gene belongs to an entirely new α-gene subfamily; the corresponding human gene has also been reported [52, 146]. Thus, the human α-gene cluster consists of three or four active genes and three pseudogenes (Table 7.4). Of the two ζ genes, only the 5′ gene (earlier known as ζ2) is active; the 3′ gene (earlier ζ1) appears to have

Table 7.4. Gene arrangement in the α- and β-globin clusters of various vertebrates

Human [59, 146]
α cluster: – ζ – ψζ – ψα2 – ψα1 – α2 – α1 – Θ1 –
β cluster: – ε – $^{G}\gamma$ – $^{A}\gamma$ – ψβ(ψη) – δ – β –

Mus domesticus[a] [129, 147]
β cluster: – y(ε) – bh0 – bh1 – bh2 – bh3 – b1 – b2 –

Peromyscus maculosus[a] [32, 114]
β cluster: – y(ε) – bh0 – b1 – b2 – b3 –

Rabbit [144]
α cluster: – ζ0 – ζ1 – α1 – Θ1 – ζ2 – ζ3 – Θ2 – ζ4
β cluster: – β4 – β3 – ψβ – β1 –

Domestic cow (R) and goat (z) β cluster [140]
R – ε^3 – ε^4 – ψ^3 – β – ε^1 – ε^2 – ψ^1 – ψ^2 – γ –
Z – ε^I – ε^{II} – $\psi\beta^X$ – β^C – ε^{III} – ε^{IV} – $\psi\beta^Z$ – β^A – ε^V – ε^{VI} – $\psi\beta^F$ – β^F –

Chicken [84]
α cluster: – π – α^D – α^A –
β cluster: – ϱ – β^H – β – ε –

Clawed frog Xenopus laevis [72]
αβ cluster: – $\alpha^L{}_a$ – $\alpha^L{}_b$ – α^A – β^a – $\beta^L{}_a$ – $\beta^L{}_b$ –

[a] bh denotes pseudogenes and b active β genes; bh1, however, codes for an embryonal globin.

been inactivated not so long ago by the nonsense mutation 6-GAG(Glu) → TAG (stop). Of the two α genes, α2 is expressed at levels two- to three-fold higher than α1 [106].

All the genes mentioned have the two introns typical of animal globin genes (see Fig. 2.9; p. 38); the plant haemoglobins contain a third, centrally located intron [95]. The introns of the various human globin genes have homologous positions but very different lengths. Thus, intron 1 of the α and β genes is only 117–130 bp long, whereas in the ζ gene it is 1265 bp long; intron 2 has a length of 140–149 bp in the α genes, 341 bp in the ζ gene, and 850–904 bp in the β-like genes [80]. The human globin genes have clearly attained the character of a model; much of our knowledge of gene structure, the control of transcription and expression, and mRNA processing and translation have been gained by their study. Regulation of the β-globin locus is achieved in the locus control region (LCR), which is localized in the region 6–20 kb in front of the ε-globin gene and contains four DNase I hypersensitive sites [66, 105, 124]. The two **changes in gene expression** (haemoglobin switches) during the development from embryonal to foetal haemoglobin and further to the adult haemoglobins are amongst the most investigated examples of gene regulation, and, because of the pathological persistence of Hb F in adults, they also have medical significance [80]. Gene duplications and deletions have often been observed in the globin gene clusters. For example, up to 10 % of the chromosomes in human populations have only one α gene or have three α genes; duplications have also been reported in the γ and ζ genes [66]. Comparison of neighbouring DNA sequences has shown that conversion between the two copies of the α and γ genes has occurred frequently in the evolution of the primates, but conversion between the β and δ genes has been less frequent [44, 89].

Globin gene mutations in man occur with a frequency of about 0.3 %; more than 400 anomalous haemoglobins are known so far, the study of which has produced valuable insight into the relationship between haemoglobin structure and function. Structural variation may involve the exchange of single amino acids; the most famous example is Hb S, where exchange of βA3-Glu for Val results in the aggregation of haemoglobin molecules upon deoxygenation and the formation of anomalous erythrocytes (sickle-cell anaemia). Other anomalous haemoglobins show shortened or lengthened chains, deletions or insertions of single amino acids, or frameshift mutations; the latter are tolerated only close to the C-terminus. Gene fusions result, for example, in δβ hybrids (Hb Lepore), βδ hybrids (Hb Antilepore) or $^{A}\gamma\beta$ hybrids (haemoglobin b Kenya). Mutations in transcription and splicing signals, nucleotide exchange in initiation codons, or the formation of new termination or splicing signals result in α- or β-thalassaemia, in which the corresponding chains are produced in insufficient amounts. The persistence of foetal haemoglobins in adults also involved changes in expression control signals [66, 70].

The location of α- and β-gene clusters on separate chromosomes is already found in the marsupials [180]. In the mouse, ψα-like pseudogenes are spread over several chromosomes. There is a large amount of comparative information available on the **β cluster of the mammals** (Table 7.4). In addition to man, the gorilla and chimpanzee also possess a ψβ gene between $^{A}\gamma$ and δ. The pseudogenes of all three species have the ATG initiator codon mutated to GTG, and 15-TGG to a stop codon GTA; a frameshift mutation in codon 20 has resulted in many stop codons in exons 2 and 3. In a genealogical tree constructed by the parsimony method, the primate ψβ and the goat ε^{II} lie on a line of orthologous genes which are separate from the ε, γ, δ and β genes and must, therefore, be considered as a different chain type; this is closer to ε and γ than to δ and β and is designated the η chain. Thus, the original mammalian β-gene cluster did not consist of four genes, as thought earlier, but rather of five (ε-γ-η-δ-β) [54]. The separation of the ε, γ, η and β, δ predecessors occurred 155–200 million years ago. Consequently, a primitive mammalian species, like the opossum *Didelphis virginiana*, possesses only two β-like globin loci, one of which is orthologically homologous to the ε, γ, and η genes of the higher mammals, and the other to the β and δ genes [88].

The functional predecessors of the η genes are encountered early in mammalian evolution; it may be, therefore, that there were originally three haemoglobin switches in mammalian development [54]. The original β-gene cluster has seen many changes during mammalian evolution. Duplication of the γ loci occurred only after the separation of the Old and New World apes (Catarrhini and Platyrrhini), in connection with the insertion of two type-Kpn transposable elements in the ε/γ region. An active δ gene is found in the New World apes as well as in man and the anthropoid apes; δ is not expressed in the Old

World apes or lemurs [173]. Inactivation of this gene in *Lemur fulvus* is the result of hybridization with the β gene [60]. Conversions between δ and β have occurred frequently in mammalian evolution and have led not uncommonly to the formation of hybrid genes, e.g. rabbit ψβ2, mouse bh2 and bh3, and goat $\psi\beta^x$ and $\psi\beta^y$ [54]. Thus, the rabbit β-gene cluster (Table 7.4) consists of two embryonal active genes, β4 and β3, which are comparable to ε and γ, and a foetal/adult gene β1, which corresponds to the β gene of other mammals [114].

Haemoglobin evolution in the rodents apparently led initially to three adult β genes. At any rate, three active β genes (b1, b2 and b3) are found in *Peromyscus maniculatus*, and several rat lines also express three adult β genes. The β-gene cluster of the mouse, on the other hand, has only two active adult genes (Table 7.4) [129, 147]. During the evolution of the ruminants, an ancestral gene cluster with the form ε, ε, ψ, β was duplicated (cattle) or even triplicated (goat, sheep) (Table 7.4). Sequence comparisons between the goat and sheep show that the rate of evolution of the haemoglobins after the two duplication steps was about threefold higher compared with that in man and the mouse. The goat produces three consecutive non-embryonal β types; the foetal chain β^F (earlier γ) is followed by the juvenile form β^C for up to 3 months after birth, when it is replaced by the adult type β^A. This third switch is reversible; β^C is again produced after extensive loss of blood [144]. Sheep of the haplotype A have the same 12 genes at the β locus as have the goat, and also produce β^C during anaemic stress. Haplotype B sheep lack four genes, including β^C [47]. The third gene block with the juvenile/stress haemoglobin chain β^C is also missing from the bovine genome (Table 7.4). Furthermore, one of the bovine pseudogenes is duplicated, and in relation to this remarkable phenomenon it should be recalled that pseudogenes may serve as gene reservoirs [144]. The haemoglobin system of the domestic pig also has some special features: there is only one adult haemoglobin, $\alpha_2\beta_2$, but four embryonal haemoglobins with the compositions $\zeta_2\varepsilon_2$, $\alpha_2\varepsilon_2$, $\zeta_2\Theta_2$ and $\alpha_2\Theta_2$. This Θ-chain, which should not be confused with the product of the Θ gene recently found in the α cluster, so far appears unique to the pig [11].

There are, as yet, no such detailed concepts of the **evolution of the mammalian α-gene cluster** (Table 7.4). It occurs at a significantly faster rate than that of the β cluster [58]. Two ζ genes are found in man and the chimpanzee. The ζ1 gene, which in man has become a pseudogene, appears to be active in the chimpanzee [185]. Duplicated α loci are found in many mammals, e.g. dogs, rodents, horses, cattle and primates, and appear to be subject to frequent gene conversion [34]. Genes in homologous positions in different species may be active or non-functional, depending upon whether or not they have been recently corrected through conversion with a functional α gene [58]. The sensational discovery of the Θ1 gene of the orang-utan as the first representative of a new α subfamily will certainly stimulate further research into the evolution of the α-gene cluster. Genes of the Θ family have been detected in various primates, the rabbit and probably also the horse. In contrast to the genes of the rabbit and the primitive primate *Galago*, the genes of the higher primates *Papio anubis* and *Pongo pygmaeus* (orang-utan) are intact and probably functional. Transcripts of the Θ1 globin gene have been found in embryonal erythroidal cells and in a human erythroleukaemia cell line [68, 102]. The Θ1 genes of *Pongo* and *Papio* have only 27 substitutions, of which 22 are synonymous. The amino acid sequences of Θ1 and α1 differ in only 55 positions, from which it can be concluded that a duplication α/Θ occurred about 260 million years ago. There is more than one Θ copy in *Papio, Pongo* and man [146]. The β-globin locus of the rabbit consists of three active genes and one pseudogene, whereas the α locus has five embryonal ζ genes in addition to one α and two Θ genes (Table 7.4); in this case, ζ1, ζ2 and ζ3 have suffered deletions and the substitution of critical amino acids, and they no longer produce functional globin chains [32, 114].

Post-translationally altered haemoglobins are widely found in the mammals. The proportion of glycosylated haemoglobins, however, appears to be dependent upon their rate of formation, which itself is determined by the glucose permeability of the erythrocyte membrane [137]. **Allelic variants** of the haemoglobins are actually no less common in other vertebrates than in man, but they have been less intensively investigated [123]. As well as in man, the sickle-cell phenomenon has been observed in numerous deer species, in sheep, goat, genets, the mongoose and even lizards; only in man and the deer has a mutation of the β chain been shown to be causal [149].

The only avian haemoglobin system to have been extensively investigated is that of the **chicken**. Some illuminating changes in the course of embryo development have received particular attention. There are four embryonal haemoglo-

bins in the chick, $\pi_2\varrho_2$, $\pi'_2\varrho_2$, $\alpha^A{}_2\varepsilon_2$ and $\alpha^D{}_2\varepsilon_2$, and two adult forms, $\alpha^A{}_2\beta_2$ and $\alpha^D{}_2\beta_2$, which altogether contain four α-like chains (π, π′, α^A and α^D) and three β-like chains (ϱ, ε and β). All these chains have been sequenced. A particular β type (β^H) is produced at the time of hatching. Mature embryos produce three so-called Koelliker haemoglobins with α^A chains lacking the terminal 141-Arg. Unfortunately, these are found in quantities sufficient for analysis only in the breeding strain "HYPECO" [33]. The chicken also contains two globin clusters. In the α cluster, as in the mammals, the genes are so arranged that the early expressed genes (π, α^D, α^A) are 5′; in contrast, the gene arrangement of the β cluster (probably ϱ, β^H, β, ε) does not appear to follow this rule [140]. The switch between embryonal and adult haemoglobins in the chicken is peculiar in that embryonal and adult genes are transcribed in haematopoietic cells towards the end of the switching period, but only adult globin is produced; this suggests the existence of post-transcriptional control.

Unlike in the mammals and birds, the α- and β-gene clusters of the clawed frog *Xenopus laevis* are close. The tetraploid species *X. laevis* has two αβ clusters and the diploid species *X. tropicalis* has only one. The two gene clusters of *X. laevis* each contain two larval and one adult gene (Table 7.4). The adult globin chains differ between the primitive *X. tropicalis* and the tetraploid species *X. laevis* and *X. borealis* in 12 % of positions; this is a larger difference than that found between mammalian species of different orders. The difference between homologous chains from the two clusters of the tetraploid species is 3–4 %. It can be estimated from these data that the genome doubling which separated *X. tropicalis* from the other species occurred about 110–120 million years ago [84, 152]. The larval globins differ from the adult forms in 40–50 % of their amino acids, and the different larval genes are expressed with different time schedules [6]. The larval α chains are more than 90 % similar; however, in the tadpole there is a further α chain (αT5) of unknown gene localization which has only 72–75 % similarity to the other larval α chains. Finally, in the earlier stages of embryo development, two other β chains are produced (βE1 and βE2) but these disappear in the feeding stage [7].

Thus, in the amphibians there is also at least one **switch in globin gene expression** which is related to metamorphosis, the timing of which, however, is very species specific. In many urodelans, the switch begins long before metamorphosis but extends over a long period. In the anurans, the switch usually occurs rapidly, mostly at the time of metamorphosis, although in *Hyla* it occurs 10 days later. Only in the case of *Bufo bufo* does the switchover begin months after metamorphosis and lasts for several months [30]. Experimental anaemic stress in *Rana* and *Xenopus* results in the renewed production of larval haemoglobins. Most of the investigated fish species showed alterations in haemoglobins during post-embryonal development. In many amphibians and fish, the availability of a whole spectrum of structurally and functionally different haemoglobins facilitates adaptation to changing environmental conditions (p. 257). The **heterogeneity of haemoglobins in fish** is very great; 71 out of 77 Amazonian fish genera examined possessed several Hb bands (2–12, with a mean of 3.9). As many as 18 different haemoglobins appear in the course of the life cycle of *Oncorhynchus* and other salmonids. The coexistence of numerous haemoglobins is made possible by the fact that the dissociation of fish haemoglobins to dimers is restrained compared with the situation in other vertebrates, and therefore heterotetramers, like $\alpha\alpha'\beta\beta'$, are stable.

7.1.6 Myoglobins

Myoglobins are present in the heart and skeletal muscle of most vertebrates and in the muscle of many invertebrates. They promote O_2 supply to the sarcosomes, their most important property in this connection being their high rate of O_2 dissociation. Even Antarctic fish which lack haemoglobin-containing erythrocytes have myoglobins in their heart muscle [41]. The concentration of myoglobin is particularly high in the muscles of aquatic mammals, where is also serves as an O_2 store. Vertebrate myoglobins are all monomers with 153–154 amino acids and a molecular mass of 17.8 kDa. The sequences of more than 60 myoglobins are now known, mainly from the mammals [110]. One myoglobin gene has been described for each of man, the seal and the mouse, species in which there are probably multiple Mb genes. Myoglobin genes, like those for all globins, have two introns, but they are unusually long: intron 1 in the species mentioned is about 4.4 kb; the human intron 2 is 3.6 kb, that of the seal is 3.2 kb, and in mouse it is only 1.5 kb [14]. Myoglobin exists in solution in two conformational isomers in which the haem is rotated 180° around the axis α–γ. One of the forms generally

predominates and constitutes the native form of the pigment. In the sperm whale, the deviant form makes up less than 10 % of the myoglobin at equilibrium; in contrast, in the tuna fish the two myoglobin forms are equally represented [10].

7.1.7 Methaemoglobin Formation and Reduction

Although the Fe(II) in the haem pocket is relatively well protected from oxidation, both erythrocyte haemoglobin and muscle myoglobin of the vertebrates is continuously converted to methaemoglobin (or -myoglobin) with Fe(III), which can bind no more O_2. The auto-oxidation rate is temperature dependent; the haemoglobins of the marine teleosts show some adaptation to the average temperatures of their habitats, such that at any given temperature cold-water fish have a tenfold higher auto-oxidation rate than warm-water fish. On the other hand, hydrostatic pressure has no influence upon auto-oxidation [187]. The reduction of methaemoglobin to functional haemoglobin in vertebrates of all classes is achieved mainly by NADH- and NADPH-specific MetHb reductases, all of which are partly membrane bound [71, 103, 145]. Vertebrate blood at equilibrium contains 1–3 % methaemoglobin. The much higher values given in the literature for the reptiles are probably artefacts.

7.2 Invertebrate Haemoglobins and Chlorocruorins

The haemoglobins of invertebrates are remarkably sporadic in their distribution. Whereas the vertebrates always have myoglobins in their heart and skeletal muscle and, with few exceptions, possess haemoglobin-containing blood cells, haemoglobins are found in many phyla of invertebrates (Table 7.2) but only in individual species or groups. Therefore, the question of the biological importance of the haemoglobins to the invertebrates is not so clearly answerable as it is for the vertebrates. Haemoglobins are found either freely dissolved or in haemoglobin-containing cells (erythrocytes) in the blood, haemolymph or body-cavity fluid (Table 7.2). Intracellular haemoglobins are also present, often simultaneously, in invertebrate muscle, nerve cells and other cell types. For example, in the echiurid *Thalassema* haemoglobins are present in cells of the coelom fluid, coelom epithelium, body wall muscles, gut wall, fat cells, nervous system, anal vesicles, and eggs. Haemoglobins are even found in several protozoans. Some animal species have several different types of respiratory pigment, e.g. haemoglobin in the muscle or body cells together with chlorocruorin dissolved in the body fluid (e.g. the polychaete *Potamilla*), or haemocyanin (e.g. Polyplacophora, Gastropoda *Busycon* and *Aplysia*). Other species possess both dissolved haemoglobin and an intracellular form in the cells of the coelom fluid (certain polychaetes), or they have both extracellular haemoglobin and chlorocruorin in the blood (the polychaete *Serpula*).

The universally used term "haemoglobin" emphasizes the presently accepted homology of all these proteins, regardless of whether they occur in plants or animals [55, 174]. Before this homology was defined, other terms were applied and these are still partly in use today. The name **"erythrocruorin"** has a complicated history. It was coined by Lankester in 1868 to distinguish the haemoglobins of certain invertebrates from green chlorocruorin; it was adopted in 1933 by Svedberg to differentiate between the intracellular, low molecular weight vertebrate haemoglobins and the then known extracellular, high molecular weight invertebrate haemoglobins. Today, when many low molecular weight invertebrate haemoglobins are known, the name erythrocruorin may no longer be used for all invertebrate haemoglobins, but should be reserved for the highly polymeric, extracellular haemoglobins of the molluscs and annelids in order to emphasize their unique character. Myoglobin refers only to the muscle haemoglobins and not to those of other body cells.

The **intracellular haemoglobins** of the invertebrates are, as a rule, mono-, di- or tetramers, and only seldom higher polymers, with subunits of 12–16 kDa and one haem group. There are, however, intracellular haemoglobins with subunits of twice this size, and these include two domains each with a haem group, e.g. in the larvae of the bot fly *Gasterophilus* and in the mussel *Barbatia reeveana*. The **extracellular haemoglobins** are divided into four types according to their structure [174]. The first type are mono-, di- or tetramers with subunits of 16 ± 2 kDa and with one haem group; this is the structure of the haemoglobins in the haemolymph of the larvae of non-biting midges (Chironomidae). Haemoglobins of the second type have subunits of 30–40 kDa with two haem-bearing domains; the haemoglobin molecule of several lower crusta-

ceans is composed of 10–24 such subunits, and that from the body cavity fluid of the roundworm *Ascaris* and its relatives has 8 subunits. The polypeptide chains of the third group each carry 8–20 haem groups; the haemoglobins of certain freshwater snails (Planorbidae), marine mussels (Astartidae and Carditidae) and the brine shrimp *(Artemia salina)* are made up of two or more such subunits. The fourth type includes the haemoglobins of the annelids and Pogonophora as well as the polychaete **chlorocruorins**, which contain a somewhat different prosthetic group but are otherwise extraordinarily similar to the haemoglobins. The pigments of this fourth group are all molecules of several thousand kilodaltons, and in the electron microscope have the characteristic appearance of two superimposed hexagons. The molecular architecture of these giant molecules has not yet been defined; they often consist of subunits of about 16 kDa with one haem group which are partly linked by disulphide bridges, but they may also have subunits of twice, or sometimes three or four times, the size and haem-free polypeptides.

The haem-bearing subunits or domains consistently have molecular masses of 16 kDa and lengths of between 132 and 157 amino acids [9]. The **amino acid sequences** of haemoglobins from very different invertebrate groups have recently been determined; these include the ciliates, annelids, pogonophorans, echiurids, crustaceans, insects, molluscs and holothurians. In contrast, the genes are known in only two cases: a globin gene from the earthworm *Lumbricus* has a structure similar to that of the vertebrates with two introns, whereas the globin genes of the non-biting midges (Chironomidae) have no introns. Pairwise comparison of the amino acid sequences of distantly related animal species usually shows only a few identical positions. The **phylogenetic history** of these haemoglobins apparently goes far back in time. Even different globin chains of the same species can vary greatly. For example, the chains I and II of the haemoglobin from the earthworm *Lumbricus terrestris* have only 59 (42 %) identical amino acids [148]; pairwise comparisons of the 12 haemoglobins in *Chironomus thummi thummi* consistently show only 50 % agreement [128]. Only when a large number of sequences are available will it be possible to derive a molecular genealogy which reflects the relationships between the various invertebrate haemoglobins. Despite the large sequence differences, there is still no doubt that all haemoglobins are homologous. For example, the sequence of the very unusual globin F–I of the echiurid *Urechis caupo* in fact agrees in only 11–20 % of positions with those of man, the agnathan *Petromyzon*, the holothurian *Molpadia*, the insect *Chironomus*, the annelids *Glycera* and *Lumbricus*, the molluscs *Anadara* and *Aplysia*, and the soybean, but 51 % of the amino acids of *Urechis* are also present in at least one of the other chains [46]. The **spatial structure** of invertebrate haemoglobin has, up to now, been examined only in the case of CTT-III from *Chironomus thummi thummi*, where there is extensive similarity to vertebrate haemoglobin (Fig. 7.4). Extrapolating from the amino acid sequences, it would appear that all known haemoglobin chains have similar secondary and tertiary structures, and this is not analogous to that of any other protein family [9]. Thus, despite fundamental changes in sequence in the course of evolution, structural parameters essential for the function have been retained [174].

Only two amino acids are invariant in all haemoglobins: CD1-Phe and the haem-binding "proximal" F8-His. The haem iron is coordinately bound, on the one hand, to the four pyrrole-N residues and, on the other hand, to the imidazole ring of this histidine; the haem itself is anchored in the haem pocket by non-polar interactions with about 15 other amino acids. The region of the **haem pocket** has been especially conserved during evolution. Position E7 is also occupied by histidine, the "distal" histidine, in almost all haemoglobin chains. On oxygenation, this amino acid swings away from the entrance to the haem pocket, thereby providing access for O_2. Apart from histidine, sterically the most suitable amino acid for this function is glutamine, which in fact replaces E7-His in some globins, e.g. the myoglobin of the African and Indian elephants, the myoglobin of the shark *Mustelus antarcticus*, the α-chain of the opossum, the β-chain of the snake *Liophis miliaris*, the globin of the agnathan *Myxine*, the larval globin CT-IIIA of the non-biting midge *Chironomus thummi thummi*, the globin of the echiurid *Urechis caupo*, and haemoglobin I of the mussel *Calyptogena soyoae* [46, 108, 118, 159, 174]. One finds E7-Leu in the monomeric globin of the polychaete *Glycera dibranchiata*, but in the monomeric myoglobins of the snails *Aplysia limacina* and *Dolabella auricularia* E7-Val occurs [156, 189]. The haemoglobin of the small liver-fluke *Dicrocoelium dendriticum* has tyrosine at position E7, and not glycine as was thought earlier [97].

Cooperativity and heterotropic modulations in the tetrameric vertebrate haemoglobins involve

interactions between heterodimers. However, these concepts cannot, as a rule, be applied to invertebrate haemoglobins. Cooperative homodimers and the resulting tetramers are known from various mussel species of the family Arcidae. In these cases, the contacts between the subunits are formed via the helices E and F; consequently, the subunits have exactly the opposite orientation to those in vertebrates where E and F are external [17, 108]. Compared with vertebrate tetramers, the homotetramers of haemoglobin F–I from *Urechis caupo* are also "inside out"; the G/H helices lie on the tetramer surface. The contacts between the subunits, however, are different to those in the Arcidae; the haemoglobin shows neither cooperativity nor a Bohr effect nor any other heterotropic interaction [46, 86]. In the giant haemoglobins of the annelids, the **Hill constants, n**, have maximal values of about 7–8, although the whole molecule contains at least 144 haem groups. In this case, the cooperative subunit apparently corresponds to one of the 12 structural units visible in electron micrographs of this molecule, which has the form of two hexagons [108]. The even larger extracellular haemoglobins of the marine mussels from the families Astartidae and Carditidae, which contain hundreds of haem groups, show no cooperativity whatsoever [163]. Many, but by no means all, invertebrate haemoglobins show a **Bohr effect**. Many of the haemoglobins show both the normal (alkaline) and the inverse (acidic) Bohr effect; the haemoglobins CTT-III and CTT-IV of *Chironomus* have only the alkaline effect, and *Dicrocoelium* shows only the acidic effect [151]. Of all the invertebrate haemoglobins, the molecular mechanism of the alkaline Bohr effect has been clarified to some extent only in *Chironomus* [151]. Heterotropic modulation of haemoglobins by organophosphates, which is of great importance in the vertebrates, is apparently not found in the invertebrates [174].

The **enthalpy values** (ΔH) of oxygenation are as variable in the invertebrates as they are in the poikilothermal vertebrates. The different haemoglobins of *Chironomus* have values between –40 and –81 kJ/mol O_2 [182]; the haemoglobin of the earthworm *Maoridrilus montanusi* has a value of –27 kJ/mol O_2, like other annelids [43]; and the tetrameric and monomeric haemoglobins of the mussel *Barbatia reeveana* have values of –22.6 and –12.1 kJ/mol O_2, respectively [57]. In spite of the basic similarity of haem pocket structures, marked deviations in the amino acid sequence result in large differences in O_2-binding kinetics between invertebrate haemoglobins, especially in terms of the rate constant for dissociation (k_{off}) rather than that for binding (k_{on}) (Table 7.5). The kinetic constants for CO or ethylisocyanide, on the other hand, are much less variable [36, 72].

Table 7.5. The rate constants according to Eq. (7.2) (p. 254) for O_2 binding (k_{on}) and release (k_{off}) of various haemoglobins [72]

	k_{on} $\mu mol^{-1}s^{-1}$	k_{off} s^{-1}
Human α chain	50	28
Human β chain	60	16
Sperm whale myoglobin	19	10
Glycera dibranchiata	190	2800
Chironomus thummi thummi	300	218
Dicrocoelium dendriticum	300	30
Ascaris lumbricoides	1.5	0.004
Aplysia limacina myoglobin	15	70
Leghaemoglobin	150	11

Little is known about the **biosynthesis and degradation** of invertebrate haemoglobins. Just as in the vertebrates, they are undoubtedly subject to inactivation by methaemoglobin formation. Superoxide dismutase, which can reduce this effect, is widely found in the invertebrates (p. 706), but methaemoglobin reduction has so far been detected only in the sipunculid *Themiste* sp.

7.2.1 Annelids, Pogonophora and Echiurids

A particularly large variety of respiratory pigments is found in the annelids. Many polychaetes, oligochaetes and leeches possess soluble haemoglobins in the blood, and sometimes also in the coelom fluid (e.g. *Nephthys hombergi*). In several polychaete families, the blood contains green chlorocruorin; this has the slightly different spirographis haem as the prosthetic group, instead of protohaem (Fig. 7.2), but is otherwise very similar to the annelid haemoglobins. Intracellular haemoglobins are also found, for example, in cells of the coelom fluid of some polychaete families, and in the muscles and nerve cells of several species, either as the sole respiratory pigment *(Aphrodite aculeata)* or in addition to soluble haemoglobin *(Abarenicola pacifica)*, soluble chlorocruorin *(Potamilla leptochaeta)* or coelom cell haemoglobin *(Glycera robusta)*. *Magelona* is probably the only polychaete to possess haemerythrin. The haemoglobins found in some pogonophoran species are very similar to annelid extracellular haemoglobins.

The **extracellular haemoglobins** of the annelids are giant molecules whose structure can be analysed with the electron microscope. They have the form of two superimposed hexagons with more-or-less spherical subunits. For technical reasons, previous studies reported widely varying values for the dimensions of these molecules. More recent investigations consistently give a diameter of 30 nm and a height of 20 nm. The molecular mass is 3.6–3.8 MDa [56, 79]. In addition to molecules with the described dimensions, the polychaetes *Ophelia bicornis* and *Euzonus mucronata* of the family Ophelidae contain aggregates with a two- to threefold larger mass [108].

In mildly alkaline conditions, the molecules dissociate into fragments of varying size but always include a fragment that is 1/12th of the original mass. Whereas Ca^{2+} stabilizes the molecular structure, ethylene diaminetetraacetic acid (EDTA) promotes **dissociation**. A few species show reversible dissociation *(Dina dubia, Haemopis sanguisuga)*, and in *Eunice aphroditis* an association-dissociation equilibrium exists over a wide pH range around the neutral point. Dissociation of the native haemoglobin molecule can also occur at acidic or neutral pH values upon addition of urea, perchlorate or certain other anions [108]. For a long time there was controversy over just how many different polypeptides exist in extracellular haemoglobins and chlorocruorins of the annelids, and which are the stoichiometric relationships between these types of polypeptides. It is now agreed that the haemoglobins of the polychaetes and oligochaetes consist of four types of haem-bearing 16-kDa chains, three of which are linked by disulphide bridges. There are, in addition, at least two types of non-haem-bearing 30-kDa chains known as "linkers". The situation in the leech *Macrobdella decora* is somewhat different. SDS-gel chromatography under non-reducing conditions in this case gives three types of subunit: monomers of about 17 kDa, non-reducible dimers of 30 kDa, and reducible haem-bearing dimers of 34 kDa. According to the N-terminal sequences, at least five different polypeptides are present [79]. Amino acid sequences have been obtained for all four haem-bearing chains of the polychaete *Tylorrhynchus heterochaetus* and of the common earthworm *Lumbricus terrestris*, and for the monomeric subunits of the freshwater oligochaete *Tubifex tubifex* and of the earthworm *Pheretima sieboldi*. The ten haem-bearing chains agree in 30–50 % of positions; of the 25 invariant amino acids, 9 are conserved in the human β-chain [56, 154]. Linkers 1 and 2 of *Tylorrhyncus* and one linker of the deep-sea pogonophoran *Lamellibrachia* have also been sequenced; they show agreement in 23–27 % of their 224–253 positions but show no relationship to the haemoglobin sequences [160]. The only annelid haemoglobin gene so far to have been sequenced, the gene for chain III (or c) of *Lumbricus terrestris*, contains two introns of 1344 and 1169 bp in positions homologous to those in vertebrate haemoglobin genes [76].

Because of the existence of polypeptides without haem, the haem (2.6 ± 0.4 %) and iron (0.23 ± 0.01 %) contents of the annelid haemoglobins are both very low compared with those of other haemoglobins; this results in the unusually high mean molecular mass of 23–26 kDa per haem [175]. The nomenclature of the haem-bearing chains is inconsistent: in *Tylorrhynchus*, the monomeric chains are labelled as I and the subunits of the trimers as IIA–IIC; the corresponding designations in *Lumbricus* are I or d and II–IV or a–c. Because of the greater sequence similarity pairwise, the designations a, A, b, B were suggested more recently for the four chains [56]. Accordingly, the monomeric subunit is termed a (*Tylorrhynchus* I, *Lumbricus* I or d), and the subunits of the trimers are called A (*Tylorrhynchus* IIA, *Lumbricus* II or b), b (*Tylorrhynchus* IIC, *Lumbricus* III or c) and B (*Tylorrhynchus* IIB, *Lumbricus* IV or a). Despite many investigations with the most advanced electron microscopic techniques, the quarternary structure of the annelid haemoglobins has not yet been unambiguously defined. One model that is acceptable to many workers in the field is similar to the "bracelet" model proposed by Vinogradov in 1986; in this, 12 complexes with the composition $(aAbB)_4$ are bound together via 24 haem-free subunits to form two sets of six linked rings. The value of 192 haem groups per molecule which emerges from this model is somewhat larger than that calculated from the haem content and molecular mass [56, 174].

The extracellular annelid haemoglobins show **cooperativity**. For earthworm haemoglobin, Hill constants, n, have been recorded as between 2.5 and 7.9, depending upon the pH; Ca^{2+} shifts the pH for n_{max} from 8.1 to 7.6, and at the same time increases both the O_2 affinity and the Bohr effect. The 1/12th fragments also show cooperativity; although the n values for these fragments are always lower than for native haemoglobin, the homotropic interaction nevertheless appears to lie mainly, or entirely, within this 1/12th structure [50, 108, 176]. The O_2 affinities of annelid

haemoglobins are in general higher than those of the vertebrate haemoglobins. An (alkaline) **Bohr effect**, where it exists, is usually only of weak intensity. Annelid haemoglobins are able to bind anions and **cations** but only the latter influence structural and functional properties. The affinity is 1000-fold higher for divalent ions, like Ca^{2+} and Mg^{2+}, than for monovalent ions. There are high-affinity binding sites which are always saturated under physiological conditions and are responsible, above all, for stabilizing contacts between the 1/12th fragments. There are also Ca^{2+}-binding sites of lower affinity which modulate functional properties such as O_2 affinity and cooperativity [108]. The haemoglobins of *Lumbricus* and *Tylorrhynchus* were found to contain 1–4 tightly bound copper and zinc atoms in addition to 50–61 calcium atoms [153]. Organophosphates have no effect on annelid haemoglobins. The **temperature sensitivity** appears to be generally low; the ΔH of about –25 kJ/mol O_2 is lower than for many vertebrate haemoglobins [43].

The structure of **pogonophoran haemoglobins** allows conclusions to be drawn about the systematic position of this unique group of animals. These sedentary worms of the seabed were initially considered to be a unique phylum close to the Hemichordata on the basis of their three body subdivisions (archimerie); more recent studies of intact Pogonophora, however, show the existence of a fourth, bristle-bearing, segmented region which was missing in the original specimens. This morphological similarity to the annelids is paralleled by marked similarities between the haemoglobins.

The giant, tube-inhabiting worms of the genera *Riftia* and *Lamellibrachia*, which are amongst the characteristic inhabitants of hydrothermal vents, have recently been divorced from the Pognophora and placed in a new phylum, the Vestimentifera. Hydrothermal vents are openings in the seabed from which hot water with a high hydrogen sulphide content emerges; they were discovered in 1977 at a depth of 2000–3000 m off the Galapagos Islands. This extreme habitat with temperatures of 6–25 °C, a pH value of about 3.5, and an H_2S concentration of over 10 mmol/l is occupied by giant, tube-dwelling worms *(Riftia pachyptila)*, which are 1–3 m long and have a diameter of 4–5 cm, and by large mussels of the species *Calyptogena magnifica*, with a diameter of 25–30 cm. *Riftia* has no gut but instead has a special organ (trophosome) that hosts chemosynthetic, symbiotic bacteria which oxidize sulphide to sulphate; the energy extracted in this way is used to convert carbon dioxide to various organic substances. Similar bacteria are to be found in the gills of the mussel and also in various deep-sea annelids [74].

Riftia has two haemoglobins in the blood and coelom fluid; the predominant Hb A form in the blood has a hexameric structure and a molecular mass of 1.5 MDa, and the Hb B of the coelom fluid has a mass of 400 kDa. Both forms have the capacity to bind SH^- ions, whereby the binding capacity of Hb A (1–2 mol sulphide/mol haem) is greater than that of Hb B (0.4–1). The haemoglobin here apparently serves to supply the symbiotic bacteria with SH^- ions and to protect the animal tissues from the highly toxic effect of a high concentration of free sulphide [108]. The second pogonophoran species with a well-investigated haemoglobin is *Lamellibrachia* sp., which lives in less extreme hydrothermal regions. These species also possess two haemoglobins with different molecular masses. The 3-MDa haemoglobin contains two haem-free linker chains, AV and AVI, of about 30 kDa, and four haem-bearing chains of about 16 kDa, of which AIII and AIV are present as monomers and AI and AII are associated as a dimer. The 440-kDa haemoglobin is made up only of the haem-bearing chains BI–BIV. The N-terminal sequences have been determined for all eight haem-bearing chains; AII agrees with BII, and AIII with BIII. The complete amino acid sequences are known for AIII and the linker chain AV. The chains AI, AII, AIII, BII and BIII all have a free cysteine residue which probably represents the sulphide-binding site. The haem-bearing chains of *Lamellibrachia* are all significantly homologous to those of the annelids and may be incorporated into the annelid dichotomous phylogenetic tree; AI, BI and AIII/BIII belong to the aA group of annelid subunits, and AIII/BII, AIV and BIV belong to the bB group. The AV linker agrees at 27 % of positions with the homologous subunits of the polychaete *Tylorrhynchus heterochaetus* [161].

The green **chlorocruorins** are present in solution in the blood of the sedentary polychaetes of the families Sabellidae, Serpulidae, Chlorhaemidae and Ampharetidae. The prosthetic group is spirographis haem, which differs from protohaem only by having a 2-formyl instead of a 2-vinyl residue (Fig. 7.2). Apart from this, the chlorocruorins are identical to the annelid extracellular haemoglobins in molecular size, electron microscopic appearance, haem content, and dissociation kinetics. Even very closely related species differ simply with respect to these two respiratory pigments, e.g. *Spirorbis borealis* has chlorocruorin,

whereas *S. corrugatus* has haemoglobin. In *Serpula vermicularis*, both pigments are present in the same animal. In this case the blood is greenish-brown with absorption maxima of 605 and 577 nm, corresponding to the α-bands for chlorocruorin and haemoglobin (Table 7.1). There is about 60 % protein-bound spirographis haem and 40 % protohaem, and as these have as yet proved impossible to separate, it may be that there is only one protein present with all the properties of a typical extracellular annelid pigment and bearing both prosthetic groups on the same molecule [174, 176].

Coelom cell haemoglobins are found in polychaetes of the families Glyceridae, Terebellidae, Opheliidae, Capitellidae and Ceratulidae, occurring together with extracellular blood haemoglobin in the terebellids *Amphitrite ornata* and *Thelepus crispus*. The coelom cell haemoglobins are all monomers of about 14 kDa in the terebellids, and 32-kDa dimers in the opheliid *Travisia*. The erythrocytes of the "blood worm" *Glycera dibranchiata* (Glyceridae) contain five to six monomeric haemoglobins of about 17 kDa and at least six globin chains of the same size; on oxygenation these form octamers and higher polymers by self-association. The monomers M-II and M-IV and the components P1, P2 and P3 of the polymer fraction have been completely sequenced and P4, P5 and P6 have been partially sequenced. The five complete sequences include 39 invariant positions. A further 44 positions are found in all three polymeric globins but not in the two monomers. In the two monomeric globins, the distal E7-His is replaced by leucine, which is no longer able to stabilize the HbO_2 by an H-bridge. Some of the haemoglobins of *Glycera* and of the larvae of the non-biting midge genus *Chironomus* include conformational isomers with the haem rotated 180°, as has already been described for vertebrate myoglobins. This isomerism has, of course, functional consequences [3, 122, 189]. The *Glycera* haemoglobin has a unique, extremely rapid O_2 dissociation (Table 7.5). In addition to blood and coelom cell haemoglobins, the annelids also have **myoglobins**. These are generally monomers with a molecular size of about 16 kDa, although those of *Travisia foetida* are dimeric. Dimeric myoglobins are otherwise only encountered in the molluscs [163].

The coelom cells of the **echiurid** *Urechis caupo* contain several tetrameric haemoglobins and the muscles possess a monomeric myoglobin. The haemoglobins show no cooperativity and no Bohr effect. At least five polypeptides, which differ only slightly from each other, are detectable in the haemoglobins, suggesting that *U. caupo* possesses several very similar globin genes. The amino acid sequence of globin F–I has been determined directly and via the cDNA; the chain of 141 amino acids is β-like with well-formed D helices, but shows no more than 20 % sequence similarity to other globins. The CD1-Phe and the proximal F8-His are conserved, whereas the distal E7-His is replaced by glutamine. Compared with human Hb A, only 2 each of the 17 $\alpha^1\beta^1$ and the 13 $\alpha^1\beta^2$ contacts are present. The subunits are so arranged in the homotetrameric F–I molecules that the G/H helices lie at the outer surface of the molecule and not at the inner surface, as in the vertebrate haemoglobins. In this respect, F–I resembles the heterotetrameric haemoglobins of the mussel family Arcidae (p. 271), although the contacts between the subunits in *Urechis* and the Arcidae are very different [117].

The haemoglobins of the echiurid *Thalassema mellita* have rather deviant properties. Here, there are two disulphide-linked dimers and smaller amounts of a monomer, which in total are made up from three different polypeptide chains of 13–16 kDa. One of the dimeric haemoglobins is a heterodimer having one chain in common with the monomeric haemoglobin; the other is a homodimer composed of chains of the remaining type. The coelom cells and their lysate show cooperative O_2 binding with $n = 1.5–1.9$, but the isolated haemoglobin has little cooperativity ($n = 1.0–1.3$) [172].

7.2.2 Haemoglobins of the Molluscs

Amongst the molluscs there are various **haemoglobin types** with characteristic distributions. High molecular weight extracellular haemoglobins are found only in the aquatic pulmonate snails of the family Planorbidae; otherwise, the typical respiratory pigment of the gastropods is haemocyanin. Mussels of the families Carditidae and Astartidae have a further type of extracellular haemoglobin. Haemoglobin-containing erythrocytes are found in mussels from at least seven families, but only a few species of the Arcidae have been examined in any detail. Myoglobins are present in the radula and buccal muscles of the polyplacophores and the gastropods, and in the shell adductors and foot muscle of the mussels; in fact, intracellular haemoglobins are also found in the heart and in gill and nerve cells of some molluscs [163].

The **extracellular haemoglobins** of the planorbids *Helisoma, Planorbis* and *Biomphalaria* have a completely unique structure. They are large molecules of 1.75–2.25 MDa which appear in electron micrographs as decagons with a diameter of 20 nm. SDS or guanidine breaks up the structure into fragments of 350–380 kDa which are composed of two disulphide-linked subunits of 175–190 kDa per haem, and thus each subunit must consist of 10–12 haem-bearing domains. On the basis of electron microscope studies, it was previously always assumed that there were up to ten subunits in the molecule [61, 163]. However, very accurate measurements of the molecular mass in *Helisoma trivolvis* give values of 2.25 MDa for the native haemoglobin and 190 kDa for its subunits; the derived model of 12 subunits arranged on a spherical surface would appear as a decagon in the electron microscope [108]. The functional interactions between the subunits and internally between the different domains have by no means been fully clarified. The native haemoglobin of *Planorbis corneus* has a p50 of 650 Pa, cooperativity ($n \leq 3.4$), and an alkaline Bohr effect, but no other heterotropic interactions. Partial proteolysis with subtilisin yields fragments with one haem; these can still bind O_2 but are no longer cooperative. Similar properties were reported for a haemoglobin present together with haemocyanin in the blood of the marine snail *Aplysia californa* [163]. The *Biomphalaria* haemoglobin is unusual in containing 3% carbohydrate; one of the two oligosaccharides is bound to an asparagine and the other to threonine or serine [171].

The blood of the marine mussels *Cardita borealis, C. affinis* (Carditidae) and *Astarte castanea* (Astartidae) contains extracellular haemoglobin with a molecular mass as large as 12 MDa. This is the largest known respiratory protein, larger even than gastropod haemocyanin, which has a mass of up to 10 MDa. These haemoglobins appear in electron micrographs as cylinders that are 20–34 nm in diameter and 36–120 nm long. Above pH 9 they break up into fragments of 1.4 MDa; subunits obtained by the usual methods are about 300 kDA in size. From the haem content it may be calculated that each subunit carries 18–20 haem groups. As in the previously mentioned case, fragments formed by partial proteolysis contain one haem and bind O_2 but show as little cooperativity as the native haemoglobin [174].

Mussels of the genera *Scapharca* and *Anadara* from the family Arcidae contain two types of haemoglobin in nucleated **erythrocytes**, homodimeric Hb I and tetrameric Hb II, the latter being made up from two heterodimers. On deoxygenation, Hb II tends to form aggregates of up to four tetramers. The amino acid sequences of various Hb I and Hb II chains are now known. All these chains carry a further helix in front of helix A; this is known as "N" or "pre-A" and lies parallel to G and H between H and the EF edge. The sequences of the haem-binding helices E and F deviate markedly from those of other haemoglobins; however, the non-polar nature of the haem pocket is preserved [108, 134, 163]. The sequence difference between the two haemoglobin types is greater than the species-specific differences. For example, the Hb I chains of *Anadara broughtoni* and *A. trapezia* agree at 82% of positions, but Hb I and Hb IIα of *A. broughtoni* are only 45% similar [108]. All these mussel haemoglobins are rather unusually dissociation resistant: 1 mol/l solutions of sodium iodide, propylurea or guanidine hydrochloride, which dissociate not only human Hb A but also the more resistant fish haemoglobin, have no effect [17].

The subunits of the homodimer Hb I are bound via contacts between the E and F helices; the tetrameric Hb II is composed of two similarly linked dimers. Whereas helices E and F point outwards in vertebrate haemoglobin dimers, they lie internally in the haemoglobins of the Arcidae and are responsible for the stability of the molecule and for interaction between the subunits [108]. All these haemoglobins have relatively low O_2 affinities, equally low Bohr effects and show no further heterotropic interactions. They are cooperative, with n values of 1.5 for the dimers and 2.1 for the tetrameric forms. The responsible homotropic interactions, however, are fundamentally different from those of vertebrate haemoglobins. In the latter, cooperativity and heterotropic modulation involve interaction between heterodimers; in the Arcidae, the dimers themselves are already cooperative. Furthermore, vertebrate haemoglobins, with values of $n \leq 3.0$, have more effective interaction than mussel haemoglobins [17, 37, 141].

The erythrocytes of other Arcidae, e.g. *Noetia ponderosa* or *Barbatia virescens*, possess only heterodimeric haemoglobins with subunits of 16 kDa. In addition to such heterodimers, the related species *Barbatia reeveana* and *B. lima* also contain polymeric haemoglobins of 430 kDa, consisting of 12–14 subunits of 32 kDa with two haem-bearing domains. Both domains apparently arose by gene duplication and consequently show high sequence similarity with each other and also

with subunits of other Arcidae haemoglobins. Although the polymeric *Barbatia* haemoglobins are the largest known intracellular haemoglobins, they have a relatively low Hill coefficient of n = 1.8 [108, 158]. Of the erythrocyte haemoglobins from the other mussel families, only those of the genus *Calyptogena* (Vesicomyidae) have been investigated. The famous *C. magnifica* of the hot hydrothermal vents possesses one tetrameric haemoglobin, whereas *C. soyoae* from cooler deep-sea regions has two homodimeric haemoglobins. The subunits of *C. soyoae* show 42 % sequence agreement with each other but only 12–20 % with other mollusc haemoglobins; they therefore represent a very primitive type of globin. The distal E7-His of Hb I has been replaced by glutamine [159].

The molluscs are the only animal phylum in which are present not only monomeric **myoglobins** (of about 16 kDa) but also dimers of twice the size. The myoglobins of the shell-closing muscle of the mussels *Mercenaria mercenaria* and *Saxidomus nuttali* are monomeric. Monomeric myoglobins are also found in the radula muscles of opisthobranchs (e.g. *Aplysia*), pulmonates (e.g. *Siphonaria, Helisoma*) and certain prosobranchs *(Patella, Nerita, Lunatia, Polinices);* other prosobranchs have dimeric myoglobins *(Buccinum, Busycon, Littorina, Nassa)*. Dimeric and monomeric myoglobins exist side by side in all examined Polyplacophora, and the amino acid composition is much more similar within each of the two classes than between the monomers and dimers of the same species. The amino acid sequences are known for the monomeric myoglobins of *Aplysia kurodai, A. limacina* and *Dolabella auricularia* and for the dimers of *Busycon canaliculatum* and *Cerithidea rhizophorarum*. All monomeric myoglobins lack the distal E7-His but it is present in the dimeric forms [156, 163]. Several of the dimeric myoglobins (in *Amaurochiton, Nassa, Buccinum*) show weak cooperativity (n = 1.2–1.5); however, this is more the result of a dissociation equilibrium between monomers and dimers than of real homotropic interactions; the dimeric myoglobin of *Busycon* is not cooperative [22, 163].

7.2.3 Haemoglobins of the Crustaceans

The typical respiratory pigment of the higher crabs (Malacostraca) is haemocyanin; haemoglobins in the crustaceans are restricted to the Copepoda, Cirripedia, Ostracoda, Phyllopoda and Anostraca. Detailed investigations have been carried out on just a few species of the latter two groups; here, haemoglobins are always found dissolved in the haemolymph and their structure in the Anostraca is completely different to those of the three subgroups of the Phyllopoda, viz. Notostraca, Conchostraca and Cladocera. In the cirripede *Briarosaccus callosus*, a parasite on brachyurans, there is a heterogenous mixture of haemoglobins with molecular masses of 250–4000 kDa, which appear on SDS gels under reducing conditions as subunits of 17–19 kDa; they have no similarity to any other arthropod haemoglobin [108]. Intracellular haemoglobins are also found in the crustaceans, e.g. in muscle, nerve and fat cells and in the eggs of the water flea *Daphnia* [174]. Three haemoglobins of about 260 kDa are present in the brine shrimp *Artemia salina* **(Anostraca)** and these differ in some functional characteristics; two are homodimers and the third is heterodimeric. The 130-kDa subunits consist of eight domains (E1 to E8) that are all different (the sequence agreement between pairs is 17–38 %) but display about 20 % similarity with vertebrate haemoglobins. The residues CD1-Phe and F8-His are present in all nine domains that have been sequenced so far [167]. Their secondary and tertiary structures are clearly similar to those of vertebrate globins. The molecules appear in electron micrographs as two superimposed disks, and in bisexual populations these haemoglobins are highly polymorphic. The related species *Parartemia zietziana* and *Streptocephalus torvicornis* also have haemoglobins of this type.

The haemoglobins of the **Phyllopoda** have very variable molecular masses. The largest are found in the Notostraca: for example, about 800 kDa in *Lepidurus apus* and about 600 kDa in *Triops longicaudatus*. Values between 420 and 700 kDa are reported for the haemoglobins of various *Daphnia* species (Cladocera); that of *D. magna* is 490 kDa and appears in electron micrographs as two superimposed octagons with a diameter of 14 nm. The haemoglobin of *Cyzicus hierosolymitanus* (Conchostraca) is a double pentagon with a diameter of 13 nm and a mass of 280 kDa. Denaturation of all these haemoglobins produces subunits of 30–34 kDa. The iron and haem contents correspond to the usual values of 15–17 kDa per haem, i.e. the subunits apparently have two haem-bearing domains. The native haemoglobin molecules are composed of 10–24 such subunits which, according to electron microscope studies, are organized in two superimposed rings. SDS electrophoresis of the haemoglobins of *Daphnia*

magna produces a complex spectrum of subunits under both reducing and non-reducing conditions; these subunits apparently arise partially by reciprocal conversion. The structure of these haemoglobins is thus more complicated than was originally assumed [130]. Crustacean haemoglobins show only weak cooperativity (e.g. for *Artemia* n= 1.6–1.9) and never a strong Bohr effect. Haemoglobin synthesis in *Daphnia* is stimulated by reductions on pO_2, and in *Artemia salina* by an increase in the salt concentration of the environment. The haemoglobin concentration in the haemolymph of both species shows corresponding wide fluctuation, reaching 15 g/l in *Daphnia* and up to 20 g/l in *Artemia salina*. Curiously, the concentration in the Australian species *Parartemia zietziana* is always very low (0.1 g/l) and is not influenced by the pO_2 [85].

7.2.4 Haemoglobins of the Insects

Intracellular haemoglobins are found in only a few insect species. Larvae of the fly *Gasterophilus intestinalis*, which live in the horse stomach, initially have haemoglobin in various body cells, but in later developmental stages haemoglobin occurs only in tracheal cells. The *Gasterophilus* haemoglobin is a non-cooperative homodimer of 34 kDa, and belongs to the few haemoglobins which have a lower affinity for CO than for O_2. Tracheal cell haemoglobin is also found in the back-swimmers (Notonectidae) of the genera *Anisops* and *Buenoa*. The haemoglobin of *Anisops assimilis* is a monomer of 16 kDa in the oxygenated state, and aggregates on deoxygenation. The Hill plot is triphasic with n = 1.1, 5.2 and 2.6. The extremely high cooperativity in physiological conditions is explained entirely by the O_2-dependent dissociation. In this case, the haemoglobin assists in the production of an O_2 gas-bubble for regulating buoyancy. *Buenoa* haemoglobin also tends to aggregate. Intracellular haemoglobins have also been reported in the accessory glands of the male genital tract of the water bug *Macrocorixa geoffroy* and in the eggs of the louse *Pediculus humanus* and the blood-sucking bug *Rhodnius prolixus* [128, 174].

Extracellular haemoglobins are restricted to the larvae of the non-biting midges (Chironomidae). These inhabit the mud regions of stagnant water and appear red due to the haemoglobin dissolved in the haemolymph. As in the case of *Daphnia* and *Artemia*, the haemoglobin concentration is higher in animals from low-O_2 biotopes. The concentration in most species is highest during the fourth larval stage, and shows a marked seasonal rhythm with a maximum in summer. In extreme cases, haemoglobins make up more than 90 % of the haemolymph protein [104]. Contrary to the rule that extracellular haemoglobins have high molecular weights, the *Chironomus* haemoglobins are monomers or dimers, and seldom tetramers *(C. strenzkei)*, of 16-kDa subunits. Dimeric forms predominate in most species. All the chironomid species possess 6–12 different haemoglobins encoded by different genes. Twelve haemoglobins have been described in the most investigated species *C. thummi thummi:* these are monomers CTT-I, -IA, -III, -IIIA, and -IV; homodimers CTT-IIβ, -VI, -VIIA, -VIIB, -VIII and -IX; and CTT-X exists as monomers and dimers in equilibrium [128].

The **amino acid sequences** are known for all 12 haemoglobins of *C. thummi thummi* [81, 108]. The chains vary in length between 136 and 151 amino acids due to extensions at both ends or to deletions. Pairwise sequence comparisons give a consistent 50 % agreement; 21 positions are invariant if CTT-IIIA is not considered, but only 12 are conserved when this haemoglobin is included in the analysis. Thus, all these haemoglobins apparently arose 200–400 million years ago from a common monomeric ancestor. The three-dimensional structure has been examined only for CTT-III (Fig. 7.4) but, including all H-bridges, haem contacts etc., the analysis is at least as thorough as for sperm whale myoglobin. Dimer formation involves the interaction of G19-His (Lys in CTT-IIβ) and G7-Glu. The monomeric forms contain G19-His but lack G7-Glu [128]. The haem has the usual proximal contacts, including F8-His, but the distal haem contacts are unusual. With the exception of CTT-I, position E11 is always occupied by isoleucine; this interacts with the haem iron and forces the E7-His out of the haem pocket. CTT-I has valine at position E11 and the Fe contact here is possibly via E7-His. E7 in CTT-IIIA is occupied by glutamine. The haemoglobins CTT-III and CTT-IV include isomers in which the haem is rotated 180°, as has already been described for vertebrate myoglobins and the haemoglobin of the annelid *Glycera* [31, 128].

It has been proposed that the site of **haemoglobin biosynthesis** in the chironomids lies not only in the fat bodies but also in the oenocytes; however, tracer experiments have now shown that haemoglobins are synthesized in subepithelial fat bodies and temporarily stored in the oenocytes.

The juvenile hormone switches on the transcription of **haemoglobin genes** and ecdysterone switches them off [125, 128]. In *C. thummi thummi*, these genes lie close together on the third of the four chromosomes, as expected for genes that have arisen by duplication. Several copies exist of CTT-III, -IV and VIIB [82]. More recently, investigations of the haemoglobins and haemoglobin genes of the subspecies *C. thummi piger* have been initiated [82, 142]. In contrast to all other globin genes from plants or animals, the sequences available so far for this family are unique in having no introns.

All the *Chironomus* haemoglobins have a high O_2 affinity (p50 = 40–170 Pa), and most have a considerable alkaline Bohr effect (φ = –0.30 to –0.94; but φ(CTT-I) = 0) and a high oxygenation enthalpy (ΔH = –40 to –81 kJ/mol) [182]. The dimeric haemoglobins also show no cooperativity (n = 0). The Bohr effect, which is also observed with the monomeric haemoglobins, results from a change in conformation which can be described in the case of CTT-III as a switch between two states: $t \leftrightarrow r + H^+$. In the t state, the distance between the haem iron and the proximal F8-His is reduced and the O_2 affinity decreases; this state is stabilized by a salt brigde between G2-His and H22-Met. This bridge, which is ultimately responsible for the Bohr effect, does not occur in CTT-I due to the exchange G2-His → Pro [31, 128].

The **biological importance** of the *Chironomus* haemoglobins lies, above all, in the improvement of O_2 diffusion in low pO_2 environments and in O_2 storage. Because, like all haemoglobins, they also have pseudoperoxidase and mono-oxygenase activity, they may also contribute to resistance against toxic substances [128]. Finally, where they constitute as much as 27 % of the dry weight of larvae, they may also have a nutritional role. It has been shown that haemoglobins are taken up by maturing oocytes, processed and stored; these altered forms disappear during embryo development [166]. The biological significance of the marked heterogeneity of *Chironomus* haemoglobins is still a puzzle.

7.2.5 Haemoglobins of Other Invertebrates

There is really very little biochemical information available about the haemoglobins in representatives of other phyla (Table 7.2). **Ciliates** of the genera *Paramecium* and *Tetrahymena* contain low concentrations (0.1–1.0 mg/g fr. wt.) of haemoglobins with molecular masses of 13–15 kDa. The globin chains of *Paramecium caudatum* and *Tetrahymena pyriformis* agree with each other in 34 % of their 116–119 amino acids but are so different from other globins that sequence alignment and comparison are very difficult. Analysis, centred upon the invariant residues CD1-Phe and F8-His, suggests the absence of the D helix and shortening of the A helix. The distal E7-His is apparently replaced by glutamine. The biological role of this pigment is unclear [73]. Haemoglobins have been detected in the parenchyme cells and the pharynx musculature of many **turbellarians and trematodes**. The only pigment to have been examined in detail, and almost completely sequenced, is that from the small liver-fluke *Dicrocoelium lanceolatum;* this is a monomeric non-cooperative haemoglobin of about 16 kDa. Its O_2 affinity is one of the highest recorded (p50 = 2–20 Pa); the explanation lies possibly in the unusual structure of the haem pocket. The distal E7-His is replaced by tyrosine and not, as initially assumed, by glycine. This haemoglobin is the only one known with an acidic but no alkaline Bohr effect (φ = +0.96) [97, 151].

Haemoglobins are widely found in both parasitic and free-living **nematodes**, either in the body cavity fluid or in various tissues. The best described are those of the round-worms (Ascaridae). The body cavity fluids of *Ascaris suum* (in the pig), *A. lumbricoides* (in humans) and *Parascaris equorum* (in the horse) contain haemoglobins of about 330 kDa, which are composed of eight subunits of 42 kDa with two haems. The haem content of up to 2.86 % corresponds to 21.6 kDa/mol haem, but is sometimes lower due to the reduced occupancy of some haem-binding sites [39]. This haemoglobin shows no cooperativity but an extremely high O_2 affinity (p50 = 0.2 Pa); this is in accordance with an O_2 dissociation constant of 0.004 s^{-1}, which is 1000-fold lower than that of mammalian haemoglobin (Table 7.5). In contrast, the dissociation constants for CO and ethylisocyanide are not significantly different from those of other haemoglobins [36]. Nematode haemoglobin is one of the few with a lower affinity for CO than for O_2. As it requires about 10 min to unload 50 % of the haemoglobin, a transport function would appear to be unlikely; however, a role as an O_2 store is not excluded. Although the roundworms have an anaerobic metabolism to supply energy, they do require oxygen for the hydroxylation of the proline in collagen. The haemoglobin in the body wall of *Ascaris* is about 40 kDa and has a low O_2 affinity. The parasite of fowl respiratory tracts,

Syngamus trachea, possesses a haemoglobin of 38 kDa which probably carries two haem groups [174]. In the marine **gastrotrich** species *Neodysis* sp. (Nemathelminthes), haemoglobins are found in special cells associated with nerve and muscle cells. These animals have such a small body size (less than 1 mm) that the function of these respiratory pigments is something of a mystery [35].

The **phoronidans** possess red blood cells from which four different globin chains of 16 kDa have been isolated in both *Phoronopsis viridis* and *P. architecta*. Two of these are always monomeric and the other two aggregate to form dimers on deoxygenation *(P. viridis)* or CO binding *(P. architecta)* [170]. Haemoglobins in the cells of the coelom fluid are widely distributed in the **holothurians** and may also occur in the **ophiurids**. Several *Cucumaria* species have been closely examined; their haemoglobins are dimers of about 35 kDa when oxygenated and aggregate to give tetramers and higher polymers on deoxygenation. They show weak cooperativity (n = 1.4), which may either involve homotropic interaction within the dimers or be the result of O_2-dependent dissociation. The coelom cells of the holothurian *Paracaudina chilensis* contain three globin chains. The sequence of 157 amino acids of globin I agrees 25 % with the human β-chain, but has an N-terminal extension of 9–10 amino acids. The N-terminus is also acetylated. Furthermore, globin I has 59 % similarity to globin D from *Molpadia arenicola*. In the latter, there are four different chain types, all of which are acetylated. In contrast, *M. oolitica* and other holothurian species have only one chain type, also with a blocked N-terminus [157].

7.3 Haemocyanins

Haemocyanins are found only in the arthropods and the molluscs (Table 7.4). Amongst the **arthropods**, haemocyanin is generally distributed in the higher crab genera (Malacostraca), the horseshoe crabs (Xiphosura) and various groups of spiders (Arachnida), but it is also present in the chilopod *Scutigera coleoptrata*. In the **molluscs**, it is found in the polyplacophores, gastropods and cephalopods and has just recently been detected in several mussel species. There have been only a few investigations of the changes in haemocyanins during development. For example, in the edible crabs of the genus *Cancer*, the haemocyanins of the embryonal and larval stages have different subunits and a lower O_2 affinity [179]. The polypeptide chains of the haemocyanins have a "binuclear" O_2-binding site with two copper atoms, Cu(A) and Cu(B), which in the deoxygenated state are present as Cu(I) and in the oxygenated state as Cu(II) (Fig. 7.1). Although both copper atoms are apparently bound to a histidine residue, the region around them may vary greatly. Sequence comparisons between haemocyanin subunits from the spiny lobster *Panulirus interruptus* (Pint-a), the spider *Eurypelma californicum* (Eury-e) and the edible snail *Helix pomatia* (β_c-Hcd) and tyrosinases from *Neurospora crassa, Streptomyces glaucescens* and the mouse give only one region with significant similarity, viz., the 42–60 amino acids around the Cu(B). Thus, whilst Cu(B) is apparently phylogenetically very old, Cu(A) appeared on several independent occasions. In contrast, the Cu-binding sites (one dinuclear and two mononuclear) of a third type of copper protein, caeruloplasmin, show no similarity at all to haemocyanins or tyrosinases [100, 108, 109, 178].

The **quaternary structures** of arthropod and mollusc haemocyanins are very different but, in both cases, extraordinarily complex. Mollusc haemocyanins are cylindrical molecules with a diameter of about 35 nm, a length of 38 nm and a molecular mass of up to 10 MDa; they consist of 10 or 20 subunits, each with seven or eight O_2-binding domains of about 55 kDa. In contrast, the arthropod haemocyanins are composed of one, two, four or eight hexameric basic units which are themselves made up of 75-kDa subunits, each with one O_2-binding site. Each molecule may consist of up to eight different types of subunit. Some haemocyanins contain carbohydrate, and the structures of the oligosaccharides differ significantly between arthropods and molluscs (see Fig. 7.10) [91, 92]. The scientific interest in the haemocyanins extends far beyond their role in the respiratory physiology of arthropods and molluscs; they not only have very well-defined and measurable functions but also exhibit molecular structures of great complexity. It is difficult to envisage a better protein for studying the way in which interaction between different subunits and domains can lead to the structural and functional properties of a very large molecule. Particularly intensive research in recent years has contributed significantly to our knowledge of the haemocyanins [108, 174]. The abbreviation "Hc" should be used for haemocyanin to avoid confusion with homocysteine "Hcy" [108].

The **O_2 affinity** of the haemocyanins is determined to a large extent by the pH, ion concentration and temperature. Because **cooperativity** is influenced by the same factors, the Hill constants for arthropod and mollusc haemocyanins usually have values of only 3–4; however, extremely high values may be achieved under optimal conditions, e.g. n = 7 for β_c-Hc of the edible snail *Helix pomatia*, and n = 7–9 for the haemocyanin of the spider *Eurypelma californicum* [108]. CO also binds to the copper atoms of haemocyanin, but in the arthropods with a much lower affinity and cooperativity than does O_2 [40]. An attempt has been made many times to describe the cooperativity of the haemocyanins in terms of the MWC model, but there are in fact few observations of relevant conformational changes. For example, the haemocyanins of *Helix pomatia* and *Eurypelma californicum* undergo changes in Cu-Cu spacing and the number of Cu ligands on oxygenation which involve conformational changes in the protein bridges between the copper atoms [99, 108]. In terms of the molecular mass of a respiratory pigment which is required for the binding of one oxygen, the haemocyanins (75 kDa for the arthropods and 55 kDa in molluscs) are significantly inferior to the haemoglobins (16 kDa). However, the arthropod haemocyanins in particular are modulated by allosteric factors to an extent found in vertebrate, but rarely in invertebrate, haemoglobins [115].

The O_2 affinity of the haemocyanins is influenced by **heterotropic interaction** with H^+, Cl^-, Ca^{2+}, Mg^{2+} and Na^+. The transport of O_2 in whole animals is in fact regulated mainly by pH changes which affect haemocyanin directly and without the damping effect of an intervening erythrocyte membrane. Organic factors, such as lactate and uric acid, also function as modulators in the crustaceans [108]. The pH effects on affinity and cooperativity are independent, suggesting the involvement of different protonated groups. The pH dependence of the p50 of haemocyanins in *Limulus* and several snails is given by a maximum curve, e.g. in *Buccinum undatum* a maximal p50 value of 7.7 kPa is reached at pH 8.1; at pH values below the maximum, an increase in pH reduces the O_2 affinity (a reverse Bohr effect). This pH effect is so great in *B. undatum* that the pigment cannot be completely saturated (reverse Root effect). The biological sense of these inverted pH effects is that under low oxygen conditions acidic metabolic products can increase the O_2 affinity of haemocyanin [23]. Most haemocyanins have a normal Bohr effect which varies greatly with the species and conditions. Particularly pronounced Bohr effects are found in some arthropods with φ values of up to –1.3; normal Root effects have been reported for *Octopus dofleini* and *Panulirus interruptus* [120]. Because the oxygenation of haemocyanins is also an exothermic process, the O_2 affinity is generally reduced with an increase in temperature. The ΔH values for various arthropod and mollusc haemocyanins were –13 to –47 kJ/mol at temperatures between 15 and 25 °C [29]. It has been demonstrated by the detection of intracellular haemocyanins or the corresponding mRNA that the **biosynthesis** of haemocyanins occurs on membrane-bound ribosomes of the branchial glands of cephalopods and on free ribosomes in so-called cyanocytes of the arthropods. The latter cells are found especially in tissues behind the eye of *Limulus polyphemus*, in the hepato-pancreas and close to the eye artery of the crustaceans *Carcinus, Cancer* and *Astacus*, and on the inner wall of the heart of the spider *Eurypelma* [48]. Sequencing of haemocyanin mRNAs via the corresponding cDNAs is currently in progress [177]. Tracer experiments indicate haemocyanin biosynthesis in the mantle tissues of the snail *Lymnaea stagnalis*. This creature secretes a mixture of air and haemolymph from the haemal pore to scare predators, losing in the process 1–3 mg haemocyanin. Although the rate of synthesis subsequently increases fivefold, it requires 14 days to make up the loss. Iodinated species-specific haemocyanin is degraded in the lobster *Homarus americanus* with a half-life of 26 days.

7.3.1 Haemocyanins of the Arthropods

The arthropod haemocyanins are hexamers, or multiples thereof, with subunits of about 75 kDa bearing a single O_2-binding site. Various aggregates with differing sedimentation coefficients can be distinguished: 16S (6-mer), 24S (12-mer), 36S (24-mer) and 60S (48-mer). One or two of these **aggregation states** predominates in each arthropod species, depending upon the number of different subunits and the possibilities for association. In the crabs, the 24S 12-mer is characteristic for the Astacura, Brachyura and Anomura; a few species of this group have predominantly or exclusively 6-mers. For example, in the semi-terrestrial crab *Ocypode quadrata*, 56 % of the haemocyanin is 6-mer and 44 % is 12-mer [77], whereas the crab *Uca urvillei* and the crayfish *Jasus* sp. have only 6-mers. Larger aggregates in

addition to 6-mers and 12-mers are reported to occur in the crayfish *Cherax destructor.* The 6-mers are the rule in the Palinura *(Panulirus, Scyllarus)*, Natantia *(Palaemon)* and Euphausiacea [108, 115, 178]. The haemocyanins of most Chelicerata are larger than those of the Crustacea, with only the Xiphosura *(Limulus, Tachypleus)* having 48-mers, the scorpions, uropygi, amblipygi and most spiders having 24-mers, and the opiliones having 12-mers [16, 108, 115]. Some spider species have haemocyanin molecules smaller than 24-mer: *Cupiennius salei* has 12-mers, and 6-mers occur in *Filistata insidiatrix* and *Dysdera crocata* [108, 115]. The haemocyanin of the chilopod *Scutigera coleoptrata* is apparently made up of six hexamers [108].

The **dissociation** of native haemocyanins into subunits is a slow process which is promoted at alkaline pH, and strongly promoted by the removal of divalent cations and the addition of denaturing reagents. Thus, dissociation of haemocyanin in *Limulus* and other chelicerates is initiated only at pH 9.6 in the presence of Ca^{2+}, but at pH 8.8 upon addition of 10 mM EDTA or 2% SDS + 1% mercaptoethanol. Some crustacean haemocyanins can resist dissociation at pH 10 in the presence of Ca^{2+}. Reaggregation at neutral pH in the presence of a sufficiently high Ca^{2+} concentration is often incomplete or results in products which are not identical to the native haemocyanin [77, 108]. The subunits of the arthropod haemocyanins are polypeptides with about 630–670 amino acids, 20 α-helices and a seven-stranded β structure; they are subdivided into three domains, the middle domain bearing the O_2-binding site. They appear in electron micrographs as kidney-shaped structures (Fig. 7.9); these interpretations of the secondary and tertiary structures have been confirmed by X-ray analysis of the haemocyanin subunits of the spiny lobster *Panulirus interruptus* [108, 115, 178].

Most arthropod species possess several electrophoretically or immunologically distinct **types of haemocyanin subunit**. In crustacean haemocyanin, three types of subunit can usually be distinguished: α, β and γ. Types α and γ are related immunologically, whilst β forms a separate class. Two different α-types may often be found. In some crustacean species or groups, the haemocyanins have a simpler organization: the 6-mers of *Uca urvillei, Euphausia superba* and *Palinurus vulgaris* appear to have only type α, the 6-mers of *Panulirus interruptus* and *Palaemon elegans* only α and γ, and the 12-mers of the Grapsidae only α and β. The larger haemocyanin molecules of the

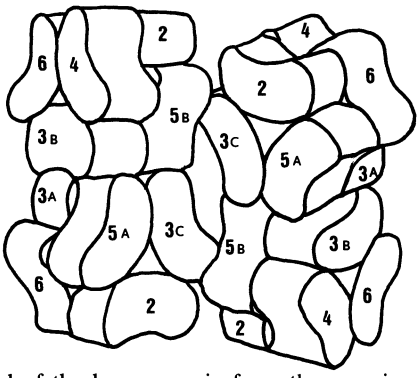


Fig. 7.9. A model of the haemocyanin from the scorpion *Androctonus australis* [150]. The molecule is a 24-mer built up from eight different subunits designated 2, 3A, 3B, 3C, 4, 5A, 5B and 6

chelicerates have a very complicated composition. The 48-mer of the xiphosuran *Limulus polyphemus* consists of eight different subunits (I, II, IIa, IIIa, IIIb, IV, V, VI), as does the 24-mer of the scorpion *Androctonus australis* (1, 2, 3A, 3B, 3C, 4, 5A; Fig. 7.9). The 24-mer of the spider *Eurypelma californicum* has seven subunits (a–h). In contrast, only two types of subunit are recognizable in the 12-mers of the spider *Cupiennius salei* and the harvestman *Leiobunum limbatum* [108, 115, 126, 155].

The **amino acid sequences** of various haemocyanin subunits have been partially or completely determined for the crustaceans *Panulirus interruptus, Palinurus vulgaris* and *Astacus leptodactylus* and for the chelicerates *Limulus polyphemus, Tachypleus tridentatus* and *Eurypelma californicum* [5, 108, 126, 127, 143, 177]. Asparagine-linked carbohydrates have been detected in two haemocyanin subunits (Pint-a and -b) of the spiny lobster *Panulirus interruptus* and in the haemocyanin of the scorpion *Androctonus australis*. In the latter, the classical type of N-linked oligosaccaride $(Man)_9(GlcNAc)_2$ is found, whereas the lobster haemocyanin contains very uncommon carbohydrate chains (Fig. 7.10a) [91, 92]. The chelicerate subunits differ from each other by 43–47% and from the α-chain of *P. interruptus* in 69–70% of positions. A significant difference between the haemocyanin sequences of these two arthropod groups is the presence in crustacean, but not chelicerate, subunits of a loop of 21 amino acids (positions 22–42) [5]. Based upon the known three-dimensional structure of the sequenced *P. interruptus* α-chain, all subunits would appear to have a similar spatial structure. Differences in sequence are greater in the terminal domains than in the central O_2-binding domain [108]. The

binding-site sequences of Cu(A) and Cu(B) of the crustaceans (HHVTWH and HNTAH) and the chelicerates (HHWHWH and HNWGH) are very similar [108]. Comparison of the presumably orthologous subunits Lim-II/Eury-a and Tachy-α/Eury-e suggests a moderate rate of evolution for the haemocyanins of 0.72–0.85 · 10^{-9} amino acid substitutions per year (see Table 4.12; p. 161). The gene duplications leading to the appearance of multiple subunits would appear to have occurred before the separation of the Arachnida and the Xiphosura [126]. Surprisingly, there is significant homology between arthropod haemocyanins and several haemolymph proteins of the lepidoptera, e.g. the SP1 protein of *Bombyx mori*, the arylphorin of *Manduca sexta* and the acidic juvenile hormone-suppressible haemolymph protein, AJSP-1, of *Trichoplusia ni* [78, 116].

The number of molecules in the different **subunits in native haemocyanin molecules**, their organization and their contribution to the coherence of the molecule have been investigated in *Limulus polyphemus, Androctonus australis* and *Eurypelma californicum* by means of four different approaches:

1. Determination of the stoichiometry of the individual subunits.
2. Analysis of the fragments after dissociation.
3. Reaggregation experiments with different combinations of subunits.
4. Immuno-electron microscopy, i.e. electron micrographic analysis of haemocyanin molecules labelled with subunit-specific antibodies or antibody fragments [115].

The 48-mer of *Limulus* dissociates under mild conditions into 24-mers which resemble those of *Androctonus* (Fig. 7.9) or *Eurypelma*. In the latter species, the two hexamers of the 12-mers are linked via two bc dimers, and the two 12-mers of the 24-mer are linked via interaction between the f subunits. This is also true for the 24-mers of *Limulus* and *Androctonus*. The hexamers in the 12-mer of the spider *Cupiennius* are held together by a disulphide bridge, as are the hexamers of the 12-mers in the crayfish *Cherax destructor* and *Astacus leptodactylus;* in the case of *Homarus americanus* and *Cancer magister*, the link appears to be non-covalent [108]. The similarity of the contacts between the subunits of chelicerates (e.g. *Limulus, Androctonus* and *Eurypelma*) is such that hybrid molecules are formed in reaggregation experiments with subunits from different species [115]; despite immunological differences, the subunits of the crayfish *Cherax* and *Jasus* can also form hybrid hexamers [108]. There are, however, exceptions to this scheme. For example, the 12-mer of the stomatopod *Squilla mantis* appears in electron micrographs as two hexamers which overlap by only 60 % [12].

The native haemocyanin of *Eurypelma californicum* is characterized by a low O_2 affinity, high cooperativity with n values of more than 9.0, and a strong alkaline Bohr effect. Isolated subunits bind O_2 with a higher affinity but exhibit neither cooperativity nor a Bohr effect. The hexamer already has the normal O_2 affinity and Bohr effect; however, cooperativity increases stepwise from the hexamer to the 12-mer and then to the 24-mer. This phenomenon is described as "nesting" in **cooperation models:** the conformation of the multimeric molecule influences ligand binding to the subunits in a hierarchical manner with several steps. According to this idea, the 24-mer of *Eurypelma* consists of two allosteric 12-mers whose properties are dependent upon the conformation of the 24-mer. The nesting model can

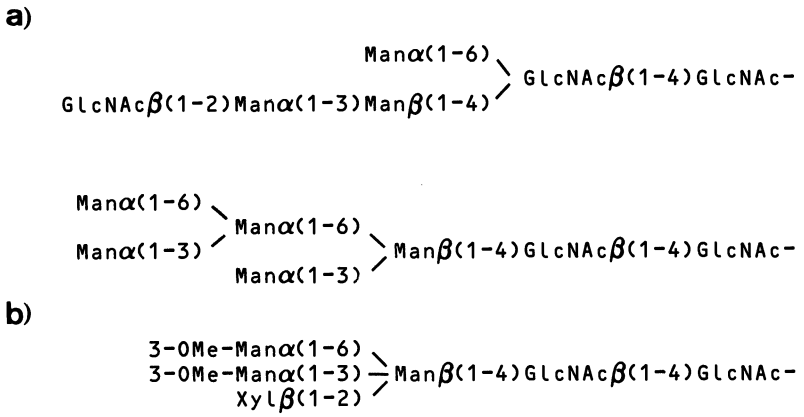


Fig. 7.10a, b. The carbohydrate chains of the haemocyanins from **a** the spiny lobster *Panulirus interruptus* [91] and **b** the pulmonate snail *Lymnaea stagnalis* [93]

also be used to describe the cooperativity of the 12-mer from *Homarus americanus* and of the 48-mer of *Limulus polyphemus*. Cooperative basic units present in both 6-mers and 12-mers may each be in either the T or the R state [25, 40]. Homohexamers can arise from certain haemocyanin subunits of various crustaceans and chelicerates; these may even show weak cooperativity, e.g. the subunit Lim-II with n = 1.8 [108, 111].

The importance of subunit interaction for the respiratory properties of the whole molecule can be examined by the use of toxic mercury salts. Each *Eurypelma* subunit binds 1–2 Hg^{2+}; 24-mers formed from reaggregated subunits treated in this way have the same binding characteristics as isolated subunits, i.e. all interaction is blocked. The use of individually poisoned subunits in the reaggregation of the 24-mers of *Eurypelma* or of the 48-mers of *Limulus* has facilitated the study of their roles in the complete molecule [108, 115]. The O_2 affinity of haemocyanins from many crustaceans is increased by L-lactate. D-lactate, glycolate and pyruvate are less active and D- and L-alanine have no effect; thus, all four substituents of the asymmetric carbon atom are involved in the interaction. The strength of the reaction varies with the species; the value of $\Delta \log p50/\Delta$[L-lactate] varies from –0.096 *(Carcinus maenas)* to –0.560 *(Palaemon elegans)*. The strength of the Bohr effect, with φ values between –0.38 *(Maja squinado)* and –1.40 *(Liocarcinus depurator)*, is not correlated to the lactate effect [108]. The haemolymph contains other unknown haemocyanin modulators (UF, unidentified factors) in addition to lactate; a part of the UF activity can be ascribed to uric acid [94, 108].

7.3.2 Haemocyanins of the Molluscs

The edible snail *Helix pomatia* possesses three different **haemocyanin forms** to which the names α-Hc, β_s-Hc and β_c-Hc were initially given. The α-Hc, which makes up 75 % of the total, is characterized by the fact that it dissociates into two halves in 1 mol/l NaCl at pH 5.7. Of the non-dissociated β-Hc, one part (30–50 %) precipitates at pH 5.3 (β_c-Hc) and the rest remains in solution (β_s-Hc). After it was discovered that α-Hc and β_s-Hc are made up from the same subunits (α and α′), the haemocyanins were renamed α_D-Hc (dissociating) and α_N-Hc (non-dissociating); the name β_c-Hc for the third component, which is made up of different (type β) subunits, was retained. β_c-Hc is the most investigated form. The haemocyanin of *Helix aspersa* may also be divided into three fractions [52]. In electron micrographs, the **haemocyanin molecules** from *Helix pomatia* and other gastropods appear as hollow cylinders that are 35 nm in diameter and 38 nm long; the ends are partly closed by two so-called collars (Fig. 7.11a). This haemocyanin sediments in the ultracentrifuge with 100S, corresponding to 9 MDa, but molecules of 120S, i.e. with 1.5 times greater mass, are also found. At weakly alkaline pH and low Ca^{2+} concentration, gastropod haemocyanin molecules split into fragments of 60S (half), 20S (1/10) and 11S (1/20). Each subunit (1/20) consists of eight domains, each with one O_2-binding site, and appears in the electron microscope as a chain of eight spheres with a diameter of 5.5 nm. The copper content of about 0.23 % gives a mass of 55 kDa per O_2-binding site. In marine snails of the family Melongenidae, which includes the genera *Melongena* and *Busycon*, the haemocyanins exist as extended, tubular aggregates of five or more 60S half molecules; they are consistently organized such that both ends of the tubulus are closed by a collar [63]. Partial tryptic digest of the **subunits** gives eight fragments (a–h) with a mass of 55 kDa, six of which form the wall of the tubulus and two form the collar. In the 1/10 fragments, the two subunits are apparently arranged in parallel (Fig. 7.11a). The eight fragments of the subunits differ greatly in their amino acid and carbohydrate composition, and in their functional properties. For example, the carbohydrate content in *Helix pomatia* varies from 17.6 % for fragment g to 0.7 % for fragment c, with a mean of 8.2 % [108, 188]. The structures of the carbohydrates have been analysed in *H. pomatia* and *Lymnaea stagnalis* (Fig. 7.10b). They differ from the N-linked oligosaccharides of other animals by the presence of xylose and O-methylated sugars. The association of further sugars can lead to their transformation into complex carbohydrates of greater molecular mass [93]. The methylated sugars are not incorporated as such, but are later methylated with adenosylmethionine as the cofactor.

The functional subunit d from the β_c-Hc of *H. pomatia* has been completely sequenced. The **sequence** of 410 amino acids agrees 42 % with the 399-amino-acid C-terminal domain of octopus Hc, which was sequenced via the cDNA [96]; however, it exhibits homology to arthropod haemocyanins only in a small, 42-amino-acid fragment. This segment corresponds to the Cu(B) region of the haemocyanin from the spiny lobster *Panulirus interruptus;* homologous sequences are

also found in various tyrosinases. The second Cu-binding site of the *Helix* subunit has not yet been identified but is, in any case, completely different to the Cu(A) of the arthropods [42]. It was previously assumed that the *Helix* Hc model was a generalization valid for all gastropod haemocyanins. However, accurate measurements of the haemocyanin of the pulmonate snail *Levantina hierosolima* give 10.4 MDa for the whole molecule and 334 kDa for the subunit; also in this case, the copper content of 0.23 % gives a mass of 55 kDa per O_2-binding site. These data do not fit the model of 20 subunits each with eight domains, but rather correspond to a molecule with 30 subunits each with six domains [108].

It has recently been discovered that members of the most primitive mussel order **(Protobranchia)** also possess haemocyanins. The pigments of the species *Acila castrensis, Yoldia thraciaefromis, Y. limatula* and *Nucula hanleyi* are very similar to the gastropod haemocyanins, both in the size of the native molecules and their subunits and in their appearance in electron micrographs [62, 164, 165]. The haemocyanins of the **Polyplacophora** and **Cephalopoda** are 60S and therefore only half the size of the gastropod haemocyanins. The structure of the haemocyanins of the polyplacophores *Katharina tunicata, Acanthopleura granulata, Crytochiton stelleri* and *Stenoplax conspicua* also corresponds to a half helical cylinder with one collar [63]. In contrast, the cylindrical haemocyanin molecule of the cephalopod *Sepia officinalis* with two collars resembles more the whole cylinder of the edible snail but in a shortened form (Fig. 7.11b). Of the eight O_2-binding domains present here, b–g lie in the cylinder wall, whilst a and h form the collars at either end. The haemocyanin molecule of the cephalopod *Nautilus pompilius* also includes two collars [63, 184]. The haemocyanin subunits of *Octopus dofleini* and *O. vulgaris* contain only seven, rather than eight, domains [63, 108, 121].

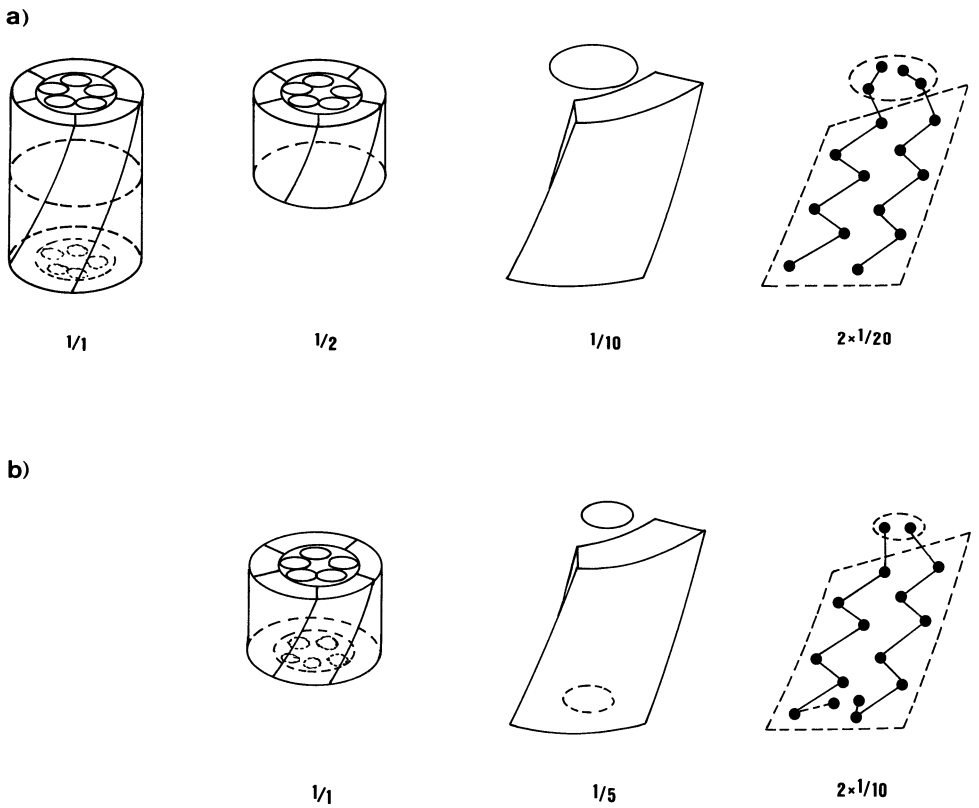


Fig. 7.11a, b. Models of mollusc haemocyanins [184]. **a** In the edible snail *Helix pomatia*, the haemocyanin molecule is a hollow cylinder closed at both ends by a collar. The first dissociation step divides the molecule transversely to the longitudinal axis into identical halves *(1/2)*. Each half consists of five wall fragments, each with a section of collar *(1/10)*, and each 1/10 unit is made up of two identical polypeptide chains *(1/20)* each with eight domains, of which six lie in the wall and two lie in the collar. **b** In the cephalopod *Sepia officinalis*, the haemocyanin molecule is similar in size to one half of the *Helix* molecule but bears a collar at both ends. The molecule is made up of five identical sections, each consisting of two parallel polypeptide chains with eight domains; the first and last domains form a collar

The majority of mollusc haemocyanins consist of only one type of subunit, but electrophoretically or immunologically distinct types are found in some species. The 100S haemocyanin of the snail *Murex fulvescens* is made up of ten each of two subunits, A and B, which differ especially in their respiratory properties. Reassociation products with the characteristics of the native molecule result only from mixtures of the two subunit types; A or B alone produces only dimers. **Heterogeneous subunits** are also found in *Nautilus pompilius*, the snail *Megathura crenulata*, and the Polyplacophora [63, 108]. Single subunits, and even isolated domains, can bind O_2 but generally exhibit no cooperativity and a reduced Bohr effect. Reassociation of the subunits usually leads to molecules with the structural and functional properties of the native haemocyanin [108].

7.4 Haemerythrins

The haemerythrins have a curious, sporadic distribution (Table 7.2). Haemerythrin-containing cells (erythrocytes) are found in the coelum fluid of all sipunculids and priapulids as well as several species of the ecardinate brachiopods (e.g. *Lingula reevei*), and are present in the enucleate blood cells and muscles of the polychaete *Magelona papillicornis*. Blood cells and muscles of the sipunculids also contain haemerythrins with somewhat different properties (myohaemerythrins). The myohaemerythrins are monomers of about 13.5 kDa, whereas the pigments in the blood and coelom cells of the sipunculids, priapulids and brachiopods are trimers, tetramers or octomers of subunits with the same size [83, 108, 174]. The major vascular haemerythrin of *Magelona* has the unusually large size for an intracellular respiratory pigment of 693 ± 32 kDa [113]. Myohaemerythrins and the subunits of the polymeric haemerythrins have a similar tertiary structure with four helical segments, A–D (Fig. 7.12). The octomers always consist of two layers, each with four subunits which are joined "head to side" by hydrogen bonds; the layers are held together mainly by electrostatic interactions [83].

The sequences are known for the 118 amino acids of myohaemerythrin from *Themiste zostericola* and for the 113 amino acids of the coelom-cell haemerythrin from *Phascolopsis gouldii* and *Themiste dyscritum*. The coelom-cell haemerythrins agree 79 % with each other and 42 % with the muscle pigments; there is no homology to the

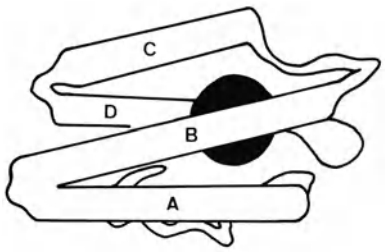


Fig. 7.12. A haemerythrin chain with the four helices *A–D* and the O_2-binding site *(black)* [83]

haemoglobins. Both the amino acid sequences and the tertiary structures exhibit marked symmetry between the two halves of the molecule, suggesting the occurrence of a duplication sometime during haemerythrin evolution. The haemerythrins are usually distinctly polymorphic: at least 20 different forms can be distinguished in *Siphonosoma* sp. and at least 10 in *Phascolosoma aggassizii* [83]. The two iron atoms constituting the active centre are complexed with three histidines and one aspartate of helices C and D as well as with two histidines and one aspartate of helices A and B [190]. In deoxygenated haemerythrin, both irons are present as Fe(II), and in the oxygenated form they occur as Fe(III); the oxygen is bound by means of oxidative addition with electron transfer (Fig. 7.1). The oxidation of one or both iron atoms leads to semi-methaemerythrin and methaemerythrin, which can still bind azide ions but not oxygen [83, 108]. A methaemerythrin-reducing system, consisting of cytochrome b_5 and an NADH-cytochrome b_5 reductase, has been demonstrated in *Phascolopsis gouldii* [168].

The p50 values for the haemerythrins consistently lie between 0.1 and 2 kPa. The haemerythrins of the blood cells (p50 = 2.0 kPa), coelom cells (p50 = 0.47 kPa) and muscles (p50 = 0.15 kPa) of *Themiste zostericola* make up a transport chain. Homotropic and heterotropic interactions generally appear to be absent in the octomeric haemerythrins, although the haemerythrin of the brachiopod *Lingula reevei* exhibits weak cooperativity (n = 1.7–1.8) at pH 7–8 and an alkaline Bohr effect. The responsible conformational changes of the octomer are partly described by the MWC model if, in addition, a T/R hybrid condition is assumed. The conformation changes between pH 7.7 (mainly oxy-R) and pH 6.3 (T + T/R + R) can be directly seen with resonance-Raman spectroscopy [138, 191]. *Magelona* haemerythrin shows cooperativity (n = 2.1)

only at high O_2 saturation levels; the Bohr effect is very small ($\varphi = -0.13$) [113].

References

1. Abbasi A. and Braunitzer G.: The primary structure of hemoglobins from the domestic cat (Felis catus, Felidae). Biol. Chem. Hoppe-Seyler 366: 699–704 (1985)
2. Abbasi A. et al.: Molecular basis for ATP/2,3-bisphosphoglycerate control switch-over (poikilotherm/homeotherm): An intermediate amino-acid sequence in the hemoglobin of the great Indian rhinoceros (Rhinoceros unicornis, Perissodactyla). Biol. Chem. Hoppe-Seyler 368: 323–332 (1987)
3. Arents G. and Love W. E.: Glycera dibranchiata hemoglobin. Structure and refinement at 1.5 Å resolution. J. mol. Biol. 210: 149–161 (1989)
4. Aschauer H., Weber R. E. and Braunitzer G.: The primary structure of the hemoglobin of the dogfish shark (Squalus acanthias): Antagonistic effects of ATP and urea on oxygen affinity of an elasmobranch hemoglobin. Biol. Chem. Hoppe-Seyler 366: 589–599 (1985)
5. Bak H. J. and Beintema J. J.: Panulirus interruptus hemocyanin – The elucidation of the complete amino acid sequence of subunit a. Eur. J. Biochem. 169: 333–348 (1987)
6. Banville D. and Williams J. G.: The pattern of expression of the Xenopus laevis tadpole α-globin genes and the amino acid sequence of the three major tadpole α-globin polypeptides. Nucleic Acids Res. 13: 5407–21 (1985)
7. Banville D. and Williams J. H.: Developmental changes in the pattern of larval β-globin gene expression in Xenopus laevis. Identification of two early larval β-globin mRNA sequences. J. mol. Biol. 184: 611–620 (1985)
8. Bartlett G. R.: Phosphate compounds in vertebrate red blood cells. Amer. Zool. 20: 103–114 (1980)
9. Bashford D., Chothia C. and Lesk A. M.: Determinants of a protein fold. Unique features of the globin amino acid sequences. J. mol. Biol. 196: 199–216 (1987)
10. Bellelli A. et al.: Haem disorder in two myoglobins: comparison of reorientation rate. Biochem. J. 246: 787–789 (1987)
11. Bieber F. A. and Braunitzer G.: Prae- and perinatal oxygen transport in mammals: The embryonic hemoglobins of the pig (Sus scrofa domestica). Hoppe-Seyler's Z. physiol. Chem. 365: 321–334 (1984)
12. Bijlholt M. and van Bruggen E. F. J.: A model for the architecture of the hemocyanin from the arthropod Squilla mantis (Crustacea, Stomatopoda). Eur. J. Biochem. 155: 339–344 (1986)
13. Biswanger H.: Theorie und Methoden der Enzymkinetik. Verlag Chemie, Weinheim 1979
14. Blanchetot A., Price M. and Jeffreys A. J.: The mouse myoglobin gene – Characterization and sequence comparison with other mammalian myoglobin genes. Eur. J. Biochem. 159: 469–474 (1986)
15. Bogusz D. et al.: Functioning haemoglobin genes in non-nodulating plants. Nature 331: 178–180 (1988)
16. Boisset N. et al.: Three-dimensional reconstruction of native Androctonus australis hemocyanin. J. mol. Biol. 216: 743–760 (1990)
17. Borgese T. A., Harrington J. P. and Hoffman D.: Anadara ovalis hemoglobins: distinct dissociation and ligand binding characteristics. Comp. Biochem. Physiol. Pt. B. 86: 155–165 (1987)
18. Borgese T. A. et al.: Haemoglobin properties and polymerization in the marine teleost Lophius americanus (goosefish). Comp. Biochem. Physiol. Pt. B 91: 663–670 (1988)
19. Braunitzer G. and Hiebl I.: Molecular aspects of high altitude respiration in birds (In German). Naturwissenschaften 75: 280–287 (1988)
20. Brittain T. and Wells R. M. G.: Characterization of the changes in the state of aggregation induced by ligand binding in the hemoglobin system of a primitive vertebrate, the hagfish Eptatretus cirrhatus. Comp. Biochem. Physiol. Pt. A 85: 785–790 (1986)
21. Brittain T.: The Root effect. Comp. Biochem. Physiol. Pt. B 86: 473–481 (1987)
22. Brittain T.: Co-operative functioning of the dimeric haemoglobin obtained from the radular muscle of the amphineurian mollusc Amaurochiton glaucus. Comp. Biochem. Physiol. Pt. B 96: 96–295 (1990)
23. Brix O.: The adaptive significance of the reversed Bohr and Root shifts in blood from the marine gastropod, Buccinum undatum. J. exp. Zool. 221: 27–36 (1982)
24. Brix O. et al.: The chloride shift may faciliatate oxygen loading and unloading to/from the hemoglobin from the brown bear (Ursus arctos L.). Comp. Biochem. Physiol. Pt. B 95: 865–868 (1990)
25. Brouwer M. and Serigstad B.: Allosteric control in Limulus polyphemus hemocyanin: Function relevance of interactions between hexamers. Biochemistry 28: 8819–27 (1989)
26. Brunori M. et al.: Is there a Root effect in Xenopus hemoglobin? FEBS Letters 221: 161–166 (1987)
27. Bunn H. F.: Regulation of hemoglobin function in mammals. Amer. Zool. 20: 199–211 (1980)
28. Bunn H. F. and Forget B. G.: Hemoglobin. Molecular, genetic and clinical aspects. Saunders, Washington 1986
29. Burnett L. E., Scholnick D. A. and Mangum C. P.: Temperature sensitivity of molluscan and arthropod hemocyanins. Biol. Bull. 174: 153–162 (1988)
30. Cardellini P and Sala M.: Developmental time of the hemoglobin transition in the anuran Bombina orientalis. Comp. Biochem. Physiol. Pt. B 75: 259–262 (1983)
31. Chacko V. P. et al.: Proton-magnetic-resonance investigation of the dynamics of the conformational transition in allosteric monomeric insect hemoglobins. Eur. J. Biochem. 161: 375–381 (1986)
32. Cheng, J. F., Krane D. E. and Hardison R. C.: Nucleotide sequence and expression of rabbit globin genes zeta-1, zeta-2 and zeta-3. Pseudogenes generated by block duplications are transcriptionally competent. J. Biol. Chem. 263: 9981–93 (1988)
33. Cirotto C. and Arangi I.: Koelliker haemoglobins in developing chick embryo. Comp. Biochem. Physiol. Pt. B 92: 103–109 (1989)
34. Clegg J. B.: Gene conversions in the horse α-globin gene complex. Mol. Biol. Evol. 4: 492–503 (1987)

35. Colacino J. M. and Kraus D. W.: Hemoglobin-containing cells of Neodasys (Gastrotricha, Chaetonotida) – II. Respiratory significance. Comp. Biochem. Physiol. Pt. A 79: 363–369 (1984)
36. Coletta M. et al.: Ligand-dependent behaviour of the hemoglobin from the ascarid Parascaris equorum. Biochim. biophys. Acta 870: 169–175 (1986)
37. Coletta M. et al.: A novel mechanism of heme-heme interaction in the homodimeric hemoglobin from Scapharca inaequivalis as manifested upon cleavage of the proximal Fe-N bond at low pH. J. Biol. Chem. 265: 4828–30 (1990)
38. Dafré A. L. and F ° D. W.: Root effect hemoglobins in marine fish. Comp. Biochem. Physiol. Pt. A 92: 267–471 (1989)
39. Darawshe S., Tsafahdya Y and Daniel E.: Quaternary structure of erythrocruorin from the nematode Ascaris suum. Evidence for unsaturated haem-binding sites. Biochem. J. 242: 689–694 (1987)
40. Decker H. and Sterner R.: Nested allostery of arthropodan hemocyanin (Eurypelma californicum and Homarus americanus). J. mol. Biol. 211: 281–293 (1990)
41. Douglas E. L. et al.: Myoglobin in the heart tissue of fishes lacking hemoglobin. Comp. Biochem. Physiol. Pt. A 81: 885–888 (1985)
42. Drexel R. et al.: Complete amino-acid sequence of a functional unit from a molluscan hemocyanin (Helix pomatia). Biol. Chem. Hoppe-Seyler 368: 617–635 (1987)
43. Ellerton H. D., Bearman C. H. and Loong P. C.: Erythrocruorin from the New Zealand earthworm Maoridrilus montanus: a multi-subunit annelid extracellular hemoglobin. Comp. Biochem. Physiol. Pt. B 87: 1017–23 (1987)
44. Fitch D. H. A. et al.: Molecular history of gene conversions in the primate fetal gamma-globin genes. Nucleotide sequences from the common gibbon, Hylobates lar. J. Biol. Chem. 265: 781–793 (1990)
45. Focesi A. jr., Ogo S. H. and Matsuura M. S. A.: Dimer-tetramer transition in hemoglobins from Liophis miliaris. II. Evidence with the stripped proteins. Comp. Biochem. Physiol. Pt. B 96: 119–122 (1990)
46. Garey J. R. and Riggs A. F.: The hemoglobin of Urechis caupo. The cDNA-derived amino acid sequence. J. biol. Chem. 261: 16446–50 (1986)
47. Garner K. J. and Lingrei J. B.: A comparison of the βA- and βB-globin gene clusters of sheep. J. mol. Evol. 28: 175–184 (1989)
48. Gelissen G., Hennecke R. and Spindler K. D.: The site of synthesis of hemocyanin in the crayfish, Astacus leptodactylus. Experientia 47: 194-/195 (1991)
49. Giardina B. et al.: Interaction of hemoglobin with chloride and 2,3-bisphophoglycerate: A comparative approach. Eur. J. Biochem. 194: 61–65 (1990)
50. Gibson Q. H. et al.: Ligand binding in a hierarchy of globin complexes: The hexagonal bilayer hemoglobin of Lumbricus terrestris and its heme-containing subunits. J. Biol. Chem. 266: 13097–102 (1991)
51. Gielens C. et al.: Identification, separation and characterization of the haemocyanin components of Helix aspersa. Comp. Biochem. Physiol. Pt. B 88: 181–186 (1987)
52. Gonzalez-Redondo J. M. et al.: Nucleotide sequence of the human theta-1 globin gene. Biochem. Genet. 26: 207–211 (1988)
53. Goodman M. et al.: Amino acid sequence evidence on the phylogeny of primates and other eutherians. In: Goodman M. (ed.): Macromolecular sequences in systematics and evolutionary biology, pp. 115–191. Plenum, New York 1982
54. Goodman M. et al.: The eta-globin gene. Its long evolutionary history in the β-globin gene family of mammals. J. mol. Biol. 180: 803–823 (1984)
55. Goodman M. et al.: An evolutionary tree for invertebrate globin sequences. J. mol. Evol. 27: 236–249 (1988)
56. Gotoh T. and Suzuki T.: Molecular assembly and evolution of multi-subunit extracellular annelid hemoglobines. Zool. Sci. 7: 1–16 (1990)
57. Grinich N. P., Terwilliger R. C. and Terwilliger N. B.: Oxygen equilibria and structural characteristics of the tetrameric and polymeric intracellular hemoglobins from the bivalve mollusc Barbatia reeveana. J. comp. Physiol. B 156: 675–682 (1986)
58. Hardison R. C. and Gelinas R. E.: Assignment of orthologous relationships among mammalian α-globin genes by examining flanking regions reveals a rapid rate of evolution. Mol. Biol. Evol. 3: 243–261 (1986)
59. Hardison R. C. et al.: A previously undetected pseudogen in the human alpha globin gene cluster. Nucleic Acids Res. 14: 1903–11 (1986)
60. Harris S. et al.: Nucleotide sequence analysis of the lemur β-globin gene family: Evidence for major rate fluctuations in globin polypeptide evolution. Mol. Biol. Evol. 3: 465–484 (1986)
61. Herskovits T. T. and Hamilton M. G.: The hemoglobin of the aquatic snail, Planorbella duryi (Wetherby). Comp. Biochem. Physiol. Pt. B 95: 321–326 (1990)
62. Herskovits T. T. et al: Light-scattering and scanning transmission electron microscopic investigation of the hemocyanin of the bivalve, Yoldia limatula (Say). Comp. Biochem. Physiol. Pt. B 96: 497–503 (1990)
63. Herskovits T. T. and Hamilton M. G.: Higher order assemblies of molluscan hemocyanins (Minireview). Comp. Biochem. Physiol. Pt. B 99: 19–34 (1991)
64. Hirsch R. E. and Noble R. W.: Intrinsic fluorescence of carp hemoglobin: a study of the R-T-transition. Biochim. biophys. Acta 914: 213–219 (1987)
65. Hombrados I. et al.: Primary structure of the minor hemoglobins from the sea lamprey (Petromyzon marinus, Cyclostomata). Biol. Chem. Hoppe-Seyler 368: 145–154 (1987)
66. Honig G. R. and Adams J. G. (eds.): Human hemoglobin genetics. Springer, Wien 1986
67. Honzatko R. B. and Hendrickson W. A.: Molecular models for the putative dimer of sea lamprey hemoglobin. Proc. Nat. Acad. Sci. USA 83: 8487–91 (1986)
68. Hsu L. et al.: Structure and expression of the human theta-globin gene. Nature 331: 94–96 (1988)
69. Huber F. and Braunitzer G.: The primary structure of the hemoglobin of the electric eel (Electrophorus electricus). Biol. Chem. Hoppe-Seyler 370: 245–250 (1989)
70. Huisman T. H. J.: A comprehensive list of all hemoglobin variants and their references. Hemoglobin 13: 221–323 (1989)
71. van Iersel A. A. J. and Blaauboer B. J.: NADH-ferrihemoglobin reductase in avian erythrocytes. Comp. Biochem. Physiol. Pt. B 81: 1027–31 (1985)

72. di Iorio E. E. et al.: Kinetics of oxygen and carbon monoxide binding to liver fluke (Dicrocoelium dendriticum) hemoglobin. J. biol. Chem. 260: 2160–64 (1985)
73. Iwaasa H., Takagi T. and Shikama K.: Protozoan hemoglobin from Tetrahymena pyriformis. Isolation, characterization, and amino acid sequence. J. Biol. Chem. 265: 8603–09 (1990)
74. Jannasch H. W.: Deepsea life on the basis of chemical synthesis (In German). Naturwissenschaften 72: 285–290 (1985)
75. Jensen F. B. et al.: A three-state MWC analysis of oxygenation in tench (Tinca tinca) hemoglobin. J. comp. Physiol. B 160: 407–411 (1990)
76. Jhiang S. M. and Riggs A. F.: The structure of the gene encoding chain c of the hemoglobin of the earthworm, Lumbricus terrestris. J. Biol. Chem. 264: 19003–08 (1989)
77. Johnson B. A.: Structure and function of the hemocyanin from a semiterrestrial crab, Ocypode quadrata. J. comp. Physiol. B 157: 501–509 (1987)
78. Jones G. et al.: Molecular cloning, regulation, and complete sequence of a hemocyanin-related, juvenile hormone-suppressible protein from insect hemolymph. J. Biol. Chem. 265: 8596–8602 (1990)
79. Kapp O. H. et al.: Quaternary structure of the giant extracellular hemoglobin of the leech Macrobdella decora. J. mol. Biol. 213: 141–158 (1990)
80. Karlson S. and Nienhuis A. W.: Developmental regulation of human globin genes. Annual Rev. Biochem. 54: 1071–1108 (1985)
81. Kleinschmidt T. and Sgouros J. G.: Hemoglobin sequences. Biol. Chem. Hoppe-Seyler 368: 579–615 (1987)
82. Kleinschmidt T., Keyl H. G. and Braunitzer G.: Comparison of insect hemoglobins (erythrocruorins) from Chironomus thummi thummi and Chironomus thummi piger (Diptera): The primary stucture of the monomeric hemoglobin CPTIII. Biol. Chem. Hoppe-Seyler 370: 839–845 (1989)
83. Klippenstein G. L.: Structural aspects of hemerythrin and myohemerythrin. Amer. Zool. 20: 39–51 (1980)
84. Knöchel W. et al.: Globin evolution in the genus Xenopus: Comparative analysis of cDNAs coding for adult globin polypeptides of Xenopus borealis and Xenopus tropicalis. J. mol. Evol. 23: 211–223 (1986)
85. Kobayashi M., Nezu T. and Tanaka Y.: Hypoxic induction of hemoglobin synthesis in Daphnia magna. Comp. Biochem. Physiol. Pt. A 97: 513–517 (1990)
86. Kolatkar P. R. et al.: Novel subunit structure observed for noncooperative hemoglobin from Urechis caupo. J. biol. Chem. 263: 3462–65 (1988)
87. Komiyama M. H. et al.: Was the loss of the D-helix in α-globin a functional neutral mutation? Nature 352: 349–351 (1991)
88. Koop B. F. and Goodman M.: Evolutionary and developmental aspects of two hemoglobin β-chain genes (epsilon-M and beta-M) of opossum. Proc. Nat. Acad. Sci. USA 85: 3893–97 (1988)
89. Koop B. F. et al.: Tarsius delta- and beta-globin genes: conversions, evolution and systematic implications. J. Biol. Chem. 264: 68–79 (1989)
90. Kortt A. A., Trinick M. J. and Appleby C. A.: Amino acid sequences of hemoglobins I and II from root nodules of the non-leguminous Parasponia rigida-Rhizobium symbiosis, and a correction of the sequence of hemoglobin I from Parasponia andersonii. Eur. J. Biochem. 175: 141–149 (1988)
91. van Kuik J. A. et al.: Primary structure of the neutral carbohydrate chains of hemocyanin from Panulirus interruptus. Eur. J. Biochem. 159: 297–301 (1986)
92. van Kuik J. A. et al.: Primary structure of the acidic carbohydrate chain of hemocyanin from Panulirus interruptus. FEBS Letters 221: 150–154 (1987)
93. van Kuik J. A. et al.: Primary structure determination of seven novel N-linked carbohydrate chains derived from hemocyanin of Lymnaea stagnalis – 3-O-Methyl-D-galactose and N-acetyl-D-galactosamin as constituents of xylose-containing N-linked oligosaccharides in an animal glycoprotein. Eur. J. Biochem. 169: 399–411 (1987)
94. Lallier F. and Truchot J. P.: Modulation of hemocyanin oxygen affinity by L-lactate and urate in the prawn Penaeus japonicus. J. exp. Biol. 147: 133–146 (1989)
95. Landsmann J. et al.: Common evolutionary origin of legume and non-legume plant haemoglobins. Nature 324: 166–168 (1986)
96. Lang W. H.: cDNA cloning of the Octopus dofleini hemocyanin: Sequence of the carboxyl-terminal domain. Biochemistry 27: 7276–82 (1988)
97. Lecomte J. T. J. et al.: Structural and electronic properties of the liver fluke heme cavity by nuclear magnetic resonance and optical spectroscopy. Evidence for a distal tyrosine residue in a normally functioning hemoglobin. J. mol. Biol. 209: 235–247 (1989)
98. Lee A. W. and Karplus M.: Structure-specific model of hemoglobin cooperativity. Proc. Nat. Acad. Sci. USA 80: 7055–59 (1983)
99. Leidescher T. and Decker H.: Conformational changes of tarantula (Eurypelma californicum) hemocyanin detected with a fluorescent probe, 7-chloro-4-nitrobenzo-2-oxa-1,3-diazole. Eur. J. Biochem. 187: 617–625 (1990)
100. Lerch K. et al.: Different origins of metal binding sites in binuclear copper proteins, tyrosinase and hemocyanin. J. inorg. Biochem. 26: 213–217 (1986)
101. Lesk A. M. and Chothia C.: How different amino acid sequences determine similar protein structures: The structure and evolutionary dynamics of the globins. J. mol. Biol. 136: 225–270 (1980)
102. Leung S. O., Proudfoot N. J. and Whitelaw E.: The gene for theta-globin is transcribed in human fetal erythroid tissues. Nature 329: 551–554 (1987)
103. Levy M. J. et al.: Isolation and characterization of methemoglobin reductase from yellowfin tuna (Thunnus albacares). Comp. Biochem. Physiol. Pt. B 81: 809–814 (1985)
104. Leyko W. and Osmulski P. A.: Seasonal variability of hemoglobin content and component composition of Chironomus thummi larvae. Comp. Biochem. Physiol. Pt. B 89: 613–616 (1985)
105. Li Q. L. et al.: Beta-globin locus activation regions. Conservation of organization, structure and function. Proc. Nat. Acad. Sci. USA 87: 8207 – 11 (1990)
106. Liebhaber S. A., Cash F. E. and Ballas S. K.: Human α-globin gene expression. The dominant role of the α2-locus in mRNA and protein synthesis. J. biol. Chem. 261: 15327–33 (1986)
107. Lima A. A. B. et al.: Allosteric effect of protons and adenosine triphosphate on hemoglobins from aquatic amphibia. J. comp. Physiol. B 155: 353–355 (1985)

108. Linzen B. (ed.): Invertebrate oxygen carriers. Springer, Berlin 1986
109. Linzen B.: Blue blood. Structure and evolution of hemocyanins (In German). Naturwissenschaften 76: 206–211 (1989).
110. Livingston D. J., Watts D. A. and Brown W. D.: Myoglobin interspecies structural differences: effects on autoxidation and oxygenation. Arch. Biochem. Biophys. 249: 106–115 (1986)
111. Makino N.: Subunits of Panulirus japonicus hemocyanin. 2. Cooperativity of the homogonous hexamers. Eur. J. Biochem. 173: 431–435 (1988)
112. Mangum C. P. (ed.): Blood and oxygen carriers. Springer, New York 1992
113. Manwell C. and Baker C. M. A.: Magelona haemerythrin: tissues specificity, molecular weights and oxygen equilibiria. Comp. Biochem. Physiol. Pt. B 89: 453–463 (1988)
114. Margot J. B., Demers G. W. and Hardison R. C.: Complete nucleotide sequence of the rabbit β-like globin gene cluster. J. mol. Biol. 205: 15–40 (1989)
115. Markl J.: Evolution and function of structurally diverse subunits in the respiratory protein hemocyanin from arthropods. Biol. Bull. 171: 90–115 (1986)
116. Markl J. et al.: Quaternary and subunit structure of Calliphora arylphorin as deduced from electron microscopy, electrophoresis, and sequence similarities with arthropod hemocyanin. J. comp. Physiol. B 162: 665–680 (1992)
117. Martin K. D. and Parkhurst L. J.: Kinetics and thermodynamics of oxygen and carbon monoxide binding to the T-state hemoglobin of Urechis caupo. Biochemistry 29: 5718–26 (1990)
118. Matsuura M. S. A., Fushitani K and Riggs A. F.: The amino acid sequences of the α and β chains of hemoglobin from the snake, Liophis miliaris. J. Biol. Chem. 264: 5515–21 (1989)
119. Mayr G. W. and Dietrich W.: The only inositol tetrakisphosphate detectable in avian erythrocytes is the isomer lacking phosphate at position 3: a NMR study. FEBS Letters 213: 278–282 (1987)
120. Miller K. I. and Mangum C. P.: An investigation of the nature of Bohr, Root, and Haldane effects in Octopus dofleini hemocyanin. J. comp. Physiol. B 158: 547–552 (1988)
121. Miller K. I., Schabtach E. and van Holde K. E.: Arrangement of subunits and domains within the Octopus dofleini hemocyanin. Proc. Nat. Acad. Sci. USA 87: 1496–1500 (1990)
122. Mintorovitch J., van Pelt D. and Satterlee J. D.: Kinetic study of the slow cyanide binding to Glycera dibranchiata monomer hemoglobin. Biochemistry 28: 6099–6104 (1989)
123. Miyashita N. et al.: Allelic constitution of the hemoglobin beta chain in wild populations of the house mouse, Mus musculus. Biochem. Genetics 23: 975–986 (1985)
124. Moon A. M. and Ley T. J.: Conservation of the primary structure, organization, and function of the human and mouse β-globin locus-activating regions. Proc. Nat. Acad. Sci. USA 87: 7693–97 (1990)
125. Myers C. R. et al.: Haemoglobin-producing tissue of larvae and pupae of Chironomus thummi (Diptera). J. Insect Physiol. 32: 845–851 (1986)
126. Nakashima H. et al.: Structure of hemocyanin II from the horseshoe crab, Limulus polyphemus. Sequences of two overlapping peptides, ordering the CNBr fragments, and the complete amino acid sequence. J. biol. Chem. 261: 10526–33 (1986)
127. Neuteboom B. et al.: Partial amino acid sequence of a hemocyanin subunit from Palinurus vulgaris. Comp. Biochem. Physiol. Pt. B 94: 593–597 (1989)
128. Osmulski P. A. and Leyko W.: Structure, function and physiological role of Chironomus haemoglobin (Review). Comp. Biochem. Physiol. Pt. B 85: 701–722 (1986)
129. Padgett R. W. et al.: The molecular organization of the beta-globin complex of the deer mouse, Peromyscus maniculatus. Mol. Biol. Evol. 4: 30–45 (1987)
130. Peeters K. et al.: The globin composition of Daphnia pulex hemoglobin and the comparison of the amino acid composition of invertebrate hemoglobins. Comp. Biochem. Physiol. Pt. B 97: 369–381 (1990)
131. Perutz M. F.: Species adaptation in a protein molecule. Mol. Biol. Evol. 1: 1–28 (1983)
132. Perutz M. F.: Species adaptation in a protein molecule. Adv. Protein Res. 36: 213–244 (1984)
133. Petruzzelli R. et al.: Amino acid sequence of α-chain of hemoglobin IV from trout (Salmo irideus). Biochim. biophys. Acta 995: 255–258 (1989)
134. Petruzzelli R. et al.: Scapharca hemoglobins, type cases of a novel mode of chain assembly and heme-heme communication. Amino acid sequence and subunit interactions of the tetrameric. FEBS Letters 259: 133–136)1889)
135. Powers D. A.: Molecular ecology of teleost fish hemoglobins: strategies for adapting to changing environments. Amer. Zool. 20: 139–162 (1980)
136. Reischl E. et al.: Bohr effect, electroin spin resonance spectroscopy and subunit dissociation of the hemoglobin components from the turtle Phrynops hilarii. Comp. Biochem. Physiol. Pt. B 78: 251–157 (1984)
137. Rendell M. et al.: An interspecies comparison of normal levels of glycosylated hemoglobin and glycosylated albumin. Comp. Biochem. Physiol. Pt. B 81: 819–822 (1985)
138. Richardson D. E. et al.: Allosteric interactions in sipunculid and brachiopod hemerythrins. Biochemistry 26: 1003–13 (1987)
139. Riggs A. F.: The Bohr effect. Annual Rev. Physiol. 50: 181–204 (1988)
140. Robinson I. B. and Ingram V. M.: Gene evolution in the chicken β-globin cluster. Cell 28: 515–521 (1982)
141. Royer W. E. jr., Hendrickson W. A. and Chiancone E.: Structural transitions upon ligand binding in a cooperative dimeric hemoglobin. Science 249: 518–521 (1990)
142. Rozynek P., Hankeln T. and Schmidt E. R.: Structure of a hemoglobin gene cluster and nucleotide sequence of three hemoglobin genes from the midge Chironomus thummi piger (Diptera, Insecta). Biol. Chem. Hoppe-Seyler 370: 533–542 (1989)
143. Schartau W. et al.: Hemocyanins in spiders. XIII. Complete amino acid sequence of subunit a Eurypelma californicum. Biol. Chem. Hoppe-Seyler 371: 557–565 (1990)
144. Schimenti J. C. and Dumcan C. H.: Structure and organization of the bovine β-globin genes. Mol. Biol. Evol. 2: 514–525 (1985)
145. Scott E. M. and Harrington J. P.: Methemoglobin reductase activity in fish erythrocytes. Comp. Biochem. Physiol. Pt. B 82: 511–513 (1985)

146. Shaw J. P., Marks J. and Shen C. K. J.: Evidence that the recently discovered theta-1-globin gene is functional in higher primates. Nature 326: 717–720 (1987)
147. Shehee W. R. et al.: Nucleotide sequence of the BALB/c mouse β-globin complex. J. mol. Biol. 205: 41–62 (1989)
148. Shishikura F. et al.: Amino acid sequence of the monomer subunit of the extracellular hemoglobin of Lumbricus terrestris. J. biol. Chem. 262: 3123–31 (1987)
149. Simpson C. F., Taylor W. J. and Jacobson E. R.: Sickling hemoglobin polymerization in iguana erythrocytes. Comp. Biochem. Physiol. Pt. A 73: 703–708 (1982)
150. Sizaret P. Y. et al.: A refined quaternary structure of Androctonus australis hemocyanin. Eur. J. Biochem. 127: 501–506 (1982)
151. Smit J. D. G. et al.: Acid Bohr effect of a monomeric haemoglobin from Dicrocoelium dendriticum – Mechanism of the allosteric conformation transition. Eur. J. Biochem. 155: 231–237 (1986)
152. Stalder J. et al.: Primary structure and evolutionary relationship between the adult α-globin genes and their 5'-flanking regions of Xenopus laevis and Xenopus tropicalis. J. mol. Evol. 28: 64–71 (1989)
153. Standley P. R. et al.: The calcium, copper and zinc content of some annelid extracellular haemoglobins. Biochem. J. 249: 915–916 (1988)
154. Stern M. S. et al.: Amino acid sequence of the monomer subunit of the giant extracellular hemoglobin of the aquatic oligochaete, Tubifex tubifex. Eur. J. Biochem. 194: 67–73 (1990)
155. Stoecker W. et al.: The quatenary structure of four crustacean two-hexameric hemocyanins: immunocorrelation, stoichiometry, reassembly and topology of individual subunits. J. comp. Physiol. B 158: 271–289 (1988)
156. Suzuki T.: Amino acid sequence of myoglobin from the mollusc Dolabella auricularia. J. biol. Chem. 261: 3692–99 (1986)
157. Suzuki T.: Amino acid sequence of a major globin from the sea cucumber Paracaudina chilensis. Biochim. biophys. Acta 998: 292–296 (1989)
158. Suzuki T. et al.: Hemoglobins from the two closely related clams Barbatia lima and Barbatia virescens. Comparison of their subunit structures and N-terminal sequence of the unusual two-domain chain. Zool. Sci. 6: 269–281 (1989)
159. Suzuki T., Takagi T. and Ohta S.: Amino acid sequence of the dimeric hemoglobin (Hb I) from the deep-sea cold-seep clam Calyptogena soyoae and the phylogenetic relationship with other molluscan hemoglobins. Biochim. biophys. Acta 999: 254–259 (1989)
160. Suzuki T., Takagi T. and Gotoh T.: Primary structure of the two linker chains of the extracellular hemoglobin from the polychaete Tylorrhynchus heterochaetus. J. Biol. Chem. 265: 12168–77 (1990)
161. Suzuki T., Takagi T. and Ohta S.: Primary structure of a constituent polypeptide chain (AIII) of the giant hemoglobin from the deep-sea tube worm Lamellibrachia. A possible hydrogen sulfide-binding site. Biochemistry 266: 221–225 (1990)
162. Tam L. T., Gray G. P. and Riggs A. F.: The hemoglobins of the bullfrog Rana catesbeiana. The structure of the β-chain of component C and the role of the α-chain in the formation of intermolecular disulfide bonds. J. biol. Chem. 261: 8290–94 (1986)
163. Terwilliger R. C. and Terwilliger N. B.: Molluscan hemoglobins (Review). Comp. Biochem. Physiol. Pt. B 81: 255–261 (1985)
164. Terwilliger N. B. et al.: Bivalve hemocyanins – a comparison with other molluscan hemocyanins. Comp. Biochem. Physiol. Pt. B 89: 189–195 (1988)
165. Toulmond A., Jouin C. and de Frescheville J.: Hemocyanin of the protobranch bivalve mollusc Nucula hanleyi Winckworth. Comp. Biochem. Physiol. Pt. B 88: 71–74 (1987)
166. Trewitt P. M., Boyer D. R. and Bergstrom G.: Characterization of maternal haemoglobins in the eggs and embryos of Chironomus thummi. J. Insect Physiol. 32: 963–969 (1986)
167. Trotman C. N. A. et al.: The polymeric hemoglobin molecule of Artemia. Interpretation of translated cDNA sequence of nine domains. J. Biol. Chem. 266: 13789–95 (1991)
168. Utecht R. E. and Kurtz jr. D. M.: Cytochrome b5 and NADH-cytochrome-b5 reductase from sipunculan erythrocytes; a methemerythrin reduction system from Phascolopsis gouldii. Biochim. biophys. Acta 953: 164–178 (1988)
169. Val A. L. et al.: Biological aspects of Amazonian fishes – I. Red blood cell phosphates of schooling fishes (genus Semaprochilodus: Prochilodontidae). Comp. Biochem. Physiol. Pt. B 78: 215–217 (1984)
170. Vandergon T. L. and Colacino J. M.: Characterization of hemoglobin from Phoronis architecta (Phoronida). Comp. Biochem. Physiol. Pt. B 94: 31–39 (1989)
171. Viana de Freitas T., Alfonso A. M. M. and Neves A. G. A.: Purification and characterization of the glycopeptide II from the hemoglobin of Biomphalaria glabrata. Comp. Biochem. Physiol. Pt. B 81: 743–747 (1985)
172. Vinson C. R. and Bonaventura J.: Structure and oxygen equilibrium of the three coelomic cell hemoglobins of the echiuran worm Thalassema mellita (Conn). Comp. Biochem. Physiol. Pt. B 87: 361–366 (1987)
173. Vincent K. A. and Wilson A. C.: Evolution and transcription of Old World monkey globin genes. J. mol. Biol. 207: 465–479 (1989)
174. Vinogradov S. N. and Kapp O. H. (eds.): Structure and function of invertebrate oxygen carriers. Springer, New York 1991
175. Vinogradov S. N., Sharma P. K. and Walz D. A.: Iron and heme content of the extracellular hemoglobins and chlorocruorins of annelids (Review). Comp. Biochem. Physiol. Pt. B 98: 187–194 (1991)
176. Vinogradov S. N. et al.: A dodecamer of globin chains is the principal functioning subunit of the extracellular hemoglobin of Lumbricus terrestris. J. Biol. Chem. 266: 13091–96 (1991)
177. Voit R. and Feldmaierfuchs G.: Arthropod hemocyanins: Molecular cloning and sequencing of cDNAs encoding the tarantula hemocyanin subunit-A and subunit-E. J. Biol. Chem. 265: 19447–52 (1990)
178. Volbeda A. and Hol W. G. J.: Crystal structure of hexameric haemocyanin from Panulirus interruptus refined at 3.2 Å resolution. J. mol. Biol. 209: 249–279 (1989)

179. Wache S., Terwilliger N. B. and Terwilliger R. C.: Hemocyanin structure changes during early development of the crab Cancer productus. J. exp. Zool. 247: 23–32 (1988)
180. Wainwright B. and Hope R.: Cloning and chromosomal location of the α- and β-globin genes from a marsupial. Proc. Nat. Acad. Sci. USA 82: 8105–08 (1985)
181. Wakabayashi S., Matsubara H. and Webster D. A.: Primary sequence of a dimeric bacterial haemoglobin from Vitreoscilla. Nature 322: 481–483 (1986)
182. Weber R. E., Braunitzer G. and Kleinschmidt T.: Functional multiplicity and structural correlations in the hemoglobin system of larvae of Cironomus thummi thummi (Insecta, Diptera): Hb components CTT I, CTT IIß, CTT III, CTT IV, CTT VI, CTT VIIB, CTT IX and CTT X. Comp. Biochem. Physiol. Pt. B 80: 747–753 (1985)
183. Weber R. E. and Jensen F. B.: Functional adaptations in hemoglobins from ectothermic vertebrates. Annual Rev. Physiol. 50: 161–179 (1988)
184. Wichertjes T. et al.: The quaternary structure of Sepia officinalis haemocyanin. Biochim. biophys. Acta 872: 183–194 (1986)
185. Willard C. et al.: Comparison of human and chimpanzee zeta-globin genes. J. mol. Evol. 22: 309–315 (1985)
186. Wills C.: Genetic variability. Clarendon Press, Oxford 1981
187. Wilson jr. R. R. and Knowles F. C.: Temperature adaptation of fish hemoglobins reflected in rates of autoxidation. Arch. Biochem. Biophys. 255: 210–213 (1987)
188. Wood E. J. et al.: Relative molecular mass of the polypeptide chain of βc-haemocyanin of Helix pomatia and carbohydrate composition of the functional units. Comp. Biochem. Physiol. Pt. B 82: 179–186 (1985)
189. Zafar R. S. et al.: The cDNA sequences encoding two components of the polymeric fraction of the intracellular hemoglobin of Glycera dibranchiata. J. Biol. Chem. 265: 21843–51 (1990)
190. Zhang K. et al.: The active site of hemerythrin as determined by X-ray absorption fine structure. Biochemistry 27: 7470–79 (1988)
191. Zimmer J. R. et al.: Kinetic study of the oxygenation process of hemerythrins from Lingula unguis and Siphonosoma crumanense. Biochim. biophys. Acta 874: 174–180 (1986)

8 Peptide Hormones

Next to the steroids, the most important signalling molecules in animals are peptides. Their high information content and unambiguous genetic coding make them particularly suitable for signal transmission in the body. In addition to their classical endocrine functions, peptides also play a role as neurotransmitters, neuromodulators and growth substances. This variety of functions is reflected in the term **"regulatory peptides"**. In the vertebrates, we now know more about regulatory peptides than all other types of signalling molecule put together (steroids, amines and amino acids), and they continue to be discovered. The concentrations of regulatory peptides, even in organs of synthesis or storage, are mostly of the order of nanograms per tissue; consequently, their isolation often requires extraction from thousands of individuals, especially when dealing with small invertebrates. The paucity of material available hinders direct sequencing; this is made still more problematic by the blocked N- and C-terminals of many peptide hormones. As a result, the investigation of regulatory peptides relies more and more upon the determination of cDNA and gene sequences. The ever-increasing sophistication of analytical methods has led to a drastic increase in the number of known peptides in the last two decades, but in many cases neither the action spectrum nor biological role has yet been defined. The following sections deal only with those peptides for which some comparative biochemical data are available [73, 81].

The **sizes of the regulatory peptides** range from tripeptides to polypeptides of over 100 amino acids, which as a rule also bear carbohydrate chains. The peptides usually arise from significantly larger precursors **(prepro-hormones)**. The primary translation products begin, as in the majority of secreted proteins, with a pre- or signalling sequence which is removed by a signal protease during translation. The peptide precursor is modified in the endoplasmic reticulum and Golgi apparatus by N- and O-glycosylation, by the phosphorylation of certain serines or threonines, and by the sulphation of certain tyrosine residues. The precursor is endoproteolytically cleaved in the secretory granules, or sometimes this has already occurred in the Golgi apparatus. The responsible proteases are highly specific, the cleavage sites in the precursor sequence being marked by pairs of basic amino acids (arginine, lysine) or by single arginine residues. In some cases, a whole series of peptides of varying length, but with the same active sequence, is produced by irregular cleavage of the pro-protein. One of the proteases involved in this process is carboxypeptidase E, which belongs to the same family as the carboxypeptidases A and B,

although it shares only 20 and 17 %, respectively, of their sequences [248]. Only one molecule of the bioactive peptide may arise from the precursor (e.g. prepro-insulin), but in most cases several are produced; these may be identical (e.g. prepro-caerulein), just similar [e.g. prepro-encephalin, prepro-endorphin, prepro-glucagon, prepro-vasoactive intestinal polypeptide (VIP)], or completely different (e.g. vasopressin/neurophysin, oxytocin/neurophysin, pro-opiomelanocortin). Precursors giving rise to several bioactive peptides are known as **polyproteins**. The first typical polyprotein to be discovered (in 1977) was pro-opiomelanocortin, and at present almost 20 mammalian polyproteins are known. Several peptide hormones of the invertebrates have also been shown to have larger precursors, e.g. the egg-laying hormone of various snails [75].

The peptides arising by cleavage of a precursor may be further modified at the N-terminus by formation of a pyroglutamyl residue or by N-acetylation, and at the C-terminus by α-amidation. The amide group orginates in a C-terminal glycine from which a carbohydrate chain is removed as glyoxylate. The reaction requires molecular oxygen and ascorbic acid, which is oxidized to dehydroascorbate. The amidating monoxygenases are specific for the amidated amino acids [24, 248]. The **genes of the prepro-hormones** often show subdivision of the exons corresponding to the organization of the polypeptide chain into individual domains or active peptides, e.g. in prepro-glucagon, vasopressin/neurophysin and calcitonin/CGRP (calcitonin gene-related peptide). In contrast the genes of the pro-encephalins and pro-opiomelanocortins show no correspondence between exon and domain boundaries. Alternative splicing of the primary transcription product of one and the same gene can lead to different mRNAs and peptides; thus, expression of the calcitonin/CGRP gene results in calcitonin in the C cells of the thyroid and CGRP in the brain [59].

By the use of highly sensitive immunological methods or the analysis of corresponding mRNAs, it has recently been shown that typical peptide hormones of the alimentary system are also present in the central nervous system and vice versa: hence the term **"neuro-intestinal peptide"**. Many of these regulatory peptides are synthesized and secreted by nerve cells. Such peptidogenic neurons are found in large numbers in, for example, the hypothalamus and the neurohypophysis, but they also occur in the gut wall. However, bioactive peptides do not only arise in neurons; for example, the adenohypophysis corresponds to an epithelial gland, and endocrine gland cells are present in the pancreas associated with the islets of Langerhans. Growth factors are produced by many different types of cell. Some peptides, e.g. cholecystokinin, are produced both in gland cells of the intestine and in neurons, e.g. in the limbic system of the brain.

Some peptides act as **neurotransmitters** during the excitation of synapses. This has been demonstrated for substance P, β-endorphin, encephalin and somatostatin, and probably also applies to VIP, pancreas polypeptide (PP), angiotensin, adrenocorticotropic hormone (ACTH), and cholecystokinin (CKK). Approximately one-half of the synapses in the human brain function with classical neurotransmitters, such as acetylcholine, biogenic amines (dopamine, noradrenalin, serotonin) and amino acids (γ-aminobutyric acid, glycine, glutamic acid and aspartic acid); the rest are possibly all peptidergic. **Neuromodulator functions** of peptides refer to their activities over longer distances, longer times or as inhibitors, in contrast to a normal neurotransmitter. Apart from nerve cells, other cell types can also effect neighbouring cells via active substances (paracrine functions). This applies to the biogenic amines, like histamine and serotonin, and also to the peptide somatostatin and the growth factors. If a peptide is to function as a hormone, neurotransmitter, neuromodulator or growth factor on a target cell, the latter must carry on its surface the appropriate acceptor molecules. In biochemical language these integral membrane proteins are referred to as "receptors", a term the biologist reserves for whole cells or even whole organs. Interestingly, there are peptides that function as sexual lures, i.e. they serve as **pheromones** for signal transfer between individuals. An example is aphrodisin in the vaginal secretions of the hamster *Mesocricetus auratus;* this is a glycoprotein of 151 amino acids from the super-family of the α_{2u}-globulins and stimulates copulatory behaviour in the males via the accessory olfactory system [77].

Also, in the case of the peptide hormones, attempts have been made to assign similar sequences to families and derive their common ancestors. Difficulties are encountered with short sequences in that convergent evolution cannot always be excluded. It is very relevant to concepts of **peptide hormone evolution** that many hormones initially discovered in vertebrates are also found in invertebrates. For example, peptides corresponding in their immunological and biolo-

gical activities to vertebrate insulin are also detected in tunicates, insects, crustaceans, gastropods, cephalopods, ciliates, and even fungi and prokaryotes. Representatives of these invertebrate groups have also been shown to possess somatostatin, encephalin, gastrin, CKK, secretin, substance P, PP, VIP, calcitonin and other vertebrate peptides. It must be stated, however, that in invertebrates these peptides have been detected almost exclusively using immunological methods, and their identity with vertebrate peptides is not completely certain. Conversely, the brain and gut of amphibians, birds and mammals apparently contain the neuropeptide FMRF-amide of molluscs, and mammals also possess the "head activator" of the freshwater polyp *Hydra*. As will be discussed later for individual peptides, only in a few cases is it known whether the vertebrate peptide has a hormone function in invertebrates and how far this corresponds to the activity in vertebrates. Because each hormone needs a corresponding specific receptor, the evolution of which does not necessarily occur simultaneously with that of the peptide, the simple presence of a particular hormone does not allow firm conclusions to be drawn about its function. It is also true that some peptide hormones are phylogenetically very old and have possibly changed in function during evolution (127, 166]. What follows is a discussion of, in particular, the structure and formation of regulatory peptides from a comparative biochemistry point of view; comparative observations on their activities can be found in the many textbooks on hormone physiology.

8.1 Hormones of the Neurohypophysis

The neurohypophysis releases **nonapeptide hormones** into the blood. The high molecular weight precursors of these peptides are synthesized in certain neurons of the hypothalamus and are proteolytically cleaved; the nonapeptides are bound to specific carrier proteins (neurophysins) and transported as neurosecretory granules via nerve processes to the neurohypophysis, where they are released. In the blood plasma they are bound to a β-globulin fraction. Ten variants of these nonapeptides have been found in investigations of more than 40 vertebrate species (Fig. 8.1). They are characteristically distributed amongst the different vertebrate classes, and the data allow the construction of a **phylogenetic tree** which provides a classical example of molecular evolution.

Arg-Vasopressin	C	Y	F	Q	N	C	P	R	G-amide
Lys-Vasopressin	.	Y	F	Q	.	.	.	K	.
Phenypressin	.	F	F	Q	.	.	.	R	.
Vasotocin	.	Y	I	Q	.	.	.	R	.
Oxytocin	.	Y	I	Q	.	.	.	L	.
Mesotocin	.	Y	I	Q	.	.	.	I	.
Isotocin	.	Y	I	S	.	.	.	I	.
Glumitocin	.	Y	I	S	.	.	.	Q	.
Valitocin	.	Y	I	Q	.	.	.	V	.
Aspargtocin	.	Y	I	N	.	.	.	L	.

Fig. 8.1. The nonapeptides of the neurohypophysis

Two classes of these peptide hormones can be differentiated: the basic variety with arginine or lysine at position 8, which act antidiuretically or are vasoconstricting in mammals, and the neutral variety, which induce uterus contractions and milk secretion. Both structural types are found in all vertebrates from the cartilaginous fish onwards, but their specific functions are not clear in all cases.

The prototype of the **basic nonapeptides** is arginine-vasotocin, which is generally distributed from the agnathans to the birds and is also present in mammalian embryos, although it is often replaced by arginine-vasopressin [114, 188]. In some mammals, lysine-vasopressin is found in addition to arginine-vasopressin or may occur alone. Amongst the higher mammals, the Placentalia, lysine-vasopressin is found only in swine (Suidae); it is the only vasopressin in the domestic pig, and occurs together with arginine-vasopressin in the warthog *Phacochoerus aethiopicus*. Contrary to earlier assumptions, no lysine-vasopressin is found in the related families of the hippopotamus (Hippopotamidae) and the peccary (Tayassuidae) [1, 183]. The egg-laying monotreme mammals, the duck-billed platypus *Ornithorhyncus anatinus* and the echidna *Tachyglossus aculeatus*, possess arginine-vasopressin [35]. In contrast, the marsupials (Marsupialia) are heterogeneous: the Australian Phalangeridae have arginine-vasopressin, the kangaroos (Macropodidae) lysine-vasopressin and phenypressin, and the American opossums (Didelphidae) both lysine-vasopressin and arginine-vasopressin [1].

Vasotocin- and vasopressin-like peptides have been detected using immunological, chromatographic and pharmacological methods in the freshwater polyp *Hydra* and in the ganglia of various insects (e.g. *Locusta migratoria)* and molluscs *(Aplysia californica, Lymnaea stagnalis, Octopus vulgaris)* [140, 147]. Two such neuropeptides are found in the migratory locust: F1, with the

sequence CLITNCPRG-amide, and F2, which is an antiparallel dimer consisting of two such chains. The sequence similarity of the last five positions is responsible for the immunological cross-reactivity with arginine-vasopressin [174]. The venom of marine gastropods of the genus *Conus* contains vasopressin-like peptides referred to as conopressins. The Lys-conopressin-G from *C. geographus* has the sequence CFIRNCPKG-amide, and the Arg-conopressin-S from *C. striatus* has the sequence CIIRNCPRG-amide [46]. Homology of these invertebrate peptides with the hypophysis hormones is not certain; their short sequences make it difficult to exclude convergent evolution.

With the exception of the agnathans, all vertebrates possess **neutral nonapeptide hormones** in addition to the basic type. Of the cartilaginous fish, the sharks have aspargtocin or valitocin, the rays glumitocin, and the chimaeras oxytocin. Isotocin (ichthyotocin) appears to be characteristic of the bony fish classes Teleostei, Holostei, Chondrostei and Polypteri; however, as only a very few members of this extensive group have as yet been examined, some surprises may be in store. The lungfish, amphibians, reptiles and birds all possess mesotocin. The Placentalia have exclusively oxytocin, as have the South American marsupials (Didelphidae). However, the three groups of Australian marsupials (Dasyuridae, Phalangeridae, Macropdidae) produce mesotocin, and the North American opossum *Didelphis virginiana* has both mesotocin and oxytocin. The occurrence of mesotocin in the marsupials could be taken as an ancestral sign were it not for the fact that the even more primitive monotreme species both possess only oxytocin [35]. The higher molecular weight **precursor of the nonapeptides** also includes the sequences for the neurophysins (NPs). Thus, individual neurons in mammals contain either vasopressin or oxytocin together with the corresponding NP, although in vitro both nonapeptides can bind either of the NPs. The approximately 17-kDa precursors of arginine-vasopressin/MSEL-NP and oxytocin/VLDV-NP from several mammals have been sequenced via their cDNA. The prepro-AVP/MSEL-NP is made up consecutively of the signal peptide (19 amino acids), arginine-vasopressin (9 amino acids), the linker peptide Gly-Lys-Arg with the glycine residue for amidation of the terminal-COOH and the basic amino acids of the cleavage site, the MSEL-neurophysin (NP-II) (93–95 amino acids), and a glycopeptide (copeptin) of 38–39 amino acids. Prepro-OT/VLDV-NP is similarly constructed with the signal sequence, oxytocin (OT), the linker peptide and NI-I, but it has a histidine residue in place of copeptin [35]. The genes for these precursors are organized into three exons. In the guinea-pig, an NP glycopeptide fragment results from incomplete processing of the AVP/NP-II precursor, an indication that different enzymes are responsible for the cleavage sites AVP/NP and NP/glycopeptide. This processing is complete in the rat, although the sequence of the precursor differs at only two positions from that of the guinea-pig [34]. In the vasotocin/NP and isotocin/NP precursors from the teleost *Catostomus commersoni* and in the duplicated prepro-vasotocin/NP from *Oncorhynchus keta*, the NP sequences have a C-terminal extension of about 30 amino acids, which, apart from one missing glycosylation site, is very similar to mammalian copeptin [146]. A new type of neurohypophysis hormone found in the anurans arises by variant processing of the vasotocin/NP precursor; these hormones are found only in anurans and are apparently involved in osmoregulation. Hydrin 1 (vasotocinyl-Gly-Lys-Arg) is found in *Xenopus laevis*, and hydrin 2 (vasotocinyl-Gly) occurs in various species of the genera *Rana* and *Bufo* [138, 184] .

The **neurophysins** are acidic proteins of about 10 kDa with 90–100 amino acids; complete or partial sequences are known for NPs of various mammals and of the cod *Gadus morrhua*. They are phylogenetically very conservative, positions 10–74 being practically invariant in the mammals. The classification suggested by Chauvet in 1976 is based on the occupation of positions 2,3,6 and 7, and distinguishes the MSEL classes linked to vasopressin and the VLDV classes belonging to oxytocin. However, the larger amount of material now available for comparison reveals considerable variability, particularly at positions 2 and 3. Whatever the case, the sequence differences between the two NP types are always greater than the species-specific differences within each type. Thus, the gene duplications leading to the appearance of the two nonapeptide-NP precursors occurred much earlier than the separation of the present mammalian species [1]. Both NP classes are to be found in birds. The only sequenced example, from the ostrich, agrees in only 60–65 % of its 93 amino acids with mammalian neurophysin [116].

8.2 Hormones of the Hypothalamus and Urophysis

8.2.1 Release and Release-Inhibiting Factors

Hormone release from the adenophysis is regulated by peptide factors from the hypothalamus: thyrotropin-releasing factor (TRF), gonadotropin- or lutropin-releasing factor (GnRF or LH-RF), growth-hormone-release-inhibiting factor (GH-RIF) or somatostatin (SST), corticotropin-releasing factor (CRF) and growth-hormone-releasing factor (GRF). These are also sometimes referred to as releasing hormones (RH) or release-inhibiting hormones (RIH) [123]. The **thyrotropin-releasing factor** stimulates the release of thyrotropin and prolactin in mammals. It is a tripeptide, blocked at both ends (Fig. 8.2a), which can already be detected in the brain of the marine lamprey, but does not have the same activity in fish and amphibians as it does in mammals. It is, nevertheless, found in large amounts in the poison glands of frog skin. *Xenopus laevis* possesses two prepro-TRFs in which the TRF sequence appears seven times; the TRF precursors of the rat have only five TRF copies [111, 117].

A **gonadotropin-releasing factor** is found in all vertebrates from the agnathan *Petromyzon marinus* onwards; it appears to be missing in the more primitive common hagfish *Myxine glutinose*. In addition to the mammalian hypothalamus, GnRF is found also in the retina, adrenal body, testis, ovary and placenta. The decapeptide known from mammals (Fig. 8.2b) is also found in amphibians and fish, where it occurs in addition to widely distributed GnRFs with deviant sequences: in five positions in agnathans *(Petromyzon marinus)*, and in two different positions in the bony fish *(Oncorhynchus keta)* and reptiles [173, 200]. Apart from in the mammals, a specific plasma protein binding GnRF has been detected only so far in the goldfish *Carassius auratus* [84]. Multiple GnRFs, presumably with different functions, are

```
a) Thyrotropin-RF (mammals):                              pEHP-amide
b) Gonadotropin-RF (mammals):                          pEHWSYGLRPG-amide
c) Somatostatin-14 (sheep, pig):                    AGCKNFFWKTFTSC
d) Corticotropin-RF (sheep):
                SQEPPISLDLTFHLLREVLEMTKADQLAQQAHSNRKLLDIA-amide

e) Substance P (mammals):                             RPKPQQFFGLM-amide
   Neurokinin A (cattle):                              HKTDSFVGLM-amide
   Physalaemin (Physalaemus fuscumaculatus): pEADPNKFYGLM-amide
   Scyliorhinin-I (Scyliorhinus canicula):     AKFDKFYGLM-amide
   Eledoisin (Eledone moschata):                    pEPSKDAFIGLM-amide

f) Neurotensin (cattle):                          pELYENKPRRPYIL

g) Urotensin-I (Catostomus commersoni):
                NDDPPISIDLTFHLLRNMIEMARIENEREQAGLNRKYLDEV-amide

h) Urotensin-II (Gillichthys mirabilis):    AGTAADCFWKYCV
     Somatostatin-14 for comparison:        AGCKNFFWKTFTSC

i) Gastrin-17 (pig):                          pEGPWLEEEEEAYGWMDF-amide
   CKK-33 (pig):          KAPSGRVSMIKNLQSLDPSHRISDRDYMGWMDF-amide
   Caerulein (Hyla caerulea):                       pEQDYTGWMDF-amide

k) Glucagon (pig):          HSQGTFTSDYSKYLDSRRAQDFVQWLMNT
   Secretin (pig):          HSDGLFTSEYSKMRGNAQVQKFIQNLM-amide
   Helospectin I            HSDATFTAEYSKLLAKLALQKYLESILGSSTSPRPPSS
     (Heloderma suspectum):

l) Calcitonin (pig):     CSNLSTCVLSAYWRNLNNFHRFSGMGFGPETP-amide

m) epidermal growth factor (mouse):
   NSYPGCPSSYDGYCLNGGVCMHIESLDSYTCNCVIGYGSDRCQTRDLRWWELR

n) "head activator" (Hydra, mammals):             EPPGGSKVILF
```

Fig. 8.2a–n. The peptide hormones of the vertebrates, and related peptides from invertebrates [16, 35, 39, 86, 172, 205]. For the peptides of the gastrin/CKK family **(i)**, the minimal sequences required for the biological activity are *underlined*. *pE*, pyroglutamic acid

found in teleosts and birds. Two GnRFs are known in the chicken, of which type I shows one difference to the mammalian GnRF and type II shows three differences. Whereas chicken GnRFs give reduced receptor binding and hormone release in rats, the chicken receptors react to mammalian GnRF to the same degree as to their own factor [249]. GnRF-like substances have been detected immunologically in the tunicate *Ciona intestinalis* but not in either crustaceans or platyhelminthes. The α mating factor of yeast coincides in 6 of its 13 amino acids with mammalian GnRF, binds to rat receptors and releases hormone from the rat hypophysis [249]. The human GnRF precursor consists of a signal peptide (23 amino acids), GnRF (10 amino acids), linker peptide Gly-Lys-Arg, and a C-terminal sequence of 56 amino acids [249].

Somatostatin is a peptide of 14 amino acids (Fig. 8.2 c) that inhibits the release of growth hormone, TRF and other hormones. SST-14 is highly conservative, the same sequence being present from the agnathan *Myxine* to the mammals. The only known exception was found in two primitive fish species: 5-Asn to Ser in the Pacific ratfish *Hydrolagus colliei* (Holocephali), and 12-Thr to Ser in the lamprey *Petromyzon marinus* (Agnatha) [4, 41]. SST-14-like substances have been detected immunologically in tunicates and insects and even in the ciliate *Tetrahymena* and in *E. coli*. In mammals, both SST-14 and SST-28, with 28 amino acids, are produced from the same SST; the relative proportions of the two products in the hypothalamus, gut and pancreas vary. Fish also contain larger SSTs with 22–37 amino acids; however, these do not arise by the tissue-specific processing of the same gene product but are encoded by a further gene. Detailed investigations of the angler fish *Lophius piscatorius* showed that gene I produces the conservative SST-14 and gene II the SST-28, which is in part hydroxylated on 23-Lys [144].

Corticotropin-releasing factor, which stimulates the secretion of adrenocorticotropins and β-endorphins, is a peptide of 41 amino acids (Fig. 8.2 d); the human and rat sequences coincide but differ from that of the sheep in seven positions. It is noteworthy that the CRF sequence of the teleost *Catostomus commersoni* differs from the human and rat sequences in only two positions at the C-terminus [159]. CRF agrees to a certain extent with sauvagin (40 amino acids) from the skin of the South American frog *Phyllomedusa sauvagei*, and with urotensin-I from the urophysis of various teleosts [35]. **Growth-hormone-releasing factors** have been isolated from the hypothalamus of various mammalian species and have been sequenced. They are peptides of 44 amino acids which are C-terminal amidated and only vary slightly in the C-terminal region. They stimulate the release of growth hormone, i.e. they are antagonists of the SSTs. Mammalian GRFs also function in fish and GRF-like substances have been detected immunologically in the salmon *Oncorhyncus keta* and *O. kisutch* [129]. The GRFs belong to the glucagon-secretin family. Rat GRF is unusual in that it is only 43 amino acids long, is not amidated at the C-terminus and differs from human GRF in no less than 14 positions. Human GRF causes the release of growth hormone in both birds and fish [123, 249].

8.2.2 Peptides from the Hypothalamus, Gut and Frog Skin

Peptides of rather similar structure are found in the hypothalamus and other parts of the central nervous system (CNS), in the gut wall and the pancreas, and in the skin of various frogs. This surprising discovery has been rationalized in that gut neurons and skin glands both arise from cells of the neural groove, which is replaced by the neural tube during embryo development. These substances may be grouped around neurotransmitters or neuromodulators, like substance P, and the neurotensins [14]. **Substance P** (SP) is an undecapeptide (Fig. 8.2 e) found especially in the hypothalamus, although it also occurs in other parts of the CNS and in the first part of the small intestine; it is released as a transmitter by pain-conducting nerve fibres. This is the best-investigated representative of the *tachykinins*, which are characterized by the C-terminal sequence -FXGLM-amide [153]. Alternative splicing of the products of the SP-precursor gene results in several prepro-tachykinins, each of which, apart from SP, contains other tachykinin sequences [74]. The same peptide family includes various species-specific peptides found at high concentrations in the skin of anurans, e.g. physalaemin from *Physalaemus fuscumaculatus*, or the four ranatachykinins from *Rana catesbeiana* [109]; other examples are scyliorhinins from the gut of the shark *Scyliorhinus canicula* [39], and eledoisin from the salivary glands of the cephalopod *Eledone moschata* (Fig. 8.2). Physalaemin-like peptides from rabbit stomach have also been isolated and sequenced [233]. **Bombesin**, from

the skin of the toad *Bombina bombina* (see Fig. 9.2, p. 324), which arises from a 107-amino-acid precursor, has a C-terminal structure similar to that of tachykinin. This peptide stimulates the secretion of gastrin and other gastrointestinal hormones in mammals; the gastrin-releasing peptide (GRP) isolated from the gut of man and other mammals is identical to bombesin in 9 out of 10 C-terminal amino acids [181]. A peptide similar to mammalian GRP has been isolated from the brain of the frog *Rana ridibunda* [42]. Bombesin-like peptides with species-specific sequences are known from the skin of other anurans (see Fig. 9.2). **Neurotensin** (NT) (Fig. 8.2 f) was discovered in bovine hypothalamus, but is also found in the alimentary canal. NT and the NT-like neuromedin N (six amino acids) stimulate gut contractions. Both arise from a common large precursor of 170 amino acids [31] and are already detectable in the chicken. **Xenopsin** (see Fig. 9.2), found in the skin of the clawed frog *Xenopus laevis*, is very similar to NT in its C-terminal sequence. Altogether, more than 40 different peptides have been detected in *Xenopus* skin, including the xenopsin-like levitide, a further peptide identical to the mammalian TRF, and **caerulein** (Fig. 8.2 i), which belongs to the gastrin-cholecystokinin family but is considered here because of its location. The presence of a further peptide (PYL^a) of unknown biological significance was postulated on the basis of cDNA analysis; so far it has been isolated only in a shortened form, PGL^a [80, 172]. The xenopsin precursor has a length of 80 amino acids and carries the sequence of the octapeptide at its C-terminus; its sequence agrees up to 86% with the levitide precursor and to 54% with that of physalaemin [172]. There are at least three caerulein precursors in *Xenopus* skin, and these are classified, according to the number of caerulein sequences they contain, into types I, III and IV. Types I and III are encoded by different genes, both of which have eight exons. The sequence of the decapeptide caerulein is encoded in small 45-bp exons. In the gene of type I, one caerulein-encoding exon is replaced by a coding sequence for a novel peptide that appears to have only six amino acids (60%) in common with caerulein but has not yet been isolated [180, 225]. The other peptides detectable in *Xenopus* skin apparently arise from the precursors of TRF, xenopsin, PGL^a and caerulein by cleavage at single arginine or lysine residues [70]. An endoprotease specific for the sequence – RXVRG – has been isolated from *Xenopus* skin [112].

8.2.3 Hormones of the Urophysis

In addition to the hypothalamus-hypophysis system, the bony fish have a further neurosecretory region in the bone marrow of the tail, the urophysis, which produces several hormonal factors known as **urotensins**. Urotensin I (Fig. 8.2 g) is 41 amino acids long and significantly homologous to the mammalian CRF. Urotensin II (Fig. 8.2 h) shows some similarity in its 12 amino acids to SST-14 and appears in the fish in several forms [86].

8.3 Hormones of the Adenohypophysis

8.3.1 Glycoproteins of the Adenohypophysis

Lutropin (LH), **follitropin** (FSH) and **thyrotropin** (TSH) of the hypophysis, and **choriongonadotropin** (CG) of the placenta are all glycoproteins of about 30 kDa made up of an α and a β chain. The α chains of all hormones in a species are identical, or at least very similar, whilst the β chains responsible for the hormone activity are homologous but quite variable. The α-chain sequences of different mammals differ in only 2–5% of their 96 amino acids, and compared with those of the teleosts by only 30–35% [9, 87, 126]. The chum salmon *Oncorhyncus keta* is so far the only vertebrate known to have two different α chains (in gonadotropin I). The two chains are only 72% similar, which is little more than the similarity to the bovine α chain [87]. Although the rate of evolution of the β chains is significantly higher than that of the α chains, the species-specific differences of the β chains of a particular hormone are much less than the sequence differences between the β chains of different hormones in one species. Thus, in humans, almost all pairwise comparisons of β chains give merely 30–40% identical amino acids; only the β chains of LH and CG agree by 80%. In contrast, the sequence similarity of the LH β chains from the sperm whale and swine is 90%, from the dog and rat is 89%, and from the sperm whale and man is 70%; the human and equine CG β chains agree by 66% [166, 213, 236]. The α and β chains of different species can combine to give biologically active hybrid molecules, independently of the normal specificity of the β subunit [187]. The car-

bohydrate content of the hormones varies widely, e.g. from 16% in bovine LH to 45% in equine CG.

The amphibians, turtles, crocodiles, birds and mammals produce two different **gonadotropins** (GTH): FSH and LH. There is apparently only one GTH present in lizards and snakes, and this is immunologically distinct from mammalian FSH but has no LH activity. On the other hand, sheep FSH is immunologically cross-reactive not only with the FSH of other mammals but also with that of birds, turtles, crocodiles and amphibians. Usually only one GTH is present in fish and this shows species-specific differences in structure and activity. For example, androgen production in mammalian testis is stimulated to the same extent by the GTH from *Tilapia* sp. as by mammalian LH, but is stimulated 500-fold less by the GTH from *Muraenesox cinereus*. The salmonids contain two gonadotropins, GTH-I and GTH-II, which are produced in different cells of the hypophysis and differ greatly in their β subunits. The subunit GTH-Iβ has only 31% sequence agreement with GTH-IIβ, but almost 40% agreement with human LHβ and FSHβ. In contrast, GTH-IIβ is very similar to the β subunits of other teleost GTHs and has higher sequence similarity to human LHβ (48%) than to FSHβ (38%). Thus, the differentiation between LH and FSH is already indicated here [87]. The in vitro association of the α and β subunits of mammalian LH proceeds very slowly, requiring several hours at normal body temperature and several days at low temperatures. Fish GTHs, in contrast, show adaptation to low temperature, e.g. the association of GTH subunits of carp is almost complete after 45 min at 20 °C. Hybrids of bovine LHα and *Cyprinus* gonadotropin β associate at an intermediate rate.

The **choriongonadotropins** have been examined in detail only in the primates and Equidae, but have also been detected in the rat, mouse, guinea-pig, rabbit and sheep. The human and equine CGβ chains are 30 amino acids longer at the C-terminus than are the 115–118 amino acids of LHβ and FSHβ; this implies a stop-codon mutation and is probably related to the localization in the placenta [213]. There is only one human LHβ gene but seven genes for CGβ, although only three are expressed. There is also only one CGα gene, but despite this there is a surplus of α chains in the placenta [213]. In the horse, the amino acid sequences of the β chains from LH and CG are identical, including the C-terminal extension typical of CG; only the carbohydrate contents are different, with 24% in LH compared with 45% in CG. As a result of these particular characters, equine LH/CG has both LH and FSH activity; furthermore, the LH/CG β subunits can be combined only with α subunits of the same species and, unlike other β subunits, cannot form hybrid molecules [23, 213]. Peptides that are immunologically similar to LH and FSH have been detected in the CNS of the cockroach *Periplaneta americana;* these substances stimulate testosterone secretion in the Leydig cells of mouse testis [223].

The gonads and the placenta produce two peptides which inhibit FSH release from the anterior lobes of the hypophysis by a negative-feedback mechanism; these are **inhibin** A and B. They are heterodimers of a glycosylated α chain of 16 kDa and a β chain of 14 kDa, which occurs in pigs and cattle as two sequence variants [35, 210]. In addition to the inhibins, the follicle fluid also contains homo- and heterodimers of the β subunits, which stimulate FSH release and are consequently referred to as **activins**. Finally, this fluid also contains a monomeric glycoprotein of 35 kDa, **follistatin**, which is apparently not related to the inhibins but nevertheless inhibits FSH release [201].

8.3.2 Pro-Opiomelanocortins

Adrenocorticotropin (ACTH), the **lipotropins** (β- and γ-LPH), the **melanotropins** (α-, β- and γ-MSH) and the opioid **endorphins** (β- and α-END) arise by proteolysis of a common precursor, pro-opiomelanocortin (POMC) (Fig. 8.3). The hormone-coding domains in this case are also flanked by pairs of basic amino acids which are attacked by specific serine proteases. After cleavage, the peptides are further modified, e.g. glycosylated (ACTH), N-acetylated (α-MSH, β-END), C-terminal amidated (α-MSH), Ser/Thr phosphorylated (CLIP), or Tyr sulphated (ACTH). The processing is tissue specific: the POMC produced in the pars distalis gives primarily ACTH and β-LPH; in the pars intermedia, ACTH is in part further cleaved to α-MSH and CLIP, and β-LPH to γ-LPH and β-END, and in some species also to β-MSH [185]. In *Xenopus laevis*, and apparently also in the salmon *Oncorhynchus keta* and other vertebrates, there are two POMC genes, and thus two sets of cleavage products [133, 249]. The POMC mRNA of *O. keta* lacks the coding sequence for γ-MSH [101].

Human ACTH differs from that of the mouse in 2, from that of the frog in 9 and from that of the

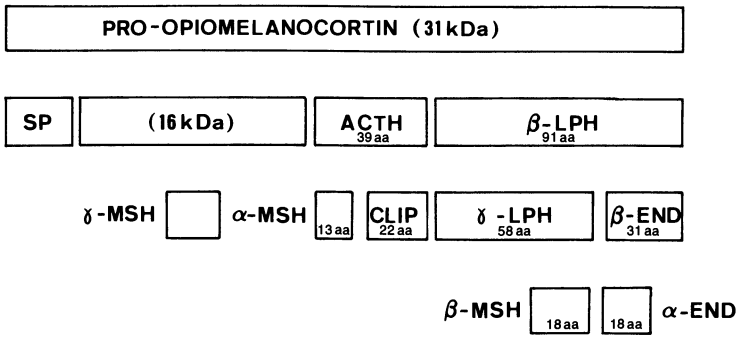


Fig. 8.3. Proteolytic products of human prepro-opiomelanocortins [43]. *ACTH*, adrenocorticotropin; *CLIP*, hormonally inactive peptide; *END*, endorphins; *LPH*, lipotropins; *MSH* melanotropins; *SP*, signal peptide

fish in 15 of its 39 amino acids [133]; the ACTHs of the ostrich *Struthio camelus* and the shark *Squallus acanthias* are identical to mammalian ACTH in the first 12 positions, but differ in 6 and 8 amino acids, respectively, in the C-terminal region. Substances with immunological similarity to ACTH have been detected in the earthworm *Lumbricus terrestris*, and MSH-like substances have been detected in the gastropods [93]. As POMC contains three copies of the **melanotropin** sequence – His-Phe-Arg-Trp –, cleavage produces three different MSHs in all vertebrates (Fig. 8.3). Mammals all contain the same α-MSH, which only differs in the shark *Squalus acanthias* at one of 13 positions. On the other hand, β-MSHs are quite variable, even in the mammals; *Squalus* β-MSH (15 amino acids) differs from the bovine variety (18 amino acids) in 8 of the comparable amino acids. In the shark, the peptide corresponding to mammalian β-MSH is hormonally inactive [35]. **β-LPH sequences** of 79–93 amino acids are known from various mammals and the ostrich; they show particularly high variability at the N-terminus [124].

8.3.3 Opioids

Peptides have been found in the brains of many vertebrates from the mammals to the bony fish which bind to the same receptors as the plant opiate alkaloids. All these opioids have the pentapeptide sequence of the encephalins (Tyr-Gly-Gly-Phe-Met/Leu) at the N-terminus, but arise from three different precursors: POMC, proencephalin and pro-dynorphin. POMC gives rise to the 31-amino-acid β-endorphin (β-END) and the C-terminal-shortened variant α-END (Fig. 8.3). Only 9 of the 31 amino acids of the β-ENDS vary from the mammals to the amphibians, e.g. human β-END differs from that of the turkey in only 6 positions. In the salmon, positions 6–12 are replaced by a completely different sequence of 9 amino acids, giving a β-END of 33 amino acids [79]. The human prepro-encephalin gene consists of four exons, of which the largest (exon IV) encodes all active peptides: four Met-ENC, one Leu-ENC and two Met-ENC extended at the C-terminus by – Arg-Phe – and – Arg-Gly-Leu –. Human and bovine prepro-encephalins differ in only 36 of 267 amino acids [249]. The prepro-encephalin of *Xenopus laevis* encodes seven Met-ENC but no Leu-ENC. The brains of turtles, crocodiles and lizards as well as the hagfish *Eptatretus stouti* contain Met-ENC, Leu-ENC and Met-ENC-Arg-Phe but no Met-ENC-Arg-Gly-Leu. *Petromyzon marinus* possesses Met-ENC and Leu-ENC but no C-terminal extended forms of the Met-ENC [54, 122]

Just as the endorphins can be looked upon as large Met-encephalins, so the mammals, and also probably the amphibians, possess large Leu-encephalins e.g. **dynorphin**, α-neoendorphin and rimorphin, which arise from a third precursor. Prepro-dynorphin, which contains three encephalin sequences, differs between man and the pig in 59 out of 256 amino acids. The similarity of the encephalin sequences near the C-terminus and of the cysteine residues near the N-terminus makes it likely that the encephalin and dynorphin precursors are homologous; however, as the central regions of the known precursors can not be definitely aligned, the evolutionary distance of the two precursors cannot be properly assessed [249]. Substances with immunological similarity to endorphin and encephalin have been found in the annelids, gastropods, cephalopods and bivalves [2]. Met-ENC and Leu-ENC have actually been isolated and sequenced from the thoracic ganglia of the shore crab *Carcinus maenas* [128]. A family of opioid peptides was recently discovered in the skin of arboreal frogs of the subfamily Phyllome-

dusinae, and these peptides have the highest known morphine-like activity. They include the μ-receptor-specific **dermorphin** (Tyr-D-Ala-Phe-Gly-Tyr-Pro-Ser-Nh_2) and the δ-receptor-specific **dermencephalins** and **deltorphins**, which all carry a D-Ala or D-Met residue in position 2. In the mRNAs of the longer precursors, however, the corresponding positions have the codons GCG or AUG, coding for L-Ala or L-Met; frog brain contains in addition to the 2-D-Met dermencephalin the corresponding L-Met compound, in fact at a 100-fold higher concentration. The D-amino acid residues apparently arise by post-translational processing of the precursor via an incompletely described isomerization reaction of low efficiency [143].

8.3.4 Growth Hormone, Prolactin and Lactogen

The growth hormone **somatotropin** (GH) and **prolactin** (PRL) from the adenohypophysis together with **lactogen** (LS) or chorionsomatomammotropin from the placenta belong to the same hormone family. GH is present in all vertebrates, with the possible exception of the agnathans, and is consistently involved in growth regulation. In contrast, the equally widely found PRL has very varied functions, from osmoregulation in the teleosts to control of lactation in mammals. The role of PRL in birds is not yet clear; in the pigeon it stimulates production of crop milk [228].

The **growth hormones**, whose amino acid sequences have been determined directly or via cDNA or gene sequences in all classes of vertebrates, are very variable. The GH genes of most vertebrates, including the carp *Cyprinus carpio*, consist of five exons; in *Salmo gairdneri* and *S. salar* there are six exons [36, 90]. GH varies in length, e.g. within the mammals from 178 to 196 amino acids; the smallest GH, with only 173 residues, is found in the flounder *Paralichthys olivaceus* [142]. The similarity of the amino acid sequences correlates with the extent of the phylogenetic relationships. For mammalian GHs, between 98% (primates) and 66% (human/sei whale) identical amino acids are found. Comparisons with the GH of the sea turtle *Chelonia midas* gave the following percentage similarities: chicken 89%, rat 79%, bullfrog *Rana catesbeiana* 69%, blue shark *Prionace glauca* 68%, and eel 58%; for the chum salmon *Oncorhyncus keta* it is only 40%. Differences in GH sequences are greatest within the teleosts; compared with the carp *Cyprinus carpio*, the salmon *Salmo salar* shows 66% agreement, but the yellow tail *Seriola quinqueradiata* shows only 49% [36, 105, 242].

The **PRL** sequences vary within and between vertebrate classes to an extent similar to that of the GH sequences. Relative to the turtle *Chelonia midas*, one finds 86% identical amino acids in the chicken, 81% in the horse, pig and fin whale, 75–71% in cattle, sheep and humans, 60–56% in the mouse and rat, and only 35–31% in various fish species [94, 228, 243]. Further members of the GH/PRL family have been discovered in the hypophyses of the Atlantic cod *Gadus morrhua* and the flounder *Paralichthys olivaceus;* these are 21–27% similar to the GH and PRL of the teleosts and are therefore referred to as somatolactins [178]. The GH/PRL family also includes peptide hormones produced in the placenta. The placental **lactogen** (PL), which stimulates milk synthesis, has so far been detected in primates, Artiodactyla, Lagomorpha and Rodentia; it is by no means certain that it is universal in the Placentalia. The bovine hormone is very different from that of all other species, e.g. with a molecular mass of 32 kDa it is much larger than the rest (20–22 kDa). Whereas human PL has 85% sequence agreement with human GH, bovine PL is more similar to PRL than to GH (51% agreement as against 20%). The similarity of bovine and human PL is only 20%, but compared with placental hormones of the rodents it is about 30% [196]. Further PRL-like placental proteins have been detected in cattle, mice and rats [239].

8.4 Hormones of the Stomach, Intestine and Pancreas

8.4.1 The Gastrin-Cholecystokinin Family

The peptides **gastrin** and **cholecystokinen** (CKK) from the alimentary canal, together with caerulein from frog skin, may be considered homologous based upon the agreements shown by their C-terminal sequences (Fig. 8.2i). As the precursors can be cleaved at various positions, there exist gastrins and cholecystokinins with N-terminal regions of different length: gastrin-34, -17 and -14; and CKK-58, -39, -33, -22 and -8. The sequence – YMGWMDF – is sufficient for binding to the CKK receptor; the gastrins contain the

homologous sequence – YGWMDF-amide [167]. The tyrosine residue of CKK is always sulphated, but in gastrin it is sulphated only in certain vertebrates. The **CKK receptor** in mammalian pancreas is highly specific for the sulphated CKK, whereas the receptors in the stomach and brain bind both sulphated and non-sulphated gastrins just as well as CKK [249]. The chicken also contains the specific pancreas receptors and the non-specific brain receptors; whereas in fish, amphibians and reptiles all the receptors are non-specific [224]. CKK is apparently phylogenetically older than gastrin; fish and amphibians already possess CKK but not gastrin. CKK is already found in both the gut and brain of the agnathans, and also in the skin of amphibians [35]. A peptide of 36 amino acids, isolated from the stomach wall of the chicken, has the C-terminal sequence – WMDF – but otherwise has no similarity to mammalian gastrin or to CKK. It stimulates secretion in the stomach but not in the pancreas; this may, in fact, be an avian gastrin. CKK has also been detected immunologically in both the gut and brain of birds [51].

The **prepro-gastrin** of mammals consists of 101 amino acids; the **CKK precursor** has a length of 114 amino acids as the result of an insertion. Otherwise, the cDNA sequences of the two precursors are very similar; 63 % of bases in codon positions 1 and 2 are identical [48]. Internal homologies indicate a gene duplication during their evolution [18]. Human gastrin-34 corresponds to positions 59–92 of the precursor; the immediate flanking sequence – Gly-Arg-Arg – provides both the cleavage site and the glycine for amidation of the terminal – COOH. Cleavage of the precursor at position 75/76 gives gastrin-17. The human and porcine precursors differ at 29 positions, of which only 4 lie within the gastrin-34 region [18]. Gastrin- or CKK-like peptides have been detected immunologically in cnidarians, annelids, insects, crustaceans, xiphosurans, gastropods, bryozoans and tunicates [61, 71, 154]. A sulphated, gastrin-like peptide has even been sequenced from the cockroach *Leucophaea maderae* (see Fig. 8.6k) [148]. Cionin (NYYGWMDF-amide) from the ascidian *Ciona intestinalis* is identical at the C-terminus with gastrin, but is sulphated on both tyrosine residues [91]. The decapeptide caerulein, found in the skin and gut wall of amphibians (Fig. 8.2i), and described with other frog-skin peptides on p. 294, differs in only one position of the active peptide from the CKK sequence and has CKK activity [167].

8.4.2 The Insulin Family

From the point of view of comparative biochemistry, the insulins are by far the best-investigated peptide hormones. In mammals, insulin is produced in the B cells of the pancreas islets tissue; here one also finds the glucagon-producing A cells and other endocrine cell types. In many teleosts, the endocrine pancreas tissue is a separate organ, the so-called Brockman'sche body, and this has facilitated isolation of the hormone mRNA, e.g. from the angler fish *Lophius americanus*. In the elasmobranchs, as in the tetrapods, the islets tissue is integrated in the pancreas [38, 60]. Human insulin is made up of an A chain with 21 and a B chain with 30 amino acids and these are linked by two disulphide bridges. The primary translation product is **prepro-insulin** with 113 amino acids. The pro-insulin which remains after cleavage of the signal peptide is made up of the B chain at the N-terminus and the A chain at the C-terminus; the intervening region (the C peptide) is excised after A and B are joined (Fig. 8.4). The storage granules of the B cells contain crystalline insulin in the form of hexamers with bound zinc. Isolated B cells from the rat form at least six other peptides in addition to insulin and these are also possibly constituents of the storage granules [206]. Insulin is dissolved in the blood as a monomer.

The **sequences** of the insulin A and B chains, and in some cases even of prepro-insulin, are known for more than 40 vertebrate species from

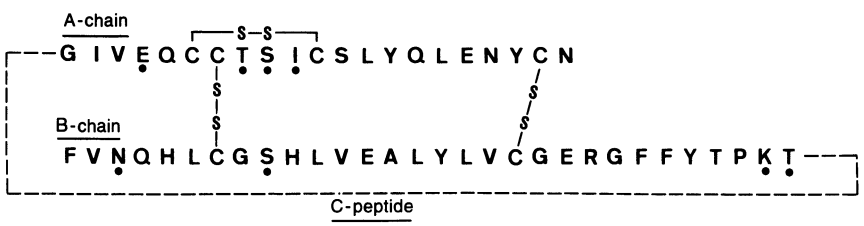


Fig. 8.4. Human insulin and the location of the C peptide. The variable positions in the mammals except the Hystricomorpha) are marked by *dots* [72]

the agnathans to the mammals. Greater differences between species are found in the C peptide than in the A and B chains; this is related to the fewer functional specifications of the former (Table 4.12; p. 161). The amino acids responsible for hormone activity, zinc binding and chain association are highly conserved; 22 of the 51 insulin amino acids are invariant in all vertebrates [35]. The insulins of most **mammals** are very similar to human insulin, differences being found at only eight positions of the two chains (Fig. 8.4) [72, 131]. The insulin of the South American douroucouli *Aotes trivirgatus* differs markedly from those of other primates; compared with that of man, *Aotes* prepro-insulin has three amino acid substitutions in the signal peptide, two in the B chain, seven in the C peptide and three in the A chain. The receptor binding affinity of *Aotes* insulin is fivefold lower than that of human insulin [198].

The most marked differences to other mammalian insulins are shown by the insulins of the Hystricomorpha, e.g. the African porcupine *Hystrix cristata*, the guinea-pig *Cavia porcellus*, the casiragua *Proechimys quarirea*, the chinchilla *Chinchilla laniger*, and the nutria *Myocastor coypus*. The extremely aberrant nutria insulin has only 5 % of the receptor affinity and 3 % of the effect on lipogenesis shown by porcine insulin in a test system involving rat fat cells. Due to amino acid substitutions at the zinc-binding site B10-His and positions responsible for molecular contacts, the hystricomorphan insulins are always monomers. Compared with bovine insulin, that of casiragua has 17 differences, that of the guinea-pig 18, and that of nutria as many as 22; the spatial structures, however, remain similar [6]. Despite large sequence differences, the insulins of the Hystricomorpha fit into the genealogical tree of the mammalian insulins and are therefore orthologous according to the definition on p. 155.

As expected, the **fish** differ most of all from the mammals [12, 40, 108, 186]. The prepro-insulin of the agnathan *Myxine glutinosa*, which with 115 amino acids is somewhat longer than that of the mammals, has the signal sequence (26 amino acids), B chain (31 amino acids), C peptide (37 amino acids) and A chain (21 amino acids) arranged consecutively. However, the mature insulin differs at 16 positions from that of the carp and at 21 positions from human insulin. It cannot bind zinc and can only form dimers, whereas insulin from the shark *Scyliorrhinus canicula* can already form zinc-stabilized hexamers. The insulin of the lamprey *Petromyzon marinus* differs from that of the common hagfish *Myxine* (17 amino acid differences) more than from porcine insulin or from that of the salmon (both with 14 differences) [168]; this underlines the large evolutionary distance between the two agnathan groups Myxinoidea and Petromyzontoidea. There are 17 differences between the insulin of the carp *Cyprinus carpio* and human insulin, including two extra N-terminal amino acids on the B chain; however, the zinc-binding B10-His and the amino acids responsible for aggregation and biological activity are all present. Carp insulin in the fat-cell test has only threefold weaker binding and has the same maximum activity as porcine insulin. Amino acid substitutions in the region of the B/C cleavage site lead to 37- to 38-amino-acid B chains in the holocephalans *Chimaera monstruosa* and *Hydrolagus colliei*, whereas that of the related species *Callorhynchus milii* has only the normal 31 amino acids [12].

The **gene sequences of human prepro-insulin** and those of a range of mammals, the chicken, various teleosts and *Myxine* are now known. All consist of three exons: the first exon codes for a part of the non-translated (NT) 5′ sequence; the second for the remaining 5′-NT, the signal sequence, B chain and a part of the C peptide; and the third for the rest of the C peptide, the A chain and the 3′-NT [209]. The sequence coding for the signal peptide occurs once again in a slightly modified form in the C-coding region. It is possible that two genes with identical signal peptide regions arose initially by gene duplication, and were then later joined by a large intron to give one gene. There are two prepro-insulin genes in the mouse and rat which lie only 100 kb apart on chromosome 1 of the rat but are found separately on chromosomes 6 and 7 of the mouse. In both rodent species, gene I is very similar to gene II in the coding sequence but lacks the second intron. As in the chicken insulin gene, the presence of two introns is probably the ancestral state; the insulin genes rat-I and mouse-I apparently arose by the reinsertion of partly processed transcripts. It is interesting that the retrotransposons of the two species differ much more than the original insulin genes rat-II and mouse-II [232].

Insulin-like substances have been detected by both immunological and pharmacological methods in a wide spectrum of very different organisms: tunicates, insects, crustaceans, gastropods and cephalopods, and even in the ciliate *Tetrahymena*, the lower fungi *Aspergillus* and *Neurospora*, and the prokaryote *E. coli* [68, 118]. Insulin-receptor proteins have been identified

immunologically in *Neurospora* and *Streptococcus* [50], and insulin-like peptides have been sequenced via their cDNA in the snail *Lymnaea stagnalis* and the insects *Bombyx mori* and *Locusta migratoria* (p. 307). Only in the case of the insects have sufficient quantities of insulin-like peptides been isolated with which to study activity. In the honey bee, reinjection of the peptide brings about a reduction in the concentrations of trehalose, glucose, diacylglycerols and phospholipids, and an increase in free fatty acids and triacylglycerols in the haemolymph. As in the vertebrates, these typical insulin effects are inhibited by the simultaneous injection of vertebrate somatostatin [22]. *Drosophila melanogaster* possesses not only a typical insulin receptor but also a specific insulin protease like those known in mammals; this suggests specific functions for insulin in the fly [56, 63].

Insulin belongs to the same super-family as the ovarian hormone relaxin, the insulin-like growth factors IGF-I and IGF-II (somatomedins), and probably also the β chain of the nerve growth factor (NFG) and the lens growth factor lentropin [8]. **Relaxins** synthesized in the ovary and also in the uterus and placenta are made up of an A chain of 22–24 and a B chain of 25–35 amino acids, and are considered to be homologous to the insulins, especially because of the positions of the disulphide bridges. In mammals, relaxin is responsible for adaptation of the uterus musculature to the foetus and for expansion of the pubic bone symphysis prior to birth [197]. The two human relaxins have only 77 % sequence agreement. Relaxins are already found in the cartilaginous fish, the relaxins of the shark species *Squalus acanthias* and *Odontaspis taurus* being 45–46 % similar to human relaxin; the relaxins of the pig and rat show an equivalent degree of similarity. The relaxin of the skate *Raja erinacea*, which has 48 % amino acids in common with that of *Squalus* and 31 % with human relaxin, has only 1 % of the activity of mammalian relaxin when tested on mammals [28]. Relaxin-like substances are even found in the ciliate *Tetrahymena pyriformis*. It also appears from this that the substance is much older than its present function in the mammals. The observation that the insulin activity of organ extracts is often not completely abolished by anti-insulin antibodies led to the discovery of the **insulin-like growth factors** (IGFs). There are two IGFs (IGF-I and IGF-II) and corresponding specific receptors. Both IGFs bind to the insulin receptor, but insulin itself binds only to the IGF-I receptor and not to that of IGF-II. IGF-I (with 70 amino acids) shows more than 60 % sequence similarity with IGF-II (67 amino acids). The three regions B, C and A are distinguishable in the IGF precursors, as in pro-insulins, but they are extended at the C-terminus by a short D region and an E region, which consists of about 35 amino acids in pro-IGF-I and of 85–90 amino acids in pro-IGF-II. Region E is removed during IGF processing, but sections C and D remain in the mature molecule. The IGFs are bound to specific proteins in the blood plasma, possibly in order to prevent interaction with receptors, e.g. the IGF concentration of 100 nmol/l in human plasma is 1000-fold higher than that of insulin and would have a 50-fold higher insulin activity [85, 119]. IGFs are known from vertebrates of all classes down to the Atlantic hagfish *Myxine glutinosa* [150]. The acranian *Branchiostoma californiensis* has an insulin-like peptide (ILP) which has approximately the same sequence similarity to insulin in both IGFs, and arises from a precursor with the regions B–C–A–D–E; thus, it is apparently very closely related to the ancestral gene of the insulin family [32].

8.4.3 The Glucagon-Secretin Family

Glucagon, glicentin and secretin, together with the vasoactive intestinal peptide (VIP), the gastric inhibitory peptide (GIP), the growth-hormone releasing factor (GRF) and the helospectins in the poison of the venomous lizard *Heloderma suspectum* (p. 324), make up a family of peptides with 9–15 of their 27–43 amino acids in common (Fig. 8.2 k) but with widely varying activities. Glucagon and glicentin stimulate gluconeogenesis and glycogenolysis, whereas secretin, VIP and the helospectins promote secretion from the pancreas. In addition, VIP brings about relaxation of the vascular and other smooth muscles; GIP inhibits acid and pepsin formation in the stomach. The **glucagon** secreted by the A cells of the islets tissue and the **glicentin** produced by the L cells of the gut mucosa arise from the same precursor by differential proteolytic cleavage. The human precursor and that of various mammals has a length of 180 amino acids and consists of a signal sequence, glicentin with 69 amino acids, including the glucagon sequence in positions 33–61, and two glucagon-like sequences GLP-1 and GLP-2. The angler fish *Lophius americanus* possesses two glucagon genes which, however, lack the GLP-2 region [3, 161]. The teleosts *Ictalurus punctata, Oncorhynchus kisutch* and

Cottus scorpius, the archaic gar-pike *Lepisosteus spatula* (Holostei), and the skate *Torpedo marmorata* all produce only one glucagon and one GLP. Mammals, and also lower vertebrates such as the bullfrog *Rana catesbeiana* and *Lepisosteus spatula*, possess not only glucagon and GLP but also the somewhat longer **oxyntomodulin** (36 amino acids), which is made up of the glucagon sequence and a C-terminal extension; this peptide is 5- to 10-fold less active than glucagon on liver cell membranes but is approximately 20-fold more effective on stomach mucous membranes [170]. Analysis of the prepro-glucagon sequences suggests that GLP-2 arose from glucagon about 1200 million years ago and GLP-1 about 800 million years ago, and that in fish the GLP-2 was lost before duplication of the glucagon gene.

Glucagon is a highly conserved polypeptide; with the exception of the guinea-pig, the sequences of all mammals to have been examined are identical. The chicken and turkey differ from the mammalian sequence at only one position, the duck at two positions, the shark *Scyliorhynchus canicula* at four positions, and the coho salmon *Oncorhynchus kisutch* at eight positions [3]. The GLPs are evolving rather more rapidly: the GLP-1 sequenced from *Ictalurus* differs from the two GLPs of the angler fish *Lophius* in 5 and 9 out of 34 amino acids, from the GLP-1 of the hamster at 11 positions, and from hamster GLP-2 at as many as 23 positions [3, 38]. In contrast to insulin, the glucagons in the storage granules of islet cells are mostly amorphous; only in the teleosts are they crystalline, probably with a trimeric structure. Glicentin apparently has trophic functions in the gut mucosa and may be looked upon as an enteroglucagon. A further GLP, synthesized in the endocrine L cells of the porcine gut, corresponds to amino acids 33–69 of glicentin and is referred to as enteroglucagon II. The special evolutionary position of the guinea-pig gastro-enteropancreatic peptides is as clearly illustrated by glucagon and VIP as by insulin: whereas all other mammals have a glucagon of 29 amino acids, the guinea-pig molecule is 40 amino acids long. Similarly, the VIP sequences of all other mammals are identical but guinea-pig VIP differs in 4 of the 28 amino acids [55]. The VIP of the chicken also differs from the typical mammalian sequence at four positions and those of the cod *Gadus morrhua* and the shark *Scyliorhinus canicula* differ at eight positions [52]. GIP is not quite as highly conserved, with the human sequence differing in 2 and 4 of 42 amino acids from the porcine and bovine sequences, respectively [214]. **Secretin** is a peptide of 27 residues that stimulates pancreas secretion. The rat secretin gene shows homology only to the other genes of this hormone family in the exons encoding the bioactive peptide, and the only translated product is secretin, whereas the genes for pro-glucagon and pro-VIP each produce several different bioactive peptides [88, 107].

8.4.4 The Family of Pancreas Polypeptides

The pancreas polypeptide (PP) of the chicken, discovered in 1975, is a C-terminal amidated peptide of 36 amino acids that functions as an antagonist of glucagon, inhibiting lipolysis in the fat cells. PP belongs to a family of polypeptides including peptide YY from the gut wall (PYY) and from the brain NPY, with which it shows 50–70 % similarity [202, 216]; also included is peptide YG from the pancreas of the angler fish *Lophius americanus* (aPY). The PP-producing cells are a specific type of endocrine pancreatic cell distinct from the glucagon-producing A cells, the insulin-producing B cells and the somatostatin-producing D cells. The human, sheep and cat PP precursor includes a C-terminal 20-residue peptide that is absent in the rat and the mouse. The guinea-pig, whose insulin and glucagon differ markedly from those of other mammals, produced a PP precursor with 89 % agreement to the human molecule in the hormone region but with only 40 % similarity in the icosapeptide part [15, 244]. The aPY precursor is identical in structure to that of PP. The 64 % similarity of the active peptide of aPY with PP and NPY is much greater than the agreement between the remaining regions; this is comparable with the higher rate of evolution of the insulin C peptide compared with the A and B chains [7]. PPs have been reported from many vertebrate species and have 9 out of 36 invariant residues [125, 171]. PP-like substances have also been detected in annelids and insects.

8.5 Peptides Regulating Blood Pressure

The juxtaglomerular cells situated in the kidney glomeruli release **renin** in response to a decrease in blood pressure or to sodium deficiency. Renins are found in several forms; as well as in the kidneys, they are produced in the salivary glands and in the female genital tract. They are all glycoproteins of about 40 kDa, made up of two chains

with unequal lengths. These arise by post-translational proteolysis from a prepro-renin of about 50 kDa. The corresponding genes of man, the mouse and rat show several special features: the mouse alone possesses two prepro-renin genes, and the human gene contains an additional, tenth exon that encodes for a 3-amino-acid insert [29]. The physiological substrates of the renins are the **angiotensinogens**, which are multiple glycoproteins of about 450 amino acids produced by the liver cells [65, 141]. Renin cleaves a decapeptide from angiotensinogen, angiotensin-I, which is further processed to the biologically active octapeptide angiotensin-II by an angiotensin-converting enzyme (Fig. 8.5a). This is a zinc-containing dipeptidase, synthesized mainly in the endothelium but also in other tissues [13]. The renin-angiotensin system appears to be present in all vertebrates with the exception of the agnathans and some elasmobranchs. The angiotensin-I sequence differs only at positions 1, 5 and 9. Position 1 is occupied by aspartic acid (in most tetrapods) or asparagine (in most teleosts); position 5 is occupied by isoleucine (in most mammals) or valine (in ruminants and all non-mammals); and position 9 is occupied by histidine (all mammals, the turtle *Pseudemys scripta*, and the teleost *Lophius litulon)*, serine (in the chicken), tyrosine (in the snake *Elaphe climocophora)*, asparagine (in the salmon *Oncorhynchus keta* and the frog *Rana catesbeiana)* or glycine (in the eel *Anguilla rostrata)* [35, 99].

Bradykinin is a blood-pressure-lowering nonapeptide that, like the blood-pressure-elevating angiotensin, is released from a larger precursor (kininogen) by the activity of a specific protease (kallikrein) (Fig. 8.5b). Amphibians contain in addition 6-T-bradykinin, which carries Thr instead of Ser at position 6, and ranakinin, which has a six-amino-acid C-terminal extension. The **ornithokinin** of the chicken differs from mammalian bradykinin by two amino acid substitutions: 6-Ser to Thr and 8-Phe to Leu. It is mainly the latter which is responsible for the species- and group-specific activities of this peptide; ornithokinin causes contraction of chicken smooth muscle but not of the rat uterus [100]. Peptides of 9–18 amino acids, which include the sequence of bradykinin or 6-T-bradykinin, are found in the venom of various Hymenoptera (see Fig. 9.3, p. 325). The blood plasma of the mammals contains two **kininogens** with different molecular weights. The larger high molecular weight kininogen (HMWK) in humans is a glycosylated chain of 114 kDa. In this case, the release of bradykinin leaves behind a cleaved kininogen consisting of an H and an L chain of 63 and 58 amino acids, respectively, bound by a disulphide bridge. This molecule associates with various components of the blood clotting system and is responsible for the initiation of the intrinsic pathway (p. 203). LMWK, the second, significantly smaller kininogen (69 kDa), forms the same H chain but a smaller L chain. The human and bovine HMWK sequences show 74 % agreement, and human and murine sequences show 61 % [98]. The rat possesses a third kininogen type, the T kininogen. In contrast, the chicken has only ornithokininogen, which corresponds to the HMWK of mammals and agrees in 13 (43 %) of the 30 N-terminal amino acids with the bovine form [100]. The **kallikreins**, which are responsible for the cleavage of kininogens, are found in the mammals, e.g. in the kidney, bladder, prostate, pancreas and sublingual salivary glands, in various tissue-specific forms which arise either by post-translational modification or as the products of different genes. In the mouse and rat, there are gene families of 12–24 members which include not only the gene for true kallikrein but also kallikrein-like or tonin-like sequences, as well as many pseudogenes [37, 234]. The kallikreins make up one of the families of serine proteases and their evolution has been discussed on p. 91. The **endothelins**

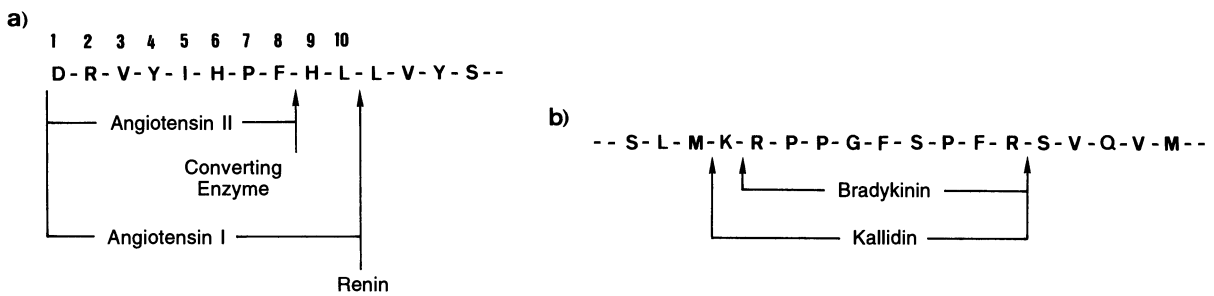


Fig. 8.5. **a** The formation of angiotensin-I and -II from angiotensinogen. **b** The formation of bradykinin and kallidin from kininogen

(ETs) also have vasoconstrictive effects; this is a family of acidic peptides with 21 amino acids which includes not only the ET isoforms ET-1, ET-2 and ET-3 but also the vasoactive intestinal contractor (VIC) of the mouse and the four sarafotoxins in the venom of the snake *Atractaspis engaddensis*. The prepro-ET-1, synthesized in the endothelial cells, has a length of 202–212 amino acids according to the species. A pro-ET-1 with 38 (human) or 39 amino acids (pig) is released proteolytically, and this gives rise to the mature ET-1 by the unusual cleavage of a 21-Trp/22-Val bond [58, 204). The homeostatic control of blood pressure and osmolality is the function of the **natriuretic peptides** from the cardiac atria (ANP) and the brain (BNP); these also arise from longer precursors. Whilst the circulating α forms of ANP vary by at the most one amino acid amongst the mammals, the related peptides of the chicken, the frog *Rana ridibunda* and the eel *Anguilla japonica* show sequence differences amongst themselves and compared with the mammalian ANPs and BNPs of about 50 %. ANP-like substances have been detected immunologically in gastropods and crustaceans [215, 245).

8.6 Calcitonins and Parathormone

Calcitonin and parathormone are involved in calcium homeostasis, with calcitonin depressing the calcium level in the blood plasma, and parathormone treatment resulting in its elevation. **Calcitonins** (CT) are synthesized in the C cells of the mammalian thyroid gland and in the ultimobranchial bodies of other vertebrates. They have been detected either immunologically or by functional tests in all vertebrates except the agnathans. The CTs consist of 32 amino acids (Fig. 8.21), of which the 7 N-terminal positions are identical in the human, bovine, porcine and salmon molecules, but the remaining residues differ widely. In spite of this, the fish CT has hormone activity in birds and mammals but not in fish. Thus, the substance itself appears to be older than its function in some of the vertebrates [169]. In the CT precursor of man, the rat and chicken, the CT sequence lies between long C- and N-terminal sequences; the cleavage sites here also consist of pairs of basic amino acids. The CT gene is made up of six exons, one of which encodes the whole CT sequence, and a further one codes for a homologous peptide of 37 amino acids, the calcitonin gene-related peptide (CGRP). The relative proportions of these two peptides are regulated in a tissue-specific manner by alternative splicing in the two mammals mentioned and in the chicken. CT predominates in the thyroid and ultimobranchial bodies, and CGRP predominates in the brain [59]. CT/CGRP-like substances have been detected immunologically in the brains of all classes of vertebrate including the agnathans, and even in the lobster *Nephrops norvegicus* [103, 238].

The parathyroid glands (small epithelial bodies) produce the **parathormone** (PTH). Of the 84 amino acids making up this molecule, there is only 8 % difference between the sequences of cattle and swine, but 33 % difference between these two mammals and the chicken. PTH is probably present in all tetrapods but not in the fish. In bovine prepro-PTH, the PTH sequence is preceded by a signal sequence of 25 amino acids and a pro-sequence of 6 amino acids [35].

8.7 Growth Factors

The growth of vertebrate cells in tissue culture depends upon peptide growth factors (GFs) which are found, for example, in blood plasma. For this reason, many tissue culture media contain foetal calf serum. In contrast to the hormones, which are stored intracellularly and secreted as required, the growth factors are released continuously from cells. These factors bind to **specific receptors** on the cell surface. This activates a tyrosine kinase which, via a cascade of further specific kinases, initiates the cell's growth response [30, 137, 240]. In many cases, the GF-producing cells themselves possess receptors and respond to their own growth factors. It may be that the evolution of signalling by peptides began with this autocrine mechanism; this then developed to produce effects on neighbouring cells (paracrine mechanism) and finally, by transmission in the extracellular fluids, to effects on distant cells (endocrine mechanism). Unlimited cell division (proliferation) is a characteristic of tumour cells and, as might be expected, there is a relationship between growth regulation by GF and carcinogenesis. There exist, for example, transforming GFs, and some oncogenes code for aberrant forms of the growth factors or their receptors.

The continuously increasing number of known growth factors can be divided into families according to structure, origin and activity. The

hormonally active β subunit of nerve **growth factor** (NGF) is produced as a dimer by the sensory and sympathetic neuron target tissues. The submandibular glands of mice contain a 230-kDa complex (7S-NGF) made up of two α chains, a β chain, two γ chains and one or two zinc atoms. Based on their sequences, the α and γ chains belong to the kallikrein families of the serine proteases, although the α chain is proteolytically inactive. The 7S-NGF and similar complexes from other tissues do not bind to the NGF receptor and their biological significance is not clear [136]. The brain-derived neurotrophic factor (BDNF) of the rat and NT-3 of mice are homologous to NGF; the three sequences have 48 % of their amino acids in common (130). The murine prepro-NGF of 307 amino acids includes the 118 amino acids of the mature β-NGF sequence in positions 188–305. The corresponding human chain differs in 11 amino acids, the bovine chain in 16 amino acids, and that of the chicken in 19 amino acids. NGFs are also found in snake venoms but their sequences are, at present, largely unknown.

The **epidermal growth factors** (EGFs) have lengths of between 48 and 53 amino acids (Fig. 8.2 m). The rat EGF sequence of 48 amino acids shows 77 % agreement with the murine EGF (53 amino acids), 69 % with the human sequence (53 amino acids), and only 67 % with the guinea-pig EGF (51 amino acids) [205]. Surprisingly, the EGF precursor is anchored in the cell membrane and contains within its 1200 amino acids the real EGF and eight EGF-like sequences. The EGF receptor is an integral membrane protein of 1186 amino acids with many N-linked oligosaccharide chains. Similarly to other growth factor receptors, the cytoplasmic domain of the EGF receptor has tyrosine kinase activity; this represents the beginning of a complex signal transduction pathway [30]. DNA sequences have been identified in *Drosophila* which encode polypeptides similar to the tyrosine kinase and the extracellular domain of the human EGF receptor [227]. The EGF family includes further GFs which bind to the EGF receptor despite their low sequence similarity, e.g. human amphiregulin, the transforming growth factor TGF-α, and the pox virus growth factor (VGF). On the other hand, TGF-β, which has also been detected in *Drosophila* eggs and embryos [11], constitutes a distinct family of multifunctional regulatory peptides [44]. EGF-like sequences are found in many proteins of both vertebrates and invertebrates, and present a particularly impressive example of the importance of exon shuffling in the evolution of proteins: 1 such sequence is found in tissue plasminogen activator and urokinase, 2 sequences are found in the clotting factors IX and X, 3 in the low density lipoprotein (LDL) receptor, 11 in the gene "lin" of the nematode *Caenorhabditis elegans*, at least 18 in the gene "uEGF-1" of the sea urchin *Strongylocentrotus lividus*, and 36 in the gene "notch" of *Drosophila melanogaster* and the related gene "xotch" from *Xenopus laevis*.

There are other peptides with regulatory effects on cell growth and differentiation: one of the most interesting of these peptides is the **head activator** of the freshwater polyp *Hydra;* this defines the anterior end during regeneration. It is an undecapeptide (Fig. 8.2 n) which not only occurs in *Hydra* and other Cnidaria but, remarkably, is also found in the gut and hypothalamus of the rat and man [16, 192]. The growth of blood vessels in mammals is stimulated by angiogenin, a homologue of the pancreas RNase [19]. The differentiation of fibroblasts to myoblasts is controlled by three related peptides known as myogenin, myoD and myf5 [64]. A structurally and functionally very similar factor, SUM-1, arises as the translation product of an mRNA from the sea urchin *Lytechinus variegatus* [222]. A bovine bone-inducing protein has been isolated and is referred to as the osteogenic protein (OP) [110]. **Prothymosin** α is an acidic peptide of 110 amino acids, found in many mammalian tissues, and has immuno-enhancing effects. The first substance of this type to be found, thymosin α, from calf thymus, corresponds to the 28 N-terminal amino acids of prothymosin α_1 and is probably an extraction artefact. **Parathymosin**, isolated from rat thymus, is antagonistic to prothymosin α and has 15 % sequence similarity [229]. Thymosin α_1-like substances have been detected immunologically in insects, crustaceans, protozoans, lower fungi and even bacteria [157]. The homologous proteins thymosine β_4, β_9 and β_{10} form a distinct family of immuno-modulatory peptides [229].

8.8 Peptide Hormones of Invertebrates

Peptide hormones are in fact found in all groups of invertebrates but have so far been mainly investigated in the molluscs and arthropods. Because of problems in obtaining sufficient quantities, their isolation and characterization has proved difficult. However, the development of more sensitive methods of peptide sequencing,

e.g. fast atom bombardment spectroscopy and the new possibilities for cDNA and gene sequencing, has led, during the last few years to rapid increases in the number of available sequences [104, 115, 166, 219]. The invertebrate peptide hormones are also produced from larger precursors by proteolytic processing and, in many cases, are also post-translationally modified. The formation of D-amino acid residues from the L-forms of the precursors, as described above for the dermorphins of the vertebrates, has also been described for the mollusc peptides, achatin and fulicin [104, 158]. The discovery of a new peptide hormone in any one animal group is often followed by evidence that this or similar substances occur in other groups, or are even ubiquitous. Examples of this are provided by the detection of vertebrate hormones in invertebrates or the discovery of peptides similar to the mollusc FMRF-amide in cnidarians, annelids and insects [104]. Comparisions across taxon borders should allow a deeper insight into the relationships between peptide structure and function. However, the evolutionary relationships between the peptide hormones cannot be determined from short sequence comparisons because of the possibilities of convergence. A much better basis for comparison are gene or precursor sequences but these are, as yet, in short supply.

Apart from mollusc and arthropod hormones, only a few regulatory peptides are known from other invertebrate groups. Amongst these are the sperm-activating peptides (SAPs) of sea urchin eggs; these stimulate sperm metabolism and motility and act as chemo-attractants for the spermatozoa. SAPs are peptides of 10–14 amino acids which are species specific but occur in several variant forms within one species. Forty different SAPs have so far been described from nine sea urchin species. The first of these substances to be found were speract (GFDLNGGGVG) from *Strongylocentrotus purpuratus* and resact (CVTGAPGCVGGGRL-amide) from *Arbacia punctulata*. Speract arises from a precursor of about 260 amino acids containing 10 speract or speract-like sequences [177, 246]. The **conjugation initiating substances** of the ciliates have also been intensively investigated. These unicellular organisms proliferate asexually by fission; occasionally, a sexual process (conjugation) occurs in which two individuals fuse, exchange genetic material, and separate. In some species, conjugation is restricted to individuals of opposite "mating types". Each mating type produces a different conjugation initiating substance, and these are known generally as mating pheromones or gamones but are sometimes named after the genus. Two gamones are known from *Blepharisma japonicum*, the first species to be investigated: blepharom, a basic glycoprotein of 20 kDa, and blepharisom, which has been identified as calcium-3-(2′-formylamino-5′-hydroxybenzoyl)acetate. Genetic analysis of several species of the genus *Euplotes* has shown that the gamones are encoded by alleles of one gene locus. The pheromones Er-1 (40 amino acids) and Er-10 (38 amino acids) of *E. raikovi* have been sequenced; they show no similarity to any other proteins [175].

8.8.1 Peptide Hormones of Molluscs

It was known for some time that ganglion extracts of various molluscs could amplify the contraction of isolated hearts of the Venus mussel; the responsible factor was later isolated from the ganglia of the mussel *Macrocallista nimbrosa* and was shown to be a C-terminal amidated tetrapeptide, **FMRF-amide**. This peptide is apparently distributed ubiquitously in molluscs, mostly in association with smaller amounts of the related peptide FLRF-amide. The precursor of these peptides from *Aplysia californica* has been sequenced; it consists of 28 copies of the FMRF sequence and only 1 FLRF sequence. Various mollusc species have been found to possess variants of these two peptides with 1- to 6-amino-acid N-terminal extensions (Fig. 8.6a). The mussel *Mytilus* produces a decapeptide which ends with the FFRF-amide sequence. Whilst the peptides mentioned so far all have stimulatory effects on muscle contraction, a hexapeptide with the C-terminal sequence LFRF-amide, found in *Fusinus*, has an inhibitory effect. FMRF-amide-related peptides (FaRPs), ending with FXRF-amide, have been found in nematodes, annelids and arthropods in addition to the molluscs, and RF-amide-terminating peptides are even more widely distributed. The evolutionary relationships between these regulatory peptides will become clear only after analysis of their gene or precursor sequences has been undertaken [104]. A gene sequenced from *Drosophila melanogaster* codes for a polyprotein with at least 13 FaRP sequences, and a cDNA from the sea anemone contains 19 precursor sequences of the antho-RF-amide (QGRF-amide) [47, 193].

A large number of other myotropic peptides and peptide amides have been found in the molluscs, e.g. the catch-relaxing peptide (CARP) of

```
a) FMRF-amide-like peptides and opioids
    (widespread)                                    FMRF-amide
    (widespread)                                    FLRF-amide
    (Helix aspersa)                               pEDPFLRF-amide
    (Lymnaea stagnalis)                            GDPFLRF-amide
    (Octopus vulgaris)                             YGGFMRF-amide
     for comparison: Met-Encephalin                YGGFM
b) ELH (Aplysia californica)
                ISINQDLKAITDMLLTEQIRERQRYLADLRQRLLEK-amide
c) CDCH (Lymnaea stagnalis)
                LSITNDLRAIADSYLYDQHWLRERQEENLRRRFLEL-amide
d) Eclosion hormone (Manduca sexta)
    NPAIATGYDPMEICIENCAQCKKMLGAWFEGPLCAESCIKFKGKLIPECEDFASIAPFLNKL
e) Proctolin                                          RYLPT
f) Paragonial peptide (Drosophila funebris)
                        DVPSANANANNQRTAAAKPQANAEASS
g) AKH I (Locusta, Schistocerca)              pELNFTPNWGT-amide
   AKH II-L (Locusta migratoria)              pELNFSAGW-amide
   AKH II-S (Schistocerca gregaria)           pELNFSTGW-amide
   CC I  (Periplaneta americana)              pEVNFSPNW-amide
   CC II (Periplaneta americana)              pELTFTPNW-amide
h) RPCH (Pandalus borealis)                   pELNFSPGW-amide
i) αPDH (Pandalus borealis)          NSGMINSILGIPRVMTEA-amide
j) PDF  (Romalea microptera)         NSEIINSLLGLPKLLNDA-amide
k) Leucosulfakinin (Leucophaea maderae)    EQFGDYGHMRF-amide
   for comparison: Gastrin-17 (pig)       -EEEEAYGWMDF-amide
```

Fig. 8.6a–k. Peptide hormones of invertebrates [57, 95, 134, 160, 166, 179, 182]. *pE*, pyroglutamic acid

Mytilus, the myomodulins of various prosobranchs, the *Mytilus* inhibitory peptides (MIPs), and the *Fusinus* and *Achatina* excitatory peptides (FEPs and AEPs). The mussel *Mytilus edulis* possesses at least 17 peptidic substances which affect the anterior byssus retractor muscle (ABRM). Achatin-1 from the pulmonate snail *Achatina*, which interestingly includes a D-amino acid residue, Gly-D-Phe-Ala-Asp, induces voltage-dependent uptake of sodium ions by the periodically oscillating neuron (PON). Achatin-II, which is also present and has only L-amino acids, has no such activity [104]. The unusually large size of the neurons of *Aplysia californica* facilitates analysis of the peptides from single ganglion cells. Neuron B15 produces two small cardioactive peptides (SCP_A and SCP_B with 11 amino acids), which stimulate the contraction of various mollusc muscles, buccalin (11 amino acids), which inhibits muscle contraction, and a 28-residue peptide, which promotes water retention in the animal. Neuron B16 synthesizes myomodulin, a peptide structurally related to the SCPs and buccalin. Neuron 14 of the abdominal ganglion produces a neuropeptide precursor of 85 amino acids which is cleaved into three peptides with cotransmitter activity. All the sensory neurons of *Aplysia* contain an mRNA which codes for an inhibitory cotransmitter with codons 46–54 [27, 45, 104, 231).

Of particular interest is the **neurohormonal control system of egg laying** which occurs in all gastropods, but has been examined in detail only in the prosobranch sea snail *Aplysia californica* and the pulmonate freshwater snail *Lymnaea stagnalis*. In *Aplysia*, egg-laying behaviour begins with persistent electrical activity of the bag cells, two clusters of neuroendocrine cells in the abdominal ganglion. These secrete at least nine peptides, including the egg-laying hormone (ELH) of 36 amino acids (Fig. 8.6b) and four bag cell peptides (BCP-α to -δ), all of which arise from the same precursor. The atrial gland, which secretes into the oviduct, contains three pro-hormones encoded by different genes which, by rather complicated processing, give rise to two main types of regulatory peptides: the peptides A, A' and B, which stimulate ELH production by the bag cells, and the three ELH-related peptides, which are linked to the 18-residue peptide A-AP by a disulphide bridge and mimic the actions of ELH. The ELH gene of the bag cells and the three genes for the ELH-related peptides of the atrial gland are homologous [151]. Egg laying by *Lymnaea stagnalis* is also a complicated process that lasts for

several hours and is controlled by about 100 neuroendocrine cells in the cerebral ganglion (caudodorsal cells). These cells secrete various peptides which probably arise from more than one large precursor. The best-known of these peptides is the ovulation hormone CDCH (Fig. 8.6 c), which agrees in 16 of its 36 amino acids with the ELH of *Aplysia*. Taking into account the fact that the prosobranchs and the pulmonates separated about 400 million years ago, these peptides appear to be very highly conserved [57].

8.8.2 Peptide Hormones of Arthropods

Although many studies have been carried out of the regulatory peptides of insects and crustaceans [20], the other arthropod groups have been rather neglected. Of the insect neuropeptides, the best-known are those of the silkworm *Bombyx mori*. Four types are found, in some cases with several variants: the prothoracicotropic hormone (PTTH), the diapause hormone, the hatching (eclosion) hormone (EH), and the melanizing and red-pigment concentrating hormone (MRCH). The enormous number of individuals needed for the isolation of these peptides is most easily attained with the silkworm; Japanese research groups made use of 10 million male insects for this purpose. The **prothoracicotropic hormone** is produced from a pair of neurosecretory cells in the brain and stimulates ecdysterone release from the prothorax glands. PTTH has 109 amino acids and arises from a precursor of 224 amino acids [97]. In addition to the species-specific PTTH, the *Bombyx* brain contains several peptides of about 5 kDa, the **bombyxins**, which in fact affect ecdysone release from the prothorax glands of *Samia cynthia* and were, therefore, earlier known as 4K-PTTHs; their function in *Bombyx mori* is, however, not yet clear. They belong to the insulin family and have the typical dimeric structure with A and B chains linked by a disulphide bridge. They are encoded by a gene family with many members which can be subdivided into three classes, A–C. The probombyxins have the same domain organization, B–C–A, as the pro-insulins [89]. An insulin-related peptide (LIRP) of unknown function from *Locusta migratoria* has been sequenced via the cDNA; LIRP shows no greater sequence similarity to the bombyxins than to the insulins [113]. Insects contain further peptides with glandotropic hormone effects, for example, allatotropin, a 13-residue peptide from *Manduca sexta* which stimulates the secretion of the juvenile hormone from the corpora allata; allostatins (8–13 amino acids), isolated from the cockroach *Diploptera punctat*, have an antagonistic effect [96, 237]. The production of sexual attractants by the females of the lepidopterans *Bombyx mori* and *Heliothis zea* is stimulated by the 33-amino-acid pheromone-biosynthesis-activating neuropeptides (PBANs) [102, 176].

The **eclosion hormone** synthesized in the neurosecretory brain cells of the silkworm controls the hatching of the adult animal from the pupal case; the earlier moultings are also regulated by EH produced in abdominal tissues. In addition to the full-length form of 62 amino acids (Fig. 8.6 d), *Bombyx* EH is also found in N- and C-terminal shortened forms. *Manduca* EH differs at 12 positions from that of *Bombyx* [82, 106]. The **paragonial peptide** from males of *Drosophila funebris* (Fig. 8.6 f) influences the courting behaviour of the female and is, therefore, effectively a pheromone. At least three variants of the **melanizing and red-pigment concentrating hormone** are found in *Bombyx mori* but, as yet, only the N-terminal sequences are known. These show only small differences to each other but, as in the case of PTTH, have some similarity to insulin and IGF-II [135].

Proctolin is a pentapeptide which was known already in 1967 to initiate contractions in the musculature of the posterior gut (proctodaeum) of the cockroach *Periplaneta americana* (hence the name); it was first characterized structurally in 1975 using material extracted from 125 000 individuals (Fig. 8.6 e). Proctolin is found in most, but by no means all, insects; for example, it is apparently absent from *Manduca sexta*. It has also been found in crustaceans, the xiphosuran *Limulus polyphemus*, the leech *Hirudo medicinalis*, and even in the hypothalmus of the rat [162]. Many other insect myotropic peptides have been found and sequenced in recent years. The pyrokinins, which stimulate contractions of the posterior gut and the oviduct, are characterized by the C-terminal sequence – FXPRL-amide. Several such peptides are known from the cockroach *Leucophaea maderae* and from *Locusta migratoria* [149, 163, 195]. *Locusta* also contains peptides, the locustatachykinins, which have some similarity to the vertebrate tachykinins and affect the gut musculature [194]. A myotropic peptide with similarity to the vertebrate gastrin has been isolated from the cockroach *Leucophaea maderae* (Fig. 8.6 k). The cardioactive peptide known from the brachyuran *Carcinus maenas* (CCAP, carcinus

cardioactive peptide) has also been found in *Locusta*. In contrast, the cardioactive peptide corazonin, from *Periplaneta americanus*, has a very different structure [208, 221].

The first-described insect neuropeptide, which was sequenced in 1976, was the 10-amino-acid **adipokinetic hormone** (AKH-I) from the corpora cardiaca of the migratory locusts *Locusta migratoria* and *Schistocera gregaria*. The hormone acts at the start of flight to ensure that diacylglycerol, necessary for metabolism in the flight muscles, is released from the fat bodies into the haemolymph. As in all Orthoptera, there is in addition to the normal AKH-I a shorter AKH-II with eight amino acids and a somewhat different sequence (Fig. 8.6 g). Surprisingly, the **red-pigment concentrating hormone** (RPCH) from the crustacean *Pandalus borealis* has a completely different function but differs from the AKH-II of *Locusta* in only one amino acid (Fig. 8.6 h), and thus belongs to the same peptide family. In fact, the AKHs have effects on the pigment cells of crustaceans and the RPCHs mobilize lipids in insect fat bodies. To date, there are 18 peptides known which belong to the AKH/RPCH family, most of which are found in more than one species. They are made up of 8–10 neutral amino acids, carry N-terminal a pyroglutamate residue, and are amidated at the C-terminus (Fig. 8.6 g & h). The AKH of the dipterans *Phormia terraenovae* and *Drosophila melanogaster* is the only one known to carry a charged amino acid (Asp) [66, 67]. The prepro-AKH of *Manduca sexta* has only one AKH sequence behind the signal sequence, and the precursors of both AKHs of the locust *Schistocera nitans* also have this organization. In the latter case, the sequence of the prepro-AKH-I (63 amino acids) agrees only 55 % with that of prepro-AKH-II (25, 155]. Regulation of the blood-sugar level in the crustaceans is not via peptides of the AKH/RPCH family, but by a special class of protein hormone which also inhibits moulting. The hyperglycaemic hormones (CHHs) from the sinus glands of *Carcinus maenas* and *Homarus americanus* agree by 61 % in their 71–72 amino acids. Lying between the signal peptide of prepro-CHH of *C. maenas* and the CHH sequence is a 28-residue peptide of unknown function [33, 230].

In addition to the RPCH, two further pigment-dispersing chromatophorotropins have been sequenced in crustaceans: the **α-PDH (pigment-dispersing hormone)** from *Pandalus borealis*, which is also known as DRPH (distal retinal pigment hormone) (Fig. 8.6 i), and the **β-PDH** from *Uca pugilator*. These hormones also appear to be widely distributed amongst the arthropods. Thus, a peptide of 18 amino acids isolated from the locust *Romalea microptera* (Fig. 8.6 j) agrees at 14 positions with β-PDH and at 7 positions with α-PDH of the crustaceans [179].

8.9 Hormone Receptors

The functional links between chemical signals arriving at the cell surface and the intracellular processes take the form of specific **signal transduction pathways**. These begin with a membrane-spanning receptor whose extracellular domain carries a hormone-binding site. The resulting conformational change affects an intracellular effector which brings about the cellular response via a consecutive series of intermediate processes. For one group of receptors, a GTP-binding protein intervenes between the receptor and the primary effector (G-protein-linked receptors). Most receptors are present as subtypes which have differing pharmacological and physiological properties, and are even sometimes linked to different effectors. The subtypes are either encoded by different genes or arise by alternative splicing of the same transcript. The presence of a very specific pattern of receptors on the cell surface leads to differential reactions of organs to the same hormone. Post-translational modification, e.g. phosphorylation, may facilitate adaptation of the receptor properties to specific requirements of the system. There are two quite different receptor types for the neurotransmitter acetylcholine. The muscarinic acetylcholine receptors are stimulated by muscarin, a poison from the fungus *Amanita muscaria*. They are found mainly in the central nervous system (CNS), the sympathetic ganglia and the heart, and belong to the G-protein-linked receptors. The nicotinic acetylcholine receptors are stimulated by nicotine and are inhibited by α-bungarotoxin; because of their rather unusual molecular structure they are placed in a special receptor family. They are found in the neuromuscular junctions of vertebrates and in the electric organs of the electric rays *Torpedo californica* and *T. marmorata* and of the electric eel *Electrophus electricus;* these organs are constructed from specialized muscle plates. Receptors from this family are also found in the CNS. In addition to the G-protein-linked receptors and the nicotinic acetylcholine receptors, the comparative biochemistry

of the receptors for insulin and insulin-like growth factors has also been quite well described.

Each receptor is associated with one of the many intracellular effectors: the adenylate cyclases produce the secondary messenger cAMP; the cGMP phosphodiesterases inactivate the messenger cGMP. Membrane-bound guanyl cyclases function also as receptors for the atrial natriuretic peptide (ANP) and for the peptides speract and resact from sea urchin eggs. The murine guanylate cyclase/ANP receptor is a polypeptide with a hormone-binding extracellular domain of 469 amino acids, a membrane segment of 21 amino acids, and an enzymically active intracellular domain of 567 amino acids [164]. The intracellular domains of the receptors for insulin, insulin-like and other growth factors have tyrosine kinase activity. Phospolipase C acts on phosphatidyl-4,5-bisphosphate to release both diacylglycerol, which activates protein kinase C, and inositol-1,4,5-trisphosphate (ins-1,4,5-P_3), which elevates the cytoplasmic concentration of the secondary messenger Ca^{2+} by mobilizing intracellular reserves and by opening ion channels. The ins-1,4,5-P_3 receptors of the mouse and rat function as calcium channels. They are polypeptides of about 2750 amino acids with seven or eight transmembrane segments close to the C-terminus; these exist as several tissue-specific variants. The corresponding receptor from *Drosophila melanogaster* agrees in 58 % of its amino acids with that of the mammals [139, 152]. Ins-1,4,5-P_3 is either dephosphorylated to ins-1,4-P_2 and recycled during lipid synthesis, or further phosphorylated to ins-P_4, ins-P_5 and ins-P_6, which are all apparently involved in the regulation of the Ca^{2+} concentration [218].

8.9.1 G-Protein-Linked Receptors

To date, more than 100 G-protein-linked receptor types and subtypes are known; they show between 20 and more than 90 % sequence agreement and make up their own super-family. The similarity is more marked between receptors binding the same ligands [5, 17, 211]. This receptor family includes the α- and β-adrenergic receptors (α1, α2, β1, β2, β3), the muscarinic-cholinergic receptors (m1-m5), the receptors for histamine, serotonin, tyramine and dopamine [69, 190, 247], the receptors for substance P, VIP and endothelin [78, 121, 207], and the receptors for the glycoproteins of the adenohypophysis LH, FSH, TSH and CG [189]. The visual-pigment (opsins) and olfactory receptors of vertebrates and insects are significantly homologous to the G-protein-linked hormone receptors [26, 49]. It appears in this case that one signal transduction system has been adapted to cover many signals in the course of evolution. The visual pigments and corresponding G-proteins (transducins) will be examined in more detail in Chapter 19. All the G-protein-linked receptors consist of a polypeptide chain of 402–590 amino acids. The genes of the opsins and various hormone receptors differ in the number of their introns. In contrast, the genes for β2-adrenergic receptors of the hamster and of man have no introns, but flanking repetitive sequences are present. Thus, we are probably dealing here with functional copies of a processed mRNA. The N-terminal domains of the receptors always carry sites for N-linked glycosylation, and the C-terminal domains have several sites which are phosphorylated by specific protein kinases [10]; phosphorylation reduces sensitivity in the case of the β-adrenergic receptor. The membrane topology corresponds to that of bacterial rhodopsin with seven transmembrane domains (Fig 8.7), the amino acid sequences of which are highly conserved. The serotonin receptor of *Drosophila melanogaster* is unique in having eight transmembrane domains [235]. The second and third transmembrane domains of the adrenergic and muscarinic-cholinergic receptors contain acidic amino acid residues which probably form the binding site for the basic ligands. The intracellular region interacts with the G-proteins, which in turn affect the associated effectors. The three β-adrenergic receptors, the serotoninergic receptor 5HT-1A and the dopaminergic receptor D1 stimulate membrane-bound adenylate cyclase via a G_S-protein, whilst the α2-adrenergic receptors, the muscarinic-cholinergic receptors m2 and m4 and the dopaminergic receptor D2 inhibit this effector enzyme via a G_i-protein. The α1-adrenergic receptors, the muscarinic-cholinergic receptors m1, m3 and m5, and the serotoninergic receptor 5HT-2 regulate the hydrolysis of phosphoinositol lipids by phospholipases A_2 or C via a G_i- or G_o-protein, and thereby modulate the production of the second messengers DAG (diacylglycerol) and ins-1,4,5-P_3; they may also regulate other effectors such as K^+ channels [211].

The **G-proteins** function in a cyclic fashion. The non-activated protein is a complex of three polypeptides, α–γ, with a GDP bound to the α chain. Activation of the protein results in the exchange of GDP for GTP and the separation of the α chain from the β/γ dimer. The free GTP-

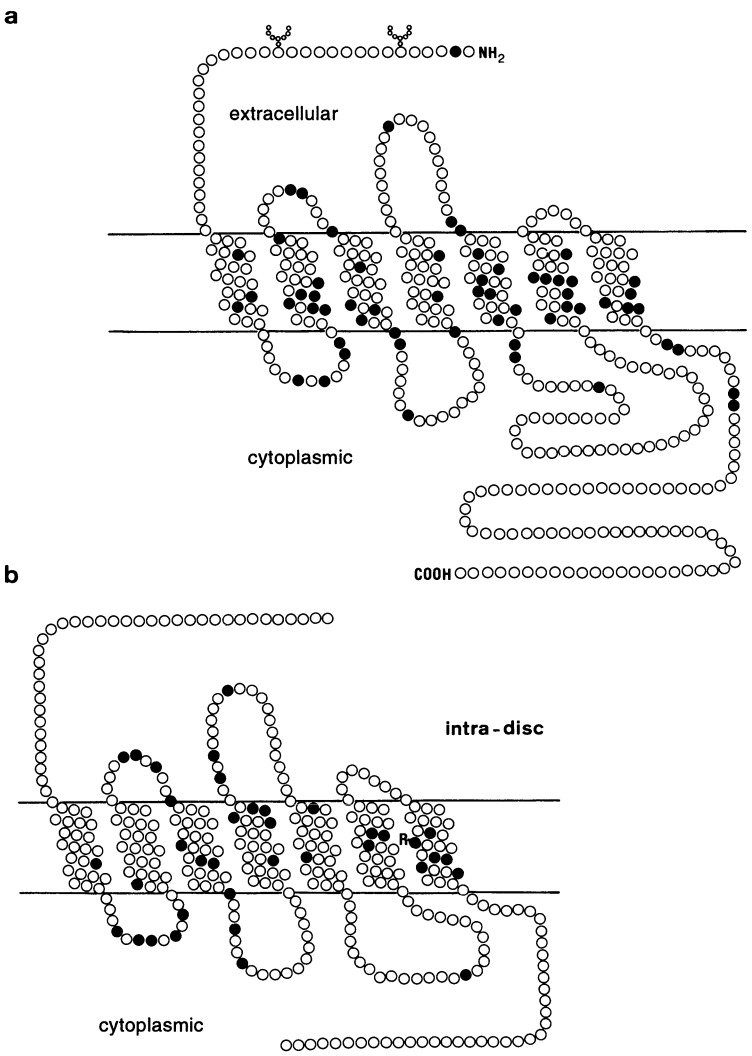


Fig. 8.7a, b. The β2-adrenalin receptor **(a)** and rhodopsin **(b)** have the same tertiary structure with seven transmembrane α helices. **a** The human β2-adrenalin receptor [53]: amino acids in common with bovine rhodopsin are shown as *black dots*; also shown are the positions of the carbohydrate chains. **b** Bovine rhodopsin [53]: amino acids that are invariant in all seven known opsins (one bovine, four human, two from *Drosophila melanogaster*) are shown as *black dots; R* indicates the position of retinal

linked α chain interacts with the effector. Autocatalytic activity of the α chain then hydrolyses the GTP, allowing restoration of the inactive α–β–γ–GDP complex. The specificity of the G-protein classes (G_s, G_i, G_o, G_t = transducin) is determined mainly by the α chains, which have a size of 39–54 kDa, bear the GTP-binding site and have intrinsic GTPase activity. Pertussis toxin uncouples binding of the α chains of many G-proteins to the receptor, thus preventing signal transduction. Cholera toxin inhibits the GTPase activity and thereby amplifies hormone activity. At least 16 different α chains have been described so far, and these can be ordered into four classes according to various structural and functional characters. They agree amongst each other in 35–94 % of their amino acids and exhibit 82 invariant residues. A particular G-protein can bind to many different receptors but shows high specificity for the effector. The α_s chains stimulate adenylate cyclase and open calcium channels. Several tissue-specific α_s variants with different functional properties are produced from the same gene by alternative splicing. The α_i chains, which exist in three forms ($\alpha_{i\text{-}1}$, $\alpha_{i\text{-}2}$ and $\alpha_{i\text{-}3}$) encoded by different genes, inhibit adenylate cyclase. The α_o ("o" for other) chains are found mainly in the nervous system and affect ion channels; they also exist as alternative splicing variants [83]. The α_t chains of the transducins activate cGMP-phosphodiesterase; the rods and cones of the retina, for example, contain transducins with different α_t chains. Olfactory mucous membranes were found to contain an α chain (α_{olf}) which stimulates adenylate cyclase [92]. One cell may contain four or five different α chains and even all three types of α_i chain together. Three types of β chain are known so far (β1 with 35, β2 with 36

and β3 with 37 kDa); these agree in 81–90 % of their amino acid sequences [120]. The β chains of *Drosophila* and *Caenorhabditis* have 80–85 % sequence agreement with those of the mammals [226, 241]. β chains form stable complexes with γ chains (8–10 kDa). Just how many different γ chains there are is not yet clear, but they are more or less tissue specific and the six known sequences agree by up to 55 % [62, 203, 217].

8.9.2 Nicotinic Acetylcholine Receptors

The nicotinic acetylcholine receptor (nAchR) from vertebrate muscle or the electric organ of the electric ray *Torpedo californica* is constructed of five polypeptide chains according to the formula $\alpha_2\beta\gamma\delta$, and these are organized symmetrically around a central channel. Each subunit carries between one and four N-linked carbohydrate chains. The four subunits within any one species are about 35–50 % homologous. In mammals (the rat and cattle), but not in the chicken, the γ subunit in adult animals is replaced by an ε subunit encoded by its own gene [235]. *Torpedo*, the chicken and all mammals that have been examined possess only one α gene, whereas Xenopus laevis has two genes which are expressed simultaneously [76]. nAChRs are also found in the CNS, where they apparently show great diversity: a whole series of nAChR subunits have been identified via their cDNAs in the rat, chicken and goldfish; for example, in the rat there are α2, α3, α4, α5, β2, β3 and nα2 (= β4). The α sequences (α1–α5) show 49–55 % sequence similarity and the β sequences (β1–β3) show 43–64 % [21]. Three α subunits have been sequenced via their cDNAs in *Drosophila melanogaster* and one has been identified in the locust *Schistocerca migratoria* [132, 191].

8.9.3 Insulin Receptors

The mammalian **insulin receptor** consists of four glycosylated polypeptide chains, linked by disulphide bridges, according to the formula $(\alpha\beta)_2$. The exclusively extracellular α subunits (125–135 kDa) make up the insulin-binding site; the β subunits (90–95 kDa) are organized into an extracellular domain associated with the α subunits, a membrane segment of about 23 nonpolar amino acids, and a cytoplasmic domain with tyrosine kinase activity. Binding of a ligand results initially in the self-phosphorylation of the chain, which in its activated form phosphorylates other proteins and sets in motion a cascade leading to the metabolic effects of insulin. The α and β subunits are produced from the same precursor of about 200 kDa. Insulin receptors are present in various very different cell types from the bony fish to man. They remain very similar in binding affinity and specificity, despite the reduced immunological similarity associated with increasing evolutionary distance [199, 212].

The **IGF-I receptor** has the same basic structure as the insulin receptor and is homologous in its subunit sequence. It also binds IGF-II but more weakly than IGF-I. The specific membrane protein identified previously as the **IGF-II receptor** has a structure very different to that of the insulin and IGF-I receptors; it consists of a single polypeptide of 2450 amino acids which shows 80 % sequence agreement with the mannose-6-phosphate receptor of the lysosomes and it therefore belongs to a completely different receptor family [145]. This receptor is unlikely to be responsible for the physiological effects of IGF-II and it may well be that a specific IGF-II receptor awaits discovery.

References

1. Acher, R. and Chauvet J.: Structure, processing and evolution of the neuro-hypophysical hormone-neurophysin precursors. Biochimie 70: 1197–1207 (1988)
2. Andersen A. C. et al.: Immunohistological detection of methionin-enkephalin-like and endorphin-like material in the digestive tract and in the nervous system of the mussel, Mytilus edulis L. Gen. comp. Endocrinol. 62: 111–119 (1986)
3. Andrews P. C. and Ronner P.: Isolation and structures of glucagon and glucagonlike peptide from catfish pancreas. J. biol. Chem. 260: 3910–14 (1985)
4. Andrews P. C. et al.: Isolation and characterization of a variant somatostatin-14 and two related somatostatins of 34 and 37 residues from lamprey (Petromyzon marinus). J. Biol. Chem. 263: 15809–14 (1988)
5. Attwood T. K., Eliopoulos E. E. and Findlay J. P. C.: Multiple sequence alignment of protein families showing low sequence homology. A methodological approach using data-base pattern-matching discriminators for G-protein-linked receptors. Gene 98: 153–159 (1991)
6. Bajaj M. et al.: Coypu insulin. Primary structure, conformation and biological properties of a hystricomorph rodent insulin. Biochem. J. 238: 345–351 (1986)
7. Balasubramaniam A. et al.: Detection of conformational isomers of anglerfish peptide YG (aPY) by reversed phase chromatography. Int. J. Peptide Prot. Res. 34: 158–160 (1989)
8. Beebe D. C. et al.: Lentropin, a protein that controls lens fiber formation, is related functionally and immu-

nologically to the insulin-like growth factors. Proc. Nat. Acad. Sci. USA 84: 2327–30 (1987)
9. Bello P. A. et al.: Cloning and DNA sequence analysis of the cDNA for the common α-subunit of the ovine pituitary glycoprotein hormones. Nucleic Acids Res. 17: 10494 (1989)
10. Benovic J. L. et al.: cDNA cloning and chromosomal localization of the human β-adrenergic receptor kinase. FEBS Letters 283: 122–126 (1991)
11. Benzakour O., Echalier G. and Lawrence D. A.: Drosophila cell extracts contain a TGF-ß-like activity. Biochem. biophys. Res. Commun. 169: 1178–84 (1990)
12. Berks B. C. et al.: Isolation and structural characterization of insulin and glucagon from the holocephalan species Callorhynchus milii (elephant fish). Biochem. J. 263: 261–266 (1989)
13. Bernstein K. E. et al.: The isolation of angiotensin-converting enzyme. J. Biol. Chem. 263: 11021–24 (1988)
14. Bevins C. L. and Zasloff M. A.: Peptides from frog skin. Annual Rev. Biochem. 59: 395–414 (1990)
15. Blackstone C. D. et al.: Novel organization and processing of the guinea pig pancreatic polypeptide precursor. J. Biol. Chem. 263: 2911–16 (1988)
16. Bodenmueller H. and Roberge M.: The head activator: discovery, characterization, immunoassays and biological properties in mammals. Biochim. biophys. Acta 825: 261–267 (1985)
17. Boege F., Neumann E. and Helmreich E. J. M.: Structural heterogeneity of membrane receptors and GTP-binding proteins and its functional consequences for signal transduction (Review). Eur. J. Biochem. 199: 1–15 (1991)
18. Boel E. et al.: Molecular cloning of human gastrin cDNA – Evidence for evolution of gastrin by gene duplication. Proc. Nat. Acad. Sci. USA 80: 2866–69 (1983)
19. Bond M. D. and Strydom D. J.: Amino acid sequence of bovine angiogenin. Biochemistry 28: 6110–13 (1989), Correction l.c. 8262
20. Borkovec A. B. and Gelman D. B. (eds.): Insect neurochemistry and neurophysiology 1986. Humana Press, Clifton 1986
21. Boulter J. et al.: α3, α5, and β4: three members of the rat neuronal nicotinic acetylcholine receptor-related gene family form a gene cluster. J. Biol. Chem. 265: 4472–82 (1990)
22. Bounias M. and Gourdoux L.: Effects of honeybee insulin-immunoreactive peptide on haemolymph lipid and carbohydrate. Interaction of vertebrate somatostatin. Insect Biochem. 16: 721–731 (1986)
23. Bousfield G. R. et al.: Structural studies on equine glycoprotein hormones. Amino acid sequence of equine lutropin β-subunit. J. biol. Chem. 262: 8610–20 (1987)
24. Bradbury A. F. and Smyth D. G.: Enzyme-catalysed peptide amidation: Isolation of an stable intermediate formed by the reaction of the amidating enzyme with the imino acid. Eur. J. Biochem. 169: 579–584 (1987)
25. Bradfield J. Y. and Keeley L. L.: Adipokinetic homone gene sequence fro Manduca sexta. J. Biol. Chem. 264: 12791–93 (1989)
26. Breer H., Raming K. and Boekhoff I.: G-proteins in the antennae of insects. Naturwissenschaften 75: 627 (1988)
27. Brunet J. F. et al.: Identification of a peptide specific for Aplysia sensory neurons by PCR-based differential screening. Science 252: 856–859 (1991)
28. Büllesbach E., Schwabe C. and Callard I. P.: Relaxin from an oviparous species, the skate (Raja cinerea). Biochem. biophys. Res. Commun. 143: 273–280 (1987)
29. Burnham C. E. et al.: Molecular cloning of rat renin cDNA and its gene. Proc. Nat. Acad. Sci. USA 84: 5605–09 (1987)
30. Carpenter G. and Cohen S.: Epidermal growth factor (Minireview). J. Biol. Chem. 265: 7709–12 (1990)
31. Carraway R. E. and Mitra S. P.: Differential processing of neurotensin/neuromedin N precursor(s) in canine brain and intestine. J. Biol. Chem. 265: 8627–31 (1990)
32. Chan S. J., Cao Q. P. and Steiner D. F.: Evolution of the insulin superfamily. Cloning of a hybrid insulin-like growth factor cDNA from Amphioxus. Proc. Nat. Acad. Sci. USA 87: 9319–23 (1990)
33. Chang E. S., Prestwich G. D. and Bruce M. J.: Amino acid sequence of a peptide with both molt-inhibiting and hyperglycemic activities in the lobster, Homarus americanus. Biochem. biophys. Res. Commun. 171: 818–826 (1990)
34. Chauvet M. T. et al.: Guinea pig neurohypophysial hormones: peculiar processing of the three-domain vasopressin precursor. FEBS Letters 210: 40–44 (1987)
35. Chester-Jones I., Ingleton P. M. and Phillips J. G.: Fundamentals of comparative vertebrate endocrinology. Plenum, New York 1987
36. Chiou C. S., Chen H. T. and Chang W. C.: The complete nucleotide sequence of the growth-hormone gene from the common carp (Cyprinus carpio). Biochim. biophys. Acta 1087: 91–94 (1990)
37. Clements, J. A., Matheson B. A. and Funder J. W.: Tissue-specific developmental expression of the kallikrein gene family in the rat. J. Biol. Chem. 265: 1077–81 (1990)
38. Conlon J. M., O'Toole L. and Thim L.: Primary structure of glucagon from the gut of the common dogfish (Scyliorhinus canicula). FEBS Letters 214: 50–56 (1987)
39. Conlon J. M. and Thim L.: Isolation of the tachykinin Des[Ser-1,Pro-2]-scyliorhinin II from the intestine of the ray, Torpedo marmorata. Gen. comp. Endocrinol. 71: 383–388 (1988)
40. Conlon J. M. and Thim L.: Isolation and primary structure of the C-peptide of proinsulin from the European eel (Anguilla anguilla). Comp. Biochem. Physiol. Pt. B 93: 359–362 (1989)
41. Conlon J. M.: [Ser-5]-somatostatin-14. Isolation from the pancreas of a holocephalan fish, the pacific ratfish (Hydrolagus colliei). Gen. comp. Endocrinol. 80: 314–320 (1990)
42. Conlon J. M., Oharte F. and Vaudry H.: Primary structures of the bombesin-like neuropeptides in frog brain show that bombesin is not the amphibian gastrin-releasing peptide. Biochem. biophys. Res. Commun. 178: 526–530 (1991)
43. Cooper P. E. and Martin J. B.: Neuroendocrinology and brain peptides (Review). Trends Neurosci. 5: 186–189 (1982)
44. Cox D. A. and Burk R. R.: Isolation and characterization of milk growth factor, a transforming-growth-factor-ß2-related polypeptide from bovine milk. Eur. J. Biochem. 197: 353–358 (1991)
45. Cropper E. C. et al.: Structure and action of buccalin: A modulatory neuropeptide localized to an identified small cardioactive peptide-containing cholinergic motor neuron of Alysia californica. Proc. Nat. Acad. Sci. USA 85: 6177–81 (1988)

46. Cruz L. J. et al.: Invertebrate vasopressin/oxytocin homologs. Characterization of peptides from Conus geographus und Conus stratus venom. J. biol. Chem. 262: 15821–24 (1987)
47. Darmer D. et al.: Primary structure of the precursor for the sea anemone neuropeptide antho-RF-amide (less-than Glu-Gly-Arg-Phe-NH2). Proc. Nat. Acad. Sci. USA 88: 2555–59 (1991)
48. Deschenes R. J. et al.: Primary structural comparison of the prehormones cholecystokinin and gastrin. FEBS Letters 182: 135–138 (1985)
49. Dhallan R. S. et al.: Primary structure and functional expression of a cyclic nucleotide-activated channel from olfactory neurons. Nature 347: 184–187 (1990)
50. Dietz E., Uhlenbruck G. and Luetticken R.: An insulin receptor in microorganisms: Fact or fiction? Naturwissenschaften 76: 269–270 (1989)
51. Dimaline R., Young J. and Gregory H.: Isolation from chicken antrum, and primary amino acid sequence of a novel 36-residue peptide of the gastrin/CCK family. FEBS Letters 205: 318–322 (1986)
52. Dimaline R. et al.: A novel vasoactive intestinal peptide (VIP) from elasmobranch intestine has full affinity for mammalian pancreatic VIP receptors. Biochim. biophys. Acta 930: 97–100 (1987)
53. Dohlman H. G., Caron M. G. and Lefkowitz R. J.: A family of receptors coupled to guanidine nucleotide regulatory proteins. Biochemistry 26: 2657–64 (1987)
54. Dores R. M. and Gorbman A.: Detection of met-encephalin and leu-enkephalin in the brain of the hagfish, Eptatretus stouti, and the lamprey, Petromyzon marinus. Gen. comp. Endocrinol. 77: 489–499 (1990)
55. Du B. H. et al.: Guinea pig has a unique mammalian VIP. Biochem. biophys. Res. Commun. 128: 1093–98 (1985)
56. Duckworth W. C. et al.: Drosophila insulin degrading enzyme and rat skeletal muscle insulin protease cleave insulin at similar site. Biochemistry 28: 2471–77 (1989)
57. Ebberink R. H. M. et al.: Purification and amino acid sequence of the ovulation neurohormone of Lymnaea stagnalis. Proc. Nat. Acad. Sci. USA 82: 7767–71 (1985)
58. Elshourbagy N. A., Kmetz P. and Sathe G. M.: Nucleotide sequence of endothelin from bovine endothelial cells. Nucleic Acids Res. 18: 4273 (1990)
59. Emeson R. B. et al.: Alternative production of calcitonin and CGRP mRNA is regulated at the calcitonin-specific acceptor. Nature 341: 76–80 (1989)
60. Epple A. and Brinn J. E.: The comparative physiology of pancreatic islets. Springer, Berlin 1987
61. Favrel P. et al.: Immunochemical and biochemical characterization of gastrin/cholecystokinin-like peptides in Palaemon serratus (Crustacea Decapoda): Intermolt variations. Gen. comp. Endocrinol. 65: 363–372 (1987)
62. Fawzi A. B. et al.: Rhodopsin and the retinal G-protein distinguish among G-protein beta/gamma-subunit forms. J. Biol. Chem. 266: 12194–200 (1991)
63. Fernandez-Almonacid R. and Rosen O. M.: Structure and ligand specificity of the Drosophila melanogaster insulin receptor. Mol. cell. Biol. 7: 2718–27 (1987)
64. Fujisawa-Sehara A. et al.: Myogenin contains two domains conserved among myogenic factors. J. Biol. Chem. 265: 15219–23 (1990)
65. Fukamizu A. et al.: Structure and expression of the human angiotensinogen gene. J. Biol. Chem. 265: 7576–82 (1990)
66. Gaede G.: The adipokinetic hormone/red pigment-concentrating hormone pigment family: structures, interrelationships and functions. J. Insect Physiol. 36: 1–12 (1990)
67. Gaede G.: The adipokinetc neuropeptide of Mantodea: Sequence elucidation and evolutionary relationships. Biol. Chem. Hoppe-Seyler 372: 193–201 (1991)
68. Galloway S. M. and Cutfield J. F.: Insulin-like material from the digestive tract of the tunicate Pyura pachydermatina (sea tulip). Gen. comp. Endocrinol. 69: 106–113 (1988)
69. Gantz I. et al.: Molecular cloning of a gene encoding the histamine H2 receptor. Proc. Nat. Acad. Sci. USA 88: 429–433 (1991)
70. Gibson B. W. et al.: Novel peptide fragments originating from PGLa and the caerulein and xenopsin precursors from Xenopus laevis. J. biol. Chem. 261: 5341–49 (1986)
71. Guissi-Kadri S., Bulet P. and Curgy J.: Polypeptides related to mammalian procholecystokinin in the brain of an invertebrate, a marine worm, Nereis diversicolor: Evidence for in ovo translation of mRNA. Gen. comp. Endocrinol. 77: 339–347 (1990)
72. Halldén G. et al.: Characterization of cat insulin. Arch. Biochem. Biophys. 247: 20–27 (1986)
73. Hardie D. G.: Biochemical messengers: Hormones, neurotransmitters and growth factors. Chapman & Hall, New York 1991
74. Harmar A. J., Hyde V. and Chapman K.: Identification and cDNA sequence of delta preprotachykinin, a fourth splicing variant of the rat substance-P precursor. FEBS Letters 275: 22–24 (1990)
75. Harris R. B.: Processing of pro-hormone precursor proteins. Arch. Biochem. Biophys. 275: 315–333 (1989)
76. Hartman D. S. and Claudio T.: Coexpresion of two distinct muscle acetylcholine receptor α-subunits during development. Nature 343: 372–375 (1990)
77. Henzel W. J. et al.: The primary structure of aphrodisin. J. Biol. Chem. 263: 16682–87 (1988)
78. Hershey A. D. and Krause J. E.: Molecular characterization of a functional cDNA encoding the rat substance P receptor. Science 247: 958–962 (1990)
79. Ho C. L., Ko J. L. and Li C. H.: β-Endorphin: peripheral opioid activity of homologues from six species. Int. J. Peptide Protein Res. 29: 521–524 (1987)
80. Hoffmann W.: A new repetitive protein from Xenopus laevis skin highly homologous to pancreatic spasmolytic polypeptide. J. Biol. Chem. 263: 7686–90 (1988)
81. Holmgren S. (ed.): The comparative physiology of regulatory peptides. Chapman & Hall, New York 1989
82. Horodyski F. M., Riddiford L. M. and Truman J. W.: Isolation and expression of the eclosion hormone gene from the tobacco hornworm, Manduca sexta. Proc. Nat. Acad. Sci. USA 86: 8123–27 (1989)
83. Hsu W. H. et al.: Molecular cloning of a new splice variant of the α subunit of the mammalian Go protein. J. Biol. Chem. 265: 11220–26 (1990)
84. Huang Y. P. and Peter R. E.: Evidence for a gonadotropin-releasing hormone binding protein in goldfish (Carassius auratus). Gen. comp. Endocrinol. 69: 308–316 (1988)
85. Humbel R. E.: Insulin-like growth factors I and II (Review). Eur. J. Biochem. 190: 445–462 (1990)
86. Ichikawa T. et al.: In situ hybridization demonstrating coexpression of urotensins I, II-alpha and II-gamma in the caudal neurosecretory neurons of the carp, Cyprinus carpio. Gen. comp. Endocrinol. 71: 493–501 (1988)

87. Itoh H., Suzuki K. and Kawauchi H.: The complete amino acid sequences of α subunits of chum salmon gonadotropins. Gen. comp. Endocrinol. 78: 56–65 (1990)
88. Itoh N. et al.: The secretin precursor gene: Structure of the coding region and expression in the brain. J. Biol. Chem. 266: 12595–98 (1991)
89. Iwami M. et al.: A novel family C of the genes that encode bombyxin, an insulin-related brain secretory peptide of the silkmoth Bombyx mori: isolation and characterization of gene C-1. Insect Biochem. 20: 295–300 (1990)
90. Johansen B., Johnsen O. C. and Valla S.: The complete nucleotide sequence of the growth hormone gene from Atlantic salmon (Salmo salar). Gene 77: 317–324 (1989)
91. Johnsen A. H. and Rehfeld J. F.: Cionin: a disulfotyrosyl hybrid of cholecystokinin and gastrin from the neural ganglion of the prochordate Ciona inestinalis. J. Biol. Chem. 265: 3054–58 (1990)
92. Jones D. T. et al.: Biochemical characterization of three stimulatory GTP-binding proteins. The large and small forms of G-s and the olfactory-specific G-protein G-olf. J. Biol. Chem. 265: 2671–76 (1990)
93. Kaloustian K. V.: Immunochemical evidence for ACTH-like immunoreactivity in tissues of the earthworm Lumbricus terrestris. Comp. Biochem. Physiol. 85A: 351–354 (1986)
94. Karatzas C. N., Zadworny D. and Kuhnlein U.: Nucleotide sequence of turkey prolactin. Nucleic Acids Res. 18: 3071 (1990)
95. Kataoka H. et al.: Isolation and primary structure of the eclosion hormone of the tobacco hornworm, Manduca sexta. Biochem. biophys. Res. Commun. 146: 746–750 (1987)
96. Kataoka H. et al.: Identification of an allatotropin from adult Maduca sexta. Science 243: 1481–83 (1989)
97. Kawakami A. et al.: Molecular cloning of Bombyx mori prothoracicotropic homone. Science 247: 1333–35 (1990)
98. Kellermann J. et al.: Completion of the primary structure of human high-molecular-mass kininogen – The amino acid sequence of the entire heavy chain and evidence for its evolution by gene triplication. Eur. J. Biochem. 154: 471–478 (1986)
99. Khosla M. C. et al.: Identification and synthesis of (1-asparagine, 5-valine, 9-glycine) angiotensin I produced from plasma of American eel, Anguilla rostrata. Gen. comp. Endocrinol. 57: 223–233 (1985)
100. Kimura M. et al.: Isolation and characterization of ornithokininogen. Eur. J. Biochem. 168: 493–501 (1987)
101. Kitahara N. et al.: Absence of a gamma-melanocyte-stimulating hormone sequence in proopiomelanocortin mRNA of chum salmon, Oncorhynchus keta. Comp. Biochem. Physiol. Pt. B 91: 365–370 (1988)
102. Kitamura A. et al.: Amino acid sequence of pheromone biosynthesis-activating neuropeptide (PBAN) of the silkworm, Bombyx mori. Biochem. biophys. Res. Commun. 163: 520–526 (1989)
103. Kline L. W. et al.: Calcitonin gene-related peptide in neural tissues: A phylogenetic study. Gen. comp. Endocrinol. 76: 408–413 (1989)
104. Kobayashi M. and Muneoka Y.: Structure and action of molluscan neuropeptides (Review). Zool. Sci. 7: 801–814 (1990)
105. Kobayashi T. et al.: The complete amino acid sequence of growth hormone of the bull frog (Rana catesbeiana). Biochim. biophys. Acta 1078: 383–387 (1991)
106. Kono T. et al.: Isolation and complete amino acid sequences of eclosion hormones of the silkworm, Bombyx mori. Insect Biochem. 21: 185–195 (1991)
107. Kopin A. S. et al.: The secretin gene: evolutionary history, alternative splicing, and developmental regulations. Proc. Nat. Acad. Sci. USA 88: 5335–39 (1991)
108. Koval A. P., Petrenko A. I. and Kavsan V. M.: Sequence of the salmon (Oncorhynchus keta) preproinsulin gene. Nucleic Acids Res. 17: 1758 (1989)
109. Kozawa H. et al.: Isolation of four novel tachykinins from frog (Rana catesbeiana) brain and intestine. Biochem. biophys. Res. Commun. 177: 588–595 (1991)
110. Kuber T. et al.: Bovine osteogenic proteins is composed of dimers of OP-1 and BMP-2, two members of the transforming growth factor-ß superfamily. J. Biol. Chem. 265: 13198–205 (1990)
111. Kuchler K. et al.: Two precursors of thyrotropin-releasing hormone from skin of Xenopus laevis. Each contains seven copies of the end product. J. Biol. Chem. 265: 11731–33 (1990)
112. Kuks P. F. M. et al.: Xenopus laevis skin Arg-Xaa-Val-Arg-Gly-endoprotease. A highly specific protease cleaving after a single arginine of a consensus sequence of peptide hormone precursors. J. Biol. Chem. 264: 14609–12 (1989)
113. Lagueux M. et al.: cDNAs from neurosecretory cells of brains of Locusta migratoria (Insecta, Orthoptera) encoding a novel member of the superfamily of insulins. Eur. J. Biochem. 187: 249–254 (1990)
114. Lane T. F., Sower S. A. and Kawauchi H.: Arginine vasotocin from the pituitary gland of the lamprey (Petromyzon marinus): Isolation and amino acid sequence. Gen. comp. Endocrinol. 70: 152–157 (1988)
115. Laufer H. and Downer R. G. H. (eds.): Endocrinology of seleted invertebrate types. Alan R. Liss, New York 1988
116. Lazure C. et al.: Complete amino acid sequence of a VLDV-type neurophysin from ostrich differs markedly from known mammalian neurophysins. Int. J. Peptide Protein Res. 30: 634–645 (1987)
117. Lee S. L., Stewart K. and Goodman R. H.: Structure of the gene encoding rat thyrotropin releasing hormone. J. Biol. Chem. 263: 16604–09 (1988)
118. LeRoith D. et al.: Insulin-related material in microbes: similarities and differences from mammalian insulins. Can. J. Biochem. Cell Biol. 63: 839–849 (1985)
119. Leroith D. and Raizada M. K. (eds.): Molecular cellular biology of insulin-like growth factors and their receptors. Plenum, New York 1989
120. Levine M. A. et al.: Moleclar cloning of β3 subunit, a third form of the G protein β-subunit polypeptide. Proc. Nat. Acad. Sci. USA 87: 2329–33 (1990)
121. Lin H. Y. et al.: Cloning and functional expression of a vascular smooth muscle endothelin-1 receptor. Proc. Nat. Acad. Sci. USA 88: 3185–89 (1991)
122. Lindberg I. and White L.: Reptilian enkephalines: Implications for the evolution of proenkephalin. Arch. Biochem. Biophys. 245: 1–7 (1986)
123. Ling N. et al.: Growth hormone releasing factors. Annual Rev. Biochem. 54: 403–423 (1985)

124. Litthauer D., Naudé R. J. and Oelofsen W.: Isolation, characterization and primary structure of two γ-LPH variants from ostrich pituitary glands. Int. J. Peptide Protein Res. 24: 309–315 (1984)
125. Litthauer D. and Oelofsen W.: Purification and primary structure of ostrich pancreatic polypeptide. Int. J. Peptide Protein Res. 29: 739–745 (1987)
126. Liu C. S. et al.: Pike eel (Muraenesox cinereus) gonadotropin. Amino acid sequences of both α and β subunits. Eur. J. Biochem. 186: 105–114 (1989)
127. de Loof A. and Schoofs L.: Homologies between the amino acid sequences of some vertebrate peptide hormones and peptides isolated from invertebrate sources (Minireview). Comp. Biochem. Physiol. Pt. B 95: 459–468 (1990)
128. Lueschen W. et al.: Isolation, sequence analysis, and physical properties of enkephalins in the neurons of the shore crab, Carcinus maenas L. Proc. Nat. Acad. Sci. USA 88: 8671–75 (1991)
129. Luo D. and McKeown B. A.: Immunological evidence of growth hormone-releasing factor-like substances in salmon (Oncorhynchus kisutch and O. keta). Comp. Biochem. Physiol. Pt. B 93: 615–620 (1989)
130. Maisonpierre P. C. et al.: Neurotrophin-3: A neurotrophic factor related to NGF and BDNF. Science 247: 1446–51 (1990)
131. Majewski T. et al.: Isolation and amino acid sequence of insulins and C-peptides of European bison (Bison bonasus) and fox (Alopex lagopus). Int. J. Peptide Protein Res. 30: 379–387 (1987)
132. Marshall J. et al.: Sequence and functional expression of a single α-subunit of an insect nicotinic acetylcholine receptor. Embo J. 9: 4391–98 (1990)
133. Martens G. J. M.: Expression of two proopiomelanocortin genes in the pituitary gland of Xenopus laevis: complete structures of the two prohormones. Nucleic Acids Res. 14: 3791–98 (1986)
134. Marti T. et al.: Microanalysis of the amino acid sequence of the eclosion hormone from the tobacco hornworm, Manduca sexta. FEBS Letters 219: 415–418 (1987)
135. Matsumoto S., Isogai A. and Suzuki A.: N-terminal amino acid sequence of an insect neurohormone, melanization and reddish coloration hormone (MRCH): heterogeneity and sequence homology with human insulin-like growth factor II. FEBS Letters 189: 115–118 (1985)
136. McDonald N. Q. and Blundell T. L.: Crystallization and characterization of the high molecular weight form of nerve growth factor (7S-NGF). J. mol. Biol. 219: 595–601 (1991)
137. Meakin S. O. and Shooter E. M.: Tyrosine kinase activity coupled to the high-affinity nerve growth factor receptor complex. Proc. Nat. Acad. Sci. USA 88: 5862–66 (1991)
138. Michel G. et al.: Evolutionary specificity of hydrins, new hydroosmotic neuropeptides. Occurrence of hydrin 2 (vasotocinyl-Gly) in the toad Bufo marinus but not in the viper Vipera aspis. FEBS Letters 264: 135–137 (1990)
139. Miyawaki A. et al.: Structure-function relationships of the mouse inositol-1,4,5-trisphosphate receptor. Proc. Nat. Acad. Sci. USA 88: 4911–15 (1991)
140. Mizuno J. and Takeda N.: Phylogenetic study of the arginine-vasotocin/arginine-vasopressin-like immunoreactive system in invertebrates. Comp. Biochem. Physiol. Pt. A 91: 739–747 (1988)
141. Moffett R. B.: Purification of multiple forms of plasma angiotensinogen: molecular weight and charge heterogeneity. Biochim. biophys. Acta 912: 1–8 (1987)
142. Momota H. et al.: Amino acid sequence of flounder growth hormone deduced from a cDNA sequence. Nucleic Acids Res. 16: 10362 (1988)
143. Mor A., Delfour A. and Nicolas P.: Identification of a D-alanine-containing polypeptide precursor for the peptide opoid, dermorphin. J. Biol. Chem. 266: 6264–70 (1991)
144. Morel A. et al.: Prosomatostatin processing in anglerfish brain, gut, and pancreas. Biochem. biophys. Res. Commun. 151: 347–354 (1988)
145. Morgan D. O. et al.: Insulin-like growth factor II receptor as a multifunctional binding protein. Nature 329: 301–307 (1987)
146. Morley S. D. et al.: Vasotocin genes of the teleost fish Catostomus commersoni: Gene structure, exon-intron boundary, and hormone precursor organization. Biochemistry 29: 2506–11 (1990)
147. Muehlethaler M., Raggenbass M. and Dreifuss J. J.: Peptides related to vasopressin in invertebrates. Experientia 40: 777–782 (1984)
148. Nachman R. J. et al.: Leucosulfakinin, a sulfated insect neuropeptide with homology to gastrin and cholecystokinin. Science 234: 71–73 (1986)
149. Nachman R. J. et al.: Active conformation of an insect neuropeptide family. Proc. Nat. Acad. Sci. USA 88: 4518–22 (1991)
150. Nagamatsu S. et al.: Evolution of the insulin gene superfamily: Sequence of a preproinsulin-like growth factor cDNA from the Atlantic hagfish. J. Biol. Chem. 266: 2397–2402 (1991)
151. Nagle G. T. et al.: Delta-bag cell peptide from the egg-laying hormone precursor of Aplysia: Processing, primary structure, and biological activity. J. Biol. Chem. 265: 22329–35 (1990)
152. Nakagawa T. et al.: The subtypes of the mouse inositol-1,4,5-trisphosphate receptor are expressed in a tissue-specific and developmentally specific manner. Proc. Nat. Acad. Sci. USA 88: 6244–48 (1991)
153. Nakanishi S.: Substance P precursor and kininogen: Their structures, gene organizations, and regulation. Physiol. Rev. 67: 1117–42 (1987)
154. Nichols R., Schneuwly S. A. and Dixon, J. E.: Identification and characterization of a Drosophila homologue to the vertebrate neuropeptide cholecystokinin. J. Biol. Chem. 263: 12167–70 (1988)
155. Noyes B. E. and Schaffer M. H.: The structurally similar neuropeptides adipokinetic hormone I and II are derived from similar, very small mRNAs. J. Biol. Chem. 265: 483–489 (1990)
156. Nozaki M. et al.: Salmon pituitary gonadotropins. II. Ontogeny of GTH I and GTH II cells in the rainbow trout (Salmo gairdneri irideus). Gen. comp. Endocrinol. 77: 358–367 (1990)
157. Oates K. K. and Erdos M.: Biochemical identification of thymosin alpha-1: its phylogenetic distribution and evolutionary implications. Comp. Biochem. Physiol. Pt. B 94: 759–763 (1989)
158. Ohta N. et al.: Fulicin, a novel neuropeptide containing a D-amino acid residue isolated from the ganglia of Achatina fulica. Biochem. biophys. Res. Commun. 178: 486–493 (1991)
159. Okawara Y. et al.: Cloning and sequence analysis of cDNA for corticotropin-releasing factor precursor

from the teleost fish Catostomus commersoni. Proc. Nat. Acad. Sci. USA 85: 8439–43 (1988)
160. Orchard I.: Adipokinetic hormones – an update (Review). J. Insect Physiol. 33: 451–463 (1987)
161. Orskov C. et al.: Complete sequences of glucagon-like peptide-1 from human and pig small intestine. J. Biol. Chem. 264: 18826–29 (1989)
162. O'Shea M. and Adams M.: Proctolin: from „gut factor" to model neuropeptide. Adv. Insect Physiol. 19: 1–28 (1986)
163. Paemen L. et al.: Isolation, identification and synthesis of Lom-AG-myotropin-II, a novel peptide in the male accessory reproductive glands of Locusta migratoria. Insect Biochem. 21: 243–248 (1991)
164. Pandey K. N. and Singh S.: Molecular cloning and expression of murine guanylate cyclase/atrial natriuretic factor receptor cDNA. J. Biol. Chem. 265: 12342–48 (1990)
165. Pankov Y. A. and Karasyov V. S.: Primary structure of sperm whale luteinizing hormone. Int. J. Peptide Protein Res. 28: 124–129 (1986)
166. Penzlin H.: Neuropeptides: Occurrence and functions in insects. Naturwissenschaften 76: 243–252 (1989)
167. Pincus M. R. et al.: On the biologically active structures of cholecystokinin, little gastrin, and enkephalin in the gastrointestinal system. Proc. Nat. Acad. Sci. USA 84: 4821–25 (1987)
168. Plisetskaya E. M. et al.: Isolation and structure of lamprey (Petromyzon marinus) insulin. Gen. comp. Endocrinol. 69: 46–55 (1988)
169. Poeschl E. et al.: The structure of procalcitonin of the salmon as deduced from its cDNA sequence. FEBS Letters 226: 96–100 (1987)
170. Pollock H. G. et al.: Isolation of alligator gar (Lepisosteus spatula) glucagon, oxyntomodulin, and glucagon-like peptide: amino acid sequences of oxyntomodulin and glucagon-like peptide. Gen. comp. Endocrinol. 69: 133–140 (1988)
171. Pollock H. G. et al.: Isolation of peptide hormones from the pancreas of the bullfrog (Rana catesbeiana). Amino acid sequences of pancreatic polypeptide, oxyntomodulin, and two glucagon-like peptides. J. Biol. Chem. 263: 9746–51 (1988)
172. Poulter L. et al.: Levitide, a neurohormone-like peptide from the skin of Xenopus laevis. Peptide and peptide precursor cDNA sequences. J. Biol. Chem. 263: 3279–83 (1988)
173. Powell R. C., King J. A. and Millar R. P.: (Trp-7, Leu-8) LH-RH in reptilian brain. Peptides 6: 223–228 (1985)
174. Proux J. P. et al.: Identification of arginine vasopressin-like diuretic hormone from Locusta migratoria. Biochem. biophys. Res. Commun. 149: 180–186 (1987)
175. Raffioni S., Luporini P. and Bradshaw R. A.: Purification, characterization, and amino acid sequence of the mating pheromone Er-10 of the ciliate Euplotes raikovi. Biochemistry 28: 5250–56 (1989)
176. Raina A. K. et al.: Identification of a neuropeptide hormone that regulates sex pheromone production in female moths. Science 244: 796–798 (1989)
177. Ramarao C. S., Burks D. J. and Garbers D. L.: A single mRNA encodes multiple copies of the egg peptide speract. Biochemistry 29: 3383–88 (1990)
178. Randweaver M. et al.: Isolation and characterization of somatolactin, a new protein related to growth hormone and prolactin from Atlantic cod (Gadus morhua) pituitary glands. Biochemistry 30: 1509–15 (1991)
179. Rao K. R. et al.: Primary structure of an analog of crustacean pigment-dispersing hormone from lubber grasshopper Romalea microptera. J. biol. Chem. 262: 2672–75 (1987)
180. Richter K., Egger R. and Kreil G.: Sequence of preprocaerulein cDNAs cloned from skin of Xenopus laevis. A small family of precursors containing one, three, or four copies of the final product. J. biol. Chem. 261: 3676–80 (1986)
181. Richter K., Egger R. and Kreil G.: Molecular cloning of a cDNA encoding the bombesin precursor in skin of Bombina variegata. FEBS Letters 262: 353–355 (1990)
182. Rothman B. S. et al.: Isolation and primary structure of the califins, three biologically active egg-laying hormone-like peptides from the atrial gland of Aplysia californica. J. biol. Chem. 261: 1616–23 (1986)
183. Rouille Y. et al.: The distribution of lysine vasopressin (lysipressin) in placental mammals. A reinvestigation of the hippopotamidae (Hippopotamus amphibius) and tayassuidae (Tayassu angulatus) families. Gen. comp. Endocrinol. 71: 475–483 (1988)
184. Rouille Y. et al.: Hydrins, hydroosmotic neurohypophysial peptides: osmoregulatory adaptation in amphibians through vasotocin precursor processing. Proc. Nat. Acad. Sci. USA 86: 5272–75 (1989)
185. Rouille Y. et al.: Particular processing of pro-opiomelanocortin in Xenopus laevis intermediary pituitary. Sequencing of alpha-melanocyte-stimulating, and beta-melanocyte-stimulating hormones. FEBS Letters 245: 215–218 (1989)
186. Rusakov Y. I. et al.: Isolation, primary structure, and biological and immunological properties of pink and chum salmon insulins. Comp. Biochem. Physiol. Pt. B 95: 477–482 (1990)
187. Ryan R. J. et al.: The glycoprotein hormones: recent studies of structure-function relationships. Faseb J. 2: 2661–69 (1988)
188. Saayman H. S. et al.: Mesotocin and vasotocin, two neurohypophysial hormones in the ostrich, Struthio camelus. Int. J. Peptide Protein Res. 28: 398–402 (1986)
189. Salesse R. et al.: Towards understanding the glycoprotein hormone receptors. Biochimie 73: 109–120 (1991)
190. Saudou F. et al.: Cloning and characterization of Drosophila tyramine receptor. Embo J. 9: 3611–17 (1990)
191. Sawruk E. et al.: Heterogeneity of Drosophila nicotinic acetylcholine receptors: SAD, a novel developmentally regulated α-subunit. Embo J. 9: 2671–77 (1990)
192. Schaller H. C. et al.: The head activator is released from regenerating Hydra bound to a carrier molecule. Embo J. 5: 1821–24 (1986)
193. Schneider L. E. and Taghert P.: Organization and expression of the Drosophila Phe-Met-Arg-Phe-NH2 neuropeptide gene. J. Biol. Chem. 265: 6890–95 (1990)
194. Schoofs L.: Locustatachykinin-I and locustatachykinin-II, two novel insect neuropeptides with homology to peptides of the vertebrate tachykinin family. FEBS Letters 261: 397–401 (1990)
195. Schoofs L. et al.: Isolation, primary structure and synthesis of locustapyrokinin: A myotropic peptide of Locusta migartoria. Gen. comp. Endocrinol. 81: 97–104 (1991)

196. Schuler L. A. et al.: Bovine placental lactogen: Molecular cloning and protein structure. Biochemistry 27: 8443–48 (1988)
197. Schwabe C. and Buellesbach E. E.: Relaxin (Minireview). Comp. Biochem. Physiol. Pt. B 96: 15–21 (1990)
198. Seino S., Steiner D. F. and Bell G. I.: Sequence of a new world primate insulin having low biological potency and immunoreactivity. Proc. Nat. Acad. Sci. USA 84: 7423–27 (1987)
199. Seino S. et al.: Structure of the human insulin receptor gene and characterization of its promoter. Proc. Nat. Acad. Sci. USA 86: 114–118 (1989)
200. Sherwood N. M., Sower S. A. and Marshak D. R.: Primary structure of gonadotropin-releasing hormone from lamprey brain. J. biol. Chem. 261: 4812–19 (1986)
201. Shimasaki S. et al.: Primary structure of the human follistatin precursor and its genomic organization. Proc. Nat. Acad. Sci. USA 85: 4218–22 (1988)
202. Sillard R. et al.: Sheep neuropeptide-Y: A third structural type of a highly conserved peptide. FEBS Letters 258: 263–265 (1989)
203. Simon M. I., Stathmann M. P. and Gautam N.: Diversity of G-proteins in signal transduction. Science 252: 802–808 (1991)
204. Simonson M. S. and Dunn M. J.: Cellular signalling by peptides of the endothelin family. Faseb J. 4: 2989–3000 (1990)
205. Simpson R. J. et al.: Rat epidermal growth factor: complete amino acid sequence – homology with the corresponding murine and human proteins; isolation of a form truncated at both ends with full in vitro biological activity. Eur. J. Biochem. 153: 629–637 (1985)
206. Sopwith A. M., Hales C. N. and Hutton J. C.: Pancreatic B-cells secrete a range of novel peptides besides insulin. Biochim. biophys. Acta 803: 342–345 (1984)
207. Sreedharan S. P. et al.: Cloning and expression of the human vasoactive intestinal peptide receptor. Proc. Nat. Acad. Sci. USA 88: 4986–90 (1991)
208. Stangier J., Hilbich C. and Keller R.: Occurrence of crsutacean cardioactive peptide (CCAP) in the nervous system of an insect, Locusta migratoria. J. comp. Physiol. B 159: 5–11 (1989)
209. Steiner D. F. et al.: Structure and evolution of the insulin gene. Annual Rev. Genetics 19: 463–484 (1985)
210. Stewart A. G. et al.: Human inhibin genes: genomic characterization and sequencing. FEBS Letters 206: 329–334 (1986)
211. Strosberg A. D.: Structure/function relationship of proteins belonging to the family of receptors coupled to GTP-binding proteins (Review). Eur. J. Biochem. 196: 1–10 (1991)
212. Stuart C. A.: Phylogenetic distance from man correlates with immunological cross-reactivity among liver insulin receptors. Comp. Biochem. Physiol. Pt. B 84: 167–172 (1986)
213. Sugino H. et al.: Structural studies on equine glycoprotein hormones. Amino acid sequence of equine chorionic gonadotropin β-subunit. J. biol. Chem. 262: 8603–09 (1987)
214. Takeda J. et al.: Sequence of an intestinal cDNA encoding human gastric inhibitory polypeptide precursor. Proc. Nat. Acad. Sci. USA 84: 7005–08 (1987)
215. Takei Y. et al.: Amino acid sequence and relative biological activity of a natriuretic peptide isolated from eel brain. Biochem. biophys. Res. Commun. 170: 883–891 (1990)
216. Tatemoto K. et al.: Isolation and primary structure of human peptide YY. Biochem. biophys. Res. Commun. 157: 713–717 (1988)
217. Taylor C. W.: The role of G-proteins in transmembrane signalling. Biochem. J. 272: 1–13 (1990)
218. Theibert A. B.: Inositol-1,3,4,5-tetrakisphosphate and inositol hexakisphosphate receptor proteins: Isolation and characterization from rat brain. Proc. Nat. Acad. Sci. USA 88: 3165–69 (1991)
219. Thorndyke M. C. and Goldsworthy G. J. (eds.): Neurohormones in invertebrates. Cambridge Univ. Press, New York 1988
220. Thwaites D. T. et al.: The isolation and chemical characterization of a novel vasoactive intestinal peptide-related peptide from a teleost fish, Gadus morhua. Biochim. biophys. Acta 999: 217–220 (1989)
221. Veenstra J. A.: Isolation and structure of corazin, a cardioactive peptide from the American cockroach. FEBS Letters 250: 231–234 (1989)
222. Venuti J. M. et al.: A myogenic factor from sea urchin embryos capable of programming muscle differentiation in mammalian cells. Proc. Nat. Acad. Sci. USA 88: 6219–23 (1991)
223. Verhaert P. and de Loof A.: Substances resembling peptides of the vertebrate gonadotropin system occur in the central nervous system of Periplaneta americana L. Insect Biochem. 16: 191–197 (1986)
224. Vigna S. R., Thorndyke M. C. and Williams J. A.: Evidence for a common evolutionary origin of brain and pancreas cholecystokinin receptors. Proc. Nat. Acad. Sci. USA 83: 4355–59 (1986)
225. Vlasak R. et al.: Conserved exon-intron organization in two different caerulein precursor genes of Xenopus laevis – Additional detection of an exon potentially coding for a new peptide. Eur. J. Biochem. 169: 53–58 (1987)
226. van der Voorn L. et al.: Characterization of a G-protein β-subunit gene from the nematode Caenorhabditis elegans. J. mol. Biol. 213: 17–26 (1990)
227. Wadsworth S. C., Vincent III W. S. and Bilodeau-Wentworth D.: A Drosophila genomic sequence with homology to human epidermal growth factor receptor. Nature 314: 178–180 (1985)
228. Watahiki M. et al.: Primary structure of chicken pituitary prolactin deduced from the cDNA sequence. Conserved and specific amino acid residues in the domains of prolactins. J. Biol. Chem. 264: 5535–39 (1989)
229. Watts J. D. et al.: Thymosins: Both nuclear and cytoplasmic proteins. Eur. J. Biochem. 192: 643–651 (1990)
230. Weidemann W., Gromoll J. and Keller R.: Cloning and sequence analysis of cDNA for precursor of a crustacean hyperglycemic hormone. FEBS Letters 257: 31–34 (1989)
231. Weiss K. R. et al.: Purification and sequencing of neuropeptides contained in neuron R15 of Aplysia californica. Proc. Nat. Acad. Sci. USA 86: 2913–17 (1989)
232. Wentworth B. M. et al.: Characterization of the two nonallelic genes encoding mouse preproinsulin. J. mol. Evol. 23: 305–312 (1986)

233. Wilson W. E. et al.: Physalaemin-like immunoreactive peptides from rabbit stomach. Int. J. Peptide Protein Res. 28: 58–66 (1986)
234. Wines D. R. et al.: Evolution of the rat kallikrein gene family: Gene conversion leads to functional diversity. J. mol. Evol. 32: 476–492 (1991)
235. Witzemann V. et al.: Primary structure and functional expression of the alpha-subunit, beta-subunit, gamma-subunit, delta-subunit and epsilon-subunit of the acetylcholine receptor from rat muscle. Eur. J. Biochem. 194: 437–448 (1990)
236. Wolf D. L. et al.: Nucleic acid and amino acid sequences of dog beta-LH – Comparison to rat, cow and human beta-LH. Nucleic Acids Res. 15: 10602 (1987)
237. Woodhead A. P. et al.: Primary structure of four allatostatins: neuropeptide inhibitors of juvenile hormone synthesis. Proc. Nat. Acad. Sci. USA 86: 5997–6001 (1989)
238. van Wormhoudt A. and Fouchereau-Peron M.: Partial characterization and amino acid composition of a high molecular weight peptide with salmon calcitonin immunoreactivity in a crustacean, Nephrops norwegicus. Biochem. biophys. Res. Commun. 148: 463–470 (1987)
239. Yamakawa M. et al.: Expression of new members of the prolactin groth hormone gene family in bovine placenta. Isolation and characterization of two prolactin-like cDNA clones. J. Biol. Chem. 265: 8915–20 (1990)
240. Yarden Y. and Ullrich A.: Growth factor receptor tyrosine kinases. Annual Rev. Biochem. 57: 443–478 (1988)
241. Yarfitz S., Provost N. M. and Hurley J. B.: Cloning of a Drosophila melanogaster guanine regulatory protein β-subunit gene and characterization of its expression during development. Proc. Nat. Acad. Sci. USA 85: 7134–38 (1988)
242. Yasuda A. et al.: The complete amino acid sequence of growth hormone from the sea turtle (Chelonia mydas). Gen. comp. Endocrinol. 73: 242–251 (1989)
243. Yasuda A., Kawauchi H. and Parkoff H.: The complete amino acid sequence of prolactin from the sea turtle (Chelonia mydas). Gen. comp. Endocrinol. 80: 363–371 (1990)
244. Yonekura H. et al.: Mosaic evolution of prepropancreatic polypeptide. II. Structural conservation and divergence in pancreatic polypeptide gene. J. Biol. Chem. 263: 2990–97 (1988)
245. Yoshihara A. et al.: Isolation and sequence determination of frog C-type natriuretic peptide. Biochem. biophys. Res. Commun. 173: 591–598 (1990)
246. Yoshino K. I. et al.: A species-specific sperm-activating peptide from the egg jelly of the sea urchin Diadema setosum. Comp. Biochem. Physiol. Pt. B 95: 423–429 (1990)
247. Zhou Q. et al.: Cloning and expression of human and rat D1 dopamine receptors. Nature 347: 76–80 (1990)
248. Special Topic: Recombinant DNA to study neuropeptide processing. Annual Rev. Physiol. 50: 305–376 (1988)
249. Symposium: Evolution of hormone diversity. Amer. Zool. 26: 923–1054 (1986)

9 Toxic Proteins and Peptides

Many animals produce poisonous substances which, even at a relatively low dose, are deleterious or lethal to other organisms, including man. They may serve on the one hand to disable or kill prey, or on the other hand as a protection or defence against predators. These substances are mostly produced by special glands, and both poison reservoirs and specialized structures for delivering the secretion, e.g. stings or modified mouthparts or fangs, are usually also present. In accordance with the aims of this book, we will deal with the comparative biochemistry of the animal toxins as natural products with special structures and functions; a discussion of the mechanisms of their pharmacological effects and the highly complicated organ structures and behavioural patterns associated with natural poisons may be found in numerous monographs [5, 20, 36, 68, 69]. Toxic proteins and oligopeptides are known from many very different animals (Table 9.1); there is, however, a host of low molecular weight compounds with considerable toxicity which will be considered along with other non-toxic products of secondary metabolism in Chapter 19.

Animal protein and peptide toxins are found with very different **structures and activities**. They may, for example, stimulate or paralyse respiration, the blood circulation or skeletal muscles; they may lead to cell lysis or tissue necrosis; or they may have neurotoxic effects, for example, by interacting in different ways with the pre- or postsynaptic membranes. Toxins from different animal groups but with very similar effects in general show little structural similarity. Particular toxin characteristics have apparently arisen several times quite independently in the evolution of the proteins with little structural convergence.

Table 9.1. The occurrence of toxic proteins and peptides in the animal kingdom [28]

Porifera	In *Suberites domuncula*
Cnidaria	In the nettle poison of all species
Gastropoda	In the "poison tongues" (Toxoglossa) of the families Conidae, Turridae and Terebridae
Cephalopoda	In the rear salivary glands of, e.g., *Octopus* and *Sepia* species
Arachnida	Scorpions: in all species Araneida: probably in all species Acarina: in the eggs of individual tick species (Ixodidae)
Chilopoda	In several genera of the Scolopendromorpha *(Scolopendria, Otostigmus, Cryptops, Scolopocryptops)*
Hymenoptera	In the parasitic wasps (Ichneumonidae), spider wasps (Pompilidae) and sand wasps (Sphecidae); bees and bumblebees (Apidae); wasps and hornets (Vespidae); and amongst the ants (Formicidae) in the primitive subfamilies (Ponerinae, Myrnicinae, Pseudomyrmicinae)
Lepidoptera	In the poisonous hairs of some caterpillars
Coleoptera	In the haemolymph of *Leptinotarsa* sp.
Chondrichthyes	In the stingrays (Dasyatidae)
Teleostei	In eels and morays (Anguillidae); the greater weevers (Trachinidae); the scorpion fish (Scorpaenidae); and many catfish species (Siluroidea)
Amphibia (Anura)	In the toads and related species (Discoglossidae); the tree frogs (Hylidae); and the real frogs (Ranidae)
Reptilia	In many snakes (Ophidia) (see Table 9.2); and lizards of the genus *Heloderma* ("venomous lizards", "gila monster")
Mammalia	In the duck-billed platypus *Ornithorhynchus anatinus*

Toxic secretions are generally complex mixtures of several toxic proteins or polypeptides, with different structures and activities, together with non-toxic proteins, enzymes, oligopeptides, free amino acids, amines and phospholipids. In many cases, the accompanying substances are synergistic to or potentiate the pharmacological effects, such that the overall activity of the toxin is more than the sum of the individual components. For example, the hyaluronidase found in the poisonous secretions of snakes, fish, bees, spiders and scorpions promotes the spread of the toxin in the tissues of the victim. In addition to serotonin and other biological amines, the main cause of the pain associated with the bite or sting of poisonous animals is bradykinin or another kinin; either these are components of the poison, e.g. in Cnidaria and Vespidae, or, as in many snakes, they are formed from the plasma kininogen of the victim by kininogenase activity in the poison. The lethal effects of snake bites involve various enzymes as well as the toxins themselves.

9.1 Snake Venoms

Only about one-third of the known 2500 snake species are poisonous (Table 9.2); 7 of the 12 snake families are composed only of non-poisonous species. In the family Colubridae (adders), only about 400 of the 1500 species are poisonous; in these species, the fangs are usually located at the back of the mouth cavity (opisthoglyphous snakes) and only come into effect during swallowing. Only the few species with fangs further forward are dangerous for humans. In the four families of effectively **poisonous snakes**, the poison apparatus is highly developed: the fangs are located right at the front of the mouth cavity (proteriglyphous snakes) and the venom either flows along a groove in the fang, as in the elapids and sea snakes, or is injected directly into the victim through the hollow fangs, as in the vipers and pit vipers. The venom apparatus of the snakes serves primarily to immobilize the prey during the long swallowing process. Death of the prey mostly involves disruption of the circulation and respiration through the effect of nerve and heart toxins; in addition, some snake venoms result in tissue destruction through necrosis or prevention of blood clotting.

Table 9.2. The main species of poisonous snakes [28]

Colubridae (adders) About 400 out of 1500 species are poisonous, e.g.:
Boiga dendrophila (tree adder) South Asia
Dispholidus typus (boomslang) South Africa
Psammophis sp. (sand adder) Africa, Near East

Elapidae (cobras and mambas) All are poisonous, 41 genera with 180 species and 300 subspecies, e.g.:
Acanthopis antarticus (death adder) Australia
Bungarus fasciatus (banded krait) Southeast Asia and Sundain Islands
Dendroaspis sp. (mambas) Africa;
e.g. *D. polylepis* (black mamba)
Hemachatus haemachatus (ringhals) South Africa
Micrurus sp. (coral snakes) mostly Central America
Naja (cobras);
e.g. *N. naja* (common cobra) Asia
N. najed (Egyptian cobra) Africa
Ophiophagus hannah (king cobra) India
Oxyuranus scutellatus (taipan) Australia

Hydrophiidae (sea snakes) All are poisonous, 16 genera with 50 species, e.g.:
Enhydrina schistosa coast of Southeast Asia
Hydrophis sp. (sea snake) coasts of South Asia and the Indo-Australian archipelago
Laticauda sp. (laticaudine sea snakes) from the Bengal coast to the Friendly Islands
Pelamis platurus (yellow-bellied sea snake) Indian and Pacific Oceans

Viperidae (vipers) All are poisonous, 10 genera with 60 species and 110 subspecies, e.g.:
Bitis sp. (puff adders) Africa and the Arabian peninsula
Cerastes cerastes (horned viper) Africa
Echis carinatus (saw-scaled viper) Africa, Asia
Vipera sp. (vipers) about 20 species in Europe, Asia and Africa;
V. berus (common viper) Europe and north Asia
V. ammodytes (sand viper) southeast Europe and west Asia
V. russelli (Russel's viper) south and southeast Asia

Crotalidae (pit vipers) All are poisonous, 6 genera with 130 species and 210 subspecies, e.g.:
Agkistrodon sp. (copperheads and cotton mouths) Asia, America, one species in Europe
Bothrops sp. (American lance-head viper) Central and South America
Crotalus sp. (real rattlesnakes) America
Lachesis mutus (bushmaster) Central and South America
Sistrurus sp. (pygmy rattlesnakes) North to Central America
Trimeresurus sp. (Asian lance-head vipers) south and Southeast Asia

9.1.1 The Toxins of Snake Venoms

Most snake toxins are classified according to their pharmacological effects as neurotoxins or membrane toxins; there are also venoms with an unknown mechanism of operation. Many **neurotoxins** act postsynaptically; they are curaromimetic in that they bind to the nicotinic acetylcho-

line receptor of the skeletal muscles, as in the case of D-tubocurarine. With dissociation constants of 10^{-10} to 10^{-11} mol/l, the neurotoxins bind more tightly than curare itself and are therefore useful tools for the isolation of the receptors. Other neurotoxins act presynaptically, disturbing the release of acetylcholine; their toxicity is greater than that of the postsynaptic variety. Poisonous effects on the heart (cardiotoxic effects) and the lysis of erythrocytes or other cells (cytotoxic effects) are brought about by further toxins that apparently act primarily on the cell membrane and are, therefore, known as **membrane toxins**. The haemolytic and toxic effects of these venom components are markedly increased in the presence of phospholipase. The snake venom toxins are referred to either by the initials of the snake's name with additional numbers or letters (e.g. "Lc-a" from *Laticauda colubrina*), or by colourful names such as "cobrotoxin" (from *Naja naja atra*), "erabutoxin" from *Laticauda semifasciata*) or "bungarotoxin" (from *Bungarus multicinctus*). Approximately 150 snake venom toxins have so far been sequenced [9, 13] and in several cases the spatial structure has also been determined [11]. Two classes are distinguishable on the basis of structure and activity spectrum, viz. the postsynaptic-acting neurotoxins and membrane toxins and the presynaptic neurotoxins.

Numerous **postsynaptic-acting neurotoxins** are known from the cobras and the sea snakes, and the **membrane toxins** are known mostly from the cobras. They are all polypeptides of 60–74 amino acids with four or five disulphide bridges, and apparently all arose from a common ancestral form with about 60 amino acids [9, 13, 51]. Several venom components belonging to this protein family show only very limited toxicity [26]. The neurotoxins may be divided into "short", with 60–52 amino acids and four disulphide bridges, and "long", with 71–74 amino acids and five disulphide bridges. However, several neurotoxins are found in *Laticauda* species with 66 amino acids and five disulphide bridges, or with 69 amino acids and only four disulphide bridges [35, 36, 48]. The membrane toxins all have about 60 amino acids. The sequences of the postsynaptic-acting neurotoxins and the membrane toxins contain ten invariant positions; considering the neurotoxins alone, there are in fact 22 conserved positions, of which only four or five are essential for the toxicity. one characteristic of these sequences is that two, or even three, identical amino acids are often found next to each other; five to ten such homodipeptides are consistently found in the short toxins (Fig. 9.1) and two to four are found the long variety. The three-dimensional structure of the short erabutoxin b from *Laticauda semifasciata* is characterized by five antiparallel β strands joined by loops [11]. Relatively little is known about the biosynthesis of the snake venom toxins; however, the precursor of the erabutoxin a from the poison gland of *Laticauda fasciata* has been sequenced via the cDNA and consists only of a 21-amino-acid signal sequence and the sequence of the mature toxin [66]. Most of the postsynaptic-acting neurotoxins bind specifically to the nicotinic acetylcholine receptor of muscle; only the recently discovered k-bungarotoxin from *Bungarus multicinctus* and the pseudonajatoxin b (with 71 amino acids) from the Australian species *Pseudonaja textilis* attack the muscarinic receptors of the sympathetic ganglia. Pseudonajatoxin a acts postsynaptically on muscle receptors and has 117 amino acids, including 14 cysteine residues, and is thus something of an exception [70].

There is no doubt that the postsynaptic-acting neurotoxins, membrane toxins and certain non-toxic proteins of snake venoms have a common origin. However, nothing can be said for certain about the phylogenetic relationships within this protein family. Genealogical trees constructed by the matrix method show separate lines of evolution for the short neurotoxins, the long neurotoxins and the membrane toxins, but leave open the question of the root of the tree and the earliest ancestral form. This is mainly due to the fact that numerous gene duplications hinder the distinction of orthologous and paralogous sequences. It is already clear from the presence of homologous toxins in the same individual that gene duplications were of frequent occurrence in the **evolution of the snake venoms**.

The second class of snake venom toxins, the **presynaptic-acting neurotoxins**, is a collection of proteins of very different structure which all belong to the **protein family of phospholipase A_2** and are homologous to the corresponding enzyme proteins. Phospholipases A_2 cleave the 2-acyl groups from various phospholipids; they have a greater activity with aggregated than with dis-

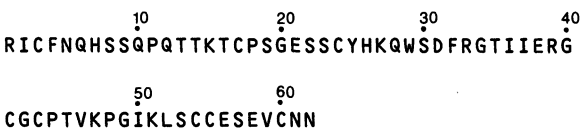


Fig. 9.1. Amino acid sequence of erabutoxin b from the venom of the sea snake *Laticauda semifasciata* [66]

solved substrates, and are Ca^{2+} dependent. Many snake venoms contain several isoenzymes with widely differing isoelectric points (pI 4–11) and different quaternary structures; only the basic isoenzymes are toxic. They are presynaptically neurotoxic on skeletal muscles through the inhibition of acetylcholine release, and are also often myonecrotic, destroying muscle fibres and causing the appearance of myoglobin in the urine [30]; they are amongst the most dangerous animal toxins. The more than 40 well-known phospholipase A_2 snake venoms are polypeptides of 108–140 amino acids with 4, 6 or 7 disulphide bridges; they can occur in solution as monomers, homodimers, heterodimers or even larger aggregates [38, 42]. They are homologous amongst themselves and with the enzymes from the mammalian pancreas; the sequence agreement is particularly high in the region of the active centre and the Ca^{2+}-binding site. The enzymes from bovine pancreas and the venom of the rattlesnake *Crotalus atrox* are very similar in their spatial structure. The Ca^{2+}-binding 49-Asp as well as 28-Tyr, 30-Gly and 32-Gly are all involved in the catalytic mechanism; the venom of the copperhead snake *Agkistrodon piscivorus*, however, contains in addition to the 49-Asp enzyme a further phospholipase A_2 with 49-Lys, which has an altered catalytic mechanism but is otherwise homologous to the other isoenzymes. Both the pancreas enzymes and the venom enzymes of the elapids and the sea snakes contain a characteristic disulphide bridge, 11-Cys-69-Cys; this is missing in the venom enzymes of the Crotalidae and Viperidae, which instead have a disulphide bridge associated with an extended C-terminus [42].

Snake venoms often contain several toxic phospholipases A_2, e.g. the sand viper *Vipera ammodytes* has three monomeric enzymes referred to as ammodytoxins A–C [53]. In addition, many venoms contain complex toxins composed of different, but homologous, subunits. Thus, the "vipoxin" from *V. ammodytes* consists of a toxic, basic phospholipase A_2 and a non-toxic acidic phospholipase inhibitor which agree in 62% of their 122 amino acids [41]. The "crotoxin" of the North American rattlesnake *Crotalus durissimus* was discovered in 1938 as the first presynaptic-acting neurotoxin, but its structure has only recently been completely determined. In this case, there is also a basic subunit, CB, of 122 amino acids which is enzymically active and toxic; the acidic subunit, however, consists of three peptides of 38, 35 and 14 amino acids linked by a total of seven disulphide bridges. These result from the partial hydrolysis of a 122-amino-acid precursor, CA, which is homologous to CB and even has an identical signal sequence of 16 amino acids [2, 7]. The most complicated structure and with the highest known toxicity (1 μg/kg mouse i.v.) is found in the "textilotoxin" of the Australian common brown snake *Pseudonaja textilis*; this toxin is made up of three peptide chains, A–C, and a homodimer DD [52]. The venom of the North American rattlesnake contains a further toxin, "crotamin", which has as yet undefined effects on the musculature. This is a basic polypeptide of 42 amino acids with three disulphide bridges, and has no sequence similarity to any other known toxin. Other *Crotalus* species have similar toxins [36, 63]. The "taipoxin" of the Australian "taipan" *Oxyuranus scutellatur* is made up of three subunits which differ greatly from each other but are all homologous with other phospholipases A_2. The basic α subunit, which is the only one showing any sort of enzyme activity, is weakly toxic in the purified form; the combination of the three subunits is 500-fold more toxic, with an LD_{50} of 2.1 μg/kg [37].

9.1.2 Enzymes and Other Protein Components

The poisonous secretions of the snakes always contain a large number of protein components with different pharmacological effects or biological activities; often more than 25 protein fractions may be distinguished by electrophoresis or chromatography. Venom secretions with similar toxicities from different populations, or even from different individuals, may show large differences in electrophoretic pattern and enzyme activity. The effects of enzyme components include local capillary damage and tissue necrosis (from phospholipases, proteases, hyaluronidases), inhibition or promotion of blood clotting (from specific proteases, phospholipases), reduction in blood pressure, and pain (from kinin-releasing enzymes).

Most of the **phospholipases** from snake venom cleave 2-acyl residues from different phospholipids, i.e. they are of type A_2. They have haemolytic effects and several of them have synergistic effects on toxin activity. A phospholipase B found in the venom of the elapid *Pseudechis australus* attacks the 1-acyl group in lysophosphatide, and several phospholipases A_2 also have a similar, albeit weak, activity. Although the phospholipases A_2 are acidic or basic, according to their

amino acid composition, their activity pH optima lie between 7.5 and 8.5 [36]. It has already been mentioned that the basic phospholipases A_2, in contrast to the acidic variety, are, without exception, toxic, and that during snake venom evolution they formed the origin of highly effective toxins at the expense of their enzyme activity.

Snake venoms may contain numerous **proteases** of varying specificity. For example, five clotting-inhibitory (haemorrhagic) Zn^{2+} proteases have been detected in the rattlesnake *Crotalus*, together with a non-haemorrhagic collagenase, two kallikrein-like arginine esterases, four other clotting-inhibitory proteases, four anticomplement factors, plus two fibrinolytic and three caseinolytic activities, i.e. a total of 21 protease activities, several of which, however, are probably identical [6]. Zn^{2+} proteases with broader specificity are widely found in viperid and crotalid venoms but are absent from those of the elapids and sea snakes, which often contain oligopeptidases. Highly specific proteases, however, make an important contribution to the effectiveness of snake venoms. Thus, for example, clotting-inhibitory or promotory enzymes result in internal bleeding (haemorrhages) or the formation of intravasal blood clots, and thereby amplify the lethal effects of the snake bite. Haemorrhagic Zn^{2+}-proteases are found, above all, in the venom of the viperids and crotalids [6], but one such enzyme is also present in the elapid *Naja nigicollis*. This enzyme destroys erythrocytes and punctures capillary walls by hydrolysing collagen IV and other protein components of the endothelium basal membrane [65].

In addition to haemorrhagic metalloenzymes, venom secretions may at the same time contain thrombin-like serine proteases. These enzymes, however, attack only the Aα chain or, less often, the Bβ chain of fibrinogen and in no case form insoluble fibrin [17, 28]. The best-known of these enzymes, batroxobin from *Bothrops atrox*, shows only 31 % sequence agreement with human thrombin and can only cleave the Aα chain of fibrinogen. The batroxobin gene is much more similar in its exon/intron organization to the genes of the kallikreins and trypsins than to the thrombin gene. Thus, batroxobin apparently originates from glandular kallikreins [28]. There are further venom components which influence blood clotting. The echistatin of the viper *Echis carinatus* (49 amino acids) or arietin from *Bitis arietans* are antagonists of the fibrinogen receptor and inhibit clotting by preventing platelet aggregation [25]. The venom of the southern copperhead snake *Agkistrodon contortrix* contains a serine protease which activates plasma protein C; this in turn hydrolyses the factors Va and VIIIa and thereby inhibits clotting [44]. The serine protease cerastobin from the horned viper *Cerastes cerastes* activates factor X, and thus promotes clotting [17].

A large proportion of the activity of venom secretions is due to **kinins and kinin-releasing enzymes**. The peptide bradykinin, and its effect of reducing the blood pressure, was in fact first discovered in 1949 in investigations of the action of the venom from *Bothrops jararaca* on mammalian blood plasma. All viperid and crotalid venoms contain enzymes which release bradykinin or kallidin from the precursor kininogen in mammalian plasma (see Fig. 8.5, p. 302). Such kininogenases are similar to the mammalian kallikrein, are unusually heat resistant, and can be separated from the arginine esterases and thrombin-like enzymes [36, 50]. The biological significance of other enzymes and protein components of the snake venoms is not completely clear. **Acetylcholine esterases** are widely found in the venoms of the elapids and hydrophiids but not in those of the viperids and crotalids [36]. The **L-amino acid oxidases** of snake venoms differ from the corresponding enzymes of mammalian tissues in their coenzyme (FAD instead of FMN) and in the more than 100-fold higher activity. These enzymes are found at significantly higher concentration in the viperids and crotalids than in the elapids and impart the yellow colour to the venom. White venom without the FAD enzyme is found in the hydrophiids and several species of the viperids and crotalids. The enzymes of the different species appear to be very similar in their catalytic and physicochemical properties [36, 71].

Protease inhibitors have been isolated from various snake venoms and in some cases have been sequenced; they apparently all belong to the pancreas-trypsin inhibitor (Kunitz) family but are in part specific for chymotrypsin [53]. Amongst the most surprising components of snake venoms are the **nerve growth factors** (NGFs). These are present at low concentrations in many vertebrate tissues, but occur at high concentration only in the submaxillary glands of several mouse species and in the venom glands of most poisonous snakes. NGFs promote the growth and differentiation of neurons, particularly those of the sympathetic nervous system; their role in snake venoms is something of a mystery. Venom secretions contain only the bioactive β chain; they lack the α and γ chains found in the high molecular weight NGF complexes of mammalian tissues. The

amino acid sequences of the β NGFs of *Naja naja* and the mouse agree in about 65 % of positions [27]. The first **venom lectin** (thrombolectin) was found in 1980 in the "fer de lance" *Bothrops atrox*; this is a β-galactoside-specific dimeric lectin of 28 kDa which is similar in specificity, quaternary structure and molecular size to other vertebrate lectins but is Ca^{2+}-dependent. Similar lectins have since been found in the venoms of many other snake species [22, 39].

9.2 Proteotoxins of Other Vertebrates

In the **fish**, one can differentiate between about 500 passively poisonous species, the consumption of which leads to poisoning symptoms, and about 250 actively poisonous species, which possess specialized mechanisms for delivering the poison. In both cases, the poisons serve as a defence mechanism or warning factor against potential predators. The passive toxicity of fish is mainly caused by low molecular weight substances, which are dealt with in Chapter 19; the blood of eels and morays, however, contains a protein which haemolyses the erythrocytes of warm-blooded animals and is toxic when eaten. The actively poisonous fish probably all contain toxic peptides or proteins, the chemical structures of which are, however, largely undetermined. For example, the poisonous secretion of the sting ray (Dasyatidae) contains various enzymes as well as several toxic proteins of about 100 kDa [20]. In order to scare off sharks, the flatfish *Pardachirus marmoratus* produces in its skin glands steroid glycosides and three related peptides of 33 amino acids, the "pardaxins", which are similar to the melittin of bee venom in their postsynaptic neurotoxic activity but not in their sequences [59].

The skin secretions of the **amphibians** also serve to scare away predators. In addition to toxic amines, alkaloids and steroids, they contain a wide spectrum of pharmacologically effective oligopeptides (Fig. 9.2). Many of them are very similar in structure and effect to peptide hormones from the mammalian gut and brain. They have, for example, neurotoxic, haemolytic, cytotoxic or myotropic effects; related, non-bioactive peptides may also be present [4, 16, 20, 47, 60]. The decapeptide caerulein was discovered in the skin of *Hyla cearuleus* but is also found in other anurans. It belongs to the same family as the peptide hormones cholecystokinin (CKK) and gastrin and was mentioned in Chapter 8. The skin of various frog species contains large amounts of kinins with unusual structures (Fig. 9.3).

Apart from the snakes, there are only two other poisonous species of reptile, the **venomous lizards** *Heloderma suspectum* ("gila monster") and *H. horridum*. These animals possess poison glands in the lower jaw, the secretions from which moisten the teeth in both jaws. Four different proteotoxins of 26–37 kDa have so far been identified in the venom, and have been sequenced; these are referred to as gilatoxin, horridumtoxin and helothermin, and apparently do not constitute a uniform protein family [46]. In addition, several proteins of 35–39 amino acids, known as helodermins, helospectins and exendin-3, have been described and these belong to the family of the peptide hormones glucagon, secre-

"Sauvagin" from Phyllomedusa sauvagei	pEGPPISIDLSLELLRKMIEIEKQEKEKQQAANNRLLLDTI
"Bombesin" from Bombina bombina and B.variegata	pEQRLGNQWAVGHLM-amide
"Alytesin" from Alytes obstetricans	pEGRLGTQWAVGHLM
"Physalaemin" from Physalaemus fuscumaculatus	pEADPNKFYGLM
"Uperolein" from Uperoleia rugosa and U.marmorata	pEPDPNAFYGLM
physalaemin-like peptide and	pEDPDRFIGLM
"Phyllomedusin" from Phyllomedusa bicolor	pENPNRFIGLM
"Caerulein" from Hyla caerulea	pEQDYTGWMDF
"Litorin" from Litoria (Hyla) aurea	pEQWAVGHFM
"Xenopsin" from Xenopus laevis	pEGKRPWIL
"Bombinin" from Bombina sp.	GIGALSAKGALKGLAKGLAZHFAN
"Bo-III" from Bombina orientalis	RPPGFSPIRGKFH
"Kassinin" from Kassina senegalensis	DFVKSDQFVGLM
"Crinia-Angiotensin" from Crinia georgiana	AGPDRIYVHPF

Fig. 9.2. The peptides of amphibian skin [4, 16, 20, 47, 60]. *pE*, Pyroglutamic acid

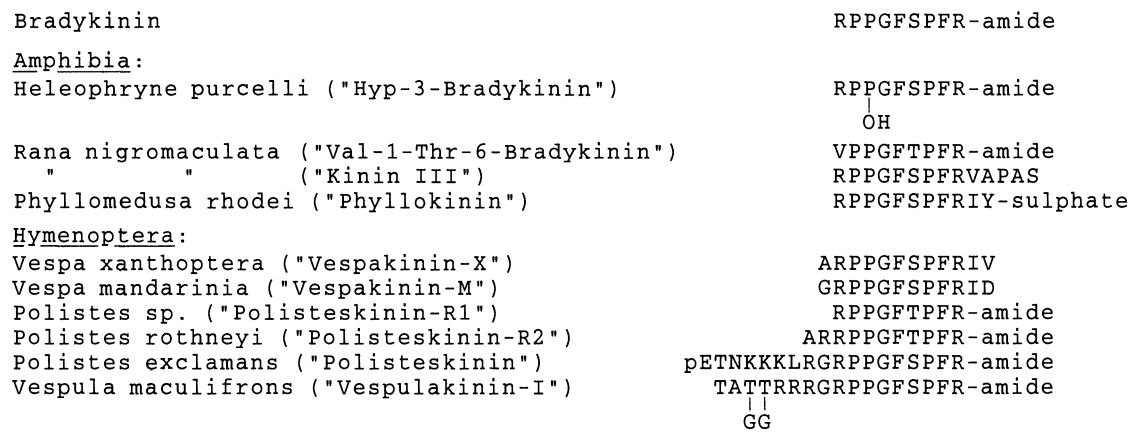


Fig. 9.3. The kinins of amphibian skin and hymenopteran venoms [55]. *P-OH*, Hydroxyproline; *T-G*, glycosylthreonine; *pE*, pyroglutamic acid

tin and vasoactive intestinal peptide (VIP). They are not only structurally similar to the other members of this family but also have similar stimulatory effects on CAMP concentration and on the secretory activity of pancreas cells [15]. The gila-monster venom also contains at least five variants of phospholipase A_2, which surprisingly show much more similarity in their sequence of 141–143 amino acids to the bee venom enzyme than to the enzyme from snake venom or mammalian pancreas [72]. The only **poisonous mammalian species** is the Australian duck-billed platypus *Ornithorhyncus anatinus*, the males of which possess a spur with a poison gland on their rear legs. The only thing known about this venom is that it is proteinaceous.

9.3 Arthropod Venoms

9.3.1 The Venoms of Scorpions and Other Arachnids

All the approximately 650 scorpion species are poisonous and possess a poison apparatus consisting of a sting at the end of the abdomen and a pair of venom glands. The secreted poison contains 15–20 electrophoretically distinct protein components, including proteotoxins and enzymes such as phospholipases A_2, hyaluronidases, acetylcholine esterases, proteases and acidic phosphatases. The majority of the **scorpion toxins** are basic proteins with 60–70 amino acids and four disulphide bridges. In the Central Asian species *Buthus eupeus* and in the African species *Androctonus mauretanicus*, however, there are also shorter insect-toxic proteins with only 35–37 amino acids and three disulphide bridges [43]. The sequences are known of numerous toxins or their precursors from about six scorpion species; they show a consistently high degree of sequence similarity [8, 20, 34, 40]. Each species possesses several (often more than ten) different proteotoxins which may be classified by the specificity of their effects into mammalian, insect or crustacean toxins [5, 20]. The venom apparatus of the scorpions is directed, on the one hand, to defence against mammalian predators and, on the other hand, to the capture of prey, i.e. the killing of insects and other arthropods.

In mammals, the scorpion sting excites the autonomous nervous system, accelerates respiration and the circulation, and produces muscle cramps. All these **effects** involve highly specific interactions with sodium channels of the excitable membranes; thus, scorpion toxins, like other neurotoxins, are exploited as molecular tools for the investigation of membrane components. On the basis of differences in binding to synaptosomes from rat brain, scorpion toxins can be divided into two classes: the α toxins bind in a potential-dependent manner to a certain site (site 3) of the sodium channel and retard its inactivation; the β toxins bind in a voltage-independent manner to a different site (site 4) and influence the activation of the sodium channel. In the scorpions of the Old World (Africa and Asia) one finds predominantly α toxins, whereas American scorpion species possess mainly β toxins; both toxin types are found, for example, in the Tityninae. The most toxic scorpion venom, with an LD_{50} of 0.18 μg/kg

in the mouse, is the α-type toxin Aa-II of the North African species *Androctonus australis*. Most insect-specific toxins differ characteristically from the mammal-specific toxins in their amino acid sequences, spatial structures and modes of action [18]; some insect toxins, however, are quite similar to the mammal-specific ones [14, 40]. Amongst the insect-specific toxins, one type blocks action potentials and thus produces a flaccid paralysis, and the other type induces repetitive firing of the motor nerves and therefore produces a spastic paralysis [14].

The approximately 25 000 **spider species** all possess poison glands on their mouth appendages (chelicera) and these glands serve in the catching of prey. The crude poison of the spiders also contains numerous components, including biogenic amines, polyamines, enzymes and toxic proteins [5, 20, 54, 62]. Only a few species are dangerous for humans, the most dangerous being *Atrax robustus*, which is found only in metropolitan Sydney and has led to 13 deaths in the last 60 years. The venom of the funnel-web spider *Agelenopsis aperta* contains both postsynaptic-acting acylpolyamines (α-agatoxins) and two classes of peptides which presynaptically affect the glutamatergic neuromuscular junction of insects. Of these latter, the μ-agatoxins (36–38 amino acids) have an activating effect, whilst the ω-agatoxins (≤ 66 amino acids) are inhibitory [1]. Further toxic peptides, with 36–77 amino acids, from other spider species have been sequenced [10, 29, 54, 64]; their phylogenetic relationships to each other are not yet known.

9.3.2 Insect Venoms

Next to those of the scorpions, the best-investigated arthropod venoms are those of the honeybee *Apis mellifera* and several wasp species. Like all poisonous secretions examined in any detail, the **venoms of the Hymenoptera** are found to be complex mixtures of substances. Amongst the low molecular weight compounds are biogenic amines (histamine, 5-hydroxytryptamine, dopamine, adrenalin, noradrenalin, acetylcholine), phospholipids, sugars and free amino acids [5, 55]. Electrophoresis distinguishes 25–40 polypeptides in the toxic secretions of different Hymenoptera. Various enzymes are found amongst the proteins. The phospholipase A_2 which makes up 12 % of the dry weight of bee venom is a basic polypeptide of 129 amino acids of known sequence. Wasp venoms also contain phospholipases of the types A_1 and B [33, 55]. The hyaluronidase of bee and other hymenopteran venoms promotes the spread of the toxic substances in the tissues [32, 55]. The hymenoptera-venoms also include esterases, glycosidases and protease inhibitors [55]. Enzyme proteins constitute approximately 13–15 % of the dry weight of worker bee venom, whereas that of the queen bee contains only traces of phospholipase and hyaluronidase.

The most investigated components of hymenoptera-venoms are the bioactive peptides, which, for example, constitute 50–60 % of the dry weight of bee venom. On the basis of their pharmacological effects, one can distinguish above all haemolytic, neurotoxic and kinin-like peptides; these are distributed quite irregularly amongst the different hymenopteran groups. The haemolytic melittin (Fig. 9.4a) predominates in bee venom; in contrast, there are no haemolytic proteins in wasp venom but various kinins are found, which are lacking in the bee (Fig. 9.3). **Melittin** is a peptide of 26 amino acids with an amidated C-terminus and an N-terminus which is formylated in about 10 % of the molecules. On the extended molecule, a non-polar N-terminal section of 20 amino acids and a very basic C-terminal section can be differentiated (Fig. 9.4a). This structure relates to the unusually strong interaction with lipid double membranes: melittin has a haemolytic effect and reduces the surface tension of water to a much greater extent than do all other peptides. Solutions of low ionic strength contain the monomer, whereas in higher ionic strength solutions and in the poison reservoir the venom is found as a tetramer with the form of a double cross [67]. Species related to the honey bee have somewhat modified sequences (Fig. 9.4a). The sequence of prepro-melittin has been determined by the analysis of mRNA from honey-bee venom glands: it consists of 70 amino acids, of which 1–21 form the signal sequence and 22–43 form the pro-sequence. Mature melittin consists of amino acids 44–70, with the terminal glycine taking part in the amidation of the C-terminus [73]. The most important neurotoxic component of bee venom is **apamin**, a basic peptide of only 18 amino acids (Fig. 9.4a). Melittin itself has a weak degranulating and histamine-releasing activity on mammalian mast cells; a much higher activity is shown by the mast-cell degranulating peptide (MCD), an extremely basic peptide (pI = 12) with 22 amino acids, of which none is acidic and no fewer than 7 are basic (Fig. 9.4a). Finally, bee venom contains small

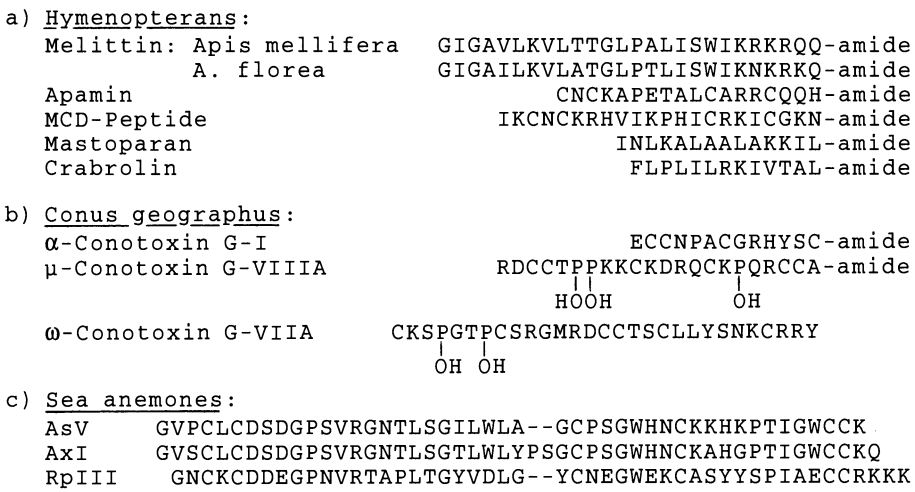


Fig. 9.4a–c. The toxic peptides of various invertebrates [3, 45, 51, 55]. *P-OH*, Hydroxyproline

amounts of other peptides with limited pharmacological effects, e.g. secapin (24 amino acids) and tertiapin (20 amino acids) [12, 55]. The pain and inflammation resulting from a bee sting are caused by the joint action of several components. MCD releases histamine from the mast cells, and phospholipase A_2 induces cytolysis, which is promoted synergistically by melittin, and the release of arachidonic acid, which as a precursor of prostaglandin and leukotriene results in the production of further bioactive compounds [3].

The **kinins** typical of **wasp venom and hornet venom** always contain the sequence of bradykinin, either unmodified or with a serine/threonine substitution, but may have a much greater activity. The kinins of *Vespula* sp. surprisingly have carbohydrate chains attached to two threonine residues [55]. Compared with those of the mammals, these kinins usually have an extended N-terminus whereas amphibian kinins have additional C-terminal amino acids (Fig. 9.3). The venom of the wasps and hornets contains species-specific mast-cell degranulating tetradecapeptide-amides, the **mastoparans**, which amongst other effects increase the activity of phospholipase A_2 (Fig. 9.4a). Crabrolin (13 amino acids), from the venom of the hornet *Vespa crabro* (Fig. 9.4a), and the five bombolitins (17 amino acids), of the bumblebee *Megabombus pennsylvanicus*, are similar in activity and basic structure to the mastoparans but have different sequences [3, 24]. In wasps and hornets there are other toxic proteins whose structures have not yet been determined. A neurotoxic protein of 20 kDa (mandarotoxin) is the most active poisonous component of the giant hornet *Vespa mandarinia*, from which a single sting can be lethal for humans; hornetin from *V. flavicornis* is a basic protein of 32 kDa with a low LD_{50} in mice of 0.42 μg/kg [23]. Toxic proteins are also produced by ichneumons (Terebrantia), spider wasps (Pompilidae) and sand wasps (Sphecidae), but have been just as little investigated as the corresponding toxins from other insect orders (Table 9.1).

9.4 Proteotoxins of Other Invertebrates

Many of the **prosobranch gastropods** from the suborder of the Toxoglossa, especially the family Conidae, are poisonous. The approximately 300 species of the genus *Conus* are all specialists in hunting fish, molluscs and polychaetes; their complicated device for hunting prey can shoot poison darts several millimetres in length. In order to paralyse their fish prey, the species *Conus geographus* and *C. magus* have several classes of toxic peptides (**conotoxins**) which block the consecutive steps in the action chain of neuromuscular transduction [19, 51]. Each class includes several peptides of similar sequence; the sequences of the different classes show little similarity (Fig. 9.4b). The ω-conotoxins (25–29 amino acids) affect presynaptic voltage-dependent Ca^{2+} channels and inhibit acetylcholine release; the α-conotoxins (14–16 amino acids) have postsynaptic effects on acetylcholine receptors; the μ-conotoxins (22 amino acids) block sodium channels and reduce action potentials. Conotoxin GS of *C. geographus* (34 amino acids) is similar to the μ type in activity but not very similar in sequence

[74]. A toxin isolated from the snail-hunting species *C. textile* corresponds to the ω-conotoxins of the piscivore species in the position of the cysteine residues but is only effective against snails and crustaceans [21].

The proboscis of many **nemertines** is equipped with poison glands. In the hoplonemertines, where the proboscis bears a stiletto, the glands produce low molecular weight toxins. Only in the case of the North Atlantic heteronemertine *Cerebratulus lacteus* have toxic proteins been examined in detail and partly sequenced. Two groups of small toxic proteins are found: the three A toxins consist of 93–98 amino acids and are strongly cytolytic; toxin A-III, for example, is fourfold more active than melittin in the haemolysis of human blood. The four B toxins are significantly smaller (49–56 amino acids) and have neurotoxin activity, although only against crustacean neurons [31]. The poison in the nematocysts of some **Cnidaria** is extraordinarily toxic with LD_{50} values against crustaceans of 1–2 μg/kg. Certain species, e.g. the sea wasp *Chironex fleckeri* (Cubozoa), can be dangerous for humans. The venom is again a mixture of various pharmacologically or enzymatically active components. The acute pain arising from contact with the sea nettle *Chrysaora* is produced by histamine, serotonin, prostaglandins E and F, and kinin-like peptides [20]. The toxins isolated from the nematocysts of different jellyfish vary greatly in both their action and structure; they include proteins and peptides, ranging in size from 3 to 240 kDa, which have cardio-, neuro- or myotoxic activities. The best investigated are the neurotoxins of the sea anemones, all of which act by retarding the inactivation of sodium channels. The proteins may be divided, according to structure, into at least two classes; within each class they are homologous. One such class contains proteins of 44–47 amino acids and includes, for example, the components AsI, AsII and AsV of Anemone sulcata, AxI of *Anthopleura xanthogrammica*, and RpII and RpIII from *Radianthus paumatoensis* (Fig. 9.4c). A second class includes peptides with a maximum of 31 amino acids, such as AsIII and a toxin isolated from *Parasicyonis actinostoloides* [45, 49, 56]. The lethal haemolytic toxins of various sea anemones have a different sort of structure, e.g. cytolysin I–IV from *Stoichactis helianthus* (153–160 amino acids), tenebrosin C from *Actinia tenebrosa* (179 amino acids) and equinatoxin from *A. equina*, but are very similar to each other in sequence [61]. The main constituent of the poison obtained from isolated nematocysts of the Portuguese man-of-war *Physalia physalis* (Siphonophora) is a haemolytic glycoprotein of 240 kDa, composed of three different, as yet unsequenced, polypeptides of 125, 53 and 34 kDa.

References

1. Adams M. E. et al.: Omega-agatoxins: Novel calcium channel antagonists of two subtypes from funnel web spider (Agelenopsis aperta) venom. J. Biol. Chem. 265: 861–867 (1990)
2. Aird S. D. et al.: The amino acid sequence of the acidic subunit B-chain of crotoxin. Biochim. biophys. Acta 1040: 217–224 (1990)
3. Argiolas A. and Pisano J. J.: Bombolitins, a new class of mast cell degranulating peptides from the venom of the bumblebee Megabombus pennsylvanicus. J. biol. Chem. 260: 1437–44 (1985)
4. Barra D. et al.: Rohdei-litorin: a new peptide from the skin of Phyllomedusa rohdei. FEBS Letters 182: 53–56 (1985)
5. Bettini S. (ed.): Arthropod venoms. Hb. exp. Pharmacol. Vol. 48. Springer, Berlin 1978
6. Bjarnason J. B. and Fox J. W.: Characterization of two hemorrhagic zinc proteinases, toxin c and toxin d, from western diamondback rattlesnake (Crotalus atrox) venom. Biochim. biophys. Acta 911: 356–363 (1987)
7. Bouchier C. et al.: Cloning and sequencing of cDNAs encoding the two subunits of crotoxin. Nucleic Acids Res. 16: 9050 (1988)
8. Bougis P. E., Rochat H. and Smith L. A.: Precursors of Androctonus australis scorpion neurotoxins. J. Biol. Chem. 264: 19259–65 (1989)
9. Breckenridge R. and Dufton M. J.: The structural evolution of cobra venom cytotoxins. J. mol. Evol. 26: 274–283 (1987)
10. Brown M. R. et al.: Amino acid sequence of versutoxin, a lethal neurotoxin from the venom of the funnel-web spider Atrax versutus. Biochem. J. 250: 401–405 (1988)
11. Corfield P. W. R., Lee T. J. and Low B. W.: The crystal structure of erabutoxin a at 2.0 Å resolution. J. Biol. Chem. 264: 9239–42 (1989)
12. Dotimas E. M. et al.: Isolation and structure analysis of bee venom mast cell degranulating peptide. Biochim. biophys. Acta 911: 285–293 (1987)
13. Dufton M. J.: Classification of elapid snake neurotoxins and cytotoxins according to chain length: evolutionary implications. J. mol. Evol. 20: 128–134 (1984)
14. Eitan M. et al.: A scorpion venom neurotoxin paralytic to insects that affects sodium current inactivation: Purification, primary structure, and mode of action. Biochemistry 29: 5941–47 (1990)
15. Eng J. et al.: Purification and structure of exendin-3, a new pancreatic secretagogue isolated from Heloderma horridum venom. J. Biol. Chem. 265: 20259–62 (1990)
16. Erspamer V. et al.: Amino acid composition and sequence of Crinia-angiotensin, an angiotensin-II-like endecapeptide from the skin of the Australian frog Crinia georgiana. Experientia 35: 1132–33 (1979)
17. Farid T. M., Tu A. T. and El-Asmar F.: Characterization of cerastobin, a thrombin-like enzyme from Cerastes vipera. Biochemistry 28: 371–377 (1989)

18. Fontecilla-Camps J. C.: Three-dimensional model of the insect-directed scorpion toxin from Androctonus australis Hector and its implication for the evolution of scorpion toxins in general. J. mol. Evol. 29: 63–67 (1989)
19. Gray W. R., Olivera B. M. and Cruz L. J.: Peptide toxins from venomous Conus snails. Annual Rev. Biochem. 57: 665–700 (1988)
20. Habermehl G. G.: Toxic animals and their weapons (in German). 4th Ed. Springer, Berlin 1987
21. Hillyard D. R. et al.: A molluscivorous Conus toxin: Conserved frameworks in conotoxins. Biochemistry 28: 358–361 (1989)
22. Hirabayashi J., Kusonoki T. and Kasai K.: Complete primary structure of a galactose-specific lectin from the venom of the rattlesnake Crotalus atrox: Homologies with Ca-dependent-type lectins. J. Biol. Chem. 266: 2320–26 (1991)
23. Ho C. L. and Ko J. L.: Hornetin: the lethal protein of the hornet (Vespa flavitarsus) venom. FEBS Letters 209: 18–22 (1986)
24. Ho C. L. and Hwang L. L.: Structure and biological activities of a new mastoparan isolated from the venom of the hornet Vespa basalis. Biochem. J. 274: 453–456 (1991)
25. Huang T. F. et al.: Purification and characterization of an antiplatelet peptide, arietin, from Bitis arietans venom. Biochim. biophys. Acta 1074: 136–143 (1991)
26. Inoue S. et al.: Amino acid sequence of a cytotoxin-like basic protein with low cytotoxic activity from the venom of the Thailand cobra Naja naja siamensis. FEBS Letters 218: 17–21 (1987)
27. Inoue S. et al.: Amino acid sequences of nerve growth factors derived from cobra venoms. FEBS Letters 279: 38–40 (1991)
28. Itoh N. et al.: Organization of the gene for batroxobin, a thrombin-like snake venom enzyme. Homology with the trypsin/kallikrein gene family. J. Biol. Chem. 263: 7628–31 (1988)
29. Junor L. R. et al.: The purification and amino acid sequence of the lethal neurotoxin Tx1 from the venom of the Brasilian armed spider Phoneutria nigriventer. FEBS Letters 263: 251–253 (1990)
30. Kaiser I. I. et al.: The amino acid sequence of a myotoxic phospholipase from the venom of Bothrops asper. Arch. Biochem. Biophys. 278: 319–325 (1990)
31. Kem W. R.: Structure and action of nemertine toxins. Amer. Zool. 25: 99–111 (1985)
32. Kemeny D. M. et al.: The purification and characterization of hyaluronidase from the venom of the honey bee, Apis mellifera. Eur. J. Biochem. 139: 217–223 (1984)
33. King T. P., Kochoumian L. and Joslyn A.: Wasp venom proteins: Phospholipase A1 and B. Arch. Biochem. Biophys. 230: 1–12 (1984)
34. Kopeyan C. et al.: Primary structure of scorpion anti-insect toxins isolated from the venom of Leiurus quinquestriatus quinquestriatus. FEBS Letters 261: 423–426 (1990)
35. Kosen P. A. et al.: Structural studies of α-bungarotoxin. 3. Corrections in the primary sequence and X-ray structure and characterization of an isotoxic α-bungarotoxin. Biochemistry 27: 2775–81 (1988)
36. Lee C. Y. (ed.): Snake venoms. Hb. exp. Pharmacol. Vol 52. Springer, Berlin 1979
37. Lind P. and Eaker D.: Amino-acid sequence of the α-subunit of taipoxin, an extremely potent presynaptic neurotoxin from the Australian snake taipan (Oxyuranus s. scutellatus). Eur. J. Biochem. 124: 441–447 (1982)
38. Liu C. S. et al.: The amino acid sequence and properties of an edema-inducing Lys-49 phospholipase-A2 homolog from the venom of Trimesurus mucrosquamatus. Biochim. biophys. Acta 1077: 362–370 (1991)
39. Lomonte B. et al.: Isolation of a galactose-binding lectin from the venom of the snake Bothrops godmani (Godman pi viper). Toxicon 28: 75–81 (1990)
40. Loret E. P. et al.: An anti-insect toxin purified from the scorpion Androctonus australis Hector also acts on the α-sites and β-sites of the mammalian sodium channel: Sequence and circular dichroism. Biochemistry 30: 633–640 (1991)
41. Manchewa I. et al.: Sequence homology between phospholipase and its inhibitor in snake venom. The primary structure of phospholipase A2 of vipoxin from the venom of the Bulgarian viper (Vipera ammodytes ammodytes, Serpentes). Biol. Chem. Hoppe-Seyler 368: 343–352 (1987)
42. Maraganore J. M. and Heinrikson R. L.: The lysine-49 phospholipase A2 from the venom of Agkistrodon piscivorus piscivorus. Relation of structure and function to other phospholipases A2. J. biol. Chem. 261: 4797–4804 (1986)
43. Massefski W. jr. et al.: Molecular structure of charybdotoxin, a pore-directed inhibitor of potassium ion channels. Science 249: 521–524 (1990)
44. McMullen B. A., Fujikawa K. and Kisiel W.: Primary structure of a protein C activator from Agkistrodon contortrix venom. Biochemistry 28: 674–679 (1989)
45. Metrione R. M., Schweitz H. and Walsh K. A.: The amino acid sequence of toxin Rp111 from the sea anemone, Radianthus paumotensis. FEBS Letters 218: 59–62 (1987)
46. Mochcamorales J., Martin B. M. and Possani L. D.: Isolation and characterization of helothermin, a novel toxin from Heloderma horridum horridum (Mexican beaded lizard) venom. Toxicon 28: 299–309 (1990)
47. Montecucchi P. C., Gozzini L. and Erspamer V.: Primary structure determination of a tryptophan-containing tridecapeptide from Phyllomedusa rohdei. Int. J. Peptide Protein Res. 27: 175–182 (1986)
48. Mori N. and Tu A. T.: Isolation and primary structure of the major toxin from sea snake, Acalyptophis peronii, venom. Arch. Biochem. Biophys. 260: 10–17 (1988)
49. Nishida S. et al.: Amino acid sequence of a sea anemone toxin from Parasicyonis actinostoloides. Eur. J. Biochem. 150: 171–173 (1985)
50. Ohtani Y. et al.: Some properties of a kininogenase from the venom of Agkistrodon caliginosus (Kankoku-Mamushi). Toxicon 26: 903–912 (1988)
51. Olivera B. M. et al.: Diversity of Conus neuropeptides. Science 249: 257–263 (1990)
52. Pearson J. A. et al.: Studies on the subunit structure of textilotoxin, a potent presyaptic neurotoxin from the venom of the Australian common brown snake (Pseudonaja textilis). 2. The amino acid sequence and toxicity studies of subunit-D. Biochim. biophys. Acta 1077: 147–150 (1991)
53. Pungercar J. et al.: Amino acid sequence of ammodytoxin C as deduced from cDNA. Nucleic Acids Res. 17: 4367 (1989)
54. Savel-Niemann A.: Tarantula (Eurypelma californicum) venom, a multicomponent system. Biol. Chem. Hoppe-Seyler 370: 485–498 (1989)

55. Schmidt J. O.: Biochemistry of insect venoms. Annual Rev. Entomol. 27: 339–368 (1982)
56. Schweitz H. et al.: Purification, sequence and pharmacological properties of sea anemone toxins from Radianthus paumatoensis. A new class of sea anemone toxins acting on sodium channel. Biochemistry 24: 3554–61 (1985)
57. Shafquat J. et al.: Primary structure and functional properties of cobra (Naja naja naja) venom Kunitz-type trypsin inhibitor. Eur. J. Biochem. 194: 337–341 (1990)
58. Shafquat J. et al.: Extensive multiplicity of the miscellaneous types of neurotoxins from the venom of the cobra Naja naja naja and structural characterization of major components. FEBS Letters 284: 70–72 (1991)
59. Shai Y. et al.: Sequencing and synthesis of pardaxin, a polypeptide from the Red Sea moses sole with ionophore activity. FEBS Letters 242: 161–166 (1988)
60. Simmaco M. et al.: A family of bombinin-related peptides from the skin of Bombina variegata. Eur. J. Biochem. 199: 217–222 (1991)
61. Simpson R. J. et al.: Complete amino acid sequence of tenebrosin-C, a cardiac stimulatory and haemolytic protein from the sea anemone Actinia tenebrosa. Eur. J. Biochem. 190: 319–328 (1990)
62. Skinner W. S. et al.: Purification and characterization of two classes of neurotoxins from the funnel web spider, Agelenopsis aperta. J. Biol. Chem. 264: 2150–55 (1989)
63. Smith L. A. and Schmidt J. J.: Cloning and nucleotide sequences of crotamine genes. Toxicon 28: 575–585 (1990)
64. Stapleton A. et al.: Curtatoxins. Neurotoxic insecticidal polypeptides isolated from the funnel-web spider Hololena curta. J. Biol. Chem. 265: 2054–59 (1990)
65. Takeya H. et al.: The complete amino acid sequence of the high molecular mass hemorrhagic protein HR1B isolated from the venom of Trimeresurus flavoviridis. J. Biol. Chem. 265: 16068–73 (1990)
66. Tamiya T. et al.: Cloning and sequence analysis of the cDNA encoding for a snake neurotoxin precursor. Biochimie 67: 185–189 (1985)
67. Terwilliger T. C. and Eisenberg D.: The structure of melittin. J. biol. Chem. 257: 6016–22 (1982)
68. Teuscher E. und Lindequist U.: Biogenic Toxins (in German). G. Fischer, Stuttgart 1987
69. Tu A. T. (ed.): Insect poisons, allergens and other invertebrate venoms. Handbook of natural Toxins Vol. 2. M. Dekker, New York 1984
70. Tyler M. I. et al.: Pseudonajatoxin b: unusual amino acid sequence of a lethal neurotoxin from the venom of the Australian common brown snake, Pseudonaja textilis. Eur. J. Biochem. 166: 139–143 (1987)
71. Ueda M., Chang C. C. and Ohno M.: Purification and characterization of the L-amino acid oxidase from the venom of Trimeresurus mucrosquamatus (Taiwan habu snake). Toxicon 26: 695–706 (1988)
72. Vandermeers A. et al.: Differences in primary structure among 5 phospholipases-A2 from Heloderma suspectum. Eur. J. Biochem. 196: 537–544 (1991)
73. Vlasak R. et al.: Nucleotide sequence of cloned cDNA coding for honeybee prepromelittin. Eur. J. Biochem. 135: 123–126 (1983)
74. Yanagawa Y. et al.: A novel sodium channel inhibitor from Conus geographus: Purification, structure and pharmacological properties. Biochemistry 27: 6256–62 (1988)

10 Proteins of Muscle and the Cytoskeleton

The contractile elements of striated vertebrate skeletal muscle, the myofibrils, contain thin filaments, which are 6 nm in diameter and consist mainly of actin, and thicker myosin filaments with a diameter of 16 nm (Fig. 10.1). During muscle contraction, the filaments undergo a sliding movement relative to each other (**sliding filament mechanism**). This is brought about by the reversible formation of bridges between the myosin molecules and the actin filaments, which bind, change their conformation and then dissociate (**bridge cycle**). The required energy is supplied by the hydrolysis of ATP. The sliding distance (step size) per molecule of ATP hydrolysed is controversial; the most recent measurements give a value of 40 nm [102, 286]. As yet, the molecular processes of bridge formation and change in conformation have not been fully defined [79, 257, 267]. Because one ATP is hydrolysed per bridge cycle, the **ATPase activity** can be used as an in vitro indication of contraction. Isolated myosin has a very low ATPase activity, but this increases up to 1000-fold after combination with actin to give actomyosin. ATPase is inhibited in relaxed muscle; the relief of ATPase inhibition and the triggering of contraction in activated muscle involves an increase in the sarcoplasmic Ca^{2+} concentration from about 0.1 μmol/l at rest to about 10 μmol/l on activation. The calcium-dependent regulation of contraction is brought about by tropomyosin and the troponin complex, which are an integral part of the thin filaments [238].

Essentially the same processes are responsible for the contraction of all muscle types, and also for the movement of non-muscle cells. Actins are abundant in all cell types, making up 20–25 % of the cell protein in muscle cells and 10–15 % in non-muscle cells. Whereas the actin molecules in striated muscle are all aggregated into filaments, they exist in non-muscle cells as both actin monomers (G-actin) and filaments (F-actin). The formation and degeneration of actin filaments is regulated by various proteins. Further actin-binding proteins bring about the association of actin filaments in non-muscle cells into bundles and networks. The myosin concentration in smooth-muscle cells (8 % of the total protein) is significantly lower than in striated muscle (36 %), and is lower still in non-muscle cells (1 %). Smooth muscle contains short myosin filaments which, unlike the situation in striated muscle, do not show any regular structural relationship to the actin filaments. The myosin of non-muscle cells can also form filaments of 12–18 molecules in vitro; however, the nature of myosin aggregation in the cytoplasm of living cells is not known [244].

The cytoplasm of all cells contains filament-like structures which together make up the **cytoskeleton**. This is a complex system of protein molecules which confers on the cell its typical form, internal organization and motility [313]. In addition to actin microfilaments (diameters of 6–8 nm), the other main components of the cytoskeleton are the tubular microtubules (exter-

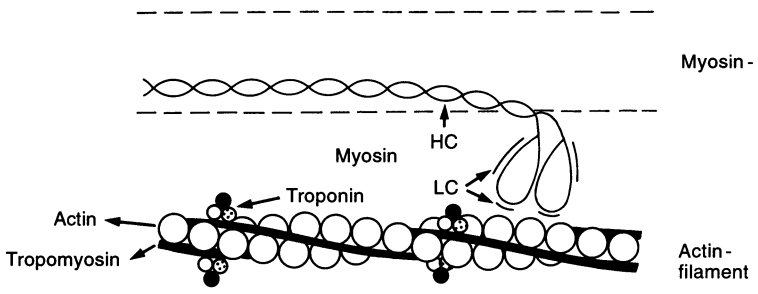


Fig. 10.1. The actin-myosin system of vertebrate skeletal muscle. For clarity, only one myosin molecule is shown in the thick myosin filament; this consists of a pair of heavy chains *(HC)* and two pairs of light chains *(LC)*. In addition to the two entwined strands of actin monomers, the thin actin filament includes tropomyosin and the troponin complex of three molecules

nal diameter 25 nm, internal diameter 15 nm) and the intermediary filaments (with a diameter of about 10 nm). The microtubules are made up mostly of tubulin but can also bind other proteins (the microtubule-associated proteins). The movement of flagella and cilia is the result of interaction between the microtubules and the ATP-hydrolysing dynein. The composition of the intermediary filaments varies markedly with the cell type, but the constituent proteins all belong to the same super-family and present a particularly interesting example of adaptive evolution. They include the cytokeratins of epithelial cells, the vimentins from cells of mesenchyme origin, the desmins of muscles, the neurofilament proteins, and the acidic protein of the glial filaments (GFAP). It has recently been demonstrated that lamins A and C from the nuclear membrane are homologous to proteins of the intermediary filaments. Specific differentiated cells contain filament-like structures which do not conform to any of the three general types (microfilaments, intermediary filaments, microtubules), e.g. the connectin/titin network of the myofibrils, the spasmonemes in the stalk of the peritrichous ciliate *Vorticella*, or the myonemes of *Stentor* [244]. The spermatozoa of the nematode *Caenorhabditis elegans* have no myosin and only a very little actin and tubulin (0.02 % of the total protein), but contain large amounts (15 %) of specific basic proteins (main sperm proteins, MSPs) of 15 kDa; these are mainly located in the pseudopodium and are therefore probably involved in the movement of these aberrant sex cells. Of the more than 30 MSP genes present, at least three are expressed [28].

The components of the cytoskeleton are involved in complex interactions with each other and with the plasma membrane, and beyond this with components of the extracellular matrix, such as collagen, fibronectin and laminin. The mechanical strength of the plasma membrane relies upon a **membrane skeleton** of specific structural proteins; this is especially well developed in vertebrate erythrocytes. Finally, the structural proteins of the cell also include clathrin, which forms the polyhedral envelope of the cytoplasmic vesicles (coated vesicles), and a special type of structural protein found in the cells of the eye lens. What follows is a comparative description of the structural variety and evolution of the proteins of muscle and the cytoskeleton; the physiological and cell biological aspects of these proteins are covered by many other textbooks and monographs.

10.1 Calcium Regulation of the Actomyosin System

In both muscle and non-muscle cells, the ATPase activity and movement of the actin-myosin system are controlled by calcium ions. The **regulation of the intracellular calcium concentration** in skeletal muscle is carried out mainly by the sarcoplasmic reticulum; extracellular calcium is of great importance for smooth muscle and heart muscle, as is calcium storage in the mitochondria of non-muscle cells. In smooth muscle and non-muscle cells, Ca^{2+} is taken up from the outside by potential-dependent or ligand-dependent calcium channels, and is set free from intracellular reserves by the action of the secondary messenger inositol-1,4,5-trisphosphate [238, 293]. The **calcium pump of the sarcoplasmic reticulum**, which transports calcium from the sarcoplasma back to the sarcoplasmic reticulum during relaxation, is stimulated both by free Ca^{2+} and by calcium complexed with the intracellular receptor calmodulin. Hormones like adrenalin and noradrenalin influence the calcium pump via β-adrenergic membrane receptors, which activate a cAMP-dependent protein kinase in association with the corresponding G proteins and adenylate cyclase. The kinase phosphorylates an integral protein of the reticulum membrane, phospholamban, which

then effects an increase in the affinity of the Ca^{2+}-ATPase of the sarcoplasmic reticulum. Phospholamban is present in skeletal and heart muscle as a pentamer of 52-amino-acid subunits; the sequence of this protein varies only slightly between different mammals, and these differ by only 22–29 % from the chicken phospholamban [195, 285]. **Calsequestrin** in the reticulum membrane binds Ca^{2+} in large amounts, although with a low affinity, and it thus maintains the free calcium concentration of the reticulum lumen at 0.5–1.0 mmol/l; this restricts the concentration gradient between the sarcoplasma and the reticulum, against which the calcium pump must work [294]. The sarcoplasmic calcium concentration is not reduced quickly enough by uptake into the reticulum to ensure rapid muscle relaxation. Instead, much of the Ca^{2+} is bound temporarily to **parvalbumin**, which functions as a soluble relaxation factor [238].

There are three main mechanisms by which calcium ions **regulate the actomyosin ATPase;** one affects the thin actin filaments (A regulation), and the other two affect the thick myosin filaments (M regulation). Vertebrate skeletal and heart muscle, for example, are regulated **actin specifically**. In this case, ATPase is inhibited at low calcium concentrations by the troponin-tropomyosin complex of the thin filaments; binding of Ca^{2+} to troponin C relieves this inhibition. Until recently, the inhibitory effect of the troponin-tropomyosin complex in vertebrate muscle was thought to be due solely to steric hindrance of the actin-myosin interaction; this was supported by the observation that tropomyosin changes its position on the actin filaments at increased calcium concentrations. However, it has now been shown that actin-myosin binding may be stable in the absence of calcium ions, even though both ATPase activity and contraction are inhibited. Apparently, the absence of Ca^{2+} inhibits a step in ATP hydrolysis which occurs after the binding of myosin to actin. In preparations of various mollusc muscles, e.g. the striped adductor muscle of the scallop *Aequipecten irradians* or the mantle muscle of the cephalopod *Loligo pealii*, myosin-actin binding also persists in the presence of ATP and the absence of calcium, although it may be weaker. In contrast, the ATPase activity declines by 90 %, in particular by reduction of the V_{max}. Ca^{2+} thus influences the kinetics of ATP hydrolysis as well as of actin-myosin binding [238].

A **myosin-specific regulation mechanism** is active in various mollusc muscles and in vertebrate smooth muscle. The contraction of mollusc muscles is triggered by direct binding of Ca^{2+} to myosin. The Ca^{2+}-binding regulatory domain includes the regulatory and the essential light chains (LCs) and a part of the heavy chain (HC) in the neck region of the myosin head. Ca^{2+} apparently binds specifically to the essential LC, which in mollusc myosins, in contrast to those not directly regulated by Ca^{2+}, carries a functional Ca^{2+}-binding site [153]. Ca^{2+} regulation of vertebrate smooth muscle is also myosin specific; here, however, the inhibitory effect of the regulatory LC is prevented through phosphorylation by a specific Ca^{2+}-calmodulin-dependent myosin light-chain kinase (MLCK). Phosphorylation of the regulatory LC at a further site by the Ca^{2+}-phospholipid-dependent protein kinase C has a deactivating effect; this does not occur in embryonal smooth muscle [158]. The regulatory LC in striated skeletal muscle can also be phosphorylated by calmodulin-dependent MLCK; phosphorylation of at least part of the LC is important for both the strength and the efficiency of muscle contraction [238]. A mutant, *mdf*$^-$, of *Drosophila melanogaster* has been isolated and shown to have threefold lower myosin ATPase activity due to a lack of phosphorylation [278].

In addition to myosin phosphorylation, **caldesmon** appears to play a role in the regulation of vertebrate smooth-muscle (chicken stomach, bovine aorta and uterus) and non-muscle (blood platelets) cells; this is a calcium-regulated dimeric protein with the ability to bind in a flip-flop manner to F-actin and calmodulin. In relaxed muscle at low calcium concentration, caldesmon is bound to actin and ATPase is inhibited; at higher calcium levels, caldesmon binds to calmodulin via an interaction with tropomyosin and the inhibition is relieved. Regulation in vertebrate smooth muscle probably depends upon cooperation between both the mechanisms described. Through its effect on phosphorylation of the regulatory chain, MLCK brings about the correct conditions for actin-myosin interaction, which is then switched on or off by the caldesmon flip-flop mechanism [238, 250]. Caldesmon occurs in two tissue-specific isoforms: h-caldesmon (h for heavy) is typical of smooth muscle, and l-caldesmon (l for light) is found in the brain and other non-muscle cells; caldesmon is absent from skeletal and heart muscle. The heavy form is composed of 771 amino acids, and l-caldesmon differs only by the deletion of 232 amino acids from the repetitively organized central region [252]. A caldesmon-like protein has also been isolated from striated

muscle of the molluscs *Pecten maximus* and *Sepia officinalis* [13]. Regulation of the smooth muscle also probably involves **calponin**, a protein of 252 or 292 amino acids which belongs to the troponin C family and inhibits actomyosin ATPase. Ca^{2+} binding has no effect on the inhibitory activity of calponin, although this disappears after phosphorylation by Ca^{2+}-calmodulin-dependent protein kinase [277, 307].

The absence of an increase in the ATPase activity of a myosin preparation following the addition of pure, troponin-free vertebrate actin indicates the presence of M regulation. According to this test, which was first performed on the adductor muscle of the scallop, the muscles of molluscs, nemertines, brachiopods and echinoderms are **M regulated**, whereas the skeletal muscles of vertebrates and tunicates are **A regulated;** most arthropod muscles and the diagonally striated muscles of nematodes, annelids and sipunculids are both A and M regulated [238]. However, all three components of the troponin complex have been isolated from the striated adductor muscle of the mussel *Chlamys nipponensis akazara*, and thus the molluscs also appear to possess two different regulation mechanisms (**dual regulation**) [207]. There are also indications of dual regulation in the tunicates. Although a troponin-tropomyosin complex is present in the smooth muscle of the ascidian *Halocynthia roretzi*, and can combine with the species' own or with mammalian actomyosin to give a calcium-regulated system, the ATPase of ascidian actomyosin is clearly calcium dependent without the troponin-tropomyosin complex, even though this M regulation is of relatively little importance [273].

In many **mollusc muscles**, such as the smooth shell adductor and the byssus retractor of the mussel, there is a special contraction condition, the **catch**, the molecular mechanism and regulation of which remain obscure despite many years of intense effort [238]. The thick filaments in these muscles have a special structure and contain large amounts of the muscle protein **paramyosin**, which is worthy of further discussion. The asynchronous **flight muscles of the insects** pose special problems; these contain a functional tropomyosin-troponin complex but their contraction is also modulated by a passive expansion. In the muscle fibres of the water bug *Lethocerus*, the large size of which makes them convenient subjects for investigation, Ca^{2+} brings about a fourfold increase in ATPase activity, but a subsequent expansion of 1% brings a further threefold increase. The molecular mechanism by which such expansion influences the ATPase activity and tension development in insect muscle is not known. Some of the many myofibrillar proteins of insect muscle are unique to insects, and it may be that one or other of them is involved in the special regulatory characteristics of the muscles. The thick filaments contain varying proportions of paramyosin: it is higher in the case of insects with low wing-beat frequencies (in *Lethocerus* it is 10%) and lower in those with high beat frequencies (e.g. in flies it is only 2%). There is also phosphorylation of the light myosin chains of insect muscles, but the biological significance of this is not completely clear [238]. In addition to actin (43 kDa), the thin filaments in the flight muscles of *Drosophila* and other insects also contain large amounts of a similar protein, **arthrin** (55 kDa), which is a combination of actin and ubiquitin [10].

10.2 Molecular Heterogeneity of Skeletal Muscle Fibres

It is not surprising that different muscle types and non-muscle cells exhibit differences in the function and molecular structure of their contractile apparatus. However, in vertebrate skeletal muscle there are, in fact, corresponding differences between single muscle fibres. Thus, in the mammals one can distinguish between the slow muscle fibres of class I (slow twitch) and the faster fibres of class II (fast twitch), which differ in their innervation, the nature of their stimuli-transmitting structures, the metabolic provision of energy, and also the structural and functional properties of their contractile apparatus. Each skeletal muscle is made up of a mosaic of different **types of fibre** in varying proportions, according to the muscle function. All muscle fibres supplied by the same nerve fibre (motor units) belong to the same class. Following a change in muscle function, fibres can convert from one type to the other; for example, during postnatal development of the rat m. soleus, type I changes to type II. Persistent electrical excitation can effect the conversion of type II to I [220]. The actin-myosin system of all fibre types, in fact of all muscle types and non-muscle cells, is basically very similar; functional differentiation is based upon variety in the **isoforms of protein components**, i.e. the actins, myosins, tropomyosins, and troponins. However, a more detailed examination shows that the composition, for example, of the tropomyosin-troponin system in different muscle

fibres is rather a continuum, which is only crudely reflected by classification into types. In addition, it should not be forgotten that each skeletal muscle fibre possesses many nuclei which may exhibit different patterns of gene expression [242, 258]. There are also characteristic differences between the fibre types in the enzymes of energy-providing metabolism, and these are the basis of a more exact classification, e.g. slow-twitch oxidative (SO), or fast-twitch glycolytic (FG) and fast-twitch oxidative glycolytic (FOG), or in the acid stability of the myosin ATPase (subclasses IIA and IIB) [220]. The heterogeneity of invertebrate muscle fibres is less marked. Thus, for example, in the mussel *Spisula solidissima* the adductor, foot muscle, mantle and heart all contain the same forms of actin, myosin and tropomyosin. On the other hand, the outer muscle cylinder of the earthworm *Lumbricus terrestris* has two different types of fibre which differ in the light myosin chain and in succinate dehydrogenase activity [55]. The claw-closing muscles of the lobster *Homarus americanus* contain three types of muscle fibre, which show qualitative and quantitative differences in their protein spectra, and are distributed characteristically between the functionally different cutter and crusher claws. Altogether, two paramyosins, three troponin-Ts, five troponin-Is, two troponin-Cs, three myosin-αLCs and one myosin-βLC can be isolated from these muscles [201].

The existence of many isoforms of almost all muscle proteins is partly due to the expression of different genes; **multi gene families** are known in particular for the actins, the heavy and light chains of the myosins, and the tropomyosins. A further mechanism for the production of protein heterogeneity is particularly characteristic of the muscle proteins, namely, the **variable splicing of the primary transcript** of one gene. Alternative splicing has in fact been shown for almost all muscle proteins: actins, heavy myosin chains, light myosin chains, tropomyosins, troponin T, troponin I and calsequestrin. Finally, muscle proteins may be **post-translationally modified** in different ways. Phosphorylation, for example, is found in light and heavy myosin chains, troponin I and T, and caldesmon, although the functional meaning is in many instances by no means clear. Methylation is also very common amongst the muscle proteins, Monomethyl-lysine and -histidine, trimethyl-lysine and histidine have been demonstrated in myosins [175]; dimethyl- and trimethyl-lysine are found in calmodulins and other calcium-binding proteins [139, 243], as well as in the actin-binding protein profilin [4].

10.3 Myosins

Two fundamentally different **types of myosin** may be distinguished. The myosins of type II are soluble at a salt concentration of 600 mmol/l and at physiological salt concentrations of 100–250 mmol/l they aggregate to form filaments. The molecules of **myosin II** have a rod-shaped tail of about 150 nm and two globular heads of about 20 × 7 nm. They are hexamers of 460–500 kDa, consisting of two heavy chains (HCs) of about 200 kDa and four light chains (LCs) of 15–27 kDa. The C-terminal ends of both HCs are twisted into an α-helical coiled-coil and form the tail region. The head region is made up of the N-terminal half of each HC together with one each of the regulatory and essential LCs, and carries binding sites for actin and adenine nucleotide (Fig. 10.1). The myosin molecule can be cleaved in various ways by partial proteolysis: the lighter fragment, LMM, corresponds to the C-terminal two-thirds of the tail, and the heavier HMM to the rest of the tail and the two heads. The subfragment designated S-1 consists of the N-terminal half of an HC, folded into a head, together with the two associated LCs; S-2 is the tail section between the LMM/HMM border and the heads. The thick filaments of vertebrate skeletal muscle each contain several hundred myosin molecules arranged in a bipolar fashion, such that the heads of both filament halves point to their respective ends. There are further proteins in the thick filaments, e.g. the C protein, found in several muscle-type-specific isoforms of 135–150 kDa, and the similar X protein. According to the results of X-ray analysis, the molecular arrangement of the thick filaments of the invertebrates differs somewhat from that of the vertebrates. **Myosin I**, which does not form filaments, consists of a single HC of 110–140 kDa and one or two LCs. As in myosin II, the N-terminal domain of the HC also forms the head, but the tail region has a completely different structure. The function of myosin I in living cells has not been determined.

Due to the relatively low myosin concentrations in smooth muscle, and especially in non-muscle cells, the myosins of the vertebrates were initially investigated in the various types of fibre found in skeletal and heart muscle. More recently, the possibilities for cDNA and gene sequencing have significantly expanded our knowledge of the myosins of vertebrate smooth muscle and non-muscle cells as well as of inver-

tebrate muscles. The myosins of the nematode *Caenorhabditis elegans* and the soil amoeba *Acanthamoeba castellani* have been particularly thoroughly investigated. In the nematodes, 22 genes are now known which code for muscle proteins; mutations in these genes lead to defective motility and abnormalities of the muscle ultrastructure. They include four myosin-HC genes, three myosin-LC genes, four actin genes, the paramyosin gene unc-15, and the gene unc-22, which encodes the myosin-associated, 600-kDa protein twitchin. Of the four HC genes, A and B are expressed in the 95 muscle cells of the body wall, and C and D only in the pharynx musculature. Myosin HC-A is found only in the central region of the thick filaments and apparently plays a role in assembly. There are present two LCs of 16 and 18 kDa; two genes are found for the regulatory 18-kDa LC, and these code for identical amino acid sequences [5, 122]. The amoeba possesses one myosin II with the usual structure, and four type-I myosins. Myosin IB is made up of an HC of 125 kDa and one LC of 27 kDa, whereas myosin IC has one 130-kDa HC and two LCs of 14 kDa. Very recently, a gene has been sequenced for a relatively large HC of 177 kDa and this apparently belongs to a high molecular weight form of myosin I (HMWMI) [111, 171].

Only by DNA-sequence analysis has it been possible to determine the amino acid sequence of giant polypeptides such as myosin HC. The first complete **HC sequence** obtained was that of myosin B from the body wall of *Caenorhabditis elegans*, which is encoded by gene unc-54. This is described here as an example of the rather uniform HCs of myosin type II. It consists of a total of 1966 amino acids; the N-terminal 850 amino acids are folded into the head region, in which are found certain amino acids responsible for ATP binding, ATPase activity, and actin and LC binding. There follows a hinge or neck region, in which the mobility of the bridges is located; and, finally, the rod region.

The amino acid sequence of the rod region shows considerable **periodicity** with period lengths of 7, 28 and 196 amino acids. The 160 repeats of seven amino acids contain non-polar amino acids at the first and fourth positions. This leads to the formation of fields of non-polar side-chains, which allow hydrophobic interaction between the two HCs, thereby stabilizing the coiled-coil structure of the rod region. This pattern is interrupted only by four inserted amino acids (skip residues) and ends at the HMM/LMM border. The LMM region of the HC carries many charged amino acids at its outer surface, such that their side-chains produce a charge pattern on the surface of the coiled-coil with periodicities of 28 and 196 amino acids; the stable structure of the thick filaments thus relies mainly upon electrostatic interaction. The HC hinge region (S-2) has no periodicity of charged groups and can therefore move away from the surface of the thick filament when the myosin head makes contact with an actin filament. The polypeptide chain ends with a non-helical, C-terminal extension of 24 amino acids, similar to those seen on other myosin HCs from *Caenorhabditis, Drosophila* and non-muscle cells but not from vertebrate skeletal muscle [65].

The HCs of all type-II myosins have a structure similar to that of *Caenorhabditis* HC [167, 181, 260]. However, they may be further subdivided on the basis of particular sequence characters into two families: the HCs of striated skeletal and heart muscle, and the HCs of smooth muscle and non-muscle cells. In the chicken, the HCs of the smooth, intestinal muscles have 81–84 % sequence similarity to non-muscle-cells HCs but only 38–41 % similarity to skeletal-muscle HCs [124]. Within the two families themselves, there are also instances of HC isoforms with different, specific, physiological functions and corresponding deviant structures. Thus, for example, in the rat there are at least seven muscle-specific HCs (embryonal skeletal muscle, neonatal, extraocular, fast IIA, fast IIB, α-ventricle/atrium, β-ventricle/slow) encoded by at least 13 genes. Altogether, six different HCs are expressed in the breast muscle of the chicken during development, and seven HC isoforms are present in the various muscles of the fish *Salvelinus alpinus* [65, 107, 177, 260]. All the smooth muscles of the mammals contain the same two HCs; a third type is found only in the lung arteries [208]. In vertebrates, the variety of HCs is based on a multi-gene family, which in the case of the chicken includes 31 members. Additional variation in the protein-coding sequences arises by alternative splicing. In *Drosophila*, all the HCs of striated muscles come from one gene and those of non-muscle myosin come from another. Five of the 18 exons in the muscle-HC gene can each be spliced in five different ways; theoretically, therefore, there are 480 possible HC isoforms in *Drosophila* [128, 129]. In spite of the generally similar geometry of the tail regions of different myosins, there is evidence for adaptive evolution of their amino acid sequences. The proportion of α-helix in the tail region, which at 0–5 °C accounts for almost 100 %, decreases

with increasing temperature in all myosins. The mean "melting point", however, varies from 33 °C for Arctic fish to 45 °C for frogs and 49 °C for rabbits [235].

Two classes of **light** chain can be distinguished; these have different functional properties and are solubilized from myosin molecules in different ways. One type of LC is termed "regulatory", in view of its involvement in calcium regulation in the scallop [88]. The second type, known as "essential", was initially thought to be an absolute requirement for ATPase function; in the meantime, it has become clear that the myosin fragment S-1 can hydrolyse ATP in the absence of an LC. The function of the essential LCs is therefore controversial [88]. In the myosin of mammalian skeletal muscle, the essential chains are represented by the alkaline LCs and the regulatory chains by DNTB LCs, which can be extracted with the Ellmann reagent 5,5′-dithiobis(2-nitrobenzoic acid) (Fig. 10.2). In contrast, the regulatory LCs of vertebrate smooth muscle are not DNTB-extractable and resemble much more those of the molluscs in their phosphorylation and calcium-binding capacity. The regulatory LCs from the myosins of the scallop *Aequipecten irradians* can be extracted with ethylenediaminetetraacetic acid (EDTA). The extraction of just one of the regulatory LCs removes the calcium dependence of ATPase activity of the whole myosin molecule. This is apparently a reflection of the interaction between various subunits of the molecule. The calcium sensitivity of such molecules can be restored by re-addition of the removed LCs, or even by the addition of regulatory LCs from other myosins, e.g. from vertebrate smooth muscle or the DNTB LC from mammalian skeletal muscle [238].

Regulatory LCs can be phosphorylated on specific serine residues by **myosin LC kinase** (MLCK). This is true not only for vertebrate smooth muscle, where the process has a regulatory function, but also for skeletal and heart muscle. Thus, the heart is often found to contain several variably phosphorylated regulatory LCs. The MLCK found in a wide variety of muscle and non-muscle cells is activated by calcium ions and is highly specific for myosin regulatory LCs. Skeletal muscle MLCKs (65–90 kDa) are smaller than those of smooth muscle (130–150 kDa). The species- and muscle-specific differences between different MLCKs lie mainly in the N-terminal region, whilst the central catalytic domain and the C-terminal calmodulin-binding domain show great similarity [100]. The MLCKs themselves may be phosphorylated by cAMP-dependent protein kinases; this is inhibitory in the case of smooth muscle and non-muscle cells, where calmodulin affinity is reduced, but has no effect on the enzyme activity of heart or skeletal muscle [88, 238].

The essential and regulatory myosin LCs belong to the same **protein family** as troponin C, calmodulin and other calcium-binding proteins. Of the four calcium-binding sites, only the first is functional in regulatory LCs and none in the essential LCs. Binding of Ca^{2+} is prevented in the non-functional domains by at least one amino acid substitution [65, 89]. In contrast, in the molluscs, the essential LCs still possess one active calcium-binding site [153]. Sequence comparisons between the various LC types lead to some interesting conclusions (Table 10.1). The regulatory DNTB chain from the crop of the chicken differs at almost 50 % of positions from skeletal muscle, and the distance from rabbit skeletal muscle is hardly any greater. Thus, the sequences appear to be more related to muscle type than to species. The same is true of the essential LCs; however, these show a slower evolution with only 30 % dif-

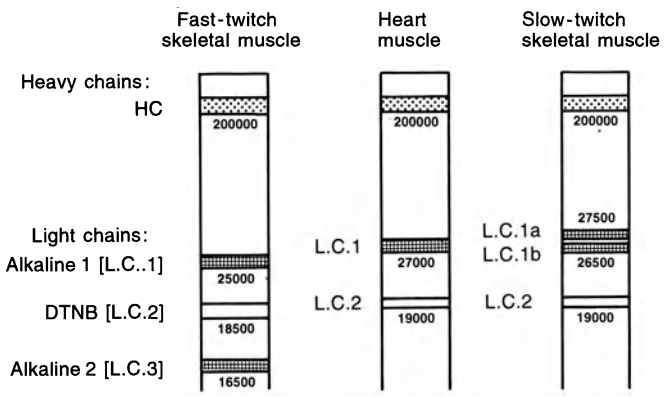


Fig. 10.2. Gel electrophoregrams of the myosin subunits in various rabbit skeletal muscles show one band for the heavy chains *(HC)* and two or three bands for the light chains *(LC)*. The *shaded* and *open bands* represent the essential and regulatory LCs, respectively; the *numbers* are the molecular masses in Da

Table 10.1. The percentage agreement between the amino acid sequences in the families of calmodulin-like proteins

Sequences compared	Agreement (%)	Reference
Regulatory light chain		
Chicken crop/skeletal muscle	53	[174]
Chicken crop/rabbit skeletal muscle	54	[174]
Chicken skeletal muscle/rabbit skeletal muscle	91	[174]
Rabbit skeletal muscle/bovine calmodulin	30	[53]
A2 chicken heart/2B chicken heart	90	[178]
Halocynthia roretzi/rabbit skeletal muscle	56	[273]
H. roretzi/*Chlamys nipponenis akazara*	39	[273]
H. roretzi/bovine calmodulin	32	[273]
H. roretzi/Tn-C rabbit skeletal muscle	27	[273]
Aequipecten irradians/rabbit skeletal muscle	31	[88]
A. irradians/chicken crop	44	[88]
Essential light chain		
Chicken crop/A2 skeletal muscle	70	[89]
Chicken crop/A1 rabbit skeletal muscle	70	[89]
A1 chicken skeletal muscle/A1 rabbit skeletal muscle	84	[173]
A2 chicken skeletal muscle/A2 rabbit skeletal muscle	83	[173]
A1/A2 chicken skeletal muscle	95	[173]
A1/A2 rabbit skeletal muscle	95	[53]
A1 rabbit skeletal muscle/bovine calmodulin	39	[53]
Chicken crop/bovine calmodulin	45	[89]
Halocynthia roretzi/rabbit skeletal muscle	58	[273]
H. roretzi/chicken crop	57	[273]
H. roretzi/bovine calmodulin	37	[273]
Troponin C		
Rabbit skeletal muscle/chicken skeletal muscle	90	[53]
Rabbit skeletal muscle/*Rana esculenta* skeletal muscle	91	[53]
Rabbit skeletal muscle/bovine calmodulin	51	[53]
Rabbit skeletal muscle/carp parvalbumin	27	[53]
Calmodulins		
Bovine brain/*Paramecium tetraurelia*	90	[243]
Bovine brain/*Tetrahymena pyriformis*	92	[318]
Bovine brain/*Metridium senile*	98	[271]
Bovine brain/*Renilla reniformis*	98	[80]
Parvalbumins		
Rana esculenta α/β	51	[115]
R. esculenta α/rabbit α	69	[115]
R. esculenta β/*Cyprinus carpio* β	68	[53]
Gadus merlangus β/*C. carpio* β	69	[53]
Raja clavata β/*C. carpio* β	50	[53]
Calbindins		
Pig/rat	80	[52]
Pig/rabbit parvalbumin	17	[53]

ference between intestinal and skeletal muscle: despite differences in length, the essential LCs from the crop, heart and skeletal muscle of the chicken have 91 out of 150–192 amino acids in common.

The LCs also show adaptation to different functional requirements by the existence of tissue-specific **isoforms**, which are either encoded by separate genes or arise by alternative splicing of the same primary transcript. The essential LCs of all mammals are probably encoded by at least four genes. The two essential LCs in the fast-twitch fibres arise from the same gene, LC-1/3, which has two promoters and produce two polypeptides with different N-termini by alternative splicing. The slow-twitch muscles contain two isoforms of essential LCs, one of which also appears in ventricular muscle. The atrial muscle expresses the same gene as embryonal skeletal muscle. The essential LCs in human smooth muscle and non-muscle cells differ only in five N-terminal amino acids out of a total of 151; this is also the result of alternative splicing. Vertebrate regulatory LCs are also encoded by four genes which are expressed in a tissue-specific manner in skeletal, heart and smooth muscle and in non-muscle cells [71, 164].

The **type I myosin** molecules are fundamentally different from the filament-forming molecules of type II. In the former, there is only one heavy chain, usually of only 100–140 kDa, and this is therefore significantly smaller than the paired HCs of type II. Four myosin-I HCs from the soil amoeba *Acanthamoeba castellanii* have been investigated in detail. Also in this case, the HC N-terminal region together with one or two LCs forms a head domain which bears the actin- and ATP-binding sites and can develop ATPase activity. The amino acid sequence of the N-terminal, 80-kDa domain of this HC is about 50 % similar to the S-1 subfragment of myosin II. The HC tail region of the type-I myosin has a completely different structure and function compared with that of type II. Myosins IA, IB and IC show 60–75 % sequence agreement in this HC region. The tail has its own actin-binding site, distinct from that of the head domain, which allows it to maintain contact with one actin molecule whilst translocating to another. The tail region also has a membrane-binding site, which binds the myosin to the plasma membrane or allows the transport of vesicles along an actin filament [129]. The unusually large tail region of the 177-kDa HC of the high molecular weight myosin I (HMWMI) from *A. castellanii* has very little in

common with the smaller myosin-I HCs [111]. The ATPase activity of myosin I is only activated by actin when a specific serine or threonine residue of the HC is phosphorylated by a specific protein kinase (MHCK). The MHCK can bind 7–8 moles of phosphate per mole of enzyme by autophosphorylation, which is promoted by phosphatidylserine but not by calcium [27]. Only a few proteins comparable with type-I myosin from *A. castellanii* are known from other animals. The product of the ninAc gene, expressed in the eyes of *Drosophila melanogaster*, has a C-terminal domain homologous to the head and tail of type-I myosin [129]. The brush-border myosin I (BB-MI), which is present in the microvilli of the gut epithelia and binds the central actin-filament bundle to the plasma membrane, apparently has a motor function. BB-MI from the chicken includes a heavy chain with a myosin-like globular head of 82 kDa and a 37-kDa tail domain which bears no resemblance to the conventional myosins [91].

The **myosin genes** sequenced so far show large variation in their exon/intron organization: the HC genes from vertebrate skeletal-muscle myosins mostly have 41 exons, the HC gene unc-54 from *Caenorhabditis elegans* has 8, and the HC gene of the muscle myosin of *Drosophila* has 9 exons. In *A. castellanii*, the HC gene of myosin II has only 2 very short introns, whereas the HMWMI gene has 17. The mammalian LC-1/3 gene has nine exons, in contrast to the six exons of the essential LC of *Drosophila*. The regulatory LC of mammals also has more exons (seven) than does that of *Drosophila* (three); the last exon of the fly corresponds to exons 3–7 of the rat [65, 71, 111, 189].

10.4 Actins

Actin is a globular protein of 43 kDa. The polypeptide chain consists of 374–376 amino acids and is always N-terminal acetylated. The amino acid sequence is very conservative. Several acidic amino acids at the N-terminus are variable, but otherwise the animal actins vary at most by 5 %. There are also multiple actins in plants which, in contrast, vary both amongst themselves and compared with those of the animals by more than 10 %. The **actin monomer** is divided into two lobes, between which is found the ATP-binding site; one lobe is formed from the N-terminal 150 amino acids and the other is formed from the remaining 225 amino acids; each domain consists of one β folded sheet surrounded by several α helices. Under physiological conditions, the actin monomers (G-actin) aggregate to form **actin filaments** (F-actin). Actin in the muscle is almost entirely in the F form, but in non-muscle cells both G- and F-actin are found. The arrangement of the monomers in an actin filament is usually described as a right-hand-threaded double helix with a pitch of 71.5 nm. Considering the tight binding between the monomers of both strands, the filament can also be seen as a simple left-handed helix with a pitch of 5.9 nm [224].

Polymerization and depolymerization of the actin monomers depends in a complex way on the external conditions, and is influenced especially by ATP, ADP, Mg^{2+} and Ca^{2+}. Because all the monomers in a filament are orientated in the same direction, the filament ends vary. In view of the arrowhead forms, which result from the binding of myosin head pieces (HMM), the ends are referred to as "pointed" (P) or "barbed" (B). The association rate is greater at B ends than at P ends, and is greater with ATP-actin than with ADP-actin. At a certain (critical) actin concentration, the association (particularly at B ends) and dissociation (particularly at P ends) are in equilibrium, and the filament length remains constant. The critical actin concentration depends upon the actin type and the external conditions and is usually 0.1–1 μmol/l [30, 224].

At least six **tissue-specific isoactins** can be distinguished in mammals and birds; these can be classified into three groups, α, β and γ, according to increasing pI values in isoelectric focusing. There are four isoforms of the muscle-specific α actins. All skeletal muscles contain the same α actin and this is therefore the only muscle protein that shows no differences amongst the various fibre types. Further α actins are found in heart muscle and in the smooth muscle of the stomach and blood-vessel walls; β and γ actins are typical of non-muscle cells. The sequences of these different actins are very similar; the α actins differ in only 4–6 out of 375 amino acids, and β and γ actins differ from each other at only 4 positions and from α actin at 25–27 positions [188]. Species-specific differences between actins of one type are rare in the warm-blooded vertebrates; for example, the α-actin sequences of cattle, rabbit, rat and chicken are identical, as are the γ actins of man, the mouse and rat [186, 288]. Only three amino acid substitutions separate the skeletal-muscle actins of agnathans and mammals [288]. Thus, the actins are changing no quicker than histone H4, which is considered to be a model of slow evolution (see Table 4.12, p. 161).

As is seen from the sequence of appearance of the different mRNAs, individual actin genes are activated in a stage- and tissue-specific manner. In *Xenopus*, γ actin is expressed from the beginning of development, whereas α skeletal muscle and α heart appear only at gastrulation, and only in mesodermal cells which later give rise to muscles. Embryonal muscles express a third α-actin gene [188]. Each amphibian species has its own set of up to eight tissue-specific, non-muscle actins [50]. Chicken myoblasts synthesize only β and γ during the division phase, with α appearing at the beginning of fusion. No β- or γ-actin mRNA is present in the complete myotubule. The heart type predominates initially in skeletal muscle, e.g. there is 80% in the chicken leg muscle at the 11th day of development and still several percent in the adult animal. Conversely, human, bovine and porcine heart ventricles contain up to 20% skeletal-muscle actin [290]. The different **isoactins** can be clearly classified on the basis of their N-terminal amino acid sequences (Fig. 10.3). Comparison of these sequences shows that the actins of invertebrate muscles are much more similar to the non-muscle than to the muscle actins of the vertebrates. Invertebrate muscle actins and the β and γ actins of vertebrates have an N-terminal cluster of three acidic amino acids, whereas vertebrate-muscle α actin always has four. This surprising similarity between invertebrate muscle actins and vertebrate non-muscle actins is confirmed by the complete sequences of several invertebrate actins: e.g. the muscle actins of the scallop *Pecten* spp. and the sea urchin *Centrostephanus robertsii* differ from the α actin of mammalian heart in 19–20 amino acids, and from mammalian β actin in only 9–10 amino acids [288]. The specific muscle actins of the vertebrates arose relatively early in the evolution of the Chordates; the actins of the ascidians and the lancelets have the four N-terminal, acidic amino acids that are typical of mammalian muscle. In the tunicates and acranians there is probably only one muscle actin; different actins in smooth and striated muscles are clearly distinguishable in the cartilaginous and bony fish. From the reptiles onwards, the two muscle actins developed further: to skeletal muscle and the heart type in the striated muscles, and to stomach and blood-vessel types in the smooth muscles [224, 288].

The actin of the dysentery amoeba *Entamoeba histolytica* is so similar to mammalian actin that it

Various invertebrates:							
Acanthamoeba sp.			G	D	E	V	Q ...
Pecten sp.			D	D	E	V	A ...
Caenorhabditis elegans			D	D	E	V	A ...
Strongylocentrotus purpuratus			D	D	E	X	X ...
" " (egg)			D	D	D	V	A ...
Centrostephanus robertsi			D	D	E	V	A ...
Chordata:							
Tunicata: Cyela clavata	D	D	E	E	X	X	...
Acrania: Branchiostoma sp.	D	D	E	E	X	X	...
Agnatha: Lampetra fluviatilis							
Skeletal and heart muscle	D	D	E	E	T	T	...
Elasmobranchii: Mustelus must.							
Skeletal and heart muscle	D	D	E	E	X	X	...
Stomach and aorta	D	E	E	D	X	X	...
Liver		D	D	E	X	X	...
Urodela: Pleurodeles waltlii							
Skeletal and heart muscle	D	D	E	E	X	X	...
Stomach and aorta	D	E	E	D	X	X	...
Anura: Xenopus Laevis							
Skeletal and heart muscle	D	D	D	E	T	T	...
	D	D	E	E	T	T	...
Stomach and aorta	D	E	E	D	X	X	...
Reptilia: Calotes emma							
Skeletal muscle	D	E	D	E	X	X	...
Heart muscle	D	D	E	E	X	X	...
Stomach and aorta		E	E	E	X	X	...
Birds and mammals:							
Skeletal muscle	D	E	D	E	T	T	...
Heart muscle	D	D	E	E	T	T	...
Stomach		E	E	E	T	T	...
Aorta	E	E	E	D	S	T	...
β-Cytoplasmic		D	D	D	I	A	...
γ-Cytoplasmic		E	E	E	I	A	...

Fig. 10.3. N-Terminal sequences of muscle and cytoplasmic actins from invertebrates and vertebrates [68, 288, 289]. In every case acetylated; X represents an unidentified neutral amino acid

can copolymerize with the actin monomers from rabbit skeletal muscle. In contrast, the *Acanthamoeba* actin is unique in several ways: it has N-terminally a blocked residue, and not aspartic or glutamic acid as in all other known actins (Fig. 10.3); there is an N-trimethyl-lysine residue at position 326 [289]. In addition to this actin, which makes up about 10 % of the total cytoplasmic protein, *Acanthamoeba* has two others, of which one only appears to be present in the nucleus. The most deviant actins are found in the ciliates. For example, actin from *Tetrahymena thermophila*, with a sequence of 275 amino acids, is 25 % different to yeast actin, and even 35 % different to the small actin from *Oxytricha fallax*. Vertebrate actin differs from yeast actin by only 13 %, from the actins of higher plants by 11–14 %, from *Acanthamoeba* actin by 10 %, and from those of *Oxytricha* by 29–33 %. In addition, the *Oxytricha* actin has only 356 amino acids, lacking the 18 amino acids which appear at positions 68–85 in the other actins. The great evolutionary distance of their actins highlights the special position of the ciliates within the eukaryotes [51, 95, 101, 104].

Actin isoforms are so far known from only a few invertebrates. Two electrophoretically distinct actins can be detected in the flight muscles of the Diptera and Hymenoptera; in *Drosophila*, the actins of the flight muscles, leg muscles and larval abdominal muscles show tissue-specific differences. Five genes identified in the sea urchin *Strongylocentrotus purpuratus* code for tissue-specific cytoskeleton actins [62]. Even though no actin isoforms are detected in most invertebrates, there are usually several **actin genes** present; these are either dispersed in the genome (as in *Drosophila*) or ordered in clusters (as in the sea urchin). Like yeast, *Tetrahymena* has only 1 actin gene, but there are 3 in the ciliate *Euplotes* and in *Acanthamoeba*, 4 in *Caenorhabditis*, 5–6 in various insects, 5–8 in echinoderms, 15–17 in the slime mould *Dictyostelium*, and 50–75 in the roundworm *Ascaris* [51, 62, 101, 147, 309]. *Drosophila melanogaster* has six actin genes, which are expressed in a stage- and tissue-specific manner, including four which are muscle specific [176]. The silkworm *Bombyx mori* also possesses at least five actin genes, three of which have been sequenced. One of these three genes is expressed in the silk glands in actin filaments involved in the ejection of the silk thread. The coded sequence is almost identical to that of the cytoplasmic actin from *Drosophila*. The other two sequenced genes are muscle specific and differ in 8 of their 374 amino acids [196].

The sea urchin *Strongylocentrotus purpuratus* has at least six variously expressed actin genes. Five genes arranged in two clusters encode the cytoplasmic actins; two have been sequenced. Just one isolated gene codes for a muscle-specific actin in the sea urchin, whereas in insects, as in the vertebrates, there are always several muscle-specific actins. The sea urchin muscle actin is more similar to the cytoplasmic actins of vertebrates than to their muscle actins; in one region of the gene there is clear evidence for a conversion event with one of the cytoplasmic-actin gene clusters [49]. The six genes are expressed to varying degrees, e.g. in the pluteus larva there are 60–190 mRNA molecules for the cytoplasmic actins compared with 650–1300 for the muscle actin. The gene for muscle actin is expressed relatively late in development and only in certain mesoderm cells. The five genes of the cytoplasmic actins fall into three groups, one or two of which are expressed in each embryonal cell line, but never all three [48].

There are about 20 actin genes in the human and murine genomes; if genes which differ by more than 20 % are included, then the number in the mouse may be much higher. Other mammals, like the golden hamster, appear to possess only five or six actin genes [218]. Why are there so many actin genes in some mammals and yet only six different proteins? The actins of human skeletal and heart muscles are each encoded by only one gene, as is the cytoplasmic β actin. The most plausible explanation is that among the human and murine actin genes there are probably many pseudogenes which are never expressed. The **arrangement of actin genes in the genome** covers the whole range of possibilities: tight linkage in *Dictyostelium*, dispersed in *Drosophila*, and tandem clusters in sea urchins and nematodes [68]. The human and murine genes for the different α actins are not linked [3]. In general, in the mouse the active genes for the various muscle proteins are widely dispersed in the genome; the genes for myosin HC are on chromosome 11, for regulatory myosin LC on chromosome 7, for the skeletal actin on chromosome 3, for heart muscle on chromosome 17, and for the non-muscle actin on chromosome 5.

About two dozen actin genes have been sequenced to date. In parallel to the great similarity of the encoded proteins, the actin genes themselves show such broad similarity that the genes of widely separated organisms can hybridize. Of course, synonymous substitutions have occurred in the **evolution of the actin genes**. However,

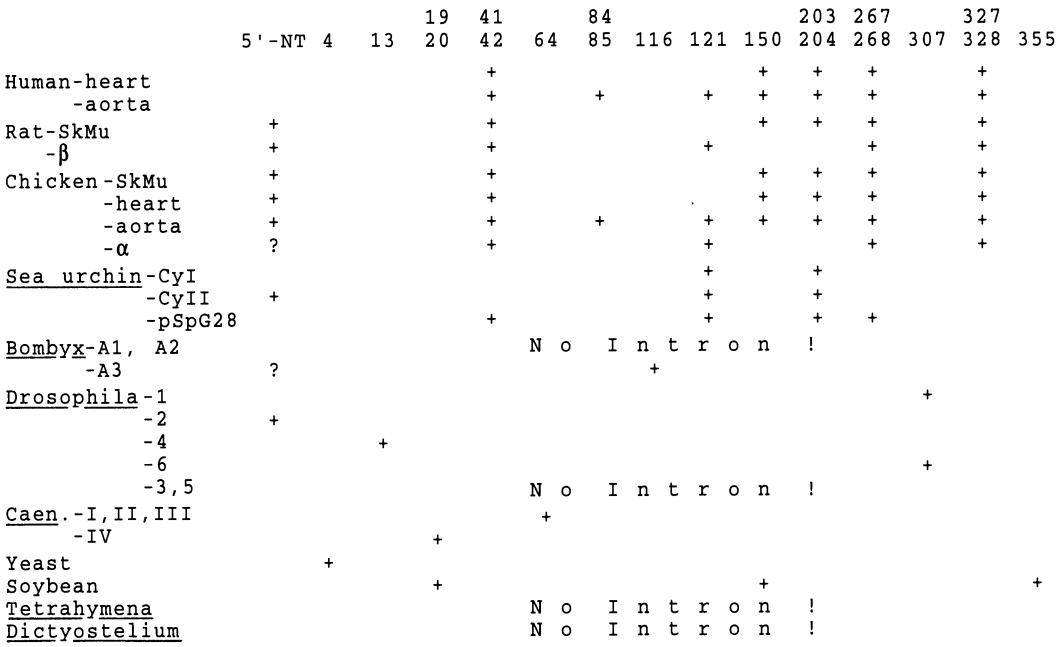


Fig. 10.4. Actin genes show large differences in the number and position of the introns [32, 49, 51, 196]. The positions of the introns between two codons (two numbers) or within a codon (one number) are given at the *top* of the figure

comparisons of isotypic actin genes from different mammalian species or from closely related genes of one mammalian species show, quite surprisingly, that the substitution rate in the 5′- and especially in the 3′-non-translated (NT) region is much less than the synonymous substitution rate of the coding region. Thus, for example, in the mouse the genes for α skeletal muscle, α heart muscle and β actin, with 7% difference in their amino acid sequences, have 13.8–16.5% synonymous substitutions in the coding sequence, but only 5.6–8.7% in the flanking NT regions. It appears that a particularly strong selection pressure affects this region of the genes in connection with differences in the expression and regulation of the individual isoactin genes [3, 34, 202, 315].

The **introns of the actin genes** vary greatly in number and position between the genes of different species, and even between the isoactin genes within one genome (Fig. 10.4). In this respect, the actin multi-gene family is unique: in all other well-known gene families the introns always occupy the same relative positions, with occasional introns being absent. The variability of the actin introns may have arisen by the loss of certain ancestral introns in different evolutionary lines. Of course, the other possibility is that introns were inserted later in evolution. The fact that the number of introns in the lower eukaryotes is consistently relatively low speaks in favour of the second explanation. Both the length and sequence of introns change very quickly during evolution, and even neighbouring genes often show great differences [32, 49, 51, 145, 196]. In contrast to the situation for the other muscle proteins, **alternative splicing** of the primary transcript has so far been described for only one actin gene: at least four mRNAs are produced from the one gene for chicken aorta actin and these vary in the length of the region [32]. The **biosynthesis of actins** exhibits some special features in the formation of the N-terminus. Initially, the initiator methionine is always acetylated and removed; the acetylation of the now terminal aspartate residue completes maturation in the case of the non-muscle actins. In the case of the muscle actins, the cysteine following the methionine is also acetylated and removed before the exposed aspartate or glutamate can be acetylated [32, 202, 224].

10.5 Actin-Binding Proteins

The **actin concentration in the cells** is several hundred micromolar, and therefore exceeds that of any other protein. At this concentration, actin should rapidly polymerize and exist almost

entirely as F-actin. In muscle this is in fact the case, but in non-muscle cells the actin filaments are constantly turned over and are in **dynamic equilibrium** with the free G-actin monomers. The regulation of this equilibrium as well as the functional organization of the actin filaments and their reorganization during normal cell function is apparently the task of a complex group of proteins found particularly in non-muscle cells in association with actin monomers and filaments. The structural and functional properties of the actin-binding proteins have by no means been determined in detail. More than 60 such proteins from various tissues have been described and more are being added to this list. Many of the known actin-binding proteins are probably ubiquitous but comparative biochemical data are not yet available. As there are as yet very few sequence data, nothing can be said at present about the evolution of these proteins, and the number of protein super-families involved cannot be estimated. It can be assumed, however, that every cell contains representatives of different families of actin-binding proteins [224, 265].

According to the types of interaction with actins, the actin-binding proteins can be ordered into **function classes** with some overlap. Many of these proteins bind actin monomers and thus inhibit polymerization. The "capping" proteins deposit themselves on only one end of the filament, usually the B end. Finally, there are proteins which bind along the length of the actin filaments and these are also of different sorts: one binds like tropomyosin to only one filament, whilst others are divalent and form filament networks or bundles [224]. In the especially well-investigated cells of *Acanthamoeba*, one or more representatives of all classes have been found [46].

The best-known monomer-binding proteins are the **profilins**, which have been detected in several isoforms in various mammalian cells (spleen cells, macrophages, brain cells), sea urchin eggs, sperm of the holothurian *Thyone*, as well as in *Tetrahymena, Acanthamoeba* and the slime mould *Physarum*. Profilins bind only to actin monomers, and not to dimers or higher polymers, and have a greater affinity for ATP-actin than ADP-actin. Their inhibitory effect on polymerization cannot involve reduction of the monomer concentration alone; the dissociation constant of the actin-profilin complex is so high that even in the presence of profilin the monomer concentration exceeds the critical concentration. Sequence data are available for human profilin, for a basic profilin from calf spleen (142 amino acids), and for two acidic and one basic isoform from *Acanthamoeba* (all with 125 amino acids). The profilins contain trimethyllysine but have no other similarities to actin [4, 64, 152, 224, 276]. Further monomer-binding proteins have been isolated from the amoeba: these include the homodimer actobindin (two units of 13 kDa), which in contrast to the *Acanthamoeba* profilin also reacts with rabbit actin, and the monomeric actophorin (15 kDa), which is not related to profilin and unlike the latter protein can also bind F-actin [46, 156]. Cofilin (166 amino acids), from various mammalian and chicken tissues, and the homologous destrin also bind F-actins in addition to actin monomers [2, 194].

Most **capping proteins** bind to the B end of the filament, but the exact mechanism of their effect is not clear. These proteins may inhibit association or accelerate dissociation at the occupied end. In particular, they may bring about the de novo synthesis of shorter filaments with occupied (capped) ends; several of them have a filament-severing effect. The biological significance of these actin-binding proteins may lie in the stabilization of a certain filament length, preventing the otherwise constant process of addition to the B end and reduction of the P end (the "actin treadmill"). The intracellular concentration of the capping proteins is sufficient to occupy all the available actin filaments [224]. There are apparently three proteins families in this functional class:

1. The capping proteins in the narrowest sense; these are heterodimers with subunits of 30–35 kDa, known not only from *Acanthamoeba*, slime moulds and vertebrate non-muscle cells but also from vertebrate skeletal muscle [29, 224].
2. The fragmin/severin family (monomers of 45 kDa) is found in slime moulds, thyroid cells and sea urchin eggs.
3. The gelsolin/villin family (monomers of 90–95 kDa) is as yet known only in the vertebrates.

Proteins from the latter two families are regulated by calcium and also have filament-cleaving properties. Brevin from the blood plasma is encoded by the same gene as that which encodes the intracellular gelsolin, but it has N-terminal extensions of 9 amino acids in the pig and 25 amino acids in man [151, 301]. The proteins binding to the P end of actin filaments are not so well known. They include, for example, acumentin from macrophages (63.5 kDa), β actinin from mammalian skeletal muscle (35–37 kDa), and also cytoskeleton proteins such as spectrin and the band-4.1 protein (p. 360). Smaller filament-cleaving proteins

(15–20 kDa) have been isolated from protozoans, echinoderms and vertebrates, but in the absence of sequence data little can be said about their relationships to each other [46, 224].

The proteins that bind sideways on a single actin filament include, in particular, tropomyosin. The rod-shaped tropomyosin molecule is monovalent, i.e. it cannot make either bundles or lattices from actin filaments, but it probably serves to stiffen them and increase their mechanical strength [114]. Bundling of actin filaments is observed in microvilli and some other cytoplasmic structures. The **actin-bundling proteins** include the fascins, found as monomers of 57 kDa in *Acanthamoeba* and slime moulds, in the eggs, sperm and coelomocytes of echinoderms, in *Limulus* sperm and in mammalian brain; fimbrim (68 kDa) from various mammalian cells; and villin (95 kDa) from the microvilli of the brush-border cells in the gut and kidney [15]. **Actin cross-linking proteins** were discovered through their drastic effect on the viscosity of actin solutions. The filamins of smooth muscle and the "actin-binding proteins" (ABPs) from non-muscle cells are homodimers (two units of 250 kDa) which carry an actin-binding site on each subunit, and hence have filament-bundling or filament-cross-linking activity [106, 224]. There are two **types of thin filament** in the smooth muscle of the chicken crop; one consists of actin, tropomyosin and caldesmon, and the other of actin, tropomyosin and filamin. The biological significance of this heterogeneity is not known [163].

Specific structural proteins link the actin microfilaments of the cytoskeleton to the plasma membrane. For example, the actin filaments of the fibroblasts are bundled into so-called stress fibres which end in adhesion plaques on the plasma membrane. The stress fibres contain periodically arranged α actinin in addition to filamin, and the plaques contain vinculin (130 kDa) and talin (215 kDa). Vinculin and talin are also found in muscle [125, 229, 302]. Desmosomes and other structures linking epithelial cells contain placoglobin (83 kDa), which is also found in cells that lack desmosomes [123]. The α **actinins** were discovered in vertebrate striated-muscle cells as a component of the Z lines, but they are also present in the functionally equivalent "thick bodies" of smooth muscles, in the non-muscle cells of vertebrates, and in sea urchin eggs and slime moulds. They are homodimers of 100-kDa subunits with almost 900 amino acids [6, 183, 224]. It has only recently become clear via sequence comparisons that α actinin has homology to the spectrins of the membrane skeleton. The spectrins are composed of 106-amino-acid repeats; three such repeats with significant (25 %) sequence similarity to spectrin make up the C-terminal domain of chicken α actinin. In contrast, the N-terminal, actin-binding domain has no similarity to spectrin [24, 298].

10.6 Tropomyosins and Troponins

The **tropomyosins** are α-helical polypeptides with a length of 284 amino acids and a mass of 34–46 kDa; they occur in the muscles of vertebrates and invertebrates, and may be smaller in non-muscle cells. Usually, two equal or different chains are intertwined to give a rod-like dimer that is 40 nm long (coiled-coil structure). Tropomyosin has been detected in all muscle types of the vertebrates, and also in the muscles of various invertebrates and in many non-muscle cells. The thin filaments of vertebrate skeletal muscle contain actin monomers, tropomyosin and the troponin complex in the ratio 7:1:1. The tropomyosin molecules lie in the grooves between the actin strands of the thin filaments, and each carries one troponin complex. This conformation is responsible for the actin-specific calcium regulation of the contractile apparatus. Tropomyosin is also found in myosin-specific regulated muscle and non-muscle cells, but is then not associated with a troponin complex; here, it appears to have the function of increasing the mechanical strength and stiffness of the thin filaments [114]. The **tropomyosin genes** have quite different lengths (rat α-TM 28 kb; rat β-TM 10 kb; *Drosophila* Tm1 22 kb and Tm2 28 kb), but always have exons of either 21 or 42 codons. One large and two small exons encode each of the actin-binding sites, of which there are seven regularly distributed over the whole length of 284 amino acids [289]. The shorter tropomyosins from equine blood platelets (247 amino acids), human fibroblasts (247 amino acids) and *Drosophila* non-muscle cells (252 amino acids) extend over only six actin monomers [92].

Just like other muscle proteins, tropomyosins exist as several **isoforms**, which are either encoded by different genes or derived by alternative splicing of the same primary transcript. Additional heterogeneity can arise by post-translational phosphorylation. There are two tropomyosin subunits, α and β, in striated and smooth vertebrate muscle; these are encoded by different genes but differ only slightly in their sequences. They mostly exist as αβ heterodimers

but can also form the two different homodimers. Each gene produces several tropomyosin isoforms, the exact number of which is not known, even in the case of the common laboratory mammals. Mammals also express the α form of the fast-twitch skeletal-muscle fibres in the heart, whereas birds have heart-specific isoforms. Different isoforms are also found in non-muscle cells. The production of isoforms by the β-TM gene of chicken is a good example of alternative splicing; at least nine different mRNAs are produced by the use of two promoters and alternative splicing of several of the 13 exons. The mRNAs of all muscle tropomyosins include exons 1–5 (amino acids 1–188) and 8–9 (amino acids 214–257), whereas exon 6 (amino acids 189–213) and exon 11 (amino acids 258–284) are found only in smooth-muscle mRNA, and the analogous exons 7 and 10 exist only in skeletal muscle. An additional exon, 1′, between exons 2 and 3 encodes, together with exons 3, 4, 5, 6A, 7, 8 and 9A, a non-muscle-cell tropomyosin of 247 amino acids [72, 166]. The species-specific and isoform-specific differences between the tropomyosins involve 15–25 % of the whole sequence but are mainly located at the C-terminal end. This region is responsible for dimer formation and the binding of actin and troponin, and the tropomyosin isoforms thus show considerable differences in these properties.

Multiple tropomyosins are also known from *Drosophila*. There are at least three muscle-specific isoforms, of which two are encoded by the same gene, Tm1. Differential splicing of exon 4 leads to two muscle-type-specific isoforms with a length of 284 amino acids, which differ only in 27 C-terminal amino acids. The third muscle-specific tropomyosin is encoded by the gene Tm2 and, as the result of a substitution in the stop codon, has a length of 285 amino acids. Alternative splicing produces a further non-muscle-cell tropomyosin isoform from the same gene with a length of 252 amino acids [92]. A tropomyosin from the locust *Locusta migratoria* agrees by 51–55 % in its sequence of 283 amino acids with various vertebrate tropomyosins, and by 87 and 64 % with Tm1 and Tm2 from *Drosophila*, respectively [148]. Tissue-specific tropomyosin isoforms have also been found in other invertebrates, e.g. there are up to five in various mussels, and two in the muscles of the earthworm *Lumbricus terrestris*. There are at least six tropomyosin isoforms in the xiphosuran *Tachypleus tridentatus*, but these are distributed equally in all tissues, i.e. tissue specificity is not so marked in this case [114, 187].

The **troponin complex** consists of three non-covalently linked proteins: Tn-T as the tropomyosin-binding component; Tn-I, which inhibits actomyosin ATPase; and the calcium-binding Tn-C. All three types from the fast-twitch skeletal-muscle fibres of rabbit have been sequenced, and they constitute the most studied model of the relationship between structure and function in the troponin complexes. The N-terminal amino acids 1–70 of Tn-T bind in an antiparallel fashion to the C-terminus of one tropomyosin molecule and to several N-terminal amino acids of the subsequent molecule. The C-terminal region of the Tn-T binds to a second site, which is 20 nm from the first site, on the tropomyosin molecule. The contact sites between Tn-T and the components Tn-C and Tn-I also lie in this region. Binding of Ca^{2+} to Tn-C disrupts the binding of Tn-T to the second site on the tropomyosin molecule and causes a conformational change which expose the attachment site of the myosin head on actin. In addition to that in the vertebrates, a complete troponin complex has been found in several invertebrates, e.g. arthropods, molluscs and the roundworm *Ascaris lumbricoides* [130]. Some non-muscle cells contain only Tn-C, and as it cannot be involved in A regulation in the absence of Tn-I and Tn-T, its function in this case is not clear.

The **Tn-T** from the muscles of the rabbit, rat and chicken have very different lengths of between 259 and 301 amino acids; Tn-T is, as a rule, longer in arthropods and mussels than in the vertebrates. Thus, in Tn-T evolution there have been insertions and deletions of larger DNA segments as well as nucleotide substitutions [25]. Isoforms of Tn-T have been detected in vertebrates of all classes and even in the ascidian *Halocynthia roretzi* [206]. Like other muscle proteins, the troponins of the vertebrates show large differences in different muscle types. Thus, in mammals and birds, there are different isoforms of Tn-T in heart muscle and in fast-and-slow-twitch skeletal muscle, and these are encoded by different genes. There are also pseudogenes in the Tn-T gene family. The number of isoforms of all tissue-specific Tn-Ts can be further increased by alternative splicing. In the fast-twitch skeletal-muscle fibres of the rat, 64 Tn-T isoforms arise by the alternative use of exons 4–8 and 16–17 [25]; the situation is very similar for the Tn-T variants in chicken skeletal muscle [249]. In the case of the heart-specific Tn-T gene of the chicken, only one of the two mRNAs formed receives the transcript of exon 5 and is translated to the 301-amino-acid isoform I of Tn-T, which predominates in the

embryonal heart and early embryonal skeletal muscle. It may be that the acidic decapeptide encoded by exon 5, which is missing from isoform II (291 amino acids) from adult heart, is only required for the early synthesis of sarcomeres. In fact, the heart-specific Tn-T gene of the chicken includes amongst its 18 exons the shortest known exon sequence: exon 17 is only 17 bp long. In the rat Tn-T gene the corresponding sequence is included in the last exon; the number and position of the introns does not completely agree in the two Tn-T genes [47].

Alternative splicing complicates comparison of the Tn-T sequences of different species. The Tn-T from the skeletal muscles of the rat and rabbit differ mainly in the N-terminal amino acids 1–36; the amino acids 37–259 are identical except at four positions. The differences between skeletal-muscle Tn-Ts of mammals and birds are also mainly restricted to the terminal regions [25]. The Tn-T from rabbit heart muscle (276 amino acids) differs in particular in three regions (amino acids 17–46, 152–169 and 237–249) from that of skeletal muscle, but is in fact very similar to chicken heart Tn-T in just these regions; it is possible that this part of the polypeptide chain is essential for heart function [215].

The **Tn-I** is also found in heart muscle, as well as in both fast- and slow-twitch skeletal muscle in different isoforms with up to 40 % differences in sequence. The Tn-I from the skeletal muscle of the chicken has a length of 182 amino acids [203], whilst Tn-I from rabbit heart bears an N-terminal extension of 26 amino acids; the phosphorylation of the 20-Ser within this segment by a cAMP-dependent protein kinase reduces the calcium affinity of the troponin complex. In normal heart, Tn-I is only 30 % phosphorylated, but after β-adrenergic stimulation it increases to 100 %. Only one Tn-I has been sequenced from invertebrates, from the crayfish *Procambarus clarkii*. It has an acetylated N-terminus, contains two trimethyllysine residues and shows only 20–25 % agreement to vertebrate Tn-Is in its sequence of 201 amino acids; in contrast, the latter show more than 60 % agreement with each other [140].

The **Tn-C** belongs, like the myosin LC, to the protein super-family of the calmodulin-like, calcium-binding proteins, and bears the typical four calcium-binding sites. In the Tn-C of fast-twitch skeletal-muscle fibres, only binding sites I and II are available for calcium regulation, whilst sites III and IV are occupied by Mg^{2+} under physiological conditions. In the Tn-C from slow-twitch muscle fibres and heart muscle, only Ca^{2+}-binding site II is functional. The heart and slow-twitch skeletal-muscle fibres of the rabbit contain the same Tn-C, and this differs in 56 of its 181 amino acids (35 %) from that of the fast-twitch muscle fibres. Despite the identity of the Tn-Cs, the troponin complexes in heart and slow-twitch muscle fibres have differences in their calcium affinities, determined partially by the different Tn-Is present. The Tn-Cs show relatively slow evolution: rabbit and bovine heart Tn-Cs differ in only one amino acid; the skeletal muscle Tn-C of the rabbit differs from that of man by one amino acid, and from that of the chicken and frog *Rana esculenta* in 13–15 of about 160 amino acids (Table 10.1). The crayfish *Procambarus clarkii* has two Tn-C isoforms which differ in 21 of 150 amino acids, i.e. they are apparently encoded by different genes. The giant barnacle *Balanus nubilis* also has two Tn-Cs of 151 and 158 amino acids and these agree by only 61 %. These are the only invertebrate Tn-Cs to have been described in detail, and they have only one-third of their sequences in common with the Tn-Cs and calmodulins of the vertebrates. They have only two functional Ca^{2+}-binding sites [44, 141]. The annelids, arthropods, molluscs, nematodes and some other invertebrates have all been shown to possess Tn-Cs, but investigations of the arthropod and mollusc molecules show only 0.75–1.3 Ca^{2+} ions are bound in vitro compared with 3–4 in the vertebrates [35].

10.7 Paramyosins

Paramyosins are rod-shaped molecules with a length of 130 nm and a diameter of 2 nm. Like the LMM region of the myosins or the tropomyosins, they consist of two α helices associated into a coiled-coil; the two chains are identical and have a molecular mass of 100–110 kDa. At low pH and in the presence of a high concentration of divalent cations they form paracrystals; paramyosin is easily identified from its characteristic axial periodicities of 72.5 and 14.5 nm. Higher or lower amounts of paramyosin are to be found in probably all invertebrate muscles, and it has been detected in, for example, molluscs, annelids, xiphosurans, insects, nematodes, trematodes and cestodes; it is absent from vertebrate muscles and non-muscle cells [103, 238]. The paramyosin gene of *Caenorhabditis elegans* (unc-16) has been sequenced. It codes for a protein of 868 amino acids which has the heptad repeats of hydro-

phobic amino acids and the 28-mer repeats of charged amino acids that are very typical of the myosin heavy-chain rod. The paramyosin molecule, however, differs from the myosin HC in that it lacks the terminal non-helical regions [122].

All the muscles of molluscs contain paramyosin, with the highest amounts being found in the smooth shell adductors, foot muscles and byssus retractors of the mussels. The thick filaments in these **paramyosin muscles** are unusual in their thickness and length; they consist of a paracrystalline nucleus of paramyosin bearing myosin on its surface. The ratio of paramyosin to myosin lies between 2:1 and 10:1 [33, 238]. The paramyosin molecule, like the HC tail region of myosin, can be reversibly phosphorylated by a cAMP-dependent protein kinase; in both cases this results in increased solubility [33]. Functionally, the paramyosin muscles are noted for the "catch" mechanism, by which a high mechanical tension can be maintained over long periods with minimal energy consumption. This phenomenon can be dramatically demonstrated with mussels, which can keep their shells closed despite many weeks of applied tension of several kg/cm^2 muscle cross-section. The catch mechanism does not involve the active generation of tension, but more a rigor-like state which is relieved by, for example, serotonin in intact fibres and cAMP in the case of sarcolemma-free fibres. The molecular mechanism of the catch system has not yet been described. One set of hypotheses assumes it to be related to the mechanical properties of the paramyosin-myosin complex and to be regulated by phosphorylation, although a connection between an increase in phosphorylation and release of the catch has by no means been established. Other hypotheses seek to explain the mechanism by changes in the kinetics of bridge formation between thick and thin filaments, i.e. increased stability or reduced breakage of these bridges. Unambiguous evidence is also lacking in this case, and neither hypothesis explains why the catch phenomenon depends upon the presence of large quantities of paramyosin [33, 238, 299].

10.8 Calcium-Binding Proteins

10.8.1 Calmodulins

The protein family to which the light myosin chains and troponin C belong also includes a whole series of proteins from muscles and non-muscle cells which have the capacity to bind Ca^{2+} and Mg^{2+} with high affinity. The prototype of these calcium-binding proteins is **calmodulin**, which was discovered at the beginning of the 1970s as the activator of certain phosphodiesterases and is now known as the universal calcium receptor of eukaryote cells [80, 263]. Calmodulin has been found in all eukaryote cells examined, from the mammals to the lower fungi, and even in several bacteria [149]. Particularly high levels are found in mammalian brain and testis as well as in the electric organ of the electric eel *Electrophorus electricus*, from which up to 1 g calmodulin per kg organ can be obtained.

Calmodulin is made up of a polypeptide chain of 16.7 kDa with 148 amino acids, including approximately 50 aspartate and glutamate residues which confer the acidic nature of the molecule (pI = 4). The N-terminus is always blocked, and many calmodulins have a post-translationally modified N^{ε}-trimethyl-lysine at position 115. The calmodulin molecule has the shape of a dumbbell, with two globular regions, each bearing two calcium-binding sites, connected by a long α helix [8]. The amino acid sequence has a distinct internal periodicity with four homologous, calcium-binding domains, I–IV, each of 30 amino acids. They have a characteristic tertiary structure referred to as EF-hand, consisting of a calcium-binding loop of 12 amino acids flanked by two α-helical sections arranged more-or-less perpendicular to each other (Fig. 10.5). The sequence similarity is especially high between domains I and III, and II and IV. It seems reasonable to assume that this structure arose by the double duplication of a gene for a primitive calcium-binding eukaryote protein. It is clear, however, that the intron/exon structure of the gene is not entirely compatible with this concept, as the introns often break up Ca^{2+}-binding domains. The evolution of the calmodulin gene has apparently been somewhat complicated [204, 217].

The calmodulins of different animal species are structurally and functionally very similar, as is shown already by analysis of the amino acid composition. Vertebrate calmodulins contain two tyrosines, whereas in the invertebrates there is always only one. The trimethyl-lysine that is characteristic of most calmodulins is missing in green plants, *Dictyostelium* and *Neurospora*, and also in *Trypanosoma*, the crayfish *Orconectes limosus* and the insects *Bombyx mori, Drosophila melanogaster* and *Mamestra configurata*. The amino acid sequence of *Drosophila* calmodulin, obtained via the cDNA, has a lysine in position

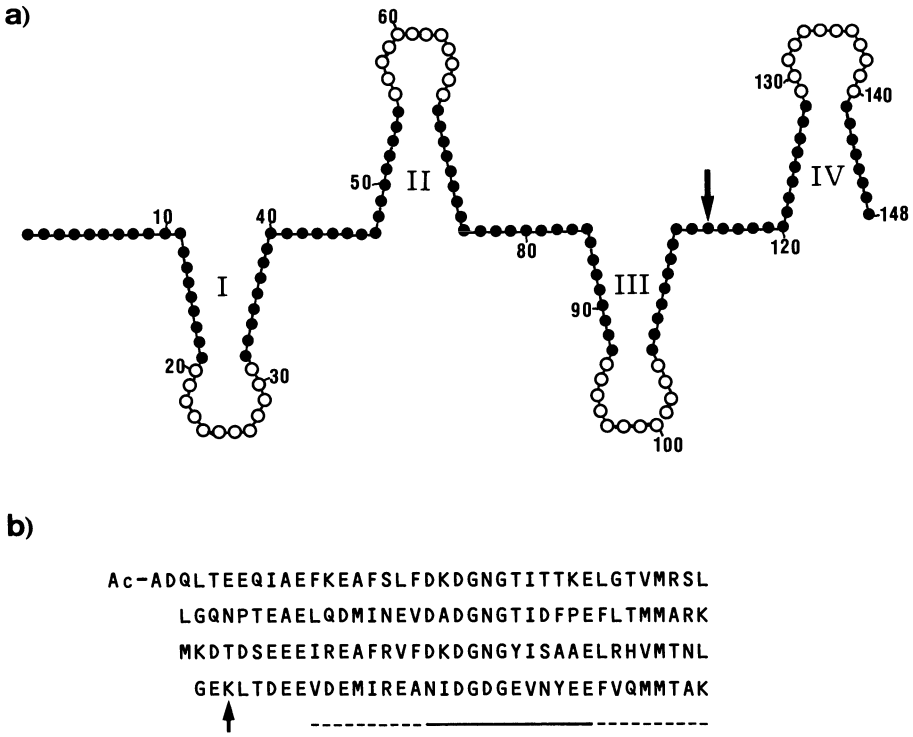


Fig. 10.5a, b. The structure of calmodulin from bovine brain. The *arrows* show the position of N-trimethyl-lysine residues [263]. **a** The internal periodicity of the molecule; four domains *(I–IV)*, each with a Ca^{2+}-binding loop of 12 amino acids *(open circles)*, are flanked by α-helical stretches of 8 amino acids and linkers of various lengths *(closed circles)*. **b** The sequence is arranged so that the four highly homologous Ca^{2+}-binding loops *(solid underlining)* and associated α-helical stretches *(dashed underlining)* lie above each other

115 but it is apparently not methylated. In contrast, the calmodulin of the Mediterranean fruit fly *Ceratitis capitata* does contain a trimethyl-lysine [23, 86, 193, 316]. Tissue-specific differences in the methylation of the 115-Lys are reported from the cephalopods of the genus *Octopus;* the central nervous system (lobus opticus) contains 1 mol trimethyl-lysine per mole calmodulin compared with 0.1 mol each of mono-, di- and trimethyl-lysine in the other tissues [23]. In the calmodulin of the ciliate *Paramecium*, not only is the 115-Lys trimethylated, but also the 13-Lys is dimethylated [243]. As is shown by the calmodulins of *Bombyx mori* and *Octopus*, the function of calmodulin as an activator of phosphodiesterase is independent of the methylation of 115-Lys [23]. The **amino acid sequences of the calmodulins** of many species are now known [7, 243, 248, 316], and in the case of various vertebrates, the gastropod *Aplysia californica, Drosophila melanogaster* and the sea urchin *Arbacia punctulata* we also know the gene sequences [94, 204, 245, 251, 269]. Calmodulin is a very conservative protein in terms of its evolution. All mammals, the chicken and the frog *Xenopus laevis* have identical calmodulin sequences. All known sequences from the plants and unicellular organisms to man differ at only 39 out of 148 positions; in the whole Metazoa, with the exception of the aberrant calmodulin of the chicken, there are only six variable amino acids [94]. The mammalian calmodulins differ from those of *Drosophila melanogaster* and the cnidarians *Metridium senile* and *Renilla reniformis* in only 3 amino acids, and even from those of the ciliate *Tetrahymena pyriformis* in only 12 positions and from *Paramecium tetraurelia* in 15 positions (Table 10.1). It seems that the universal calcium receptor, calmodulin, with its many functional restrictions, is subject to much slower evolution than are the calcium-binding proteins troponin C and parvalbumin, which are adapted to quite specific functions (Table 10.1).

The conservative evolution of the calmodulins is also demonstrated at the gene level; the cDNA of calmodulin from *Electrophorus* hybridizes not only with the DNA of various vertebrates but also with that from wheat germ. The number and structure of **calmodulin genes** differs quite widely between species. In some species, e.g. *Drosophila melanogaster*, there is only one calmodulin gene; two are found in the chicken, *Xenopus laevis* and humans, three in the flagellate *Trypanosoma brucei* and the xiphosuran *Limulus polyphemus*, and four in the rat. The coding sequence is divided into three exons in *Limulus*, four in *Drosophila*, five in the gastropod *Aplysia californica*, and five (gene CaM II) or six (genes CaM I and III) in the rat [60, 204, 216, 269]. Whereas the multiple calmodulin genes of the mammals all

code for the same amino acid sequence, the sea urchin *Arbacia punctulata* produces two quite similar isocalmodulins [94]. Retropseudogenes identified in the multi-gene family of the rat, humans and the chicken would code for calmodulins with quite deviant sequences, but are probably not expressed [142]. A protein of unknown function, isolated from the optic lobe of the squid *Loligo pealei*, is 68 % similar in its 149-amino-acid sequence to mammalian calmodulin, and contains four calcium-binding sites [96].

Ca^{2+}-calmodulin takes part in the regulation of enzymatic and other cellular reactions through involvement of the secondary messenger Ca^{2+}, either as a regulatory component of the enzyme itself (MLCK, phosphorylase kinase), or indirectly, and hence more slowly, by influencing the activity of protein kinases. Many enzymes which are directly or indirectly regulated by Ca^{2+}-calmodulin are simultaneously substrates of cAMP-dependent protein kinases; this is also true for MLCK. Furthermore, as an activator of cAMP phosphodiesterase, Ca^{2+}-calmodulin interacts with the cAMP secondary messenger system. Ca^{2+}-calmodulin also has a functional relationship with contractile mechanisms and the cytoskeleton. Caldesmon of the smooth muscle alternates, according to the calcium concentration, between calmodulin and F-actin with a sort of flip-flop mechanism, and the microtubule-associated tau factor does the same but with calmodulin and tubulin. Finally, calmodulin is involved in the regulation of calcium pumps in the plasma membrane and the endoplasmic reticulum membrane, and thus participates in the calcium homeostasis of the cell [238].

Despite the extensive structural similarities between all calmodulins, there are sometimes **differences in function**. Thus, for example, the guanylate cyclase of *Tetrahymena* is only activated by the calmodulin of this ciliate, although it differs at only 12 positions from other calmodulins [318]. The calmodulins of *Tetrahymena*,the mussels *Patinopecten* and *Mytilus*, and the fly *Drosophila* are much less active in the activation of mammalian brain phosphodiesterase than is mammalian calmodulin itself. Thus, is would appear that coevolution of calmodulin and the enzymes has occurred [86, 185]. In this connection, it is noteworthy that calmodulin from muscles of the migratory locust *Schistocera gregaria* can activate phosphodiesterase from bovine brain but not the enzyme from its own muscles.

10.8.2 Other Calcium-Binding Proteins

Apart from the calmodulins, essential and regulatory myosin-LCs and troponin C, muscles and non-muscle cells contain a series of further proteins which may be included in this same **protein superfamily** on the basis of their EF-hand structure: the parvalbumins of vertebrate muscle, the sarcoplasm calcium-binding proteins (SCBP) of invertebrate muscles, the large (28 kDa) and small (9 kDa) cholecalciferol (vitamin D)-induced calbindins of birds and mammals, the S-110 proteins, the calcium vector protein (CaVP) of the acranian *Branchiostoma*, the spec proteins of sea urchin embryos, the calcium-dependent proteases (calpains), the muscle protein α actinin, and the fluorescent protein aequorin. A phylogenetic tree, constructed by the parsimony method, of the 153 known EF-hand proteins shows that a least 12 evolutionary lines separated quite early and that further subfamilies probably exist [190, 223]. From comparisons of the calcium-binding domains, it seems likely that the common **ancestral form** had two calcium-binding sites, and independent duplications in different evolutionary lines led to proteins with four binding sites [213]. The parvalbumins lack the N-terminal 40 amino acids and, therefore, calcium-binding site I; domain II is no longer able to bind Ca^{2+}, and sites III and IV have increased affinities. Amongst the SCBPs of invertebrates are some with two or three functional binding sites: I and II in the scallop *Patinopecten*, I–II in the Crustacea, and I, III and IV in the polychaete *Perinereis* [272]. In troponin C, all four binding sites are present but only I and II are calcium specific; III and IV have changed so much that they preferentially bind Mg^{2+}. Site I of the heart type of troponin C can no longer bind Ca^{2+}. Amongst the myosin LCs, the regulatory LCs can bind Ca^{2+} at site I, whilst all four domains of the essential LCs have completely lost the ability to bind divalent ions. The 28-kDa calbindins of birds and mammals contain six EF-hand structures, of which four can bind calcium; in contrast, the 9-kDa variety from mammalian gut has only two calcium-binding sites. The two calbindins show no great similarity, either to each other or to other calcium-binding proteins, and quite clearly deviated early from the evolutionary line of the calmodulin-like proteins [7, 76, 112].

The **parvalbumins** are acidic cytosolic proteins of 108–110 amino acids and a molecular mass of 12 kDa. They were discovered in 1934 in frog muscle, and are found at the highest concentrations in the white muscles of fish, amphibians and

small mammals; they also occur at lower levels in the brain and other non-muscle tissues of vertebrates. They are hardly detectable in the heart and gut muscles [81, 241]. The chicken thymus contains a peptide that induces T-cell markers on cultured bone-marrow cells and is referred to, therefore, as avian thymic hormone. The sequence shows, however, that it is really a non-muscle parvalbumin [26, 211]. The parvalbumins in muscles have the function of a soluble relaxation factor: the calcium released from the sarcoplasmic reticulum in excited muscle binds initially to troponin C, thereby activating the actomyosin ATPase and triggering contraction. Although parvalbumin has a higher calcium affinity than troponin C, the excess of magnesium in the relaxed muscle results in it being fully loaded with magnesium, the dissociation of which requires a certain time. The calcium pump of the sarcoplasmic reticulum has an affinity even higher than that of parvalbumin, but it must first be activated. Muscle relaxation thus begins with the transfer of Ca^{2+} from troponin C to parvalbumin. After a certain delay, calcium is returned to the sarcoplasmic reticulum by the calcium pump, and the Ca^{2+} bound to parvalbumin exchanges with Mg^{2+} due to the declining Ca^{2+}: Mg^{2+} ratio [238].

As the relaxation rate of a muscle is mainly determined by the transfer of Ca^{2+} from troponin C to parvalbumin, the rate should be correlated to the parvalbumin concentration. In actual fact, comparison of different muscle fibres in an animal shows the highest parvalbumin concentration in type IIB (fast-twitch glycolytic), and in comparisons of the muscles of different mammalian species the highest values are found for the gastrocnemius of the mouse (4.9 g/kg) and the lowest values for the muscles of man and the horse (≤ 0.001 g/kg). The parvalbumin of rat brain and testis is identical to that in muscle. In non-muscle cells, parvalbumin may be involved in calcium homeostasis as a calcium buffer. More than 20 parvalbumin sequences from mammals, reptiles, amphibians, Crossopterygei and teleosts are known, and indicate a relatively rapid evolution (Table 10.1). Based on sequence characters and the isoelectric point, two different types of parvalbumin are recognized (α and β) and these can occur together in the same muscle [81, 241].

Instead of parvalbumins, invertebrate muscles contain **sarcoplasmic calcium-binding proteins** (SCBPs) with very heterogeneous structures; they have been reported, for example, from annelids, crustaceans, mussels and lancelets. Alone in the Crustacea, there are dimeric SCBPs with two 22-kDa chains and altogether six calcium-binding sites, four of which also competitively bind magnesium. In the decapods *Astacus leptodactylus, Homarus gammarus* and *Hippolyte zostericola*, two types of chain are found in different species-specific ratios, from which homo- and heterodimers with the same calcium-binding properties are formed [116]. The SCBP of the scallop *Patinopecten yessoensis* has a length of 176 amino acids, corresponding to 20.1 kDa; of the four binding sites present only I and III are functional. In the polychaete *Nereis diversicolor* and *Perinereis vancaurica SCBPs*, which are more-or-less the same size, domain IV can also bind Ca^{2+} bind Ca^{2+} [279]. In addition to calmodulin, the cilia of *Tetrahymena thermophila* contain two further calcium-binding proteins of 23 and 25 kDa, each with four EF-hand domains [43, 272]. The flagella of *Trypanosoma cruzi* and *T. brucei* have homologous calcium-binding proteins of 23 kDa [161].

Four sequenced SCBPs isolated from amphioxus (*Branchiostoma lanceolatum*) differ in only 1–7 of their 185 amino acids; they bear four EF-hand domains, of which I–III are functional calcium-binding sites, and one of them also competitively binds magnesium [275]. Amphioxus possesses a further calcium-binding protein of 161 amino acids with an acetylated N-terminus. The C-terminal half of the molecule, with the two functional calcium-binding domains III and IV, agrees 45 % with human calmodulin and 40 % with troponin C of rabbit skeletal muscle. The N-terminal half contains two non-functional Ca^{2+} domains and shows only 24 and 30 % similarity, respectively, to the two mammalian proteins. Because of its similarity to calmodulin, this protein is known as the **calcium vector protein** (CaVP) [139]. It builds a complex with a target protein (CaVPT) of 243 amino acids which can interact with thick myofilaments via two immunoglobulin-like domains [274]. The direct transduction of Ca^{2+} signals to an effector system has also been described for the S-100 proteins, which are widely distributed in the animal kingdom, and for the spec proteins of the sea urchin. The **S-100 proteins** make up a family of small Ca^{2+}-binding modulator proteins of 10–12 kDa; they were discovered in mammalian brain but have since been reported in many different animals from protozoans to man. In the mammals, they are homo- or heterodimers of α and β subunits, which show about 50 % sequence difference to each other, but contain two EF-hand domains each. They bind to various effector proteins and are involved in the process of cell differentiation and interac-

tions between the cytoskeleton and the plasma membrane. They also act extracellularly: the ββ homodimer is identical to the neurite extension factor (NEF) which stimulates neurite growth; the αβ heterodimer has no such activity [135]. The sea urchin *Strongylocentrotus purpuratus* has about eight **spec genes** which are expressed exclusively in the aboral ectodermal cells of the embryo. They encode about ten different spec proteins of 14–17 kDa with four EF-hand domains, which are produced during embryogenesis in the same cells but at different times and in different amounts. The spec1 protein is, for a time, one of the most predominant proteins of the embryo. Its precise function is not known; however, it appears to bind to actin and effect regular changes in the cell form. If the spec proteins really play an important role in embryo development, corresponding proteins would be expected in other sea urchin species. In fact, a protein of 34 kDa with eight EF-hand domains in *Lytechinus pictus* could have arisen by the duplication of a spec-like protein [93, 314].

The duodenum of birds and mammals contains calcium-binding proteins with a similar function but different structures; these are involved in vitamin D-dependent calcium absorption and are now mostly known as **calbindins**. The so-called 28-kDa calbindin, isolated from the gut wall of the chicken and sequenced via the cDNA, consists of 262 amino acids, corresponding to a molecular mass of 30 kDa, and includes six functional EF-hand domains [159]. In the gene there are 11 exons whose borders do not coincide with the 6 domains; 3 poly(A) signals are present and 3 mRNAs of different length are produced [306]. Sequence comparisons of human, bovine and chicken 28-kDa calbindin suggest a low rate of evolution of $0.3 \cdot 10^{-9}$ substitutions per amino acid per year [213]. Instead of the 28-kDa calbindin, the mammalian gut contains an equivalent molecule of 9 kDa. The small calbindin isolated from rat gut consists of 79 amino acids and possesses only two calcium-binding domains, which apparently correspond to domains I and II of calmodulin. The 9-kDa calbindin from porcine and bovine intestine differ by about 20 % from that of the rat [52]. The kidneys of mammals, birds, reptiles and amphibians express the 28-kDa calbindin, and only in the bat *Rousettus aegyptiacus* has a 9-kDa version been found. The 28-kDa molecule is also present in the cerebellum, and the 9-kDa form occurs in the placenta [209].

There are several families of calcium-binding proteins which do not belong to the super-family of the EF-hand proteins. The **calsequestrins** are the predominant Ca^{2+}-binding protein of the cisternae of the sarcoplasmic reticulum in the skeletal muscle of the mammals, birds and amphibians, and in mammalian heart. They bind 40–50 calcium ions per molecule with a relatively low affinity, and ensure that the Ca^{2+} concentration in the lumen of the endoplasmic reticulum does not exceed 1 mmol/l, thereby reducing the task of the calcium pump. Several such proteins have been sequenced via cDNA. The calsequestrin from the fast-twitch skeletal muscle of the rabbit is an aspartate/glutamate-rich glycoprotein of 63 kDa with 367 amino acids. The calsequestrins from the slow-twitch skeletal muscles and heart muscle differ in their glycosylation and molecular mass [317, 319]. The sarcoplasmic reticulum of smooth muscle and the endoplasmic reticulum of the non-muscle cells contain no calsequestrin; the predominant Ca^{2+}-binding proteins in this case are the **calreticulins**, which are located in their own storage compartments, the calciosomes. However, calreticulins are also found together with calsequestrins in the sarcoplasmic reticulum of striated muscle. Calreticulin from rabbit skeletal muscle consists of 418 amino acids, 109 of them acidic, and binds 43 Ca^{2+} ions with low affinity. It shows clear agreement with calsequestrin in two parts of its sequence [69, 184, 287]. Calsequestrin-like proteins have also been found in plant cells [146]. More recently, a further group of calcium-binding proteins has been discovered; these have no EF-hand structures in their sequences and have a high affinity for phospholipids. These proteins were originally given names like lipocortin, calpactin, calelectrin, endonexin, synexin, chromobindin and calcimedin; the members of the new super-family of calcium-binding proteins are now referred to as **annexins** [118]. The annexins are found in very different types of vertebrate cells, but have also been reported in *Drosophila* and are probably widely distributed in the animal kingdom. Some annexins are highly tissue specific; for example, annexin IV (32.5-kDa calelectrin) is found only in the epithelial cells of the gall and pancreas ducts, and annexin II (lipocortin 2) is found mainly in the microvilli. In contrast, annexin VI (67-kDa calelectrin) is found in all cells. The amino acid sequences of the smaller annexins (ca. 32 kDa) always show four repeats of about 80–85 amino acids, and in the larger annexins (ca. 67 kDa) there are eight such repeats [14, 82, 118, 138].

10.9 Microtubule Proteins

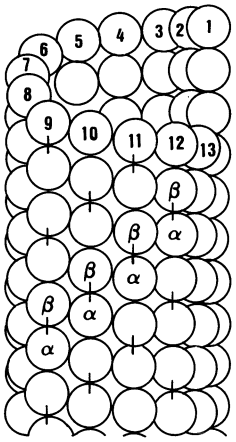


Fig. 10.6 Microtubles. One microtubule is composed of 13 rows of dimeric tubulin molecules (protofilaments), aligned such that they give the impression of a spiral structure. Each tubulin dimer consists of one α and one β chain

The **microtubules** are tube-like structures of 25-nm outer diameter and 15-nm inner diameter; they are normally made up of 13 rows of globular tubulin molecules. As the building blocks of each row (protofilaments) are shifted somewhat in relation to each other, a spiral structure results (Fig. 10.6). Some microtubules consist of more than or fewer than 13 protofilaments; e.g. there are 12 in the nerve axons of the crayfish *Procambarus clarkii* compared with 13 in the enveloping glial cells; there are 13 in bovine brain but 14 after their dissociation and reassociation in vitro; there are 14 in the brain of the shark *Scyllium*, 15 in the cortical organ of the guinea-pig, and 11 and 15 in the nematode *Caenorhabditis* [63]. **Tubulin molecules** are heterodimers of about 100 kDa made up of the polypeptides α and β, the binding of which in the dimer is so strong ($K < 1$ μmol/l) that they do not separate [244]. α and β tubulin are clearly homologous; in the microtubules of the pig and chicken they agree in 40–42 % of their 445–450 amino acids. A γ tubulin has been identified in the clawed frog *Xenopus laevis*, in plants and in lower fungi in the centrosomes, which are the origin of spindle formation in nuclear division; this tubulin agrees in only 35 % of its sequence with α and β tubulins [259]. In addition to tubulins, the microtubules often include large amounts of other proteins, the microtubule-associated proteins (MAPs) of 280–330 kDa, and smaller proteins of 55-70 kDa which are collectively called tau factors [41, 63, 244].

Above a certain critical concentration, the **tubulin dimers spontaneously polymerize** under physiological conditions to form microtubules. In the process, both ends may elongate or a directed movement can be generated by the extension of one end and simultaneous shortening of the other end. Microtubule formation requires the presence of Mg^{2+} and is inhibited by Ca^{2+}-calmodulin. Each tubulin dimer can bind two GTPs, whereby the one on the β chain is hydrolysed during polymerization. The **GTP cleavage** is, however, not essential for microtubule formation as polymerization/depolymerization of tubulin occurs close to thermodynamic equilibrium. MAPs are not necessary for **spontaneous microtubule formation** but, nevertheless, appear to have some modifiying effects [41, 63, 244]. As the process is driven by entropy, polymerization of microtubules in birds and mammals predominates in the temperature range 30–37 °C, and depolymerization predominates between 0 and 4 °C. A subpopulation of these microtubules is cold stable, even if no elongation occurs at the low temperature; this involves the activity of an associated protein (cold-stabilizing factor, 64 kDa). The tubulins from cold-adapted animals, e.g. various Antarctic fish, can polymerize at low temperature. The difference between these tubulins and those of fish, birds and mammals from temperate climates lies mainly in the α chain [54, 63].

The tubulins are evolutionally very conservative. Analysis of the sequences of 31 α and 31 β tubulins showed 48 % invariant amino acids, 23 % being found in both chains. Identical, or at least very similar, tubulins are found in very different mammalian and avian species. The largest known differences between tubulins of different eukaryotes are of the order of 62–63 % (α tubulin *Drosophila*/yeast; β tubulin mouse/yeast). Even the Protozoa, which are so heterogeneous at the molecular level, fit into this pattern. Comparisons with the β tubulins of the malaria parasite *Plasmodium falciparum* shows 94 % agreement with *Toxoplasma* and *Tetrahymena*, about 87 % with vertebrates and 72 % with yeast [21, 169, 303]. In view of the slow evolution of the tubulins, it is rather surprising that within the same organism, or even the same cell, they may show considerable heterogeneity. Multiple tubulins are apparently ubiquitous; they may be the product of different genes or arise by post-translational modification. Isoelectric focusing separates 7 α and 10–14 β tubulins from extracts of mammalian brain. The mouse possesses six functional genes each for α and β tubulins, from which in each case four or five different polypeptides are produced. In the chicken, the amino acid sequences encoded by the five active α-tubulin genes agree

by about 83–96 %. In the ciliate *Tetrahymena* and the flagellate *Trypanosoma* the cilia and the cytoplasm contain different tubulins [11, 63, 227].

Multiple tubulin genes are the rule in the eukaryotes. *Toxoplasma gondii*, on the other hand, possesses only one α- and one β-tubulin gene, and *Tetrahymena pyriformis* has one α and two β genes [11, 303]. The genome of the ciliate *Stylonychia lemnae* contains two each of the α- and β-tubulin genes, which are amplified 20000- to 150000-fold in the macronucleus [45]. *Drosophila* has 4 genes each for α and β tubulin, all of which are active; the sea urchin has about 13 of each, and mammals more than 20, although in this case they include many pseudogenes. There are three active human α genes and four β genes; in the mouse six of each are active, and in the chicken five of the seven to nine β genes are active [45]. The tubulin genes of the Metazoa are always widely dispersed in the genome, with the exception of several linked genes in the sea urchin. In contrast, *Trypanosoma brucei* and *T. evansi* have tubulin loci containing numerous αβ pairs in tandem; these loci are apparently transcribed as a unit, but are then rapidly cleaved into individual mRNAs with the typical mini-exons. The tubulin genes of *T. cruzi* and *T. rangeli* are not so strictly organized; in addition to tandem clusters, these species also have isolated αβ gene pairs [113, 172]. In the related species *Leishmania major*, there are separate α and β clusters and also dispersed tubulin genes [256]. Vertebrate α- and β-tubulin genes both contain three introns but in different positions. Comparisons within the eukaryotes show that α genes contain one to seven introns and β genes two to eight introns, the positions of which are extremely variable. Seventeen different intron positions have been reported for α-tubulin genes and 20 for β genes, only one of which is common to the two types [56, 303].

The **expression** of many tubulin genes is tissue specific and dependent on the developmental stage. This has been examined in detail in *Drosophila*. Here, there are four α and four β genes, of which, in each case, two are expressed constitutively, whereas α2 is active only in the testes of adult males, α4 in early embryos and the ovaries of mature females, one β gene only in the testes and early embryos, and another β gene only during the intermediate stages of embryo development. The sequence differences between these isotubulins are very large; α1 differs in only 2 amino acids from α3, but in 21 from α2, and in 149 from α4 [41, 283]. Several β-tubulin genes of the chicken have been sequenced and the encoded amino acid sequences show up to 8.7% difference. Tubulin cβ6 from erythrocytes differs, in fact, by 17% from the rest. β1 is the predominant tubulin of skeletal muscle, and β2 predominates in the brain, which is the only organ to contain β4. β3 is found only in the testis; β5 is everywhere except in neurons; and β6 occurs only in erythrocytes [199, 268]. In the mouse, two of the six α-tubulin genes are expressed only in the testis [292], and the blood platelets also contain a specific β tubulin [169]. The sequence differences between isotubulins are much greater than the differences between the same isoform in different species. The differences are not distributed randomly over the β chain; for example, the sequences between positions 401 and 425 in all β tubulins from mammals to yeast are identical, but the C-terminal regions from position 430 onwards are very variable in both sequence and length [41]. It would appear that the sequence differences between the isotubulins are the result of adaptive evolution, but so far very little is known about the functional differences between microtubules [41, 199].

The heterogeneity of tubulins is increased by **post-translational modifications** such as phosphorylation and glycosylation. Tyrosylation of the C-terminal glutamate residue of α tubulin by an ATP-dependent tubulin tyrosine ligase is very common, and is detectable in all mammalian cells [16]. Worthy of comment is the fact that most α chains are synthesized with a C-terminal tyrosine which is later cleaved off by a specific carboxypeptidase; this is an enzyme of 90 kDa which has been examined in detail only in the brain. Tubulin tyrosylation is found from the mammals to *Trypanosoma* but not, for example, in *Tetrahymena*. The ligases and tubulin from many different species, e.g. mammal and sea urchin, may be combined, but *Tetrahymena* tubulin is not processed by the mammalian enzyme [264].

Flagella and cilia (undulipodia) have a typical 9 + 2 structural pattern in cross-section: two central tubules are surrounded by nine peripheral doublet-tubules, each of which is made up of a complete A tubule and an incomplete B tubule; the A and B tubules have three protofilaments in common. The A tubules bear two rows of hook-like dynein arms (outer and inner) over their whole length, and these can form bridges to the B tubules of the neighbouring duplex (Fig. 10.7). The **dyneins** are energy-releasing ATPases, similar in this respect to myosin. They are protein complexes of 1.2–1.9 MDa, made up of two or three heavy chains (α, β and γ) of 350–430 kDa and with ATPase activity, and several medium (57–85 kDa) and light (6–22 kDa) chains. It is not yet understood why the dyneins are so much

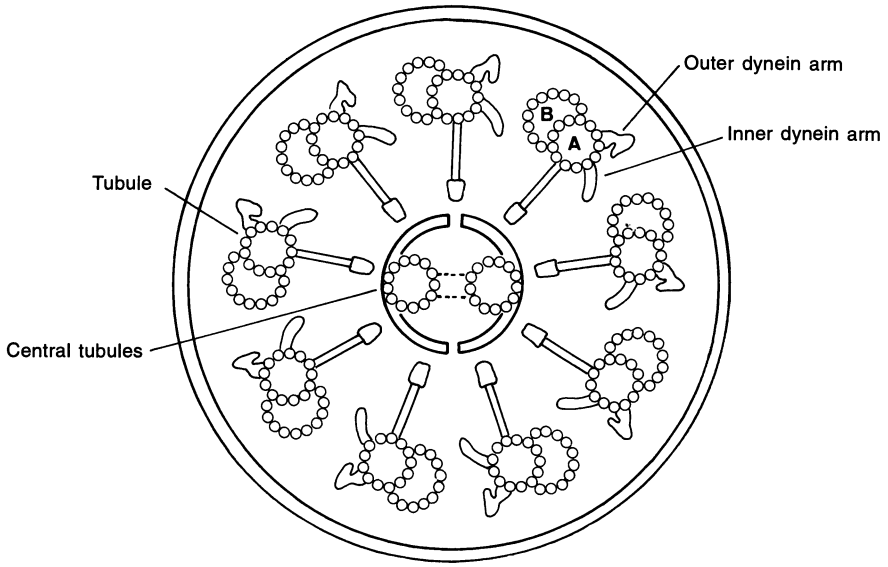


Fig. 10.7. Cross-section of a cilium. The nine outer duplex tubules are made up of the *A* tubule, with dynein arms, and the *B* tubule, which has three protofilaments in common with the *A* tubule (not five as shown in the figure!)

larger and more complex than the head region of the myosin molecule. Their large size has hindered analysis by the usual methods of protein chemistry. The outer dynein arms of the axonemes and the inner dynein arms, which may be present in two different forms, have basically the same structure but a very different polypeptide composition. The inner arms are not very well described, the best-known being the outer dynein arms from the cilia of *Tetrahymena*, the flagella of *Chlamydomonas* and the tails of sperm from sea urchins and trout [78, 83, 179]. Electron microscope pictures of the outer-arm dyneins show two or three globular heads arising from a common point on flexible stems. The number of heads corresponds to the number of heavy chains (two in sperm tails, and three in *Chlamydomonas* and *Tetrahymena*). In a beating axoneme, a gliding movement between the outer doublet tubules is translated by elastic elements into a transverse bending movement. An ATP-driven bridge cycle occurs between the dynein arm on the A tubule of a doublet and the B tubule of the neighbouring doublet. The heavy chains in *Tetrahymena* may carry up to six phosphate residues, with the phosphorylation presumably having some regulatory function [36].

Cytoplasmic dyneins appear to be ubiquitous, even occurring in *Caenorhabditis*, which, like all nematodes, has no cilia or flagella. The dyneins detected in the cytoplasm of sea urchin eggs probably represent, on the one hand, the maternal contribution to cilia formation in the larva, but perhaps also have egg-specific functions such as cytoplasmic movement. A further microtubule motor protein, **kinesin**, was discovered in the squid giant axon in 1985; this protein appears also to be widely distributed. The best-investigated kinesin, i.e. that from bovine brain, has a mass of about 400 kDa and consists of two heavy chains of 135 kDa and two light chains of 60 kDa. The heavy chains form two heads with ATP- and tubulin-binding sites, and a stem-like tail with a coiled-coil structure which bears the light chains on the C-terminal end [144, 179]. A family of at least 30 genes has been found in *Drosophila* and codes for various kinesin-like proteins of as yet unknown function [66]. In axons, the kinesins are apparently the driving force for the transport of cytoplasmic particles away from the soma (anterograde), and the cytoplasmic dyneins are responsible for movement in the opposite direction (retrograde). The dyneins and kinesins are possibly joined together in a common supramolecular structure, in which the opposing movements can be switched on or off as required [83]. Neither of the two proteins appears to be involved in the interactions between the cytoplasmic microtubules. This is apparently the responsibility of another protein, **dynamin**, which can unite microtubules in an ATP-dependent manner into extensive bundles. A polypeptide of 851 amino acids has been isolated as part of this newly discovered microtubule-dependent ATPase; its sequence shows no similarity to the proteins of the kinesin super-family [205, 246].

The **microtubule-associated proteins** are a very heterogenous mixture of proteins of which, as yet, only two types have been described in any detail. The high molecular weight MAPs are

molecules of more than 250 kDa. These include MAP1 with 330–350 kDa and MAP2 with 280–300 kDa; MAP1 is bound to three light chains of 34, 31 and 18 kDa in the bovine brain. The second group of MAPs are the tau factors, which are tissue- and stage-specific proteins of 55–62 kDa that include in their sequence of 341–383 amino acids three or four typical, C-terminal 31-residue repeats [85, 150, 160, 244, 305]. In the parasitic flagellate *Giardia*, the protofilaments formed from microtubules bear band-like protein structures. The predominant protein of the bands, β giardin, consists of 259 amino acids whose sequence includes 35 seven-residue repeats (heptads) and forms an α-helical coiled-coil [110].

10.10 Proteins of Intermediary Filaments

In addition to actin microfilaments and tubulin, the cytoskeleton of almost all eukaryotic cells contains a third type of filament which, with diameters of 8–12 nm, are larger than microfilaments (5–6 nm) but smaller than microtubules (23–25 nm), and are therefore known as **intermediary filaments** (IF). In contrast to the actin filaments and microtubules, these are stable structures which are not subject to constant switching between elongation and shortening. On the basis of their occurrence in cells of different types and of the properties of their protein components, six **classes of IF proteins** are recognized:

1. About 30 different cytokeratins of 40–70 kDa in epithelial cells and their derivatives, and also in liver cells.
2. Desmins (52 kDa) in different muscle types.
3. Vimentins (53 kDa) in cells of mesenchymal origin.
4. Glia fibril acidic proteins (GFAPs) of 45–55 kDa in glial cells (astrocytes).
5. The triplet of the neurofilament (NF) proteins: NF-L (68 kDa), NF-M (150 kDa) and NF-H (200 kDa).
6. Lamins (60–80 kDa) from the lamina of the nucleus.

A special feature of the vimentins is that they occur in embryonic or tissue-culture cells together with, or instead of, the cell-specific IF proteins. Thus, vimentin is found with desmin in embryonic myotubules, and together with cytokeratins in cultures of epithelial cells. The cells of the mammalian germline are amongst the few cell types without IFs; in contrast, the oocytes of amphibians contain both vimentins and cytokeratins [87, 212, 244].

Despite their many differences, the IF proteins have some common characteristics; in particular, they are all weakly soluble under normal physiological conditions, and both in vivo and in vitro they form morphologically similar filaments. Copolymeric filaments of desmin and vimentin not only arise in vitro but are also found naturally in the wall cells of blood vessels and in the myoblasts of chicken embryos. The basic structural unit of the IF is apparently a complex of four chains which, in pairs, form two **coiled-coil** structures. All IF proteins have in their centre a conservative α-helical rod domain of about 310 amino acids, flanked by variable, non-helical domains. The central domains are made up of four segments of uniform size (1A, 1B, 2A and 2B), which consist of seven-amino acid repeats (heptads) of similar sequences and which produce the coiled-coil structure of the dimers, and three linkers of varying lengths (L1, L12 and L2), which have no tendency to form coiled coils (Fig. 10.8).

In order for two α-helical peptide chains to form a coiled-coil, they must be built repetitively of heptads with non-polar amino acids in the first and fourth positions. IF proteins can be categorized into four **types** according to their heptad sequences and the structure of their terminal domains. The cytokeratins can be divided into acidic (type I) and basic (type II); vimentin, desmin and GFAP make up type III and the NF proteins type IV [74, 244, 261]. The amino acid sequences of type-I cytokeratins have less in common with type II (26–28 %) than with vimentin and GFAP (36–37 %); in contrast, desmin, vimentin and GFAP agree in 60–70 % of positions [9, 136]. It was only recently discovered that the lamins A and C from the lamina of the cell nucleus also belong to the super-family of the IF proteins. They also form dimers with a coiled-coil structure; however, these do not associate into long filaments. Because of differences in the structure of the rod domains, lamins A and C do not fit into any of the classes I–IV, and they are designated type V [180, 214].

All the known **IF genes** occur as only one copy in the genome. However, several mRNAs arise from the chicken vimentin gene by use of at least three poly(A)-regions, which all code for the same protein [261, 320]. The genes for cytokeratins, desmin, vimentin and GFAP have similar structures with eight or nine (cytokeratin I) exons; the genes for NF-L with four exons and NF-M with three exons have a very different

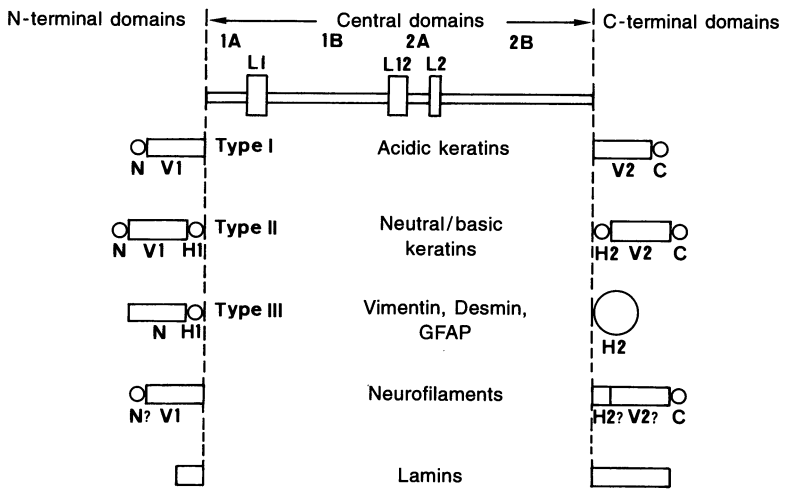


Fig. 10.8. The proteins of the intermediary filaments (IF) [261] and the homologous lamins A and C from the nuclear envelope [180] have comparable domain structures. All IF subunits are composed of one central α-helical rod domain and two non-helical end domains. The rod domain consists of four segments (*1A, 1B, 2A* and *2B*) which, in all IF subunits, are made up of the same sequence of seven amino acids (heptads) and three linkers; the length of the linkers varies in different IF proteins: *L1* consists of 8–14 amino acids, *L12* 16–17, and *L2* always 8. In the tetrameric IF components, the rod domains of two antiparallel subunits form a coiled-coil structure. The end domains of different IF proteins each have a specific organization of subdomains

organization [9, 121, 165, 200, 320]. In the **evolution of the IF proteins**, the coiled-coil-forming rod domain probably arose first, with the non-helical end domains appearing later. The internal periodicity of the rod domains is evidence for their creation by repeated duplication of heptads; at least four heptads are needed for the formation of a coiled-coil structure. The most ancient IF proteins were probably lamin-like. The NF proteins were apparently the first to branch from this line of evolution, followed by the ancestral forms of desmin, vimentin and GFAP [22]. Immunological and biochemical investigations of molluscs, annelids and nematodes suggest that invertebrate IF proteins have a consistently lower level of complexity; however, two main types can be distinguished: neuronal and non-neuronal. NF proteins have been detected in various invertebrates. Of the non-neuronal IF proteins, all cases investigated in detail have shown two related forms which can each form homopolymeric IFs. These have been sequenced in the snails *Helix pomatia* and *H. aspersa*, and the nematode *Ascaris lumbricoides*. They contain a typical coiled-coil rod domain, with a segment 1B that is six heptads longer than in vertebrate IF proteins, and is thus quite similar to lamin. This finding fits the concept that lamins are the ancestral IF proteins [58, 300]. Vimentins, desmins, all three NF proteins and some cytokeratins include several **phosphorylation sites**, which can be phosphorylated by specific, cAMP-dependent or Ca^{2+}-calmodulin-dependent protein kinases. The biological significance of this IF phosphorylation is not yet clear [244]. As in the case of the actin filaments and microtubules, the IFs also have **associated proteins**, which in some unknown way influence the formation and function of the filaments. For example, the desmin filaments of the muscles are associated with synemin (230 kDa) and paranemin (280 kDa), which can also bind to vimentin. Filaggrin from epithelial cells, which binds keratin filaments to macrofibrils, consists in fact of a whole family of basic, histidine-rich proteins of various sizes (in the rat 49 kDa, in humans 35 kDa and in the mouse 26 kDa) which arise from a 250-kDa polyprotein. Reduced or increased expression of the corresponding intron-less gene is the cause of many human skin diseases [74, 244].

10.10.1 Cytokeratins

All horny substances are usually known as **keratin**. However, some such substances include additional components from other protein families, in addition to IF proteins. The IF protein families themselves are therefore better termed cytokeratins. Electrophoretic and immunological methods or in vitro translation of the mRNA of higher vertebrates has revealed about 30 different cytokeratin polypeptides encoded by just as many genes.

The variation amongst the keratins is increased by cell-specific phosphorylation [244]. This heterogeneity of the proteins corresponds to the morphological and evolutionary plasticity of horned epidermal structures. Usually, the distinction is drawn between the soft keratins in vertebrate skin and hard keratin structures like hair, feathers, spines, claws, nails and hooves, horns, beaks and whalebone. The formation of specialized keratin structures in various parts of the epidermis first appeared after the transition of the vertebrates to a terrestrial existence, i.e. with the reptiles, birds and mammals; only a few exceptional amphibians have keratinized warts, claws or horned jaws [20].

The **keratinocytes** in the stratum germinativum of human epidermis have up to 30 % of their total protein in the form of prekeratins, which can be extracted into acidic buffers. In the space of a few days, these living cells are converted into the dead, horny cells of the outer skin layer (stratum corneum), in which 70 % of the dry weight consists of keratins; here, they are cross-linked into a fibrous state by disulphide bridges and N′-(γ-glutamyl)-lysine isodipeptide bonds, and can be solubilized only by a combination of reducing and denaturing reagents. The formation of the cross-links is catalysed by keratinocyte-specific transglutaminases which occur as different isoenzymes [221]. In the course of its development, each keratinocyte expresses a series of different cytokeratins. Cytokeratins 1 and 10 predominate at the start of terminal differentiation. These form a highly complex aggregate of cross-linked fibrils beneath the plasma membrane, the cross-linked cell envelope (CE). Construction of the CE involves the histidine-rich, 26-kDa protein **filaggrin**, the sequence of which is very similar between the rat and mouse but is quite different in man. Special granules in immature keratinocytes contain a high molecular weight (>350 kDa) pro-filaggrin, made up of numerous filaggrin copies that are bound together by linkers and phosphorylated on many serine residues. During the first phase of terminal differentiation, the pro-filaggrin is dephosphorylated and then processed further by specific proteases [237]. In addition to cytokeratins, the CE of terminally differentiated, flattened keratinocytes contains further, cross-linked proteins, e.g. the small pancornulins, keratolinin, loricrin [109, 222], and **involucrin**, the structure and evolution of which has received particular attention. A large part of the involucrin chain consists of a segment of tandem repeats, and this varies in position and repeat length between different mammals. In the dog and pig and in the prosimians *Lemur catta* and *Galago crassicaudatus*, there is a "premodern" segment, located at a site P near the N-terminus, which has 6–13 repeats each of 20–31 amino acids. The higher primates possess a "modern" segment of 10-residue repeats located at a site M nearer the C-terminus. In the tarsioids, which lie close to the roots of the modern apes, there is in fact a P segment with 14 repeats of 17–18 amino acids, and also an M segment. The modern segment of the higher primates can be divided into three regions, of which the early region is present in all species, the middle region is found only in the hominoids, and the late region shows large differences between man, gorilla and orang-utan [57]. The continuing evolution of the late region is indicated by the distinct polymorphism of this region in the gorilla [282].

As well as in the ectodermal epidermis cells, **cytokeratins** are also detectable in other multilayered epithelia, for example, in the mouse in the mucous membranes of the palate and cheeks, on the tongue and in the lining of the oesophagus and the cardia portion of the stomach, i.e. in cells of endodermal origin [137]. The cytokeratins have the typical structure of IF proteins with a central rod domain, which can form a coiled-coil structure, and a non-helical terminal domain. Two types of cytokeratin can be distinguished. Type I includes the acidic proteins (pI 4.5–5.5) of about 40–60 kDa, and type II the neutral to basic proteins (pI 6.5–7.5) of about 60–70 kDa. The charge differences are due to the large proportion of basic amino acids in segments 1A and 2B of the central domain; the size difference results from the absence from type I of the terminal subdomains H1 and H2 (Fig. 10.8). The keratin filaments are I/II heteropolymers, with the cells forming one or more specific I/II pairs at each step in their differentiation [262]. The genes of both cytokeratin types are arranged in separate clusters [226]. The immunological similarity within one type in different species is much greater than between types I and II in the skin of one animal. Hybridization experiments with cDNAs for the keratin mRNAs synthesized in human epidermis cell cultures have shown that genes for both types are present in all vertebrates from *Myxine* to man. Distantly related sequences have even been found in the *Drosophila* genome. Little is known about the structure of cytokeratins of the lower vertebrates; nevertheless, a non-epidermal, 56-kDa cytokeratin from the oocytes and embryos and from cells in the liver, gut and kidney of the clawed frog *Xenopus laevis* has been sequenced via the

cDNA; this protein is highly homologous to mammalian keratins of type II [73].

The **hard keratins of the mammals** are complex structures which contain other, unrelated proteins in addition to cytokeratins. The best-investigated example is **sheep wool**. The keratinocytes in the hair follicles of the sheep produce filaments or fibrils from low-sulphur (LS) proteins that have a mass of 45–58 kDa and only 5 % cysteine. In contrast to the skin cells, the filaments of the hair-follicle cells are embedded in a matrix of proteins from two further families: the high-sulphur (HS) proteins (11–26 kDa and up to 30 % cysteine), and the glycine/tyrosine-rich (HGT) proteins of 6–9 kDa and with up to 30 % glycine and 15 % tyrosine. The compressibility of the keratin structures depends upon the proportion of matrix proteins. For example, the ratios of LS:HS:HGT proteins in sheep are 62:29:9 in the wool, 76:11:13 in horn and 79:11:10 in the hooves [236]. The composition of mammalian hair, and hence its mechanical properties, differs between species and races, and is also dependent upon external and internal conditions. The LS proteins are similar in amino acid sequence and tertiary structure to the cytokeratins. In sheep wool, there are two groups, each of four LS proteins, which apparently correspond to the types I and II of the cytokeratins [254]. Two HGT proteins and 18 HS proteins have been sequenced so far; they are not related to each other or to the cytokeratins. The genes of both protein families contain no introns.

The differences between the hard keratins of the mammals and those of both reptiles and birds are so fundamental (Table 10.2) that the nomenclature originally formulated for the mammals is worthless in a discussion of comparative biochemistry. In the mammals, filaments of about 7-nm diameter and made up of low-cysteine, predominantly α-helical proteins, are embedded in a non-filamentous matrix of cysteine-, glycine- and tyrosine-rich proteins; these keratins are known as α-keratins. The **hard keratin structures of the reptiles and birds**, on the other hand, contain filaments that are only 3 nm in diameter and are predominantly made up of a protein with a β-sheet structure; this type is termed φ- or β-keratin [84].

The best-investigated examples of β-keratins are the **feather keratins** of the chicken and other bird species. The feathers and the scales of the leg arise from very similar embryonal sources; it is assumed, therefore, that feathers evolved from scales. A comparison of the feather keratin gene B and the scale keratin gene III from chicken embryos shows that about 75 % of the sequence

Table 10.2. Amino acid composition in mol % of various keratins. Only the amino acids showing characteristic differences are given

	1	2	3	4	5	6	7	8
Ala	4.8	7.0	6.1	6.1	0.8	8.4	3.4	3.3
Asp	4.8	5.9	3.7	4.1	2.9	0.8	0.8	1.6
Cys	13.0	0	0	0	0	5.1	22.0	6.6
Glu	3.4	7.0	3.2	3.8	0.7	0.5	2.3	0
Gly	28.2	15.6	33.9	30.8	40.1	6.0	7.0	23.0
Leu	3.6	7.8	5.4	5.8	5.3	12.8	2.9	3.3
Phe	2.3	3.0	3.7	3.6	5.0	3.2	1.6	9.8
Pro	9.1	11.1	9.6	9.4	3.9	1.4	11.2	6.6
Tyr	4.1	3.9	9.4	9.0	18.9	3.2	2.3	18.0

1, *Varanus gouldii:* claw outer layer [84]; 2–4, Kookaburra bird *Dacelo novaeguineae* [84]: 2, feather; 3, beak; 4, claw; 5, echidna *Tachyglossus aculeatus:* spine [84]; 6–8, sheep wool [53]: 6, low-sulphur (LS) protein; 7, high-sulphur (HS) protein; 8, glycine/tyrosine-rich (HGT) protein

can be aligned; 70 % of the amino acid sequences encoded by these regions coincide. The main difference between the two keratins is that the scale keratin has a glycine-rich repetitive sequence of 4 × 13 amino acids that feather keratin lacks. The scale keratin, therefore, has a molecular mass of about 15 kDa, compared with 10.5 kDa for the feather keratin; the claw and beak keratins are also about 14–15 kDa. Both feather and scale keratins consist of a central region with a β structure and terminal regions in which almost all the cysteines are to be found. Neighbouring filaments are bound to each other by non-polar interactions between the central regions and by disulphide bridges between the tail regions to give a two-dimensional grid (feather) or a three-dimensional network. The feather keratins are encoded by a multi-gene family consisting of about 20 genes [228, 304]. The hard keratin structures of the reptiles correspond in their molecular structure to those of the birds. Consequently, the claws of the monitors and other lizards are morphologically very similar to those of the mammals, but are biochemically completely different. In their amino acid composition with its high glycine content, the keratins of reptile claws resemble more the keratins of bird claws and beaks than the feather keratins (Table 10.2). Additional components of reptile claws are tyrosine-rich, low-cysteine keratins with an α-helical central domain similar to that of the mammals [84].

10.10.2 Proteins of Other Intermediary Filaments

Vimentin filaments were discovered in fibroblasts but also occur in other cell types. In some cells, vimentin is found along with cytokeratin, desmin

or GFAP; gene expression is therefore not as strictly regulated as for other IF proteins. There is apparently only one vimentin gene in birds and mammals, and from this several mRNAs are transcribed with varying 3′ regions [244]. The evolution of the vimentins is relatively slow, e.g. the sequences of chicken and hamster agree in 80 % of positions [320]. The species-specific differences lie mainly in the N-terminal, non-helical domains, as is commonly found for IF proteins. In addition to the filamentous, insoluble vimentins, some cells also contain soluble, tetrameric vimentins at lower concentrations [253]. **Desmin filaments** are found in many different types of vertebrate muscle. Extraction of the contractile proteins by use of high-ionic-strength buffers leaves behind desmin in the Z lines of striated muscle, in the intercalated discs of heart muscle, and in the dense bodies of smooth muscle. In embryonic muscle, desmin filaments are found as a cytoplasmic network [244].

The **IF proteins of the glial cells** (GFAPs) vary in size in different species; e.g. human 49 kDa, bovine 54 kDa, and murine 45 and 55 kDa [244, 283]. In contrast, the **neurofilaments** of mammals, birds and reptiles apparently always contain the same **protein triplet** of NF-L (62 kDa), NF-M (102 kDa) and NF-H (110–116 kDa), although the relative proportions vary with the species. The corresponding human and murine genes have been sequenced. The organization of the three genes is very similar (although with two introns in the NF-M gene and three introns in NF-L and NF-H), and quite different from the other IF genes. Thus, the NF proteins appear to have separated at quite an early stage from the evolutionary line of the IF proteins. Compatible with this concept is the fact that neuronal IF proteins have been found in all invertebrates examined [121, 162]. NF-L is necessary but sufficient for the formation of filaments in vitro; the other two proteins are integrated into the filament at regular intervals as required [105]. NF-M and NF-H have C-terminal extensions which build cross-bridges between the filaments. These extensions contain numerous Lys-Ser-Pro triplets and the serines are sometimes phosphorylated. This high degree of phosphorylation explains the discrepancy between the molecular masses measured by gel filtration (150 kDa for NF-M and 220 kDa for NF-H) and those, given above, which have been calculated from sequence data [121, 162]. NFs may contain other proteins in addition to the three NF proteins, e.g. periferin (57 kDa) in the peripheral nervous system, and α-internexin in the brain (66 kDa). Both the latter proteins belong to the IF family but do not form filaments in vitro [70, 157].

10.10.3 Lamins

The nuclear envelope consists of an inner and an outer membrane and a fibrillar layer (lamina) that is 10–20 nm thick. The lamina serves to anchor the chromosomes during interphase and it is broken down at prophase. In vertebrates, the lamina is built mainly from polypeptides of 60–80 kDa, the **lamins**. Mammalian nuclei have three lamins: A (70 kDa), B (68 kDa) and C (60 kDa). The sequences of A (702 amino acids) and C (572 amino acids) are identical up to amino acid 566; C then has only 6 further amino acids whilst A has 136. Lamins A and C are so similar to the IF proteins in structure, having a central rod domain with a coiled-coil structure and flanking non-helical domains, that they must surely be considered members of the same family (Fig. 10.8). On the other hand, they show significant differences to the other IF proteins and very probably departed quite early from the main line of IF evolution. For example, the central domains of A and C are 42 amino acids longer than other IF proteins, and the rod region is consequently 6 nm longer (53 nm). Cross-linking occurs via the terminal domains. Two further lamins, D and E, have been found in several very different types of cells in the rat, including lymphoid cells which lack A and C. The lamins B, D and E apparently also belong to the IF family but are structurally and immunologically quite different to lamins A and C [126].

Two or three lamins are also detected electrophoretically in the cell nuclei of birds and reptiles; e.g. lamin A and two different forms of B are found in the chicken [219]. Five lamins have been described in the clawed frog *Xenopus laevis* and these are distinctly cell and developmental-stage specific. The lamins L-I and L-II are present in almost all somatic cells, L-I from the blastula stage onwards, and L-II from the gastrula stage. Oocytes contain L-III, which is also found in highly specialized cells of adult animals, e.g. muscle and nerve cells. L-IV is restricted to cells of the male germline. The fifth lamin of this series to be identified from the gene sequence is expressed in all somatic cells except erythrocytes [59, 126]. In contrast to the situation in vertebrates, the molluscs and arthropods appear to possess only one lamin. Sequence analysis is only available for *Drosophila melanogaster;* in this

case, two different polypeptides are produced by alternative splicing from one lamin gene. In flies, a further protein associated with the inner nuclear membrane, otefin (45 kDa), shows no similarity to lamins or to other IF proteins [126, 210].

10.11 Further Intracellular Structural Proteins

Various proteins associated with the cell plasma membrane link the cytoskeleton with integral membrane proteins and are very important for attachment of the cell to various surfaces, for cell movement and for the structural organization of the cell surface. The most easily approachable system, and therefore the most investigated, is that of isolated membranes of vertebrate erythrocytes (erythrocyte ghosts). Erythrocytes show extraordinary mechanical stability; during their lifetime of 120 days, human erythrocytes are forced through the capillary bed 200 000 times and are radically deformed in the process. As they forfeit most of their cytoskeleton during differentiation, their mechanical strength depends largely on the structure of the cell membrane [42]. The central role amongst the **structural proteins of the erythrocyte membrane** is taken by spectrin. This forms a network on the cytoplasmic side of the membrane and is bound to other structural components in a rather complex manner (Fig. 10.9). A connection is made via ankyrin to the cytoplasmic domain of an integral membrane protein (band-3 protein), further to short actin-tropomyosin filaments, and via the band-4.1 protein to glycophorins and the band-3 protein. The names of these proteins are derived from their SDS-gel mobilities. Interactions also exist between spectrin, ankyrin and the microtubules, as well as between spectrin and IFs. Structural proteins that are similar, but not identical, to spectrin, ankyrin and the hand-4.1 protein are found in other cells, e.g. brain cells [42].

Spectrin from mammalian erythrocytes forms rod-shaped dimers, of about 100 nm, made up of two antiparallel chains: α (260 kDa) and β (220 kDa). Two dimers are associated head-to-head such that the α N-terminus of one dimer is bound to the β C-terminus of the other (Fig. 10.9). Each human erythrocyte contains about 100 000 such tetramers. The α and β subunits of spectrin bear binding sites for calmodulin and ankyrin, respectively. The two chains are homologous; their sequences are divided into repeated segments of 106 amino acids, each of which is folded into a triple helix (Fig. 10.10). The α and β chains of human erythroid spectrin, with 2429 and 2137 amino acids, respectively, each have 17 such segments. The membrane skeleton of almost all the body cells contains tissue-specific spectrins. The α chain of human non-erythroid spectrin agrees only 58 % in its 2472 amino acids with erythrocyte spectrin. The β-like chains vary greatly in size according to the cell type: 260 kDa in the TW 240/260 proteins from the microvillous border of gut cells; 235 kDa in the αγ spectrins (fodrins) from brain, liver, lymphocytes and fibroblasts; and 220 kDa in the spectrins from erythrocytes and heart and skeletal muscle [121, 240, 308]. The different spectrin types are not only cell-specific but also development-dependent; for example, αγ spectrin (fodrin) is synthesized initially in the myoblasts of the chicken embryo, and later αβ spectrin is produced. The nucleated erythrocytes of the shark contain a related pro-

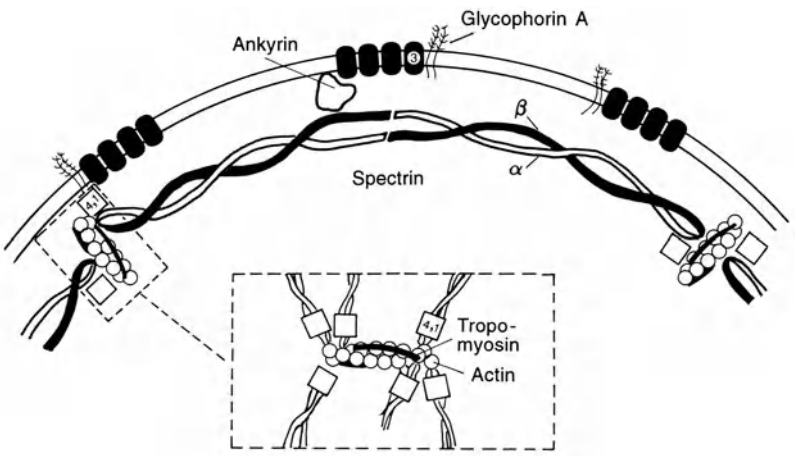


Fig. 10.9. The membrane skeleton of human erythrocytes [19]. See text for further details

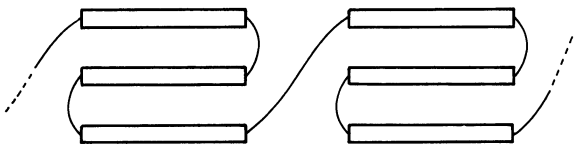


Fig. 10.10. The spectrin subunits are built up of repeated segments of 106 amino acids, each of which is folded into a triple helix with short, non-helical linkers [255]

tein which, immunologically, is more closely related to fodrin than to mammalian erythroid spectrin [31].

A human erythrocyte contains about 100 000 molecules of the phosphoprotein **ankyrin** (band-2.1 protein), which includes 23 repeats of 33 amino acids in a total of 1879 amino acids (206 kDa) [155]. The **band-4.1 protein** of human erythrocytes consists of 588 amino acids; non-erythroid isoforms of this protein are also found [281]. One of the integral glycoproteins of the erythrocyte membrane linked to the membrane skeleton is the **band-3 protein**, which is present in about 1 million copies per cell. The polypeptide chain of more than 900 amino acids is divided into two domains of different function. One serves as the anchor of the erythrocyte cytoskeleton but also binds, for example, glycolytic enzymes; the other domain forms an anion antiport built of at least 12 non-polar transmembrane segments. In the erythrocytes of mammals, birds and fish the size of the band-3 protein varies between 95 and 116 kDa. Here, there are also non-erythroid forms which are much more similar to the erythroid isoforms in their transmembrane than in their cytoplasmic domains [168, 182]. Binding of ankyrin and the band-3 protein is stabilized by the **band-4.2 protein**, which shows considerable homology to various transglutaminases in its sequence of 691 amino acids [143]. The best-known sialoglycoproteins from human erythrocyte membranes are the **glycophorins**, of which five forms have so far been described. Glycophorins A (131 amino acids), B (71aa), C (128aa), D and E (59aa) differ, above all, in the C-terminal region; the N-terminal 26 amino acids of A, B and E are identical [291]. Glycophorin A represents the original, ancestral form; glycophorin B is found only in man, the chimpanzee and gorilla [231]. Compatible with this evolutionary concept is the fact that glycophorin MK (144 amino acids), from the Japanese monkey *Macaca fuscata*, is particularly similar to human glycophorin A [198]. All the glycophorins bear numerous N- and O-bound oligosaccharides on their N-terminal extracellular domains. The polymorphism of 1-Ser/Leu and 5-Gly/Glu in glycophorin A determines the blood groups M/N, and 29-Met/Thr determines the blood groups S/s. The glycophorins, together with the band-3 proteins, constitute the binding site for lectins and viruses, and the recognition site of the malarial agent *Plasmodium falciparum* [291]. The family of submembrane proteins responsible for the attachment of the cytoskeleton to the plasma membrane in many different cell types also includes **talin** from the mouse; this protein has some N-terminal similarity to band-4.1 protein in its 2541 amino acids, but otherwise shows no noticeable relationships to other proteins [232]. The intermediary filaments of each cell are bound to those of neighbouring cells by specific intercellular junctions known as desmosomes, and thus a continuous network is formed throughout the whole tissue. The main components of a desmosome are the **desmoplakins** DPI (240–285 kDa) and DPII (210–225 kDa). These are made up of a C-terminal globular domain and a rod-like region, of about 130 nm, whose coiled-coil structure allows interaction with the IF proteins [90].

The passive elasticity of vertebrate skeletal muscle arises from the presence of filaments of the proteins **titin** (or connectin) and **nebulin**, which lie parallel to the contractile apparatus. The extended titin molecule of almost 3 MDa is unusually long. It links the ends of the myosin filaments with the Z line [296]. Nebulin is associated with actin filaments and controls their length. The thin filaments of a muscle are made up of equal proportions of actin, tropomyosin and troponin molecules; nebulin apparently functions as a measure of excessive filament tension. Accordingly, the mass of the nebulin molecule increases in parallel with the length of the thin filament in the order: rabbit (1.05 μm), chicken (1.1 μm), and cattle (1.3 μm); heart muscle with filaments of variable length contains no nebulin [154]. **Dystrophin** anchors the sarcolemma to the cytoskeleton of the muscle fibres, and is also present in large amounts in the electric organ of the electric eel *Torpedo californica*. This is a large protein of 3685 amino acids, the gene for which includes more than 65 exons in a length of 200 kb [67]. The largest known muscle protein is **twitchin** from the nematode *Caenorhabditis elegans;* its sequence of 6049 amino acids consists mainly of two types of 100-residue repeats. Its function is not yet clear, but mutations of the corresponding gene unc-22 give rise to the movement anomaly illustrated by the name [17].

The uptake of macromolecules like the epidermal growth factor (EGF), transferrin and asialoglycoproteins into the cell by receptor-mediated endocytosis is an extraordinarily intensive process; a membrane area equivalent to that of the whole cell is taken up within 30 min. The resulting transport vesicles (coated vesicles) have a protein envelope which is preformed in the coated pits of the plasma membrane. This envelope is made up of a characteristic, outer layer of **clathrin**, which is polyhedric in structure, and an inner layer of assembly proteins (17–100 kDa), which form complexes of 250–300 kDa. Quite surprisingly, one of the assembly proteins, β-adaptin (100 kDa), is identical throughout its whole 937-amino-acid sequence in man, the rat and cattle [133, 225]. The clathrin envelope consists of three-armed molecules (triskelions), each of which is built of three heavy chains, of about 190 kDa, and three light chains of 23–25 kDa (Fig. 10.11).

The complete sequence of 1675 amino acids of the rat clathrin HC is known and there are no significant homologies to other proteins. Comparisons between this and partially sequenced HCs of other mammals point to a relatively low rate of evolution [132]. The LCs are present in two homologous versions, LC-A and LC-B, which agree, for example, in bovine non-brain cells in 123 (51 %) of the 243 amino acids. Tissue-specific isoforms of both LC types are found, of which one occurs only in the brain, and the other occurs in various cell types, e.g. fibroblasts, adrenal cells and liver cells. All four human, bovine and rat LCs have been sequenced. The α-helical, central domain of all the molecules has a repeated sequence of ten heptads, and this is homologous to the rod domain of cytokeratins. The partial homology to the IF proteins suggests that exon shuffling has played an important role in clathrin evolution. The central domains bind to the HCs and the terminal domains to other proteins. Almost the only difference between bovine brain and non-brain sequences is the presence of an insert in both brain LCs (30 and 18 amino acids in LC-A and LC-B, respectively, where the LC-B insert corresponds to the N-terminus of the LC-A insert). Thus, the brain and non-brain LCs apparently arise by alternative splicing from the same gene transcript [127]. The presynaptic vesicles of nerve cells also bear a clathrin envelope. Two of the proteins in this particularly complicated structure have been examined in detail. **Synaptophysin** (33 kDa) is a Ca^{2+}-binding protein, with four transmembrane domains, which is involved in the Ca^{2+} regulation of transmitter release [12]. The **synapsins** bind the synaptic vesicles to the actin of the microfilaments, the NF-L of the neurofilaments, tubulin of the microtubules, and spectrin of the membrane skeleton. There are two synapsins, 1 and 2, encoded by different genes; both genes can produce two forms with different C-termini by alternative splicing. The polypeptide chain of about 700 amino acids is divided into a conservative N-terminal globular domain (amino acids 1–401) and a tail domain, which differs markedly between synapsins 1 and 2. Synapsin 1 is a very conservative protein and the human, bovine and rat sequences agree by about 95 %. It has at least three phosphorylation sites that are sensitive to various protein kinases; phosphorylation is presumably of regulatory significance [18, 266].

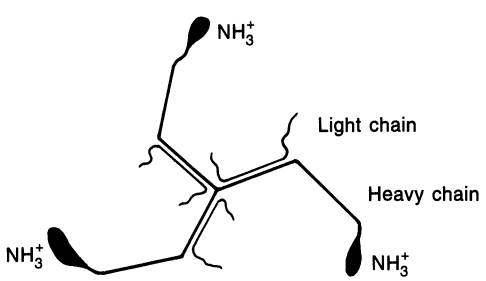


Fig. 10.11. The basic component of the clathrin envelope of cytoplasmic vesicles is a three-armed structure (triskelion), made up of three heavy and three light chains [131]

The product of the **"glued" locus** of *Drosophila melanogaster* appears to be a filament-building cytoskeleton protein. Glued heterozygotes have severe optical defects. The glued cDNA codes for a polypeptide of 1319 amino acids (148 kDa), and this is subdivided into a central α-helical domain and two terminal non-helical domains. The central domain is composed of heptad repeats and can apparently form a coiled-coil structure. There are two N-terminal serine clusters which, as in lamin, are probably phosphorylation sites; the C-terminal domain includes several possible glycosylation sites [270]. The **main sperm proteins** (MSPs) of the nematode *Caenorhabditis elegans* may also have cytoskeletal functions, and are probably involved in the movement of these unusual spermatozoa. As in other nematodes, *C. elegans* spermatozoa have no flagella but move in a gliding fashion with the aid of pseudopodia. They have almost no actin and no detectable myosin. The MSPs are small basic polypeptides of about 120 amino acids, which exist in various forms and together constitute about 15 % of the total sperm protein. They are encoded by a multi-gene family, the members of which are arranged in six clusters on chromo-

somes II and IV, but are coordinately expressed. There are several pseudogenes in the family but at least 37 genes are transcribed. The proteins are synthesized in the primary spermatocytes and aggregate to form a fibrillar cell organelle; they break up again into monomers in the spermatids and reaggregate to form the fibrils of the pseudopodium in the mature spermatozoon [297].

10.12 Lens Proteins (Crystallins)

The lens of the vertebrate eye is the most protein-rich organ of the body; the dry weight is almost exclusively that of the constituent proteins and the **protein content** may reach 50 % of the fresh weight. The refractive index of the lens increases with increasing protein content, but the deformability is reduced. Avian eye lenses with their particularly wide accommodation spectrum have the lowest protein concentration of about 20 %. Because of the optical characteristics of water, fish eyes require lenses of extremely high refractive index and therefore have protein concentrations of about 50 %; this means, however, that adjustments to distance cannot be achieved by deformation but only by moving the lens. The lenses of mammalian eyes have medium protein concentrations of 30–40 %. More than 90 % of lens protein is accounted for by crystallins. In the vertebrates these are restricted to the eye, where they also occur in small amounts in tissues other than the lens.

Lenses are built up from cellular lens fibres, derived from a layer of epithelial cells on the anterior surface of the lens. Newly synthesized fibres are laid down on the preceding fibres so that a concentric layering occurs with the oldest fibres in the centre. As there is no turnover of cells and proteins in the lens, the proteins in the oldest fibres are already synthesized in the embryo. Crystallins are thus the vertebrate proteins with the longest life and are, therefore, excellent objects for the study of **ageing processes**. The solubility of the lens proteins declines with age; the lens of a child of 10 years contains only 3 % insoluble protein compared with 40 % in the lens of an 80-year-old person. The insoluble proportion may increase pathologically (cataract) to 70 %, the main cause being increased cross-linking due to the formation of disulphide bridges and other, as yet unknown, cross-linking structures; contrary to previous hypotheses, bistyrosine is not involved. Further age-dependent changes in crystallins are the deamidation of specific asparagines and glutamines [295], and chain cleavage close to the C-terminus. Racemization of aspartate residues to the D-isomer also occurs in crystallins; the rate of this change is 0.14 % per year and cannot therefore be detected in proteins with a higher turnover rate. The main cause of all these alterations is probably light-induced oxidation.

Many different types of **crystallins** have been described for the eye lenses of the vertebrates (Table 10.3). The main forms are crystallins α, β and γ. Little or no γ crystallin is found in birds and reptiles but, instead, up to 70 % δ crystallin is present. More recently, the lenses of some vertebrates have been found to contain components other than the predominant crystallins; these are restricted to certain species or groups (Table 10.3). Future investigations will undoubtedly reveal other taxon-specific crystallins. The lens proteins are extraordinarily **heterogeneous** and are thus suitable markers for population genetics and taxonomy. This heterogeneity is based on differences in the proportions of different classes present, on the relative proportions of different aggregated states of α and β crystallins and in the chain composition of their multimers, on sequence differences between the polypeptide chains, and particularly on post-translational modifications.

The situation in the rat is described here as an interesting example in which a comparison of extractable lens proteins with mRNA translation products can be used to differentiate between primary and post-translationally modified sequences. The various highly aggregated α crystallins are made up five different chains; of these αA2,αA^{ins} and αB2 are primary gene products,

Table 10.3. The main types of crystallins in the lenses of vertebrate eyes [77, 312]

Type	Subunits		Distribution
	Number	Size (kDa)	
α	30–50	20	All vertebrates
β	2–8	23–35	All vertebrates
γ	1	20	All vertebrates (low in birds/ reptiles)
δ	4	48–50	All birds and reptiles
ε	4	35	Some birds, crocodiles
ζ	1	36	Guinea-pig, camel
η	4	54	Elephant shrews
λ	1	35	Rabbit, hare
μ		35	Wallaby and other marsupials
ϱ	1	39	Frogs of the genus *Rana*
τ	1	46–48	Agnathans and also certain fish, reptiles and birds

and αA1 and αB1 have been phosphorylated post-translationally. In theory, more than 100 different α crystallins could arise in one individual by variations in the molecular size and chain composition. The β crystallins of the rat occur as at least ten chain variants, of which four are primary gene products; seven variants of γ crystallins are known, of which only one arises by post-translational modification [239]. To the species-specific and individual differences one may also add ontogenetic alterations. In the rat, all the crystallin genes are active during early embryogenesis and they are then switched off at various moments during further development. Only two genes are later reactivated: the αB gene on day 18 and the $\beta_s(\gamma_s)$ gene post-natally [1]. In mammals, γ crystallin is the first to be produced in embryo development, whereas in birds it is δ crystallin; in both cases, α and β crystallins later predominate. This leads to differences in the crystallin spectrum between different layers of the lens.

The four main classes of crystallins belong to three **protein super-families** (α, β/γ and δ). The homology of β and γ crystallins clearly ensues from the similarity of their amino acid sequences and the corresponding subdivision of their polypeptide chains into two domains, each with two structural motifs. The sequence similarities of 37 % for the β and γ chains of the rat, and 29 % for the bovine βBp and γII chains are highly significant [61]. The amino acid sequences of the α and δ crystallins show no similarity to each other or to the β and γ crystallins, although the hydropathy profiles of the different chains have a similar pattern. The exon/intron structures of members of the three super-families are also quite different. Comparisons of orthologous sequences from different species suggest that the **rate of evolution** of all crystallins is very low (see Table 4.12, p. 161). Other lens proteins are also quite conservative, e.g. the lens fibre proteins, which probably arise from the gap junctions, have very similar peptide cleavage patterns from the amphibians to the mammals [280]. The maintenance of maximum transparency through closely packed protein molecules probably requires specific spatial relationships between the side-chains of the polypeptide chains, and this presumably restricts the possibilities for amino acid substitution.

As crystallins are found only in the vertebrates it may be assumed that they originated from non-lens proteins. In fact, a partial fragment of 76 amino acids from the αA crystallin of various mammals shows significant sequence agreement with the small heat-shock protein (hsp26) of *Drosophila, Caenorhabditis* and soybean. There is evidence of a relationship between the β/γ crystallins and the Ca^{2+}-binding protein S of the bacterium *Myoxococcus xanthus* [110, 134]. The taxon-specific crystallins are all related to, or are even identical to, enzymes of cell metabolism: δ with arginine succinate lyase, ε with lactate dehydrogenase B_4, ζ with alcohol dehydrogenase, η with aldehyde dehydrogenase, λ with hydroxyacyl-CoA-dehydrogenase, ϱ with aldose reductase and aldehyde reductase, and τ with α-enolase. The crystallins δ, ε and τ are even catalytically active [120, 312].

The characteristically high molecular weight α crystallins (Table 10.3) consist of two chain types, αA and αB, which differ in about 40 % of their 170–175 amino acids. The relative proportion of the two types varies from 9:1 in the kangaroo to 4:1 in the shark. The amino acid sequences of the αA chains are now known for more than 40 mammals, 21 birds, 2 reptiles, 2 amphibians, and the shark *Squalus acanthias* (Fig. 10.12); the αB chains are known for only a few mammals. Because there is apparently only one αA gene, the αA sequences are orthologous and therefore suitable for the study of taxonomic relationships; genealogical trees of the αA sequences have, to date, always been constructed on the maximum parsimony principle and are thus subject to the usual criticism of this method. However, the data suggest an average rate of evolution of the αA chains of about only 3 % per 100 million years. The human and bovine αA chains differ at only ten positions and the αB chains at only three; thus, the rate of evolution of the αB chains is probably even lower than that of αA. As a result of the slow evolution of the α crystallins, even distantly related species have identical αA sequences, e.g. dog/cat, rat/golden hamster/the gerbil *Meriones ungulatus*, giraffe/hippopotamus, and guinea-pig/springhaas *Pedetes caffer* [119]. The conservative nature of the lens proteins is also apparent from immunological data. Of the antigen determinants of undissociated bovine α crystallin, 42 % are present in the agnathans, a further 11 % in the cartilaginous fish, and a further 11 % in the bony fish, 21 % in the amphibians and reptiles, and only 15 % are restricted to the mammals.

Three different α-crystallin chains are produced in the mouse (αA, αA^{ins} and αB), although there are only two genes each with three exons. The 1376-bp first intron of the A gene, between codons 63 and 64, contains 288 bp downstream

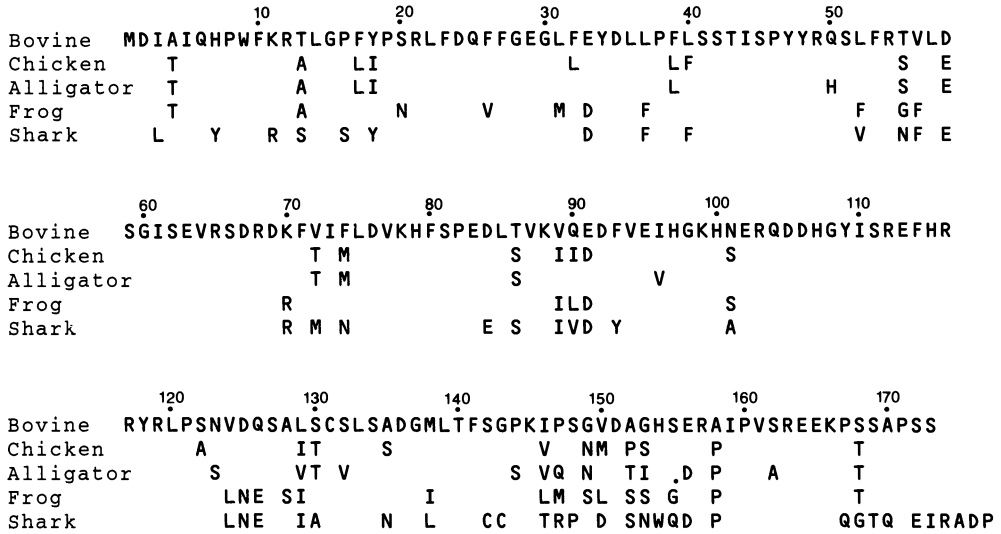


Fig. 10.12. The amino acid sequences of bovine αA crystallin chains, and the variations found in the chicken, *Alligator mississippiensis*, the frog *Rana esculenta* and the shark *Squalus acanthias* [119]

from the start of a coding sequence which corresponds to the insertion sequence of 23 amino acids present in the αA^{ins} chain variant; thus, the mRNAs for αA and αA^{ins} arise from the same gene by variable splicing of the same primary transcript (see Fig. 2.12b, p. 41). αA^{ins}-like peptides are immunologically detectable not only in all rodents but also in the hedgehog *Erinaceus europaeus*, the bat *Artibeus jamaicensis* and the lagomorph *Ochotona pinceps*. The only active human αA gene contains an αA^{ins}-homologous sequence which, however, displays a frameshift and two internal stop codons due to a nucleotide deletion. The αA^{ins} exon is therefore apparently very old in evolutionary terms, and has been inactivated in most mammals [117].

The rudimentary lenses of the mole *Talpa europaes* and the mole rat *Spalax ehrenbergi* contain immunologically detectable α crystallins, and also β and γ crystallins in the case of *Talpa*. The αA gene of the mole rat has been sequenced. It carries the same transcription and translation signals as a fully functional gene, but shows a markedly increased rate of evolution of 13% amino acid substitutions per 100 million years compared with 3% for other mammals. Whereas in all other αA sequences from the shark to man, positions 12, 29, 60 and 163 are invariant, in *Spalax* they have suffered substitutions (Fig. 10.12). If one assumes that the rate of evolution first began to increase at the start of their subterranean existence about 25 million years ago, the substitution rate in *Spalax* can be calculated as $0.9 \cdot 10^{-9}$ amino-acid-exchanging substitutions per nucleotide per year; this is much lower than the value of $5 \cdot 10^{-9}$ for synonymous substitutions [98].

Crystallin genes are expressed only in the eye. Microinjection experiments with the chicken αA gene have shown that **tissue-specific expression** is regulated by enhancer-like sequences which lie 189–242 bp upstream of the initiation site. The δ1 crystallin gene of the chicken is expressed in mouse lens epithelium just as well as in the chicken, but it is not expressed in non-lens cells [120].

Of all the lens proteins, β **crystallin** is the most heterogeneous. There are at least seven different bovine primary translation products and ten post-translationally modified chains; these are associated into di- to octomers such that there are two size classes of β crystallins (β_L and β_H). There are at least six different human β chains of 18.5 to 55 kDa, and at least four β chains of various lengths in the chicken [97, 108]. The variation in chain length is based upon the presence or absence of C- and N-terminal extensions (arms) arising through variation in the positions of initiation and termination signals [120]. Thus, for example, the bovine β chains can be divided into two groups. The first group includes the chains βB1 and βA and the predominant βBp, all of which have C-terminal extensions. Those of the second group, βB, βC, βD and βE, have no C-terminal extensions, βD differs from βB only by the possession of an N-terminal extension of 17 amino acids, and this probably arises by differential splicing of the same primary transcription

product. The bovine orthologous chains differ from those of the rat or mouse by only 4–5 %, giving an estimate of about 4 % per 100 million years for the rate of evolution. Thus, the separation of the different chain types probably occurred about 300 million years ago. Homologous chains should be present in all mammals, birds and reptiles, and most of them also occur in amphibians and fish [97]. A crystallin sequenced via the cDNA from the frog *Rana temporaria* is highly homologous to the βA1 crystallins of the mammals and birds [170]. The so-called β_s crystallin of the mammalian lens is rather unique in the β crystallin family; it behaves electrophoretically like a β crystallin but is always present as a monomer and corresponds more in its sequence to a γ crystallin [234].

The γ **crystallins** are always monomers (Table 10.3) and, like the β crystallins, show marked heterogeneity in all vertebrates. They have a typical tertiary structure, which is also seen in the β-crystallin subunits; there are two globular domains, each with two structural motifs consisting of 40–42 amino acids in four strands of an antiparallel β sheet (Fig. 10.13). Despite the great similarity in spatial structure, the sequence of the four motifs, I to IV, are identical in only 18–36 % of positions; 30–45 % of the remaining amino acids consists of conservative substitutions which have negligible influence on the chain conformation. The marked agreement between I and III and between II and IV suggests that the chains arose by two successive gene duplications. The structures of the β and the γ genes are compatible with this concept. In the β gene of the mouse, one exon is present for each of the four structural motifs; one human β gene has two further 5′ terminal exons coding for an N-terminal extension of 32 amino acids [108]. The γ gene, on the other hand, consists of only three exons, the first of which encodes only three amino acids, and the other two each encode one domain [61]. The γ crystallin of the shark *Scoliodon walbeemi* has quite a different structure that is rich in α helices, in contrast to the typical β structures of all other γ crystallins [39].

All the vertebrates examined so far were found to have multiple γ genes, of which several from mammals and the frog have been sequenced. In man and the rat there are six closely associated genes, all of which are expressed in the rat; two of the human genes are pseudogenes with internal stop codons [61, 247]. The γ crystallins also evolved slowly; the N-terminal sequences of the γ chains from the carp, the American bullfrog, the cayman and the rat agree in 12–14 out of 20 compared positions [37]. The tetrameric δ **crystallins** appear to be present only in reptiles and birds, where they substitute for the γ crystallins. Their concentration, as a proportion of the total crystallin, varies widely with the species; in the chicken they make up 50 % of the total lens protein, whereas the chimney swift *Chaetura pelagica* has no δ crystallin but instead 20 % ε crystallin [311]. The δ crystallin is homologous to arginine succinate lyase. As the result of a gene duplication, the chicken and duck possess two δ-crystallin genes, δ1 and δ2. The crystallin encoded by δ1 has a much lower enzyme activity than that encoded by δ2. The two genes are expressed more or less equally in the duck lens, whereas δ1 predominates in the chicken. The lens of the duck thus contains a much higher enzyme activity than that of the chicken, although in the latter the activity in the lens is nevertheless higher than in the liver.

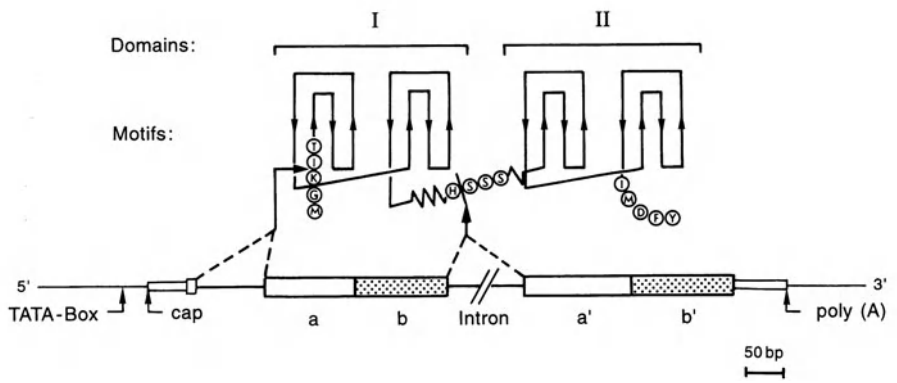


Fig. 10.13. The exon/intron structure of the mammalian γ-crystallin gene coincides with the domain structure of the polypeptide chain. The chain is composed of two domains, each with two structural motifs; the γ-crystallin gene includes three exons, of which the first codes for only three amino acids, and the other two each encode one of the two domains [192]

The δ crystallin of the ostrich *Struthio camelus* also has high arginine succinate lyase activity [40, 310]. ε **Crystallin** was discovered in the Peking duck but also occurs in other avian species, e.g. vultures and stilts, in proportions of up to 23% of the lens protein; it is not found in the chicken. It is also present in the eye lenses of crocodiles but does not occur in any other reptiles. ε Crystallin is homologous to the heart-specific isoenzyme B_4 of lactate dehydrogenase (LDH-B_4). Ducks and chickens possess only one LDH-B gene which codes for both ε crystallin and the cell enzyme. Thus, the gene product in this case must be adapted to two quite different functions and this "gene sharing" results in a radically reduced rate of evolution. Like LDH-B_4, ε crystallin is a homotetramer and differs from the cell enzyme only by the deamidation of the asparagine residues in positions 163 and 265 [99].

Taxon-specific crystallins have only quite recently been discovered in the mammals; the first to be found was λ crystallin, which makes up about 7–8% of the soluble protein in the lenses of rabbits and hares and shows 30% sequence agreement with L-3-hydroxacyl-CoA-dehydrogenase from mitochondria [197]. The ζ crystallin from the lens of the quinea-pig *Cavia pocellus* belongs to the family of zinc-containing long-chained alcohol/polyol dehydrogenases. However, adaptation to the function of lens protein has led not only to the loss of the zinc-binding amino acids, and therefore loss of the enzyme activity during evolution, but also to other changes. Thus, the lens protein is tetrameric, like mammalian sorbitol dehydrogenase, and not dimeric, like the zinc-containing alcohol dehydrogenases. Surprisingly, ζ crystallin is found not only in the guinea-pig and other hystricomorphs but also in *Camelus dromedarius*, a representative of a different mammalian order [77]. The eye lenses of three species of the primitive elephant shrews (Macroscleidae) have been shown to contain up to 24% soluble η crystallin, which is apparently identical to cytosolic aldehyde oxidase. However, the lens shows no enzyme activity, probably as the result of ageing processes [312]. In the tammar wallaby *Macropus eugenii* and the tiger cat *Dasyurops maculatus* (Marsupalia), the lens contains a μ crystallin which appears to have a nucleotide-binding structure but has no homology to known dehydrogenases [312]. The ϱ crystallin isolated from the lenses of *Rana temporaria* and *R. catesbeiana* shows high sequence homology to aldehyde reductase, aldose reductase and prostaglandin-F synthase. The frog lens contains an N-terminal acylated (RHO-I) and a non-blocked (RHO-II) form of the lens protein, neither of which has enzyme activity [75]. The α-enolase-homologous τ crystallin is found in the lamprey and also in certain species of birds, reptiles and teleosts. The duck has only one α-enolase/τ crystallin gene, which thus codes for a protein with two different functions. However, the enzyme activity in the lens is reduced by ageing processes [120].

The invertebrates have no crystallins, not even the cephalopods, whose eyes are so similar to those of the vertebrates as to serve as a remarkable example of convergent evolution. The **lenses of cephalopod eyes** contain an electrophoretically complex spectrum of proteins which are, however, also termed crystallins. The main lens proteins of *Ommastrephes sloani* are homo- or heterodimers, with subunits of 222–223 amino acids, encoded by a gene family with at least three members. Two of these genes show only 50% agreement in their amino acid sequences but both are significantly homologous to rat glutathione-S-transferase. Similar lens proteins have been detected in *Octopus vulgaris* and *Sepia esculenta*, which belong to very different cephalopod groups [38, 284].

References

1. Aarts H. J. M., Lubsen N. H. and Schoenmakers J. G. G.: Crystallin gene expression during rat lens development. Eur. J. Biochem. 183: 31–36 (1989)
2. Abe H. et al.: Sequence of cDNAs encoding actin depolymerizing factor and cofilin of embryonic chicken skeletal muscle: Two functionally distinct actin-regulatory proteins exhibit high structural homology. Biochemistry 29: 7420–25 (1990)
3. Alonso S. et al.: Comparison of three actin-coding sequences in the mouse: evolutionary relationships between actin genes of warm-blooded vertebrates. J. mol. Evol. 23: 11–22 (1986)
4. Ampe C. et al.: The primary structure of the basic isoform of Acanthamoeba profilin. Eur. J. Biochem. 170: 597–601 (1988)
5. Anderson P.: Molecular genetics of nematode muscle. Annual Rev. Genet. 23: 507–525 (1989)
6. Arimura C. et al.: Primary structure of chicken skeletal muscle and fibroblast α-actinins deduced from cDNA sequences. Eur. J. Biochem. 177: 649–655 (1988)
7. Baba M. L. et al.: The early adaptive evolution of calmodulin. Mol. Biol. Evol. 1: 442–455 (1984)
8. Babu Y. S. et al.: Three-dimensional structure of calmodulin. Nature 315: 37–40 (1985)
9. Balcarek J. M. and Cowan N. J.: Structure of the mouse glial fibrillary acidic protein gene: implications for the evolution of the intermediate filament multigene family. Nucleic Acids Res. 13: 5527–43 (1985)
10. Ball E. et al.: Arthrin, a myofibrillar protein of insect flight muscle, is an actin-ubiquitin conjugate. Cell 51: 221–228 (1987)

11. Barahona I. et al.: Sequence of one α- and two β-tubulin genes of Tetrahymena pyriformis. Structural and functional relationships with other eukaryotic tubulin genes. J. mol. Biol. 202: 365–382 (1988)
12. Bargou R. C. E. F. and Leube R. E.: The synaptophysin-encoding gene in rat and man is specifically transcribed in neuroendocrine cells. Gene 99: 197–204 (1991)
13. Bartegi A. et al.: Isolation, characterization and immunocytochemical localization of caldesmon-like protein from molluscan striated muscle. Eur. J. Biochem. 185: 589–595 (1989)
14. Barton G. J. et al.: Amino acid sequence analysis of the anexin super-gene family of proteins. Eur. J. Biochem. 198: 749–760 (1991)
15. Bazari W. L. et al.: Villin sequence and peptide map identify six homologous domains. Proc. Nat. Acad. Sci. USA 85: 4986–90 (1988)
16. Beltrama D. M., Arce C. A. and Barra H. S.: Tubulin, but not microtubules, is the substrate for tubulin:tyrosine ligase in mature avian erythrocytes. J. biol. Chem. 262: 15673–77 (1987)
17. Benian G. M. et al.: Sequence of an unusually large protein implicated in regulation of myosin activity in C. elegans. Nature 342: 45–50 (1989)
18. Bennett A. F., Hayes N. V. L. and Baines A. J.: Site specificity in the interactions of synapsin-I with tubulin. Biochem. J. 276: 793–799 (1991)
19. Bennett V.: The membrane skeleton of human erythrocytes. Annual Rev. Biochem. 54: 273–304 (1985)
20. Bereiter-Hahn J., Matolsky A. G. and Richards K. S. (eds.): Biology of the integument, 2 vol. set. Springer, Berlin 1986
21. di Bernardo M. G. et al.: Nucleotide sequence of a full length cDNA clone encoding for β-tubulin of the sea urchin Paracentrotus lividus. Nucleic Acids Res. 17: 5851 (1989)
22. Blumenberg M.: Evolution of homologous domains of cytoplasmic intermediate filament proteins and lamins. Mol. Biol. Evol. 6: 53–65 (1989)
23. Bodnaryk R. P. and Morishima I.: Isolation and characterization of a calmodulin lacking trimethyllysine from the fat body of the silkworm, Bombyx mori. Insect Biochem. 14: 11–17 (1984)
24. Boron M. D. et al.: The sequence of chick α-actinin reveals homologies to spectrin and calmodulin. J. biol. Chem. 262: 17623–29 (1987)
25. Breitbart R. E., Andreadis A. and Nadal-Ginard B.: Alternative splicing: A ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annual Rev. Biochem. 56: 467–495 (1987)
26. Brewer J. M., Wunderlich J. K. and Ragland W. L.: The amino acid sequence of avian thymic hormone, a parvalbumin. Biochimie 72: 653–660 (1990)
27. Brzeska H., Lynch T. J. and Korn E. D.: Acanthamoeba myosin I heavy chain kinase is activated by phosphatidylserine-enhanced phosphorylation. J. Biol. Chem. 265: 3591–94 (1990)
28. Burke D. J. and Ward S.: Identification of a large multigene family encoding the major sperm protein of Caenorhabditis elegans. J. mol. Biol. 171: 1–29 (1983)
29. Caldwell J. E. et al.: cDNAs encoding the β subunit of Cap Z, the actin-capping protein of the Z line of muscle. J. Biol. Chem. 264: 12648–52 (1989)
30. Carlier M. F.: Actin. Protein structure and filament dynamics. J. Biol. Chem. 266: 1–4 (1991)
31. Carlin R. K., Scheele G. A. and Cohen W. D.: Similarities between the Mr 245,000 calmodulin-binding protein of the dogfish erythrocyte cytoskeleton and α-fodrin. Arch. Biochem. Biophys. 230: 13–20 (1984)
32. Carroll S. L., Bergsma D. J. and Schwartz R. J.: Structure and complete nucleotide sequence of the chicken α-smooth muscle (aortic) actin gene. An actin gene which produces multiple messenger RNAs. J. biol. Chem. 261: 8965–76 (1986)
33. Castellani L. and Cohen C.: Rod phosphorylation favors folding in a catch muscle myosin. Proc. Nat. Acad. Sci. USA 84: 4058–62 (1987)
34. Chang K. S., Rothblum K. N. and Schwartz R. J.: The complete sequence of the chicken α-cardiac actin gene: a highly conserved vertebrate gene. Nucleic Acids Res. 13: 1223–37 (1985)
35. Cheng Q. et al.: Cloning, sequencing and expression of a full-length rabbit fast skeletal tropinin-C cDNA. FEBS Letters 228: 22–26 (1988)
36. Chilcote T. J. and Johnson K. A.: Phosphorylation of Tetrahymena 22 S dynein. J. Biol. Chem. 265: 17257–66 (1990)
37. Chiou S. H. et al.: Sequence comparison of gamma-crystallins from the reptilian and other vertebrate species. FEBS Letters 221: 134–138 (1987)
38. Chiou S. H.: A novel crystallin from octopus lens. FEBS Letters 241: 261–264 (1988)
39. Chiou S. H. et al.: Comparison of the gamma-crystallins isolated from eye lenses of shark and carp. Unique secondary and tertiary structure of shark gamma-crystallins. FEBS Letters 275: 111–113 (1990)
40. Chiou S. H. et al.: Ostrich crystallins: Structural characterization of delta-crystallin with enzymic activity. Biochem. J. 273(2): 295–300 (1991)
41. Cleveland D. W. and Sullivan K. F.: Molecular biology and genetics of tubulin. Annual Rev. Biochem. 54: 331–365 (1985)
42. Cohen C. M. and Palek J.: Cellular and molecular biology of normal and abnormal erythroid membranes. Wiley-Liss, New York 1990
43. Collins J. H., Cox J. A. and Theibert J. L.: Amino acid sequence of a sarcoplasmic calcium-binding protein from the sandworm Nereis diversicolor. J. Biol. Chem. 263: 15378–85 (1988)
44. Collins J. H. et al.: Amino acid sequences and calcium-binding properties of two isoforms of barnacle troponin-C. Biochemistry 30: 702–707 (1991)
45. Conzelmann K. K. and Helftenbein E.: Nucleotide sequence and expression of two β-tubulin genes in Stylonychia lemnae. J. mol. Biol. 198: 643–653 (1987)
46. Cooper J. A. et al.: Purification and characterization of actophorin, a new 15.000 dalton actin-binding protein from Acanthamoeba castellanii. J. biol. Chem. 261: 477–485 (1986)
47. Cooper T. A. and Ordahl C. P.: A single cardiac troponin T gene generates embryonic and adult isoforms via developmentally regulated alternative splicing. J. biol. Chem. 260: 11140–48 (1985)
48. Cox K. H. et al.: Cell lineage-specific programs of expression of multiple actin genes during sea urchin embryogenesis. J. mol. Biol. 188: 159–172 (1986)
49. Crain W. R. jr. et al.: The sequence of a sea urchin muscle actin gene suggests a gene conversion with a cytoskeletal actin gene. J. mol. Evol. 25: 37–45 (1987)
50. Cross G. S. et al.: Cytoskeletal actin gene families of Xenopus borealis and Xenopus laevis. J. mol. Evol. 27: 17–28 (1988)

51. Cupples C. G. and Pearlman R. E.: Isolation and characterization of the actin gene from Tetrahymena thermophila. Proc. Nat. Acad. Sci. USA 83: 5160–64 (1986)
52. Darwish H. M. et al.: Molecular cloning of the cDNA and chromosomal gene for vitamin D-dependent calcium-binding protein of rat intestine. Proc. Nat. Acad. Sci. USA 84: 6108–11 (1987)
53. Dayhoff M. O. (ed.): Atlas of protein sequence and structure, Vol. 5 and Suppl. 1–3. Nat. Biomed. Res. Foundation, Washington 1972–78
54. Detrich III H. W., Johnson K. A. and Marchese-Ragona S. P.: Polymerization of Antarctic fish tubulins at low temperatures: Energetic aspects. Biochemistry 28: 10085–93 (1989)
55. D'Haese J. D. and Carlhoff D.: Localization and histochemical characterization of myosin isoforms in earthworm body wall muscle. J. comp. Physiol. B 157: 171–179 (1987)
56. Dibb N. J. and Newman A. J.: Evidence that introns arose at proto-splice sites. Embo J. 8: 2015–21 (1989)
57. Djian P. and Green H.: Involucrin gene of tarsioids and other primates. Alternatives in the evolution of the segment of repeats. Proc. Nat. Acad. Sci. USA 88: 5321–25 (1991)
58. Dodemont H., Riemer D. and Weber K.: Structure of an invertebrate gene encoding cytoplasmic intermediate filament (IF) proteins. Implications for the origin and the diversification of IF proteins. Embo J. 9: 4083–94 (1990)
59. Doring V. and Stick R.: Gene structuere of nuclear lamin-LIII of Xenopus laevis. A model for the evolution of IF proteins from a lamin-like ancestor. Embo J. 9: 4073–81 (1990)
60. Doyle K. E. et al.: Drosophila melanogaster contains a single calmodulin gene. Further structure and expression studies. J. mol. Biol. 213: 599–605 (1990)
61. den Dunnen J. T. et al.: Concerted and divergent evolution within the rat gamma-crystallin gene family. J. mol. Biol. 189: 37–46 (1986)
62. Durica D. S.: DNA sequence analysis and structural relationships among the cytoskeletal actin genes of the sea urchin. J. mol. Evol. 28: 72–86 (1989)
63. Dustin P.: Microtubules, 2. ed. Springer, Berlin 1984
64. Edamatsu M. et al.: The primary structure of Tetrahymena profilin. Biochem. biophys. Res. Commun. 175: 543–550 (1991)
65. Emerson C. P. jr. and Bernstein S. I.: Molecular genetics of myosin. Annual Rev. Biochem. 56: 695–726 (1987)
66. Endow S. A. and Hatsumi M.: A multimember kinesin gene family in Drosophila. Proc. Nat. Acad. Sci. USA 88: 4424–27 (1991)
67. Ervasti J. M., Kahl S. D. and Campbell K. P.: Purification of dystrophin from skeletal muscle. J. Biol. Chem. 266: 9161–65 (1991)
68. Files J. G., Carr S. and Hirsch D.: Actin gene family of Caenorhabditis elegans. J. mol. Biol. 164: 355–375 (1983)
69. Fliegel L. and Michalak M.: Fast-twitch and slow-twitch skeletal muscles express the same isoform of calreticulin. Biochem. biophys. Res. Commun. 177: 979–984 (1991)
70. Fliegner K. H., Ching G. Y. and Liem R. K. H.: The predicted amino acid sequence of an α-internexin is that of a novel neuronal intermediate filament protein. Embo J. 9: 749–755 (1990)
71. Fodor W. L.: Human ventricular/slow twitch myosin alkali light chain gene characterization, sequence, and chromosomal location. J. Biol. Chem. 264: 2143–49 (1989)
72. Forryschaudies S. and Hughes S. H.: The chicken tropomyosin-I gene generates 9 messenger RNAs by alternative splicing. J. Biol. Chem. 266: 13821–27 (1991)
73. Franz J. K. and Franke W. W.: Cloning of cDNA and amino acid sequence of a cytokeratin expressed in oocytes of Xenopus laevis. Proc. Nat. Acad. Sci. USA 83: 6475–79 (1986)
74. Fraser R. D. B.: Focus on intermediate filaments. Trends biochem. Sci. 12: 43–45 (1987)
75. Fujii Y. et al.: Purification and characterization of rho-crystallin from Japanese common bull frog lens. J. Biol. Chem. 265: 9914–23 (1990)
76. Fullmer C. S. and Wasserman R. H.: Chicken intestinal 28-kilodalton calbindin-D: complete amino acid sequence and structural considerations. Proc. Nat. Acad. Sci. USA 84: 4772–76 (1987)
77. Garland D. et al.: Zeta-crystallin is a major protein in the lens of Camelus dromedarius. Arch. Biochem. Biophys. 285: 134–136 (1991)
78. Gatti J. L. et al.: Outer arm dynein from trout spermatozoa. Purification, polypeptide composition, and enzymatic properties. J. Biol. Chem. 264: 11450–57 (1989)
79. Geeves M. A.: The dynamics of actin and myosin association and the crossbridge model of muscle contraction (Review Article). Biochem. J. 274(1): 1–14 (1991)
80. Gerday C., Gilles R. and Bolis L. (eds.): Calcium and calcium binding proteins. Springer, Berlin 1988
81. Gerday C., Collin S. and Gerardin-Otthiers N.: The amino acid sequence of the parvalbumin from the very fast swimbladder muscle of the toadfish (Opsanus tau). Comp. Biochem. Physiol. Pt. B 93: 49–55 (1989)
82. Gerke V. and Koch W.: The cDNA sequence of chicken annexin II. Nucleic Acids Res. 18: 4246 (1990)
83. Gibbons I. R.: Dynein ATPases as microtubule motors. J. Biol. Chem. 263: 15837–40 (1988)
84. Gillespie J. M., Marshall R. C. and Woods E. F.: A comparison of lizard claw keratin proteins with those of avian beak and claw. J. mol. Evol. 18: 121–129 (1982)
85. Goedert M. et al.: Cloning and sequencing of the cDNA encoding an isoform of microtubule-associated protein tau containing four tandem repeats. Differential expression of tau protein mRNAs in human brain. Embo J. 8: 393–399 (1989)
86. Görlach M. et al.: Characterization of calmodulin from Drosophila heads. Biochim. biophys. Acta 832: 228–232 (1985)
87. Goldman R. D. and Steinert P. M. (eds.): Cellular and molecular biology of intermediate filaments. Plenum. New York 1990
88. Goodwin E. B., Szent-Györgyi A. G. and Leinwand L. A.: Cloning and characterization of the scallop essential and regulatory myosin light chain cDNAs. J. biol. Chem. 262: 11052–56 (1987)
89. Grand R. J. A. and Perry S. V.: Preparation of the alkali and P light chains of chicken gizzard myosin. Amino acid sequence of the alkali light chain. Biochem J. 211: 267–272 (1983)
90. Green K. J. et al.: Structure of the human desmoplakins. Implications for function in the desmosomal plaque. J. Biol. Chem. 265: 2603–12 (1990)

91. Halsall D. J. and Hammer J. A.: A 2nd isoform of chicken brush border myosin-I contains a 29-residue inserted sequence that binds calmodulin. FEBS Letters 267: 126–130 (1990)
92. Hanke P. D., Lepinske H. M. and Storti R. V.: Characterization of a Drosophila cDNA clone that encodes a 252-amino acid non-muscle tropomyosin isoform. J. biol. Chem. 262: 17370–73 (1987)
93. Hardin P. E. et al.: Spec2 genes of Strongylocentrotus purpuratus. Structure and differential expression in embryonic aboral ectoderm cells. J. mol. Biol. 202: 417–431 (1988)
94. Hardy D. O., Bender P. K. and Kretsinger R. H.: Two calmodulin genes are expressed in Arbacia punctulata. An ancient gene duplication is indicated. J. mol. Biol. 199: 223–227 (1988)
95. Harper D. S. and Jahn C. L.: Differential use of termination codons in ciliated protozoa. Proc. Nat. Acad. Sci. USA 86: 3252–56 (1989)
96. Head J. F.: Amino acid sequence of a low molecular weight, high affinity calcium-binding protein from the optic lobe of the squid Loligo pealei. J. Biol. Chem. 264: 7202–09 (1989)
97. Hejtmancik J. F. et al.: cDNA and deduced protein sequence for the βB1-crystallin polypeptide of the chicken lens. Conservation of the PAPA sequence. J. biol. Chem. 261: 982–987 (1986)
98. Hendricks W. et al.: The lens protein αA-crystallin of the blind mole rat, Spalax ehrenbergi: Evolutionary change and functional constraints. Proc. Nat. Acad. Sci. USA 84: 5320–24 (1987)
99. Hendriks W. et al.: Duck lens epsilon-crystallin and lactate dehydrogenase B4 are identical. A single-copy gene product with two distinct functions. Proc. Nat. Acad. Sci. USA 85: 7114–18 (1988)
100. Herring B. P., Stull J. T. and Gallagher P. J.: Domain characterization of rabbit skeletal muscle myosin light chain kinase. J. Biol. Chem. 265: 1724–30 (1990)
101. Hightower R. C. and Meagher R. B.: The molecular evolution of actin. Genetics 114: 333–343 (1986)
102. Higuchi H. and Goldman Y. E.: Sliding distance between actin and myosin filaments per ATP molecule hydrolysed in skinned muscle fibers. Nature 352: 352–354 (1991)
103. Hinkelaust S., Hinkel P. and Beinbrech G.: Four cross-bridge strands and high paramyosin content in the myosin filaments of honey bee flight muscles. Experientia 46: 872–874 (1990)
104. Hirono M. et al.: Tetrahymena actin. Cloning and sequencing of the Tetrahymena actin gene and identification of its gene product. J. mol. Biol. 194: 181–192 (1987)
105. Hisanaga S. I., Ikai A. and Hirokawa N.: Molecular architecture of the neurofilament. I. Subunit arrangement of neurofilament L protein in the intermediate-sized filament. J. mol. Biol. 211: 857–869 (1990)
106. Hock R. S., Davis G. and Speicher D. W.: Purification of human smooth muscle filamin and characterization of structural domains and functional sites. Biochemistry 29: 9441–51 (1990)
107. Hofmann S., Duesterhoeft S. and Pette D.: Six myosin heavy chain isoforms are expressed during chick breast muscle development. FEBS Letters 238: 245–248 (1988)
108. Hogg D. et al.: Characterization of the human β-crystallin gene HuβA3/A1 reveals ancestral relationships among the β″-crystallin superfamily. J. biol. Chem. 261: 12420–27 (1986)
109. Hohl D. et al.: Characterization of human loricin. Structure and function of a new class of epidermal cell envelope proteins. J. Biol. Chem. 266: 6626–36 (1991)
110. Holberton D., Baker D. A. and Marshall J.: Segmented α-helical coiled-coil structure of the protein giardin from the Giardia cytoskeleton. J. mol. Biol. 204: 789–795 (1988)
111. Horowitz J. A. and Hammer J. A.: A new Acanthamoeba myosin heavy chain. Cloning of the gene and immunological identification of the polypeptide. J. Biol. Chem. 265: 20646–52 (1990)
112. Hunziker W.: The 28-kDa vitamin D-dependent calcium-binding protein has six-domain structure. Proc. Nat. Acad. Sci. USA 83: 7578–82 (1986)
113. Imboden M. A. et al.: Transcription of the intergenic regions of tubulin gene cluster of Trypanosoma brucei: evidence for a polycistronic transcription unit in a eukaryote. Nucleic Acids Res. 15: 7357–68 (1987)
114. Ishimoda-Tagaki T. and Kobayashi M.: Molecular heterogeneity and tissue specificity of tropomyosin obtained from various bivalves. Comp. Biochem. Physiol. 88B: 443–452 (1987)
115. Jauregui-Adell J. et al.: Amino acid sequence of an α-parvalbumin, pI = 4.88, from frog skeletal muscle. Eur. J. Biochem. 123: 337–345 (1982)
116. Jauregui-Adell J., Wnuk W. and Cox C. A.: Complete amino acid sequence of the sarcoplasmic calcium-binding protein (SCP-I) from crayfish (Astacus leptodactylus). FEBS Letters 243: 209–212 (1989)
117. Jaworsji C. J. and Piatigorsky J.: A pseudo-exon in the functional human αA-crystallin gene. Nature 337: 752–754 (1989)
118. Johnston P. A. et al.: Two novel annexins from Drosophila melanogaster. Cloning, characterization, and differential expression in development. J. Biol. Chem. 265: 11382–88 (1990)
119. de Jong W. W. et al.: α-Crystallin A sequences of Alligator mississippiensis and the lizard Tupinambis teguixin: Molecular evolution and reptilian phylogeny. Mol. Biol. Evol. 2: 484–493 (1985)
120. de Jong W. W. et al.: Evolution of the eye lens crystallins: The stress connection. Trends biochem. Sci. 14: 365–368 (1989)
121. Julien J. P. et al.: Sequence and structure of the mouse gene coding the largest neurofilament subunit. Gene 68: 307–314 (1988)
122. Kagawa H. et al.: Paramyosin gene (unc-15) of Caenorhabditis elegans. Molecular cloning, nucleotide sequence and models for thick filament structure. J. mol. Biol. 207: 311–333 (1989)
123. Kapprell H. P., Cowin P. and Franke W. W.: Biochemical characterization of the soluble form of the junctional plaque protein, plakoglobin, from different cells. Eur. J. Biochem. 166: 505–517 (1987)
124. Katsuragawa Y. et al.: Two distinct nonmuscle myosin-heavy-chain mRNAs are differentially expressed in various chicken tissues. Identification of a novel gene family of vertebrate non-sarcomeric myosin heavy chains. Eur. J. Biochem. 184: 611–616 (1989)
125. Kaufmann S. et al.: Talin binds to actin and promotes filament nucleation. FEBS Letters 284: 187–191 (1991)
126. Kaufmann S. H.: Additional members of the rat liver lamin polypeptide family. Structural and immunological characterization. J. Biol. Chem. 264: 13946–55 (1989)

127. Keen J. H.: Clathrin and associated assembly and disassembly proteins. Annual Rev. Biochem. 59: 415–438 (1990)
128. Ketchum A. S. et al.: Complete sequence of the Drosophila nonmuscle myosin heavy-chain transcript. Proc. Nat. Acad. Sci. USA 87: 6316–20 (1990)
129. Kiehart D. P.: Molecular genetic dissection of myosin. Cell 60: 347–350 (1990)
130. Kimura K. et al.: Troponin from nematode: purification and characterization of troponin from Ascaris body wall muscle. Comp. Biochem. Physiol. Pt. B 88: 399–407 (1987)
131. Kirchhausen T. et al.: Clathrin light chains LCA and LCB are similar, polymorphic, and share repeated heptad motifs. Science 236: 320–324 (1987)
132. Kirchhausen T. et al.: Clathrin heavy chain: Molecular cloning and complete primary structure. Proc. Nat. Acad. Sci. USA 84: 8805–09 (1987)
133. Kirchhausen T. et al.: AP17 and AP19, the mammalian small chains of the clathrin-associated protein complexes show homology to Yap17p, their putative homolog in yeast. J. Biol. Chem. 266: 11153–57 (1991)
134. Klemenz R. et al.: αB-Crystallin is a small heat shock protein. Proc. Nat. Acad. Sci. USA 88: 3652–56 (1991)
135. Kligman D. and Hilt D. C.: The S100 protein family. Trends biochem. Sci. 13: 437–443 (1988)
136. Klinge E. M. et al.: Evolution of keratin genes: Different protein domains evolve by different pathways. J. mol. Evol. 24: 319–329 (1987)
137. Knapp B. et al.: Nonepidermal members of the keratin multigene family: cDNA sequences and in situ localization of the mRNAs. Nucleic Acids Res. 14: 751–763 (1986)
138. Kobayashi R., Hidaka H. and Tashima Y.: Purification, characterization, and partial sequence analysis of the 32-kDa calcimedin from chicken gizzard. Arch. Biochem. Biophys. 277: 203–210 (1990)
139. Kobayashi T. et al.: The primary structure of a new Mr 18.000 calcium vector protein from amphioxus. J. biol. Chem. 262: 2613–23 (1987)
140. Kobayashi T. et al.: Amino acid sequence of crayfish troponin I. J. Biol. Chem. 264: 1551–57 (1989)
141. Kobayashi T. et al.: Amino acid sequences of the two major isoforms of troponin C from crayfish. J. Biol. Chem. 264: 18247–59 (1989)
142. Koller M. and Stehler E. E.: Characterization of an intronless human calmodulin-like pseudogene. FEBS Letters 239: 121–128 (1988)
143. Korsgren C. et al.: Complete amino acid sequence and homologies of human erythrocyte membrane protein band 4.2. Proc. Nat. Acad. Sci. USA 87: 613–617 (1990)
144. Kosik K. S. et al.: The primary structure and analysis of the squid kinesin heavy chain gene. J. Biol. Chem. 265: 3278–83 (1990)
145. Kowbel D. J. and Smith M. J.: The genomic nucleotide sequences of two differentially expressed actin-coding genes from the sea star Pisaster ochraceus. Gene 77: 297–308 (1989)
146. Krause K. H. et al.: Plant cells contain calsequestrin. J. Biol. Chem. 264: 4269–72 (1989)
147. Krause M. et al.: Wild-type and mutant actin genes in Caenorhabditis elegans. J. mol. Biol. 208: 381–392 (1989)
148. Krieger J. et al.: Cloning, sequencing and expression of locust tropomyosin. Insect Biochem. 20: 173–184 (1990)
149. Kuehn G. D. and Hageman J. H.: Calmodulin-like protein from Bacillus subtilis. Biochem. biophys. Res. Commun. 134: 212–217 (1986)
150. Kuznetsov S. A. and Gelfland V. I.: 18 kDa microtubule-associated protein: identification as a new light chain (LC-3) of microtubule-associated protein 1 (MAP-1). FEBS Letters 212: 145–148 (1987)
151. Kwiatkowski D. J. et al.: Plasma and cytoplasmic gelsolins are encoded by a single gene and contain a duplicated actin-binding domain. Nature 323: 455–458 (1986)
152. Kwiatkowski D. J. and Bruns G. A. P.: Human profilin. Molecular cloning, sequence comparison, and chromosomal analysis. J. biol. Chem. 263: 5910–15 (1988)
153. Kwon H. et al.: Isolation of the regulatory domain of scallop myosin. Role of the essential light chain in calcium binding. Proc. Nat. Acad. Sci. USA 87: 4771–75 (1990)
154. Labeit S. et al.: Evidence that nebulin is a protein-ruler in muscle thin filaments. FEBS Letters 282: 313–316 (1991)
155. Lambert S. et al.: cDNA sequence for human erythrocyte ankyrin. Proc. Nat. Acad. Sci. USA 87: 1730–34 (1990)
156. Lambooy P. K. and Korn E. D.: Purification and characterization of actobindin, a new actin monomer-binding protein from Acanthamoeba castellanii. J. biol. Chem. 261: 17150–55 (1986)
157. Landon F. et al.: Multiple mRNAs encode peripherin, a neuronal intermediate filament protein. Embo J. 8: 1719–26 (1989)
158. de Lanerolle P. and Nishikawa M.: Regulation of embryonic smooth muscle myosin by protein kinase C. J. Biol. Chem. 263: 9071–74 (1988)
159. Leathers V. L. et al.: Calbindin D-28K, a 1⁻,25-dihydroxyvitamin D3-induced calcium-binding protein, binds five or six calcium ions with high affinty. J. Biol. Chem. 265: 9838–41 (1990)
160. Lee G., Cowan N. and Kirschner M.: The primary structure and heterogeneity of tau protein from mouse brain. Science 239: 285–288 (1988)
161. Lee M. G. S. et al.: A putative flagellar Ca-binding protein of the flagellum of trypanosomatid protozoan parasites. Nucleic Acids Res. 18: 4252 (1990)
162. Lees J. F. et al.: The structure and organization of the human heavy neurofilament subunit (NF-H) and the gene encoding it. Embo J. 7: 1947–55 (1988)
163. Lehman W., Sheldon A. and Madonia W.: Diversity in smooth muscle thin filament composition. Biochim. biophys. Acta 914: 35–39 (1987)
164. Lenz S. et al.: The alkali light chains of human smooth and non-muscle myosins are endoded by a single gene. Tissue-specific expression by alternative splicing pathways. J. Biol. Chem. 264: 9009–15 (1989)
165. Levy E. et al.: Structure and evolutionary origin of the gene encoding mouse NF-M, the middle-molecular-mass neurofilament protein. Eur. J. Biochem. 166: 71–77 (1987)
166. Libri D. et al.: A nonmuscle tropomyosin is encoded by the smooth/skeletal β-tropomyosin gene and its RNA is transcribed from an internal promoter. J. Biol. Chem. 265: 3471–73 (1990)
167. Liew C. C. et al.: Complete sequence and organization of the human cardiac β-myosin heavy chain gene. Nucleic Acids Res. 18: 3647–51 (1990)

168. Lindenthal S. and Schubert D.: Monomeric erythrocyte band-3 protein transports anions. Proc. Nat. Acad. Sci. USA 88: 6540–44 (1991)
169. Little M. and Seehaus T.: Comparative analysis of tubulin sequences. Comp. Biochem. Physiol. Pt. B 90: 655–670 (1988)
170. Lucin S. V. et al.: Frog lens βA1-crystallin: the nucleotide sequence of the cloned cDNA and computer graphics modelling of the three-dimensional structure. Biochim. biophys. Acta 916: 163–171 (1987)
171. Lynch T. J. et al.: Purification and characterization of a third isoform of myosin I from Acanthamoeba castellanii. J. Biol. Chem. 264: 19333–39 (1989)
172. Maingon R. et al.: The tubulin genes of Trypanosoma cruzi. Eur. J. Biochem. 171: 285–291 (1988)
173. Maita G., Maita T. and Umegane T.: The primary structure of L-1 light chain of chicken fast skeletal muscle myosin and its genetic implication. FEBS Letters 126: 111–113 (1981)
174. Maita T., Chen J. I. and Matsuda G.: Amino-acid sequence of the 20,000-molecular-weight light chain of chicken gizzard-muscle myosin. Eur. J. Biochem. 117: 417–424 (1981)
175. Maita T. et al.: The primary structure of the myosin head. Proc. Nat. Acad. Sci. USA 84: 416–420 (1987)
176. Manseau L. J., Ganetzky B. and Craig E. A.: Molecular and genetic characterization of the Drosophila melanogaster 87E actin gene region. Genetics 119: 407–420 (1988)
177. Martinez I. et al.: Comparsison of myosin isoenzymes present in skeletal and cardiac muscles of the Arctic charr Salvelinus alpinus (L.). Sequential expression of different myosin heavy chains during development of the fast white skeletal muscle. Eur. J. Biochem. 195: 743–753 (1991)
178. Matsuda G. et al.: Amino acid sequences of the cardiac L-2A, L-2B and gizzard 17,000-Mr light chains of chicken muscle myosin. FEBS Letters 135: 232–236 (1981)
179. McIntosh J. R. and Porter M. E.: Enzymes for microtubule-dependent motility. J. Biol. Chem. 264: 6001–04 (1989)
180. McKeon F. D., Kirschner M. W. and Caput D.: Homologies in both primary and secondary structure between nuclear envelope and intermediate filament proteins. Nature 319: 463–468 (1986)
181. McNally E. M. et al.: Full-length rat alpha and beta cardiac myosin heavy chain sequences. Comparisons suggest a molecular basis for functional differences. J. mol. Biol. 210: 665–671 (1989)
182. Michel F. and Rudloff V.: Isolation and characterization of the rainbow trout erythrocyte band-3 protein. Eur. J. Biochem. 181: 181–187 (1989)
183. Millake D. B. et al.: The cDNA sequence of a human placental α-actinin. Nucleic Acids Res. 17: 6725 (1989)
184. Milner R. E. et al.: Calreticulin, and not calsequestrin, is the major calcium binding protein of smooth muscle sarcoplasmic reticulum and liver endoplasmic reticulum. J. Biol. Chem. 266: 7155–65 (1991)
185. Mimura T., Yamanobe T. and Sugi H.: Isolation and characterization of calmodulin from a molluscan smooth muscle. Comp. Biochem. Physiol. Pt. B 81: 559–563 (1985)
186. Miwa T. and Kamada S.: The nucleotide sequence of a human smooth muscle (enteric type) gamma-actin cDNA. Nucleic Acids Res. 18: 4263 (1990)
187. Miyazaki J. I., Sekiguchi K. and Hirabayashi T.: Tissue specificity of tropomyosin from a horseshoe crab, Tachypleus tridentatus. Comp. Biochem. Physiol. Pt. B 85: 679–685 (1986)
188. Mohun T. et al.: A third striated muscle actin gene is expressed during early development in the amphibian Xenopus laevis. J. mol. Biol. 202: 67–76 (1988)
189. Molina M. I. et al.: The sequence of an embryonic myosin heavy chain gene and isolation of its corresponding cDNA. J. biol. Chem. 262: 6478–88 (1987)
190. Moncrief N. D., Kretsinger R. H. and Goodman M.: Evolution of EF-hand calcium-modulated proteins. I. Relationships based on amino acid sequences. J. mol. Evol. 30: 522–562 (1990)
191. Moon R. T. and McMahon A. P.: Generation of diversity in nonerythroid spectrins. Multiple polypeptides are predicted by sequence analysis of cDNAs emcompassing the coding region of human nonerythroid α-spectrin. J. Biol. Chem. 265: 4427–33 (1990)
192. Moormann R. J. M. et al.: Strict co-linearity of genetic and protein folding domains in an intragenically duplicated rat lens γ-crystallin gene. J. mol. Biol. 171: 353–368 (1983)
193. Morishima I. and Bodnaryk R. P.: Amino acid composition of insect calmodulins lacking trimethyllysine. Comp. Biochem. Physiol. Pt. B 80: 419–423 (1985)
194. Moriyama K. et al.: Destrin, a mammalian actin-depolymerizing protein, is closely related to cofilin. Cloning and expression of porcine brain destrin cDNA. J. Biol. Chem. 265: 5768–73 (1990)
195. Morris G. L. et al.: Phospholamban regulation of cardiac sarcoplasmic reticulum (Ca-Mg)-ATPase. J. Biol. Chem. 266: 11270–75 (1991)
196. Mounier N., Gaillard J. and Prudhomme J. C.: Nucleotide sequence of the coding region of two actin genes in Bombyx mori. Nucleic Acids Res. 15: 2781 (1987)
197. Mulders J. W. M. et al.: lambda-Crystallin, a major rabbit lens protein, is related to hydroxyacyl-coenzyme A dehydrogenase. J. Biol. Chem. 263: 15462–66 (1988)
198. Murayama J. I., Utsumi H. and Hamada A.: Amino acid sequence of monkey erythrocyte glycophorin MK. Its amino acid sequence has a striking homology with that of human glycophorin A. Biochim. biophys. Acta 999: 273–280 (1989)
199. Murphy D. B. et al.: The sequence and expression of the divergent β-tubulin in chicken erythrocytes. J. biol. Chem. 262: 14305–12 (1987)
200. Myers M. W. et al.: The human mid-size neurofilament subunit: a repeated protein sequence and the relationship of its gene to the intermediate filament gene family. Embo J. 6: 1617–26 (1987)
201. Mykles D. L.: Heterogeneity of myofibrillar proteins in lobster fast and slow muscles – Variants of troponin, paramyosin, and myosin light chains comprise 4 distinct protein assemblages. J. exp. Zool. 234: 23–32 (1985)
202. Nakajima-Iijima S. et al.: Molecular structure of the human cytoplasmic β-actin gene: interspecies homology of sequence in the introns. Proc. Nat. Acad. Sci. USA 82: 6133–37 (1985)
203. Nikovits jr. W., Kuncio G. and Ordahl C. P.: The chicken fast skeletal muscle troponin I gene: exon organization and sequence. Nucleic Acids Res. 14: 3377–90 (1986)
204. Nojima H.: Structural organization of multiple rat calmodulin genes. J. mol. Biol. 208: 269–282 (1989)
205. Obar R. A. et al.: Molecular cloning of the microtubule-associated mechanochemical enzyme

dynamin reveals homology with a new family of GTP-binding proteins. Nature 347: 256–261 (1990)
206. Ohshima S. et al.: Generation of multiple troponin T isoforms is a common feature of the muscles in various chordate animals. Comp. Biochem. Physiol. Pt. B 90: 779–784 (1988)
207. Ojima T. and Nishita K.: Troponin from akazara scallop striated muscles. J. biol. Chem. 261: 16749–54 (1986)
208. Okaimatsuo Y. et al.: A novel myosin heavy chain isoform in the vascular smooth muscle. Biochem. biophys. Res. Commun. 176: 1365–70 (1991)
209. Opperman L. and Ross F. P.: The adult fruit bat (Rousettus aegyptiacus) expresses only calbindin-D9K (vitamin-D-dependent calcium-binding protein) in its kidney. Comp. Biochem. Physiol. Pt. B 97: 295–299 (1990)
210. Padan R. et al.: Isolation an characterization of the Drosophila nuclear envelope otefin cDNA. J. Biol. Chem. 265: 1808–13 (1990)
211. Palmisano W. A. and Henzl M. T.: Partial nucleotide sequence of the parvalbumin from chicken thymus designated „avian thymus hormone". Biochem. biophys. Res. Commun. 167: 1286–93 (1990)
212. Pankov R. G. et al.: Vertebrate liver cytokeratins: a comparative study. Biochem. cell. Biol. 65: 547–557 (1987)
213. Parmentier M., Lawson D. E. M. and Vassart G.: Human 27-kDa calbindin complementary DNA sequence – Evolutionary and functional implications. Eur. J. Biochem. 170: 207–215 (1987)
214. Parry D. A. D., Conway J. F. and Steiner P. M.: Structural studies on lamin. Similarities and differences between lamin and intermediate-filament proteins. Biochem. J. 238: 305–308 (1986)
215. Pearlstone J. R., Carpenter M. R. and Smillie L. B.: Amino acid sequence of rabbit cardiac troponin T. J. biol. Chem. 261: 16795–810 (1986)
216. Pegues J. C. and Friedberg F.: Multiple messenger RNAs encoding human calmodulin. Biochem. biophys. Res. Commun. 172: 1145–49 (1990)
217. Perret C., Lomri N. and Thomasset M.: Evolution of the EF-hand calcium-binding protein family: Evidence for exon shuffling and intron insertion. J. mol. Evol. 27: 351–364 (1988)
218. Peter B. et al.: Mouse cytoskeletal gamma-actin: Analysis and implications of the structure of cloned cDNA and processed pseudogenes. J. mol. Biol. 203: 665–675 (1988)
219. Peter M. et al.: Cloning and sequencing of cDNA clones encoding chicken lamins A and B1 and comparison of the primary structures of vertebrate A- and B-type lamins. J. mol. Biol. 208: 393–404 (1989)
220. Pette D.: Metabolic heterogeneity of muscle fibers. J. exp. Biol. 115: 179–189 (1985)
221. Phillips M. A. et al.: Primary structure of keratinocyte transglutaminase. Proc. Nat. Acad. Sci. USA 87: 9333–37 (1990)
222. Phillips S. B. et al.: The pancornulins: a group of basic low molecular weight proteins in mammalian epidermis and epithelium that may function as cornified envelope precursors. Comp. Biochem. Physiol. Pt. B 95: 781–788 (1990)
223. Pochet R., Lawson D. E. M. and Heizmann C. W. (eds.): Calcium binding proteins in normal and transformed cells. Plenum, New York 1990
224. Pollard T. D. and Cooper J. A.: Actin and actin-binding proteins. A critical evaluation of mechanisms and functions. Annual Rev. Biochem. 55: 987–1035 (1986)
225. Ponnambalam S. et al.: Conservation and diversity in families of coated vesicle adaptins. J. Biol. Chem. 265: 4814–20 (1990)
226. Powell B. C. et al.: Clustered arrangement of keratin intermediate filament genes. Proc. Nat. Acad. Sci. USA 83: 5048–52 (1986)
227. Pratt L. F. and Cleveland D. W.: A survey of the α-tubulin gene family in chicken: unexpected sequence heterogeneity in the polypeptides encoded by five expressed genes. Embo J. 7: 931–940 (1988)
228. Presland R. B., Whitbread L. A. and Rogers G. E.: Avian keratin genes. II. Chromosomal arrangement and close linkage of three gene families. J. mol. Biol. 209: 561–576 (1989)
229. Price G. J. et al.: Primary sequence and domain structure of chicken vinculin. Biochem. J. 259: 453–461 (1989)
230. Ramaekers F. et al.: Classification of rat lens crystallins and identification of proteins encoded by rat lens mRNA. Eur. J. Biochem. 128: 503–508 (1982)
231. Rearden A. et al.: Evolution of the glycophorin gene family in the hominoid primates. Biochem. Genet. 28: 209–222 (1990)
232. Rees D. J. G. et al.: Sequence and domain structure of talin. Nature 347: 685–688 (1990)
233. Reeves S. A. et al.: Molecular cloning and primary structure of human glial fibrillary acidic protein. Proc. Nat. Acad. Sci. USA 86: 5178–82 (1989)
234. van Rens G. L. M. et al: Structure of the bovine eye lens gamma-s-crystallin gene (formerly βs). Gene 78: 225–233 (1989)
235. Rodgers M. E. et al.: Thermal stability of myosin rod from various species. Biochemistry 26: 8703–08 (1987)
236. Rogers G. E. et al. (eds.): The biology of wool and hair. Chapman & Hall, New York 1989
237. Rothnagel J. A. and Steinert P. M.: The structure of the gene for mouse filaggrin and a comparison of the repeating units. J. Biol. Chem. 265: 1862–65 (1990)
238. Rüegg J. C.: Calcium in muscle activation. A comparative approach, corr. 2nd printing. Springer, Berlin 1988
239. Ruiz-Opazo N. and Nadal-Ginard B.: α-Tropomyosin gene organization. Alternative splicing of duplicated isotype-specific exons accounts for the production of smooth and striated muscle isoforms. J. biol. Chem. 262: 4755–65 (1987)
240. Sahr K. E. et al.: The complete DNA and polypeptide sequences of human erythroid α-spectrin. J. Biol. Chem. 265: 4434–43 (1990)
241. Sasaki T., Tanokura M. and Asaoka K.: The complete amino acid sequence of bullfrog (Rana catesbeiana) pavalbumin Pi4.97. FEBS Letters 268: 249–251 (1990)
242. Schachat F. H., Bronson D. D. and McDonald O. B.: Heterogenity of contractile proteins. A continuum of troponin-tropomyosin expression in mammalian skeletal muscle. J. biol. Chem. 260: 1108–13 (1985)
243. Schaefer W. H. et al.: Amino acid sequence of a novel calmodulin from Paramecium tetraurelia that contains dimethyllysine in the first domain. J. biol. Chem. 262: 1025–29 (1987)
244. Schliwa M.: Cytoskeleton. An introductory survey. Springer, Wien 1986

245. Sen Gupta B., Friedberg F. and Detera-Wadleigh S. D.: Molecular analysis of human and rat calmodulin complementary DNA clones. Evidence for additional active genes in these species. J. biol. Chem. 262: 16663–70 (1987)
246. Shpetner H. S. and Vallee R. B.: Identification of dynamin, a novel mechanochemical enzyme that mediates interactions between microtubules. Cell 59: 421–432 (1989)
247. Siezen R. J. et al.: Human lens gamma-crystallins: Isolation, identification and characterization of the expressed gene products. Proc. Nat. Acad. Sci. USA 84: 6088–92 (1987)
248. Simmen R. C. M. et al.: The structural organization of the chicken calmodulin gene. J. biol. Chem. 260: 907–912 (1985)
249. Smillie L. B., Golosinska K. and Reinach F. C.: Sequence of complete cDNA encoding four variants of chicken skeletal muscle troponin T. J. Biol. Chem. 263: 18816–20 (1988)
250. Smith C. W. J., Pritchard K. and Marston S. B.: The mechanism of calcium regulation of vascular smooth muscle thin filaments by caldesmon and calmodulin. J. biol. Chem. 262: 116–122 (1987)
251. Smith V. L. et al.: Structure and sequence of the Drosophila melanogaster calmodulin gene. J. mol. Biol. 196: 471–485 (1987)
252. Sobue K. and Sellers J. R.: Caldesmon, a novel regulatory protein in smooth muscle and nonmuscle actomyosin systems (Minireview). J. Biol. Chem. 266: 12115–18 (1991)
253. Soellner P., Quinlan R. A. and Franke W. W.: Identification of a distinct soluble subunit of an intermediate filament protein: Tetrameric vimentin from living cells. Proc. Nat. Acad. Sci. USA 82: 7929–33 (1985)
254. Sparrow L. G. et al.: The amino acid sequence of component 7c, a type II intermediate-filament protein from wool. Biochem. J. 261: 1015–22 (1989)
255. Speicher D. W. and Marchesi V. T.: Erythrocyte spectrin is comprised of many homologous helical segments. Nature 311: 177–179 (1984)
256. Spithill T. W. and Samaras N.: The molecular karyotype of Leishmania major and mapping of α- and β-tubulin gene families to multiple unlinked chromosome loci. Nucleic Acids Res. 13: 4155–69 (1985)
257. Squire J. M. (ed.): Molecular mechanisms in muscle contraction. MacMillan, New York 1990
258. Srihari T. et al.: Purification and characterization of myosins from human and rabbit skeletal muscles by using specific monoclonal antibodies. Comp. Biochem. Physiol. Pt. B 85: 523–529 (1986)
259. Stearns T., Evans L. and Kirschner M.: Gamma-Tubulin is a highly conserved component of the centrosome. Cell 65: 825–836 (1991)
260. Stedman H. H. et al.: The human embryonic myosin heavy chein. Complete primary structure reveals evolutionary relationships with other developmental isoforms. J. Biol. Chem. 265: 3568–76 (1990)
261. Steinert P. M., Steven A. C. and Roop D. R.: The molecular biology of intermediate filaments. Cell 42: 411–419 (1985)
262. Steinert P. M.: The two-chain coiled-coil molecule of native epidermal keratin intermediate filaments is a type I-type II heterodimer. J. Biol. Chem. 265: 8766–74 (1990)
263. Stevens F. C.: Calmodulin, an introduction. Can. J. Biochem. Cell Biol. 61: 906–910 (1983)
264. Stieger J., Wyler T. and Seebeck T.: Partial purification and characterization of microtubular protein from Trypanosoma brucei. J. biol. Chem. 259: 4596–4602 (1984)
265. Stossel T. P.: From signal to pseudopod. How cells control cytoplasmic actin assembly (Minireview). J. Biol. Chem. 264: 18261–64 (1989)
266. Suedhof T. C.: The structure of the human synapsin I gene and protein. J. Biol. Chem. 265: 7849–52 (1990)
267. Sugi H. and Pollack G. (eds.): Molecular mechanism of muscular contraction. Plenum, New York 1988
268. Sullivan K. F. et al.: Sequence and expression of the chicken β3 tubulin gene. A vertebrate testis β-tubulin isotype. J. biol. Chem. 261: 13317–22 (1986)
269. Swanson M. E., Sturner S. F. and Schwartz J. H.: Structure and expression of the Aplysia californica calmodulin gene. J. mol. Biol. 216: 545–553 (1990)
270. Swaroop A., Swaroop M. and Garen A.: Sequence analysis of the complete cDNA and encoded polypeptide for the glued gene of Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 84: 6501–05 (1987)
271. Takagi T. et al.: The amino acid sequence of the calmodulin obtained from sea anemone (Metridium senile) muscle. Biochem. biophys. Res. Commun. 96: 377–381 (1980)
272. Takagi T., Kobayashi T. and Konishi K.: Amino-acid sequence of sarcoplasmic calcium-binding protein from scallop (Patinopecten yessoensis) adductor striated muscle. Biochim. biophys. Acta 787: 252–257 (1984)
273. Tagaki T., Kudoh S and Konishi K.: The amino acid sequence of ascidian (Halocynthia roretzi) myosin light chains. Biochim. biophys. Acta 874: 318–325 (1986)
274. Takagi T. and Cox J. A.: Primary structure of the target of calcium vector protein of amphioxus. J. Biol. Chem. 265: 19721–27 (1990)
275. Takagi T. and Cox J. A.: Amino acid sequences of four isoforms of amphioxus sarcoplasmic calcium-binding proteins. Eur. J. Biochem. 192: 387–399 (1990)
276. Takagi T. et al.: Primary structure of profilins from 2 species of echinoidea and Physarum polycephalum. Eur. J. Biochem. 192: 777–781 (1990)
277. Takahashi K. and Nadalginard B.: Molecular cloning and sequence analysis of smooth muscle calponin. J. Biol. Chem. 266: 13284–88 (1991)
278. Takahashi S., Takano-Ohmuro H. and Maruyama K.: Regulaton of Drosophila myosin ATPase activity by phosphorylation of myosin light chains. II. Flightless mfd-fly. Comp. Biochem. Physiol. Pt. B 95: 183–185 (1990)
279. Takemasa T. et al.: The third calmodulin family protein in Tetrahymena. Cloning of the cDNA for Tetrahymena calcium-binding protein of 23 kDa (TCBP-23). J. Biol. Chem. 265: 2514–17 (1990)
280. Takemoto L. J., Hansen J. S. and Horwitz J.: Interspecies conservation of the main intrinsic polypeptide (MIP) of the lens membrane. Comp. Biochem. Physiol. Pt. B 68: 101–106 (1981)
281. Tang T. K. et al.: Selective expression of an erythroid-specific isoform of protein 4.1. Proc. Nat. Acad. Sci. USA 85: 3713–17 (1988)
282. Teumer J. and Green H.: Divergent evolution of part of the involucrin gene in the homoids. Unique intragenic duplications in the gorilla and human. Proc. Nat. Acad. Sci. USA 86: 1283–86 (1989)
283. Theurkauf W. E. et al.: Tissue-specific and constitutive α-tubulin genes of Drosophila melanogaster code

for structurally distinct proteins. Proc. Nat. Acad. Sci. USA 83: 8477–81 (1986)
284. Tomarev S. I. and Zinovieva D.: Squid major lens polypeptides are homologous to glutathione S-transferase subunits. Nature 336: 86–88 (1988)
285. Toyofuku T. and Zak R.: Characterization of cDNA and genomic sequences encoding a chicken phospholamban. J. Biol. Chem. 266: 5375–83 (1991)
286. Toyoshima Y. Y., Kron S. J. and Spudich J. A.: The myosin step size. Measurement of the unit displacement per ATP hydrolyzed in an in vitro assay. Proc. Nat. Acad. Sci. USA 87: 7130–34 (1990)
287. Treveso S. et al.: Frog brain expresses a 60 kDa calcium-binding protein similar to mammalian calreticulin. Biochem. biophys. Res. Commun. 175: 444–450 (1991)
288. Vandekerckhove J. and Weber K.: Chordate muscle actins differ distinctly from invertebrate muscle actins. The evolution of the different vertebrate muscle actins. J. mol. Biol. 179: 391–413 (1984)
289. Vandekerckhove J., Lal A. A. and Korn E. D.: Amino acid sequence of Acanthamoeba actin. J. mol. Biol. 172: 141–147 (1984)
290. Vandekerckhove J., Bugaisky G. and Buckingham M.: Simultaneous expression of skeletal muscle and heart actin proteins in various striated muscle tissues and cells. A quantitative determination of the two actin isoforms. J. biol. Chem. 261: 1938–43 (1986)
291. Vignal A. et al.: A novel gene member of the human glycophorin A and B gene family: Molecular cloning and expression. Eur. J. Biochem. 191: 619–625 (1990)
292. Vilasante A. et al.: Six mouse α-tubulin mRNAs encode five distinct isotypes: Testis-specific expression of two sister genes. Mol cell. Biol. 6: 2409–19 (1986)
293. Volpe P. et al.: „Calciosome", a cytoplasmic organelle: the inositol 1,4,5-triphosphate-sensitive Calcium-store of nonmuscle cells? Proc. Nat. Acad. Sci. USA 85: 1091–95 (1988)
294. Volpe P. and Simon B. J.: The bulk of calcium released to the myoplasm is free in the sarcoplasmic reticulum and does not unbind from calsequestrin. FEBS Letters 278: 274–278 (1991)
295. Voorter C. E. M. et al.: Age-dependent deamidation of chicken αA-crystallin. FEBS Letters 221: 249–252 (1987)
296. Wang S. M., Sun M. C. and Jeng C. J.: Location of the C-terminus of titin at the Z-line region in the sarcomere. Biochem. biophys. Res. Commun. 176: 189–193 (1991)
297. Ward S. et al.: Genomic organizations of major sperm protein genes and pseudogenes in the nematode Caenorhabditis elegans. J. mol. Biol. 199: 1–13 (1988)
298. Wasenius V. M. et al.: α-Actinin and spectrin have common structural domains. FEBS Letters 221: 73–76 (1987)
299. Watabe S. and Hartshorne D. J.: Paramyosin and the catch mechanism. Comp. Biochem. Physiol. Pt. B 96: 639–646 (1990)
300. Weber K., Plessmann U. and Ulrich W.: Cytoplasmic intermediate filament proteins of invertebrates are closer to nuclear lamins than are vertebrate intermediate filament proteins; sequence characterization of two muscle proteins of a nematode. Embo J. 8: 3221–27 (1989)
301. Weeds A. G. et al.: Preparation and characteriziation of pig plasma and platelet gelsolins. Eur. J. Biochem. 161: 69–76 (1986)
302. Weller P. A. et al.: Complete sequence of human vinculin and assignment of the gene to chromosome 10. Proc. Nat. Acad. Sci. USA 87: 5667–71 (1990)
303. Wesseling J. G. et al.: Nucleotide sequence and expression of a β-tubulin-encoding gene from Plasmodium falciparum, a malarial parasite of man. Gene 83: 301–309 (1989)
304. Whitbread L. A., Gregg K. and Rogers G. E.: The structure and expression of a gene encoding chick claw keratin. Gene 101: 223–229 (1991)
305. Wiche G.: High-Mt microtubule-associated proteins: properties and functions. Biochem. J. 259: 1–12 (1989)
306. Wilson P. W. et al.: Structure of chick chromosomal genes for calbindin and calretinin. J. mol. Biol. 200: 615–625 (1988)
307. Winder S. J. and Walsh M. P.: Smooth muscle calponin. Inhibition of actomyosin MgATPase and regulation by phosphorylation. J. Biol. Chem. 265: 10148–55 (1990)
308. Winkelmann J. C. et al.: Full-length sequence of the cDNA for human erythroid β-spectrin. J. Biol. Chem. 265: 11827–32 (1990)
309. Winrow M. A. and Sodja A.: Initial characterization of several actin genes in the parasitic nematode, Ascaris suum. Biochem. biophys. Res. Commun. 178: 578–585 (1991)
310. Wistow G. J. and Piatigorsky J.: Gene conversion and splice-site slippage in the argininosuccinate lyases delta-crystallins of the duck lens: Members of an enzyme superfamily. Gene 96: 263–270 (1990)
311. Wistow G., Anderson A. and Piatigorsky J.: Evidence for neutral and selective processes in the recruitment of enzyme-crystallins in avian lenses. Proc. Nat. Acad. Sci. USA 87: 6277–80 (1990)
312. Wistow G. and Kim H.: Lens protein expression in mammals. Taxon-specificity and the recruitment of crystallins. J. mol. Evol. 32: 262–269 (1991)
313. Wohlfarth-Bottermann K. E. (ed.): Nature and function of cytoskeletal proteins in motility and transport. G. Fischer, Stuttgart 1987
314. Xiang M. et al.: Tandem duplication and divergence of a sea urchin protein belonging to the troponin C superfamily. J. Biol. Chem. 263: 17173–80 (1988)
315. Yaffe D. et al.: Highly conserved sequences in the 3' untranslated region of mRNAs coding for homologous proteins in distantly related species. Nucleic Acids Res. 13: 3723–37 (1985)
316. Yamanaka M. K. et al.: Structure and expression of the Drosophila calmodulin gene. Nucleic Acids Res. 15: 3335–48 (1987)
317. Yazaki P. J. et al.: Calsequestrin, an intracellular calcium-binding protein of skeletal muscle sarcoplasmic reticulum, is homologous to aspartactin, a putative laminin-binding protein of the extracellular matrix. Biochem. biophys. Res. Commun. 166: 898–903 (1990)
318. Yazawa M. et al.: The amino acid sequence of the Tetrahymena calmodulin which specifically interacts with guanylate cyclase. Biochem. biophys. Res. Commun. 99: 1051–57 (1981)
319. Zarain-Herzberg A., Fliegel L. and MacLennan D. H.: Structure of the rabbit fast-twitch skeletal muscle calsequestrin gene. J. Biol. Chem. 263: 4807–12 (1988)
320. Zehner Z. E. et al.: The chicken vimentin gene. Nucleotide sequence, regulatory elements, and comparison of the hamster gene. J. biol. Chem. 262: 8112–20 (1987)

11 Extracellular Structural and Secretory Proteins

The **extracellular matrices** of animal tissues consist of networks of collagens and elastins, the meshes of which are filled with structural glycoproteins and glycosaminoglycans. The cells themselves are in direct contact with the basement membranes, which are specific regions of the extracellular matrix with a thickness of 20–200 nm. The very varied compositions of the different extracellular matrices determine their mechanical characters and, in particular, their interactions with the cells [73]. The binding of cells to various matrix components is mediated by specific receptors on the cell surface and these recognize certain peptide sequences in the matrix proteins [141]. Several particularly important and well-known **protein types of the extracellular matrix** will be dealt with in Sections 11.1 to 11.4, and the glycosaminoglycans will be discussed in Chapter 13. There are very few comparative biochemical data on the other matrix components. Next to the collagens, the most predominant cartilage protein is the cartilage matrix protein (CMP), a homotrimer with subunits of 470 amino acids, the gene for which has been sequenced in the chicken [69]. Tendons, skin and cornea contain the collagen-binding fibromodulin, and the gut wall includes the homologous proteoglycans PG-S1 and PG-S2 (decorin); the 357- to 375-amino-acid polypeptide chains of these latter proteins are mainly built up of 23-residue repeats [92]. In addition to fibronectin, the other cell-binding proteins of the extracellular matrix are cytotactin and tenascin, which are mainly produced during the phases of histo- and organogenesis. In contrast to the dimeric fibronectin molecule, these two proteins form oligomers with a six-armed structure (hexabrachions). Their sequences of 1777 and 2203 amino acids, respectively, are highly repetitive; some of the repeats resemble the type-III repeats of fibronectin, and the others are similar to the repeats of the epidermal growth factor (EGF) [19, 90]. Endothelial cells, fibroblasts and smooth-muscle cells produce the large glycoprotein thrombospondin, which is able to bind to fibronectin, collagen and heparin in the extracellular matrix [25].

Typical **components of the basement membranes** are type-IV collagens, glycosaminoglycans, laminin, merosin and nidogen [129]. The insolubility of the basement membrane makes for large problems in the characterization of these proteins; significant progress has only been possible through analysis of cDNA and gene sequences. Laminin can be extracted with ethylenediaminetetraacetic acid (EDTA) and consists of three chains, A, B1 and B2. The N-terminal regions of the subunits take the form of three short, free arms; the C-terminal regions form a rod-like coil on the end of which sits a large globule made up of the C-terminus of the long A chain. Collagen IV binds to both the short terminal globules of the short arms and the large C-terminal globule of the A chain. The N-terminus of the A chain carries a cell-binding site. The binding of neuronal cells to this site stimulates the outgrowth of neurites. Further cell-binding sites, as

well as the sites responsible for the binding of nidogen/entactin and for calcium-dependent aggregation, lie on the short arms of the B chains. The large globule at the C-terminus of the A chains binds glycosaminoglycans [8]. The sequences of the human and murine A chains, with 3058–3060 amino acids, agree by 76 % and, with the exception of the globular region, are significantly homologous to the B chains. The various regions of the human B1 (1765 amino acids) and B2 (1576 amino acids) chains agree in their sequences by only 16–40 %. The B1 gene includes 34 exons and the B2 gene only 28; of the introns, only three occupy the same positions in both genes [60, 90]. Laminins have been detected in various invertebrates, but only the B chain of *Drosophila melanogaster* has been sequenced. The B2 chain of the fly agrees at 40–41 % of positions with human and murine B2, but at only 29–30 % with its own B1 chain or the B1 of the mammals [18]. The basement membranes of the trophoblasts from human placenta, of striated skeletal muscle and of the Schwann cells from the peripheral nervous system contain a specific laminin made up of the polypeptide merosin and the laminin chains B1 and B2. The C-terminal region of merosin, as yet the only characterized region of the molecule, shows 40 % sequence similarity to the laminin A chain [30]. The laminin of the basement membranes is often complexed with nidogen (entactin). Murine nidogen is a dumb-bell-shaped molecule carrying binding sites for cells, laminin and collagen IV. The globular N- and C-terminal domains of nidogen consist of 641 and 328 amino acids, respectively, and the connecting rod domain has 248 amino acids [77]. Glutactin from the basement membranes of *Drosophila* may be functionally similar to nidogen but structurally it is very different; the N-terminal half of its sequence of 1023 amino acids resembles serine enzymes, but catalytic activity is absent; the C-terminal half contains no fewer than 44 % glutamate and glutamine residues [93].

11.1 Fibronectins

Fibronectins are found in vertebrates as insoluble structural proteins on the surface of fibroblasts and other cells, in the connective tissue matrix, and between the basement membrane and the associated cells; they are also found in soluble form as cell-adhesion factors in the blood plasma. They are present at a much greater concentration in embryonal tissues than in the adult, and are thought to have an organizing function in tissue formation. Only the structures of the human, bovine and rat soluble forms are known in any detail. In these cases the proteins are dimeric, extended, globular glycoproteins of 220 kDa with the two chains bound by disulphide bridges close to the C-terminus (Fig. 11.1). The chains are identical, or nearly identical, and consist of 2146–2325 amino acids. The sequence is periodic with three types of repeats: 12 type-I domains and 9 type-II domains are present, each of which is composed of 45–50 amino acids with internal disulphide bridges; and there are also 15 type-III domains with 90 amino acids but no bridges. The species-specific carbohydrate fraction of 4.5–9.5 % is made up of three to six N-glycosidic-bound oligosaccharides [109, 116].

The fibronectins are extraordinarily varied proteins. They are involved in the maintenance of normal cell form, cell adhesion, spreading and movement, the stimulation of phagocytosis (opsonification) and wound healing; they bind collagen, fibrin, actin and other proteins of the membrane skeleton, heparin, hyaluronic acid, gangliosides, polyamines, the complement com-

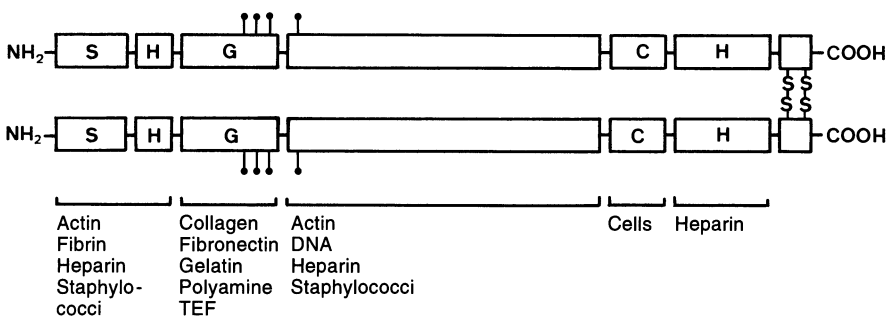


Fig. 11.1. The domain structure of human plasma fibronectin [131]. The domains are labelled according to their ligands as *S* (staphylococci), *H* (heparin), *G* (gelatin) and *C* (cell); glycosylation sites are marked with *drumsticks*

ponents C3 and C1q, C-reactive proteins, amyloid P, acetylcholinesterase, DNA, and bacterial and animal cells [109, 116]. Furthermore, they constitute the recognition and binding sites for the Chagas disease agent *Trypanosoma cruzi*. Each function has a **specific binding site** somewhere on the long polypeptide chain (Fig. 11.1). Fibronectins have been found in all vertebrates examined, and related proteins appear to be present in all Metazoa from the Porifera onwards. An antibody against human fibronectin recognizes a glycoprotein of about 230 kDa in the haemolymph of the fly *Drosophila*, and this glycoprotein is apparently quite similar to the vertebrate fibronectin [45].

The soluble plasma fibronectins of the rat and other mammals can be clearly distinguished from the insoluble cellular forms by means of monoclonal antibodies. Both show marked variability, which is partly due to post-translational modifications, such as glycosylation or sulphation of tyrosine residues, but also partly due to differences in the primary sequences [109]. Man and the rat have only one **fibronectin gene**, from which, however, 12 different mRNAs are formed by **alternative splicing** at three positions (see Fig. 2.12a, p. 41). It is not known how far the different fibronectins exercise different functions [94, 109]. The rat fibronectin gene extends for about 70 kb and has almost 50 exons. Each repeated unit of the type-I and -II fibronectins is encoded by a specific exon, whereas the repeats of type III arise from two somewhat different exons. The primary translation product contains a signal sequence of 20 amino acids and a pro-sequence of 5 amino acids [47]. The gene structure confirms the participation of gene duplication and exon shuffling in **fibronectin evolution**, but this is already suggested by the periodic amino acid sequence [96]. The blood plasma of various mammals and the chicken contains a further cell-adhesive glycoprotein, **vitronectin**, which, with a length of 458 amino acids (75 kDa), is significantly smaller than fibronectin and shows no sequence similarity [70].

11.2 Collagens

In the early 1970s only one collagen type was recognised, and that is now known as collagen I from the extracellular matrix of the vertebrate connective tissues. Today, no fewer than 14 vertebrate collagen types are known, and these are made up of 25 different polypeptide chains. Further collagen types are found in the invertebrates. Collagens form fibrils in the extracellular matrix and networks in the basement membranes; they occur, for example, in the cuticula of the annelids and nematodes, in the egg capsules of sharks, skates and rays, in the cocoon silks of some insect larvae, in the byssus threads of the edible mussels, and in the gemmules of the sponges. Quantity-wise, they are amongst the most prevalent proteins; 25–30 % of total human protein consists of collagen, and there are collagen-like sequences in other proteins. For example, the asymmetric, basement-membrane-bound form of acetylcholinesterase has a 40-kDa segment with the triple-helix structure typical of collagen, and shows immunological cross-reactivity with collagen I. The three chains of the complement factor C1q also have a short triple helix close to the N-terminus, which suggests the combination of a collagen sequence with a non-collagen sequence during evolution. The neurophysins and various salivary proteins possess collagen-like fragments [35, 150].

11.2.1 Vertebrate Collagens

The typical **collagen fibre** of the mammalian connective-tissue matrix is a bundle of fibrils with more than 1 million collagen molecules per centimetre, each of which is shifted 67 nm with respect to the others by one so-called D unit, corresponding to 234 amino acids, giving the typical cross-striping of the fibrils (Fig. 11.2b). Each **collagen molecule** is a stiff, rod-shaped structure of about 300 kDa with a length of 300 nm and a diameter of 1.5 nm; it is made up of three α chains which are, for the most part, twisted into a triple helix (Fig. 11.2a). The central part of each chain consists predominantly of a repetitive sequence of about 330 triplets with the structure -Gly-X-Y-, in which X and Y are often proline and 4-hydroxyproline, respectively. At both ends of the α chains are so-called telopeptides, of 15–20 amino acids, which have no triple-helix structure. As in the case of other secreted proteins, the newly synthesized α chains have an N-terminal signal sequence. After cleavage of the latter in the cisternae of the endoplasmic reticulum, three chains associate to form the **pro-collagen molecule** which is then released to the outside. The α chains of pro-collagen carry N- and C-terminal extensions which are cleaved off by specific amino-and carboxypro-collagen peptidases during fibril formation. Further **post-**

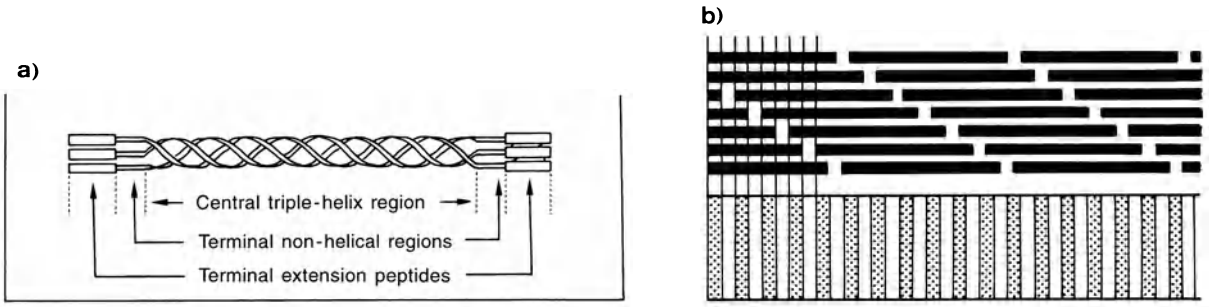


Fig. 11.2a, b. Collagen structure. **a** Pro-collagen – the terminal extensions or propeptides are cleaved during fibril formation by specific pro-collagen peptidases. **b** Collagen fibrils – the characteristic cross-striping arises from the staggered arrangement of the collagen molecules

translational modifications of collagen involve the hydroxylation of proline and lysine residues, and the glycosylation of particular hydroxylysines [35].

In the **collagen fibrils**, two, three or four molecules are connected together by special structures; thus, the fibres are really networks. Up to ten different **connecting structures (cross-links)** may be present on one molecule, and not all of these have yet been identified (Fig. 11.3). In collagen, as in elastin, the cross-linking is based on lysine or hydroxylysine, although at the most 3 lysine residues are involved in each collagen molecule compared with more than 30 in the elastin molecule. Invertebrate collagens, in particular, may also be cross-linked by disulphide bridges or by structures based on tyrosine (Fig. 11.4). The formation of the connecting structures begins with the conversion of certain lysine or hydroxylysine residues of collagen to the corresponding aldehydes allysine and hydroxyallysine by oxidation of the side-chain (Fig. 11.3a–b). Lysyloxidase is the only enzyme required for the cross-linking of collagen and elastin. This is a copper enzyme of 710 amino acids with pyridoxalphosphate as the cofactor, which interacts with growing fibrils more readily than with collagen monomers. The sequence of this enzyme shows only 18–19% agreement with that of the functionally related prolyl-4-hydroxylase [87]. In most connective tissues the cross-links are formed via hydroxyallysine, and in the skin via allysine; either of these two possibilities may predominate in tendons, e.g. allysine occurs in rat-tail tendons but hydroxy-

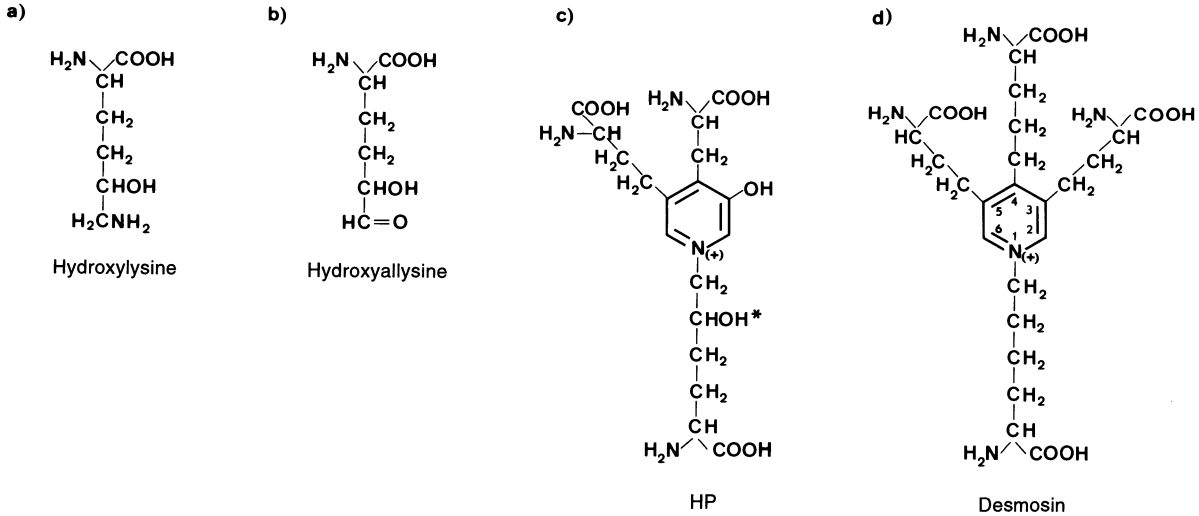


Fig. 11.3a–d. Cross-linking structures of collagen and elastin molecules and their precursors [32]. **a** Hydroxylysine; **b** hydroxyallysine; **c** cross-linking of collagens by the hydroxyallysine route involves formation of the 3-hydroxypyridine derivatives HP and LP, which differ only by the absence from LP of the hydroxyl group marked with an *asterisk*; **d** cross-linking of elastin results in the formation of desmosin and isodesmosin, the latter having a C2 instead of a C4 side-chain

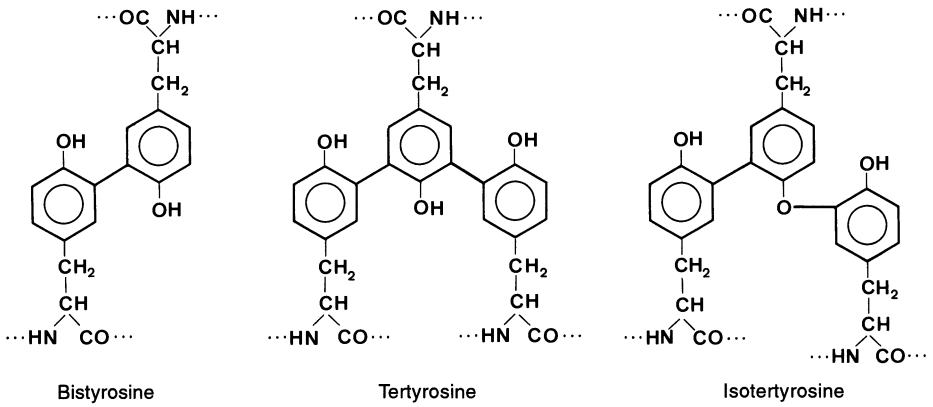


Fig. 11.4. The linking structures bistyrosine, tertyrosine and isotertyrosine, which arise from tyrosine residues

allysine is found in human Achilles tendon. In the hydroxyallysine route, the trifunctional 3-hydroxypyridine (HP) bonds are formed from three hydroxylysines and LP is formed from two hydroxylysines and a lysine (Fig. 11.3 c). The bonding structures of the allysine route are not known in detail. The formation of cross-bridges in collagens I–III always involves the same four sites: one each near the N- and the C-terminus and one each 90 positions distant in the triple-helix region. Only the terminal amino acid is oxidizable and reacts with the helical amino acids of a neighbouring molecule, which is shifted by 4 D units, i.e. by 268 nm [32, 150].

The 14 known **collagen types of mammals and birds** can be divided into three groups according to structure and localization. Types I–III, V and XI belong to the fibril-forming collagens. Their molecules consist of α chains of more than 95 kDa with a 300-nm uninterrupted triple helix, and form cross-striped fibrils with the typical D unit of 67 nm. In non-fibrillar collagens, the triple-helix domains are interrupted by non-helical segments. The α chains of the large collagens, IV, VI, VII, XII and XIV, have a mass of more than 95 kDa and those of the short-chain collagens, VIII, IX, X and XIII, are always smaller than this [99]. Amongst the **fibril-forming collagens**, type I predominates in the skin, bone, tendon and cornea; it is often associated with small amounts of type V and forms large-diameter fibrils with distinct cross-striping. Type II is found together with small amounts of type XI in cartilage and in the vitreous humour of the eye; the fibrils of type II are finer with less obvious periodicity. There are cartilaginous structures, e.g. the meniscus, which consist mainly of collagen I. Collagen type III occurs together with type I in the skin and blood-vessel walls. This is the only fibril-forming vertebrate collagen to have disulphide bridges in addition to the other bonding structures. The molecules of collagen I consist of two, or even three, different α chains; those of collagens II and III are homotrimers. Although together they can form a triple helix, the subunits of collagen I differ considerably from each other; e.g. human α1(I) and α2(I) differ in 34 % of their amino acids. The α chains from different collagen types show even larger differences, e.g. there is 42 % difference between human α2(I) and α1(III), and are unable to form triple helices with each other. In contrast, the species-specific difference between homologous chains is always much smaller, e.g. 7.9 % between human and chicken α1(I); bovine and rat α2(I) differ by 9 %, and the human/chicken difference is 21 %. These data suggest that the gene duplications leading to the separation of the different collagen types and subunits of any one type took place long before the origins of the birds and the mammals, and that the rate of evolution of the α2 chains is much greater than that of α1 [108]. In contrast to the higher vertebrates, the agreement between α1(I) and α2(I) in the agnathan *Entosphenus japonicus* is greater than that between α1(I) and α1(II). Thus, the separation of the subunits α1(I) and α2(I) probably occurred after the appearance of the types I and II. However, due to the higher rate of evolution of the α2 chains, the difference between α1 and α2 increases in parallel to the evolutionary rank of the vertebrates.

The collagens of type I mostly consist of two chain types according to the formula $(\alpha 1)_2\alpha 2$. Collagen I from the skin of many teleosts, however, includes a third chain type and has the for-

mula α1α2α3. The α3 chain is encoded by its own (third) gene which apparently arose by duplication of the α1(I) gene. The α3 chain has so far not been detected in the agnathans, cartilaginous fish or higher vertebrates [103]. The lampreys *Entosphenus japonicus* and *Eptatretus burgeri* possess two genetically different type-I collagens, one of which predominates in the skin and the other in the gut wall, muscle and cartilage [79]. The homologous collagen types I and II not only are up to 40 % different in amino acid sequence, but type II shows up to tenfold higher hydroxylation and glycosylation of the lysine residues. Both collagens show an increasing proline hydroxylation in higher vertebrates, and thus a higher temperature stability. The denaturing temperature for collagens I and II in mammals and birds is 37–39 °C; in fish, type II has a value of 32–34 °C and is relatively stable, whereas type I is already denatured at 28–29 °C in the sharks and cartilaginous fish, and at 22 °C in the agnathan *Myxine* [107]. Sharks also have predominately type I in the skin and type II in cartilage; however, the transparent, elastic fin fibres (ceratotrichia) contain **elastoidin**, which is a complex of collagen and a tyrosine-rich, non-collagen protein. This substance, which by the way is the basis of the famous delicacy shark-fin soup, contains a homotrimeric collagen with a component α1(E), which is very similar, but not identical, to α1(I) [67].

Collagens III, V and XI can alone form fibrils in vitro, but usually occur in addition to collagen I or II in complex (heterotypic) fibrils. About 5–20 % of collagens in various tissues are of **type III**, the proportion increasing with age [1]. Type V is usually associated with collagen I, and type XI with collagen II of cartilage, but fibrils composed of types V and II are also known. **Collagen V** exists as either $(\alpha1)_2\alpha2$, α1α2α3 or $(\alpha1)_3$, and a further chain form, α4(V), has been found in the chicken [91]. **Collagen XI** contains three polypeptides, of which α1(XI) and α2(XI) are encoded by their own individual genes, whilst α3(XI) is possibly a product of the α1(II) gene [68]. A collagen isolated form bovine bone contains α1(XI) in addition to α1(V) and α2(V) [91]. In fact, collagens V and XI are very similar; for example, the triple-helix sequences of the human collagen chains α1(V) and α1(XI) have 82 % identical amino acids [125].

Unlike collagens I–III and V, the basement membrane **collagen IV** does not form fibrils but forms a network. Its molecules usually consist of two α1(IV) subunits and one α2(IV), but homotrimers of α1(IV) have also been described. More recently, other tissue-specific type-IV collagen chains have been discovered; they are designated α3(IV), α4(IV) and α5(IV) and probably represent adaptations to the functions of the different tissues [55]. The human α1(IV) chain consists of 1642 amino acids, including 20 cysteine residues. These are subdivided into a short N-terminal region (15 amino acids), a central triple-helix domain of about 350 nm (1398 amino acids), and a large C-terminal globular domain NC1 (229 amino acids). The collagen network arises by binding of the terminal domains: in each case, four molecules are bound together at the N-terminal domains with 30-nm overlaps, and two molecules are bound head-to-head via the C-terminal domains. In the central domain, the repetitive Gly-X-Y sequence is interrupted at 21 positions. The NC1 domain is made up of repeats of two homologous sequences which agree in about 45 % of positions. The human and murine α2(IV) chains have a structure similar to that of human α1(IV) [13, 98]. Whereas the interstitial collagens I–III contain only about 1–2 % sugars, mainly galactose and glucose, bound O-glycosidic to hydroxylysines, the basement membrane collagen IV has a much larger and more complex carbohydrate component. The sequence agreement between homologous subunits of both human and murine α1(IV) and α2(IV) is about 65 % in the NC1 domain, and is therefore much greater than in the central domain (45 %); but also in this case, homologous subunits of different species are much more similar than are the two subunits in the same individual. For example, the human and murine α2(IV)-NC1 sequences agree in 219 of 227 amino acids [55].

The collagen fibrils of the extracellular matrix are often associated with further collagens. Thus, the fibrils of cartilage consist of types VI and IX in addition to collagens II and XI, and the fibrils of the cornea have type VI as well as types I and V. Collagen types VII, VIII and X-XIV are more-or-less tissue-specific side components of connective tissue. **Collagen VI** is localized in cartilage close to the chondrocytes and has some special features that are not found in other collagens. The monomer has the form of a dumb-bell made up of a short triple helix with a length of 105 nm and two large globular structures, which together make up more than 80 % of the mass of the molecule. It is composed of three α chains, of which α3(VI) bears a particularly large N-terminal, globular domain and, with a total of 2648 (chicken) to 2943 (human) amino acids, is much longer than chains α1(VI) and α2(VI), each of which has

about 1000 amino acids. The triple helix contains cysteine residues, which can form disulphide bridges between the monomers, as well as several putative cell-binding sites with the sequence Arg-Gly-Asp and N-glycosylation sites with the sequence Asn-X-Thr. The N-terminal domains of the chains α1(VI) and α2(VI) contain one sequence, and the large α3(VI) chains have eight sequences of 190–210 amino acids, which resemble the type-A domains of the Willebrand factor and, like the latter, also bind collagen I. Thus, collagen VI apparently has the function of a bridge between cells and collagen fibres [11, 20, 50]. The molecules of **collagen VII** consist of three identical subunits and occur in connective tissue as disulphide-linked dimers [83]. The short-chain **collagen VIII** is found in the majority of endothelial cells and, in particular, in the thick basement membrane of the cornea (Descemet's membrane). The molecule is a heterotrimer with the composition $(\alpha1)_2\alpha2$. The α chain is made up of a short triple helix of 454–457 amino acids and two terminal, non-helical domains of 117 and 173 amino acids [85].

Collagens IX and X are deposited in various cartilaginous structures formed from fibrils of collagens II and XI. **Collagen X**, which is closely related to type VIII, is expressed only at mineralization sites in hypertrophic chondrocytes; it is a homotrimer of short α chains which have triple helices with eight interruptions; the sequences of the chicken and bovine chains agree in 67 % of their 674 amino acids [127]. **Collagen IX** is a heterotrimer of three different α chains of 72–115 kDa and these include three triple-helix domains, COL1–COL3, and four flexible, non-helical domains, NC1–NC4 (the numbering beginning at the C-terminus). α2(IX) bears a covalently bound glycosaminoglycan on domain NC3. Collagen IX alone is unable to form fibrils, but it is covalently bound with collagen II of cartilage fibrils via lysine-derived cross-links. In this way, the collagen fibrils of the cartilage are completely surrounded by glycosaminoglycan, and this contributes to the high hydration state of the cartilage matrix. Two forms of the α1(IX) chain arise by alternative splicing. The cartilaginous form bears a large, globular domain, NC4, of 266 amino acids at the N-terminus; this extends beyond the fibril surface and can interact with other components of the matrix. The form found in the cornea and in the vitreous humour has a much shorter N-terminus [84, 140]. In addition to collagen IX, there are other fibril-associated collagens with interrupted triple helices (FACITs), e.g. **type XII** in the tendons of chicken embryos, **type XIII** in the endothelium of the human umbilical vein, and **type XIV** in calf tendon and skin [28, 43, 99].

The **genes** of the fibril-forming collagens I–III have 5–8 exons in the region of the N-propeptide and 44 exons in the triple-helix region, i.e. they have more interruptions than most other genes. Because correct splicing apparently requires a minimal intron length of about 80 bp, these genes with pro-collagen-coding sequences of less than 5 kb often occupy more than 40 kb of chromosomal DNA. Most of the exons have a length of 54 bp, and some have derived values like 45 (= 54–9), 108 (= 2 × 54) or 99 (= 108–9) bp. One exon of 54 bp encodes 18 amino acids or six G-X-Y triplets. Exons of 54 bp also predominate in collagens V and XI. The characteristic gene structure of the fibril-forming collagens has led to the concept that the genes of all collagens arose by the multiplication of a primordial sequence of 54 bp. However, further collagen genes that have been sequenced in the meantime do not fit this 54-bp hypothesis: the pro-collagen genes of type VI mostly have exons of 63 bp in addition to some with 27, 36, 45, 54 or 90 bp, i.e. multiples of a 9-bp unit. The human and murine α1(IV) genes contain 52 and 47 exons, respectively, with very variable lengths between 27 and 192 bp. The complete triple helix in the α1(VIII) and α2(VIII) genes of mammals and the α1(X) gene of the chicken is encoded by a single large exon [16, 50, 68, 117, 133].

It is difficult to put together a detailed picture of the evolution of vertebrate collagens. Collagens of types I and II have been identified in agnathans, elasmobranchs and teleosts, and in some cases have been partially characterized; little is known about other fish collagens. The agnathans, which are at the root of vertebrate evolution, possess a cartilaginous skeleton, the matrix of which contains chondroitin-6-sulphate as the glycan component, as in the higher vertebrates. However, the main protein components in the lamprey are not collagen and elastin as in other vertebrates; in this case collagen is subsidiary to a newly discovered structural protein, **lamprin**, which makes up 44–51 % of the dry weight. This protein differs in its amino acid spectrum from both vertebrate collagen and elastin. It consists of over 80 % of the non-polar amino acids proline, glycine, alanine, valine and leucine, and contains 6.3 % tyrosine and 4.2 % histidine. Lamprin is cross-linked into fibres but the linking mechanism is not known; desmosin and isodesmosin are in

any case not detectable. Based on its chemical structure, agnathan cartilage cannot be considered a normal hyaline vertebrate cartilage, but represents a further unique type of connective tissue [138].

11.2.2 Invertebrate Collagens

Collagens are not found in plants or protozoans but are present in all metazoans from the Porifera onwards. The extracellular matrix of invertebrate animals, such as the annelids, molluscs and arthropods, also consists of collagens and anionic glycans. In addition to the interstitial collagens of mesenchymal origin, the invertebrates also possess collagens of epithelial or partly ectodermal origin in basement membranes, in the cuticle, e.g. of the nematodes and annelids, and in secretions such as the byssus threads of the edible mussel *Mytilus* and the silks of insect larvae. Only a few invertebrate collagens have been sequenced so far and consequently very little can be said about their evolutionary relationships. Comparisons are possible only at the level of the amino acid composition (Table 11.1). As invertebrate collagens also almost invariably show triple-helix structures, they contain about one-third glycine and high proportions of proline and alanine. On the other hand, the proportions of proline and alanine. On the other hand, the proportions of hydroxylated prolines and lysines vary widely [150]. Most invertebrate collagens form typical cross-striped fibrils like those of type-I and -II vertebrate collagens. However, the sponges already contain collagens which correspond to vertebrate basement-membrane collagens in some characters [Table 11.1 (1b)]: many proline residues are hydroxylated, the cysteine content is high, and a large portion of the hydroxylysine residues bear complex oligosaccharides [150]. The lysine-derived bridge structures LP and HP (Fig. 11.3) have been detected in marine cnidarians, annelids, echinoderms, molluscs and arthropods [89].

In view of their chitinous exoskeleton, it was assumed for a long time that **insects** have no collagen. In fact, collagen is present in significant amounts subepidermally and in the extracellular matrix of all organs. Insect collagen fibrils are quite similar to those of the vertebrates; they have clear cross-striping with a periodicity of 55–61 nm. The interstitial collagens of insects are composed of three identical subunits which are very similar to collagen I of vertebrates in size

Table 11.1. Amino acid composition of various collagens (residues per 100 amino acids) [37, 150]

	1a	1b	2a	2b	3a	3b	4	5	6
Ala	106	33	64	83	53	63	100	64	138
Arg	50	33	45	42	28	38	23	44	40
Asx	92	51	89	110	55	57	58	66	58
Cys	0	12	7	16	30	8	0	–	–
Glx	71	84	104	83	54	80	81	87	98
Gly	331	310	316	323	265	326	348	284	275
His	5	10	2	4	7	3	1	2	5
Ile	11	30	27	19	10	24	15	31	20
Leu	24	54	34	23	14	54	27	41	22
Lys	28	10	9	25	30	19	16	14	17
HyLys	6	44	13	25	0	40	0	54	18
Met	8	10	0	2	8	6	0	6	–
Phe	12	27	10	10	6	14	6	18	4
Pro	121	61	64	67	357	103	9	106	97
3-HyPro	–	–	24	13	} 20	}122	2	–	0
4-HyPro	93	141	79	79			160	81	0
Ser	46	37	43	19	19	16	88	40	110
Thr	20	23	44	24	16	13	49	29	35
Tyr	3	6	4	9	2	6	1	3	25
Trp	0	–	–	–	–	–	–	–	0
Val	22	29	40	24	12	13	16	30	38

HyLys, hydroxylysine; HyPro, hydroxyproline; –, not determined

1, Mammalian collagens; 1a, type I; 1b, type IV; 2, collagens from the sponge *Ircinia* sp.; 2a, spongin A; 2b, spongin B; 3, collagens of the nematode *Ascaris lumbricoides*; 3a, cuticula; 3b, muscle layer; 4, cuticula collagen of the earthworm *Lumbricus terrestris*, 5, mesenteric collagen of the mealworm *Tenebrio molitor*; 6, collagen silk of the sawfly *Nematus ribesii*

(280 kDa), length (290 nm) and amino acid composition [Table 11.1 (5)] [37, 150]. A collagen isolated from *Drosophila melanogaster* is homologous to the vertebrate basement-membrane collagens. This is a homotrimer of α chains of 1775 amino acids stabilized by disulphide bridges. The triple-helix domain includes 9 cysteine residues and has 21 interruptions, of which 11 have the same positions as in mammalian collagen IV. There is a C-terminal, globular domain of 230 amino acids which agrees in 59 % of its sequence with the α1(IV) and α2(IV) chains of the mouse. The corresponding gene has nine exons and therefore differs in general from the many small exons and large introns of mammalian collagen IV genes [9]. The larvae of the sawfly *Nematus ribesii* spin a cocoon composed of a **silk** which is very similar to vertebrate collagens in its amino acid composition, except that it contains no hydroxyproline [Table 11.1 (6)]. It is assumed that the molecule has a triple-helix structure but this has so far not been demonstrated [150]. The **prawns** *Penaeus chinensis* and *P. japonicus* contain collagens in muscle connective tissues with

the composition $(\alpha 1)_3$ and a high content of glycosylated hydroxylysine (19–23 residues per 1000) [81, 144].

The collagens of the **Porifera** are known as **spongins**. A distinction is drawn between fibrils with a diameter of 20 nm and clear cross-stripes with an interval of 63–65 nm (spongin A) and thinner fibrils with no distinct stripes (spongin B). The spongins contain one-third glycine and large amounts of hydroxyproline, hydroxylysine and cysteine [Table 11.1 (2a/b)]. Partial cDNA and gene sequences show that the sponges have both fibrillar collagens, with almost uninterrupted triple-helix sequences, and short-chain collagens, with shorter triple helices and terminal, non-helical domains. The exons encoding the triple helices almost exclusively have lengths of 54 bp or multiples thereof [31].

The mesogloea of the **Cnidaria** contains striped collagen fibrils with a diameter of 22–66 nm; these are made up from two or three different collagen chains and are rich in hydroxylysine-bound carbohydrate [82, 150]. The extracellular matrix of the **Platyhelminthes** also contains disulphide-linked, highly glycosylated 500-kDa collagens, i.e. about twice the size of interstitial vertebrate collagens and therefore possibly dimers [150]. The fibrils of the extracellular matrix of the **Mollusca** contain collagens with one or two chain types. The head cartilage of the cephalopod *Todarodes pacificus* is histologically similar to vertebrate cartilage, and contains a collagen which resembles vertebrate type II in its high content of glycosylated hydroxylysine, but otherwise has the amino acid composition of vertebrate collagen type I [66]. The byssus threads of the edible mussel *Mytilus edulis* are composed of a type-I-like collagen with two different subunits, but also contain a tyrosine-rich non-collagen protein. Collagens of type I, III and IV are immunologically detectable in the **Echinodermata**. The sea urchin *Paracentrotus lividus* possesses tissue-specific collagen fibrils with a much smaller periodicity than vertebrate collagens I–III: 55 nm in the foot, 47 nm in the epidermis, and 44 nm in the peristome. A partial cDNA sequence from this sea urchin encodes an uninterrupted triple helix of 480 amino acids and a C-terminal pro-peptide of 252 amino acids. The gene organization is apparently similar to that of the fibrillar vertebrate collagens [14, 24]. The collagens of the **Tunicata** and the **Acrania** differ from vertebrate interstitial collagens by having a high content of glycosylated hydroxylysines; they therefore resemble more closely the basement-membrane collagens [150].

The cuticular collagens of the **Annelida** *Lumbricus, Pheretima, Nereis* and *Neanthes* are rather unusual. The molecules are 950 nm long with a mass of 1700–1900 kDa and are thus several times larger than the classical vertebrate collagens. The cuticular collagen of *Nereis* is made up of two different subunits of about 470 kDa with 6000 amino acids, i.e. five to six times larger than the subunits of vertebrate collagens, and is one of the largest known polypeptides. The cuticular collagens are not striped, in contrast to the collagens of the internal organs of annelids, which in their molecular size and periodicity are very similar to the fibril-forming collagens of the vertebrates. Furthermore, annelid cuticular collagens are unique in that 4-hydroxyproline is found not only in position Y but also in position X of the sequence Gly-X-Y. Thus the 4-hydroxyproline content is particularly high at 14–16.5 %; 3-hydroxyproline is also present [Table 11.1 (4)]. The prolylhydroxylase of the earthworm is very similar to the vertebrate enzyme but, in contrast, is also able to hydroxylate the synthetic substrate $(Gly\text{-}Pro\text{-}Ala)_n$. The cuticular collagens contain 5–15 % carbohydrate, mainly galactose; however, this is bound not to hydroxylysine but to serine and threonine [150].

The collagen from the cuticle of the **nematode** *Ascaris lumbricoides*, with a length of 410 nm and a mass of 900 kDa, is also significantly larger than vertebrate fibrillar collagens. However, reduction with, for example, mercaptoethanol increases the solubility and releases three different chains of 52 kDa. The native collagen is thus cross-linked by disulphide bridges and also by bis- and isotertyrosine residues (Fig. 11.4); cross-linking structures based on allysine or hydroxyallysine are not detectable. The total proline content of cuticular collagens is very high (22 % in *Caenorhabditis*, 23 % in *Panagrellus* and 34 % in *Ascaris*) but the proportions of 4-hydroxyproline differ (1 in 2 prolines in the case of *Caenorhabditis,* 1 in 7 in *Panagrellus*, and 1 in 20 in *Ascaris*). 3-Hydroxyproline is not present and, as in the annelid cuticle, hydroxylysine is also absent [Table 11.1 (3a)]; the carbohydrate content is low. Lysylhydroxylase shows no activity with the cuticular collagen of *Ascaris*, although this contains 45 lysines per chain. The collagen fibrils of the connective tissues of internal organs are not striped but have the typical X-ray pattern; they are also much more similar to the vertebrate collagens in their amino acid composition than are the cuticular collagens [Table 11.1 (3b)]. The basement membrane of the gut epithelium con-

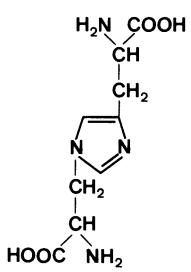
tains collagen molecules which, like the corresponding vertebrate collagens, are linked head-to-head by disulphide bridges [150].

The cuticle of the nematode *Caenorhabditis elegans* is a multi-layered extracellular structure consisting mainly of collagen. It is replaced after each of the post-embryonal ecdyses, changing each time in its ultrastructure and protein composition. Many collagen components of 30 to more than 210 kDa may be extracted from the cuticle following reduction. The pattern is specific for each ecdysis stage and includes at least 18 different chains in the adult cuticle; all have the typical amino acid composition of collagens but varying peptic digest patterns, i.e. they are the products of different genes. In vitro translation of collagen mRNAs in fact yields more than 30 different peptides. *Caenorhabditis elegans* possesses at least 50, perhaps even 150, collagen genes which encode relatively small polypeptides of 310–460 amino acids. The nine sequenced col genes for fibrillar collagens contain only one or two introns, i.e. they do not fit the 54-bp hypothesis. All encode 30-kDa polypeptides with five domains: an N-terminal non-helical domain of about 100 amino acids; two triple-helical domains of 30–33 and 127–132 amino acids, separated by a non-helical spacer; and a C-terminal non-helical domain. The longer triple-helix sequences include one to three short interruptions. This domain structure is somewhat similar to that of vertebrate collagen IX [23]. The neighbouring genes col-12 and col-13 differ in only 5 out of 951 bp and encode the same amino acid sequence. Col-13 mRNA, however, bears a trans-spliced leader, and col-12 a cis-spliced leader. As illustrated in Chapter 2, about 10% of all *Caenorhabditis elegans* mRNAs are trans-spliced [95]. In addition to the genes for fibrillar collagens, two genes (clb-1 and clb-2) for basement-membrane collagens have been sequenced. As in the mammals, the nematode basement-membrane collagen is a heterotrimer. The encoded polypeptides have distinct sequence similarity in the C-terminal non-helical domain to the NC1 domains of mammalian α(IV) chains, e.g. clb-1/α2(IV) 72% and clb-2/α1(IV) 63%. The gene duplication α1(IV)/α2(IV) thus apparently occurred early in the evolution of the Metazoa. The existence of only a single homotrimeric basement-membrane collagen in *Drosophila* stems probably from the loss of one of the two genes [46].

11.3 Matrix Proteins of Calcifying Tissues

The organic matrix plays an important, if not always obvious, role in the growth of crystals in mineralizing tissues. It is apparently organized quite similarly in vertebrates and invertebrates, with the main components being collagens and specific acidic phosphoproteins and polysaccharides; phospholipids may also be involved in calcification [57]. In addition to the acidic proteins, there may also be a network of relatively non-polar proteins that has mainly a mechanical function. The bones of all vertebrates, from the teleosts to the mammals, have a complex spectrum of **phosphoproteins** with molecular masses between 4–5 and over 100 kDa; these contain not only phosphoserine and phosphothreonine but also glutamic and aspartic acid. The largest also contain hydroxyproline and hydroxylysine and may therefore be associated with collagens [44]. The phosphoproteins are linked via histidinoalanine (Fig. 11.5), a cross-linking structure found in vertebrate bone and tooth and also in cartilage, tendon and other connective tissues [78]. The great heterogeneity of phosphoproteins in bone is probably partly related to proteolytic artefacts. In addition to phosphoproteins, vertebrate bone always contains a further type of protein, the small **osteocalcin**, which is composed of 46–50 amino acids, depending upon the species. The sequence of the protein from the teleosts to the mammals always includes glutamate residues at positions 17, 21 and 24; these are post-translationally γ-carboxylated by a vitamin K-dependent process and can then bind calcium. The similarly highly conserved central region with a disulphide bridge between 23-Cys and 29-Cys is responsible for the tight binding of osteocalcin to the hydroxyapatite of bone [75].

Calcium-binding proteins with phosphothreonine and phosphoserine are also found in calcifying cartilage, dentin, the calcifying tendons of birds, and the scales of the cartilaginous fish



N^{τ}-Histidinoalanine

Fig. 11.5. The phosphoproteins of calcifying tissues are linked via N^{τ}-histidinoalanine [78]

[111]. In contrast, the organic matrix of tooth enamel contains only phosphoserine and small amounts of acidic amino acids. Enamel, the especially highly mineralized and hard coating of vertebrate teeth, is synthesized and secreted by specialized cells (ameloblasts). The organic matrix consists of two families of enamel-specific proteins, the amelogenins (22–28 kDa) and enamelins (40–72 kDa). In reality, the enamelins appear to be tightly bound blood-plasma proteins with serum albumin as the main component [122]. The **amelogenins** of different mammals are practically identical in the N-terminal half of their 180-amino-acid chain but are very variable in the C-terminal region. Proteins that are immunologically similar to mouse amelogenins have been detected in the agnathan *Eptatretus stoutii* [148]. In the mammals, there is only one amelogenin gene per haploid genome, localized in either the X or the Y chromosome. Several different mRNAs are produced from the former by alternative splicing [40]. Enamel contains the largest hydroxyapatite crystals found in the mammalian body; at the end of the mineralization process, the matrix proteins are almost completely reabsorbed. Human and mammalian saliva contains various proteins which, through calcium binding, result in the oversaturation of the saliva with Ca^{2+}, the prevention of calcium phosphate precipitation and the stabilization of the hydroxyapatite of the enamel. Amongst them is the tyrosine-rich **statherin** [110] and numerous **acidic, proline-rich phosphoglycoproteins** (PRPs) which are encoded by a multi-gene family. The sequences of the PRPs show considerable periodicity and are significantly homologous throughout the primates and rodents, although they may differ in length: 100–120 amino acids in humans and *Macaca fascicularis* and 200–240 amino acids in the mouse and rat [5, 61].

Two classes of macromolecular components may be distinguished in the **organic matrix of mollusc shells**. About 70–87 % of the soluble proteins are aspartate-rich proteins with a β-folded structure and bind Ca^{2+} on the charged side-chains. The remainder are serine-rich proteins which are associated with large quantities of polysaccharide and can also bind calcium [137]. The calcium metabolism of several mussels *(Rangia, Macrocallista)* has a noteworthy feature. The extrapallial (mantle) fluid has calcium-containing particles, about 40 nm in diameter, from which several phosphoproteins of more than 140 kDa can be isolated. These proteins are rich in aspartate, phosphoserine, phosphothreonine and histidine, and contain up to 10 % of the cross-linking structure histidinoalanine (Fig. 11.5) [78]. Acidic matrix proteins have also been found in **other invertebrates**, e.g. the foraminifers and echinoderms, but have been little investigated. Surprisingly, antibodies against proteins of the rat incisor cross-react with the tooth matrix of the sea urchin *Lytechinus variegatus;* presumably, common evolutionally conserved or convergent molecular structures are present [21, 132].

11.4 Elastic Proteins

All the movements of the animal body rely upon some sort of bendable or stretchable structure. It is, therefore, to be expected that all vertebrates and invertebrates have substances with **elastic properties**, i.e. with the capacity to release potential energy conferred during passive stretching with minimal losses on return to the starting position. Rubber-like materials consist of kinetically free, randomly folded chains; their elasticity mainly involves changes in conformation entropy and not, for example like steel, changes in internal energy resulting from the distortion of atomic bonds. The chains must be cross-linked to each other to avoid viscous flowing movements during deformation. The proteins of animal tissues are able to meet such physical specifications. So far, only four **types of elastic protein** are known from the whole of the animal kingdom: vertebrate elastin, insect resilin, abductin from the shell ligaments of the mussels, and the elastic proteins from the arteries of the cephalopod *Octopus dofleini*. All these proteins characteristically have small amino acids; one imagines that large, unwiedly components would only hinder free movement of the polypeptide chains, and thus limit the deformability of the protein (Table 11.2). However, this alone does not explain the elastic properties as there are sequences of small amino acids in very stable, non-expansive chain conformations, e.g. those of the collagens and, in particular, the silk fibroins. Apart from this, the amino acid composition of the four protein types mentioned varies so widely that they cannot really be considered as homologous. Thus, their similar elastic properties appear to be the result of convergent evolution [113].

The best-known elastic protein is **elastin**, which is found in all vertebrates, except the agnathans, but not in the invertebrates. Elastins together with collagens are the most important structural

Table 11.2. Amino acid composition of various elastic proteins (residues per 100 amino acids) [101, 113]

	1	2a	2b	3	4
Ala	211	50	42	72	112
Arg	11	4	22	46	36
Asx	4	23	39	91	102
Cys	4	1	13	7	0
Glx	18	14	37	121	43
Gly	310	627	411	85	411
His	0	2	4	21	8
Ile	26	6	31	60	13
Leu	68	3	72	74	22
Lys	52	8	34	68	4
Met	0	92	97	21	0
Phe	23	83	22	42	25
Pro	121	10	47	55	75
Ser	10	61	89	72	77
Thr	10	10	14	64	29
Tyr	10	1	4	36	21
Val	123	5	22	62	23

1, Bovine elastin a; 2, shell ligaments of mussels: 2a, *Aequipecten irradians*; 2b, *Mytilus edulis*; 3, elastomers from the aorta of *Octopus dofleini*; 4, resilin from the wing joints of the locust *Schistocerca gregaria*

proteins of the vertebrates and are found associated in many tissues, e.g. in the skin, lungs, and elastic ligaments. The walls of the large arteries in vertebrates are particularly elastin-rich; this helps maintain the blood pressure of the heart ventricle during the filling phase. The elastic fibres of connective tissues contain cross-linked, highly polymeric aggregates of elastin; these are formed from soluble **tropoelastin** (68–72 kDa) that is released into the extracellular space by fibroblasts and other cell types. The amino acid composition of elastin presents the same picture from the cartilaginous fish to the mammals [Table 11.2 (1)]. The main constituent is glycine (31–41 %), with the non-polar amino acids alanine, valine and proline making up a further 35–41 %. The lysine content varies between 4.0 (swan aorta) and 8.8 % (*Latimeria* aorta); the hydroxyproline content is ten fold lower than in collagen, and hydroxylysine is absent. Elastin is an extremely non-polar protein, and the hydrophobicity and elastic properties have increased markedly during evolution of the vertebrates. As in all rubbery materials, the chains of elastin are cross-linked to each other; this involves the same oxidative linking of lysine side-chains by lysyloxidase as is found in collagen. However, in the case of elastin, the cross-linking structures are the pyridine derivatives desmosin and isodesmosin (Fig. 11.3 d). The bovine ligamentum nuchae was recently found to contain a further structure, allodesmosin, which carries a branched side-chain based upon two lysine molecules at the C-4 position [123]. Of the 37 lysine residues in tropoelastin, only 5 have free side-chains in mature elastin.

Because of the intensive cross-linking, direct sequencing of mature elastin is impossible. Although the sequences of almost all the individual tryptic cleavage peptides of porcine tropoelastin have been described, their order was never very certain. The **sequences of elastin** from man, various other mammals and the chicken have since been determined via the cDNA [7]. There are three bovine tropoelastins, a, b and c, which have lengths of 747, 733 and 713 amino acids, respectively, and differ only in the presence or absence of insertion sequences of 20 or 20 + 14 amino acids, introduced by alternative splicing. Multiple human, sheep and chicken tropoelastins arise by a similar process [100]. The elastin sequence contains two types of alternating domain. One type includes many lysine residues arranged in pairs with two or three intervening small amino acids (Fig. 11.6); this domain type gives rise to the cross-linking structures but has no elastic properties. The latter are localized in the second type of domain, which consists of randomly intertwined polypeptide chains of mainly non-polar amino acids [56, 101]. The chicken elastin sequence, despite its similar length, cannot be fully aligned with the bovine sequence. The comparable 516 amino acids show 69 %, and in some regions over 90 %, agreement. It would appear that the evolution of these elastins has involved several duplications and the deletion of partial sequences [102]. The cross-linking structures desmosin and isodesmosin (Fig. 11.3 d), which are typical of elastin, are also found in the membranous **egg-shells of birds and reptiles**, the proteins of which have an amino acid composition completely different from that of elastin. These proteins are richer in proline in the reptiles com-

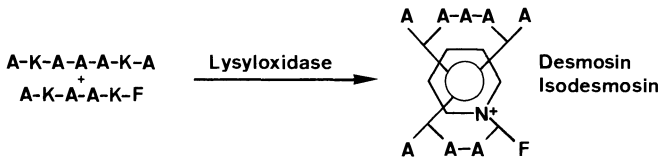


Fig. 11.6. The linking of two elastin chains and the formation of desmosin [101]

pared with in the birds (*Iguana iguana* 44 %, *Anolis limifrons* 34 %, chicken 12 %), but reptiles lack the hydroxyproline and -lysine found in chicken egg-shell proteins [145].

Insect **resilin** is a protein with almost perfect elastic properties; 96–97 % of the energy stored on stretching is released on shortening. It is found in various insect organs, e.g. in wing joints and tendons, in feather-like elastic structures in fleas, locusts and beetles, in the proboscis of butterflies and in the sucking apparatus of some bugs; it is also present in millipedes and crustaceans. Resilin has a very simple amino acid composition, with more than one-half of positions occupied by either glycine or alanine [Table 11.2 (4)]. The chains are linked by bis- and tertyrosine residues (Fig. 11.4), which are formed from tyrosines by the activity of phenoloxidases or peroxidases. Resilin contains about 3 % free tyrosine, 0.8 % bistyrosine and 0.4 % tertyrosine [29].

The **ligament of the mussel shell** is formed from mantle tissue, just like the solid part of the shell, and was originally part of the complete shell structure. It is covered externally by the periostracum and is divided into two layers. The outer ligament is made of non-calcified but quinone-treated proteins, about which little is known. The inner ligament of most mussels contains calcium carbonate to about 60–80 % of the dry weight; only that of the scallop (Pectinidae) is low in calcium. In this latter type of swimming mussel, the outer ligament is reduced to a thin band and the inner ligament has the form of a wedge-shaped cushion which functions as antagonist to the single adductor muscle present. The high elasticity of the inner ligament allows up to three flapping movements of the shell per second during recoil swimming. The inner mussel ligament consists of fibrillar elastic proteins, the **abductins** [Table 11.2 (2a/b)]. The polypeptide chains of the abductins, like those of all elastic proteins, are cross-linked but in this case the linking structures are not known. The ligament proteins of the mussels in the superfamily Mactracea contain 20–25 % methionine sulphoxide in addition to 45–50 % glycine; this can be demonstrated spectrometrically in intact ligaments. The two diastereomers of this unusual amino acid appear to be present in equal proportions. The Pectinidae, in contrast, have a much lower methionine fraction, of which only 40–50 % is oxidized [64].

The cephalopods have a vascular system with a relatively high blood pressure. Their aortas and large arteries, like those of the vertebrates, help maintain the blood pressure and have a layer of elastic fibrils in their walls. The elastic protein was first isolated from the aorta of *Octopus dofleini* and, consequently, received the name **octopus arterial elastomer** (OAE) [113]. The OAE fibrils have considerable elastic properties, although their amino acid composition is not as extreme as that of the other elastic proteins [Table 11.2 (3)].

11.5 Albumen Proteins of Birds and Reptiles

The egg-white (albumen) of the chicken and other avian species contains at least 50 different, electrophoretically distinct protein components, only a few of which, however, have been investigated (Table 11.3). It has been shown for various **albumen proteins**, e.g. ovalbumin, conalbumin (ovotransferrin), lysozyme, ovomucoid and avidin, that they are synthesized in the tubular glands of the oviduct, under the control of steroid hormones. Albumen, which surrounds the yolk as a gelatinous protective layer, owes its special mechanical properties to the cross-linking of ovomucin and lysozyme. Lysozyme has carbohydrate-hydrolyzing activity, and will therefore be discussed together with the other carbohydrases in Chapter 13. The content of this enzyme in egg-white varies with the species, from 3–4 % of total protein in the chicken and duck to only trace amounts in the penguins. **Ovomucins** are sialoglycoproteins which can interact electrostatically with the positively charged lysine side-chains of lysozymes via their negatively charged sialic acids. The chicken has two different ovomucins: an α type of 210 kDa with only 1 % sialic acid, and a β type of 720 kDa and more than 10 % sialic acid. During brooding, the carbohydrate chains of ovomucin are partially cleaved by the

Table 11.3. The proportions (%) of different proteins in the egg-white of the chicken and the penguin *Pygoscelis adeliae* [33]

	Chicken	Penguin
Ovalbumin	55	30
Penalbumin	<0.01	25
Ovomucoid	11	10
Ovotransferrin	12	4.5
Lysozyme	3.5	0.05
Ovomucin	1.5	1.0
Ovomacroglobulin	0.5	1.0
Flavoprotein	0.8	0.3
Avidin	0.05	0.02

lysozyme, and the egg-white gradually liquefies. Like lysozyme, other egg-white components, which belong functionally or structurally to other protein families, are discussed in various sections of this book. For example, conalbumin (ovotransferrin) of the chicken egg differs from serum transferrin in the blood plasma only in its carbohydrate fraction, and not in the amino acid sequence, and it has already been introduced in Chapter 5. The protease inhibitors of egg-white, ovomucoid, ovoinhibitor and ovostatin were mentioned in Chapter 3. Therefore, what follows is mainly concerned with the ovalbumin family and with avidin.

Ovalbumin is the predominant protein of the egg-white. It is easy to isolate and is, therefore, well investigated, but its biological role is not clear. Chicken ovalbumin is a polypeptide of 383 amino acids with a mass of 42.7 kDa. The polypeptide chain is post-translationally modified at four sites: the N-terminal glycine is acetylated, the residues Ser-68 and Ser-344 are phosphorylated, and Asn-292 is N-glycosylated. More than ten different oligosaccharides have been identified in the ovalbumin of the chicken, and other avian species also show marked heterogeneity in the carbohydrate fraction. The **ovalbumin gene** was the first gene in which interruptions of the coding sequence by introns were described in 1978. The promoter activity is normally suppressed and this inhibition is relieved by steroids [39]. Two further homologous genes, the **X and the Y gene**, occupy the same 40-kb DNA segment on which the ovalbumin gene is located; they show the same divisions into eight exons and are also expressed in the oviduct in an oestrogen-dependent manner. However, their protein products have not yet been identified. The sequence of 388 amino acids encoded by the Y gene is 58 % similar to that of ovalbumin, and the sites of post-translational modification, with the exception of the substitution 69-Ser to Thr, are also the same [51]. In the egg-white of the penguins, e.g. *Pygoscelis adeliae*, about half the ovalbumin is replaced by a protein, **penalbumin**, which is found in trace amounts in other avian eggs (Table 11.3). This is a glycoprotein of 61 kDa which, with a carbohydrate fraction of 15 %, is twice as highly glycosylated as ovalbumin; the sequence is not yet known [33].

Avidin is a basic glycoprotein (pI = 10) with 128 amino acids and 10 % carbohydrate, and it binds biotin with an extremely high affinity. Avidin molecules are homotetramers of 66-kDa subunits, each of which carries a biotin-binding site. This protein is synthesized in the oviducts of amphibians, reptiles and birds, but not in mammals. Its concentration in the egg-white varies more than 100-fold between different avian species, with especially large amounts occurring in the egg of the turkey *Meleagris gallopava*. Avidin synthesis in the oviduct cells is regulated by steroid hormones, and in some circumstances other cells of both males and females can synthesize the protein, e.g. following wounding or during inflammation, bacterial infection or retroviral transformation; avidin apparently has an overall protective function [42]. In addition to avidin, other vitamin-binding proteins are found in chicken egg-white; for example, there is a cobalamin-binding protein and a riboflavin-binding protein which is also present in the yolk [147, 149]. The yolk also contains a biotin-binding protein which has no immunological relationship to avidin [86]. The albumen of the eggs of the crocodile *Crocodylus porosus* contains significantly less dry matter than that of chicken eggs (3.6 % compared with 12 %). About 80 % of the total protein is comprised of a glycoprotein which resembles ovomucin of avian eggs in its amino acid spectrum and other properties. There is also a large amount of a further protein with the characteristics of ovomacroglobulin. Proteins that are comparable with ovalbumin and ovotransferrin are not found in the crocodile; instead, a novel 59-kDa protein with the name **crocalbumin** has been demonstrated [15].

11.6 Milk Proteins

11.6.1 Caseins

Caseins are the major proteins of mammalian milk. They provide the young animal not only with amino acids but also with calcium phosphate for skeleton growth; these are bound in colloidal particles referred to as **micelles**. Bovine micelles are formed by calcium-dependent aggregation of molecules of type α_{s1}, α_{s2} and β caseins, and are stabilized and protected from premature precipitation by casein κ. Interactions between non-polar regions of the polypeptide chains are responsible for the aggregation of casein molecules. Negative charges concentrated in particular parts of the sequence bring about interaction between Ca^{2+} and casein κ. Caseins α_s and β can be phosphorylated on serines which lie mainly in

the N-terminal region [6]. All the phosphorylation sites have similar sequences, with several serines followed by two glutamic acid residues. Compared with the α caseins, the β caseins have a lower phosphate content and are precipitated at much higher Ca^{2+} concentrations; the κ caseins are calcium insensitive. The caseins are clotted by **chymosin** in young animals or by pepsin in the adult; κ casein is cleaved into an N-terminal fragment, para-κ-casein, and the casein macropeptide. Many caseins are glycosylated, and κ casein is especially carbohydrate rich. The α casein of the rat is one of the few glycoproteins to have glucose as the sugar component.

As for many other proteins, extensive sequence comparisons amongst the caseins became possible only with the availability of DNA sequence analysis, e.g. of the cDNAs or genes of human, bovine, sheep and murine caseins as well as of those of the rat, guinea-pig and rabbit [2, 12, 27, 48, 52, 59, 76, 143]. The milk from most of the mammal species investigated contains the known bovine casein types, although they may have received other names before the sequence data became available. The caseins may also show differences from the bovine spectrum. For example, human and other primate caseins lack type α_s. In the rat there is an α casein that is homologous to bovine α_{s1} but which includes a unique insert of 60 amino acids that is composed of 10 amino acid repeats. The rat also possesses a so-called γ casein, which apparently belongs to the α-casein group but has only limited sequence similarity to bovine α_{s1} [59]. Mouse milk contains no fewer than eight different calcium-sensitive caseins, three of which are known as α_s, two as β, and one each as γ, δ and ε; the ε casein is similar to bovine α_{s1} [52, 143].

The genes of the four bovine casein types are arranged in the genome in a 200-kb-long cluster [128]. Of particular interest is the synchronous regulation of their expression by the concerted action of steroid and peptide hormones; this occurs at the level of both transcription and translation. In fact, all casein genes of the rat and bovine α_{s1} have a very similar organization in the 5′ region [59, 146]. Sequence comparison indicates that all calcium-sensitive caseins have a common origin: their DNA sequences show clear homology, especially in the 5′ non-translated (NT) region and in the region of the signal peptide (Fig. 11.7). In contrast, sequence similarity in the coding region and in the amino acid sequence is limited to the areas which include the phosphorylation sites. For example, the signal sequences of sheep pre-α_{s2} casein and guinea-pig pre-A

```
                                    -1 +1
Sheep-αs1            MKLLILTCLVAVALA R...
Sheep-αs2            MKVLMKACLVAVALA K...
Sheep-β              MKVLILACLVAVALA R...
Rat-β                MKVFILACLVAVALA R...
Mouse-ε              MKFIILTCLLAVALA K...
Sheep-κ      MRKS[IL/FF]LVVTILALTLPFLIA Q...
```

Fig. 11.7. The signal sequences of various caseins [52]

casein agree by 93 %, but the mature proteins agree by only 33 %. The α_{s1} caseins of the rabbit and mouse have 34 % sequence similarity; human β casein agrees 45–62 % with the bovine, sheep, rat and mouse sequences.

The gene for bovine κ casein is tightly linked to the α_{s1}-, α_{s2}- and β-casein genes but it is not closely related, as may be seen from the very different sequence of the 5′-NT region and the signal peptide (Fig. 11.7). There are also very large species-specific differences in the mature proteins; the human, bovine, sheep, goat and rat κ caseins have only 22 out of 158–169 amino acids in common. Many gaps are required for optimum alignment of the casein sequences, e.g. 22 in the case of a comparison between the rat α casein and bovine α_{s1} casein. This presents problems for estimations of genetic distances. However, it is clear that the rate of evolution of the caseins is quite high. Comparison of the human, rat and mouse signal peptides suggests that the various calcium-sensitive caseins arose by gene duplication at the time of the primitive mammals about 300 million years ago, long before the evolutionary spread of the mammals 75 million years ago. According to this interpretation, all living mammals should have the same casein genes but, in fact, humans lack a homologue to the bovine α_s casein and there is no bovine homologue to the γ casein of the rat. Casein-homologous sequences in non-mammals have not so far been described.

11.6.2 Whey Proteins

In addition to the caseins, milk typically contains low concentrations of other proteins, which are known as whey proteins. These include, in particular, α-lactalbumin, β-lactoglobulin, milk transferrin (lactotransferrin) and lysozyme, but also include serum albumin and serum transferrin. α-Lactalbumin is the regulatory chain of lactose synthase and will be discussed further in relation to this enzyme (p. 470); it is found in the milk

glands of all lactating mammals which produce lactose or lactose-derived oligosaccharides (Marsupialia), and is also released into the blood. It is not found in the milk glands and milk of aquatic mammals which do not produce lactose. The distribution and function of the **β-lactoglobulins** remains a mystery. They are found only in the milk of ruminants, the horse, ass, pig, dog, dolphin *(Tursiops truncatus)*, manatee *(Trichechus manatus)* and kangaroo, and their occurrence in human milk is controversial [97]. In the horse, ass, pig, dog, dolphin and manatee they are monomeric proteins, but they are dimeric in the ruminants and the kangaroo. The horse and the dolphin have two β-lactoglobulins with very different sequences, and the cat actually has three lactoglobulin loci (I–III) [49, 97]. The β-lactoglobulins of the ruminants agree with each other by more than 90 % but with those of other mammals by only 30–55 % [49]. The polypeptide chains of 155–166 amino acids have a spatial structure of two layers of antiparallel folded sheets, similar to that of the retinol-binding proteins of vertebrate blood, the bilin-binding protein of insect haemolymph, and other transport proteins for hydrophobic molecules in the protein super-family of the lipocalins. Pairwise sequence comparisons within the family show agreements of 25–30 % [41].

The **acidic whey proteins** (WAP) are unusual milk proteins in that they are found only in the milk of the rat, mouse and rabbit, and occur at concentrations many times higher than that of the α-lactalbumins. In the milk glands of the lactating rat, WAP mRNA makes up no less than 15 % of the total poly(A)-RNA; this protein is not found at all in the ruminants. The WAP genes of the mouse and rat are highly homologous; they coincide over a sequence which begins 325 bp upstream of the cap-binding site and extends 20 bp into the 3′-flanking region. The conservative evolution of this 5′ region probably involves selection for a similar hormone-regulating mechanism. However, the rat, mouse and rabbit amino acid sequences (127–137 amino acids) show at most 64 % agreement [26].

11.7 Proteins of the Arthropod Cuticula

The **arthropod cuticula** is an extracellular structure secreted by the underlying epidermis. It is made up in a complex manner from several layers that differ in both fine structure and composition. Exterior to the procuticula (endo- and exocuticula) of at least 100 μm lies the multilayered epicuticula of only 1–2 μm. The epicuticula is chitin-free but its molecular structure is otherwise largely unknown. The procuticula consists of α-chitin fibres in a protein matrix. Depending on its exact function, the procuticula may be hard, flexible or elastic. As will be described in more detail in Chapter 12, the hardening (sclerotization) of the cuticula, which is usually accompanied by a brown coloration, involves the cross-linking of protein molecules by phenolic substances. As the proteins are then no longer extractable, examination of the cuticula proteins of the insects has been restricted almost entirely to the non-sclerotized cuticula, i.e. that of larvae, pupae, and hatching or freshly hatched adults.

The insect cuticula typically consists of **many different proteins**, the number of which varies greatly according to the separation and detection methods used. The differences in complexity of the cuticula protein spectrum between the various insect orders will only be appreciated fully after a large number of insect species have been compared using the same method. Two-dimensional electrophoresis of extractable cuticula proteins from the locusts *Locusta migratoria* and *Schistocerca gregaria* gives more than 100 spots, most of which occupy the same position in both species [3]. Translation of the mRNAs from wing epithelium of the moth *Manduca sexta* gave 160 different polypeptides in the pupa and 120 in the adult; only some of these had similar values for molecular size and pI in both developmental stages [120]. Simpler electrophoretic or chromatographic methods reveal only about 10–20 different extractable cuticula proteins [3, 118, 121, 135]. The protein spectrum is unusually simple in the fly *Ceratitis capitata*, where more than 80 % of the total protein from pupal cuticula consists of one glycoprotein of about 100 kDa [10]. The variety of cuticula proteins varies not only with the species but also with the developmental stage within a species [65, 88, 118, 121, 135]. In many species, even different regions of the cuticula of one individual show differences in the protein spectrum.

Despite all these differences, **cuticula proteins have quite a few common features**. They are usually very small, with masses of 8–37 kDa. About 40 % of the total cuticula protein of *Drosophila* consists of one component of 90 kDa, which is covalently bound to chitin and has prophenoloxidase and peroxidase activity [53]. The predominant proteins of many species are acidic with pI

values of 3–6 [118]. The **amino acid composition** of the cuticula proteins is always rather biased; this is seen, for example, in the particularly well-investigated protein number 38 of *Locusta*, in which only 14 of the 20 common protein amino acids are found. The proteins are consistently rich in alanine and/or glycine, but are low in charged amino acids and often entirely lacking in methionine and cysteine [3, 65]. Similarities in the amino acid composition and distribution are found in both the sclerotinized proteins (oothecins) of cockroach egg capsules (ootheca) and the egg-membrane proteins (chorion proteins) of flies and moths. This does not necessarily indicate phylogenetic relationships between the proteins but, more likely, suggests similar functional adaptations. With the exception of sclerotization, the cuticula proteins are subject to little post-translational alteration [22].

Only a few cuticula proteins have been sequenced so far, e.g. those of the locust *Locusta migratoria*, the lepidopterans *Bombyx mori, Manduca sexta* and *Hyalophora cecropia*, and the dipterans *Drosophila melanogaster* and *Sarcophaga bullata* [53, 54, 88, 104, 136]. The data show quite clearly that the cuticula proteins do not belong to any of the known protein superfamilies. The partial sequences from the two fly species agree at 50–60 % of positions, but have no significant homology to the locust proteins. A cuticula-gene cluster found in the region 44D of *Drosophila melanogaster* chromosome 2 includes four active genes (I–IV) and one pseudogene. The active genes are organized in pairs with the genes of each pair being read in opposite directions. The encoded amino acid sequences agree by 95 % (I/II) and 33 % (I/III). There is significant sequence agreement between the cuticula proteins of the lepidoptera *Manduca* and *Hyalophora* and those of the Diptera *Drosophila* and *Sarcophaga* [104, 136]. A pupal cuticula protein from the silk moth *Bombyx mori*, sequenced via the cDNA, contains nine repeats of the sequence $APAX_1X_2X_3WN$ in its total sequence of 253 amino acids [88]. Little is known about the **cuticular proteins of other arthropod groups**. Numerous acidic proteins of 10–70 kDa have been extracted from the cuticula of the crustacean *Scylla serrata;* the cuticula of the crayfish *Astacus leptodactylus* contains mainly larger proteins which do not differ significantly between soft and calcareous cuticulae. Similarly, the cuticula proteins from different body regions of the shrimp *Pandalus borealis* appear the same on 2-D electropherograms [4].

11.8 Chorion Proteins

The insect egg is surrounded by an integument which must fulfil several functional requirements. It must allow entry of spermatozoa, and for this there are one or more narrow canals (micropyles). It must be permeable to respiratory gases, and an example here is the "aeropyles" of silkworm eggs; these are arranged either in bands *(Antheraea polyphemus)* or over the entire egg surface *(A. pernyi)*, and in some species they are surrounded by a small crown. The egg membrane must be flexible enough for the egg-laying process to occur but must be sufficiently protective, especially during the diapause of the winter months. This membrane is the secreted product of several thousand cells (follicle epithelium) which surround the oocyte together with its nutritional cells. Structurally, an inner thin vitellin membrane, the composition of which is not entirely known, can be distinguished from the outer chorion, which consists almost entirely of proteins. The **chorion** has a predominantly lamellar structure, the lamellae being two-dimensional networks of protein fibrils. Formation of the 60-μm-thick chorion of *Antheraea polyphemus* requires about 2 days, and the 0.8-μm-thick chorion of *Drosophila melanogaster* is synthesized in only 5–6 h; the chorion of *Bombyx mori* is intermediate between these two extremes and has a thickness of 20–25 μm. During choriogenesis, the follicle epithelial cells are given over entirely to the production of chorion proteins. These proteins, together with the corresponding genes and their regulation, were studied first in the tussah moth *Antheraea polyphemus* and later in *Bombyx mori* and *Drosophila melanogaster.* The chorion proteins of the silk moth have a characteristic composition of 30–37 % glycine, high proportions of alanine, valine, leucine and tyrosine, and 6–12 % cysteine. The corresponding *Drosophila* proteins have only 16 % glycine and 0.6 % cysteine and, instead, are high in charged, especially basic amino acids. Thus, the chorion proteins of the flies are at best only distantly related to those of the moths [105].

SDS electrophoresis separates the **chorion proteins** of *Antheraea pernyi* into four classes with different molecular masses: 38 % A proteins (9–12 kDa), 50 % B (12–14 kDa), 9 % C (16–20 kDa) and trace amounts of the much larger D proteins. A fifth type, the E protein, is unusual in that it is not involved in the cross-linking of A–D and can therefore still be

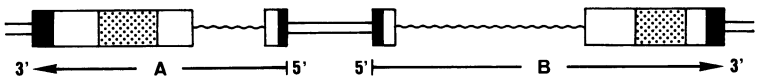


Fig. 11.8. A gene pair for the chorion proteins A and B from the tussah silk moth *Antheraea polyphemus* [58]. The *arrows* show the direction of transcription. Sequences included in the mature mRNAs are shown as *large rectangles:* the *black areas* are the 5'- and 3'-NT regions; the translated regions are shown as *white* (the terminal arms) and *dotted* (the central domains) areas. Introns are shown as *wavy lines* and the spacers as *small rectangles*

extracted from the mature chorion. This class includes the two proteins E1 (15 kDa) and E2 (85 kDa). *Bombyx mori* is apparently alone in also containing the high-cysteine (Hc) proteins which consist of about 30 % cysteine and constitute as much as 23 % of the total chorion proteins. **The polypeptide chains of proteins A–C** are subdivided into three domains (Fig. 11.8): a conserved central domain, which forms the core of the chorion fibres, and the very variable terminal domains (arms). The A proteins fall into two subfamilies with 1.0–7.6 % sequence difference within the families and 40–50 % difference between them. Thus, it appears that gene duplications occurred at two distinct times during evolution. The B family is subdivided in a similar fashion. The A and B proteins are only very distantly related. In the early expressed C proteins, both A-like (CA) and B-like (CB) sequences can be distinguished. The Hc proteins of *Bombyx mori* likewise belong to either the A or the B subfamily. Therefore, with the exception of the E proteins, all chorion proteins of the silk moths belong to either the α branch (A, CA, HcA) or the β branch (B, CB, HcB) of the super-family [74]. The E proteins form a sponge-like network ("filler") in the crowns of the aeropyles. E1 consists of 150 amino acids; the almost fourfold longer E2 protein (542 amino acids) is explained by the repetitive structure of positions 128–541. E2 is 62 % similar to E1 in some regions, but only 20 % similar in others. Neither protein is recognizably related to any other chorion protein [105].

The chorion proteins are extremely heterogeneous. Several years ago, 186 different proteins in the chorion of *A. polyphemus* had already been detected by 2-D electrophoresis: 26 A, 42 B, 69 D and 8 E proteins. More sophisticated methods would probably increase the number detected. Part of this heterogeneity undoubtedly results from protein polymorphism within the investigated populations; however, even in a genetically uniform inbred line of *Bombyx mori*, 71 protein fractions could be separated by isoelectric focusing, and several more could be distinguished in a second fractionation step. The number of chorion proteins in *Drosophila melanogaster* is much lower than in the silk moth; in this case, there are 16–19 main and 30–35 subsidiary components with molecular masses between 15 and 150 kDa [106]. The chorion proteins are subject to some post-translational modification. Apart from cleavage of the N-terminal signal peptide, cyclization of the N-terminal glutamine to pyrrolidoncarboxylate occurs in chorion proteins B and C of *A. polyphemus*. In contrast, glycosylation is not significant. The chorion proteins can be extracted up to the end of choriogenesis, after which they become insoluble within hours as the result of cross-linking with each other by disulphide bridges; A binds to B, B also binds to itself, and E1 and E2 bind to each other and to themselves [106]. The *Drosophila* chorion contains little cysteine, and here bis- and tertyrosines (Fig. 11.4) function as the cross-linking structures.

Expression of the chorion genes follows a precise time-space programme. The main proteins synthesized initially are C followed by A and B and, in *Bombyx mori*, by the Hc proteins. The proteins of the aeropyle crowns, including the E proteins and certain A–D proteins, are synthesized only later. The differences in chorion structure found in different regions of the egg may be the result of variations in gene regulation in individual follicle epithelial cells. The genes of the chorion proteins A and B are arranged in a tandem cluster in the genome of *A. polyphemus;* the 7-kb repeats each consist of one A and one B gene read in opposite directions (Fig. 11.8). Each gene contains only one intron, which divides the region of the signal peptide. The common 5'-NT region contains regulatory elements. *Bombyx mori* possesses about 200 chorion genes, arranged in two clusters, which are expressed at different times. The HcA/HcB genes also form 15 opposing pairs with common 5'-NT regions of about 270 bp [36, 119]. In *Drosophila melanogaster,* there is one cluster of three chorion genes with the same orientation on the X chromosome, and a further cluster of four genes on chromosome 3. The 5'-NT sequences of these genes containing the regulatory elements are very similar

amongst all *Drosophila* species and also in the medfly *Ceratitis capitata* [34, 71]. The number of chorion genes in the follicle cells is amplified (see Fig. 2.6, p. 28) about 20-fold in the X-chromosome cluster and 60- to 80-fold in the cluster on chromosome 3 [105]. In both clusters, a stretch of 80–100 kb is amplified and this far exceeds the chorion protein locus (8 kb) in size. The amplification is caused by an "amplification control element" of 510 bp, which is probably the start of replication, as well as by several amplification enhancers [124]. Amplification of the chorion genes has also been detected in *Ceratitis capitata* [71]. The gene families of the silk moth chorion proteins, with up to 200 genes, belong to the largest known multi-gene families; they are particularly suitable for the study of evolutionary processes such as gene duplication, rearrangement and conversion [119]. *Bombyx mori* has one pseudogene each for the type A and Hc proteins; the latter makes an mRNA which, however, is not translated [36].

Table 11.4. Amino acid compositions of various silks in mol % [63, 114]

	1a	1b	1c	2	3	4	5	6
Ala	29.8	16.9	6.0	22.5	21.2	22.5	23.9	28.5
Arg	0.2	3.8	3.1	0.8	0.5	4.9	1.0	1.9
Asx	0.7	15.4	16.7	7.6	6.0	5.9	6.2	3.2
Cys		1.6	0.2					
Glx	0.7	8.4	4.4	7.4	0.8	6.7	8.5	9.7
Gly	49.4	10.0	13.5	10.4	24.6	18.7	7.9	11.8
His	0.1	1.6	1.3	0.7		2.0		0.1
Ile	0.1	7.3	0.7	2.8		1.9	1.4	2.5
Leu	0.1	7.2	1.1	3.0		2.9	5.3	6.4
Lys	0.1	1.5	3.3	0.7		2.4	1.3	1.1
Met		0.4	0.04	0.7				
Phe	0.4	2.7	0.5	0.3			1.1	
Pro	0.3	3.0	0.7	2.2				
Ser	11.3	7.9	33.4	28.3	42.7	17.1	30.1	22.4
Thr	0.5	2.8	0.5	4.3	3.1	3.2	3.3	4.3
Tyr	4.6	3.4	2.6	2.4	0.8	1.1	3.2	3.7
Val	2.0	7.4	2.8	4.8		5.4	5.4	2.3

1, Proteins from the silk glands of *Bombyx mori:* 1a, large fibroin subunit; 1b, small fibroin subunit (25 kDa); 1c, sericin; 2, *Hypera postica* (Coleoptera): larval silk; 3, *Chrysopa flava* (Neuroptera): ovipositor; 4, *Lasciocampa quercus* (Lepidoptera): cocoon; 5, *Araneus diadematus* (Araneae): egg cocoon; 6, *Nephila senegalensis* (Araneae): egg cocoon

11.9 Silk Proteins

Silk is produced by spiders, some mites and, in particular, insect larvae; it is found only rarely in adult insects, e.g. in the ovipositor of the green lacewing *Chrysopa* and other Neuroptera, the cocoons of *Hydrophilus* and other beetles, or the ootheca of the praying mantis (Mantidae). The silk of the spiders is synthesized in special spinning glands in the abdomen, and that of insect larvae is usually produced in silk glands in the head region; these are derived from salivary glands. The silk-like material of adult insects is produced in glands associated with the oviduct *(Hydrophilus, Mantis)* or the Malpighian tubules *(Chrysopa)*. In some beetles (e.g. *Ptinus*) it is actually material of the peritrophic membranes, and is a form of chitin (γ chitin) not otherwise found in the arthropods. Silk and silk production have, for obvious reasons, been studied most thoroughly in the silk moth *Bombyx mori*. At the end of the fifth stage, the silkworm makes a cocoon, in which it pupates, by spinning a single thread that is 2 μm thick and 1000–1500 m long. The cocoon has a dry weight of 300–500 mg, of which two-thirds consists of the fibrillar protein fibroin. The fibroin thread which makes up the cocoon is enveloped in a gelatinous layer, formed largely of the protein sericin. Both these proteins have a very unusual amino acid composition (Table 11.4). **Fibroin** is synthesized in the posterior part of the silk gland and stored in the central section (Fig. 11.9); there, the fibroin granules are embedded in sericin, which is synthesized in the middle section. The anterior section of the silk gland is a narrow delivery tube, out of which the silk thread is forced at a velocity of 0.4–1.5 cm/s; the shear forces arising in this process are important for the development of the mechanical properties of the silk. In preparation for fibroin synthesis, the cells of the posterior silk gland enlarge about 50 000-fold and reach the extraordinary size of 0.3 mm. In parallel, the DNA, and thus the number of copies, of the fibroin gene increases during 17–18 replication steps by 200 000-fold; the nuclei of the mature cells have multiple branches. The sericin genes are amplified 400 000-fold. The silk glands adapt to the special demands of fibroin synthesis in that they accumulate the tRNAs and aminoacyl-tRNA-synthases specific for glycine, alanine and serine. Whereas the 20–30 constitutive $tRNA^{Ala}$ genes are dispersed in the genome, the 20 silk-gland-specific $tRNA^{Ala}$ genes form a tightly linked cluster. They are not amplified, and the increase in $tRNA^{Ala}$ concentration involves regulation of their transcription [130]. The giant fibroin molecule (365–415 kDa) is encoded in a correspondingly large mRNA,

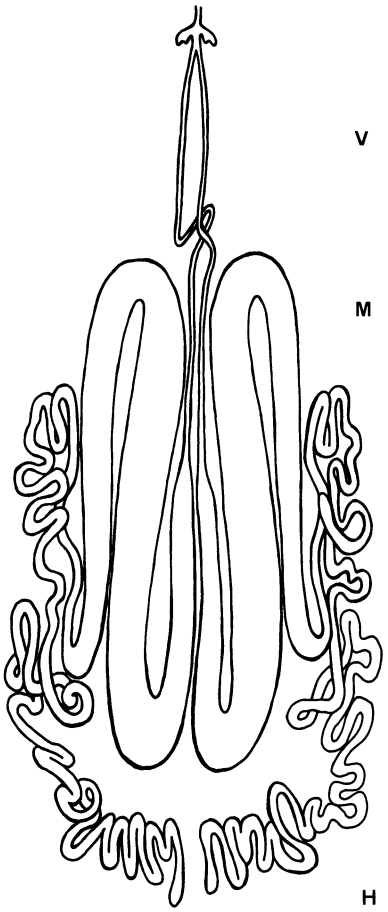


Fig. 11.9. The silk glands of *Bombyx mori* consist of three regions with different structures and functions [114]. *H*, Dorsal, *M*, middle and *V*, ventral region. Further details are given in the text

which is processed simultaneously by 45–122 ribosomes. The completion of a fibroin molecule of about 4700 amino acids takes about 25 min, i.e. four amino acids are coupled per second [114]. Translation, however, is discontinuous and is interrupted at particular sites on the mRNA (p. 81).

The long **fibroin chain** of *Bombyx mori* consists of alternating highly organized "crystalline" segments and "amorphous" segments, with the former making up about 60% of the molecular mass. The highly organized segments consist of repeats of 59 amino acids with the sequence -GAGAGSGAAG(SGAGAG)$_8$Y-, and the less organized regions have repeats of -GX-, where X may be one of many different amino acids. The gene sequence indicates that the fibroin chain is made up of ten crystalline blocks, each of four to seven times 59 amino acids, alternating with amorphous segments of 200 amino acids. Because of its high proportion of highly organized structures, the silk fibroin of *Bombyx mori* is particularly suitable for X-ray analysis. It was with this protein that, in the 1950s, Pauling and Corey discovered and described the classical β-sheet structure. These are neighbouring, antiparallel-orientated peptide chains that are linked by H-bonding into a folded sheet, which bears glycine residues on one side and alanine/serine residues on the other side. The folded sheets are stacked one above the other so that either the smaller glycines or the larger alanine/serine residues lie opposite each other; the differences in space required produce alternating distances between the sheets of 0.35 and 0.57 nm (Fig. 11.10). The fibroin stored as granules in the central section of the silk gland is not yet highly organized; the supramolecular structure which gives silk its special mechanical properties is conferred by the forces involved in ejection of the thread [114].

The **fibroin gene**, which exists as only one copy in the haploid genome, has a length of 16 kb and includes a single intron towards the 5′ end. The (G+C) content of the fibroin genes of the silk moths *Bombyx mori, Antheraea pernyi, A. yamamai* and *Philosamia cynthiae* is about 60%, which corresponds to the codon preference of GGN for alanine and GCN for glycine [126]. Comparison of various lines of *Bombyx mori* reveals considerable polymorphism of the fibroin gene. In particular, variation in the number and length of the crystalline blocks results in molecular masses of 365–415 kDa [114]. For a long time it was thought that the fibroin molecule was a single large polypeptide. However, in 1972 it was discovered that reduction of a disulphide bridge causes the appearance of a **small subunit** of 25 kD with an

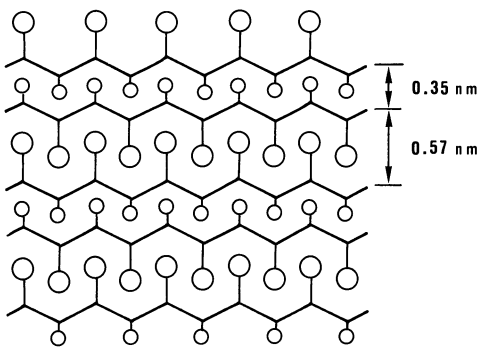


Fig. 11.10. The organization of the β-sheets in silk fibres from *Bombyx mori*. The distance between the layers varies depending on the presence of small glycine or larger alanine/serine residues

amino acid composition quite different from that of the **large subunit** [Table 11.4 (1b)]. The light fibroin chain, encoded by its own gene, is 244 amino acids in length. This includes three cysteine residues, of which two form an internal disulphide bridge, and the third is responsible for the link to the heavy chain [142]. Fibroin contains an asparagine-linked oligosaccharide of N-acetylglucosamine and mannose. A gene for the large fibroin chain, sequenced from the Japanese oak eggar *Antheraea yamamai*, consists of two exons, as in *Bombyx mori*. The two genes show close agreement in a region covering 300 bp of the 5′-NT region and the 70 bp of the first small exon, but do not show close agreement in the second exon, which encodes the major part of the fibroin sequence. Thus, the 5′-NT region is important for expression of the fibroin gene in the posterior silk gland, and is highly conserved [126].

Much less is known about **sericin**, the adhesive protein of the silk thread. Two-thirds of the amino acids are serine, glycine and aspartic acid [Table 11.4 (1c)]; in addition to N-glycosidic asparagine-bound carbohydrate, as found in fibroin, sericin also contains O-glycosidic serine- or threonine-bound oligosaccharides of N-acetylgalactosamine and galactose. Six sericin polypeptides with molecular masses between 65 and 400 kDa have been isolated from the silk glands of *Bombyx mori;* these are the result of alternative splicing of transcripts from only two genes [80].

The amino acid compositions, conformations and mechanical properties of the **various silks** are extraordinarily varied. The amino acid spectrum does not always correspond to the scheme in *Bombyx mori*, where glycine, alanine and serine predominate (Table 11.4). Thus, there are silks high in glutamine, as in those of *Digelansinus diversiipes* and *Arge ustalata* [114], and the silk of the lepidopteran *Calpodes ethlius* contains no less than 25.5 % proline. The fibroins can be ordered into eight groups according to their glycine content (Table 11.5). For groups 1–6, one can assume a classical β-sheet structure with variable distances between one sheet and the next (c value). The higher the proportion of amino acids with large side-chains, and thus the number and extent of the amorphous regions in the molecule, the greater the elasticity of the silk. Whereas classical silk is made up of only one chain type and in X-ray diagrams shows a periodicity with two layers, there are also parallel-β silks with two or three different chain types, which have a period of three layers according to the schemes AAB or ABC. Such unusual silks are found, for example, in the caddice fly *Olinga fereday*, parasitic wasps of the genus *Macrocentrus* and sawflies of the genus *Cladius* [114].

Table 11.5. Classification of silks according to their glycine content (mol %) and conformation. The c value (nm), i.e. the distance between one sheet and the next but one, is given for the parallel β-sheet silks [114]

Species	Glycine	Group	Conformation
Phymatocera aterrima	66	0	Polyglycine II
Bombyx mori	43	1	β (c = 0.92)
Anaphe moloeyi	37	2	β (c = 1.00)
Antheraea mylitta	25	3	β (c = 1.06)
Thaumetopoea pictyocampta		4	β (c = 1.50)
Nephila senegalensis		5	β (c = 1.57)
Digelansinus diversiipes	2	6	β (c = 1.38)
Arge ustalata	6	7	α helix
Nematus ribesii	36		Collagen-like
Chrysopa sp.	25		Xβ (cross-β)

There are also silks with structures other than the parallel-β, in which case the **secondary structure of the silk proteins** is critical for the mechanical properties. The cross-β conformation is derived from the stretched β-sheet structure by regular folding of the chains perpendicular to the axis of the thread (Fig. 11.11). The cross-β structure is found in the ovipositors of the green lacewings *Chrysopa* and in the silks of the beetles *Hydrophilus piceus* and *Hypera* sp. and the fungus gnat *Arachnocampa luminosa;* they are probably quite widely distributed. There are silks with an α-helical structure, e.g. in the ootheca of the praying mantis (Mantidae), in the cocoon of the flea *Xenopsylla*, as a four-stranded coiled-coil in the cocoon of the honey bee and, in addition to parallel-β structures, in the silk of the moth *Arge ustalata*. Finally, a silk with a collagen structure has been described for the cocoon of the sawfly *Nematus ribesii*, and silk with the helix structure of polyglycine has been found in the moth *Phymatocera aterrima*. Passive stretching can change these conformational states, e.g. cross-β can change into the stretched parallel-β form [63, 114]. The silks of some emperor moths (Saturniidae) contain a tanning system. The newly formed silk remains white so long as it is dry; moisture initiates the tanning process and the colour changes to yellow or brown. The tanning agents have been identified as 2-amino-3-hydroxybenzoic acid-3-O-glucoside in *Actias selene*, 2,5-dihyroxybenzoic acid-5-O-glucoside in *Antheraea pernyi*, both of these in *Hyalophora cecropia* and *H. gloveri*, and N-(3,4-dihydroxyphenyllactyl)-dihydroxyphenylalanine in *Dictyo-*

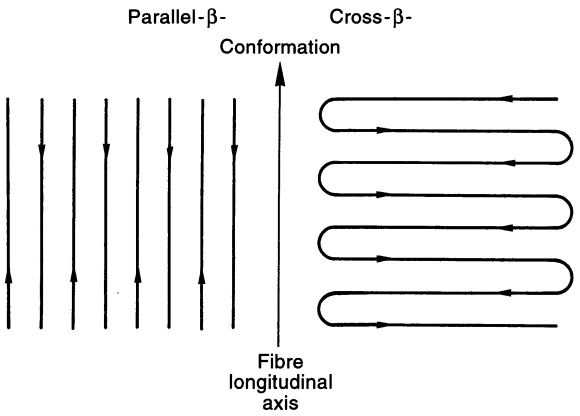


Fig. 11.11. The orientation of the polypeptide chains in the stretched parallel-β-sheet structure and in the folded cross-β (Xβ) structures of different silks. The *arrow* indicates the direction of the silk-fibre longitudinal axis [63]

ploca japonica [62]. The biochemistry of the spider's web unfortunately remains fairly obscure. *Nephila claviceps* can produce six different silks in its six spinning glands. The major ampullate-gland (drag-line) silk has a tensile strength of 29 000 kPa, higher than that of steel, and an elasticity of 35 %. The threads are constructed of pseudocrystalline regions with antiparallel β-sheets interspersed with elastic, amorphous segments. The repeat unit of 34 amino acids includes a polyalanine sequence of five to seven residues and five GGX tripeptides in which the position X is occupied by either alanine, tyrosine, leucine or glutamine [139].

11.10 Insect Secretory Proteins

Proteinaceous secretions are produced by many animals. Apart from the slimes (mucins), which will be discussed together with other glycoproteins in Chapter 13, only a few insect secretory proteins have been examined from the point of view of comparative biochemistry. In the larvae of the non-biting midges of the genus *Chironomus*, a secretion from the salivary gland is used in the construction of the nest tubes and the capture funnels. This consists of a mixture of proteins of greatly varying mass (17–1200 kDa). Of the 12–15 main components, the best known are the SPI or Balbiani-ring proteins, which are phosphoglycoproteins of 800–1200 kDa that can be spun into elastic threads. They are encoded by the genes BR1, BR2.2 and BR6, and are named after the so-called Balbiani rings, which appear as thickened "puffs" in the light-microscope image of the polytene chromosomes, and show active RNA synthesis. The SPI genes form multi-gene families, several members of which have been completely or partially sequenced from *Chrionomus tentans, C. pallidivittatus* and *C. thummi*. The genes have a hierarchical repetitive structure. In each gene there are about 100 almost identical repeats of 180–300 bp, each with a constant region without internal periodicity and a subrepeat region made up of 3–12 subrepeats of 9–33 bp. This gene structure apparently arose by multiple duplication and diversification. In addition to the four SPI proteins, a further three secretory proteins have been found and named after their mass in kDa: SP-195 contains numerous charged amino acids in 25-residue repeats, SP-140 is rich in glycine and charged amino acids and occurs in 14-residue repeats, and SP-185 contains no mutually homologous repeats but has repeats of three consecutive cysteines every 22 amino acids [38, 134].

The larvae of *Drosophila melanogaster* produce a secretion in the salivary glands and this is used to fasten the pupa to the substratum. The main components are the **Sgs or glue proteins**, of which eight are now known. The Sgs genes are distributed on various chromosomes but are coordinately expressed. The genes and the encoded polypeptides are very diverse. Gene Sgs-5 contains two introns; the genes Sgs-3, -7 and -8 have only one intron; and gene Sgs-4 has no introns. Sgs-3 and -4 consist of about 300 amino acids, have a high threonine content and are highly glycosylated. Sgs-5 has a length of only 163 amino acids and contains only a little carbohydrate. Sgs-7 and -8 are not glycosylated. However, they all contain many cysteines and can be cross-linked by disulphide bridges to form networks [72].

The **accessory glands of the male genital tract** in all insects produce a secretion with numerous species-specific peptide and protein components, some of which stimulate egg-laying in the females and reduce copulatory behaviour. The best investigated of these secretions are those of *Drosophila* species, in which 2-D electrophoresis of secretions from the paragonial glands revealed several hundred different proteins in species-specific patterns. Little is known about the structure of these proteins. One of the peptides from *Drosophila melanogaster* and *D. sechellia* has been sequenced and shown to have three differences between the species in its sequence of 36 amino acids [17]. Within a few minutes, copulation results in a drastic increase in protein biosyn-

thesis in the accessory glands, not as the result of increased transcription of the structural genes but from increased synthesis of rRNAs and ribosomal proteins [112]. In the mealworm *Tenebrio molitor*, as in many other insects, the secretion from the male accessory gland is used in the construction of a spermatophore. This is a complicated object with a multilayered outer wall and an inner space filled with sperm-containing fluid. Several "spermatophorins", with masses from 14 to more than 100 kDa, can be extracted from the spermatophores of the mealworm, but none have been characterized in any detail [115].

References

1. Ala-Kokko L. et al.: Structure of cDNA clones coding for the entire preproα1(III) chain of human type III procollagen. Biochem. J. 260: 509–516 (1989)
2. Alexander L. J. et al.: Isolation and characterization of the bovine kappa-casein gene. Eur. J. Biochem. 178: 395–401 (1988)
3. Andersen S. O.: Comparison between cuticular proteins from two locust species, Locusta migratoria and Schistocerca gregaria. Insect Biochem. 18: 751–760 (1988)
4. Andersen S. O.: Cuticular proteins from the shrimp, Pandalus borealis. Comp. Biochem. Physiol. Pt. B 99: 453–458 (1991)
5. Ann D. K., Smith K. M. and Carlson D. M.: Molecular evolution of the mouse proline-rich protein multigene family. Insertion of a long interspersed repeated DNA element. J. Biol. Chem. 263: 10887–93 (1988)
6. Barth C. A. and Schlimme E. (eds.): Milk proteins. Nutritional, clinical, functional and technological aspects. Springer, New York 1989
7. Bashir M. M. et al.: Characterization of the complete human elastin gene. Delineation of unusual features in the 5'-flanking region. J. Biol. Chem. 264: 8887–91 (1989)
8. Beck K., Hunter I. and Engel J.: Structure and function of laminin: anatomy of a multidomain protein. Faseb J. 4: 148–160 (1990)
9. Blumberg B., MacKrell A. J. and Fessler J. H.: Drosophila basement membrane procollagen α1(IV). II. Complete cDNA sequence, genomic structure, and general implications for supramolecular assemblies. J. Biol. Chem. 263: 18328–37 (1988)
10. Boccaccio G. L. and Quesada-Allué L. A.: In vivo biosynthesis of a stage-specific cuticle glycoprotein during early metamorphosis of the medfly Ceratitis capitata. Biochem. biophys. Res. Commun. 164: 251–258 (1989)
11. Bonaldo P. et al.: Structural and functional features of the α3 chain indicate a bridging role for chicken collagen VI in connective tissues. Biochemistry 29: 1245–54 (1990)
12. Bonsing J. et al.: Complete nucleotide sequence of the bovine β-casein gene. Aust. J. biol. Sci. 41: 527–537 (1988)
13. Brazel D. et al.: Completion of the amino acid sequence of the α1 chain of human basement membrane collagen (type IV) reveals 21 non-triplet interruptions located within the collagenous domain. Eur. J. Biochem. 168: 529–536 (1987)
14. Burke R. D., Bouland C. and Sanderson A. I.: Collagen diversity in the sea urchin, Strongylocentrotus purpuratus. Comp. Biochem. Physiol. Pt. B 94: 41–44 (1989)
15. Burley R. W. et al.: Proteins of the albumen and vitelline membrane of eggs of the estuarine crocodile, Crocodylus porosus. Comp. Biochem. Physiol. Pt. B 88: 863–867 (1987)
16. Butticè G. et al.: Evolution of collagen IV genes from a 54-base pair exon: A role for introns in gene evolution. J. mol. Evol. 30: 479–488 (1990)
17. Chen P. S. and Balmer J.: Secretory proteins and sex peptides of the male accessory gland in Drosophila sechellia. J. Insect Physiol. 35: 759–764 (1989)
18. Chi H. S. and Hui C. F.: Primary structure of the Drosophila laminin B2 chain and comparison with human, mouse, and Drosophila laminin B1 and B2 chains. J. Biol. Chem. 264: 1543–50 (1989)
19. Chiquet-Ehrismann R.: What distinguishes tenascin from fibronectin? Faseb J. 4: 2598–2604 (1990)
20. Chu M. L. et al.: Mosaic structure of globular domains in the human type VI collagen α3 chain: similarity to von Willebrand factor, fibronectin, actin, salivary proteins and aprotinin type protease inhibitors. Embo J. 9: 385–393 (1990)
21. Constantz B. and Weiner S.: Acidic macromolecules associated with the mineral phase of scleractinian coral skeletons. J. exp. Zool. 248: 253–258 (1988)
22. Cox D. L. and Willis J. H.: Post-translational modifications of the cuticular proteins of Hyalophora cecropia from different anatomical regions and metamer. Insect Biochem. 17: 469–484 (1987)
23. Cox G. N. et al.: Sequence comparisons of developmentally regulated collagen genes of Caenorhabditis elegans. Gene 76: 331–344 (1989)
24. D'Alessio M. et al.: Structure and developmental expression of a sea urchin fibrillar collagen gene. Proc. Nat. Acad. Sci. USA 86: 9303–07 (1989)
25. Dardik R. and Lahav J.: Multiple domains are involved in the interactions of endothelial cell thrombospondin with fibronectin. Eur. J. Biochem. 185: 581–588 (1989)
26. Devinoy E. et al.: Sequence of the rabbit whey acidic protein cDNA. Nucleic Acids Res. 16: 8180 (1988)
27. Devinoy E. et al.: Sequence of the rabbit α-s1-casein cDNA. Nucleic Acids Res. 16: 11813 (1988)
28. Dublet B. and Vanderrest M.: Type-XIV collagen, a new homotrimeric molecule extracted fom fetal bovine skin and tendon, with a triple helical disulfide-bonded domain homologous to type-IX and type-XII collagens. J. Biol. Chem. 266: 6853–58 (1991)
29. Edwards H. A.: Occurrence of resilin in elastic structures in the food-pump of reduviid bugs. J. exp. Biol. 105: 407–409 (1983)
30. Ehrig K. et al.: Merosin, tissue-specific basement membrane protein, is a laminin-like protein. Proc. Nat. Acad. Sci. USA 87: 3264–68 (1990)
31. Exposito J. Y. and Garrone R.: Characterization of a fibrillar collagen gene in sponges reveals the early evolutionary appearance of two collagen gene families. Proc. Nat. Acad. Sci. USA 87: 6669–73 (1990)

32. Eyre D. R., Paz M. A. and Gallop P. M.: Cross-linking in collagen and elastin. Annual Rev. Biochem. 53: 717–748 (1984)
33. Feeney R. E.: Penguin egg-white and polar fish blood-serum proteins (invited Review). Int. J. Peptide Protein Res. 19: 215–232 (1982)
34. Fenerjian M. G. et al.: Evolution of the autosomal chorion cluster in Drosophila. II. Chorion gene expression and sequence comparisons of the s16 and s19 genes in evolutionary distant species. J. mol. Evol. 29: 108–125 (1989)
35. Fleischajer R., Olsen B. R. and Kuehn K. (eds.): Structure, molecular biology, and pathology of collagen. New York Acad. Sci. 1990
36. Fotaki M. E. and Iatrou K.: Identification of a transcriptionally active pseudogene in the chorion locus of the silkmoth Bombyx mori. Regional sequence conservation and biological function. J. mol. Biol. 203: 849–860 (1988)
37. Francois J.: Biochemical and biophysical characterization of the collagen from an insect, Tenebrio molitor L. (Coleoptera, Tenebrionidae). Biochimie 67: 1035–42 (1985)
38. Galli J. et al.: A new member of a secretory protein gene family in the dipteran Chironomus tentans has a variant repeat structure. J. mol. Evol. 31: 40–50 (1990)
39. Gaub M. P. et al.: The chicken ovalbumin promoter is under negative control which is relieved by steroid hormones. Embo J. 6: 2313–20 (1987)
40. Gibson C. et al.: Structure and expression of the bovine amylogenin gene. Biochemistry 30: 1075–79 (1991)
41. Godovac-Zimmermann J.: The structural motif of β-lactoglobulin and retinol-binding protein: a basic framework for binding and transport of small hydrophobic moleculea. Trends biochem. Sci. 13: 64–66 (1988)
42. Gope M. L. et al.: Molecular cloning of the chicken avidin cDNA. Nucleic Acids Res. 15: 3595–3606 (1987)
43. Gordon M. K. et al.: Type XII collagen. A large multidomain molecule with partial homology to type IX collagen. J. Biol. Chem. 264: 19772–78 (1989)
44. Gotoh Y. et al.: Comparison of two phosphoproteins in chicken bone and their similarities to the mammal bone proteins, osteopontin and bone sialoprotein-II. Biochem. biophys. Res. Commun. 173: 471–479 (1990)
45. Gratecos D. et al.: Drosophila fibronectin: a protein that shares properties similar to those of its mammalian homologue. Embo J. 7: 215–223 (1988)
46. Guo X. and Kramer J. M.: The two Caenorhabditis elegans basement membrane (type IV) collagen genes are located on separate chromosomes. J. Biol. Chem. 264: 17574–82 (19989)
47. Gutman A., Yamada K. M. and Kornblihtt A.: Human fibronectin is synthesized as a pre-pro-polypeptide. FEBS Letters 207: 145–148 (1986)
48. Hall L.: Nucleotide sequence of Guinea-pig kappa-casein cDNA. Nucleic Acids Res. 18: 6129 (1990)
49. Halliday J. A., Bell K. and Shaw D. C.: The complete amino acid sequence of feline β-lactoglobulin-II and a partial revision of the equine β-lactoglobulin-II sequence. Biochim. biophys. Acta 1077: 25–30 (1991)
50. Hayman A. R., Koppel J. and Trueb B.: Complete structure of the chicken α2(VI) collagen-gene. Eur. J. Biochem. 197: 177–184 (1991)
51. Heilig R. et al.: The ovalbumin gene family: complete sequence and structure of the Y gene. Nucleic Acids Res. 10: 4363–82 (1982)
52. Hennighausen L. G., Steudle A. and Sippel A. E.: Nucleotide sequence of cloned cDNA coding for mouse epsilon-casein. Eur. J. Biochem. 126: 569–572 (1982)
53. Henzel W. J. et al.: Sarcophagid larval proteins: partial sequence homologies among three cuticle proteins and related structures of drosophilids. J. mol. Evol. 22: 39–45 (1985)
54. Hojrup P., Andersen S. O. and Roepstorff P.: Isolation, characterization, and N-terminal sequence studies of cuticular proteins from the migratory locust, Locusta migratoria. Eur. J. Biochem. 154: 153–150 (1986)
55. Hostikka S. L. et al.: Identification of a distinct type IV collagen α chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome. Proc. Nat. Acad. Sci. USA 87: 1606–10 (1990)
56. Indik Z. et al.: Alternative splicing of human elastin mRNA indicated by sequence analysis of cloned genomic and complementary DNA. Proc. Nat. Acad. Sci. USA 84: 5680–84 (1987)
57. Isa Y. and Okazaki M.: Some observations on the Ca-binding phospholipid from scleractinian coral skeletons. Comp. Biochem. Physiol. Pt. B 87: 507–512 (1987)
58. Jones C. W. and Kafatos F. C.: Accepted mutations in a gene family: Evolutionary diversification of duplicated DNA. J. mol. Evol. 19: 87–103 (1982)
59. Jones W. K. et al.: The rat casein multigene family. Fine structure and evolution of the β-casein gene. J. biol. Chem. 260: 7042–50 (1985)
60. Kallunki T. et al.: Structure of the human laminin B2 gene reveals extensive divergence from the laminin B1 chain gene. J. Biol. Chem. 266: 221–228 (1991)
61. Kauffman D. L. et al.: Basic proline-rich proteins from human parotid saliva. Relationships of the covalent structures of the proteins from a single individual. Biochemistry 30: 3351–56 (1991)
62. Kawasaki H. and Sato H.: The tanning agent in the silk of the Japanese giant silkmoth Dictyoploca japonica Butler. Insect Biochem. 15: 681–684 (1985)
63. Kenchington W.: The larval silk of Hypera spp. (Coleoptera: Curculionidae). A new example of the cross-ß protein conformation in an insect silk. J. Insect Physiol. 29: 355–361 (1983)
64. Kikuchi Y. and Tamiya N.: Chemical taxonomy of the hinge-ligament proteins of bivalves according to their amino acid compositions. Biochem. J. 242: 505–510 (1987)
65. Kimbrell D. A. et al.: Cuticle protein gene expression during the third instar of Drosophila melanogaster. Insect Biochem. 18: 229–235 (1988)
66. Kimura S. and Karasawa K.: Squid cartilage collagen: isolation of type I collagen rich in carbohydrate. Comp. Biochem. Physiol. Pt. B 81: 361–365 (1985)
67. Kimura S., Uematsu Y and Miyamuchi Y.: Shark (Prionace glauca) elastoidin: characterization of its collagen as α1(E)-homotrimers. Comp. Biochem. Physiol. Pt. B 84: 305–308 (1986)
68. Kimura S. et al.: The human α2(XI) collagen (COL11A2) chain. Molecular cloning of cDNA and genomic DNA reveals characteristics of a fibrillar collagen with differences in genomic organization. J. Biol. Chem. 264: 13910–16 (1989)

69. Kiss I. et al.: Structure of the gene for cartilage matrix protein, a modular protein of the extracellular matrix. Exon/intron organization, unusual splice sites, and relation to α chains of β-integrins, von Willebrand factor, complement factors B and C2, and epidermal growth factor. J. Biol. Chem. 264: 8126–34 (1989)
70. Kitagaki-Ogawa H. et al.: Diversities in animal vitronectins. Differences in molecular weight, immunoreactivity and carbohydrate chains. Biochim. biophys. Acta 1033: 49–56 (1990)
71. Komitopoulou K. et al.: The chorion genes of the medfly, Ceratitis capitata, 1. Structural and regulatory conservation of the S36 gene relative to two Drosophila species. Nucleic Acids Res. 18: 1731–37 (1990)
72. Kress H. and Swida U.: Drosophila glue protein expression. Naturwissenschaften 77: 317–324 (1990)
73. Labatrobert J., Biharivarga M. and Robert L.: Extracellular matrix. FEBS Letters 268: 386–393 (1990)
74. Lecanidou R. et al.: Evolution of the silk moth chorion gene superfamily: gene families CA and CB. Proc. Nat. Acad. Sci. USA 83: 6514–18 (1986)
75. Lian J. et al.: Structure of the rat osteocalcin gene and regulation of vitamin D-dependent expression. Proc. Nat. Acad. Sci. USA 86: 1143–47 (1989)
76. Lonnerdal B. et al.: Cloning and sequencing of a cDNA encoding milk beta-casein. FEBS Letters 269: 153–156 (1990)
77. Mann K. et al.: Amino acid sequence of mouse nidogen, a multidomain basement membrane protein with binding activity for laminin, collagen IV and cells. Embo J. 8: 65–72 (1989)
78. Marsh M. E.: Histidinoalanine, a naturally occurring cross-link derived from phosphoserine and histidine residues in mineral-binding phosphoproteins. Biochemistry 25: 2392–96 (1986)
79. Matsui R., Ishida M. and Kimura S.: Characterization of two genetically distinct type I-like collagens from lamprey (Entosphenus japonicus). Comp. Biochem. Physiol. Pt. B 95: 669–675 (1990)
80. Michaille J. J.; Garel A. and Prudhomme J. C.: Cloning and characterization of the highly polymorphic Ser2 gene of Bombyx mori. Gene 86: 177–184 (1990)
81. Minamisako K. and Kimura S.: Characterization of muscle collagen from fleshy prawn Penaeus chinensis. Comp. Biochem. Physiol. Pt. B 94: 349–353 (1989)
82. Miura S and Kimura S.: Jellyfish mesogloea collagen. Characterization of molecules as α1α2α3 heterotrimers. J. biol. Chem. 260: 15352–56 (1985)
83. Morris N. P. et al.: The tissue form of type VII collagen is an antiparallel dimer. J. biol. Chem. 261: 5638–44 (1986)
84. Muragaki Y. et al.: The complete primary structure of 2 distinct forms of human α1(IX) collagen chains. Eur. J. Biochem. 192: 703–708 (1990)
85. Muragaki Y. et al.: The α2(VIII) collagen gene: A novel member of the short chain collagen family located on the human chromosome-1. J. Biol. Chem. 266: 7721–27 (1991)
86. Murthy C. V. R. and Adiga P. R.: Purification of biotin-binding protein from chicken egg yolk and comparison with avidin. Biochim. biophys. Acta 786: 222–230 (1984)
87. Myllyla R. et al.: Molecular cloning of chick lysyl hydroxalase: Little homology in primary structure to the two types of subunit of prolyl 4-hydroxylase. J. Biol. Chem. 266: 2805–10 (1991)
88. Nakato H. et al.: Structure and expression of messenger RNA for a pupal cuticle protein of the silkworm, Bombyx mori. Insect Biochem. 20: 667–678 (1990)
89. van Ness K. P., Koob T. J. and Eyre D. R.: Collagen cross-linking: distribution of hydroxypyridinium cross-links among invertebrate phyla and tissues. Comp. Biochem. Physiol. Pt. B 91: 531–534 (1988)
90. Nies D. E.: The complete cDNA sequence of human hexabrachion (tenascin)-A multidomain protein containing unique epidermal growth factor repeats. J. Biol. Chem. 266: 2818–23 (1991)
91. Niyibizi C. and Eyre D. R.: Identification of the cartilage α1(XI) chain in type V collagen from bovine bone. FEBS Letters 242: 314–318 (1989)
92. Oldberg A. et al.: A collagen-binding 59-kd protein (fibromodulin) is structurally related to the small intestitial proteoglycans PG-S1 and PG-S2 (decorin). Embo J. 8: 2601–04 (1989)
93. Olson P. F. et al.: Glutactin, a novel Drosophila basement membrane-related glycoprotein with sequence similarity to serine esterases. Embo J. 9: 1219–27 (1990)
94. Oyama F. et al.: Patterns of alternative splicing of fibronectin pre-mRNA in human adult and fetal tissues. Biochemistry 28: 1428–34 (1989)
95. Park Y. S. and Kramer J. M.: Tandemly duplicated Caenorhabditis elegans collagen genes differ in their modes of splicing. J. mol. Biol. 211: 395–406 (1990)
96. Patel R. S. et al.: Organization of the fibronectin gene provides evidence for exon shuffling during evolution. Embo J. 6: 2565–72 (1987)
97. Pervaiz S and Brew K.: Purification and characterization of the major whey proteins from the milk of the bottlenose dolphin (Tursiops truncatus), the Florida manatee (Trichechus manatus latirostris), and the beagle (Canis familiaris). Arch. Biochem. Biophys. 246: 846–854 (1986)
98. Pihlajaniemi T., Pohjolainen E. R. and Myers J. C.: Complete primary structure of the triple-helix region and the carboxy-terminal domain of a new type IV collagen chain, α5(IV). J. Biol. Chem. 265: 13758–66 (1990)
99. Pihlajaniemi T. and Tamminen M.: The α1 chain of type XIII collagen consists of three collagenous and four noncollagenous domains, and its primary transcript undergoes complex alternative splicing. J. Biol. Chem. 265: 16922–28 (1990)
100. Pollock J. et al.: Chick tropoelastin isoforms. From the gene to the extracellular matrix. J. Biol. Chem. 265: 3697–3702 (1990)
101. Raju K. and Anwar R. A.: Primary structures of bovine elastin a, b and c deduced from the sequences of cDNA clones. J. biol. Chem. 262: 5755–62 (1987)
102. Raju K. and Anwar R. A.: A comparative analysis of the amino acid and cDNA sequences of bovine elastin a and chick elastin. Biochem. Cell Biol. 65: 842–845 (1987)
103. Ramshaw J. A. M., Werkmeister J. A. and Bremner H. A.: Characterization of type I collagen from the skin of the blue grenadier (Macruronus novaezelandiae). Arch. Biochem. Biophys. 267: 497–502 (1988)
104. Rebers J. E. and Riddiford L. M.: Structure and expression of a Manduca sexta larval cuticle gene homologous to Drosophila cuticle genes. J. mol. Biol. 203: 411–423 (1988)
105. Regier J. C.: Evolution and higher-order structure of architectural proteins in silkmoth chorion. Embo J. 5: 1981–89 (1986)

106. Regier J. C. and Wong J. R.: Assembly of silkmoth chorion proteins; in vivo patterns of disulfide bond formation. Insect Biochem. 18: 471–482 (1988)
107. Rose C., Kumar M. and Mandal A. B.: A study of the hydration and thermodynamics of warm-water and cold-water fish collagens. Biochem. J. 249: 127–133 (1988)
108. Runnegar B.: Collagen gene construction and evolution. J. mol. Evol. 22: 141–149 (1985)
109. Ruoslahti E.: Fibronectin and its receptor. Annual Rev. Biochem. 57: 375–413 (1988)
110. Sabatini L. M., He Y. Z. and Azen E. A.: Structure and sequence determination of the gene encoding human salivary statherin. Gene 89: 245–251 (1990)
111. Sabsay B. et al.: Domain structure and sequence distribution in dentin phosphophoryn. Biochem. J. 276: 699–707 (1991)
112. Schmidt T., Stumm-Zollinger E. and Chen P. S.: Protein metabolism of Drosophila melanogaster male accessory glands – III. Stimulation of protein synthesis following copulation. Insect Biochem. 15: 391–401 (1985)
113. Shadwick R. E. and Gosline J. M.: Molecular biomechanics of protein rubbers in molluscs. In: Wilbur K. M. (ed.): Mollusca, Vol. 1, pp. 399–430. Acad. Press, New York 1983
114. Shimura K. et al.: The physiology and biology of spinning in Bombyx mori (Review). Experientia 39: 441–473 (1983)
115. Shinbo H., Yaginuma T. and Happ G. M.: Purification and characterization of a proline-rich secretory protein that is a precursor to a structural protein of an insect spermatophore. J. biol. Chem. 262: 4794–99 (1987)
116. Skorstengaard K. et al.: Purification and complete primary structures of the heparin-, cell- and DNA-binding domains of bovine plasma fibronectin. Eur. J. Biochem. 154: 15–29 (1986)
117. Soininen R. et al.: Structural organization of the gene for the α1 chain of human type IV collagen. J. Biol. Chem. 264: 13565–71 (1989)
118. Souliotis V. et al.: Cuticle proteins during the development of Daucus oleae. Insect Biochem. 18: 485–492 (1988)
119. Spoerel N. A. et al.: Gene evolution and regulation in the chorion complex of Bombyx mori. Hybridization and sequence analysis of multiple developmentally middle A/B chorion gene pairs. J. mol. Biol. 209: 1–19 (1989)
120. Sridhara S.: Evidence that pupal and adult cuticular proteins are coded by different genes in the silkmoth Antheraea polyphemus. Insect Biochem. 15: 333–339 (1985)
121. Stiles B. and Leopold R. A.: Cuticle proteins from the Anthonomus grandis abdomen: stage specificity and immunological relatedness. Insect Biochem. 20: 113–125 (1990)
122. Strawich E. and Glimcher M. J.: Tooth „enamelins" identified mainly as serum proteins. Major „enamelin" is albumin. Eur. J. Biochem. 191: 47–56 (1990)
123. Suyama K. and Nakamura F.: Isolation and characterization of new crosslinking amino acid „allodesmosine" from hydrolysate of elastin. Biochem. biophys. Res. Commun. 170: 713–718 (1990)
124. Swimmer C. et al.: Evolution of the autosomal chorion cluster in Drosophila. III. Comparison of the s18 gene in evolutionary distant species and heterospecific control of chorion gene amplification. J. mol. Biol. 215: 225–235 (1990)
125. Takahara K. et al.: Complete primary structure of human collagen α1(V) chain. J. Biol. Chem. 266: 13124–29 (1991)
126. Tamura T. and Sakate (+) S.: Isolation and characterization of fibroin mRNAs from the saturniid silkworms, Antherea yamamai, Antherea pernyi and Phiolsamia cynthia ricini. Insect Biochem. 18: 169–175 (1988)
127. Thomas J. T. et al.: Isolation of cDNAs encoding the complete sequence of bovine type-X collagen. Evidence for the condensed nature of mammalian type-X collagen genes. Biochem. J. 273: 141–148 (1991)
128. Threadgill D. W. and Womack J. E.: Genomic analysis of the major bovine milk protein genes. Nucleic Acids Res. 18: 6935–42 (1990)
129. Timpl R.: Structure and biological activity of basement membrane proteins. Eur. J. Biochem. 180: 487–502 (1989)
130. Underwood D. C. et al.: Silk gland-specific tRNA-Ala genes are tightly clustered in the silkworm genome. Mol. cell. Biol. 8: 5504–12 (1988)
131. Vartio T. and Vaheri A.: Fibronectin: chains of domains with diversified functions. Trends biochem. Sci. 8: 442–444 (1983)
132. Veis D. J. et al.: Matrix proteins of the teeth of the sea urchin Lytechinus variegatus. J. exp. Zool. 240: 35–46 (1986)
133. Vuorio E. and de Crombrugghe B.: The family of collagen genes. Annual Rev. Biochem. 59: 837–872 (1990)
134. Wellman S. E. and Case S. T.: Disassembly and reassembly in vitro of complexes of secretory proteins from Chironomus tentans salivary glands. J. Biol. Chem. 264: 10878–83 (1989)
135. Willis J. H.: Curicular proteins: The neglected component. Arch. Insect Biochem. Physiol. 6: 203–215 (1987)
136. Willis J. H.: Partial amino acid sequences of cuticular proteins from Hyalophora cecropia. Insect Biochem. 19: 41–46 (1989)
137. Worms D. and Weiner S.: Mollusc shell organic matrix – fourier transform infrared study of the acidic macromolecules. J. exp. Zool. 237: 11–20 (1986)
138. Wright G. M., Keeley F. W. and Youson J. H.: Lamprin: a new vertebrate protein comprising the major structural protein of adult lamprey cartilage. Experientia 39: 495–497 (1983)
139. Xu M. and Lewis R. V.: Structure of a protein superfiber: Spider dragline silk. Proc. Nat. Acad. Sci. USA 87: 7120–24 (1990)
140. Yada T. et al.: Occurrence in chick embryo vitreous humor of a type IX collagen proteoglycan with an extraordinarily large chondroitin sulfate chain and short α1 polypeptide. J. Biol. Chem. 265: 6992–99 (1990)
141. Yamada K. M.: Adhesive recognition sequences (Minireview). J. Biol. Chem. 266: 12809–12 (1991)
142. Yamaguchi K. et al.: Primary structure of the silk fibroin light chain determined by cDNA sequencing and peptide analysis. J. mol. Biol. 210: 127–139 (1989)
143. Yoshimura M, Banerjee M. R. and Oka T.: Nucleotide sequence of a cDNA encoding mouse beta casein. Nucleic Acids Res. 14: 8224 (1986)
144. Yoshinaka R. et al.: Two genetically distinct types of collagen in kuruma prawn Penaeus japonicus. Comp. Biochem. Physiol. Pt. B 96: 451–456 (1990)

145. Youtz S. L., Packard M. J. and Leach jr. R. M.: Total desmosine content in the organic matrix of various avian and reptile eggs. Comp. Biochem. Physiol. Pt. B 77: 27–29 (1984)
146. Yu-Lee L. et al.: Evolution of the casein multigene family: conserved sequences in the 5' flanking and exon regions. Nucleic Acids Res. 14: 1883–1902 (1986)
147. Zaman K. and Zak Z.: Purification and properties of chicken egg-white cobalamin-binding protein. Biochimie 72: 711–714 (1990)
148. Zeichner-David M. et al.: Biosynthesis and characterization of rabbit tooth enamel extracellular matrix proteins. Biochem. J. 251: 631–641 (1988)
149. Zheng D. B. et al.: Chicken riboflavin-binding protein. cDNA sequence and homology with milk folate-binding protein. J. Biol. Chem. 263: 11126–29 (1988)
150. NATO Advanced Research Workshop: Biology of invertebrate and lower vertebrate collagens. Plenum, New York 1985

12 Small Nitrogenous Compounds

12.1 Free Amino Acids

Because of the constant turnover of proteins, protein-bound and free amino acids exist in a dynamic equilibrium. The intracellular pool of free amino acids, which is replenished by the hydrolysis of existing proteins, by uptake from the intercellular space and by de novo synthesis, is available for protein synthesis and for the many other metabolic processes dependent upon amino acids. The concentration of free amino acids is always lower than that of the protein-bound residues, one limiting factor being the strong osmotic effects of such low molecular weight compounds. Thus, there is no specific amino acid store in an organism; it is more the case that enzymes and structural proteins themselves represent the reserve of amino acids. Insect larvae, which require large amounts of amino acids for metamorphosis and sclerotization of the cuticula, make use of extracellular proteins as a source of amino acids; examples include the larval haemolymph proteins of some dipterans (p. 191) and possibly the haemoglobins of the non-biting midges (Chironomidae) (p. 273).

12.1.1 The Chemical Nature of Free Amino Acids

As the biosynthesis of cell proteins requires 20 **standard amino acids**, these must all be present in the pool of free amino acids. Previous publications about certain organisms or tissues in which fewer amino acids were typically found were based upon the insufficiently sensitive methods of paper and thin-layer chromatography. In addition to the standard amino acids, the pool of small nit-

rogenous substances contains **"rare" amino acids** which, despite the name, in fact occur quite often. They include protein-associated amino acids which were chemically altered post-translationally and released by proteolysis, in particular methyl, hydroxyl and halogen derivatives. There are, however, rare amino acids which are never protein bound. Some of the latter arise during normal metabolism in animal cells and are, therefore, more-or-less ubiquitous, e.g. ornithine, which is a product of arginine cleavage by arginase, and β-amino-isobutyric acid and β-alanine, which are degradation products of pyrimidine bases (Fig. 12.29); other such amino acids are formed in subsidiary pathways of metabolism. As higher plants are particularly rich in the products of secondary metabolism, it is often the case that unusual amino acids found in herbivores originate from nutrients. One example of this is the first report of its kind from the animal kingdom of S-methylcysteine in the caterpillar of the moth *Prodenia eridania*. It was later found that this methyl amino acid was derived from S-methylcysteine sulphoxide in the cabbage leaves on which the insects fed (Fig. 12.1); the amino acid was not found after the insects were fed with potatoes. Citrulline as an intermediary product of arginine biosynthesis should occur only in animals in which this biosynthetic pathway is functional. However, some plants are very rich in citrulline and, as a consequence, the concentration may be very high in herbivorous snails.

More than 200 rare amino acids have been identified to date in plants and animals. These are mostly single, independent studies about which further information can be found in the literature [62]. We will restrict ourselves here to an examination of cases in which such amino acids have assumed special functions or have an interesting distribution. Thyroxin and other iodinated tyrosine derivatives formed in the thyroid are of particular interest from a comparative biochemistry point of view and will be dealt with in a separate section

In the case of the **N-acetylated amino acids**, which are found in high amounts in some vertebrates, it is not clear whether the differences that exist have some adaptive significance and, if so, how this functions. N-acetylhistidine and N-acetylaspartic acid (Fig. 12.1) serve as a store of acetyl residues in vertebrate brain, e.g. for the synthesis of acetylcholine. Both amino acids are found in the brains of teleosts, but amphibians and reptiles have only acetylhistidine, and birds and mammals have only acetylaspartic acid. Searches for these two amino acids in the brains of chondrostei, agnathans and cephalopods were unsuccessful. The lower vertebrates also have high concentrations of acetylhistidine in the lens of the eye. The significance of the histidine cycle found in the carp eye is not known. Histidine dif-

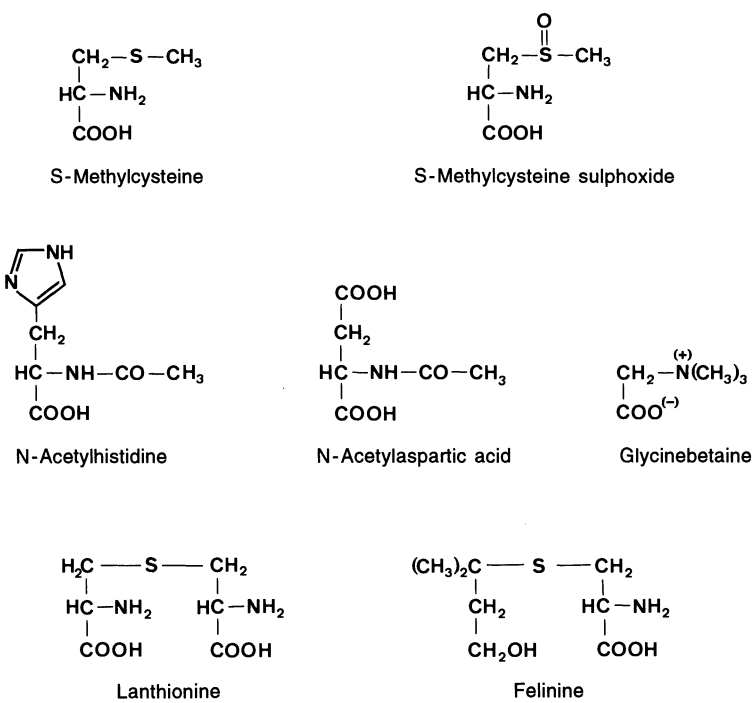


Fig. 12.1. Rare amino acids

fuses out of the eye fluid into the lens, where it is acetylated; it then diffuses back into the eye fluid and is cleaved by a deacetylase. All freshwater animals have large amounts of acetylhistidine in the heart; marine teleosts, in contrast, have much less. In fish that undertake spawning migration, the concentration of acetylhistidine in the heart correlates to the habitat in which embryo development occurs, e.g. it is tenfold higher in the salmon than in the eel [82].

Methylated amino acids are formed post-translationally in numerous proteins and are released by hydrolysis. They include, in particular, N-mono-, di- and trimethyllysine, 3-methylhistidine, and mono- and dimethylarginine. 1-Methylhistidine, on the other hand, arises by methylation of the free amino acid. Thus, for example, the silkworm *Bombyx mori* contains eight different methyl derivatives of lysine, histidine and arginine, often at high concentrations. Trimethyllysine exceeds free lysine during the fifth larval stage, where it enters the ovary of the pupa and accumulates in the eggs; adult males contain only trace amounts. In this case, the trimethyllysine apparently has a specific role, perhaps as the starting point for carnitine synthesis [174]. Mammalian erythrocytes also contain high concentrations of trimethyllysine [153]. The marked, species-specific differences in the concentration of methylated amino acids are exploited to identify certain sorts of meat in foodstuff chemistry. For example, the presence of 3-methylhistidine is an indicator of pork in meat products. A large number of N-trimethyl amino acids (betaines) are known in plants, whereas animals contain mainly glycinebetaine (Fig. 12.1), which arises by oxidation of choline and is an important donor of methyl groups. It is of great importance for intracellular osmoregulation in many marine invertebrates. Thus, its concentration in the tissues of marine molluscs and crustaceans can reach very high levels of 80–120 μmol/g [10].

The sulphur-containing amino acid **taurine** (Fig. 12.8) has various biological functions: it is an important osmoregulatory substance in many marine animals; it is conjugated with bile acids, retinoic acid or certain foreign substances in vertebrates; it is bound in peptide form in the synaptosomes of calf brain; and it is a constituent of the phosphagen taurocyamine. It is therefore not surprising to find marked species- and tissue-specific variations in taurine content. The concentration in skeletal muscle is usually higher than in other tissues of the same animal [262]. The highest recorded concentration of about 100 μmol/g fresh weight (1.3 % of the total fresh weight) is found in the giant nerve fibres of the cephalopod *Loligo* sp. Values of around 10 μmol/g are found in the tissues of many marine invertebrates and teleosts, but values are much lower in the freshwater fish, crayfish and insects. Deamination of the taurines produces **isethionic acid** (Fig. 12.8), which is also found at its highest level in cephalopod nerves: the giant axons of *Loligo* contain more than 150 μmol/g fresh weight, i.e. no less than 20 % of the dialysable anions. Much lower concentrations have been recorded in the ganglions of the snail *Aplysia* and the mussel *Spisula*, and in various mammalian organs; about 40–200 μmol pen day are excreted in human urine [82, 271]. Free **L-lanthionine** (Fig. 12.1) has been found in most insects examined, but not in crustaceans, annelids, molluscs, echinoderms, tunicates or vertebrates. In contrast, the structurally related amino acid cystathionine is widely found as an intermediate in the metabolism of methionine to cysteine (Fig. 12.7). An enzyme discovered in the fat body of *Bombyx mori* produces lanthionine from cysteine and is apparently very similar, or even identical, to cystathionine synthase, which produces cystathionine from homocysteine and serine. The fat body also contains an enzyme that cleaves lanthionine to cysteine, ammonia and pyruvate [225]. The sulphur-containing amino acid **felinine** (Fig. 12.1) has an especially interesting distribution: it is excreted in amounts of up to 120 mg per day in the urine of the domestic cat and somewhat less in the urine of the ocelot; it is absent from other cat species, such as the lion, tiger and puma, and is found nowhere else, except in some bacteria [62, 82].

12.1.2. Concentration and Spectrum of the Free Amino Acids

The ninhydrin reagent normally used for the determination of free amino acids also reacts to varying degrees, depending on the conditions, with unsubstituted amino groups such as amines, amino-sugars and ammonia. In contrast, peptides react less strongly than their constituents released by hydrolysis. Therefore, when considering concentration data it is necessary to distinguish between free amino acids, non-protein-bound amino acids (after hydrolysis of the oligopeptides) and total ninhydrin-positive substances. With the exception of certain insect species, the concentration of peptides in the tissue and body

fluids is always so low that the values for free and non-protein-bound amino acids hardly differ; on the other hand, the value for total ninhydrin-positive material may be much higher. The **concentration of free amino acids** differs markedly with the species and the tissue, and is usually higher intracellularly than extracellularly. High extracellular amino acid concentrations are found in the haemolymph of many insects. The more primitive groups of the dragonflies, cockroaches and locusts, however, have relatively low concentrations of 25–50 mmol/l; the highest values of more than 120 mmol/l are to be found in the lepidopterans, where the difference between intra- and extracellular fluids is minimal.

The **spectrum of free amino acids** in tissues and body fluids has been investigated in many species. The introduction of paper chromatography, which allowed amino acid analyses to be carried out by the analytically inexperienced, released a flood of investigations, mainly with taxonomic goals. Reliable quantitative data, however, are obtainable only by means of column or gas chromatography. Of the total free amino acids, the non-essential amino acids alanine, glycine, proline, serine, glutamic acid and glutamine are almost always in the majority, whereas asparagine, aspartic acid, cysteine and all the essential amino acids are present at lower concentrations. In contrast to the protein-bound amino acids, the spectrum of free amino acids varies so widely with the tissue and species that generalizations are difficult to formulate (Table 12.1). Older publications often quote very high glutamic acid concentrations in arthropod haemolymph. This would be surprising because glutamic acid serves as a neurotransmitter in arthropods and produces strong pharmacological effects. In fact, in these species glutamic acid is found exclusively in the blood cells, as is the case for the majority of the free amino acids in arthropod blood [46].

The concentration of each individual amino acid in the cellular pool is the resultant of influx and efflux, and can therefore be homeostatically maintained at a constant level or changed as required. Each analysis of free amino acids represents only the dynamic equilibrium at one point in time, and numerous internal and external factors can lead to changes in the spectrum. This is the reason why the taxonomic use of the amino

Table 12.1. The spectrum of free amino acids (mol %) in tissues and body fluids

	Aur	Hym	Ach		Gec		Ast	Cal		Ict			Hum
			HL	Ggl	HL	Mus	HL	HL	T	Ser	Mus	H	Ser
Ala	2	42	20	22	14	14	34	13	7	7	19	9	12
Arg	1	1	<1	11	3	3	5	3	14	3	3	2	3
Asx	1	2	5	7	1	1	1	–	8	2	1	6	} 24
Glx	8	16	7	13	1	2	1	7	26	4	3	31	
Gly	44	13	11	4	11	34	14	5	7	11	29	7	8
His	<1	2	5	3	2	2	1	6	9	2	2	5	3
Ile	1	1	2	1	1	<1	2	1	2	8	4	4	2
Leu	1	3	2	1	2	1	3	1	3	12	5	7	4
Lys	2	2	6	13	2	<1	6	–	2	10	9	7	6
Met	<1	1	<1	<1	<1	<1	2	–	–	1	2	1	<1
Phe	<1	1	1	1	1	<1	1	1	<1	4	2	2	2
Pro	1	2	5	3	25	20	10	52	13	9	7	4	8
Ser	2	5	34	19	30	21	8	6	7	6	4	5	4
Thr	1	1		3	<1		9	3	<1	7	5	6	6
Val	1	2	3	2	3	1	5	1	1	10	4	5	6

Aur, *Aurelia aurita* (Cnidaria) laboratory grown polpys in 2 % salt solution with 12.9 µmol free amino acids (faa) per g fr. wt. [259].
Hym, *Hymenolepsis diminuta* (Cestoda) with 13.9 µmol faa per g fr. wt. [36].
Ach, *Achatina fulica* (Gastropoda): haemolymph (HL) with 0.49 µmol faa per ml; and suboesophageal ganglion (Ggl) with 28 µmol faa per g fr. wt. [148].
Gec, *Gecarcinus lateralis* (Crustacea): HL with 3.8 µmol faa per ml; and claw muscles (Mus) with 190 µmol faa per g fr. wt. [277].
Ast, *Astacus leptodactylus* (Crustacea): HL with 0.63 µmol faa per ml [208].
Cal, *Calliphora erythrocephala* (Insecta): HL with 22 µmol faa per ml; and tissue (T) with 71 µmol faa per g fr. wt. (calculated from the values for whole animal and haemolymph) [55].
Ict, *Ictalurus punctatus* (Teleostei): serum (Ser) with 2.3 µmol faa per ml; muscle (Mus) with 5.6 µmol faa per g fr. Wt.; and heart (H) with 6.6 µmol faa per g fr. wt. [264].
Hum, human: serum (Ser) with 2.8 µmol faa per ml [220].

acid spectrum, for which one had great hopes 30 years ago, eventually proved to be unsuitable. Of the factors which can influence the concentration and variety of free amino acids, nutrition plays a significant role, in that it directly changes the input into the pool. Fundamental changes in the amino acid pool are observed following changes in the physiological situation (e.g. during metamorphosis of insects [46] and amphibians, and in the ecdysis cycle of the crustaceans), at times of increased protein synthesis (e.g. silk production, and formation of sex substances), or in response to parasite infection. Changes in internal conditions are the basis of seasonal and diurnal fluctuations in the amino acid pool. External factors that may influence the concentrations of free amino acids include temperature, oxygen partial pressure and osmolality.

12.1.3 Osmotic Functions of Small Nitrogenous Compounds

Inorganic ions constitute usually more than 98 % of the **osmotically active substances** of the extracellular fluids. Organic molecules make significant contributions to the extracellular osmolality in only a few animal groups, e.g. amino acids or sugars in insect haemolymph or urea, and trimethylamine oxide in the blood of certain fishes. In tissues, inorganic ions usually contribute about half the osmolality and the rest mainly arises from the presence of nitrogenous substances such as aminocarbonic acids, taurine, glycinebetaine, trimethylamine oxide and urea. The dominant role of **urea as an osmotic substance** in the marine elasmobranchs and crossopterygians is an interesting problem. Here, the urea concentrations in the blood plasma are generally about 300–450 mmol/l, and in the case of the ray *Trygonorhina* as high as 600 mmol/l; the intracellular concentrations are even higher. Such high urea concentrations should, in fact, disturb the functional and structural integrity of the cell proteins. It appears, however, that the simultaneous presence of methylamino compounds, such as trimethylamine oxide, glycinebetaine and N-methylglycine (sarcosine), compensates for the destabilizing effects of urea [253].

The osmolality of metazoan cells always agrees more or less with that of the extracellular environment, and the cell must react to all changes in the extracellular osmolality. Adaptation **(isosmotic intracellular regulation)** results above all from changes in the concentration of nitrogenous compounds. This has the advantage that the concentrations of inorganic ions, which have other specific biological effects in addition to their osmotic functions, are subject to minimal change [90, 149, 253]. Comparable intracellular regulation has been observed following dehydration, e.g. in the edible snail *Helix pomatia*. Exactly which nitrogenous compound is most involved in osmoregulation depends upon both the species and the tissue (Table 12.2), and may also be determined by the speed at which the changes in salinity of the external medium occur. In various mussels, long-lasting changes in salinity lead to regulatory changes in taurine and amino acid concentration; this does not happen, for example, in response to the 12-hour rhythm of the tides in river estuaries.

In view of the variety and number of factors that affect the intracellular pools of small nitrogenous compounds, it is not surprising that there is no general explanation for the mechanisms of isosmotic intracellular regulation. Osmoregulatory increases in intracellular nitrogenous substances **(hyperosmotic regulation)** can occur by uptake from the extracellular space, through increases in de novo synthesis, reduced degradation, or a shift in the rate of protein turnover in favour of free amino acids. Several of these mechanisms are probably always involved simultaneously, depending upon the species and the tissue [82]. That some amino acids have their origins in cell proteins is suggested by the reported increases in essential amino acids, although the role of protein hydrolysis in intracellular regulation is still controversial. Interesting data have been reported in this connection for the edible mussel *Mytilus edulis*. Here, aminopeptidase I is encoded by the polymorphic Lap locus; animals with the Lap^{94} allele have a 20 % higher proteolytic enzyme activity and do, in fact, show more rapid increases in the amount of free amino acids following hyperosmotic stress [63].

Evidence for regulatory changes in the de novo synthesis of amino acids has been sought, on the one hand, in tracer experiments, which unfortunately often give ambigous results, and, on the other hand, in studies of the enzymes of amino acid metabolism in vitro. Glutamate dehydrogenase, which is without doubt a key enzyme of amino acid metabolism, is activated by NaCl and other inorganic salts. However, the ion sensitivity of this enzyme in an animal with a marked capacity for intracellular osmoregulation, such as the shore crab *Carcinus maenas*, has in fact been found to be lower than in the much less adaptable

Table 12.2. The osmoregulatory changes in the intracellular concentration of free amino acids in mmol/kg water (μmol/g dry wt. in *Bunodosoma*, Bun)

	Aca[a]	Lit[b]	Bal[c]	Bun[d]
Ala	2.4 → 80	17.7 → 3.9	1.2 → 16.3 → 24.7	6.2 → 13.2
β-Ala				1.5 → 41.9
Arg	0.1 → 0.5	6.3 → 5.3	1.5 → 7.6 → 5.7	
Asp	0.1 → 2.2	0.9 → 1.9	0.3 → 0.3 → 2.1	6.6 → 14.1
Glu	0.4 → 4.9	4.0 → 2.4	1.5 → 6.4 → 9.1	7.2 → 24.2
Gly	0.4 → 4.1	9.0 → 5.5	0.9 → 29 → 19.5	12.5 → 10.8
His	0.1 → 0.6	Sp → 0.3	Sp → 0.7 → tr	
Ile	0.1 → 1.1	0 0	Sp → 0.6 → 0.7	
Leu	0.2 → 1.8	0.8 → 1.0	0.1 → 0.8 → 1.4	
Lys	0.3 → 1.1	tr → 0.8	0.2 → 1.4 → 0.6	
Met	0.1 → 0.8	0 0	Sp → 0.6 → tr	
Phe	0.1 → 0.6	0.4 → 4.1	0 → 0.2 → 0	
Pro	0.4 → 33	19.0 → 10.8	0.2 → 119 → 510	
Ser	0.2 → 5.1	2.9 → 3.7	0.4 → 3.7 → } 7.6	
Thr	0.2 → 3.1	1.0 → 2.1	0.1 → 1.1	
Tyr	0.1 → 0.9	0.6 → 0.5	0 → 0.6 → 0	
Val	0.2 → 2.0	0.9 → 1.3	0.2 → 1.6 → 1.8	
Taurin		40 → 22	0.5 → 16.6 → 2.9	85.4 → 83.5

tr, trace

[a] Aca, *Acanthamoeba castellani* (Protozoa): amino acid-free medium with cane sugar increased from 40 to 500 mOsmol/kg water [69].

[b] Lit, *Littorina littorea* (Gastropoda): muscles of animals from seawater and 50 % seawater [82].

[c] Bal, *Balanus improvisus* (Crustacea): thorax muscle from animals in media with 4, 1031 or 1877 mOsmol/kg water [86].

[d] Bun, *Bunodosoma cavernata* (Cnidaria) after adaptation to a change in seawater composition from 2.6 to 4.0 % salt [124].

lobster, or in the rat [82]. The effect of enzyme inhibitors on osmoregulation has been examined during hyperosmotic stress in the isolated heart of the mussel *Modiolus demissus;* the transaminase inhibitors aminooxyacetic acid and L-cycloserine could not prevent increases in the intracellular amino acid concentration, but only produced certain shifts in the spectrum of additional amino acids produced. The question also remains of the origin of the nitrogen needed for the de novo formation of the amino acids. In the prawn *Macrobrachium rosenbergii*, uptake of ammonia from the medium increases during adaptation to increased osmolality. Detailed results are available for the sea anemone *Bunodosoma cavernata*, which responds to a jump in salinity from 2.6 to 4 % with a 20-fold increase in the concentration of free β-alanine, from 1.5 to 41.9 μmol/g dry weight. The rate of synthesis from aspartate increases threefold, whilst that from uracil remains constant; the degradation rate is reduced by a factor of 2.5–3.0 [124].

Reductions in intracellular nitrogenous substances during **adaptation to lower osmolalities** may involve lowered biosynthesis, increased degradation or efflux from the cell, reduced uptake from the extracellular space, increased incorporation of free amino acids into protein, or some combination of these mechanisms. The substances which decrease during hyposmotic stress are in many, but not all, cases the same as those that increase during hyperosmotic stress. Thus, in the euryhaline *Paramecium calkinsi*, the concentrations of proline and alanine are regulated in both situations [59]. In contrast, in *Acanthamoeba castellanii*, transfer from 240 to 500 mosmol/kg results in increases in proline, alanine and glutamate, whereas transfer from 240 to 40 mosmol/kg brings about a decline, especially of α-aminobutyrate and proline [89]. The role of protein synthesis in hyposmotic conditions is just as controversial as that of protein hydrolysis in the hyperosmotic state. In the shore crab *Carcinus maenas*, transfer to a dilute medium results in an increase in labelled leucine incorporation in the haemocyanin; however, apart from increased protein synthesis, this effect could also be due to non-covalent binding of leucine to the hydrophobic side-chain of the protein [280].

Increased amino acid degradation in connection with osmoregulatory processes may upset the nitrogen balance. In the polychaetes *Neanthes succinea* and *Leonereis culveri*, part of the amino-N is retained by amination of glutamic acid to glutamine. A rapid reduction in the extracellular osmolality causes entry of water into the cell and

an increase in volume. The immediate volume regulation that ensues relies upon reductions in the osmotically active substances of the cell by the mechanisms mentioned above.

12.2 Nutritional Aspects of Amino Acid Metabolism

The amino acid metabolism of animals is characterized by the facts that nitrogen for the synthesis of amino acids can only be used at the oxidation level of the amino group, and certain amino acids (essential amino acids) cannot be synthesized de novo. Thus, the nutrients must contain sufficient amino acids to replace those lost by excretion and to supply those required for further growth; they must include all the essential amino acids in at least the amounts needed for protein synthesis. As the cell has no significant reserves of essential amino acids, a nutritional deficiency of just one of these amino acids causes the limitation of protein synthesis and degradation of the excess of remaining amino acids. Although the non-essential amino acids are in fact all dispensable, feeding of a mixture of only essential amino acids results in a suboptimal nitrogen balance; at least one of the non-essential amino acids must be supplied.

Two methods are available for the determination of **amino acid requirements**. In the classical diet method, animals are kept, or even reared, on a chemically defined diet for long periods and are examined for signs of deficiency: reduced growth, delayed development, the occurrence of malformations, reduced longevity, increased nitrogen excretion, alterations in the blood amino acid spectrum, reduced incorporation of ^{14}C-leucine into protein, etc. In the ^{14}C method, a suitable labelled nutrient, such as ^{14}C-glucose, is administered and the radioactivity of isolated amino acids is determined; only non-essential amino acids are labelled. The latter method is used for animals that cannot be maintained on a chemically defined diet, e.g. many parasites, blood-suckers, microphages, or highly specialized plant eaters. The results from the two methods may be quite different. It is often the case that the capacity to synthesize an amino acid is insufficient in a certain stage of the life cycle; such "semi-essential" or "growth-promoting" amino acids are obviously recognized only by the diet method. For animals with yolk-rich eggs, the results of diet experiments may be confused by the transfer of essential amino acids from the yolk proteins to the developing animal. The supply of essential amino acids by symbiotic microorganisms may lead to false results, even in ^{14}C experiments, especially when the labelled substrate is supplied as feed and not injected into the bloodstream.

Numerous investigations, in particular with mammals, insects and protozoans, have shown that histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan and valine are all apparently **essential**. Thus, the ability to synthesize these amino acids was already lost very early in evolution. Only a few animal groups are able to make arginine, and even here the availability of arginine for protein biosynthesis is so limited by arginase activity that supply of this amino acid is growth stimulating, e.g. in many mammals and human babies. In several animal groups, insufficient amino acid metabolism creates further essential or semi-essential amino acids: some ciliates and the beetle *Trogoderma granarium* require tyrosine in addition to phenylalanine, i.e. the hydroxylation of phenylalanine is insufficient. Proline is semi-essential for many insects, e.g. for *Culex pipiens*, some lines of *Phormia regina* and *Cochliomyia hominivorax*, and perhaps for *Bombyx mori*. Crayfish of the genus *Astacus*, the freshwater prawn *Macrobranchium rosenbergii* and the mosquito *Culex pipiens* lack the ability to make asparagine from aspartic acid; the amide must be present in the diet. Some insects cannot meet their requirements for sulphur-containing amino acids with methionine alone and also require cysteine; examples include the butterfly *Argyrotaenia velutinana* and several aphids. Glycine and serine in the diet improve the development and survival of *Culex pipiens*. The domestic cat is unable to make taurine and a deficiency leads to severe symptoms, such as degeneration of the retina; humans and other primates also require taurine in their diet [46, 82, 227].

So long as the appropriate aminotransferases are present, the essential amino acids can all be replaced by the analogous keto acids, although these are always only partly aminated. **D-amino acids** can be used after they have been deaminated by D-amino acid oxidases. In this regard, there are species-specific differences: many insects can make use of D-histidine, D-methionine and D-phenylalanine; *Aedes* uses D-tryptophan and D-leucine; and *Tribolium* uses D-lysine. None of the D-amino acids is growth stimulating for *Bombyx mori;* in fact, some are weakly inhibitory. In the Peking duck, D-methionine is also unable to replace completely the L-enantiomer

[74]. Compounds structurally related to the essential amino acids **(amino acid antagonists)** inhibit metabolism of the natural forms and produce effects very similar to those of amino acid deficiency. In many cases, they are also used for protein synthesis and dysfunctional proteins are produced. Amongst the amino acid antagonists are natural compounds such as canavanine, produced by the Leguminosae (Fig. 12.2), which as an arginine analogue has proved toxic for all animals tested [136]. Most investigations of amino acid antagonists have been carried out on bacteria and the rat, and there are very few comparative data. Experiments with ethionine (Fig. 12.2) may be mentioned; this competes with methionine for incorporation into protein, e.g. in the parasitic dipteran *Agria housei*, and also disturbs sea urchin development.

In some groups of animals, **other N compounds** in addition to amino acids may be used as nitrogen sources with the help of symbiontic microorganisms. The best-known example is that of the ruminants, in which most of the urea circulating in the blood is not excreted in the urine but is released into the rumen, where it is converted into all 20 standard amino acids by symbionts. **Recycling of urea** between the blood and a symbiont-containing fermentation chamber of the alimentary canal is also a feature of other mammals, e.g. the wallaby kangaroo *Macropus eugenii*, the hyrax *Procavia habessinica*, and the American lagomorph *Sylvilagus floridanus*. The extent of the recycling varies with the species; it involves only about 50 % of the urea produced by

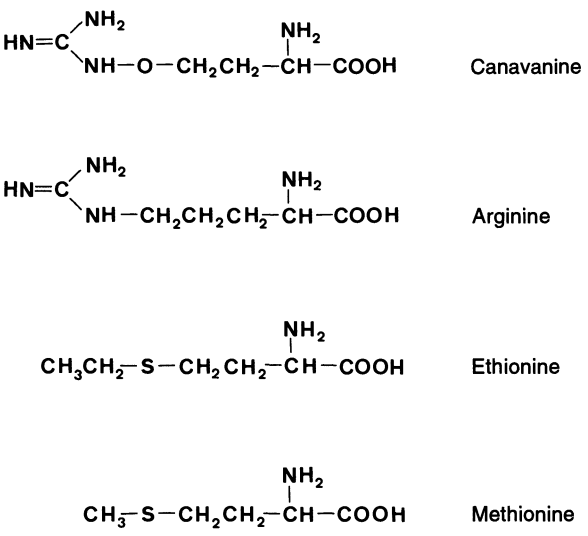


Fig. 12.2. Amino acid antagonists: canavanine to arginine, and ethionine to methionine

cattle, sheep and goats, but 74 % of that produced by the camel. In *Procavia*, the recycling increases to 70 % during conditions of water deficiency [213].

An interesting question is how the wood-eating termites satisfy their amino acid requirements, given their very protein-deficient diet. Gut symbionts certainly play an important role. In tracer experiments, essential amino acids also became labelled, and a reduction in the number of symbiontic bacteria and protozoans led to a decline in the concentrations of the free amino acids and in the incorporation of ^{14}C-acetate into amino acids. Such symbionts even allow the termite to **fix elementary nitrogen**. In most species this occurs to an extent which hardly influences the nitrogen balance; however, in *Nasutitermes corniger* the amount of nitrogen bound in 1 year is equivalent to the total nitrogen of the animal [22]. Nitrogen fixation has also been detected in mussels of the family Teredinidae (ship-worms), and a nitrogen-fixing bacterium has been isolated from one species.

12.3 End Products of Nitrogen Metabolism

Most of the excreted nitrogen-containing substances are end products of amino acid metabolism and only a few percent are degradation products of purines. The predominant end products of amino acid degradation are ammonia (ammoniotelic), urea (ureotelic) and uric acid (uricotelic). The pathway preferred during evolution depended upon the habitat of the animal. Primarily aquatic vertebrates and invertebrates which have sufficient water available to dilute the highly toxic ammonia are **ammoniotelic**. The fairly soluble urea can be released only in aqueous solution, where its concentration is limited by osmotic effects. The more-or-less insoluble uric acid and other purines can be excreted almost dry. The evolutionary alternatives of **ureotelic** or **uricotelic** are determined not so much by the water relations of the adult animal but much more by the possibility for the transport of substances through the egg membrane. Animals which produce eggs with an impermeable membrane, such as birds, terrestrial gastropods and arthropods, are characteristically uricotelic. In contrast, all terrestrial animals whose embryos can release their nitrogenous excretions into the environment or into the body of the mother are ureotelic, e.g. the

mammals and the terrestrial and semi-terrestrial amphibians. The situation of the reptiles is interesting but, unfortunately, inadequately investigated. The soft-shelled eggs of many snakes and turtles and also the hard-shelled eggs of the crocodile are permeable, and the developing embryos in these eggs always produce urea, although the adult animals are uricotelic. Surprisingly, in the eggs of the turtle *Trionyx spiniferus*, 70 % of the excreted nitrogen is urea, 24 % ammonia and only 6 % uric acid, despite the fact that these eggs are completely impermeable [176]. Owing to special adaptations, the elasmobranchs and crossopterygians are also ureotelic, as are the lungfish and terrestrial oligochaetes under certain conditions. In fact, animal excretions usually contain all three substances in various proportions, as well as further nitrogenous compounds, such as amino acids, allantoin or allantoic acid, trimethylamine oxide and other amines, creatine, creatinine, and nitrogenous substances that are partly of exogenous origin. The proportions of these substances greatly depend upon internal and external conditions.

The **Platyhelminthes** are primarily ammoniotelic. However, the terrestrial turbellarian *Bipalium kewense*, the endoparasitic trematodes *Fasciola hepatica* and *Schistosoma mansoni*, and the cestode *Hymenolepsis diminuta* produce and excrete substantial amounts of urea. Similarly, amongst the ammoniotelic **Nemathelminthes** several species excrete up to 20 % urea. *Ascaris lumbricoides* maintained in a small amount of external medium significantly increases its output of urea. All investigated nematodes excrete about 20–30 % of their nitrogenous waste as amino acids; *Trichinella spiralis* larvae and *Ditylenchus* sp. additionally produce small amounts of amines [18, 82].

Cephalopods and **bivalves** are strictly ammoniotelic. Pelagic squids of the family Cranchiidae, such as *Helicocranchia pfefferi*, make special use of their ammoniotelic nitrogen metabolism. A coelom sac, which takes up most of the mantle cavity, is filled with a fluid containing about 480 mmol/l ammonium chloride. Owing to its lower density relative to seawater, this fluid exerts a buoyancy that allows the animal to float. Cephalopod urine also contains guanine, and that of the mussels contains 25–38 % of the nitrogen in the form of amino acids, with an additional 7–13 % urea in the case of *Mya* and *Crassostrea* [26, 82].

In parallel with the great variety of their habitats, the **gastropods** show a wide spectrum of nitrogenous excretions. Aquatic snails excrete mainly ammonia, but in addition excrete species- and season-dependent amounts of urea. Uric acid is stored for a long time in the snails, and very variable results are obtained from analysis of the nephridium content and the excretions. The amounts of uric acid stored in aquatic forms are often just as high as in terrestrial varieties. Terrestrial gastropods produce urea and uric acid, and part of the latter is retained for long periods. Members of the families Helicidae and Limacidae also produce guanine and xanthine. Urea production depends upon the species and the conditions, e.g. *Helix pomatia* appears to produce only uric acid, whereas in *Lymnaea* urea makes up about 14 %, and in *Biomphalaria* in some situations it constitutes more than 50 % of the excreted nitrogen [26, 82]. The aquatic pulmonate snail *Biomphalaria glabrata* is the intermediate host for the trematode *Schistosoma mansoni*, the agent for bilharziosis. Increased protein degradation following the withdrawal of nutrients or infection with *Schistosoma* larvae causes an increase in the urea fraction from 20 to 55–60 % in *Biomphalaria* excretions. Under normal conditions terrestrial snails experience dramatic de- and rehydration events, but the resulting alterations in the osmolality of the extra- and intracellular fluids are tolerated. In the dry resting phase, several species markedly increase their synthesis and storage of urea; examples include *Bulimulus dealbatus, Limax flavus, Strophocheilus oblongus* and *Anguispira alternata*. The urea concentration in the haemolymph may increase to over 400 mmol/l. In other species, e.g. *Olata lactea* and *Helix aspersa*, urea is synthesized in large amounts but is continuously hydrolysed by a highly active urease; these species therefore excrete gaseous ammonia.

Amongst the **annelids**, polychaetes and leeches are all ammoniotelic. Amino acids are excreted by some leeches and oligochaetes, and urea excretion occurs in starving oligochaetes. The proportion of urea in the excretions and its increase with hunger varies with the species; in *Eisenia foetida* it is always less than 22 % and increases little with hunger, whereas in *Lumbricus terrestris* and *Allolobophora calliginosa* urea constitutes up to 90 % of the excreted nitrogen. The situation is similar in the freshwater oligochaetes. In *Lumbricus rubellus*, urea excretion increases with hunger, although ammonia is still excreted in large amounts. In contrast, *Eisenoides lonnbergi* remains purely ammoniotelic. The conditions bringing about the switch to ureotelic

excretion have been examined in more detail in *Lumbricus terrestris;* the critical factor is a decline in glycogen reserves from the normal 14 mg/g to less than 4 mg/g during starvation [255].

The typical excretory product of the **arachnids** is guanine, and this predominates in the real spiders, mites, uropygi, solifugi and amblypygi. The situation is more complicated in the scorpions. *Vejovis mexicanus* excretes only guanine, but *Lycas tricarinatus* also produces some urea. The excretions of *Paruroctonus mesaensis* contain over 90 % xanthine and some hypoxanthine, and this is the first known xanthotelic animal [278]. As expected, the **crustaceans** are strictly ammoniotelic; this is also true of the terrestrial isopods, which in fact excrete gaseous ammonia. The only **myriapod** investigated in any detail *(Cylindroiulus longinensis)* excreted 57 % ammonia, 38 % amino acids and 5 % uric acid. The **insects** in general are a model for uricotelic excretion, although systematic investigation of over 400 species showed that the activity of uricolytic enzymes resulted in the excretion also of allantoin, allantoic acid and urea [51]. Excretion of allantoin in fact predominates in *Dysdercus fasciatus*, and hypoxanthine and xanthine are found in addition to uric acid in some species. The butterfly *Colias croceus* stores guanine in the wings, and the tsetse fly *Glossina morsitans* excretes relatively large amounts of amino acids, especially arginine and histidine. Uric acid is sometimes stored in the specialized fat-body cells of insects; in *Periplaneta* and other cockroaches the storage is indefinite, and the uric acid is not released during the insect's lifetime [50, 51].

Echinoderms excrete large quantities of amino acids, e.g. 24–30 % of excreted nitrogen in the cases of *Asterias* and *Diadema*. In many echinoderms, more than 10 % of the excreted nitrogen is in the form of urea [237]. The **tunicates** are ammoniotelic. In the eggs of the ascidian *Corella willmeriana*, ammonium salts at a concentration of about 250 mmol/l serve to reduce the buoyant density, as in the pelagic squids. Some ascidians excrete ammonia but in addition store uric acid in the form of concretions in the nephridia [82].

In the **fish**, the elasmobranchs are ureotelic, as is the only living crossopterygian *Latimeria chalumnae*. The urea, however, is retained by the kidneys and accumulates in the intra- and extracellular fluids, where it is the most important substance of osmoregulation. In the freshwater ray *Potamotrygon* sp., the urea level in the plasma is relatively low and remains almost unchanged when the animal acclimatizes to dilute seawater. The bony fish excrete almost exclusively ammonia, although in the case of the mudskipper *Periophthalmus* up to 33 % urea is excreted. The nitrogen metabolism of the three dipnoan genera varies greatly in parallel with their abilities to withstand dry periods. The Australian lungfish *Neoceratodus*, which never leaves the water and whose lungs are used only for accessory air breathing, is almost purely ammoniotelic, with no more than 6 % of the excreted nitrogen occurring as urea. On the other hand, the South American lungfish *Lepidosiren*, which can live out of water but does not build a cocoon like *Protopterus*, excretes 38 % urea when living in water. The African lungfish *Protopterus* already has a urea fraction of over 50 % when in the water. During aestivation, the tissue urea concentration is as much as 1 % of the fresh weight. *Oreochromis alcalicus*, a species of fish living in an African soda lake at pH 10, has all the enzymes of the urea cycle in its liver; it excretes only urea, and no ammonia [200].

The rule for the **amphibians** is that the larvae and adults of persistently aquatic species are ammoniotelic, whilst the amphibious and terrestrial species are increasingly ureotelic [269]. The larvae of the ovoviviparous European salamander, which develop in the uterus up to the 14th month, are ureotelic. The larvae of the frog *Leptodactylus bufonis*, which develop in the foam of a mud nest, also produce mainly urea. During dehydration or an increase in the salt concentration of the environment, the internal osmolality of *Xenopus laevis*, various *Rana* species and *Bufo viridis* rises due to increased synthesis and retention of urea. The South African clawed frog *Xenopus laevis* can acclimatize to a 300-mmol/l sodium chloride solution; this corresponds to 60 % of the seawater osmolality. Acclimatization involves a three- to sixfold increase in activity of glutamate dehydrogenase and the enzymes of the urea cycle in the liver; the urea content of the blood plasma rises from about 4 to 200 mmol/l, and the biological half-life of the urea in the blood extends from 2 to 42 h. *Xenopus* larvae can also survive at 300 mosmol per kg water, but there are relatively minor changes in the urea concentration and its half-life in the plasma. Such adaptation to increased osmolality by an increase in urea concentration is also reported for other anurans; for example, *Rana cancrivora* from the mangrove swamps of Thailand survives 700 mosmol per kg water. Osmoregulatory changes in urea production are observed only in spring and summer animals of *Rana pipiens;* here, arginine

succinate lyase is inhibited by an increase in the level of aldosterone. Hormonal regulation of the osmoregulatory process has also been shown for *Rana temporaria*, but the exact mechanism is not known.

The South African tree frog *Chiromantis xerampelina* (Racophoridae) and the South American species *Phyllomedusa sauvagii* (Hylidae) possess a uricotelic mechanism which protects against dehydration. Uric acid, as a proportion of the total nitrogen excreted, is 50–57 % in *C. xerampelina*, about 80 % in *P. sauvagii*, but only 24–25 % in the related species *P. pailona, P. iherengi* and *P. hypochondrialis;* in other genera of the Hylidae, such as *Agalychnia, Pachymeduse* or *Hyla*, it is in fact always less than 3.5 %. In the liver of the uricotelic species, the activities of the enzymes of urea synthesis (arginase) and uricolysis (uricase) are relatively low, and those of the enzymes of uric acid synthesis (xanthine dehydrogenase) are relatively high. *Hyperolius viridiflavus* (Hyperoliidae) produces mainly guanine and hypoxanthine, which are deposited in iridophores. Thus, the osmotic problems accompanying water loss are solved and, at the same time, the high reflectance of these cells reduces heat absorption [218]. *H. nasutus*, which is also very resistant to dehydration, is, in contrast, purely ureotelic.

Amongst the **reptiles**, the snakes and lizards are all uricotelic, and the crocodile excretes mainly urea but also ammonia. Turtles and tortoises may be ammoniotelic, ureotelic or uricotelic, depending on their habitat. In addition to these species-specific differences there is also variation in the regulatory mechanisms. Aquatic and semi-aquatic species excrete ammonia and urea in about equal proportions; terrestrial species from wet biotopes produce much more urea than ammonia; and species from dry biotopes produce about 60 % uric acid and 40 % urea. Transfer of the turtle *Trionyx spiniferus* from freshwater to brackish water with 0.7 % salt results in an accumulation of urea in the tissues.

The urine of **birds** consists almost entirely of ammonium urate, i.e. it contains uric acid nitrogen and ammonia nitrogen in the ratio 4:1. Needham (1931) in his pioneering work *Chemical Embryology* described how, during chick development, ammonia is excreted initially (maximum at day 3), then urea (days 5–9) and, finally, uric acid; he cited this as an example of the biogenetic rule also being valid in biochemistry. However, already by 1935, he found that the urea in fact originated from preformed arginine. Today, it is quite clear that in the chicken, as in other birds, there is no functioning urea cycle.

Mammals are ureotelic. In addition to urea, their urine always contains uric acid or allantoin, as an end product of purine metabolism, and many other nitrogenous compounds. Hippuric acid, which is a product of the detoxification of benzoic acid, is found at high concentration in the urine of herbivores but only in trace amounts in the carnivores and omnivores.

12.4 Nitrogen Metabolism of the Amino Acids

The degradation of most amino acids begins with the cleavage of the amino group and formation of the corresponding keto acid. The L-amino acid oxidases that may be present in tissues are not significantly involved in this process; the usual situation in animals is that **deamination of amino acids** occurs by one of two more indirect pathways, one via glutamate dehydrogenase (GDH) and the other via the purine nucleotide cycle. In both pathways, the amino groups of different amino acids are initially transferred to a common intermediate (glutamic or aspartic acid), and only in a second step, if at all, are they released as ammonia. The **GDH pathway** occurs typically in vertebrate liver. Here, the glutamic acid intermediate is oxidatively deaminated by mitochondrial glutamate dehydrogenase. The fate of the ammonia released in the liver mitochondria differs markedly in different species. In ureotelic animals, it is bound to carbamylphosphate synthetase I and then incorporated into citrulline, which is released into the cytosol and there used for urea synthesis. The second amino group required comes from aspartic acid. In uricotelic animals, the ammonia is converted to glutamine by mitochondrial glutamine synthase and is then used in the cytosol for uric acid synthesis. The **purine nucleotide cycle** (Fig. 12.3) is typical of vertebrate muscle, which contains only low GDH activity. The product of the transamination is aspartic acid and its nitrogen then serves in the conversion of IMP to AMP; ammonia is released from the AMP by an AMP deaminase, giving again IMP.

Little is known of the relative importance of these two pathways of amino acid deamination in different invertebrate animals. Apart from in birds and mammals, the complete enzyme system of the purine nucleotide cycle has been found only in the goldfish and in the hepatopancreas of

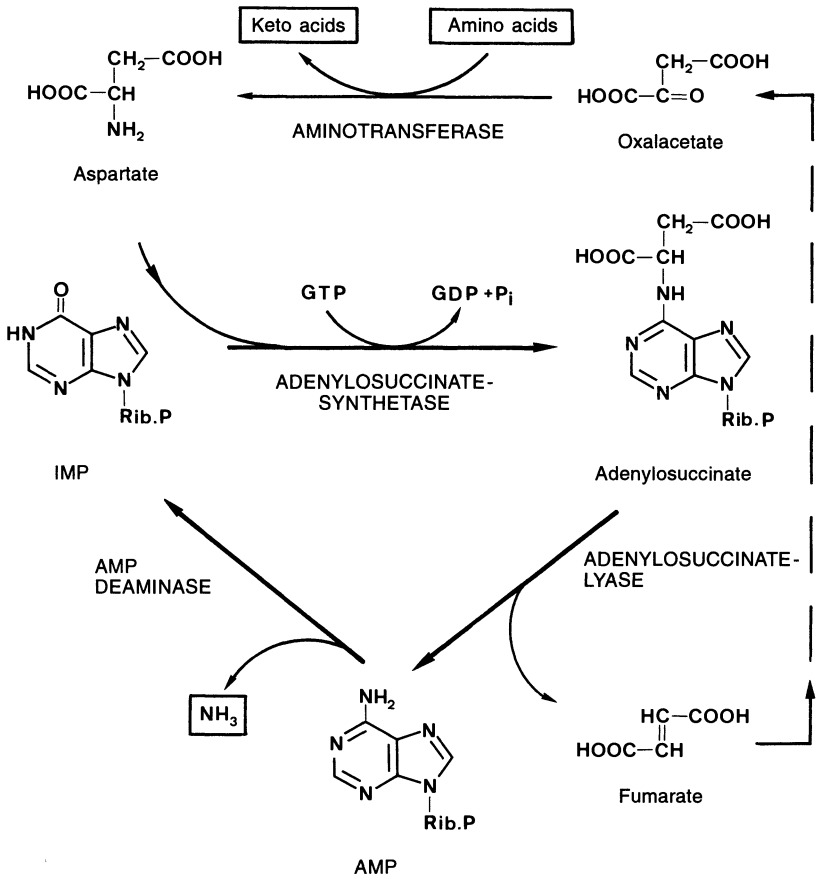


Fig. 12.3. Deamination of L-amino acids in the purine nucleotide cycle. The corresponding keto acids and aspartate are produced by amino transfer between and L-amino acid and oxaloacetate; aspartate and IMP are then converted to fumarate and AMP with adenylosuccinate as intermediate. Deamination of AMP releases ammonia and recycles IMP; the fumarate is oxidized in the citric acid cycle to give oxaloacetate. As a result, the L-amino acid is deaminated oxidatively to the corresponding keto acid

the pulmonate snail *Helix aspersa*. The white muscles of the goldfish contain GDH (0.09 U/g fresh weight) and the enzymes of IMP amination (0.11 U/g fresh weight) at approximately the same levels, and contain much higher levels of AMP deaminase (226 U/g fresh weight). In the liver, IMP amination enzymes (<0.02 U/g fresh weight) are insignificant compared with GDH (2.3 U/g fresh weight). As expected, isolated liver mitochondria only produce ammonia aerobically because GDH deaminates oxidatively; in muscle mitochondria, much more aspartic acid than ammonia is produced and this occurs also to some extent anaerobically [253]. In the *Helix* hepatopancreas, the activity of AMP deaminase is relatively low (about 2 U/g fresh weight); the adenylosuccinate synthase and the extremely labile adenylosuccinate lyase have even lower activities. However, there is no doubt that the purine nucleotide cycle is active in the hepatopancreas; the ammonia released can be bound by cytosolic glutamine synthetase and used for uric acid synthesis.

12.4.1 L-Amino Acid Oxidases

The rat kidney contains an enzyme with FMN as the prosthetic group; this enzyme deaminates, above all, leucine, methionine, proline and isoleucine with the formation of H_2O_2. However, it has such a low turnover rate as to be insignificant, and it is not found in other rat tissues or in the kidneys of other mammals. A completely different, very active form of D-amino acid oxidase is present in turkey liver; this enzyme is specific for basic amino acids such as arginine, histidine, lysine and ornithine. Chicken liver has a related amino acid oxidase with a wider specificity. Easily the most active L-amino acid oxidase of the vertebrates is the FAD enzyme in snake venoms (p. 323). Such enzymes are also found at high activities in the hepatopancreas of many molluscs. In contrast to the enzymes of venoms and mammalian kidney, the mollusc enzymes are especially active with the basic amino acids arginine, histidine, lysine, ornithine and citrulline, but are inactive with proline. Two types of enzyme can be distinguished and these differ to some extent in their specificity patterns. The *Mytilus* type is par-

ticle bound and is also present in other molluscs like *Sepia* and *Aplysia*. The *Cardium* type is soluble and specifically activated by magnesium ions. In the freshwater snail *Lymnaea stagnalis*, a soluble amino acid is found not only in the hepatopancreas, but also in the foot tissues. This enzyme is specific for basic amino acids [82].

12.4.2 Aminotransferases (Transaminases)

For most amino acids, transfer of the amino group is both the first step in degradation and the last step in de novo synthesis. Usually the 2-oxoglutarate/glutamate pair plays the role of amino group acceptor/donor. When amino transfer to 2-oxoglutarate is tested with a large number of amino acids, aspartic acid and alanine, and often also leucine and tyrosine, are found to be by far the most active donors. The ratio of aspartate transaminase to alanine transaminase differs with the species and the tissue; in most cases, the aspartate transaminase has the higher activity and has special responsibilities for transport across the mitochondrial membrane [48]. Alanine, aspartate, tyrosine and histidine transaminases are all homologous, but agree pairwise in only 16–25 % of the 400–500 positions [115, 151, 205].

Of all the aminotransferases, the **aspartate transaminase** (AspAT or GOT) is the best known. It has been studied in detail in vertebrates, protozoans, fungi and bacteria. All vertebrates possess both mitochondrial and cytoplasmic aspartate transaminases (m-AspAT and c-AspAT), which differ in all species in about 50 % of their sequences, but have identical tertiary structures. Together with mitochondrial and cytoplasmic malate dehydrogenases, they make up the most important mechanism for introducing reduction equivalents into the mitochondria; this mechanism is the malate-aspartate shuttle. The two isoenzymes of aspartate transaminase are dimers with 45-kDa subunits of about 400 amino acids. They have been sequenced via cDNAs in various mammals, the chicken and *E. coli* [65, 186]. According to these data, the mitochondrial and cytoplasmic forms had already separated about 1000 million years ago and should, therefore, also be widespread in the invertebrates. The rate of evolution of c-AspAT appears to be somewhat higher than that of m-AspAT [68]. More detailed information on the structure and intracellular localization of the AspATs of invertebrates does not appear to be available. The enzyme of the parasitic flagellate *Trichomonas vaginalis* is unusual in that, in contrast to the mammalian enzyme, it is not highly specific for the pairs glutamate/2-oxoglutarate and aspartate/oxaloacetate; it also shows high activity with the aromatic amino acids phenylalanine, tyrosine and tryptophan [141].

Cytoplasmic and mitochoncrial isoenzymes are also known for vertebrate **alanine transaminase** (AlaAT or GPT). In the mammals, the cytoplasmic form predominates in carbohydrate catabolizing tissues, such as muscle and heart, and the mitochondrial form predominates in organs with intensive gluconeogenesis, such as the liver and kidney. It may be concluded from this that in vivo the mitochondrial form catalyses the reaction alanine to pyruvate, and the cytoplasmic form does the reverse. The only alanine transaminase to have been isolated and characterized from an invertebrate is that from *Drosophila nigromelanica* [46]. Here, the enzyme is a homodimer of 113 kDa and, apart from its approximately five-fold higher affinity for alanine (K_m = 5.2 mmol/l), corresponds in its catalytic properties to the mammalian enzymes of the rat liver, and porcine and bovine heart. The enzyme is involved in energy extraction from proline in the flight muscles; the high activities in the fat body aid proline synthesis, in that the alanine transported from the muscle to the fat body provides nitrogen for the synthesis of glutamic acid and proline from 2-oxoglutarate (Fig. 12.6). In the flight muscles of *Schistocerca gregaria*, the cytoplasmic alanine transferase is monomeric and dimeric with subunits of 63 kDa, and the mitochondrial form is dimeric and tetrameric [61].

12.4.3 Glutamate Dehydrogenases

Organs with central metabolic functions, such as the liver, kidney, hepatopancreas and fat body, have high activities of GDH, whereas the activities are lower in most muscles. High activities of GDH are found in some muscles in association with the use of proline in energy-yielding metabolism, e.g. in the flight muscles of various insects and in the mantle muscles of the cephalopod *Loligo pealeii* (see Table 15.7a, p. 582). The typical animal GDH, as isolated and characterized from the liver and kidneys of mammals, birds, amphibians and fish, and from the fat bodies or flight muscles of insects [244], is a molecule of 310–340 kDa with six identical subunits of 50–57 kDa and is localized in the mitochondrial matrix. The sequences of the approximately

500-amino-acid chains have been determined, for example, for the liver enzymes of humans, several other mammals and the chicken, and for the human brain enzyme [3, 150, 163, 191]. In the frog liver, an equilibrium between forms of 250 and 500 kDa is shifted by NADH in favour of the smaller form and by NAD^+ in the opposite direction. The enzyme of bovine liver differs from that in, e.g., rat liver in being organized into aggregates of 2 to 3 MDa.

Most animal GDHs are active with both coenzymes, although the rate of reaction with NADH is always higher than with NADPH. The **activity ratio NADH:NADPH** differs widely with the origin of the enzyme, pH and other parameters: it is less than 2 for the bovine and chicken liver enzymes, 6–8 for the two isoenzymes of the beetle *Popillia japonica* [144], and 15 for the liver enzyme of the tuna fish. Reductive amination is, as a rule, more rapid than oxidative deamination, with the **activity ratio NADH:NAD^+** varying from 2.5–6.0 for GDH from various invertebrates and the gills and kidneys of fish to over 20 for the liver GDH of various vertebrates. GDHs are not absolutely specific for glutamic acid. Whereas the bovine enzyme has only minor activity against other L-amino acids, like glutamine, norvaline and 2-aminobutyrate, the frog-liver GDH is much less specific. The enzyme from both tadpoles and adults of the bullfrog has, for example, with glutamine 38 % of the activity shown with glutamic acid. The pH optima of animal GDHs are all very similar: pH 7–8 for glutamate synthesis, and somewhat higher at pH 8–9 for glutamate catabolism.

The typical GDH is **regulated allosterically**, with ADP and AMP as activators, GTP in particular as an inhibitor, and ATP with negligible effect. NADH also inhibits under certain conditions; its effect is relieved by ADP. There are many strong interactions between allosteric effectors. One would expect that cells suffering an energy shortage would catabolize glutamate, and synthesize it during periods of surplus energy. In fact, GDH is stimulated during an energy shortage by low GTP:ADP ratios; situations of surplus energy, however, are characterized by high ratios, and GDH should be inhibited. Thus, glutamate synthesis is apparently regulated more by substrate supply than by allosteric effects [144]. In some cases, tissue-specific GDH isoenzymes with varying kinetic-regulatory properties have been described. Thus, in the Amazon fish *Arapaima gigas* and *Osteoglossum bicirrhosum*, ATP activates GDH in the kidney but inhibits the enzyme in the gills; the activation by ADP is much greater for the kidney GDH than for the gill enzyme. The biological significance of these differences is not clear. In contrast, the differences in kinetic-regulatory properties of the GDH isoenzymes from the flight muscle and fat body of the beetle *Popillia japonica* are clearly interpretable. Proline is used as an energy source in the flight muscles of this species. The glutamic acid formed in the process is deaminated by the muscle GDH, giving free ammonia; in the flight muscles of the tsetse fly *Glossina morsitans* and the potato beetle *Leptinotarsa decemlineata*, on the other hand, alanine is produced by transamination (Fig. 12.6). The 2-oxoglutarate produced from the glutamate enters the citrate cycle as malate and pyruvate, and is released as water and CO_2. The 2-oxoglutarate produced from carbohydrate in the fat body of *Popilla* is aminated by GDH to glutamate, which is then used in the de novo synthesis of proline. The muscle GDH, with a relatively low substrate affinity and marked nucleotide regulation, is specifically adapted to glutamate oxidation, and is strictly regulated by the energy status of the flight muscle. In contrast, the fat-body GDH is almost independent of the nucleotide status of the cell and is specialized for glutamate synthesis by its high affinity for 2-oxoglutarate in the presence of NADPH [144].

Some animals have GDH enzymes which deviate from the normal pattern in their location or in being specific for only one coenzyme. The GDHs from the hepatopancreas of the mussels *Mytilus edulis* and *Modiolus demissus* and from the mantle muscle of the cephalopod *Loligo pealeii* are absolutely specific for NADH/NAD^+ [212]. The *Loligo* enzyme, in contrast to all other GDHs, shows the same activity in both reaction directions. The Cnidaria and Ctenophora have GDHs that are only NADPH/$NADP^+$ active and show no nucleotide regulation but are strongly influenced by the salt concentration [70]. Two very different GDHs are found in the coral *Acropora latistella:* an NADPH-specific enzyme localized in the cytoplasm, and a mitochondrial enzyme that differs from the GDH of higher animals by its high specificity for NADH/NAD^+ and the absence of allosteric effects with ADP and GTP [70]. *Trypanosoma cruzi* contains two completely different GDHs. One is absolutely specific for NADPH/$NADP^+$ and is not affected by adenine nucleotide; it is similar in all characters to the enzyme from *Plasmodium chabaudi* and corresponds to the NADP-specific GDH of bacteria, fungi and higher plants. The second *Trypa-*

nosoma GDH, which is very unstable and was at first overlooked, is NAD-specific and is inhibited by ATP and GTP as well as by ADP and AMP.

12.4.4 Glutamine Synthetases and Glutaminases

Glutamine is an important product of ammonia detoxification; it is the storage and transport form of ammonia, as well as an amino-N donor for many biosynthetic processes. **Glutamine synthetase** catalyses the formation of glutamine from glutamic acid and ammonia with the consumption of ATP. The cytoplasmic enzyme from the liver, kidney and brain of mammals and birds consists of eight identical subunits of about 360–370 amino acids. The isoenzymes from rat liver and kidney are very different in sequence [193, 229]. Mammalian liver contains a further ammonia-detoxifying enzyme in addition to glutamine synthetase, namely, carbamylphosphate synthetase I. However, the two enzymes are not found in the same liver cells; glutamine synthetase is restricted to the hepatocytes in a small region around the terminal venules, and these contain no carbamylphosphate synthetase. This alternative expression of two enzymes is apparently something unique to the mammals; glutamine synthetase is not restricted to specific liver cells in other vertebrates [228, 256]. The glutamine synthetase in the liver of uricotelic reptiles and birds is found in the mitochondria, but releases its products into the cytoplasm where uric acid synthesis occurs [8, 38]. Mitochondrial glutamine synthetase are also found in the tortoise, the liver of which produces both urea and uric acid [37]. The glutamine synthetase in the liver and kidney of purely ureotelic elasmobranchs is also mitochondrial. In contrast to the situation in reptiles and birds, the glutamine formed in the liver mitochondria is not released into the cytoplasm but is further metabolized in situ by the available mitochondrial glutamine-dependent carbamylphosphate synthetase (CPS III). The other organs of the elasmobranchs contain a cytoplasmic glutamine synthetase which is encoded by the same gene as that for the mitochondrial enzyme [8]. *Drosophila melanogaster* also possesses both a mitochondrial (430 amino acids) and a cytoplasmic (355 amino acids) glutamine synthetase; in this case, these are encoded by different genes and show only about 60% sequence agreement [35]. The hydrolysis of glutamine is achieved in insects, as in vertebrates, not only by the hydrolytic action of mitochondrial **glutaminase I** but also by the system known as "glutaminase II", which consists of glutamine aminotransferase and 2-oxoglutarate amidase [67].

12.4.5 Production of Ammonia

Various enzymes are involved in ammonia formation in ammoniotelic animals: amino acid oxidases, glutamate dehydrogenases, serine dehydrases, glutaminases, asparaginases, ureases, AMP deaminases and adenosine deaminases. Little is known comparatively about the relative contributions of these enzymes. In the **bony fish**, the gills are much more important than the kidneys as a site of ammonia release. However, only a minor part of the ammonia released is produced in gill cells, most comes from the blood. Contrary to early assumptions, glutamine has no role in the delivery of ammonia to the gills. For example, in the rainbow trout *Salmo gairdneri*, the glutamine concentration in the blood before and after the gills is about the same (59 versus 57 μmol/l); in contrast, both glutamic acid and ammonia concentrations are lowered (28 versus 17 μmol/l and 230 versus 100 μmol/l, respectively). The ammonia in the blood comes partly from the liver and kidneys, which contain high activities of glutamate dehydrogenase and glutaminase, and partly from active muscles, in which, e.g. in the goldfish, the purine nucleotide cycle and glutamate dehydrogenase contribute equally to ammonia production [47, 253].

In the marine **crustaceans** *Praunus flexuosus* and *Crangon crangon*, ammoniogenesis is due solely to the activity of glutamate dehydrogenase [20]. In the earthworm *Lumbricus terrestris*, the crayfish *Cherax destructor* and the estuarine mussel *Corbicula japonica*, L-serine is the preferred substrate for ammoniogenesis. In *Lumbricus*, however, AMP deaminase and L-serine dehydrase are involved in ammonia production [105]. In the corals, brachiopods, molluscs and echinoderms, the ammonia released in the skeleton-forming tissues is involved in the deposition of calcium carbonate by helping to maintain the optimal pH of 8.6–8.9. In many of these animals, ammoniogenesis depends upon the activity of a very active urease, which simultaneously produces carbon dioxide; in the shell-bearing pulmonate snails, the adenosine deaminase of the mantle tissue is often more important. This enzyme is also present in non-shelled animals but does not

have the typical pH optimum of the shell-bearing forms, which are important for calcification. Ammonia formation in the tapeworm *Gyrocotyle fimbriata*, which infests cartilaginous fish, is due mainly to hydrolysis by ureases of the large quantity of urea in the gut cavity; worms maintained in urea-free saline mostly excrete amino acids, which are thus the end product of its own protein metabolism.

Ureases are widely found in the invertebrates, but only a few have been purified and characterized, e.g. those from the polychaete *Arenicola cristata*, the sea urchin *Lytechinus variegatus* and the snail *Otala lactea*. Ureases have pH optima of 8–9 and K_m values of 0.1–4.5 mmol/l. However, the enzyme from the tapeworm *Lacistorhynchus tennuis*, which infests sharks, has a K_m of 15 mmol/l, corresponding to the urea concentration of its environment (300 mmol/l). The origin of the urease substrate, urea, is in many cases obscure. For example, the coral *Acropora acuminata* has a highly active urease but neither a functioning urea cycle nor the enzymes of uricolysis [26, 82].

12.4.6 Urea Synthesis

Urea is formed during various metabolic reactions: directly from preformed arginine by arginase action; indirectly by the conversion of arginine to γ-guanidinobutyrate or agmatine by specific ureohydrolases (Fig. 12.4); from the arginine synthesized de novo in the Krebs-Henseleit cycle; and from allantoic acid by uricolysis. Urea production is therefore widely distributed in the animal kingdom, but alone is not necessarily an indication of the existence of a functioning urea cycle. Evidence for the presence of all five enzymes of the cycle is needed or, even better, tracer experiments can be used, e.g. to provide evidence for the incorporation of ^{14}C-hydrogen carbonate into urea. There is unambiguous evidence for the cycle in the liver of mammals, amphibians and elasmobranchs, as well as in various turtle species and the three genera of lungfish (Dipnoi), in which the activity decreases in the order *Protopterus* > *Lepidosiren* > *Neoceratodus*. The five enzymes have also been found in the liver of the crossopterygian *Latimeria chalumnae* and are present in about the same amount as in the elasmobranchs and Dipnoi. Equally clear is the absence of the cycle in the liver and other organs of the birds, lizards and snakes, and agnathans. It has been continuously claimed that a functioning urea cycle is present in the insects, but it has never been satisfactorily demonstrated. The Trypanosomatidae have the complete cycle only when they host symbionts [18, 36, 51, 117, 154, 269].

In some animal groups, only individual species possess a functioning urea cycle. Amongst the Platyhelminthes, tracer experiments have shown the presence of the cycle in the terrestrial turbellarian *Bipalium kewense* and the rat tapeworm *Hymenolepsis diminuta*, and its absence in *Fasciola hepatica*. In the nematodes, a low-activity urea cycle has only been convincingly demonstrated by enzyme assay and tracer methods in *Panagrellus redivivus;* the results were negative for *Ascaris, Caenorhabditis* and *Toxocara* [18]. Unambiguous evidence is available only for the pulmonate molluscs *Helix aspersa, Limax flavus, Otala lactea* and *Strophocheilus oblongus;* on the other hand, various mussel species lack carbamylphosphate synthetase. The annelid *Lumbricus terrestris* was the first invertebrate in which the whole cycle was demonstrated in 1950 [189]. A functioning urea cycle is found in the teleosts *Opsanus tau* and *O. beta* (toadfishes). In this

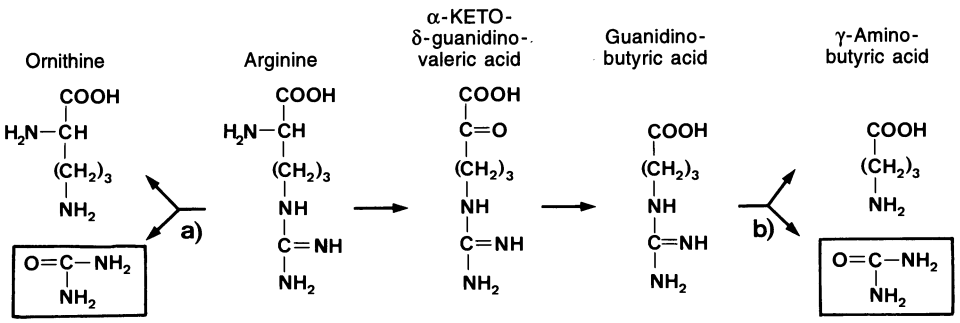


Fig. 12.4. Urea formation by arginase *(a)* and γ-guanidinobutyrate ureohydrolase. *(b)* γ-Guanidinobutyric acid is produced by the decarboxylation of α-keto-δ-guanidinovaleric acid, which itself is the result of the transamination or oxidative deamination of arginine

case, urea accumulates in the blood; it is not excreted but is hydrolysed by a urease in the gut. Thus, the toadfishes are also ammoniotelic. The air-breathing catfish *Heterpneustes fossilis* also possesses all five enzymes of the urea cycle [154]. Urea synthesis in the vertebrates is confined to the liver; the activities of the enzymes in sheep lung, heart, kidney and rumen epithelium are at least 1000-fold lower than in the liver. In the pulmonate snails, urea synthesis takes place mainly in the hepatopancreas, and in the earthworm it occurs in the gut tissues.

The five enzymes of the urea cycle (Fig. 12.5) have mainly been isolated from mammals, and characterized and sequenced via the cDNAs [117]. Information about the enzymes of the lower vertebrates and invertebrates is available only for carbamylphosphate synthetase and arginase. **Ornithine transcarbamylase** (OCT) is a mitochondrial enzyme which in humans is expressed only in the liver and gut. The native enzyme of the mammals is a trimer with subunits of 36–39 kDa; the human and rat sequences of 322 amino acids agree by 93 % [117]. Rather unexpectedly, OCT activity is found in the kidney mitochondria of certain strains of the domestic chicken, although these have no functioning urea cycle. The activity is, however, 20-fold lower than in bovine liver. The active chicken OTC is, like that in the mammals, a trimer with 36-kDa subunits, and has similar kinetic properties. A pre-OTC has also been detected in the liver, heart, brain, gut and muscles of these animals, but this is converted into an active enzyme only in the kidney. Apparently, the OTC gene in birds is not always completely expressed but it has been retained in a functional form during evolution [250]. **Argininosuccinate synthase** is present in the cytoplasm of all mammalian cells except the erythrocytes; it is a tetramer of extremely basic 46-kDa subunits which in the human and bovine forms contain 21 arginine and 31–33 lysine residues. **Argininosuccinate lyase** is expressed in all human cells including the erythrocytes, although the activity in the liver is about tenfold higher than that of other organs. The native enzyme is a homotetramer which loses activity on dissociation to dimers. The amino acid sequences of the human and rat subunits agree by 97 % [147].

Three types of **carbamylphosphate synthetase** (CPS) can be distinguished in the vertebrates. The mitochondrial CPS I requires ammonia and N-acetylglutamic acid as cofactors. This is the entry-point enzyme of the urea cycle and makes up about 20 % of the total protein of liver mitochondria. The protein is a monomer of about 160 kDa [117]. CPS II is located in the cytoplasm and requires glutamine as the amino-group donor, but no N-acetylglutamic acid. This is the initial enzyme of pyrimidine biosynthesis and is found in various tissues of ureotelic and non-ureotelic vertebrates. The CPS III of elasmobranchs is similar to CPS I in that it is localized in the mitochondria and uses N-acetylglutamic acid as the cofactor, but it is specific for glutamine. This enzyme is highly active in the livers of marine species, but is active at much lower concentrations in the freshwater ray *Potamotrygon circula-*

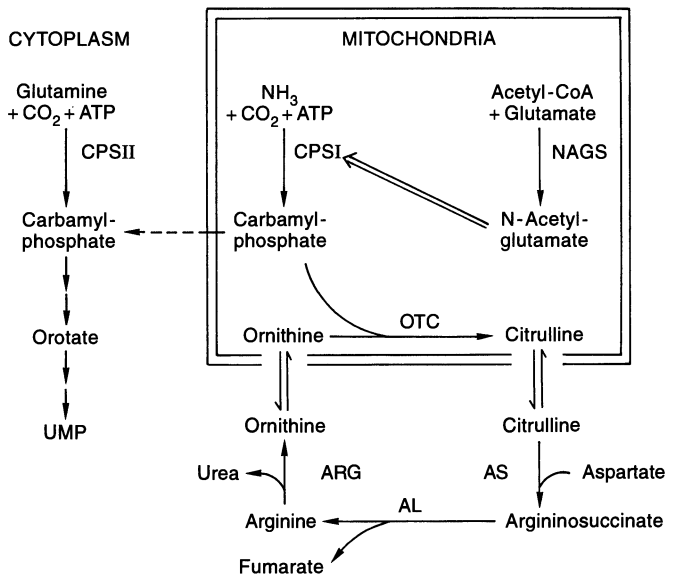


Fig. 12.5. The intracellular localization of the enzymes of the urea cycle and pyrimidine biosynthesis [117]. *CPSI* and *CPSII*, carbamylphosphate synthetases I and II; *NAGS*, N-acetylglutamate synthase; *OTC*, ornithine transcarbamylase; *ARG*, arginase; *AL*, argininosuccinate lyase; *AS*, argininosuccinate synthetase

ris. In the teleosts, CPS III has been detected in the liver of *Micropterus salmoides* and at lower activities in several other species [8]. In parallel with the CPS activity, the activity of glutamine synthase is much higher in the liver of marine elasmobranchs than in the freshwater rays and the teleosts. Attempts have been made to classify invertebrate CPSs according to substrate, dependence upon N-acetylglutamic acid, and intracellular localization, but this is not always possible. In the absence of sufficient sequence data, very little can be said about the phylogenetic relationships of the different CPSs. The earthworm *Lumbricus terrestris* and the turbellarian *Bipalium kewense* possess CPSs with the properties of types I and III, but part of the activity in the earthworm is cytosolic. CPS III-like enzymes have also been detected in various terrestrial gastropods, and CPS II-like activity has been detected in the parasitic worms *Ascaris* and *Schistosoma* as well as in some insects.

Arginases are widely distributed in the animal kingdom, although they may have grossly different activities in different species. As expected, they have the highest activity in organs that synthesize urea, but are also found at high activity in different tissues of ureotelic and non-ureotelic animals, e.g. in the kidneys of vertebrates of all classes, in the gut and epidermis of some vertebrates, in the liver of several bird species, and in some invertebrates. Arginases have many other functions in addition to their role in the urea cycle: the introduction of arginine into the citric acid cycle via ornithine, glutamic acid and 2-oxoglutarate; the furnishing of ornithine for the biosynthesis of proline and polyamines; and osmoregulatory urea formation [110]. Thus, the enzyme in the fat bodies of the Indian silk moth *Hyalophora cecropia* serves mainly in the formation of proline from arginine, whilst that in the flight muscles is involved in energy provision. There is a corresponding 20- to 30-fold increase in arginase activity during metamorphosis of this species, compared with the non-flying cockroach *Blaberus cranifera*, where the activity remains relatively low.

The liver arginases of ureotelic animals are cytosolic; in contrast, the arginases of other mammalian tissues and all the arginases of uricotelic and ammoniotelic animals are localized in the mitochondria, e.g. in the livers of birds, reptiles and teleosts [42], and in the fat body of the fly *Aldrichina grahami*. The liver arginase of the elasmobranchs is also mitochondrial, despite the fact that in these animals the synthesis of arginine from citrulline takes place in the cytoplasm. It may have something to do with the osmoregulatory role of urea in the elasmobranchs that the localization of their liver enzymes is intermediate, i.e. like the ureotelic vertebrates they have mitochondrial carbamylphosphate synthases, but like the non-ureotelic vertebrates their arginases and glutamine synthases are also mitochondrial [44]. The arginases of mammalian liver have molecular masses of 110–115 kDa with subunits of 35 kDa and probably a trimeric structure. The sequence of the human liver arginase agrees in 87 % of its 322 amino acids with the liver enzyme of the rat, and 41 % with yeast arginase. Human kidney, small intestine and lactating milk glands contain isoenzymes which differ structurally to those of the liver [100, 133]. Various avian and reptilian arginases are reported to have molecular masses of 280 kDa, and that of the liver enzyme of the clawed frog *Xenopus laevis* is 76 kDa with 18-kDa subunits. The enzymes of the earthworm *Pheretima communissima* and the sea snail *Concholepas concholepas* are monomers of 25–27.5 kDa [43, 114]. It was assumed for a long time that liver arginases occurred only in the livers of ureotelic vertebrates ("Clementis rule", 1937). After the discovery of liver arginases in the birds and lizards, Mora proposed in 1965 that a distinction be drawn between ureotelic and uricotelic liver arginases, of which the ureotelic are lower in molecular weight, higher in substrate affinities and have higher activities against the arginine analogue L-canavanine (Fig. 12.2). However, this concept is not completely justified. It is true that the liver arginases of ureotelic and uricotelic animals show no immunological cross-reactivity, but the enzymes isolated from the liver and kidney of the chicken and from the liver, kidney and brain of the lizard *Calotes versicolor* agree in many of their properties with the arginases of ureotelic animals. In particular, it is not true that uricotelic arginases, with K_m values of 100–200 mmol/l, always have lower substrate affinities than the ureotelic variety (K_m 10–20 mmol/l); in fact, there are also arginases with relatively low K_m values in non-ureotelic animals, e.g. in the lizard *Calotes*, the frog *Xenopus*, the teleost *Genypterus*, the molluscs *Chiton* and *Concholepas*, and the beetle *Tribolium* [42, 43]. Most arginases are activated by Mn^{2+}, Ni^{2+} or Co^{2+} and inhibited by Mg^{2+}, Cd^{2+} or Zn^{2+} [43, 114]. The pH optimum in the presence of Mn^{2+} ions is always about 9–10, but is only pH 7 in the presence of Co^{2+} or Ni^{2+}. Almost all arginases are inhibited by their substrates as well as by

ornithine, lysine, proline and branched-chain amino acids (valine, isoleucine, leucine); the type of inhibition (competitive or non-competitive) varies with the species [42].

In animals, there is a second **ureohydrolase** which hydrolyses γ-guanidinobutyric acid to γ-aminobutyric acid and urea, but is inactive against arginine. The γ-guanidinobutyric acid arises by the decarboxylation of α-keto-β-guanidinovaleric acid, which itself is produced by the transamination (mammalian liver) or oxidative deamination (avian liver, hepatopancreas of marine molluscs and crustaceans) of arginine (Fig. 12.4). The γ-guanidinobutyrate hydrolase has been isolated from the liver and kidney of the ray *Raja clavata*, and has been separated from the arginases. Like arginase, it is activated by Mn^{2+} and has a pH optimum of 9.5. The enzyme is widely distributed, having been reported in many gastropods, bivalves and crayfish. It is apparently absent from cephalopods, marine crustaceans and the scorpion *Androctonus*. Out of ten investigated mammalian species, only the pig and rabbit had activity in the liver; all the bird species examined, except the pigeon, had the enzyme in the liver and some species also possessed it in the kidney. In reptiles and amphibians, as well as in bony and cartilaginous fish, the enzyme is detectable in both liver and kidney.

12.5 Metabolism of Individual Amino Acids

12.5.1 Proline

The most interesting comparative biochemical aspect of proline metabolism lies in its importance for **energy metabolism in the insects.** According to the species, insect flight muscles receive their energy supplies from carbohydrates, triacylglycerol or proline (p. 581). The haemolymph of most insects has a high proline content (Table 12.1), the highest known being that of the tsetse fly *Glossina morsitans*, which at rest has as much as 170 mmol/l. High proline concentrations in the flight muscles, its reduction during flight, and the preferred oxidation of proline in the sarcosomes is found in particular in the blood-sucking dipterans *Glossina morsitans* and *Stomoxys calcitrans*, as well as in the beetles *Leptinotarsa decemlineata* and *Popillia japonica* [46, 260]. The pathway of proline degradation was first determined in the flight muscles of *Glossina* using specifically labelled substrates, but applies to all proline-oxidizing muscles (Fig. 12.6). Proline is converted to glutamic acid via Δ^1-pyrroline-5-carboxylate and glutamic acid semialdehyde; this is then deaminated to 2-

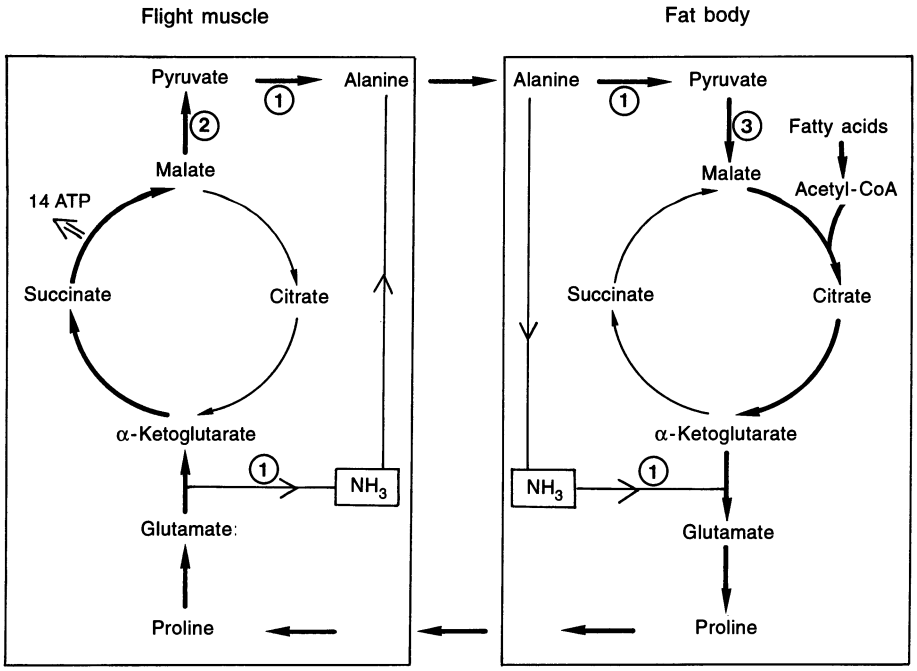


Fig. 12.6. Energy-yielding proline catabolism in the flight muscles, and proline synthesis in the fat body of the potato beetle *Leptinotarsa decemlineata* [261]. *1*, Alanine transaminase; *2*, NAD-specific malic enzyme; *3*, NADP-specific malic enzyme

oxoglutarate and introduced into the citric acid cycle. The amino group of glutamic acid is mostly transferred by an alanine transaminase to give pyruvate. As alanine formation can also take place in isolated mitochondria, pyruvate cannot arise from the cytosolic glycolysis system; it is mainly produced from malate by the NAD-dependent malic enzyme which is particularly active in the mitochondria (see Table 15.7a (4), p. 582). The regulation of malate metabolism is not yet clear; malate dehydrogenase, e.g. in the mitochondria of *Leptinotarsa*, has a tenfold higher activity than malic enzyme, which in addition is inhibited by ATP and pyruvate during proline oxidation [260]. There are species-specific differences in the fate of the amino nitrogen released from glutamic acid: in the flight muscles of *Leptinotarsa* just as much alanine is produced as proline is degraded; in *Glossina* about 20 % is released as ammonia by the action of glutamate dehydrogenase, and in *Popillia* it is 50 % [260].

The proline content of the haemolymph is kept constant through the synthesis of proline in fat bodies. In *Glossina* and *Leptinotarsa*, proline synthesis in the fat bodies makes use in particular of triacylglycerol and the alanine released by muscle; less so of glutamic acid, aspartic acid or leucine; and hardly at all of glucose. Thus, the system muscle–haemolymph–fat body allows the use of the fatty acids of the fat bodies for the energy metabolism of the flight muscles. A comparison of the enzyme constitution of *Glossina* fat bodies with that of the fat bodies of other flies clearly illustrates the specific adaptation towards proline as a substrate for energy metabolism [126, 261]. In a similar fashion, proline is oxidized in preference to other substrates in isolated sarcosomes from the leg muscle of the dung beetle *Heliocopris dilloni*. The enzyme constitution here and in other Scarabaeidae also shows that proline is the preferred energy source for the leg muscles. Tracer experiments with the fly *Aldrichina* have shown that only arginine is used for proline biosynthesis by the larvae, whilst glutamic acid is used more than arginine in the adults. The rate-limiting enzyme of proline synthesis from arginine is ornithine-δ-transaminase, which converts ornithine to glutamic acid semialdehyde and is much more active in the larvae than in the adults. As in the rat, this enzyme is mitochondrial, and its activity decreases in the order fat body > gut > Malpighian tubules > muscles.

The **enzymes of proline metabolism** have not been completely described, even for the mammals. However, it is known that the insects have a mitochondrial proline oxidase, which uses atmospheric oxygen to oxidizes proline to Δ^1-pyrroline-5-carboxylate; they also have a particle-bound Δ^1-pyrroline-5-carboxylate reductase, which reduces Δ^1-5-carboxylate to proline. The only enzyme of proline metabolism to have been isolated from invertebrates is the Δ^1-pyrroline-5-carboxylate reductase from *Drosophila melanogaster;* the main difference between this and the rat enzyme is a higher activity with NADPH than with NADH [88].

12.5.2 Sulphur-Containing Amino Acids

Of the two sulphur-containing standard amino acids, methionine is always essential, whereas cysteine can be synthesized by most animals via the **cystathionine pathway** (Fig. 12.7) [57]. In this case, the sulphur is obtained from methionine and the C skeleton from serine (trans-sulphation). ^{14}C-serine, or its precursor ^{14}C-formiate, is incorporated into cysteine in the sawfly *Acantholyda nemoralis* but only into cystathionine in the stick insect *Carausius morosus*. This suggests that the sawfly has a functioning cystathionine pathway, whereas in the stick insect the pathway is blocked at the cystathionine step. A fully functional cystathionine pathway is also present in several dipterans and lepidopterans, but is missing in other lepidopterans and some aphids [224]. During human embryo development, there is no active cystathionase and cysteine is therefore required [57]. Mammals can resynthesize methionine from homocysteine; the required methyl-group donor varies with the species: in the rat it is mainly betaine (trimethylglycine), and in sheep 5-methyltetrahydrofolate [275].

Cysteine is the starting point for the synthesis of glutathione, coenzyme A and taurine. The **synthesis of taurine from cysteine** in animals can follow one of several routes, e.g. via cysteamine, cysteine sulphinic acid or cysteinic acid (Fig. 12.8). Cysteamine does not apparently arise simply by the decarboxylation of cysteine but, in mammals, is formed mainly indirectly via coenzyme A [57]. The formation of cysteamine from lanthionine is possible in mammals but, in view of the low concentration of the latter amino acid, is of little importance. Some insects have large amounts of lanthionine but the synthesis of cysteamine in these cases has not been investigated [57]. Hypotaurine, the intermediate of taurine synthesis via cysteamine and cysteine sulphinic acid, has been detected in Porifera, Cnidaria,

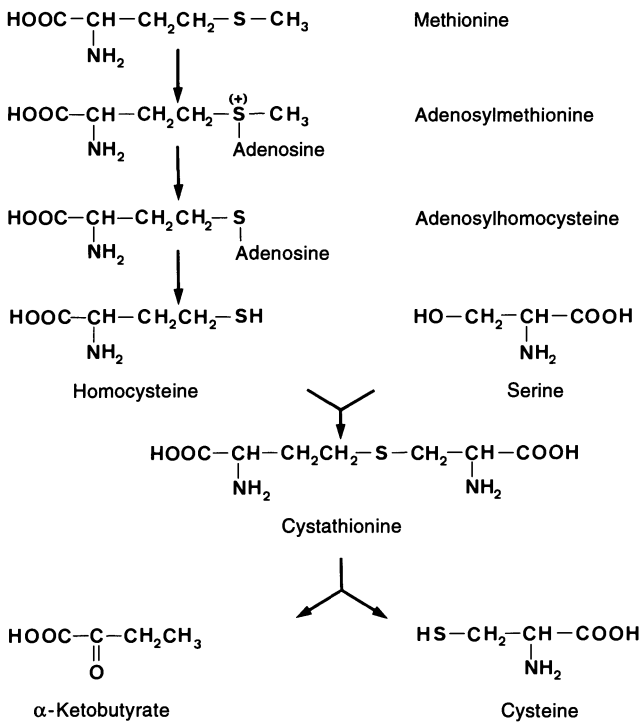


Fig. 12.7. Biosynthesis of cysteine from methionine and serine by trans-sulphation in the cystathionine pathway. The sulphur of the cysteine comes from methionine and the carbon skeleton comes from serine

molluscs, annelids and arthropods; its concentration is always lower than that of taurine itself. Many insect larvae contain large amounts of taurine. In insects with carnivorous larvae, e.g. the blow-flies *Phormia* or *Calliphora*, the taurine comes from the nutrients. In caterpillars of the butterfly *Mamestra configurata*, whose vegetable diet contains no taurine, the taurine is synthesized via the cysteamine pathway; up to 90% is stored in the newly formed flight muscles during metamorphosis. The cysteamine pathway also predominates in the mammals and in the edible mussel *Mytilus*. The cysteine sulphinic acid pathway is a subsidiary pathway in some mammalian species but forms the main route in several insect species, e.g. *Musca domestica* and *Blattella germanica*. It is not clear, however, whether the differences observed among insects are species- or stage-specific. The synthesis of taurine from cysteine sulphinic acid has also been observed, by use of tracer experiments, in isolated abdominal fat and the tail muscle of the lobster *Homarus americanus*. **Isethionic acid**, which is present at high concentrations particularly in nerve cells, may also be synthesized from taurine, but this has not yet been clearly demonstrated in either mam-

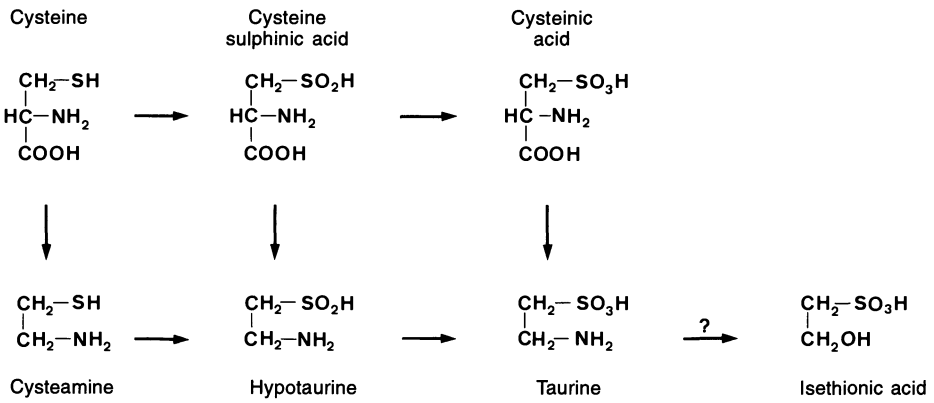


Fig. 12.8. Biosynthesis of taurine and isethionic acid. Taurine can be produced via cysteamine, cysteine sulphinic acid or cysteinic acid; the conversion of taurine to isethionic acid has not been definitively shown

mals or invertebrates. In the cephalopod *Loligo pealei*, the sulphur from labelled cysteine is incorporated into isethionic acid but the carbon atoms are not [112].

12.5.3 Serine

The degradation of L-serine can occur by two routes. L-serine is either converted directly to pyruvate by the action of L-serine dehydratase (SDH), or hydroxypyruvate is produced by L-serine: pyruvate aminotransferase (SPAT) and then introduced into glycolysis via D-glycerate and 2-phosphoglycerate [231]. The relative activities of these two pathways in vertebrates vary with the species and the organ involved. SDH activity in mammals is confined to the liver; it is present at a low level in the liver of the turtle *Testudo graeca* but is lacking in *Emys europaea*, various amphibians and the goldfish. The sequence of mammalian SDH has some similarities to microbial threonine dehydratase, and this suggests a common origin. Compared with human SDH (328 amino acids), the homologous rat enzyme is extended by an insertion of 36 amino acids between positions 64 and 65 [170]. SPAT is found in mammalian kidney and in the liver. The liver of carnivorous mammals has a consistently higher activity (1600–4200 mU/g) than that of omnivorous or herbivorous animals (always less than 390 mU/g). The activity in the carnivorous turtle *Emys* is 180 mU/g, in the herbivorous *Testudo* only 50 mU/g, and in various amphibians and the goldfish about 800 mU/g. The SPAT activity of the kidneys is also correlated to the feeding habit. The mitochondria and the peroxisomes of rat liver contain different SPATs, the mRNAs of which arise from the same gene by alternative splicing [169].

12.5.4 Tryptophan

This amino acid has the most complicated **spectrum of metabolism** (Fig. 12.9). In mammals, the main pathway concerns the oxidative opening of the ring of 3-hydroxyanthranilic acid to give the unstable 2-amino-3-carboxy-muconic acid semialdehyde, and from this compound, on the one hand, to yield acetoacetyl-CoA and CO_2 and, on the other hand, to give nicotinamide. Other metabolites, e.g. 5-hydroxyindole acetic acid, are found only in trace amounts, and part of this represents metabolism by gut bacteria. Insects are not able to cleave the benzene ring of 3-hydroxyanthranilic acid and completely degrade tryptophan to CO_2. This is shown by the accumulation of intermediates that occur before the cleavage reaction, the fact that radioactive CO_2 is not produced from ^{14}C-tryptophan labelled in the benzene ring and, in contrast to mammals, the absence of any sparing effect on nicotinic acid by tryptophan in the diet. In the insects, tryptophan metabolism from 3-hydroxykynurenine leads mainly to ommochromes, a group of brown, red or violet pigments which occur in the eye and epidermis, and also are found in the excrement (p. 748). We can only speculate about the evolutionary significance of this fundamental difference in tryptophan metabolism between insects and mammals. Tryptophan is in fact one of the rarer of the 20 protein amino acids, making up less than 1 mol%; an excess disturbs development and can be carcinogenic. It is feasible that ommochrome synthesis is primarily a mechanism for detoxifying tryptophan, and has evolved in those animals that have lost the ability to completely degrade tryptophan [137].

Over and above comparative biochemistry, great interest in the biosynthesis of ommochrome has developed owing to the existence of enzyme-deficient mutants with which the mechanisms of gene action can be studied. The investigations of Kühn on the flour moth *Ephestia* and those of Beadle and Ephrussi on *Drosophila* are among the fundamental studies which, at the beginning of the 1930s, led to the formulation of the "one gene – one enzyme" hypothesis. Thus, the tryptophan metabolism of the insects has been relatively thoroughly studied, but otherwise there is very little in the way of comparative biochemistry. The formation of 3-hydroxyanthranilic acid has been observed in tracer studies with all animals examined, but oxidative degradation of the benzene ring is found only in vertebrates of all classes, the snail *Otala* sp., and the oligochaete *Eudrilus engineae;* it does not occur in the crayfish *Cambarus virilis* and the insects. However, the first three enzymes of oxidative degradation (3-hydroxyanthranilic acid-3,4-dioxygenase, 2-amino-3-carboxymuconic acid semialdehyde decarboxylase and 2-aminomuconic acid semialdehyde dehydrogenase) have been detected in the liver and kidneys only of vertebrates and not in the invertebrates, although one must presume their existence in *Otala* and *Eudrilus*. As well as in various arthropods, ommochrome is found in cnidarians, polychaetes, echiurida and cephalopods.

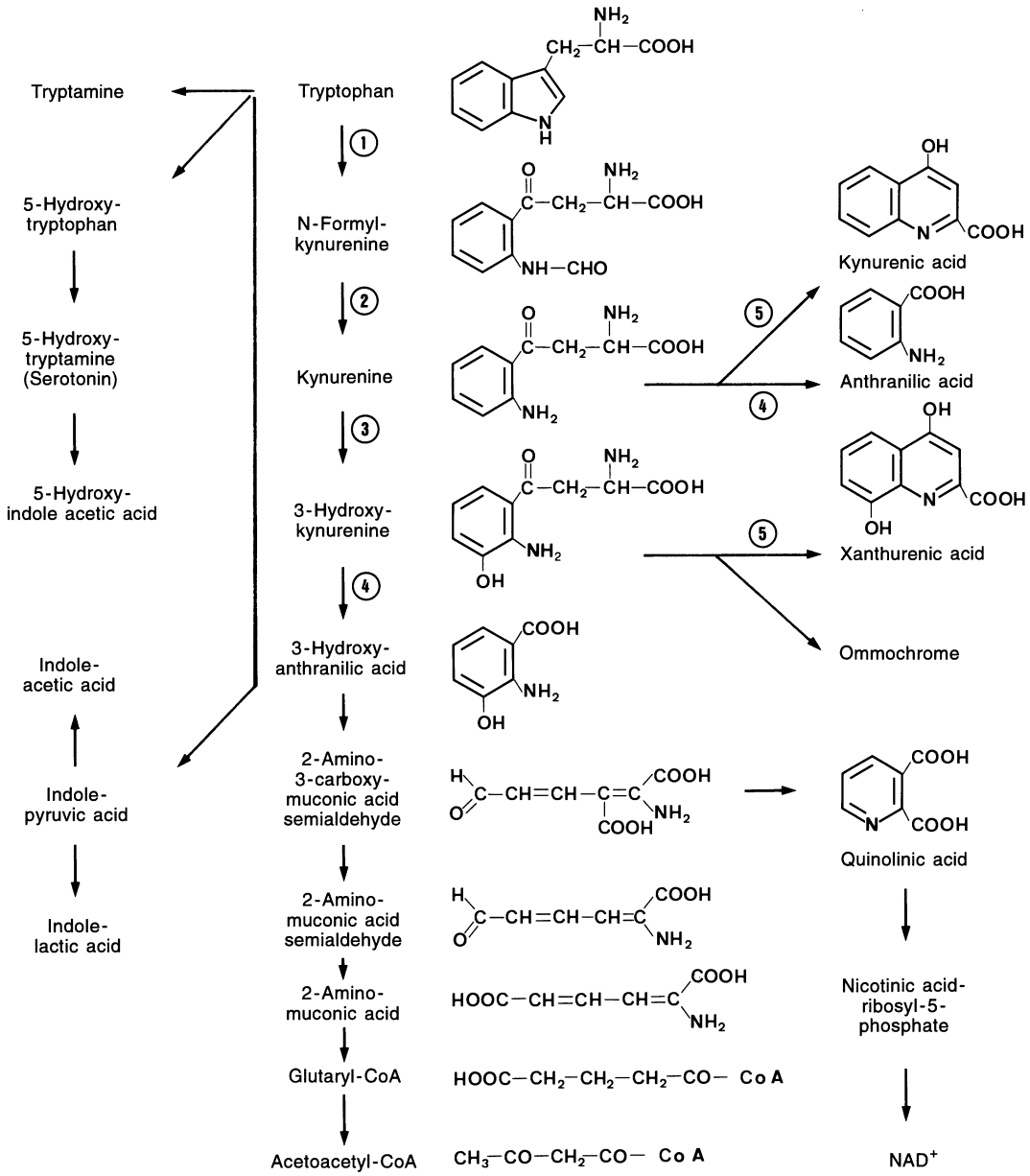


Fig. 12.9. Tryptophan metabolism. *1*, Tryptophan-2,2-dioxygenase; *2*, kynurenine formamidase; *3*, kynurenine-3-monooxygenase (kynurenine-3-hydroxylase); *4*, kynureninase; *5*, kynurenine transaminase. See text for further details

Tryptophan degradation in some *Trypanosoma* species is completely different from that of the multicellular animals. The end product in *T. lewisi* and *T. musculi* is exclusively indole lactic acid, whilst *T. brucei gambiense* and *T. equiperdum*, which belong to another subgenus, also produce indole propionic acid, indole acetic acid and indole ethanol. Here, tryptophan metabolism is apparently involved in regulating the redox state; the appearance of oxidized end products may be related to the fact that *T. brucei gambiense* and *T. equiperdum*, in contrast to the other species, have no cytochrome-dependent electron transport system but reoxidize NADH by means of an α-glycerophosphate oxidase.

A large number of **tryptophan metabolites** are known from mammals and insects, and only the most important will be dealt with here. Eye-colour mutants of insects often accumulate intermediary products of tryptophan metabolism; in the wild type these are absent or present only in trace amounts [79]. The number of end products can be increased still further by secondary metabolism, in the case of mammals by the production

of β-glucuronides and methylated and acetylated derivatives, and in insects by the formation of β-glucosides, sulphate esters and amino acid conjugates [137]. 3-Hydroxykynurenine is stored by some insects: for example, in *Drosophila* in the Malpighian tubules of the larva, which are not destroyed during metamorphosis; in *Bombyx mori* in the oocytes; and in *Drosophila* in the yolk spheres of the eggs. In all cases, these probably represent reserves for the later synthesis of ommochromes. This substance has an interesting fate in the parasitic wasp *Habrobracon juglandis:* the 3-hydroxykynurenine taken from the host is metabolized in the larvae to xanthurenic acid and is excreted; only the 3-hydroxykynurenine synthesized by the pupa is used for ommochrome synthesis. Mutants without kynurenine-3-monooxygenase activity therefore produce no ommochrome, even when they are rich in 3-hydroxykynurenine during the larval period [137]. 3-Hydroxykynurenine has also been detected in gorgonian corals of the genus *Eunicella* as well as in the urine of the cephalopod *Octopus*.

From kynurenine and 3-hydroxykynurenine, ring closure produces kynurenic acid and xanthurenic acid, which are found in trace amounts in many insects [137]. The stick insect *Carausius morosus* produces only kynurenine, which makes up about 0.1% of the faecal dry weight [137, 202]. In the presence of the relatively unspecific kynureninase, e.g. in *Bombyx mori*, anthranilic acid and 3-hydroxyanthranilic acid may be produced, and in some cases their glycine conjugates and β-glucosides may also be produced. Neither the enzyme nor its products were found in the dipterans examined [137].

In the vertebrates, tryptophan metabolism occurs primarily in the liver, whereas insects have relatively high activities of the relevant enzymes in the eyes, the epidermis and other organs. In the honey bee *Apis*, tryptophan metabolism takes place only in the eye; eyeless mutants accumulate tryptophan. A change in the concentrations of metabolites and in the activities of the enzymes of tryptophan metabolism in the course of development is often correlated to the programme of ommochrome synthesis. During metamorphosis, the body of a holometabolic insect is a closed system, for which a complete metabolite balance sheet can be drawn up. Thus, in the case of *Bombyx mori*, about 20% of the tryptophan consumed is converted to kynurenine, more than 50% to free and conjugated 3-hydroxykynurenine, about 20% to ommochromes, and traces to xanthurenic acid; in contrast, in *Phormia* 27% goes to kynurenine, 13% to 3-hydroxykynurenine and 60% to xanthommatin [137].

The first enzyme of the kynurenine pathway is **tryptophan-2,3-dioxygenase**, known previously as tryptophan pyrrolase (Fig. 12.9). In the mammals, this is a copper-containing haem protein; the prosthetic haem can be removed by dialysis and the apoenzyme reactivated with haematin. In the fly *Protophormia terrae-novae*, the coenzyme cannot be cleaved by dialysis, and in *Drosophila* the dialysed enzyme cannot be reactivated with haematin; this raises the question whether the insect enzyme is a haem protein. However, the partially purified enzyme from *Protophormia* shows the light-reversible CO inhibition typical of haem enzymes. The enzymes of insects, bacteria and mammals all appear to be homotetrameric with allosteric properties. Tryptophan can bind both to the catalytic centre and to the allosteric binding site; this increases the oxygen affinity of the enzyme. The enzyme activity at constant oxygen partial pressure depends upon the tryptophan concentration and shows marked sigmoid kinetics; in this way, tryptophan as an essential amino acid is metabolized only when in excess.

The next enzyme of the kynurenine pathway, **kynurenine formamidase**, has its highest activities in insects in the fat bodies, Malpighian tubules and gonads. The pH optima and the Michaelis constants vary greatly with the species [137]. **Kynurenine-3-monooxygenase** (kynurenine-3-hydrolase) is inactivated by light and was for a long time difficult to detect. As in the case of tryptophan-2,3-dioxygenase, this enzyme is also concentrated in different organs in different species: in *Bombyx mori* and *Calliphora* it occurs especially in the Malpighian tubules, in *Schistocera* in the eyes and integument, in *Apis* only in the eyes, and in *Ephestia* in the ovaries and the eggs. In insects, bacteria and mammals, the enzyme is localized in the outer mitochondrial membrane [137]. The enzyme from the eggs of *Ephestia kühniella*, which has been examined in detail, is active only with NADPH and not with NADH; this is in contrast to the mammalian enzyme. The affinities for substrate and cosubstrate of the enzyme from the eggs of the flour moth are about ten fold lower than those of the liver enzyme; in the stick insect, the affinities of the egg and liver enzymes are approximately the same. **Kynureninase** and **kynurenine transaminase** can also metabolize 3-hydroxykynurenine. In mammals, kynureninase catalyses the conver-

sion of 3-hydroxykynurenine to 3-hydroxyanthranilic acid in the main pathway of tryptophan catabolism, and in *Bombyx* this enzyme brings about the formation of anthranilic acid and its conjugates from kynurenine [137]. Kynurenic acid and xanthurenic acid are produced by amino transfer from kynurenine or 3-hydroxykynurenine to 2-oxoglutarate, in proportions that depend upon the species. The relevant enzymes of *Schistocerca gregaria* and *Musca domestica* have been examined in some detail.

12.5.5 Iodamino Acids

The synthesis of the thyroid hormone involves the iodination of the tyrosine residues of a specific protein, **thyroglobulin**, through the activity of a peroxidase system. The initial products are mono- and diiodotyrosine residues (MIT and DIT), which give rise to a tetraiodothyronine (thyroxin, T_4) by the combination of two DITs, and to a triiodothyronine residue (T_3) from the combination of a DIT and an MIT (Fig. 12.10). The prethyroglobulin of about 2750 amino acids, synthesized in the thyroid gland cells, is greatly modified in the Golgi apparatus and released into the gland lumen as a homodimeric glycoprotein of 660 kDa. Only then is a fraction of the tyrosine residues iodinated by the action of a membrane-bound peroxidase and linked to give iodothyronine [125]. The iodinated thyroglobulin is returned to the gland cells by endocytosis and is cleaved at specific peptide bonds in the lysosomes. Although it behaves initially like a typical secretory protein, the dimeric molecule carries two mannose-6-phosphates; this is the usual signal for transport into the lysosomes [108]. The iodinated amino acids are released and the rest of the thyroglobulin is degraded. Free MIT and DIT are mainly converted back to tyrosine in the gland cells by a microsomal iodotyrosine deiodase; T_4 and T_3 are released into the blood.

In the process of hormone synthesis from thyroglobulin, an iodophenoxyl group is transferred from a donor to an acceptor iodotyrosine residue; the acceptor becomes a hormone molecule and the donor is converted to a dehydroalanine residue. Only 25–30 of the approximately 140 tyrosine residues of the thyroglobulin dimer are iodinated, and only a few of them take part in hormonogenesis, which is highly efficient even at very low iodine concentrations. The human, rat and bovine acceptors have been identified as residues 5, 2555, 2569 and 2748, which lie at the ends of the chain; in the rabbit, residue 1291 is also involved. Residues 5, 926, 986 (or 1008) and 1375 can function as donors. Tyrosine 5 has the highest iodine affinity of all these residues and is iodinated first [172, 177]. The thyroglobulin gene of the rat has a length of 170 kb; the 9-kb coding sequence is distributed over 42 exons, of which 40 have very similar sizes of 150–200 bp [160]. About 70 % of the sequence is made up of three different types of repeats. The non-repetitive part of the polypeptide chain shows significant agreement with the sequence of acetylcholine esterase from the ray *Torpedo californica;* this amounts to up to 64 % in some regions [145, 242]. Many of the introns lie on the borders of repeats, as would be expected if the repeats arose by segment duplication. Other introns lie in the middle of repeat sequences at positions which in other repeats are interrupted by insertions. This may have arisen by "exonization" of parts of the introns. The thyroglobulin gene thus offers an interesting case study of intron evolution [178].

Thyroglobulins are found in all vertebrates down to the agnathans. In most cases they have sedimentation constants of 19S, corresponding to a size of 660 kDa, as for the dimeric mammalian protein. However, the ammocoetic larvae of the

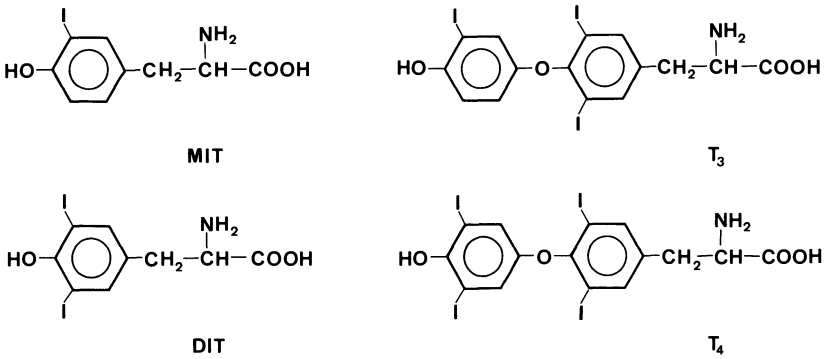


Fig. 12.10. Iodamino acids. *MIT*, monoiodotyrosine; *DIT*, diiodotyrosine; T_3, 3′,3,5,-triiodothyronine; T_4, thyroxin (3′,5′,3,5-tetraiodothyronine)

agnathans have a 3–8S form of about 160 kDa which is replaced during metamorphosis by a 12S form (in *Petromyzon fluviatilis* and *Entosphenus tridentatus)* or a 19S form (in *Lampetra reissneri)*. Iodoproteins isolated from the thyroid of the hagfish *Eptatretus burgeri* differ from thyroglobulin in many properties but can also form hormone, albeit at a lower efficiency. They contain 36–54 % carbohydrate (thyroglobulin <10 %) and up to 5.9 % iodine (thyroglobulin <1 %), of which, however, only 8–17 % (thyroglobulin 20–40 %) is associated with hormone residues [171].

The **iodination of thyroglobulin-like proteins** has also been observed in the endostyle tissue of the tunicates and acranians; this tissue is homologous to the thyroid. The production of T_4 and T_3 has been confirmed in the acranians but is still controversial in the tunicates. MIT and DIT are also produced in the tunica of ascidians. However, it should be borne in mind that MIT and DIT can arise in vitro in an oxidizing medium through non-specific reactions with iodide ions; this occurs especially in tyrosine-rich scleroproteins. Such processes have been observed in vivo in representatives of practically all animal phyla. The system of quinones and oxidases, as found in sclerotizing structures, are particularly active in this respect. However, T_4 and T_3 are produced only in trace amounts, and these processes are fundamentally different from those which occur in the thyroid. Iodinated, but also brominated and chlorinated, tyrosine derivatives are especially common in marine organisms.

More than 99 % of the thyroxin (T_4) in the blood is bound to plasma proteins. The most important human **T_4-binding protein** is the thyroxin-binding globulin (TBG), a glycoprotein of 54 kDa with 20 % carbohydrate. The T_4-binding pre-albumin (transthyretin) is also important. Human TBG shows significant similarity in its sequence of 395 amino acids to the protease inhibitors α_1-antichymotrypsin and α_1-antitrypsin but not to transthyretin [81]. TBG appears to be the most important T_4-binding protein of the mammals, but it is apparently absent from some mammals and from birds and reptiles. In the latter cases, the iodamino acids are bound to albumin and pre-albumin. The blood plasma iodamino acids are deiodinated in different tissues. The 3,5,3′-T_3 produced by cleavage of the iodine of T_4 appears to be the actual hormonally active substance, as the hormone receptors in the cell nucleus have significantly higher affinities for T_3 than for T_4. The T_4 **deiodination** takes place mainly in the liver, but also occurs in the kidneys and other tissues. The mammals also produce the isomer 3,3′,5′-T_3 (reverse T_3), which has no hormone activity, as well as several diiodothyronines. In fish, on the other hand, only 3,5,3′-T_3 has been observed as the product of deiodination of T_4 in different tissues. In the anurans, the deiodination of T_4 greatly increases in intensity during metamorphosis. Deiodination of DIT and MIT takes place in the liver of all classes of vertebrates, and in urodelans and teleosts it also occurs in the kidneys [72]. The snakes lack the required enzymes and their blood plasma contains correspondingly higher concentrations of MIT and DIT [268]. The elimination of the iodamino acids is mainly biliary, in the form of β-glucuronides or sulphate esters, and occurs in part after deiodination to diiodothyronine.

12.6 Aromatic Amino Acids and the Sclerotization of Insect Cuticulae

The diphenols (catechols) produced during the metabolism of the aromatic amino acids phenylalanine and tyrosine have important functions as neurotransmitters and neuromodulators, as precursors of the widely found pigment class of the melanins, and as cross-linking structures for the sclerotization of insect cuticulae and the tanning of various proteins. The insects have received particular attention in this respect. The phenylalanine ingested by insects is rapidly hydroxylated to tyrosine. The **phenylalanine hydroxylase** responsible for this process in insects, as well as in mammals, is an iron-containing, multifunctional oxygenase which requires a reduced pterin cofactor and atmospheric oxygen. As shown by the cDNA and gene sequences, phenylalanine hydroxylase belongs to the same protein family as the corresponding enzymes that act on tyrosine and typtophan [32, 135]. In mammals, excess tyrosine is deaminated by **tyrosine transaminase** to give *p*-hydroxyphenylpyruvate; this is completely catabolized by the opening of the benzene ring with homogentisinic acid as an intermediate. This transaminase is also present in insects. The enzyme isolated from *Drosophila hydei* is very similar to the mammalian enzyme. The tyrosine metabolism of the blowfly *Calliphora* switches over shortly before pupation. Instead of the predominant transamination to *p*-hydroxyphenylpyruvate, hydroxylation and decarboxylation results in dihydroxyphenylalanine (dopa) as the

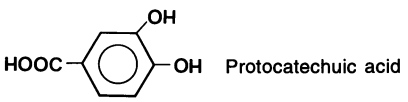
precursor of sclerotizing substances. This switch in tyrosine metabolism is controlled by the hormone ecdysone. The peptide hormone bursicon is also involved in the triggering of the sclerotization process in the adult animal [128].

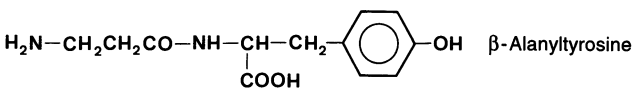
12.6.1 Sclerotizing Substances

The sclerotization of the cuticula after each moult requires large reserves of aromatic amino acids in the larva. There are, however, limits to the solubility of these amino acids, and thus phenylalanine and tyrosine are converted to more soluble compounds for **storage in the haemolymph or in the tissues**. At the same time, they are protected from attack by oxidases and other enzymes, and cause less disturbance to osmoregulation and the transport of amino acids through the plasma membrane. The Lepidoptera and the Mediterranean fruit fly *Ceratitis capitata* produce tyrosine-O-β-glucoside which is tenfold more soluble in water than the free amino acid. Most *Drosophila* species store tyrosine-O-phosphate but *D. busckii* accumulates tyrosine-β-glucoside [128, 192].

The aromatic amino acids can also be stored in the form of dipeptides or polypeptides. In the larvae of the blowfly *Sarcophage bullata*, 70–90 % of the non-protein-bound tyrosine is in the form of the dipeptide **β-alanylthyrosine** ("sarcophagine") (Fig. 12.11), which is more soluble than either free tyrosine or α-alanyltyrosine. The β-alanine required for the synthesis of this dipeptide comes partly from the catabolism of pyrimidines and partly from the decarboxylation of aspartic acid. β-Alanyltyrosine reaches its maximum concentration of 21 μmol/g fresh weight in *Sarcophaga* shortly before formation of the pupal case (puparium). Its synthesis in the fat body involves a special synthetase which requires ATP and Mg^{2+}. The fat body also contains a special peptidase, which is induced by 20-hydroxyecdysone and reaches its maximum activity at the start of sclerotization in the newly formed puparium. The peptidase present in the integument reaches its maximum 1 day later and is probably a different enzyme. Both components of the dipeptide are incorporated into the cuticula. β-Alanyltyrosine is found in all species of the genus *Sarcophaga* and in only one of the tachinid species examined, *Voria ruralis*. It is not present in other dipterans [27]. The dipeptide **γ-glutamylphenylalanine**, found in the housefly *Musca domestica*, has similar functions; it reaches a maximum concentration of 14 μmol/g fresh weight and makes up about 80 % of the non-protein-bound phenylalanine (Fig. 12.11). It is synthesized by γ-glutamyl transpeptidase, and only the phenylalanine portion is incorporated into the cuticula. This dipeptide is found in all species of the genus *Musca* and in *Stomoxys calcitrans* but does not occur in other dipterans [27]. In connection with sclerotization, it should be recalled that the larval haemolymph proteins, discussed in Chapter 5, are particularly rich in aromatic amino acids, hence the name **arylphorins**. However, they are apparently not simply a source of free phenylalanine or tyrosine; immunological investigations of the lepidopteran *Manduca* and the fly *Calliphora* suggest that arylphorins are directly incorporated unmodified into cuticula structures [128].

In insects, tyrosine is the starting point for numerous **metabolic reactions** which involve, for example, hydroxylation of the ring or side-chain, transamination, and decarboxylation or desaturation of the side-chain. The first step in the formation of sclerotizing substances is the hydroxyla-





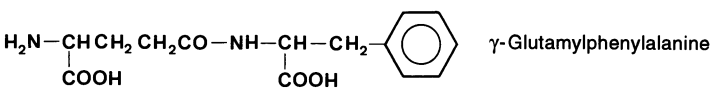


Fig. 12.11. In the quinone tanning of cuticular and non-cuticular proteins of insects the following substances are involved in addition to N-acetyldopamine (Fig. 12.12): 3,4-dihydroxybenzoic acid (protocatechuic acid) in cockroaches, β-alanyltyrosine in dipterans of the genus *Sarcophage*, and γ-glutamylphenylalanine in dipterans of the genus *Musca*

tion of tyrosine to dihydroxyphenylalanine (dopa). This reaction in insects, unlike the situation in mammals and molluscs [113], does not involve a tyrosine hydroxylase, but is catalysed by tyrosinase or phenoloxidase; it is not clear in that case why dopa is not further oxidized to *o*-quinone by the same enzyme [4, 128]. Dopa does not accumulate in the tissues but is converted immediately to dopamine or N-acetyldopamine. The metabolism of dopa to dopamine is brought about by dopa decarboxylase, which is present not only in the haemolymph cells but also especially in the epidermis. It is induced by the moulting enzyme 20-hydroxyecdysone. Dopa decarboxylase of various dipteran larvae, like that of the mammals, is a dimer of about 100 kDa. The insect enzyme is more specific than that of the mammals and is inactive against tyrosine and phenylalanine; it is competitively inhibited by N-acetyldopamine [146]. The enzyme responsible for the acetylation of dopamine has been isolated from *Drosophila* and *Aedes* [230]. In addition to N-acetyldopamine (NADA), the cuticula of many insects contains N-β-alanyldopamine (NBAD), which is synthesized from β-alanine and dopamine by the action of NBAD synthase [130]. Hydroxylation at position 2 (β) of the side-chain gives rise to N-acetylnorepinephrine (NANE) from NADA, and to N-β-alanylnorepinephrine (NBANE) from NBAD [157]. The corresponding *o*-quinone is then produced by phenoloxidase; this is converted to the unstable *o*-quinone methide, which reacts spontaneously with water to give β-hydroxylated products, or becomes involved in β-sclerotization [5, 6, 238]. The cuticula of the Coleoptera is unusual in that it contains 3,4-dihydroxyphenylacetic acid (DOPAC) in large amounts. According to the results of tracer experiments, this arises from tyrosine via *p*-hydroxyphenylpyruvate and 3,4-dihydroxyphenylpyruvate as intermediates [19]. The spectrum of catecholamine derivatives varies with the species, the developmental stage, and the exact function of the cuticula [60].

The **tanning of non-cuticular protein structures** involves other phenolic substances, e.g. 3,4-dihydroxybenzoic acid (protocatechuic acid; Fig. 12.11) and the corresponding alcohol in the egg capsule of the cockroaches, NADA, N-malonyldopamine (NMAD) and other N-acyldopamines in the egg capsule of the praying mantids (Mantodea) [128, 239], 3,6-dihydroxybenzoic acid (gentisinic acid) in the silk of *Antheraea pernyi*, and 2-amino-3-hydroxybenzoic acid (3-hydroxaanthranilic acid) in the silk of *Hyalophora cecropia* [4]. The synthesis of these compounds is still somewhat obscure. There are most likely several pathways, which probably begin with the transamination of tyrosine to *p*-hydroxyphenylpyruvate, decarboxylation to tyramine, or hydroxylation to dopa.

Whilst the cuticula contains predominantly catechols, the haemolymph has mainly **conjugates** such as β-glucosides, sulphates, or phosphates; these are less subject to oxidation. For example, NBAD, NADA and dopamine are present in the haemolymph of newly hatched individuals of *Manduca sexta* almost exclusively as glucosides; dopamine and NADA are found as sulphates in the haemolymph of *Periplaneta americana* [128]. Sulphatases with the appropriate specificity are synthesized in the skin glands of cockroaches. β-Glucosides of protocatechuic acid, the corresponding aldehydes and alcohols, and various N-acyldopamines are found in the left colleterial gland of the cockroaches and the praying mantids (Mantodea); this gland is involved in the production of the egg capsule [276].

12.6.2 The Process of Sclerotization

The newly formed insect cuticula is plastic and flexible; its tensile strength is mainly due to chitin fibrils which can slide within the protein matrix. Greater mechanical strength of the cuticula is achieved by the process of **sclerotization**, which is initiated immediately following moulting. This involves cross-linking of cuticular proteins to each other and to chitin fibrils, reorientation of the chitin fibrils, reduction of protein solubility, reduction in water content, and often the appearance of brown pigmentation. The extent of sclerotization and the resulting reinforcement varies in different regions of the cuticula. For example, the mandibles of the locust *Schistocera gregaria* have 20-fold higher cross-linking than the cuticula of the abdominal terga. In insect larvae, only the outer layer of the cuticula (exocuticula) is usually sclerotized, but in adults the endocuticula is also included. Those parts of the cuticula which also contain resilin (an ideal rubbery material; see p. 388) have special mechanical properties.

Two **types of model** have been designed to describe the phenomenon of sclerotization. According to one model, sclerotization involves the covalent linking of single amino acid residues on different polypeptide chains by aromatic or quinoid bridges. The aryl-substituted amino acids in cuticular hydrolysates predicted by this bridge

model do, in fact, exist, and the model is accepted by a majority of researchers [4, 128, 138]. The alternative model explains sclerotization by non-polar interactions between protein chains and an intervening melanin- or lignin-like polymer. According to this model, the dehydration associated with sclerotization is the cause of increased hydrophobic interactions between the components and the higher mechanical strength of the cuticula [4, 138]. The following arguments are offered in favour of the second model: nuclear magnetic resonance (NMR) spectra of the pupal cuticula of *Manduca sexta* have so far not revealed any quinone bridges [182]. The number of amino acids available for bridge building, i.e. lysine residues and N-termini, amounts to only one in 20–36 amino acids, and is insufficient to account for the observed increase in mechanical strength. Finally, it has been found that dehydration of isolated larval cuticulae from *Calliphora vomitoria* always produces the same values for tensile strength, irrespective of whether the cuticula is treated before or after sclerotization.

The continuing uncertainty about the basic principles of sclerotization can be explained by the extraordinary chemical and structural complexity of the insect cuticula and the multiplicity of processes that occur. The individual processes of sclerotization take place within nanoseconds to several hours or days. Only part of the components involved are present before the process begins; some appear only during sclerotization and are very short-lived. Little is known about the cuticular proteins involved in sclerotization (p. 391), and it is also not at all clear how far quinones serve as bridges in the binding of protein and chitin [128, 138].

Protein tanning by quinone was discovered in 1940 by Pryor in the egg cocoon of the cockroach *Blatta orientalis*. In the cockroaches, the oviduct ends in a genital cavity to which the paired colleterial glands are connected. The secretion of the left gland contains dissolved proteins, diphenoloxidase and various catechol-β-glucosides, and that of the right gland contains a β-glucosidase. Depending upon the species, the catechol portion of the glucosides consists of differing proportions of 3,4-dihydroxybenzoic acid (protocatechuic acid) and the corresponding aldehyde or alcohol (Fig. 12.11). After cleavage of the glucoside bond, the diphenol is oxidized to *o*-quinone, which causes the white solution of dissolved proteins to solidify into the brown egg cocoon (ootheca). As this contains no chitin, its mechanical strength depends entirely upon the tanning of proteins. The egg cocoon of the praying mantids is formed in a similar manner. In this case, the left colleterial gland contains a β-glucosidase, and the right gland contains acyldopamine-β-glucoside, with species-specific differences in the acyl residues: acetyl- in *Statilia maculata*, malonyl- and acetyl- in *Tenodera aridifolia* and *T. angustipennis*, and N-acetyl-β-alanyl- in *Mantis religiosa; Hierodula petelliphera* has these three plus N-β-alanyl- and N-malonyl-β-alanyl- [276].

Considering the functional adaptations required, it is to be expected that **sclerotization of insect cuticulae** is associated with great heterogeneity in the bridge structures between the polypeptides and between proteins and chitin fibrils [138]. The hydrolysis products of sclerotized cuticulae and in vitro experiments with catechols or quinones, proteins and phenoloxidases give the following picture of the process of bridge formation (Fig. 12.12a): N-acetyldopamine (NADA) is oxidized to the corresponding quinone by phenoloxidase, and binds to the free amino group of a protein to give a catechol-protein combination. This is further oxidized to quinone and binds to a second polypeptide to form a bridge, the structure of which is uncertain [4, 138]. Dopa- and dopamine-quinone tend to form precursors of melanin by intramolecular condensation; this reaction is prevented in the sclerotizing cuticula of insects by blocking of the amino group (N-acyldopamine) or its absence (protocatechic acid). Quinone bridges can also be formed from protein-bound dopa-quinone, which arises from a tyrosine residue of the polypeptide chain by the action of phenoloxidase (self-tanning). Finally, cuticular proteins can also be linked by bistyrosine residues (see Fig. 11.4, p. 380) [138].

The formation of a completely different type of bridge structure was described by Andersen in 1970, and this was termed **β-sclerotization** because the protein in this case is bound to the β carbon atom of the catechol side-chain and not to the ring. In the meantime, it has been found that protein bonding can also occur via the α carbon atom of the side-chain (Fig. 12.12b), but the name β-sclerotization has been retained. Andersen assumed that the β-sclerotizing agent was α,β-dehydro-NADA (Fig. 12.12b), which was supposedly formed by a desaturase. On the other hand, Sugumaran's research group favoured quinone methide as the β-sclerotizing agent and they rechristened the process "quinone methide sclerotization". In fact, cuticular preparations from various insects contain not only the quinone methides of NADA and NBDA but also dehydro-

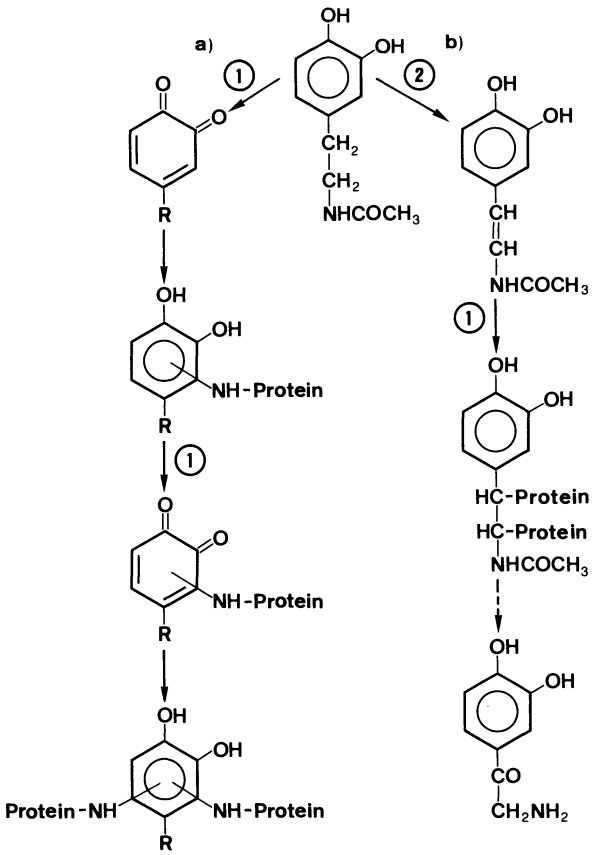


Fig. 12.12a, b. Sclerotization of insect cuticulae [4, 138]. **a** Quinone tanning. N-acetyldopamine (NADA) is oxidized to the corresponding quinone by phenoloxidase and binds to protein; the protein-catechol complex is further oxidized to quinone and is linked to a further polypeptide by a bridge of incompletely defined structure. **b** β-sclerotization. Protein chains are bound to the β and α carbon atoms of the NADA side-chain; various β-keto compounds are produced from these bridge structures by hydrolysis. The β-sclerotization probably involves an intermediate with an unsaturated side-chain. *1*, Phenoloxidase; *2*, N-acetyldopamine desaturase

NADA; however, this is probably not the result of direct desaturation of NADA but is due to the tautomerization of NADA-*o*-quinone methide. Dehydro-NADA-quinone can form benzodioxin dimers [6, 215, 238]. It is completely feasible that ring and side-chain bonds arise on the same catechol molecule so that the bridge structure has a protein:catechol ratio of 4:1 or 3:1 instead of 2:1 [138].

The ratio of quinone tanning to β-sclerotization differs markedly with the species and developmental stage, e.g. in the pupa of the lepidopteran *Manduca sexta* it is 75:25, in the puparium of the fly *Sarcophaga bullata* it is 85:15 and in the adults it is 97:3, in the ootheca of the cockroach *Periplaneta americana* it is 91:9, whilst in the cuticula of the adults it is 17:83, and in the femur of the migratory locust *Locusta migratoria* it is 10:90 [4]. There is some correlation between the type of bridge and the colour of the cuticula; in weakly pigmented skeletal structures one finds mainly β-sclerotization of the side-chains of catecholamine, and in darkly coloured structures there exists mainly quinone bridges, but the ratio of the two bridge types can vary widely even in dark regions of the cuticula [4, 157]. The degree of pigmentation of the cuticula is probably greatly dependent upon the spectrum of the acylcatecholamines, which serve as substrates for the oxidative enzymes. Thus, for example, the sclerotized, transparent cuticula of *Manduca sexta* and other insects contains mainly N-acetyldopamine but the hard, dark brown regions have mainly N-β-alanyldopamine [128].

Many insect cuticulae contain **firmly bound β-alanine**. The incorporation of this amino acid into the cuticula has been examined in detail for the puparium of flies in which the haemolymph of larvae just prior to pupation contains large quantities of free β-alanine or β-alanyl dipeptides [138]. During sclerotization, the β-alanine is incorporated into the cuticula. This process is disturbed in some mutants of the fly *Drosophila melanogaster* and the beetle *Tribolium castaneum* [210]. The *Drosophila* mutant "black" is unable to produce β-alanine, and the mutant "ebony" cannot incorporate it into the epidermis; in both cases, the cuticula of the adult animal is black instead of brown. In isolated cuticulae, dopamine alone produces a black colour, whereas N-β-alanyldopamine gives the normal brown colour. Thus, β-alanine apparently blocks the amino group of dopamine, inhibits melanin production, and promotes the formation of sclerotizing structures from catecholamine [4, 118]. The localization and the chemical structure of β-alanine bonds in sclerotized cuticulae are not known [4, 138]. It is also not clear whether, and in what way, the differences in catechol spectra between different insect species and between different regions of the cuticula influence the process of sclerotization and the properties of the sclerotized cuticula [157].

Compared with sclerotization of the insect cuticula, the hardening of the crustacean shell has been relatively little investigated [23]. In the fiddler crab *Uca pugilator*, dopamine is converted to N-acetyldopamine and incorporated into cuticular proteins of the post-moulting animal. In con-

trast, the cuticula of *Cancer pagarus* contains bis- and tertyrosine as well as its bromo-derivative. The coupling of protein-bound tyrosine residues brings about the rapid (within a few minutes) hardening of the fertilization membrane of sea-urchin eggs. In addition to bis- and tertyrosine, the isomer of isotertyrosine, pulcherosine (see Fig. 11.4, p. 380), is also formed [167]. Representatives of almost all animal phyla contain dopa-rich proteins which either form networks by quinone tanning or adsorb on mineral or metallic substrata via dopa side-chains and thereby function as adhesives [254].

12.6.3 Phenoloxidases

Apart from the pterin-dependent phenylalanine and tyrosine hydroxylases and the haem-containing peroxidases that have already been mentioned, the biocatalysts of the oxidation of aromatic compounds include in particular the copper-containing phenoloxidases, of which two types can be distinguished. One type, such as the **tyrosinase** of fungi, is active against monophenols and *o*-diphenols, is inhibited by thiourea and phenylthiourea, and has a neutral pH optimum. The second type, such as the **laccase** of the fungi and higher plants, attacks *p*- and *o*-diphenols, but is inactive against monophenols. This type mostly has a pH optimum in the weak acid region and is insensitive to the inhibitors mentioned above [4]. Many phenoloxidases occur as inactive pro-enzymes which may be activated by proteolysis.

The sclerotizing cuticula of insects contains non-extractable phenoloxidases which may be brought into solution by trypsin. Little is known about their structure. They oxidize *o*- and *p*-diphenols but not monophenols, and thus belong to the laccase group. Their pH optima usually lie between 4.5 and 5.5 but are sometimes closer to neutrality [4, 246].

The haemolymph of many insects contains numerous pro-phenoloxidases which, after activation by cuticular enzymes, are active against monophenols and *o*-diphenols, have neutral pH optima, and are inhibited by thiourea, i.e. they are tyrosinase-like. The cuticula also includes oxidases with a similar specificity which are not identical to the haemolymph enzymes. Sclerotizing cuticular regions always contain laccases and most also possess tyrosinases; in contrast, non-sclerotizing regions contain only the tyrosinases [4, 12].

The haemolymph of many insects and crustaceans contains multifactorial defence systems which recognize foreign particles and encapsulate them in melanin; at the same time, they function opsonically and activate the synthesis of immune proteins in the fat body. Amongst these systems are the zymogens of tyrosinase-like phenoloxidases and both their activators and inhibitors. Some of these components are synthesized and stored in blood cells. The in vitro activation of pro-phenoloxidases can be brought about simply by mild detergent treatment [232]. The natural activators are calcium-dependent serine proteases, which also occur as zymogens. Corresponding protease inhibitors have been detected in the lepidopteran *Manduca sexta* and the crayfish [9, 11, 13]. The phenoloxidase system of the arthropods is activated by bacterial substances such as peptidoglucans and β(1,3)-glucans. Corresponding β(1,3)-glucan recognition proteins have been isolated from the haemolymph of the silkworm *Bombyx mori* and the cockroach *Blaberus cranifer* [168, 233].

12.7 D-Amino Acids and Their Metabolism

For methodological reasons, there are very few data on the occurrence in animals of amino acids with the D-configuration. Until recently, the only stereospecific evidence was that obtained by enzymic cleavage with D-amino acid oxidases; because of the specificity of this enzyme, this method only provided information about particular amino acids. Nowadays, there are chromatographic methods of broad application, but these have not yet been used for the systematic investigation of D-amino acids in animals. Using the purchasable amino acid oxidase from mammals, 18 out of 43 examined marine animals from eight phyla were found to contain D-amino acids at concentrations of 1-2 μmol/g, and in some cases up to 44 μmol/g (coelomocytes of *Glycera dibranchiata)* [190]. Up to now, D-amino acids in animals have been found only in the free form or as constituents of low molecular weight compounds, and never as components of proteins.

D-Serine is found in the earthworm *Lumbricus* in the free form as well as in the phosphagen lombricin and its precursor serine-ethanolamine-phosphoric acid diester; it also occurs as the free amino acid in the larvae and pupae of the silk-

worm *Bombyx mori* and almost all species of Lepidoptera. Some moth larvae also contain **D-alanine**, as do the nymphs and adults of the bug *Oncopeltus fasciatus*. In contrast, neither D-serine nor D-alanine were detected in the honey bee or several species of beetle [36]. D-alanine is found in the muscles of all crustaceans examined, and in some species it also occurs in the hepatopancreas or the heart, but not in the haemolymph or the nervous system. D-Alanine has also been found in sea-urchin eggs and embryos [61]. The cysteine residue in the luciferin of all fireflies examined has the D-configuration but free **D-cysteine** was not found [36]. **D-2,3-diaminopropionic acid** occurs in the haemolymph and in the gut of the larvae of *Bombyx mori* and in several other caterpillars before pupation [36].

L-Aspartic acid converts spontaneously to the D-isomer at a temperature-dependent rate of about 0.1–0.25 % per year. Long-lived proteins, e.g. in human teeth or eye lenses, therefore contain increasing proportions of **D-aspartate** with age. However, spontaneous racemization does not explain, for example, why the cerebrum of young rats has 8.4 % of total free aspartate as the D-isomer, or why there is 4.9 % D-aspartate in human foetal blood and the value decreases with age. It is not clear whether the D-aspartate in these cases originates in the diet or from bacterial symbionts, or whether it is synthesized de novo during cell metabolism. Perhaps D-aspartate really has a transmitter function, as was claimed several years ago, although the idea was largely rejected [71]. In agreement with this concept is the fact that D-asparaginase is found only in the nervous system of the octopus and other cephalopods and not in other organs. The presence of considerable quantities of D-aspartate in the gills and foot of the mussel *Solenomya reidi* [78] is probably the result of uptake of the D-amino acid from seawater, which contains D-aspartate and other D-amino acids.

So far, there is little evidence for the **production of D-amino acids in metabolism**. In the silkworm, D-serine is produced from ^{14}C-labelled glucose or from L-serine. This reaction also takes place in homogenates and has a pH optimum of 9. The production of D-2,3-diaminopropionic acid from glucose has been demonstrated using the same technique [36]. The peptides dermorphin, dermencephalin and deltorphin, which are highly potent opioids in the skin of frogs from the subfamily Phyllomedusinae, contain a D-alanine residue which is formed from an L-alanine by post-translational modification (p. 297). The D-amino acids detected in the blood and various organs of man and other mammals have been thought to have a microbial origin; however, they also are found in axenically reared mice [161].

More is known about the catabolism of the D-amino acids than about their synthesis. As already mentioned, the D-isomers can be used to satisfy the requirements for several essential amino acids. There are two types of **D-amino acid oxidases** in animals, both of which contain FAD as the coenzyme but otherwise differ in their specificities. The first type oxidizes a series of amino-monocarbonic acids, the exact spectrum of which varies with the species. Enzymes of this type have been found in the kidney and liver of almost all vertebrates examined, although with widely different activities; they are localized in the peroxisomes. They have also been found in certain regions of the brain and in neutral leucocytes. The 345–347-amino-acid sequence of the kidney enzyme agrees between different mammals at about 77 % of positions and shows no relationship to other proteins [241, 252]. Enzymes with similar properties have also been found in the Malpighian tubules and fat bodies of insects and in the hepatopancreas of molluscs; they are probably widely distributed. Their biological significance is unclear; the main function of the enzymes from mammalian liver and kidney is that of a glycine oxidase, although this would account for less than 1 % of the total glycine turnover [36]. The second type of D-amino acid oxidase, which is highly specific for D-aspartic acid and D-glutamic acid, has so far been found only in the kidney of some mammals, the hepatopancreas of the cephalopod *Octopus vulgaris* and the antennal gland of the crayfish *Orconectes limosus* [164]. This enzyme also has FAD as the coenzyme but, in contrast to the enzymes of the first type, is not inhibited by benzoates. Comparative investigations have shown that D-aspartate oxidase is present in all eight-armed cephalopods (order Octopoda), but is consistently absent from the ten-armed species (order Decapoda).

12.8 Amines

Many amines have specific biological functions as neurotransmitters, neuromodulators or neurohormones. Biogenic amines are therefore found especially in the nervous system, although in very different, species-specific concentrations and combinations. As components of toxins, they are

found in widely varying concentrations in the toxin-producing posterior salivary glands of cephalopods and in the skin of amphibians. The **formation of amines from amino acids** requires a decarboxylation step (Fig. 12.13), but very little is known from a comparative biochemistry point of view about the corresponding enzyme. Porcine kidney contains a non-specific decarboxylase which can convert histidine, phenylalanine, tyrosine, tryptophan and 5-hydroxytryptophan; an enzyme with similar properties is also present in the insect brain. Two enzymes have been characterized in some detail from the fly *Sarcophaga barbata;* these decarboxylate glutamic acid to γ-aminobutyric acid (GABA). The mitochondrial enzyme in the flight muscle is similar to the glutamate decarboxylase II in mammalian kidney, whereas the enzyme in the fly brain is very different from enzyme I in the synaptosomes of mammalian brain. The GABA concentration in the brain of the migratory locust *Locusta migratoria*, as in mammalian brain, is regulated by the opposing actions of glutamate decarboxylase and GABA:2-oxoglutarate aminotransferase [14, 36, 46].

Part of the neurotransmitter released into the extracellular space in vertebrates, molluscs and arthropods is reabsorbed by the nerve endings. Excess amines in vertebrate brain are oxidized to the corresponding aldehydes by a mitochondrial **monoaminoxidase** (MAO). Biogenic amines can also be oxidized by other enzymes. The MAOs are FAD enzymes which are localized in the outer mitochondrial membrane and produce hydrogen peroxide. They are particularly active in liver, kidney and the gut wall, where they function in a detoxifying capacity, and in the brain, where they are involved in regulation of the neurotransmitter concentration. There are two MAO types in mammals, birds, reptiles and adult anurans: type A oxidizes serotonin and is inhibited by clorgylin, whereas type B prefers benzylamine and is insensitive to the inhibitor. The two types are mostly found together in proportions that depend upon the tissue; however, rat spleen and testis contain only type A, and murine liver, porcine brain and human blood platelets have only type B. The type B appears to be the evolutionally derived form: it is absent in the teleosts, makes up less than 10% of the total MAO activity in the urodelans, appears in the anurans only after metamorphosis, and arises later than type A in the mammalian foetus [188]. MAOs are widely distributed in the animal kingdom down to the unicellular organisms [155]. The edible snail *Helix pomatia* contains several such enzymes which are localized above all in the hepatopancreas, ganglia and nephridia. In insects, they are found with significant activity only in the Malpighian tubules and not at all in the brain or other tissues [98, 155]. As MAOs are absent from insect brain or, if present, have very low activities, amines in this case are inactivated by N-acetylation or conjugation with sugars, sulphate or phosphate [36, 46]. Similar metabolites are formed in the crustaceans, e.g. sulphate esters have been identified.

12.8.1 N-Methylated Bases

Choline (Fig. 12.14) is a component of the neurotransmitter acetylcholine and, at the same time, is an important synthetic element of lipids. The synthesis and catabolism via glycinebetaine as the intermediate has been investigated in detail only in the mammals. The enzyme betaine:homocysteine methyltransferase involved in choline catabolism is also found in fish and other lower vertebrates. As homocysteine is lacking in fish nutrients, glycinebetaine often accumulates. Choline is usually an essential nutrient of insects, although it is synthesized by some species. However, the corresponding tracer experiments with insects did not exclude the possibility of participation by symbionts. An exception is the unambiguous investigation of choline synthesis in the cabbage white *Pieris brassicae*. Caterpillars reared on an aseptic, partly synthetic diet incorporated labelled ethanolamine and the methyl group of methionine into choline, but they required the supply of some choline for optimal growth. The capacity for synthesis is hardly exploited when the diet contains sufficient choline and is, in any case, unable to provide all the choline required. As in other insects, cabbage white caterpillars are

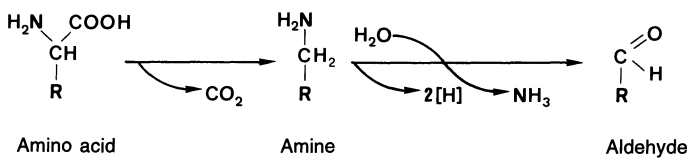


Fig. 12.13. Amine biosynthesis by decarboxylation of amino acids and amine catabolism by monoaminoxidase (MAO)

able to substitute various analogues for choline during lipid synthesis [30]. Acetylation of choline uses acetyl-CoA as the donor; the corresponding acetyltransferases have been isolated from fish, insects, the xiphosuran *Limulus*, molluscs and mammals [10]. These enzymes are all very similar in their physical and catalytic properties; for example, both the enzyme from the brain of *Loligo pealei* and that of bovine brain show a tendency to aggregate. A cDNA for choline acetyltransferase has been cloned from *Drosophila melanogaster*. The sequence of 728 amino acids is about 50–100 amino acids longer than that predicted by the 67-kDa mass of the active enzyme; it would appear that a larger, enzymatically inactive pro-enzyme is formed [116].

Trimethylamine oxide (TMAO) (Fig. 12.14) is widely found in marine animals and is one of the most important osmotically active substances. The highest concentrations are found in the muscles of marine elasmobranchs, although there are significant differences between the species; e.g., 150–200 μmol/g fresh weight (4–7 % of dry weight) in *Mustelus manazo*, and only 20–30 μmol/g fresh weight in *Squalus acanthias*. The concentrations are much lower in the marine and freshwater teleosts where, in some cases, the excretion of TMAO is accelerated by active tubular secretion in the kidney. High concentrations of more than 100 μmol/g fresh weight are also found in some cephalopods and crustaceans. The concentration of trimethylamine (TMA) is always very much lower than that of TMAO [10]. What is the origin of these large quantities of TMAO? The zooplankton all have a lot of TMAO. The copepod *Calanus finmarchicus*, which is widely found in the plankton, is able to synthesize TMAO but hardly excretes it, and the compound thus accumulates in the body to a concentration of 140 μmol/g fresh weight. Because TMAO excretion by many fish is very slow, it accumulates in the food chain. The biosynthesis of TMAO from choline has been demonstrated by tracer experiments in mammals, cartilaginous fish and crustaceans. Biosynthesis involves an NADPH-dependent TMA monooxygenase, which has been isolated from the liver of *Ginglymostoma cirratum* and other sharks and from cod liver. They are all microsomal FAD enzymes similar to that found in porcine liver. In contrast, *Calanus finmarchicus* has a cytoplasmic enzyme, which further differs from the vertebrate enzyme in having a much higher substrate affinity. Various teleosts and some marine invertebrates possess a glutathione-dependent TMAO reductase [82].

Carnitine (Fig. 12.14) is important for the transport of fatty acids across the mitochondrial membrane, and can probably be synthesized by most animals [10]. Biosynthesis in the fly *Musca domestica*, as in the mammals, begins with the N-methylation of protein-bound lysine. Carnitine supplied to *Musca*, which like all dipterans is unable to synthesize choline, is decarboxylated to β-methylcholine and incorporated into lipids; the decarboxylase, however, is active only in the larvae and pupae. The flour beetles (Tenebrionidae) cannot make carnitine and it is therefore an essential nutrient for them. *Tenebrio molitor* is also not able to decarboxylate carnitine to β-methylcholine, but it can incorporate exogenous β-methylcholine into lipids [30].

Homarine (N-methylpicolinic acid) (Fig. 12.14) is widely found in the marine arthropods and molluscs but is lacking in the freshwater invertebrates and the vertebrates. In the abdominal tissues of *Limulus* it makes up 17 % of the osmotically active substances, but plays no part in osmoregulation. De novo synthesis begins with the condensation of glycine with succinyl-CoA, but picolinic acid does not appear as an intermediate; the methyl donor is S-adenosylmethionine. The methyl group of homarine can be used in the synthesis of TMAO, choline and other methylamines; the resulting picolinic acid can be remethyl-

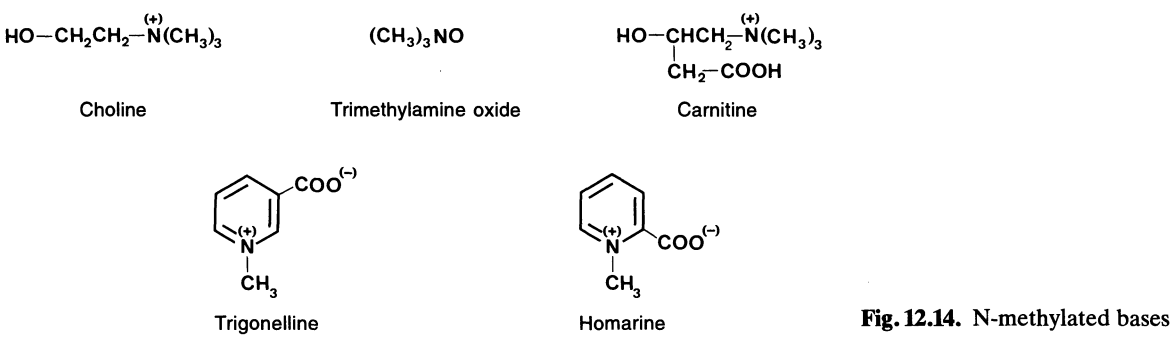


Fig. 12.14. N-methylated bases

ated to homarine by reversal of the transfer reaction. Some marine crustaceans contain the isomer trigonelline (Fig. 12.14), and further N-methylated bases are known in the molluscs [10].

12.8.2 Amines Derived from Tyrosine, Tryptophan and Histidine

Tyrosine is the starting point of a metabolic pathway via dopamine to noradrenalin and adrenalin (norepinephrine and epinephrine), and a further pathway via tyramine to octopamine (Fig. 12.15). Vertebrates mostly have noradrenalin and adrenalin, whereas invertebrate animals have dopamine and variable quantities of noradrenalin and octopamine; adrenalin has not been unambiguously detected in the invertebrates. **Octopamine** was discovered in 1951 in the toxin-producing posterior salivary glands of the cephalopod *Octopus vulgaris*, where it reaches concentrations of more than 1 mg/g fresh weight; shortly thereafter, it was recognized as a neurotransmitter. The use of sensitive detection methods has revealed concentrations of 10–100 ng/g fresh weight in the mammalian brain and in sympathetic innervated organs such as the heart, spleen and salivary glands; the concentrations in invertebrates are consistently much higher (Table 12.3). The natural form is the D- isomer [14]. With the exception of their involvement in sclerotization of insect cuticulae, there have been very few comparative biochemical investigations of the amines derived from tyrosine. Experiments with radioactive tyrosine on the nerve tissues of various insects, crustaceans and molluscs have regularly yielded labelled tyramine and octopamine, but at the most only traces of dopamine and noradrenalin. The tyramine concentration in the brain of the cephalopod *Octopus* is almost 100-fold that in the rat brain. It would appear, therefore, that the most important initial step in tyrosine metabolism in invertebrates is decarboxylation, in contrast to the hydroxylation in vertebrate brain [14]. A tyramine-β-hydroxylase has been isolated from the lobster, which, in fact, also metabolizes dopamine. In all vertebrates including the fish, catechol-O-methyltransferase is involved in the inactivation of adrenalin and noradrenalin, but this enzyme appears to play no role in the invertebrates.

The richest source of indolamines, derived from tryptophan, is the mammalian pineal organ (the epiphysis), which contains a dozen or so of such compounds. **5-Hydroxytryptamine** (serotonin) (Fig. 12.16) is apparently widely distributed in both vertebrates and invertebrates. The gills of the edible mussel *Mytilus edulis*, the ciliary action of which is regulated by serotonin, contain a specific oxidase that oxidizes serotonin and a series of other amines. The enzyme is inhibited by cyanide and azide but not by the typical inhibitors of the monoaminoxidases; thus, it is apparently an oxidase which does not require NADPH but which forms hydrogen peroxide. **Melatonin**, which is produced in the epiphysis from seroto-

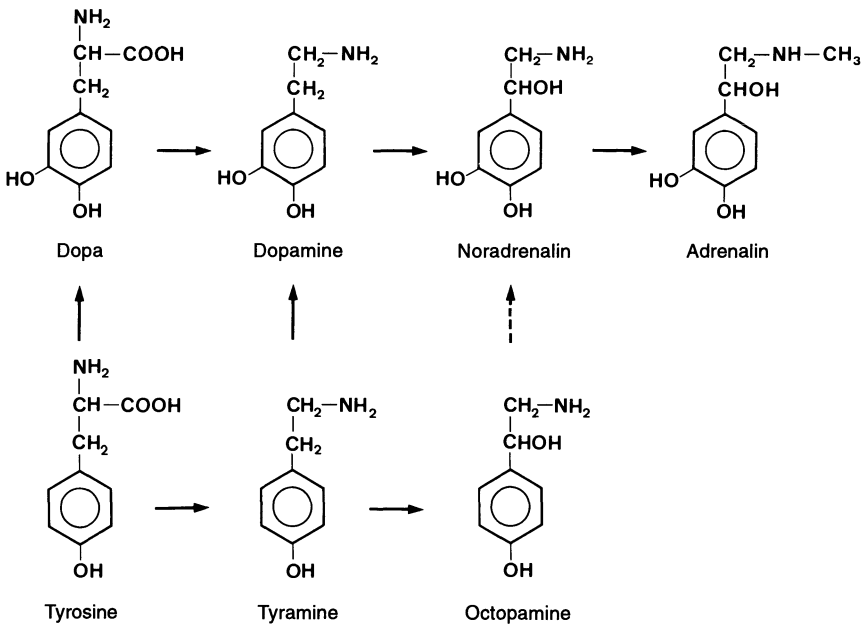


Fig. 12.15. Amines derived from the amino acid tyrosine

Table 12.3. The concentration (μg/g fresh weight) of octopamine (O) and noradrenalin (N) in various animal species

Species and tissue	O	N	Reference
Molluscs			
Helix aspersa CNS	0.13	0.07	[121]
Peristernia nassatula hypobranchial gland	125	–	[271]
Octopus vulgaris CNS	1.2	4.4	[121]
Posterior salivary gland	1300	–	[14]
Annelids and arthropods			
Lumbricus terrestris CNS	5.3	1.4	[121]
Homarus sp. CNS	0.23	0	[121]
Periplaneta americanus CNS	3.9	0.37	[121]
Echinoderms			
Pycnopodia helianthoides arm nerve	0.26	2.13	[121]
Vertebrates			
Rat brain	0.005	0.43	[121]

CNS, central nervous system

nin, is apparently found only in the vertebrates, whereas the **histamine** derived by decarboxylation of histidine (Fig. 12.16) is also found in many invertebrates.

12.8.3 Polyamines

The polyamines spermidine and spermine, together with their precursors diaminobutane and putrescine (Fig. 12.16), are involved in various fundamentally important cell functions and are universally distributed in prokaryotes and eukaryotes. As polyvalent cations they can interact with, and change the conformation of, nucleic acids. Hence, they are important for DNA replication and transcription and for mRNA translation. They apparently also have the function of secondary messengers and regulate both Ca^{2+} flux and transmitter release in synaptosomes. Through interactions with the endoplasmic reticulum (ER) membrane, they influence steroid metabolism, lipid peroxidation and prostaglandin synthesis [180, 240]. In the invertebrates, they have been detected in, for example, embryos of sea urchins and in polychaetes, in the tissues of adult echinoderms and in *Drosophila melanogaster* [258]. Particular polyamines, including the lysine derivative cadaverine (Fig. 12.16), are found in relatively high concentrations in the venom of spiders and scorpions. In the amoeba *Acanthamoeba castellanii*, 1,3-diaminopropane reaches the relatively high concentration of 15–25 mmol/l, whereas it is normally found in only trace amounts as a byproduct of polyamine metabolism [187].

In the mammals, as in the bacteria, the **biosynthesis of polyamines** begins with the decarboxylation of ornithine to putrescine, which is then successively converted to spermidine and spermine by the transfer of propylamine residues from decarboxylated S-adenosylmethionine (Fig. 12.17). This biosynthetic pathway is presumably present in all invertebrates but there have been very few comparative studies. Ornithine and S-ade-

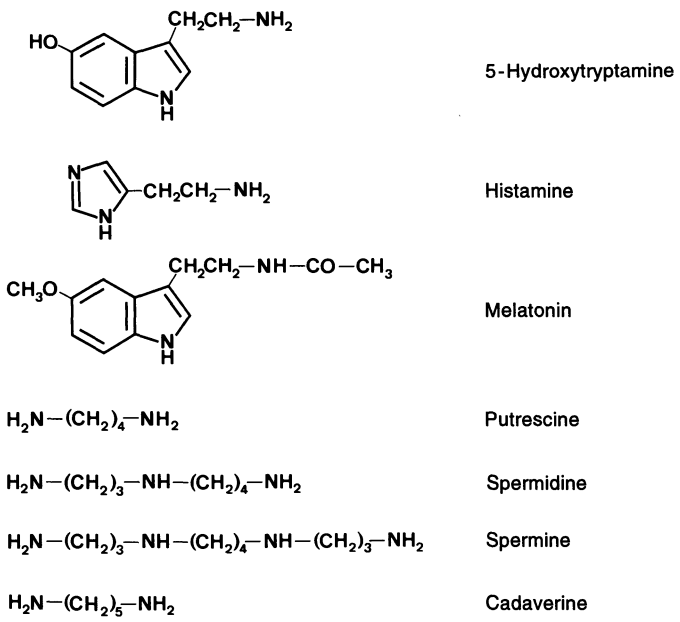


Fig. 12.16. Various mono- and polyamines

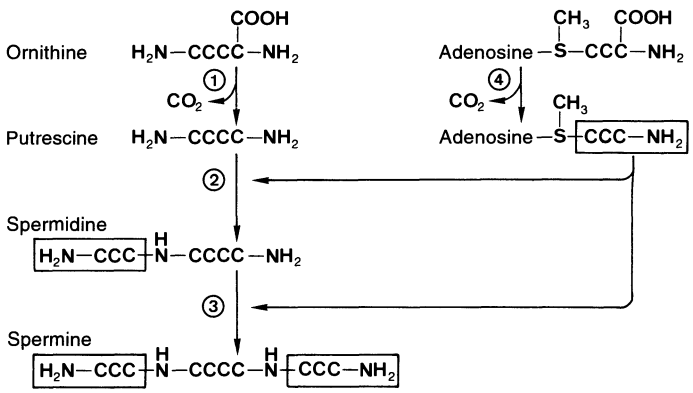


Fig. 12.17. The biosynthesis of polyamines. For clarity, the carbon skeletons are shown without hydrogen atoms. *1*, Ornithine decarboxylase; *2*, spermidine synthase; *3*, spermine synthase; *4*, S-adenosylmethionine decarboxylase

nosylmethionine decarboxylases have been found in the larvae of *Drosophila melanogaster*, with the highest activities occurring in the youngest individuals. The unexpectedly high putrescine concentrations of up to 375 μg/ml measured in the haemolymph of *Heliothis zea* caterpillars have been shown by tracer studies with ^{14}C-arginine to originate in the gut microflora. Ornithine decarboxylases are ubiquitously distributed from the bacteria to the mammals. In the flagellate *Trypanosoma*, as in the mouse, they are cytoplasmic dimeric proteins of 2 × 45 kDa. In contrast, in the nematodes *Caenorhabditis elegans* and *Haemonchus contortus* they are membrane bound. The enzyme from the ciliate *Tetrahymena pyriformis* is activated by phosphorylation [129, 140, 183, 216]. The S-adenosylmethionine decarboxylase isolated from the mantle, hepatopancreas and foot of the edible mussel *Mytilus edulis* has a mass of 65 kDa and a pH optimum of 7–8; it is therefore very similar to the mammalian enzyme in its physical and catalytic properties [17]. The spermidine and spermine synthases which function as aminopropyltransferases have not yet been investigated in the invertebrates. The catabolism of spermidine and spermine in the mammals begins with an N-acetylation. By the action of polyaminoxidase, 3-acetamidopropanol is then cleaved from the N^1-acetylspermine to give spermidine; N^1-acetylspermidine is similarly metabolized to putrescine. Polyaminoxidase was previously reported from only the rat and man, but has more recently been found in the gut wall, liver, kidney and brain of the teleost *Parasilurus asotus;* it differs from the mammalian enzyme mainly by its lower substrate affinity [132]. An enzyme isolated from *Ascaris suum* can acetylate putrescine and cadaverine but not spermidine or spermine [266].

12.9 Phosphagens and Other Guanidine Compounds

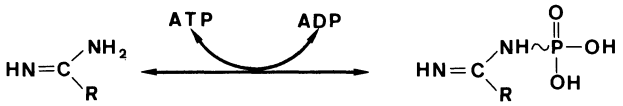
Creatine phosphate and other guanidine phosphates are referred to as phosphagens because they yield orthophosphate on mild hydrolysis. Their bound, energy-rich phosphate groups originate from ATP and can be retransferred to ADP. Thus, in muscle they constitute an energy store which is many times greater than that of ATP. Phosphagens are also involved in the regulation of glycogenesis in resting muscle, in that they bind inorganic phosphate and thus reduce its concentration to rate-limiting levels. The important role of creatine phosphate in the intracellular transport of energy-rich, bound phosphate, the **creatine-creatine phosphate shuttle**, was first discovered in 1978 [24]. Mammals have both mitochondrial and cytoplasmic forms of creatine kinase and these are found, for example, in all muscle types, brain cells and fat cells. The mitochondrial creatine kinase (mCK) is located on the outer surface of the inner mitochondrial membrane, produces creatine phosphate with mitochondrial ATP, and provides ADP for oxidative phosphorylation. The mCK interacts with the adenine nucleotide translocase of the inner mitochondrial membrane and is practically inaccessible for cytoplasmic ATP. Cytoplasmic creatine kinase is bound to the myofibrils of muscle, particularly in the region of the M lines. Isolated muscle fibres contract more rapidly when treated with 10 mmol/l creatine phosphate and 4 μmol/l ADP than with 100 μM ATP. This is not just due to the somewhat faster diffusion of creatine phosphate, but rather that creatine kinase and myosine ATPase are apparently functionally connected. Creatine kinase associated with the microsomes and polysomes is involved in lipid and

protein biosynthesis, whereas that associated with the ER and the plasma membrane has a function in ion pumping [24, 211]. Creatine phosphate-dependent energy transport has also been detected in sea-urchin spermatozoa [270]. It can be assumed that the other phosphagens of the invertebrates also take part in shuttle systems, although there is at present no direct evidence for this.

12.9.1 Biosynthesis and Occurrence

A classical thesis of comparative biochemistry states that creatine phosphate occurs only in vertebrates, and that all other animals ("invertebrates") have arginine phosphate. This simple concept was destroyed in the 1950s by the discovery in some sea urchins of creatine phosphate together with arginine phosphate, and the identification of a series of other phosphagens in annelids. Today, nine different phosphagens are recognized (Fig. 12.18), the distribution of which does not coincide entirely with the systematic divisions of the animal kingdom (Table 12.4). Even species of the same family may have different phosphagens. A correlation with the ecological conditions is also not apparent: at the same site on the seabed one finds the polychaetes *Travisia forbesi* with arginine phosphate, *Ophelia bicornis* with D-lombricine phosphate, and *O. neglecta* with opheline phosphate; all these species belong to the same family (Ophelidae) [82].

In the **biosynthesis of guanidine compounds**, the decisive step is the transfer of the amidine group from arginine to an amine as the acceptor (Fig. 12.19). Exactly which guanidine compound is produced depends only upon the nature of the



– R:	
Glycocyamine	–NH–CH_2–COOH
Creatine	–N(CH_3)–CH_2–COOH
Arginine	–NH–$(CH_2)_3$–CH(NH$_2$)–COOH
Taurocyamine	–NH–CH_2CH_2–$SO_3^{(-)}$
Opheline	–NH–CH_2CH_2–O–P(=O)(OH)–O–CH_3
L-Lombricine	–NH–CH_2CH_2–O–P(=O)(OH)–O–CH_2–CH(NH$_2$)–COOH
D-Lombricine	–NH–CH_2CH_2–O–P(=O)(OH)–O–CH_2–CH(NH$_2$)–COOH
Thalassemine	–NH–CH_2CH_2–O–P(=O)(OH)–O–CH_2–CH(N(CH_3)$_2$)–COOH

Fig. 12.18. Phosphagens

Table 12.4. Phosphagens or the corresponding phosphagen kinases in muscles or whole animals

Protozoa: A [82]		
Porifera: A or K [36, 82, 185]		
Plathelminthes: n.d. [18]; *Fasciola gigantica* and *Paramphistomum explanatum:* A + G + K [91]		
Nemertini: A [82]		
Nemathelminthes: n.d. or A [18]		
Mollusca: A [82]		
Annelida [82]		
Polychaeta:	Nereidae:	G and/or K
	Aphroditidae:	G and/or K
	Glyceridae:	K
	Eunicidae:	K
	Terebellidae:	mostly A
	Sabellidae:	A or K or T
	Ophelidae:	D–L or O
Oligochaeta: D–L; *Tubifex* sp. D–L + A		
Hirudinea: n.d.		
Sipunculida: *Sipunculus nudus:* A, various species: H + T [82, 95]		
Echiurida: *Bonellia viridis:* n.d., *Urechis caupo:* L–L [82], *Thalassema neptuni:* Th [245]		
Arthropoda: A [82]		
Echinodermata [82, 207, 257]		
Crinoidea:	A	
Echinoidea:	A or A + K	
Ophiuroidea:	K	
Asteroidea:	A or K	
Holothurioidea:	A or A + K	
Hemichordata: *Saccoglossus kowalevskii:* A + K [185]		
Chordata		
Tunicata:	A [82]	
Acrania:	K [185]	
Vertebrata:	K	

A, Arginine; G, glycocyamine; H, hypotaurocyamine; K, creatine; L–L, L-lombricine; D–L, D-lombricine; O, opheline; T, taurocyamine; Th, thalassamine; n.d., no phosphagen detectable

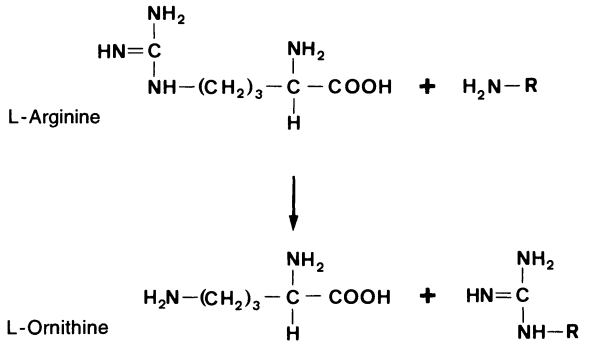


Fig. 12.19. Guanidine compounds are formed by transfer of the guanidino group of arginine to an amine

acceptor, as the amidine transferase is more-or-less unspecific. Which of the guanidine compounds is phosphorylated to a phosphagen is decided by the specificity of the available kinase. An extreme example of this is provided by the eggs of the polychaete *Audouinia tentaculata;* these contain arginine, glycocyamine, creatine, D-lombricine and taurocyamine, but only lombricine is phosphorylated [82]. The biosynthesis of creatine commences with the production by amidine transfer of guanidinoacetate (glycocyamine), which is then methylated. The enzymes involved, arginine:glycine amidinetransferase and guanidinoacetate methyltransferase, have so far not been detected in any of the creatine-containing invertebrates; tracer experiments with ^{14}C-arginine and ^{14}C-glycine all produced negative results. One or both of these enzymes is missing in certain agnathans, cartilaginous and bony fish, amphibians and reptiles. Most likely, the creatine requirement of these animals is satisfied by their diets [185]. Spontaneous ring closure converts creatine or creatine phosphate to creatinine, which is found in the urine of all vertebrates [24]. Hypotaurocyamine is produced by amidine transfer to hypotaurin, and can be oxidized to taurocyamine. Lombricine arises from serine-ethanolamine-phosphoric acid diester. In the annelids, the serine portion is in the D-configuration and D-lombricine is produced. Echiurids, on the other hand, make L-lombricine, which may be methylated to thalassemine [82, 245]. L-Serine-ethanolamine-phosphoric acid diester is also widely found in vertebrate muscle, although lombricine is never present.

Like creatine phosphate, the other phosphagens are not restricted to muscle. Arginine phosphate is found, for example, in the eyes of various arthropods. The spermatozoa of the annelids always contain creatine phosphate, but the eggs generally have the same phosphagen as the muscles. In the polychaete *Arenicola marina*, the muscles contain taurocyamine phosphate and the eggs contain this phosphagen in addition to creatine phosphate [123]. Echinoderm spermatozoa always contain creatine phosphate and the eggs usually have arginine phosphate, regardless of the phosphagen present in the muscles [123]. Several animal species possess **further guanidine compounds**, which are also formed by amidine transfer but are not phosphorylated. These include, for example, phascoline, phascolosomine and bonellidine of the echiurids *Phascolion strombi, Phascolosoma* sp. and *Bonellia viridis* (Fig. 12.20) [82, 95]. The diguanidine base, arcaine, of the mussel *Arca novae* arises from putrescine; audouine from the polychaete *Audouinia tentaculata* is formed from cadaverine; and hirudonine from the leech *Hirudo medicinalis* is formed from spermidine by a double amidine

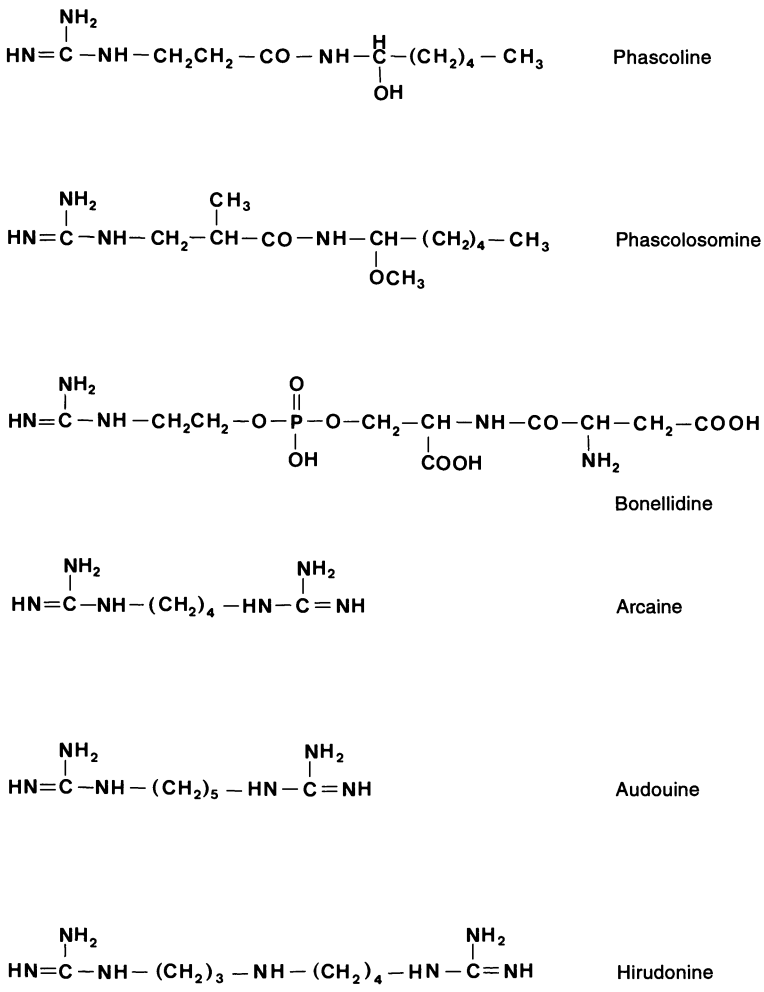


Fig. 12.20. Various guanidine derivatives

transfer. The biological importance of all these substances is unclear. *Hirudo* and *Bonellia* appear to completely lack phosphorylated guanidine compounds [82]. The turbellarian *Polycelis nigra* also lacks both phosphagens and phosphagen kinases.

12.9.2 Phosphagen Kinases

The various phosphagen kinases have numerous **catalytic properties** in common. They all require Mg^{2+}, which can be replaced by Mn^{2+} but not by Ca^{2+}, and have pH optima for the reactions with ATP and ADP of 8 and 7, respectively. The Michaelis constants lie generally around 1 mmol/l. The creatine kinases of the vertebrates and the glycocyamine kinases of the polychaete *Nephthys coeca* are strictly substrate specific. The arginine kinases of all arthropods, the mussel *Solen ensis* and the sipunculid *Sipunculus nudus* are highly specific for L-arginine. On the other hand, the arginine kinases of certain platyhelminths, molluscs and annelids can also convert, for example, D-arginine and glycocyamine [82, 221, 257]. Much less specific are the hypotaurocyamine kinase from *Phascolosoma vulgare*, the lombricine kinase from *Lumbricus terrestris*, the opheline kinase from *Ophelia neglecta* and the taurocyamine kinase from *Arenicola marina;* each shows some activity with the substrates of the others but not with arginine or creatine. All known lombricine kinases convert both L- and D-lombricine [82, 111]. The phosphagen kinases have no marked allosteric characters. Cooperativity between nucleotide- and guanidine-binding at the same active centre is seen in mammalian creatine kinase as well as in the arginine kinase of *Holothuria forskali* and the honey bee, but not in that of the crustaceans or the polychaete *Sabella*. In addition to direct phosphate transfer from one substrate to another, a second mechanism exists

```
Creatine kinase:      ...NHLGYVLTCP-SNLGTGLR...
Arginine kinase:           ...QTCPTSNLGT-VR...
Lombricine kinase:     ...LGTI-TCPGSNLGT-LR...
Taurocyamine kinase:   ...LGTLGTCP-TNIGTGLR...
```

Fig. 12.21. The amino acid sequences of various phosphagen kinases close to the essential cysteine [29]

for several enzymes in which a phosphorylated enzyme intermediate is formed. This mechanism is shown by the arginine kinases of *Homarus*, several other crustaceans and *Holothuria*, but not by those of *Sabella, Limulus* or *Apis;* it is also not observed with the vertebrate creatine kinases.

All phosphagen kinases are homologous. This is already clear from their structural and functional properties, but is seen in particular from the agreement of their **amino acid sequences** in the region of the active cysteine (Fig. 12.21). Complete sequences are available for the creatine kinases of various mammals, the chicken and the rays *Torpedo marmorata* and *T. californica*. The arginine kinase of the lobster *Homarus americanus* and the lombricine kinase of the earthworm *Lumbricus terrestris* have been partially sequenced [38]. The close relationship of the different phosphagen kinases is particularly clearly demonstrated by the creation in vitro of a dimeric hybrid enzyme from a mammalian creatine kinase and the arginine kinase of a holothurian [221].

The **arginine kinases** of all arthropods and many molluscs are monomers of about 40 kDa; the echinoderms, sipunculids and several marine molluscs have dimeric enzymes of 80 kDa. The body wall of the sedentary polychaetes *Spirographis spallanzani* and *Sabella pavonina* (Sabellidae) contains an arginine kinase of 160 kDa; in addition, the spermatozoa have a dimeric enzyme which differs electrophoretically and immunologically from that of the body wall. A tetrameric arginine kinase is also reported for the sea anemone *Condylactis auriantiaca* [201, 203]. Whereas the mitochondrial creatine kinases are well known, there are few reports on other phosphagen kinases associated with the mitochondria. Mitochondrial arginine kinases are apparently not present in the molluscs but have been described in various arthropods. Mitochondria from the heart of the horseshoe crab *Limulus polyphemus* contain two arginine kinases, one of which (AK_{mf}) appears to be identical to the cytoplasmic form (AK_c), whilst the other (AK_{ms}) is mitochondria specific. In contrast to the electrostatically bound mitochondrial creatine kinase of the vertebrates, the AK_{ms} carries a hydrophobic targeting sequence for the mitochondrion and is not functionally linked to the adenine nucleotide translocase [66]. The **lombricine kinases** are dimers of 69 kDa but, in contrast to other dimeric and tetrameric phosphagen kinases, they have only one active centre [111]. The **taurocyamine kinases** isolated from *Arenicola marina* have been estimated to have molecular masses of about 60 kDa. The **creatine kinases** are superior to all other phosphagen kinases in that they can buffer the ATP concentration in the cell at a high ATP:ADP concentration ratio [75]. They are usually dimers of about 80 kDa. As in the case of the dimeric arginine kinases, the subunits of the creatine kinases also appear to be arranged asymmetrically. Hence, hybridization of creatine and arginine kinases results in two electrophoretically distinct types of hybrid molecules with different substrate specificities. Two creatine kinases can be isolated from the sperm of the sea urchin *Strongylocentrotus purpuratus;* the F-type, located in the tail, is a monomer of 145 kDa and the K-type, of the sperm-head mitochondria, is a multimer of 240 kDa made up of 47-kDa subunits. Both form a shuttle system. The polypeptide chain of 1174 amino acids of the 145-kDa enzyme is divided into three regions, each of which bears an active centre and is highly homologous to the vertebrate creatine kinase (CK). It appears, therefore, that this structure arose by triplication of a CK gene [270]. In the sea urchin, which has two different phosphagen kinases, CK predominates in muscles with a tonic function and arginine kinase (AK) predominates in those with a phasic function. For example, in the tonic spine muscles of *Echinus esculentus* the CK:AK ratio is 20:1, but in the fast muscles of the tridactylous pedicellarians it is 1:2.6 [180]. A gene of the trematode *Schistosoma mansoni* encodes a CK of 675 amino acids with two domains which agree 49 % with each other and 33 % with the CK-M of the rabbit [236].

In the mammals, there are four CKs encoded by different genes, two mitochondrial and two cytosolic. The mitochondrial subunits consist of 416 amino acids, including a 38-residue transit peptide for mitochondrial import. They form octamers in equilibrium with dimers. Of the two types of mitochondrial CK, one is specific for the cross-striped heart and skeletal muscle, and the other is ubiquitous. In their sequences they agree 80 % with each other but only 66 % with the cytosolic isoenzymes [97]. The two cytosolic CKs are designated muscle type, M (or A), and brain

type, B. Types B and M can form heterodimers but cannot associate with the mitochondrial subunit. In addition to its occurrence in the brain, the homodimer BB is found in kidney, lung and smooth muscle; MM is present in fully differentiated skeletal muscle and, together with MB, in the heart. Embryonal skeletal muscle has exclusively type B, which is later exchanged for type M during differentiation. Chicken heart and brain also contain different types of mitochondrial subunits; these differ in 10 of their 30 N-terminal amino acids but can form heterodimers and -octomers. As in all higher vertebrates, the chicken also has cytosolic creatine kinases of type M and B. However, in this case, the two isoenzymes with differing N-terminal sequences arise by alternative splicing of the B-CK gene [265, 274]. Polyploid vertebrate species, e.g. *Xenopus laevis* and some teleosts, have even more complicated patterns of creatine kinases [206]. The gene duplication leading to the separation of the B and M types took place quite early in vertebrate evolution. This is seen from the fact that the sequence agreement between enzymes of the same type in different vertebrate classes is higher than that between isoenzymes from the same individual. For example, human creatine kinase M agrees 89–91 % with the M chain of other mammals, 81 % with that of the chicken and 80 % with that of the rays *Torpedo marmorata* and *T. californica*, but only 76–79 % with the B chain of various mammals. Canine B agrees 96 % with rat B and 90 % with chicken B, but only 81 % with canine M [25, 181]. From these sequence comparisons, a "unitary evolutionary period" (UEP) can be calculated as 30 million years for the formation of a 1 % sequence difference, i.e. a relatively slow **rate of evolution**. It may be assumed that a stabilizing selection pressure arose both from the prerequisites of the enzyme activity and from structural relationships, e.g. to the M lines of skeletal muscle and to membrane-bound ATPases. According to the UEP value, the B and M subunits separated about 600 million years ago [33].

The gene duplication of the cytoplasmic creatine kinases took place relatively early in the evolution of the chordates. While the tunicates have only one locus, the lamprey *Petromyzon marinus* already has two. More than two loci for cytoplasmic creatine kinase are found only in the teleosts and some anurans [92, 267]. All teleosts possess four loci, which show increasing tissue specificity during evolution. Isoenzyme A is predominantly or exclusively expressed in skeletal muscle, B in the brain and eye, C in the stomach musculature, and D only in the mature testis. Quite complicated electrophoretic patterns of creatine kinases show up to six bands in the skeletal muscle of various salmonids; these variants arise by post-translational modifications of the enzyme. There are also four isoenzyme loci in *Xenopus laevis*. All four are expressed in the eye, and several in the stomach (I, II, IV) and heart muscle (II, III, IV), whilst the skeletal muscle switches from III in the tadpole to II in the adult; the testis and kidney express only IV. All *Xenopus* species, including the diploid *X. tropicalis*, express the same four loci but with species-specific tissue distributions; thus, in the tetraploid species diploidization appears to be complete. *Hymenochirus boettgeri* and *Pipa cavalhoi*, which also belong to the family Pipidae, express only three loci and type III is absent [267].

12.10 Oligopeptides and Their Metabolism

Oligopeptides, i.e. peptides with at most a dozen amino acids, can arise by the hydrolytic cleavage of proteins: However, such peptides can also arise by de novo synthesis, in which case they often show structural characteristics that are not usually encountered in proteins, e.g. peptide bonds with the γ-carboxyl group of glutamic acid, or β-amino acid residues. Some insects have been found to contain large amounts of short-chain peptides which apparently function as a store of amino acids. The respective roles of the dipeptides β-alanyltyrosine and γ-glutamylphenylalanine in the sclerotization of insect cuticulae have already been discussed. A general storage function, especially of essential amino acids, can also be assumed for the numerous small peptides found in the tissues and haemolymph of *Drosophila* larvae and adult *Calliphora erythrocephala*. In both cases these consist of several hundred chromatographically distinct compounds which probably arise by intracellular proteolysis before being released into the haemolymph. The haemolymph of *Calliphora* contains at least five different peptidases which can release amino acids from these peptides [27, 46, 56]. The basic di- and tripeptides detected in the haemolymph of larvae of *Phormia regina*, *Sarcophaga bullata* and *Musca domestica* probably also have a storage function [27, 46]. Glutathione and several histidine peptides, on the other hand, have other specific functions.

12.10.1 Glutathione

The tripeptide **glutathione** (Fig. 12.22) is almost ubiquitous in animals, although its concentration varies over a wide range. In view of its numerous important functions, it is surprising to come across organisms which apparently manage without glutathione, e.g. the amoeba *Entamoeba histolytica* [76]. The glutathione molecule has an SH group and a γ-glutamyl residue and because of these special features it takes part in many reactions as substrate or cofactor; in particular, it participates in the metabolism of alien substances. Reduced glutathione (GSH) is an effective reducing agent and is able, for example, to protect the labile SH groups of enzymes. GSH is a coenzyme of glyoxylase, DDT-dechlorinase and thyroxin deiodase. The reduction of organic peroxide, e.g. from fatty acids, is catalysed by glutathione peroxidase, the only known selenoprotein of the animals. It is found in all vertebrate classes, the crayfish *Orconectus limosus* and a non-identified pulmonate snail, but not in the insects *Antherea pernyi*, *Sarcophaga* sp. or the earthworm *Lumbricus terrestris*. In fish and mammals, the muscles, liver and blood have the highest activities. The reduction of the oxidized form, glutathione disulphide (GSSG), is brought about by the NADPH-specific GSSG reductase, which is present in all prokaryotes and eukaryotes examined. This is a homodimeric FAD enzyme of 2×52 kDa which has significant sequence agreement with the lipoamide dehydrogenase involved in the oxidative decarboxylation of pyruvate and 2-oxoglutarate [249, 251]. Instead of glutathione, the trypanosomatids have the 24-membered, macrocylic disulphide trypanothione (N^1, N^8-bis[glutathionyl]-spermidine). The corresponding reductase from *Trypanosoma congolense* has been sequenced via the gene and shows more than 50 % sequence agreement with the glutathione reductase of human erythrocytes [222].

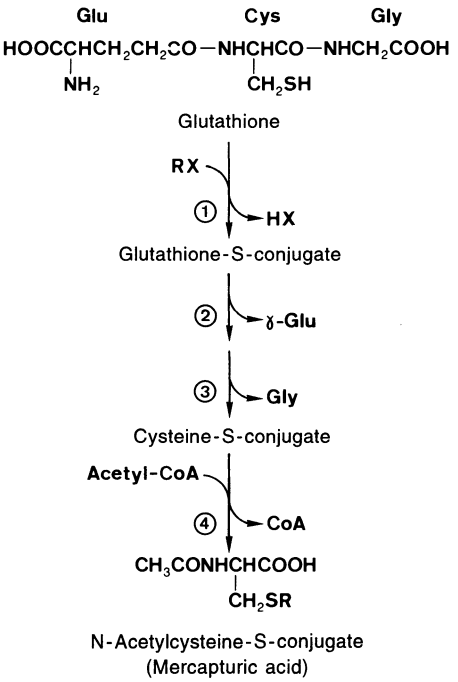


Fig. 12.22. Glutathione can form S- conjugates with various electrophilic compounds *(RX)*. These can be further metabolized to cysteine-S-conjugates and N-acetylcysteine-S-conjugates (mercapturic acids). *1*, Glutathione-S-transferase; *2*, γ-glutamyl-transpeptidase; *3*, cysteineglycine peptidase; *4*, N-acetyltransferase

Glutathione-S-transferases are a widely distributed family of multifunctional enzymes which can form S-conjugates of glutathione with many different electrophilic substances. Some of the mammalian enzymes of this type also have glutathione peroxidase or steroid isomerase activity; others bind organic anions like sulphobromphthalein, bilirubin or bile acids but do not conjugate them with glutathione. This latter type are sometimes referred to as "T proteins" or "ligandins" [184]. Mammals possess multiple isoenzymes of glutathione-S-transferase; these are encoded by a multi-gene family and, because of their differing specificities, interact with a wide range of substrates. The human and rat enzymes have been divided, according to functional and structural characters, into three classes: the basic α-, the neutral μ- and the acidic π- class. A fourth class (ϑ) has recently been added [152]. Because of their great importance in the metabolism of alien substances and environmental chemicals, glutathione-S-transferases have been investigated in the chicken [45], fish [165, 199] and many invertebrates [49], e.g. insects [16, 52, 54, 127], a mite [53], *Daphnia magna* [134], molluscs [15, 102], annelids [134], platyhelminths [31, 173], rotifers [28] ans the ciliate *Tetrahymena* [175]. Almost all the species examined in detail possess several isoenzymes, although in *Tetrahymena* only one enzyme was found. Most glutathione-S-transferases are homo- or heterodimers with subunits of 22–28 kDa; in *Schistosoma* (a trematode), *Brachionus* (a rotifer) and *Tetrahymena* (a ciliate) the enzyme is monomeric [28, 173, 175]. The complete amino acid sequences are known for several human and rodent isoenzymes and for one glutathione-S-transferase from *Drosophila* [34].

γ-Glutamyl-transferases transfer the γ-glutamyl residue from glutathione or its conjugate to an amino acid or a peptide; however, the enzyme also catalyses the hydrolytic cleavage of the γ-glutamyl bond. γ-Glutamyl-transpeptides are apparently widely distributed and are found in protozoans, echinoderms, molluscs and insects as well as in the vertebrates [46, 119, 204]. They are membrane-bound glycoproteins which occur, e.g. in the rat, as several tissue- and development-specific isoenzymes [93]. Glutathione-S-transferase and γ-glutamyl-transpeptidase catalyse the first two steps in a detoxification pathway which leads to the production of S-arylcysteine in insects and mercapto acids in vertebrates. The transpeptidase, which is located in the plasma membrane, was thought for a time to have an important role in the uptake of amino acids into cells (the γ-glutamyl cycle). However, there is also experimental evidence against such a role [214]. Human transpeptidase is a heterodimer with a heavy subunit of 42 kDa, which is anchored in the membrane, and a light subunit of 19 kDa, which carries the γ-glutamyl binding site. The two subunits arise from a common precursor of 569 amino acids [198]. In larvae of the housefly *Musca domestica*, γ-glutamyl-transpeptidase is responsible for the synthesis of the dipeptide γ-glutamylphenylalanine, whose role in sclerotization has already been described. The enzyme, which has its highest activities in the Malpighian tubules and the midgut, can utilize various amino acids in vitro, despite the fact that in vivo only the γ-glutamylphenylalanine is found. γ-Glutamylcyclotransferase, which can cleave the γ-glutamyl peptide into 5 oxoproline and free amino acids, has only low activity in larvae; thus, γ-glutamylphenylalanine is not cleaved but is released into the haemolymph. The activity of the enzyme increases considerably in the pupa; it accumulates and, in contrast to the mammalian enzyme, is highly specific for γ-glutamylphenylalanine. The release of phenylalanine from the dipeptide prior to sclerotization also involves γ-glutamyl-transpeptidase of the epidermal cells; this increases markedly in activity at this time and functions hydrolytically [27, 46].

12.10.2 Histidine Peptides

Three compounds have been described, especially from the skeletal muscles of the vertebrates, whose distribution makes them particularly interesting from a comparative biochemistry point of view: carnosine, anserine and ophidine (Fig. 12.23). Carnosine and anserine are found in almost all vertebrate muscles, although often in very different relative proportions. Birds, marsupials and rodents have more anserine than carnosine. Of the ape species examined, *Aeotes trivirgatus* has only anserine, and *Erythrocebus patas* only carnosine. Until recently, ophidine (balenine) was reported to occur only in snakes and whales from the families Balenopteridae and Delphinidae, which have no anserine. However, it is now clear that ophidine is present in trace quantities in humans and all other mammals investigated, and also in the chicken [101]. Histidine and histidine peptides apparently have mainly a buffering function. High-performance swimmers, like the tuna fish which produce up to 100 μmol lactate/g muscle during short periods of swimming at maximum speed, require an appropriate buffering capacity. In the species *Katsuwonus pelamis*, the white muscles contain free histidine at 94 μmol/g, anserine at 51 μmol/g and carnosine at 3 μmol/g, but the red muscles have altogether only 21 μmol/g [1]. Traces of carnosine are also found in vertebrates in the gut mucosa and eye lens, and somewhat higher amounts are present in the skin of various frogs of the family Leptodactylidae. Anserine is also found in the chicken brain and eye, and ophidine occurs in whale pancreas. Carnosine has been detected in the larvae of *Phormia regina* and several crustaceans, and probably is also present in molluscs. A related compound, homocarnosine (Fig. 12.23),

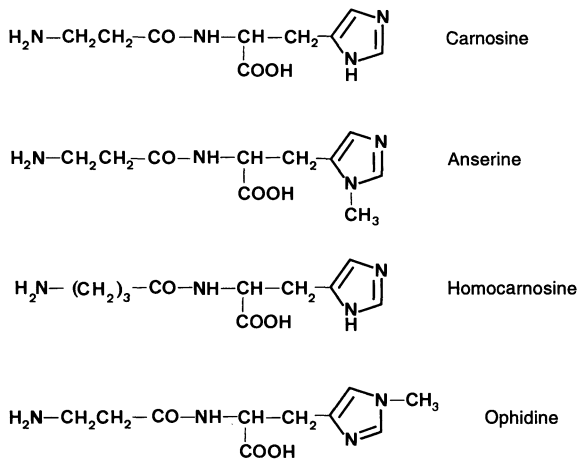


Fig. 12.23. Histidine dipeptides. Carnosine is β-alanylhistidine, anserine is β-alanyl-N^{π}-methylhistidine, homocarnosine is 4-aminobutyrylhistidine, and ophidine (balenine) is β-alanyl-N^{τ}-methylhistidine (not, as assumed previously, β-alanyl-2-methylhistidine)

is present in vertebrate brain with the highest concentrations occurring in the amphibians.

Carnosine is synthesized by carnosine synthetase; this forms an enzymes-β-alanyladenyl complex with the consumption of ATP, and this then reacts with histidine. The highest enzyme activities are found in the rat, mouse and chicken in the skeletal muscle and heart, and in the leopard frog *Rana pipiens* in the liver. Carnosine can be methylated to anserine by an S-adenosylmethionine:carnosine methyltransferase which, however, is not found in all vertebrates. The histidine dipeptides are cleaved by dipeptidases in the kidney, liver, muscle, brain and other organs. Porcine kidney contains a homocarnosinase with a broad specificity and a carnosinase with a narrower specificity. The so-called anserinase and N-acetylhistidinase from fish muscle and brain resemble porcine homocarnosinase in their broad specificity [1].

12.11 Purine and Pyrimidine Compounds

Every cell contains a complex spectrum of low molecular weight purine and pyrimidine compounds, which are components or degradation products of nucleic acids; they function variously as coenzymes or cosubstrates, secondary messengers, and nitrogenous waste products. Little variation is encountered and there is usually no explanation of its biological significance. The nucleosides with unusual structures, found in several sponges and sea snails, are dealt with as products of secondary metabolism in Chapter 19.

12.11.1 Biosynthesis of Purine Nucleotides

Purine nucleotides are synthesized either de novo or from preformed purine bases or nucleosides. De novo synthesis is not possible in many animals, in which case purine compounds are essential nutrients. In other cases, both synthetic pathways are made use of and their relative importance is usually difficult to estimate. The primary product of **de novo purine synthesis** is IMP. The origins of the individual atoms of the purine ring were determined in 1956, mainly by investigations on the pigeon liver (Fig. 12.24). In the meantime, this scheme has been shown to be universal by tracer experiments, and by the identification of intermediates and enzymes in lower and higher plants, insects, crustaceans and vertebrates [156]. The same is true for the conversion of nucleotide diphosphates into the corresponding deoxyribonucleotides by the iron-containing ribonucleotide reductase. It first became clear through molecular genetic studies with *Drosophila melanogaster* that multifunctional enzymes play an important role in both purine and pyrimidine biosynthesis. The activities of phosphoribosylamineglycine ligase (GARS), phosphoribosylglycine amide formyltransferase (GART) and phosphoribosylaminoimidazol synthase (AIRS) are located in a large polypeptide; in addition, there is a smaller protein with only GARS activity and this arises by alternative splicing. The GART gene is unusual in that the first intron contains a functional gene that is read in the opposite direction and codes for a cuticular protein (p. 15). There is also a human gene which codes for these three enzyme activities; in the avian liver, two enzyme activities of tetrahydrofolate metabolism are apparently also present on the equivalent protein complex [107]. The dual function of purine as a nucleic acid component and excretion product is

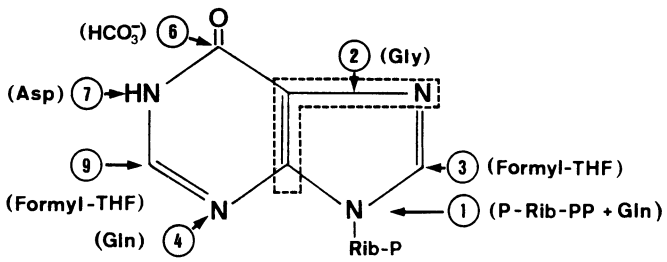


Fig. 12.24. Biosynthesis of the purine ring. The *circled numbers* indicate the consecutive steps of the reaction; the origin of each atom is given in *parentheses*. *1*, Amido transfer from glutamine to 5-phosphoribosyl-1-pyrophosphate; *2*, glycine transfer; *3*, formyl transfer from formyltetrahydrofolic acid; *4*, amido transfer from glutamine; (*5*, closing of the pentagon); *6*, carboxylation with HCO_3^-; *7*, linkage to aspartate; (*8*, cleavage of the aspartate C atoms as fumarate); *9*, formyl transfer from formyltetrahydrofolic acid; (*10*, closing of the hexagon)

reflected in organ specialization in the snail *Helix aspersa*; in the nephridium, it is mainly uric acid which is produced from ^{14}C-glycine, whereas in the hepatopancreas it is guanine. The appearance in terrestrial snails of guanine and xanthine, in addition to uric acid, as excretion products is probably due to reduced specificity of the excretion process.

Many of the parasitic protozoans belong to the class of animals which are unable to synthesize purine de novo. Of the Kinetoplastida, *Crithidia oncopelti* and *Trypanosoma equiperdum* can produce purine de novo, whereas the bloodstream form of *T. cruzi*, which in fact carries out no cell division, *T. lewisi* and various *Leishmania* species cannot. This synthesis capacity is also absent from *Giardia lamblia*, *Toxoplasma gondii* and various *Trichomonas* and *Plasmodium* species [21, 131, 194, 243]. Amongst the Platyhelminthes, larvae of the tapeworm *Mesocestoides corti* can carry out purine synthesis but the schistosome *Schistosoma mansoni* cannot; positive results are reported for the nematode genera *Angiostronglyus* and *Metastronglyus* [18]. Among the arthropods, de novo purine synthesis is not found in the brine shrimp *Artemia salina* or in certain aphids and dipterans [40]. The ascidian *Molgula manhattensis*, which like related species deposits uric acid in special storage organs, cannot make purines de novo but apparently obtains them preformed from its diet.

All cells can apparently make use of **preformed purine bases and nucleosides** (the salvage pathway). The most important of all the synthetic pathways for producing purine nucleotides is phosphoribosyl transfer (Fig. 12.25). The cells of higher animals contain two phosphoribosyltransferases, one of which is specific for adenine and the other for hypoxanthine, xanthine and guanine. These enzymes have been particularly thoroughly investigated in the erythrocytes of all vertebrate classes, where de novo purine synthesis is blocked. In contrast to the situation in the mammals, hypoxanthine/guanine-phosphoribosyltransferase activities are much lower in avian liver than in avian kidney; the enzyme would otherwise compete with xanthine dehydrogenase for the substrate hypoxanthine and would interfere with uric acid synthesis. The human erythrocyte enzyme has the relatively rare structure of a trimer with subunits of 30 kDa. Phosphoribosyltransferase from humans, the trematode *Schistosoma mansoni* and the malarial agent *Plasmodium falciparum* show up to 48 % agreement in their sequences of 218–231 amino acids [58].

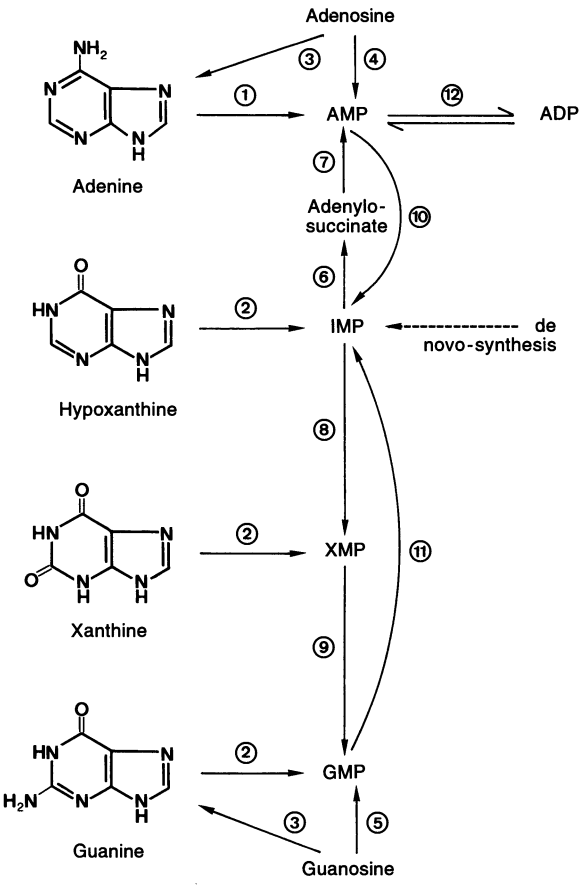


Fig. 12.25. Synthesis and conversion of purine nucleotides. *1*, Adenine-phosphoribosyltransferase; *2*, hypoxanthine/guanine-phosphoribosyltransferase; *3*, nucleosidase; *4*, adenosine kinase; *5*, guanosine kinase; *6*, adenylosuccinate synthase; *7*, adenyloscuccinate lyase; *8*, IMP dehydrogenase; *9*, GMP synthetase; *10*, AMP deaminase; *11*, GMP reductase (deaminating), *12*, adenylate kinase

There are four purine phosphoribosyltransferases in various *Leishmania* species; these are specific for adenine, hypoxanthine, guanine and xanthine [103]. In *Crithidia fasciculata* there is one each for hypoxanthine, xanthine and adenine; one enzyme is found in *Eimeria tenella* and this converts xanthine as well as hypoxanthine and guanine.

The enzymes responsible for the production of nucleotides from nucleosides and the reciprocal conversion of the purine nucleotides (Fig. 12.25) have rarely been examined comparatively. Adenylosuccinate synthase, which catalyses the first step of AMP synthesis from IMP, is part of the purine nucleotide cycle together with adenylosuccinate lysase and AMP deaminase (p. 413). Comparative investigations of the adenylosuccinate synthase from various tissues of mammals and

birds have revealed considerable differences in the kinetic properties, apparently as the result of adaption to specific functional requirements. For example, the Michaelis constants vary between 20 and 440 μmol/l for aspartic acid, and 12 to 171 μmol/l for IMP. Metabolic balance sheets have been constructed for the various purine bases and nucleosides of *Drosophila melanogaster* larvae. Adenosine is converted to inosine (57 %) and AMP (43 %), and guanosine to GMP (92 %); the free base guanine, on the other hand, is converted to xanthine (84 %) and GMP (only 16 %); IMP is converted to AMP (65 %), XMP and further products (26 %), and inosine (only 10 %).

The production of the nucleoside diphosphates and triphosphates with their energy-rich bonds comes about by interaction of the processes of substrate chain phosphorylation and oxidative phosphorylation with numerous enzymatic transphosphorylations. Adenylate and guanylate kinases are apparently present in all mammals as isoenzymes that are encoded at different loci. The adenylate kinase of muscle (myokinase) catalyses the reaction $2ADP \leftrightharpoons ATP + AMP$. In this way, it not only provides additional ATP but, in particular, increases the AMP concentration in working muscle, thereby activating phosphofructokinase and eventually glycolysis. The myokinase of the cephalopod *Loligo pealeii* is inhibited by NADH; thus, reduction of NADH in working muscle results in the activation of the enzyme.

Some **unusual features of purine metabolism** are found in the brine shrimp *Artemia salina*, which is unable to synthesize purines de novo. In unfavourable conditions, shrimp embryos in the gastrula stage can withdraw from development by forming dehydrated cysts. These contain only 2–5 % cell fluid and can survive for several years without significant metabolism; upon rehydration, they can continue their development to form nauplius larvae. Protein synthesis resumes after 10 min, and is quickly followed by RNA synthesis. The purine reserves of the cysts exist mainly in the form of diguanosine tetraphosphate (Gp_4G); in addition, compounds such as Gp_3G, Gp_2G, Gp_4A and Ap_4A are found. The Gp_4G is produced from GTP by a GTP:GTP-guanyltransferase, and the Ap_4A is formed from aminoacyl-AMP and ATP by aminoacyl-tRNA-synthetase [39]. The formation of AMP from Gp_4G involves the enzymes dinucleoside tetraphosphatase, GMP reductase, adenylosuccinate synthase and adenylosuccinate lyase. The enzymes IMP dehydrogenase and GMP synthase are needed for the conversion of the IMP to GMP. Some of these enzymes have been isolated and characterized in detail [39]. The Ap_4A produced in the oocytes of *Xenopus laevis* by the action of various aminoacyl-tRNA-synthetases is bound to a specific protein and has signal functions in DNA synthesis and cell division [281]. Production of Ap_4A is also catalysed by the luciferase of the firefly *Photinus pyralis* [96].

12.11.2 Adenylate and Guanylate Cyclases

These two cyclases catalyse the conversion of ATP or GTP to the cyclic 3′, 5′-nucleoside monophosphates cAMP and cGMP, which play an important role as secondary messengers in the transduction of extracellular signals into the cell. Adenylate cyclases stimulated by biogenic amines, peptide hormones, protein hormones or prostaglandins are present in probably all animal cells. Signal transduction between the corresponding specific receptor and the adenylate cyclase is achieved by the G-proteins (p. 309). Several receptor complexes may compete in the cell membrane for the same adenylate cyclase (floating receptor model). The substrate of the cyclase is a complex of ATP and a divalent metal ion, with Mn^{2+} being more effective than Mg^{2+}. Most animal adenylate cyclases are activated by fluoride ions.

As well as in the mammals, **adenylate cyclases** have been found in insects, crustaceans, xiphosurans, platyhelminths, ciliates and Kinetoplastida, and are apparently universally distributed. Several hormone-dependent adenylate cyclases from insects have been functionally characterized but rarely investigated biochemically. The central nervous system of various insect species contains octopamine-dependent adenylate cyclase, which probably corresponds to the catecholamine-dependent enzyme of mammalian brain, but has a tenfold higher affinity for octopamine than for noradrenalin or tyramine. However, insect nervous systems additionally contain specific dopamine- and serotonin-dependent adenylate cyclases. Octopamine-dependent adenylate cyclases are also present in insect fat bodies, where they probably play the same role in glycogen metabolism as does the adrenalin-dependent enzyme of the mammalian liver [87, 272].

The genetic analysis of *Drosophila melanogaster* showed that in addition to enzymes linked to G-proteins or monoamine-receptor-G-protein complexes, there are also Ca^{2+}-calmodulin-

dependent adenylate cyclases [139]. There are only isolated results on hormone-dependent adenylate cyclases of other invertebrate groups: an octopamine-dependent adenylate cyclase was found in the central nervous system of *Limulus*, and serotonin-dependent enzymes have been found in various platyhelminths. The hyperglycaemic hormone of the crustaceans also affects an adenylate cyclase [64]. Like the adenylate cyclases of yeast and *Neurospora*, the enzyme of *Trypanosoma* is only loosely associated with the cell membrane and requires Mn^{2+}. However, it can interact in vitro with G-proteins from the membrane of avian erythrocytes and it is then activated by GTP and Mg^{2+}; activation by Ca^{2+}-calmodulin can also be shown in vitro. Thus, the flagellate enzyme differs from those of the higher vertebrates more in the type of binding to regulatory components than in catalytic properties [248]. The adenylate cyclase of the ciliate *Paramecium tetraurelia* is also not regulated by one specific hormone and is not linked to a GTP-binding protein; it therefore resembles an enzyme found in mammalian spermatozoa [219].

Guanylate cyclases appear to be just as ubiquitous in the animal kingdom. These enzymes resemble the adenylate cyclases in that they require a divalent metal ion, but they differ in that Mn^{2+} can be replaced and fluoride is not stimulatory. Membrane-bound guanylate cyclases in mammals serve as receptors for the atrial natriuretic peptide, and in sea-urchin sperm they serve as receptors for the chemotactic peptides resact and speract (p. 305). Mammals also possess soluble enzymes of this type; they are heterodimers of homologous subunits of 70 and 82 kDa. The membrane-bound guanylate cyclases from the sperm of the sea urchin *Strongylocentrotus*, on the other hand, is a single polypeptide chain of 1104 amino acids and is subdivided into an N-terminal, extracellular domain (485 amino acids), a transmembrane segment (25 amino acids) and a C-terminal domain (594 amino acids). The C-terminal section of 202 amino acids agrees by up to 42% with the soluble guanylate cyclase of bovine lung. In contrast, the membrane-bound enzyme from the sperm of the sea urchin *Arbacia punctulata* has a C-terminal sequence which is very different from that of the soluble mammalian enzyme [162, 247]. The concentration of cGMP is usually 10- to 100-fold lower than that of cAMP; however, higher cGMP concentrations, and high activities of cGMP-dependent protein kinases, are found in the insects. In the fat body of the silkworm *Bombyx mori*, the ratio of guanylate cyclase to adenylate cyclase increases during pupation from 0.5:1 to about 15:1. Extremely high cGMP concentrations of up to 50 µmol/l are found in the spermatophores of crickets of the genera *Gryllus* and *Acheta*, although the guanylate cyclase activity of the accessory sex glands is not unusually high. It is remarkable that such high cGMP values are found only in representatives of a single subfamily of the orthopterans, the Gryllinae. The activities of adenylate cyclases in the sperm of vertebrates and invertebrates are more or less the same, whereas guanylate cyclase activities may be 100- to 1000-fold higher in many invertebrates than in the vertebrates.

12.11.3 Catabolism of Purine Nucleotides and Uric Acid Synthesis

The IMP produced during nitrogen metabolism in uricotelic animals is immediately metabolized further to uric acid. This takes place in the liver and kidneys of birds and reptiles, in the fat body of insects, and in the hepatopancreas of the terrestrial gastropods. Uric acid is also the primary end product of purine metabolism in all other animals, but it is often degraded further (uricolysis). The **formation of uric acid** from IMP or other purine nucleotides requires the cleavage of the phosphate and glycosidic bonds as well as deaminations and oxidations on the purine ring (Fig. 12.26). The phosphate residues are separated by specific 5′-nucleotidases or non-specific

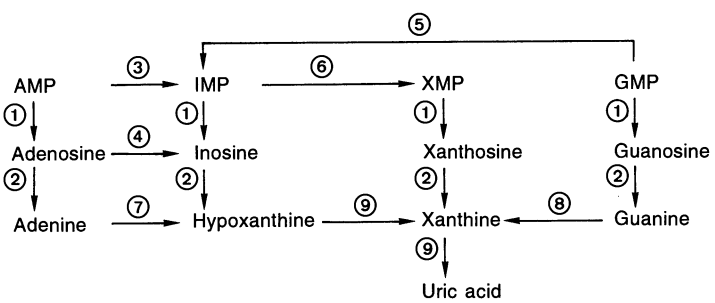


Fig. 12.26. The catabolism of purine nucleotides. *1*, Acidic and alkaline phosphatases or 5′-nucleotidases; *2*, purine nucleoside phosphorylases; *3*, AMP deaminase; *4*, adenosine deaminase; *5*, GMP reductase; *6*, IMP dehydrogenase; *7*, adenine deaminase; *8*, guanine deaminase (guanase); *9*, xanthine oxidase or xanthine dehydrogenase

phosphatases. The phosphorolytic cleavage of the glycosidic bond usually occurs at the level of the nucleoside. Purine nucleoside phosphorylases are undoubtedly widely distributed but their comparative biochemistry is largely unknown [217]. In all vertebrates down to the agnathan *Lampetra planeri*, these enzymes have the seldom found protein structure of a trimer. The deamination and oxidation steps can occur at the levels of the nucleotide, nucleoside or free base (Fig. 12.26). For guanine compounds, deamination occurs either with the nucleotide, by the action of GMP reductase, or with the free base, catalysed by guanine deaminase (guanase), which is apparently widely found but comparatively little investigated [36]. In contrast, for adenine compounds all three possibilities are encountered in the animal kingdom. Whilst adenine deaminase is restricted to the Kinetoplastida, AMP deaminase and adenosine deaminase are widely found. The oxidation of hypoxanthine to xanthine and further to uric acid is brought by xanthine dehydrogenases or oxidases.

AMP deaminase serves, on the one hand, to release ammonia from amino acids in the purine nucleotide cycle (p. 413) and, on the other hand, it plays an important role in the regulation of the adenylate pool. The AMP produced in white vertebrate muscle by the combined action of ATPase and adenylate kinase is immediately metabolized further to IMP by AMP deaminase. Due to the relatively low activity of the 5′-nucleotidase, the IMP is only slowly cleaved to give inosine. AMP deaminase is not present in the adductor muscles of the scallop *Pecten* and other bivalves; hence, the adenylate pool in these organisms remains unchanged even after their death. The mantle muscle of the cephalopods is also rather low in AMP deaminase activity, but has high activities of adenosine deaminase and an AMP-specific 5′-nucleotidase. In this case, AMP is metabolized to inosine via adenosine, a pathway that predominates in other aerobic muscles like red skeletal muscles and heart muscles of the vertebrates [80].

Isoenzymes of AMP deaminase have been reported from all classes of vertebrates; in humans and the rat they are encoded by at least two different genes [158]. The muscle enzymes of the rat, rabbit, chicken, frog *Rana esculenta* and fish *Lucioperca lucioperca* are all tetramers of 280 kDa and contain two zinc atoms per molecule. They require K^+, are activated by ATP and inhibited by P_i and GTP. In the goldfish, the white muscles used only for short periods contain a different enzyme to that found in the continuously active red muscle; the enzyme from the white muscles is allosterically regulated within broad limits. The muscles of all vertebrates have high activities of AMP deaminase; however, in fish the activities in the gills may be much higher than those in the muscles [196]. In the dogfish *Scyliorrhinus canicula*, the erythrocytes contain high activities of an AMP deaminase that is labile but otherwise similar to the muscle enzyme; in other vertebrates the red blood cells contain relatively low activities of this enzyme [195]. The activity in the liver of frogs of the genus *Rana* is tenfold higher than in the liver of other vertebrates. In the chicken and the lizard *Lacerta agilis*, the liver contains two types of AMP deaminase, one of which is activated not only by ATP and ADP but also by GTP; in contrast, GTP is inhibitory to the liver enzymes of the rat and the frog. It may be that the existence of an additional, GTP-activated liver enzyme in the birds and reptiles is associated with their urotelic metabolism [234].

AMP deaminase appears to be completely absent from many mussels and arthropods; it is present at relatively low activities in the muscles of the lobster, in the hepatopancreas of the snail *Helix aspersa* and the mussel *Meretrix meretrix*, and in the body wall of the polychaete *Arenicola cristata*. In contrast, its activity is 18-fold higher in the gut of the earthworm *Lumbricus terrestris* than in the body wall, and in the muscular foot of the edible snail it is 5- to 10-fold higher than in vertebrate muscle. The *Helix* enzyme has a very unusual substrate specificity, in that it is able to act upon ADP, ATP and NADH as well as AMP [235]. The AMP deaminases of the crustaceans *Orconectes limosus* and *Palaemon serratus* also appear to be very different to the vertebrate enzymes in their structural and functional properties [197].

Adenosine deaminases have been demonstrated in vertebrates of all classes, molluscs, annelids and arthropods, and are probably ubiquitous. They deaminate deoxyadenosine as well as adenosine. In the muscles of the vertebrates the activity of adenosine (< U/g fresh weight) is always lower than that of the AMP deaminase (10–50 U/g fresh weight); in the crustaceans, gastropods and bivalves, the two enzymes both have relatively low activities (<1 U/g fresh weight). High activities of adenosine deaminase (12–27 U/g fresh weight) are found only in the muscles of various cephalopods which, in contrast, have low AMP deaminase activities (<1 U/g fresh weight) [80]. Human adenosine deaminase, which is a polypeptide of 41 kDa with 363 amino acids, has been sequenced via the cDNA [263]. The only

invertebrate adenosine deaminase to have been characterized in any detail is that from the hepatopancreas of the edible mussel *Mytilus edulis*; it has a mass of 140 kDa but is otherwise similar in its catalytic properties and inhibitory reactions to the enzymes of the vertebrates and other invertebrates [2].

The oxidation of hypoxanthine to xanthine and of xanthine to uric acid is brought about by enzymes which are named, according to their electron acceptors, as xanthine:O_2 oxidoreductases **(xanthine oxidases, XO)** or xanthine:NAD^+ oxidoreductases **(xanthine dehydrogenases, XDH)**, although these are apparently just different forms of the same enzyme. The form that is active in the cell is possibly XDH, which can be converted to the XO form. Which of the two enzyme activities is detected depends upon the animal species examined. In mammals, XO is consistently found, whereas in birds XDH is found in addition to XO; XO activity is lacking in the columbiformes. Fish, amphibians and reptiles, and also the insects have XDH. Many other invertebrates appear to possess neither of the two activities [41, 279]. The much investigated XO of cow milk is immunologically closely related to the enzyme from bovine liver. The XDH enzymes of many insects are, with the exception of minor differences in specificity, very similar to the XDH from avian liver. In the insects, as in the vertebrates, these enzymes are dimers of 280 kDa with two electron-transport subunits, each with one molybdenum, four iron ions and an FAD [106]. Because of their broad specificities, these enzymes are difficult to distinguish from aldehyde oxidases (AO). The substrate affinities of XDH in the ureotelic mammals and the ammoniotelic fish are much higher than those in the uricotelic birds and reptiles [122, 142]. In mammals, XDH is found mainly in the liver, but in the clawed frog *Xenopus laevis* it occurs only in the kidney [142]. The liver of the camel has an unusually low XDH activity and contains no guanase. Hence, purine catabolism stops at the level of hypoxanthine, which is very suitable for recycling in purine metabolism (salvage pathway). In this way, the camel is adapted to long periods of minimal nutrition [159].

12.11.4 Uricolysis

Uric acid can be catabolized to the level of free ammonia by the sequential reactions of uratoxidase (uricase), allantoinase, allantoicase and urease. In the textbooks, information about the **distribution of the individual enzymes and products in the animal kingdom** (Fig. 12.27) mainly concerns investigations that were carried out before 1950 on individual species. Systematic studies with modern detection methods are urgently needed and would probably alter and enrich our view. Thus, for example, allontoin has been detected in the sponge *Tethya aurantia*, although it is otherwise hardly known among marine invertebrates. Whilst the oxidative degradation of purine to the level of uric acid takes place in the cytoplasm of liver cells, the enzymes of uricolysis are located partially or entirely in the peroxisomes. Uratoxidase in the fish and amphibians is in a soluble form in the peroxisomal matrix, whereas in mammals it is securely bound to the peroxisome core. In some fish, allantoinase is present in the peroxisomal matrix and in the cytosol, and in others it occurs only in the cytosol. Allantoicase is, for example, bound to the outer membrane of the peroxisome in the mackerel but dissolved in the peroxisomal matrix in the sardine [85, 104]. The **uratoxidase** of mammals is a homotetramer with subunits of 303–304 amino acids and these

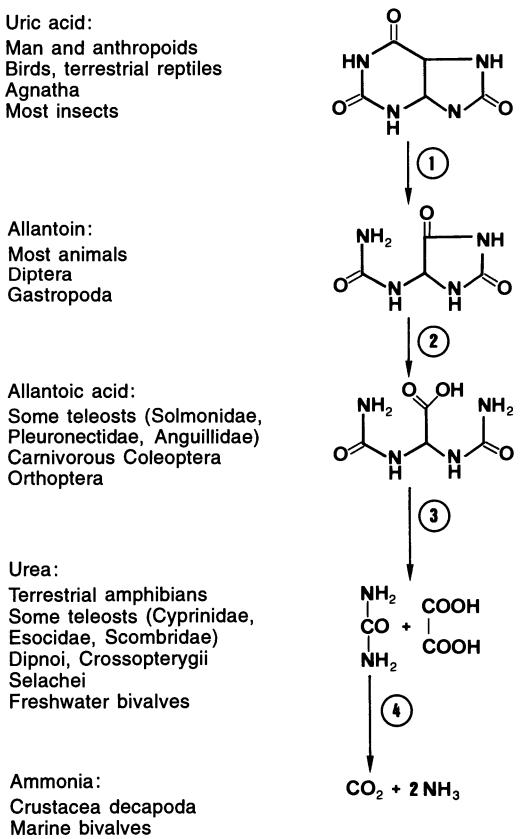


Fig. 12.27. Uricolysis. *1*, Uricase; *2*, allantoinase; *3*, allantoicase; *4*, urease

are significantly homologous to the enzyme from soybean. This enzyme is not found in human liver or, in fact, in any anthropoid, but is present in both the Old World and New World apes. Loss of the enzyme in humans is due to mutation of codons 33-Arg and 187-Arg (CGA or AGA) to a stop codon (TGA) [273]. The selective advantage of the loss of uratoxidase in the evolution of the hominoids may be that uric acid is an effective antioxidant which can absorb singlet oxygen and oxygen radicals, and thereby lower the increased risks of carcinogenesis that accompany the elevated longevity of hominoids [84, 273]. Just why the activity of uratoxidase in the liver of teleosts varies by a factor of 40 without any apparent correlation to systematic position or habitat is not known.

The **allantoinase** in the liver of the mackerel *Pneumatophorus japonicus* is a monomeric protein of 54 kDa, and the **allantoicase** of the sardine *Sardinops melanosticta* is a dimer with subunits of 48 kDa. Both activities are localized in the same protein molecule in the liver of the frog *Rana catesbeiana*, but can be separated by SDS electrophoresis into two different subunits, one of which has a mass of 54 kDa and functions as an allantoinase, and the other has a mass of 48 kDa and shows allantoicase activity. Thus, the separate enzyme proteins of the fish have become associated in the evolution of the amphibians. In the only study of an invertebrate, the two proteins were found to be separate in the mussel *Anadara broughtoni* [166].

12.11.5 Metabolism of Pyrimidine Nucleotides

As in the case of the purines, in the production of pyrimidine nucleotides a distinction can be drawn between de novo synthesis and the use of preformed pyrimidine bases or nucleosides (salvage pathway); the relative importance of the two pathways differs greatly with the species or group. For example, protozoans of the genera *Trypanosoma*, *Leishmania*, *Toxoplasma* and *Plasmodium* can synthesize the pyrimidine ring de novo from aspartate, glutamine and hydrogen carbonate, but the *Trichomonas* species and *Lamblia vaginalis* cannot [99, 109]. In *Drosophila melanogaster*, the pyrimidines taken up during the larval stage suffice for practically the whole live cycle; de novo synthesis is required only during egg production [77]. Some species of the platyhelminths, nemathelminths and dipterans appear to lack the capacity for de novo synthesis; however, in many cases this conclusion is based on the failure to detect the often very unstable carbamylphosphate synthetase, or on the observation that pyrimidine supplementation of the diet promotes growth.

De novo synthesis, when it occurs, is always via the same pathway with orotate as the intermediate (Fig. 12.28); this has been shown in protozoans, platyhelminths, nemathelminths, molluscs, insects, echinoderms and vertebrates of all classes [36, 99]. In all animals, the first three of the six enzymes of UMP synthesis (carbamylphosphate synthetase, CPS; aspartate transcarbamylase, ATC; synthetase dihydroorotase, DHO) and the last two (orotate phosphoribosyltransferase, OPRT; OMP decarboxylase, ODC) are all bound to multifunctional cytoplasmic proteins known as pyr 1–3 and pyr 5, 6; the fourth enzyme (dihydroorotate dehydrogenase DHO-DH) is localized on the outer face of the inner mitochondrial membrane. In lower eukaryotes, like yeast and *Neurospora*, only CPS and ATC are associated in a protein (pyr 1, 2); in the prokaryotes, there are six separate enzyme proteins. Thus, it appears that several fusions between

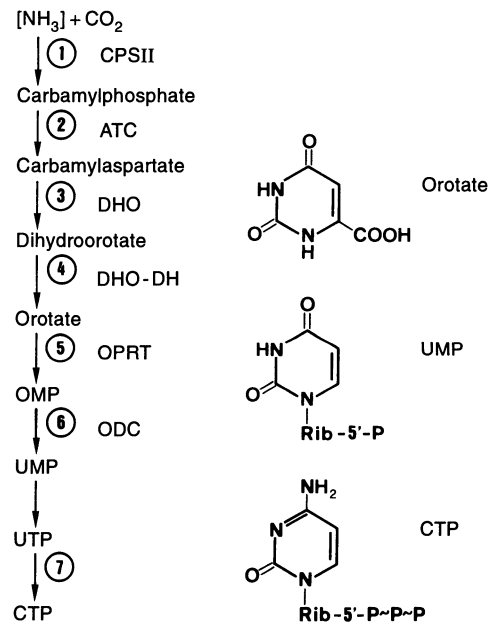


Fig. 12.28. Biosynthesis of the pyrimidine ring. Carbamylphosphate is combined with aspartate to give carbamylaspartate; closing of the ring leads to dihydroorotate, which is oxidized to orotate. Combination with phosphoribosyldiphosphate and cleavage of pyrophosphate gives OMP, which is decarboxylated to UMP. *1*, Carbamylphosphate synthetase II (glutamine dependent); *2*, aspartate transcarbamylase; *3*, dihydroorotase; *4*, dihydroorotate dehydrogenase; *5*, orotate phosphoribosyl-transferase; *6* OMP decarboxylase; *7*, CTP synthetase

genes of the orotate pathway have occurred during evolution. In the "rudimentary" gene of *Drosophila melanogaster*, the three enzyme activities of pyr 1–2 are arranged in the coding sequence of 7168 bp in the order CPS-DHO-ATC. The gene "rudimentary-like" encodes the activities of OPRT and ODC [73, 83]. The golden hamster also has the activities of CPS, DHO and ATC together in one polypeptide of 230 kDa [226].

Very little is known comparatively about the catalytic and regulatory properties of the various enzymes of the orotate pathway. The elasmobranchs possess two types of CPS, of which the mitochondrial type in the liver is active in urea synthesis, and only the cytosolic type, e.g. from the spleen, is involved in pyrimidine synthesis [7]. The ATC sequence contained in the complex pyr 1–3 of the hamster agrees by up to 44 % with the isolated enzyme from *E. coli*. A bacterial strain which, after gene fusion, produced a hybrid enzyme from the N-terminal half of mammalian ATC and the C-terminal half of the bacterial enzyme, showed normal growth [143]. In the Kinetoplastida *Trypanosoma brucei* and *Leishmania mexicana*, the step from dihydroorotate to orotate is catalysed not by a mitochondrial dehydrogenase but by a cytoplasmic, tetrahydrobiopterine-dependent dihydroorotatoxidase; OPRT and ODC, which are usually cytoplasmic, are in this case bound to the glycosomes [99].

The reduction of ribose to deoxyribose takes place on the ribonucleotides. Deoxyribonucleosides, however, can be phosphorylated; thymidine kinase is especially well known in this respect, and is found already in *Tetrahymena*, together with a non-specific nucleoside phosphotransferase. The methylation of dUMP to dTMP is catalysed by thymidylate synthase (TS). *Drosophila* TS is very similar to the mammalian enzyme. In contrast, the TS of *Leishmania tropica*, which is present at a particularly high activity, is linked to tetrahydrofolate reductase (THFR) in a bifunctional polypeptide of 520 amino acids. The N-terminal THFR domains agree in 37 % of their 220 amino acids with the corresponding human enzyme, and show 33 % agreement with that from *E. coli*. The 275-amino-acid C-terminal TS domain is 59 % identical to its human counterpart, and 49 % identical to the *E. coli* enzyme. The two domains are linked by a stretch of 37 amino acids [94]. The 132-kDa TS of the malarial agent *Plasmodium berghei* is much larger than the mammalian enzyme of 68 kDa [179].

As in the case of the purines, the degradation of pyrimidine nucleotides and deoxyribonucleotides occurs by the cleavage of phosphate and glycosidic bonds. The bases that are released are not oxidized like the purine bases, but are reductively catabolized to β-alanine and β-aminoisobutyric acid (Fig. 12.29). However, it should be borne in mind that β-alanine can also arise from the decarboxylation of aspartic acid, and its presence is not necessarily an indication of pyrimidine catabolism.

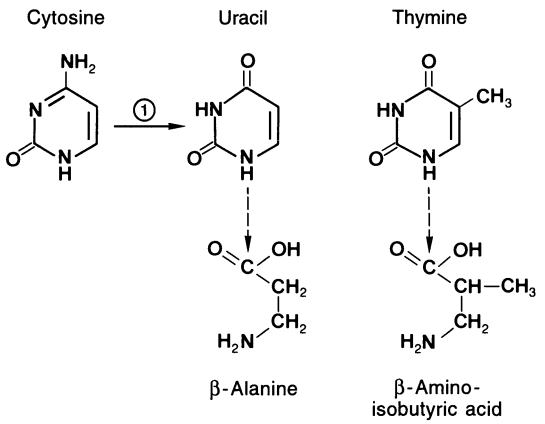


Fig. 12.29. Pyrimidine catabolism. *1*, Cytosine deaminase

References

1. Abe H., Brill R. W. and Hochachka P. W.: Metabolism of L-histidine, carnosine, and anserine in skipjack tuna. Physiol. Zool. 59: 439–450 (1986)
2. Aikawa Y. and Aikawa T.: Adenosine deaminase in the marine mussel Mytilus edulis: purification, characterization and inhibition by a substrate analog, coformycin. Comp. Biochem. Physiol. Pt. B 77: 167–173 (1984)
3. Amuro N et al.: Nucleotide sequence of rat liver glutamate dehydrogenase cDNA. Nucleic Acids Res. 17: 2356 (1989)
4. Andersen S. O.: Sclerotization and tanning of the cuticle. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 3, pp. 59–75. Pergamon Press, Oxford 1985
5. Andersen S. O.: Investigation of an ortho-quinone isomerase from larval cuticle of the American silkmoth, Hyalophora cecropia. Insect Biochem. 19: 803–808 (1989)
6. Andersen S. O.: Oxidation of N-ß-alanyldopamine by insect cuticles and its role in cuticular sclerotization. Insect Biochem. 19: 581–586 (1989)
7. Anderson P. M.: Glutamine-dependent carbamoylphosphate synthetase and other enzyme activities related to the pyrimidine pathway in spleen of Squalus acanthias (spiny dogfish). Biochem. J. 261: 523–529 (1989)

8. Anderson P. M.: Glutamine-dependent urea synthesis in elasmobranch fishes (Commentary). Biochem. Cell Biol. 69: 317–319 (1991)
9. Andersson K. et al.: Purification of the prophenoloxidase from Hyalophora cecropia and four proteins involved in its activation. Insect Biochem. 19: 629–637 (1989)
10. Anthoni U. et al.: Quaternary ammonium compounds in the biosphere: An eyxample of a versatile adaptive strategy (Review). Comp. Biochem. Physiol. Pt. B 99: 1–18 (1991)
11. Ashida M., Kinoshita K. and Brey P. T.: Studies on prophenoloxidase activation in the mosquito Aedes aegypti L. Eur. J. Biochem. 188: 507–515 (1990)
12. Aso Y. et al.: Characterization of haemolymph protyrosinase and a cuticular activator from Manduca sexta (L.). Insect Biochem. 15: 9–17 (1985)
13. Aspan A. et al.: Purification and characterization of a prophenoloxidase activating enzyme from cray fish blood cells. Insect Biochem. 20: 709–718 (1990)
14. Axelrod J. and Saavedra J. M.: Octopamine (Review). Nature 265: 501–504 (1977)
15. Balabaskaran S., Chew S. and Segaran M.: Studies on glutathion S-transferase in molluscs. Comp. Biochem. Physiol. Pt. B 85: 183–192 (1986)
16. Balabaskaran S., Chuen S. S. and Muniandy S.: Glutathione S-transferase from the diamond back moth (Plutella xylostella Linnaeus). Insect Biochem. 19: 435–443 (1989)
17. Balaña-Fouce R. et al.: S-Adenosyl-L-methionine decarboxylase activities in Mytilus edulis. Comp. Biochem. Physiol. Pt. B 83: 459–462 (1986)
18. Barrett J.: Biochemistry of parasitic helminths. Macmillan, London 1981
19. Barrett M.: Incorporation of various putative precursors into cuticular 3,4-dihydroxyphenylacetic acid of adult Tenebrio molitor. Insect Biochem. 20: 645–652 (1990)
20. Batrel Y. and Regnault M.: Metabolic pathways of ammoniogenesis in the shrimp Crangon crangon L.: possible role of glutamate dehydrogenase. Comp. Biochem. Physiol. Pt. B 82: 217–222 (1985)
21. Baum K. F. et al.: Purine deoxynucleoside salvage in Giardia lamblia. J. Biol. Chem. 264: 21087–90 (1989)
22. Bentley B. L.: Nitrogen fixation in termites: fate of newly fixed nitrogen. J. Insect Physiol. 30: 653–655 (1984)
23. Bereiter-Hahn J., Matoltsy A. G. and Richards K. S. (eds.): Biology of the integument, 2 vol. set. Springer, Berlin 1986
24. Bessmann S. P. and Carpenter C. L.: The creatine-creatine phosphate energy shuttle. Annual Rev. Biochem. 54: 831–862 (1985)
25. Billadello J. J. et al.: The complete nucleotide sequence of canine brain B creatine kinase mRNA: homology in the coding and 3' noncoding regions among species. Biochem. biophys. Res. Commun. 138: 392–398 (1986)
26. Bishop S. H., Ellis L. L. and Burcham J. M.: Amino acids in molluscs. In: Wilbur K. M. (ed.): The Mollusca, Vol 1, pp. 243–327. Acad. Press, New York 1983
27. Bodnaryk R. P.: Structure and function of insect peptides. Adv. Insect Physiol. 13: 69–132 (1978)
28. Bowman B. P., Snell T. W. and Cochrane B. J.: Isolation and purification of glutathione S-transferases from Brachionus plicatilis and B. calyciflorus (Rotifera). Comp. Biochem. Physiol. Pt. B 95: 619–624 (1990)
29. Brevet A., Zeitoun Y. and Pradel L. A.: Comparative structural studies of the active site of ATP: guanidine phosphotransferases. The essential tryptic peptide of taurocyamine kinase from Arenicola marina. Biochim. biophys. Acta 393: 1–9 (1975)
30. Bridges R. G.: Choline metabolism in larvae of Pieris brassicae. Insect Biochem. 17: 61–69 (1987)
31. Brophy P. M., Southan C. and Barrett J.: Glutathione transferases in the tapeworm Moniezia expansa. Biochem. J. 262: 939–946 (1989)
32. Brown E. R., Coker III G. T. and O'Malley K. L.: Organization and evolution of the rat tyrosine hydroxylase gene. Biochemistry 26: 5208–12 (1987)
33. Buskin J. N. et al.: The mouse muscle creatine kinase cDNA and deduced amino acid sequences: comparison to evolutionary related enzymes. J. mol. Evol. 22: 334–341 (1985)
34. Butera L. et al.: Isolation of a human liver ligandin cDNA clone and demonstration of sequence homology at ligandin loci in rats and humans. Biochem. biophys. Res. Commun. 142: 986–992 (1987)
35. Caizzi R. et al.: Homologous nuclear genes encode cytoplasmic and mitochondrial glutamine synthetase in Drosophila melanogaster. J. mol. Biol. 212: 17–26 (1990)
36. Campbell J. W. (ed.): Comparative biochemistry of nitrogen metabolism, 2 vol. set. Acad Press, New York 1970
37. Campbell J. W., Smith D. D. jr. and Vorhaben J. E.: Avian and mammalian mitochondrial ammonia-detoxifying systems in tortoise. Science 228: 349–351 (1985)
38. Campbell J. W., Vorhaben J. E. and Smith D. D. jr.: Uricotely: Its nature and origin during the evolution of tetrapod vertebrates. J. exp. Zool. 243: 349–363 (1987)
39. Canales J. et al.: IMP dehydrogenase from Artemia embryos: Molecular forms, purification and properties. Comp. Biochem. Physiol. Pt. B 81: 837–844 (1985)
40. Candy D. J.: Intermediary metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 1–41. Pergamon Press, Oxford 1985
41. Carpani G. et al.: Purification and characterization of mouse liver xanthine oxidase. Arch. Biochem. Biophys. 279: 237–241 (1990)
42. Carvajal N., Kessi E. and Ainol L.: Subcellular localization and kinetic properties of arginase from the liver of Genypterus maculatus. Comp. Biochem. Physiol. Pt. B 88: 229–231 (1987)
43. Carvajal N. et al.: Properties of arginase from the foot muscle of Chiton latus. Comp. Biochem. Physiol. Pt. B 90: 385–388 (1988)
44. Casey C. A. and Anderson P. M.: Submitochondrial localization of arginase and other enzymes asssociated with urea synthesis and nitrogen metabolism, in liver of Squalus acanthias. Comp. Biochem. Physiol. Pt. B 82: 307–315 (1985)
45. Chang L. H. et al.: Characterization of glutathione S-transferases from day-old-chick livers. Biochemistry 29: 744–750 (1990)
46. Chen P. S.: Amino acid and protein metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 177–217. Pergamon Press, Oxford 1985

47. Chew S. F. and Ip Y. K.: Ammoniagenesis in mudskippers Boleophthalmus boddaerti and Periophthalmodon schlosseri. Comp. Biochem. Physiol. Pt. B 87: 941–948 (1987)
48. Christen P. and Metzler D. E. (eds.): Transaminases. Wiley & Sons, New York 1985
49. Clark A. G.: The comparative enzymology of the glutathione S-transferases from non-vertebrate organisms. Comp. Biochem. Physiol. Pt. B 92: 419–446 (1989)
50. Cochran D. G.: Nitrogen excretion in cockroaches. Annual Rev. Entomol. 30: 29–49 (1985)
51. Cochran D. G.: Nitrogen excretion. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 4, pp. 467–506. Pergamon Press, Oxford 1985
52. Cochrane B. J., Morrisey J. J. and LeBlanc G. A.: The genetics of xenobiotic metabolism in Drosophila. IV. Purification and characterization of the major glutathione S-transferase. Insect Biochem. 17: 731–738 (1987)
53. Cohen E. and Gerson U.: Isolation and purification of glutathione S-transferase from the bulb mite Rhizoglyphus robini. Insect Biochem. 16: 449–454 (1986)
54. Cohen E.: Purification of glutathione S-transferase from the flour beetle Tribolium castaneum. Comp. Biochem. Physiol. Pt. B 88: 675–679 (1987)
55. Collett J. I.: Some features of the regulation of the free amino acids in adult Calliphora erythrocephala. J. Insect Physiol. 22: 1395–1403 (1976)
56. Collett J. I.: Characterization of the peptidases of Calliphora: many features allow the utilization of small peptides as an amino acid reservoir. Insect Biochem. 19: 535–547 (1989)
57. Cooper A. J. L.: Biochemistry of sulfur-containing amino acids. Annual Rev. Biochem. 52: 187–222 (1983)
58. Craig III. S. P. et al.: Analysis of cDNA encoding the hypoxanthine-guanine phosphoribosyltransferase (HGPRTase) of Schistosoma mansoni; a putative target for chemotherapy. Nucleic Acids Res. 16: 7087–7101 (1988)
59. Cronkite D. L. and Pierce S. K.: Free amino acids and cell volume regulation in the euryhaline ciliate Paramecium calkinsi. J. exp. Zool. 251: 75–284 (1989)
60. Czapla T. H., Hopkins T. L. and Kramer K. J.: Catecholamines and related o-diphenols in cockroach hemolymph and cuticle during sclerotization and melanization: Comparative studies on the order Dictyoptera. J. comp. Physiol. B 160: 175–181 (1990)
61. D'Aniello A. et al.: Occurrence of D-alanine in the eggs and the developing embryo of the sea urchin Paracentrotus lividus. Comp. Biochem. Physiol. Pt. B 97: 291–294 (1990)
62. Davies J. S.: Amino acids and peptides. Chapman & Hall, London 1985
63. Deaton L. E., Hilbish T. J. and Koehn R. K.: Protein as a source of amino nitrogen during hyperosmotic volume regulation in the mussel Mytilus edulis. Physiol. Zool. 57: 609–619 (1984)
64. Dieberg G. and Sedlmeier D.: Properties of crayfish abdominal muscle adenylate cyclase. Comp. Biochem. Physiol. Pt. B 77: 687–691 (1984)
65. Doonan S. et al.: The complete amino acid sequences of cytosolic and mitochondrial aspartate aminotransferases from horse heart, and inferences on evolution of the isoenzymes. J. mol. Evol. 23: 328–335 (1986)
66. Doumen C. and Ellington W. R.: Mitochondrial arginine kinase from the heart of the horseshoe crab, Limulus polyphemus. I./II. J. comp. Physiol. B 160: 449–468 (1990)
67. Dowton M. and Kennedy I. R.: Glutamine metabolism in fleshfly (Parasarcophaga crassipalpis) tissues. Comp. Biochem. Physiol. Pt. B 85: 593–600 (1986)
68. Doyle J. M. et al.: The amino acid sequence of cytosolic aspartate aminotransferase from human liver. Biochem. J. 270: 651–657 (1990)
69. Drainville G. and Gagnon A.: Osmoregulation in Acanthamoeba castellanii. – I. Variations of the concentration of free intracellular amino acids and of water content. Comp. Biochem. Physiol. Pt. A 45: 379–388 (1973)
70. Dudler N., Yellowlees D. and Miller D. J.: Localization of two L-glutamate dehydrogenases in the coral Acropora latistella. Arch. Biochem. Biophys. 254: 368–371 (1987)
71. Dunlop D. et al.: The presence of D-aspartic acid in rodents and man. Biochem. biophys. Res. Commun. 141: 27–32 (1986)
72. Eales J. G.: The peripheral metabolism of thyroid hormones and regulation of thyroidal status in poikilotherms (Review). Can. J. Zool. 63: 1217–31 (1985)
73. Eisenberg M. et al.: Molecular cloning of the UMP synthase gene rudimentary-like from Drosophila melanogaster. Mol. Gen. Genet. 222: 1–8 (1990)
74. Elkin R. G., Lyone M. L. and Rogler J. C.: Comparative utilization of D- and L-methionine by the white pekin duckling (Anas platyrhynchos). Comp. Biochem. Physiol. Pt. B 91: 325–329 (1988)
75. Ellington W. R.: Phosphocreatine represents a thermodynamic and functional improvement over other muscle phosphagens. J. exp. Biol. 143: 177–194 (1989)
76. Fahey R. C. et al.: Entamoeba histolytica: A eukaryote without glutathione metabolism. Science 224: 70–72 (1984)
77. Falk D. R. et al.: A genetic and dietary study of the physiology of pyrimidine synthesis in Drosophila melanogaster. J. Insect Physiol. 26: 735–740 (1980)
78. Felbeck H.: Occurrence and metabolism of D-aspartate in the gutless bivalve Solemya reidi. J. exp. Zool. 234: 145–149 (1985)
79. Ferré J. et al.: Xanthurenic acid 8-O-ß-glucoside, a novel tryptophan metabolite in eye-color mutants of Drosophila melanogaster. J. biol. Chem. 260: 7509–14 (1985)
80. Fijisawa K and Yoshino M.: Activities of adenylate-degrading enzymes in muscles from vertebrates and invertebrates. Comp. Biochem. Physiol. Pt. B 86: 109–112 (1987)
81. Flink I.L. et al.: Complete amino acid sequence of human thyroxine-binding globulin deduced from cloned DNA: Close homology to the serine antiproteases. Proc. Nat. Acad. Sci. USA 83: 7708–12 (1986)
82. Florkin M. and Scheer B. T. (eds.): Chemical zoology, 11 vol. set. Acad. Press, New York 1967–79
83. Freund J. N. and Jarry B. P.: The rudimentary gene of Drosophila melanogaster encodes four enzymic functions. J. mol. Biol. 193: 1–13 (1987)
84. Friedman T. B. et al.: On the loss of uricolytic activity during primate evolution – I. Silencing of urate oxidase in a hominoid ancestor. Comp. Biochem. Physiol. Pt. B 81: 653–659 (1985)
85. Fujiwara S. et al.: Subcellular distribution of hepatic allontoinase varies among fishes. Comp. Biochem. Physiol. Pt. B 93: 213–215 (1989)

86. Fyhn H. J.: Holoeurohalinity and its mechanism in a cirreped crustacean, Balanus improvisus. Comp. Biochem. Physiol. Pt. A 53: 19–30 (1976)
87. Garcia J. L. et al.: Regulatory properties of adenylate cyclase preparations of pharate adult from the insect Ceratitis capitata. Comp. Biochem. Physiol. Pt. B 80: 437–440 (1985)
88. Gates R. L. V. et al.: Purification and kinetics of delta-1-pyrroline-5-carboxylate reductase from Drosophila melanogaster. Insect Biochem. 13: 39–43 (1983)
89. Geoffrion Y. and Larochelle J.: The free amino acid contribution to osmotic regulation in Acanthamoeba castellanii. Can. J. Zool. 62: 1954–59 (1984)
90. Gilles R., Hoffmann E. K. and Bolis L. (eds.): Volume and osmolality control in animal cells. Springer Verlag, New York 1991
91. Goil M. M.: Study of phosphagens in two trematodes. Z. Parasitenk. 61: 271–275 (1980)
92. Gray K. A., Grossman S. H. and Summers D. D.: Purification and characterization of creatine kinase isozymes from the nurse shark Ginglymostoma cirratum. Comp. Biochem. Physiol. Pt. B 83: 613–620 (1986)
93. Green J. A., Cook N. D. and Manson M. M.: Monoclonal antibodies against rat kidney gamma-glutamyl transpeptidase show species and tissue specificity. Biochem. J. 238: 913–917 (1986)
94. Grumont R. et al.: Bifunctional thymidylate synthase-dihydrofolate reductase from Leishmania tropica: sequence homology with corresponding monofunctional proteins. Proc. Nat. Acad. Sci. USA 83: 5387–91 (1986)
95. Guillou Y. and Robin Y.: Phascoline (N-(3-guanidinopropionyl)-2-hydroxy-n-heptylamine) and phascolosomin (N-(3-guanidinoisobutyryl)-2-methoxy-n-heptylamine), two new guanidino compounds from sipunculid worms. J. biol. Chem. 248: 5668–72 (1973)
96. Guranowski A., Sillero M. A. G. and Sillero A.: Firefly luciferase synthesizes P1,P4-bis(5'-adenosyl) tetraphosphate (Ap4A) and other dinucleoside polyphosphates. FEBS Letters 271: 215–218 (1990)
97. Haas R. C. and Strauss A. W.: Separate nuclear genes encode sarcomere-specific and ubiquitous human mitochondrial creatine kinase isoenzymes. J. Biol. Chem. 265: 6921–27 (1990)
98. Hall R. and Urueña G.: Pharmacology and physiology of monoamine oxidase activity in vertebrates – a comparative study. Comp. Biochem. Physiol. Pt. B 76: 393–397 (1983)
99. Hammons D. J. and Gutteridge W. E.: UMP synthesis in the Kinetoplastida. Biochim. biophys. Acta 718: 1–10 (1982)
100. Haraguchi Y. et al.: Molecular cloning and nucleotide sequence of cDNA for human liver arginase. Proc. Nat. Acad. Sci. USA 84: 412–415 (1987)
101. Harris C. I. and Milne G.: The identification of the N-tau-methyl histidine-containing dipeptide, balenine, in muscle extracts from various mammals and the chicken. Comp. Biochem. Physiol. Pt. B 86: 273–279 (1987)
102. Harris J.: The isolation and characterization of the major glutathione S-transferase from the squid Loligo vulgaris. Comp. Biochem. Physiol. Pt. B 98: 511–515 (1991)
103. Hassan H. F. and Coombs G. H.: Purine phosphoribosyltransferases of Leishmania mexicana mexicana and other flagellate Protozoa. Comp. Biochem. Physiol. Pt. B 82: 773–779 (1985)
104. Hayashi S., Fujiwara S. and Noguchi T.: Degradation of uric acid in fish liver peroxisomes. Intraperoxisomal localization of hepatic allantoicase and purification of its peroxisomal membrane-bound form. J. Biol. Chem. 264: 3211–15 (1989)
105. Hayashi Y. S. et al.: Activities of the three ammonia-forming enzymes in the tissue of the brackish-water bivalve Corbicula japonica. Comp. Biochem. Physiol. Pt. B 83: 721–724 (1986)
106. Hayden T. J. and Duke E. J.: Purification and characterization of xanthine dehydrogenase from Locusta migratoria L. Insect Biochem. 9: 583–588 (1979)
107. Henikoff S. et al.: Multiple purine pathway enzyme activities are encoded at a single genetic locus in Drosophila. Proc. Nat. Acad. Sci. USA 83: 720–724 (1986)
108. Herzog V., Neumüller W. and Holzmann B.: Thyroglobulin, the major and obligatory exportable protein of thyroid follicle cells, carries the lysosomal recognition marker mannose-6-phosphate. Embo J. 6: 555–560 (1987)
109. Heyworth P. G., Gutteridge W. E. and Ginger C. D.: Pyrimidine metabolism in Trichomonas vaginalis. FEBS Letters 176: 55–60 (1984)
110. Hird F. J. R.: The importance of arginine in evolution. Comp. Biochem. Physiol. Pt. B 85: 285–288 (1986)
111. Hoffmann K. H.: Phosphagens and phosphokinases in Tubifex. sp. J. comp Physiol. B 143: 237–243 (1981)
112. Hoskin F. C. G. and Kordik E. R.: Hydrogen sulfide as a precursor for the synthesis of isethionate in the squid giant axon. Arch. Biochem. Biophys. 180: 583–586 (1977)
113. Ichikawa S., Sasaoka T. and Nagatsu T.: Primary structure of mouse tyrosine hydroxylase deduced from its cDNA. Biochem. biophys. Res. Commun. 176: 1610–16 (1991)
114. Iino T. and Shimadate T.: Purification and properties of gut arginase from earthworm Pheretima communissima. Comp. Biochem. Physiol. Pt. B 83: 79–84 (1986)
115. Ishiguro M. et al.: Complete amino acid sequence of rat liver cytosolic alanine aminotransferase. Biochemistry 30: 6048–53 (1991)
116. Itoh N. et al.: Cloning of Drosophila choline acetyltransferase cDNA. Proc. Nat. Acad. Sci. USA 83: 4081–85 (1986)
117. Jackson M. J., Beaudet A. L. and O'Brien W. E.: Mammalian urea cycle enzymes. Annual Rev. Genetics 20: 431–464 (1986)
118. Jacobs M. E.: Role of beta-alanine in cuticular tanning, sclerotization, and temperature regulation in Drosophila melanogaster. J. Insect Physiol. 31: 509–515 (1985)
119. Johnson D. B. and Glynn B. P.: Differential activation and inactivation of the hydrolytic and transpeptidation reactions of γ-glutamyltransferase purified from pyloric caeca of starfish. Comp. Biochem. Physiol. Pt. B 85: 407–412 (1986)
120. Jones M. E.: Pyrimidine nucleotide biosynthesis in animals: Genes, enzymes, and regulation of UMP biosynthesis. Annual Rev. Biochem. 49: 253–279 (1980)

121. Juorio A. V. and Robertson H. A.: Identification and distribution of some monoamines in tissues of the sunflower star, Pycnopodia helianthoides (Echinodermata). J. Neurochem. 28: 573–579 (1977)
122. Kaminski Z. W. and Jezewska M. M.: Xanthine:-NAD+ oxidoreductase in the liver of the teleostean fish Cyprinus carpio. Comp. Biochem. Physiol. Pt. B 80: 371–375 (1985)
123. Kamp G. and Juretschke H. P.: Phosphagen heterogeneity during reproductive cycle in the lugworm, Arenicola marina. Naturwissenschaften 76: 275–276 (1989)
124. Kasschau M. R. et al.: β-Alanine metabolism and high salinity stress in the sea anemone, Budonosoma cavernata. J. comp. Physiol. B 154: 181–186 (1984)
125. Kim P. S. and Arvan P.: Folding and assembly of the newly synthesized thyroglobulin occurs in a pre-golgi compartment. J. Biol. Chem. 266: 12412–18 (1991)
126. Konji V. N., Olembo N. K. and Pearson D. J.: Proline synthesis in the fat body of the tsetse fly Glossina morsitans, and its stimulation by isocitrate. Insect Biochem. 18: 449–452 (1988)
127. Kotze A. C. and Rose H. A.: Purification and properties of the glutathione S-transferases from the larvae of the Australian sheep blobfly, Lucilia cuprina (Wiedemann). Insect Biochem. 19: 703–713 (1989)
128. Kramer K. J. and Hopkins T. L.: Tyrosine metabolism for insect cuticle tanning. Arch. Insect Biochem. Physiol. 6: 279–301 (1987)
129. van Kranen H. J. et al.: Cloning and nucleotide sequence of rat ornithine decarboxylase cDNA. Gene 60: 145–155 (1987)
130. Krueger R. A. et al.: N-ß-alanyldopamine levels and synthesis in integument and other tissues of Manduca sexta (L.) during the larval-pupal transformation. Insect Biochem. 19: 169–175 (1989)
131. Krug E. C., Marr J. J. and Berens R. L.: Purine metabolism in Toxoplasma gondii. J. Biol. Chem. 264: 10601–07 (1989)
132. Kumazawa T. and Suzuki O.: Occurrence of polyamine oxidase in tissues of the catfish. Comp. Biochem. Physiol. Pt. B 88: 727–728 (1987)
133. Kuwamoto S. et al.: Complete nucleotide sequence of cDNA and deduced amino acid sequence of rat liver arginase. J. biol. Chem. 262: 6280–83 (1987)
134. LeBlanc G. A. and Cochrane B. J.: Identification of multiple glutathione S-transferases from Daphnia magna. Comp. Biochem. Physiol. Pt. B 88: 39–45 (1987)
135. Ledley F. D. et al.: Mouse phenylalanine hydroxylase. Homology and divergence from human phenylalanine hydroxylase. Biochem. J. 267: 399–406 (1990)
136. Lenz C., Dahlmann D. L. and Rosenthal G. A.: The effect of L-canavanine and L-canaline on the hemolymph amino acid composition of the tobacco hornworm, Manduca sexta (L.) (Sphingidae). Arch. Insect Biochem. Physiol. 3: 265–275 (1986)
137. Linzen B.: The tryptophan-ommochrome pathway in insects. Adv. Insect Physiol. 10: 117–246 (1974)
138. Lipke H., Sugumaran M. and Henzel W.: Mechanism of sclerotization in dipterans. Adv. Insect Physiol. 17: 1–84 (1983)
139. Livingstone M. S.: Genetic dissection of Drosophila adenylate cyclase. Proc. Nat. Acad. Sci. USA 82: 5992–96 (1985)
140. Lougovoi C. P. and Kyriakidis D. A.: Interconversion of Tetrahymena pyriformis ornithine decarboxylase from inactive to active form by phosphorylation. Biochim. biophys. Acta 996: 70–75 (1989)
141. Lowe P. N. and Rowe A. F.: Aspartate:2-oxoglutarate aminotransferase from Trichomonas vaginalis: Comparison with pig heart cytoplasmic enzyme. Comp. Biochem. Physiol. Pt. B 85: 93–99 (1986)
142. Lyerla T. A. and Fornier P. C. jr.: Xanthine dehydrogenase activity in the clawed frog, Xenopus laevis. Comp. Biochem. Physiol. Pt. B 76: 497–502 (1983)
143. Major J. G. et al.: Molecular evolution of enzyme structure: Construction of a hybrid hamster/Escherichia coli aspartate transcarbamoylase. J. mol. Evol. 28: 442–450 (1989)
144. Male K. B. and Storey K. B.: Tissue specific isozymes of glutamate dehydrogenase from the Japanese beetle, Popillia japonica: Catabolic vs anabolic GDH's. J. comp. Physiol. B 151: 199–205 (1983)
145. Maltiéry Y. and Lissitzky S.: Primary structure of human thyroglobulin deduced from the sequence of its 8448-base complementary DNA. Eur. J. Biochem. 165: 491–498 (1987)
146. Mappouras D. G. and Fragoulis E. G.: Purification and characterization of L-dopa decarboxylase from the white puparia of Ceratitis capitata. Insect Biochem. 18: 369–376 (1988)
147. Matsubasa T. et al.: Structure of the rat argininosuccinate lyase gene: Close similarity to chicken delta-crystallin genes. Proc. Nat. Acad. Sci. USA 86: 592–596 (1989)
148. Matsumoto M. et al.: Amounts of inorganic ions and free amino acids in the hemolymph and ganglion of the giant African snail, Achatina fulica Ferussac. Comp. Biochem. Physiol. Pt. A 48: 465–470 (1974)
149. Matsushima O., Khan H. R. and Saleuddin A. S. M.: Changes in free amino acid concentrations in tissues of the freshwater pulmonate, Helisoma duryi, during hypertonic stress. Comp. Biochem. Physiol. Pt. A 94: 653–657 (1989)
150. Mavrothalassitis G. et al.: Isolation and characterization of cDNA clones encoding human liver glutamate dehydrogenase: Evidence for a small gene family. Proc. Nat. Acad. Sci. USA 85: 3494–98 (1988)
151. Mehta P. K., Hale T. I. and Christen P.: Evolutionary relationships among aminotransferases: Tyrosine aminotransferase, histidinol-phosphate aminotransferase, and aspartate aminotransferase are homologous proteins. Eur. J. Biochem. 186: 249–253 (1989)
152. Meyer D. J. et al.: Theta, a new class of glutathione transferases purified from rat and man. Biochem. J. 274: 409–414 (1991)
153. Mizobuchi M. et al.: High concentration of free trimethyllysine in red blood cells. Biochim. biophys. Acta 1033: 119–123 (1990)
154. Mommsen T. P. and Walsh P. J.: Evolution of urea synthesis in vertebrates: The piscine connection. Science 243: 72–75 (1989)
155. Mondovi B. (ed.): Structure and functions of amine oxidases. CRC Press, Boca Raton USA 1986
156. Montero C. and Llorente P.: Artemia purine phosphoribosyltransferases: Purification and characterization. Biochem. J. 275: 327–334 (1991)
157. Morgan T. D. et al.: N-ß-alanylnorepinephrine: biosynthesis in insect cuticle and possible role in sclerotization. Insect Biochem. 17: 255–263 (1987)
158. Morisaki T., Sabina R. L. and Holmes E. W.: Adenylate deaminase. A multigene family in humans and rats. J. Biol. Chem. 265: 11482–86 (1990)

159. Mura U. et al.: Purine salvage as metabolite and energy saving mechanism in Camelus dromedarius: The recovery of guanine. Comp. Biochem. Physiol. Pt. B 87: 157–160 (1987)
160. Musti A. M. et al.: The complete structure of the rat thyroglobulin gene. Proc. Nat. Acad. Sci. USA 83: 323–327 (1986)
161. Nagata Y. and Akino T.: D-amino acids in mouse tissues are not of microbial origin. Experientia 46: 466–468 (1990)
162. Nakane M. et al.: Molecular cloning and expression of cDNAs coding for soluble guanylate cyclase from rat lung. J. Biol. Chem. 265: 16841–45 (1990)
163. Nakatani Y. et al.: Comparison of human brain and liver glutamate dehydrogenase cDNAs. Biochem. biophys. Res. Commun. 149: 405–410 (1987)
164. Negri A., Massey V. and Williams C. H. jr.: D-Aspartate oxidase from beef kidney. Purification and properties. J. biol. Chem. 262: 10026–34 (1987)
165. Nimmo I. A.: The glutathione S-transferase activity in the pyloric caeca of the rainbow trout, Salmo gairdneri. Comp. Biochem. Physiol. Pt. B 83: 831–835 (1986)
166. Noguchi T., Fujiwara S. and Hayashi S.: Evolution of allantoinase and allantoicase involved in urate degradation in liver peroxisomes. A rapid purification of amphibian allantoinase and allantoicase complex, and its comparison with fish allantoinase and allantoicase. J. biol. Chem. 261: 4221–23 (1986)
167. Nomura K., Suzuki N. and Matsumoto S.: Pulcherosine, a novel tyrosine-derived, trivalent cross-linking amino acid from the fertilization envelope of sea urchin embryo. Biochemistry 29: 4525–34 (1990)
168. Ochiai M. and Ashida M.: Purification of a β-1,3-glucan recognition protein in the prophenoloxidase activating system from hemolymph of the silkworm, Bombyx mori. J. Biol. Chem. 263: 12056–62 (1988)
169. Oda T., Funai T. and Ichiyama A.: Generation from a single gene of two mRNAs that endode the mitochondrial and peroxisomal serine:pyruvate aminotransferase of rat liver. J. Biol. Chem. 265: 7513–19 (1990)
170. Ogawa H. et al.: Human liver serine dehydratase. cDNA cloning and sequence homology with hydroxyamino acid dehydratases from other sources. J. Biol. Chem. 264: 15818–23 (1989)
171. Ohmiya Y., Suzuki S. and Kondo Y.: Isolation and characterization of hagfish thyroid iodoprotein by its non-thyroglobulin nature, very high iodine and carbohydrate contents and low hormone/iodine ratio. Eur. J. Biochem. 182: 11–18 (1989)
172. Ohmiya Y. et al.: Location of dehydroalanine residues in the amino acid sequence of bovine thyroglobulin. Identification of „donor" tyrosine sites for hormonogenesis in thyroglobulin. J. Biol. Chem. 265: 9066–71 (1990)
173. O'Leary K. A. and Tracy J. W.: Purification of three cytosolic glutathione S-transferases from adult Schistosoma mansoni. Arch. Biochem. Biophys. 264: 1–12 (1988)
174. Osanai M and Yonezawa Y.: The occurrence of free methylated amino acids, especially N-trimethyllysine, in the silkworm, Bombyx mori. Comp. Biochem. Physiol. Pt. B 82: 569–574 (1985)
175. Overbaugh J. M. et al.: Purification and preliminary characterization of a monomeric glutathione S-transferase from Tetrahymena thermophila. Arch. Biochem. Biophys. 261: 227–234 (1988)
176. Packard G. C. and Packard M. J.: Patterns of nitrogen excretion by embryonic softshell turtles (Trionyx spiniferus) developing in cleidoic eggs. Science 1049–50 (1983)
177. Palumbo G. et al.: The earliest site of iodination in thyroglobulin is residue number 5. J. Biol. Chem. 265: 8887–92 (1990)
178. Parma J. et al.: Structural organization of the 5' region of the thyroglobulin gene, evidence for intron loss and „exonization" during evolution. J. mol. Biol. 196: 769–779 (1987)
179. Pattanakitsakul S. N., Ratanaphan A. and Ruenwongsa P.: Comparative studies on thymidylate synthetase from P. berghei and mouse reticulocytes. Comp. Biochem. Physiol. Pt. B 80: 583–587 (1985)
180. Pegg A. E.: Recent advances in the biochemistry of polyamines in eukaryotes. Biochem. J. 234: 249–262 (1986)
181. Perryman M. B. et al.: Isolation and sequence analysis of a full-length cDNA for human M creatine kinase. Biochem. biophys. Res. Commun. 140: 981–989 (1986)
182. Peter M. G., Grun L. and Forster H.: CP-MAS-C13-NMR spectra of sclerotized insect cuticle and of chitin. Angew. Chem. Int. ed. engl. 23: 638–639 (1984)
183. Phillips M. A., Coffino P. and Wang C. C.: Trypanosoma brucei ornithine decarboxylase: enzyme purification, characterization, and expression in Escherichia coli. J. Biol. Chem. 263: 17933–41 (1988)
184. Pickett C. P. and Lu A. Y. H.: Glutathione S-transferases: Gene structure, regulation and function. Annual Rev. Biochem. 58: 743–764 (1989)
185. van Pilsum J. F., Stephens G. C. and Taylor D.: Distribution of creatine, guanidino-acetate and the enzymes for their biosynthesis in the animal kingdom. Biochem. J. 126: 325–345 (1972)
186. Pol S. et al.: Nucleotide sequence and tissue distribution of the human mitochondrial aspartate aminotransferase mRNA. Biochem. biophys. Res. Commun. 157: 1309–15 (1988)
187. Poulin R., Larochelle J. and Nadeau P.: Polyamines in Acanthamoeba castellanii: Presence of an unusually high, osmotically sensitive pool of 1,3-diaminopropane. Biochem. biophys. Res. Commun. 122: 388–393 (1984)
188. Powell J. F. et al.: The primary structure of bovine monoamine oxidase type A. Comparison with peptide sequences of bovine monoamine oxidase type B and other flavoenzymes. Biochem. J. 259: 407–413 (1989)
189. Prents P.: Distribution of arginase and other ornithine cycle enzymes in the gut of the earthworm Lumbricus terrestris L. and some physiological and comparative implications. Comp. Biochem. Physiol. Pt. B 93: 509–515 (1989)
190. Preston R. L.: Occurrence of D-amino acids in higher organisms: A survey of the distribution of D-amino acids in marine invertebrates. Comp. Biochem. Physiol. 87B: 55–62 (1987)
191. Prezioso G., Indiveri C. and Bonvino V.: Kinetic characterization of L-glutamate dehydrogenase isolated from Drosophila melanogaster larvae. Comp. Biochem. Physiol. 80B: 1–4 (1985)
192. Psarianos C. G., Marmaras V. J. and Vournakis J. N.: Tyrosine-4-O-ß-glucoside in the Mediterranean fruitfly Ceratitis capitata. Insect Biochem. 15: 129–135 (1985)

193. Pu H. and Young A. P.: The structure of the chicken glutamine synthetase-encoding gene. Gene 81: 169–175 (1989)
194. Queen S. A., van der Jagt D. L. and Reyes P.: Characterization of adenine phosphoribosyltransferase from the human malarial parasite, Plasmodium falciparum. Biochim. biophys. Acta 996: 160–165 (1989)
195. Raffin J. P.: AMP deaminase from dogfish erythrocytes: Purification and some properties. Comp. Biochem. Physiol. Pt. B 75: 461–464 (1983)
196. Raffin J. P.: AMP demaninase from the gill of Salmo gairdnerii Richardson: Effects of anions, cations and buffers. Comp. Biochem. Physiol. Pt. B 79: 499–504 (1984)
197. Raffin J. P. and Thebault M. T.: Purification and partial characterization of an AMP deaminase from the marine invertebrate Palaemon serratus. Comp. Biochem. Physiol. Pt. B 88: 1071–76 (1987)
198. Rajagopalan S. et al.: Cloning and analysis of the rat gamma-glutamyltransferase gene. J. Biol. Chem. 265: 11721–25 (1990)
199. Ramage P. I. N., Rae G. H. and Nimmo I. A.: Purification and properties of the hepatic glutathione S-transferases of the Atlantic salmon (Salmo salar). Comp. Biochem. Physiol. 83: 23–29 (1986)
200. Randall D. J. et al.: Urea excretion as a strategy for survival in a fish living in a very alkaline environment. Nature 337: 165–166 (1989)
201. Ratto A., Shapiro B. M. and Christen R.: Phosphagen kinase evolution: Expression in echinodermes. Eur. J. Biochem. 186: 195–203 (1989)
202. Real M. D. and Ferré J.: Biosynthesis of xanthurenic acid 8-O-ß-D-glucoside in Drosophila. Characterization of the xanthurenic acid:UDP-glucosyltransferase activity. J. Biol. Chem. 265: 7407–12 (1990)
203. Reddy S. R. R., Roustan C. and Benyamin Y.: Purification and properties of two molecular forms of arginine kinase from the adductor muscle of the scallop, Pecten maximus. Comp. Biochem. Physiol. Pt. B 99: 387–394 (1991)
204. Repetto Y. et al.: The gamma-glutamyltranspeptidase of Trypanosoma cruzi. Comp. Biochem. Physiol. Pt. B 87: 73–78 (1987)
205. Rettenmeier R. et al.: Isolation and characterization of the human tyrosine aminotransferase gene. Nucleic Acids Res. 18: 3853–61 (1990)
206. Robert J. and Kobel H. R.: Purification and characterization of cytoplasmic creatine kinase isozymes of Xenopus laevis. Biochem. Genet. 26: 543–555 (1988)
207. Robertson J. D.: Osmotic constituents of some echinoderm muscles. Comp. Biochem. Physiol. Pt A 67: 535–537 (1980)
208. Rogala A., Michalak W. and Gondko R.: Free amino acids in the hemolymph of three species of freshwater crayfish. Comp. Biochem. Physiol. Pt. A 60: 445–446 (1978)
209. Roseghini M. and Alcala A. C.: Occurrence of octopamine in the soft tissues of the gastropod mollusc Peristernia nassatula Lamarck. Experientia 32: 14–15 (1976)
210. Roseland C. R., Kramer K. J. and Hopkins T. L.: Cuticular strength and pigmentation of rust-red and black strains of Tribolium castaneum. Correlation with catecholamine and β-alanine content. Insect Biochem. 17: 21–28 (1987)
211. Rossi A. M. et al.: Muscle-type MM creatine kinase is specifically bound to sarcoplasmic reticulum and can support calcium uptake and regulate local ATP/ADP ratios. J. Biol. Chem. 265: 5258–66 (1990)
212. Ruano A. R. et al.: Some enzymatic properties of NAD-dependent glutamate dehydrogenase of mussel hepatopancreas (Mytilus edulis L.) – Requirement of ADP. Comp. Biochem. Physiol. Pt. B 82: 197–202 (1985)
213. Ruebsamen K., Hume I. D. and von Engelhardt W.: Physiology of the rock hyrax. Comp. Biochem. Physiol. Pt. A 72: 271–277 (1982)
214. Sastre J. et al.: Significance of gamma-glutamyltranspeptidase in exocrine pancreatic amino acid transport. Biochim. biophys. Acta 1065: 213–216 (1991)
215. Saul S. J., Dali H. and Sugumaran M.: Quinone and quinone methide as transient intermediates involved in the side chain hydroylation of N-acetyldopamine derivatives by soluble enzymes from Manduca sexta. Arch. Insect Biochem. Physiol. 16: 123–138 (1991)
216. Schaeffer J. M. and Donatelli M. R.: Characterization of a high-affinity membrane-associated ornithine decarboxylase from the free-living nematode Caenorhabditis elegans. Biochem. J. 270: 599–604 (1990)
217. Schimandle C. M. et al.: Purine nucleoside phosphorylase of the malaria parasite, Plamodium lophurae. J. biol. Chem. 260: 4455–60 (1985)
218. Schmuck R., Kobelt F. and Linsenmair K. E.: Adaptations of the reed frog Hyperolius viridiflavus (Amphibia, Anura, Hyperolidae) to its arid environment. V. Iridiphores and nitrogen metabolism. J. comp. Physiol. B 158: 537–546 (1988)
219. Schultz J. E., Uhl D. G. and Klumpp S.: Ionic regulation of adenylate cyclase from the cilia of Paramecium tetraurelia. Biochem. J. 246: 187–192 (1987)
220. Seiter P., Schultheiss H. and Hanke W.: Osmotic stress and excretion of ammonia and urea in Xenopus laevis. Comp. Biochem. Physiol. Pt. A 61: 571–576 (1978)
221. Seals J. D. and Grossman S. H.: Purification and characterization of arginine kinase from the sea cucumber Caudina arenicola. Comp. Biochem. Physiol. Pt. B 89: 701–707 (1988)
222. Shames S. L. et al.: Trypanothione reductase of Trypanosoma congolense: Gene isolation, primary sequence determination and comparison to glutathione reductase. Biochemistry 27: 5014–19 (1988)
223. Shampengtong L., Wong K. P. and Ho P. C.: N-Acetylation of dopamine and tyramine by mosquito pupae (Aedes togoi). Insect Biochem. 17: 111–116 (1987)
224. Shinbo H.: Cystathionine synthetase and cystathionase activities in the fat body during fifth instar larval stage of the silkworm, Bombyx mori. Insect Biochem. 13: 489–492 (1983)
225. Shinbo H.: An enzyme that catalyses the cleavage of lanthionine in the silkworm, Bombyx mori. Comp. Biochem. Physiol. Pt. B 91: 309–314 (1988)
226. Simmer J. P. et al.: Mammalian dihydroorotase: Nucleotide sequence, peptide sequences, and evolution of the dihydroorotase domain of the multifunctional protein CAD. Proc. Nat. Acad. Sci. USA 87: 174–178 (1990)
227. Smith B. R. et al.: Biosynthesis of asparagine and taurine in the freshwater prawn, Macrobrachium rosenbergii (De Man). Comp. Biochem. Physiol. Pt. B 87: 827–831 (1987)

228. Smith D. D. jr. and Campbell J. W.: Distribution of glutamine synthetase and carbamoyl-phosphate synthetase I in vertebrate liver. Proc. Nat. Acad. Sci. USA 85: 160–164 (1988)
229. Smith E. M. and Watford M.: Molecular cloning of a cDNA for rat hepatic glutaminase. Sequence similarity to kidney-type glutaminase. J. Biol. Chem. 265: 10631–36 (1990)
230. Smith T. J. et al.: Characterization of dopamine N-acetyltransferase from Aedes aegypti: probable role in cuticular sclerotization, but not in ovarian development. Comp. Biochem. Physiol. Pt. B 93: 721–725 (1989)
231. Snell K.: The duality of pathways for serine biosynthesis is a fallacy. Trends biochem. Sci. 11: 241–243 (1986)
232. Söderhäll K. and Smith V. J.: The prophenoloxidase activating system: The biochemistry of its activation and role in arthropod cellular immunity with special reference to crustaceans. In Brehélin M. (ed.): Immunity in invertebrates, pp. 208–223. Springer, Berlin 1986
233. Soederhaell K. et al.: The properties and purification of a Blaberus cranifer plasma protein which enhances the activation of haemocyte prophenoloxidase by a β-1,3-glucan. Insect Biochem. 18: 323–330 (1988)
234. Spychala J. and Marszalek J.: Two forms of AMP deaminase from the lizard (Lacerta agilis) liver. Comp. Biochem. Physiol. Pt. B 83: 169–171 (1986)
235. Stankiewicz A. J.: Direct deamination of AMP, ADP, ATP and NADH by non-specific adenylate deaminase in the foot muscle of the snail Helix pomatia. Biochem. J. 215: 39–44 (1983)
236. Stein L. D., Harn D. A. and David J. R.: A cloned ATP:guanidino kinase in the trematode Schistosoma mansoni has a novel duplicated structure. J. Biol. Chem. 265: 6582–88 (1990)
237. Stickle W. B.: Patterns of nitrogen excretion in the phylum Echinodermata. Comp. Biochem. Physiol. Pt. A 91: 317–321 (1988)
238. Sugumaran M., Saul S. J. and Dali H.: On the mechanism of side chain oxidation of N-beta-alanyldopamine by cuticular enzymes from Sarcophaga bullata. Arch. Insect Biochem. Physiol. 15: 255–269 (1990)
239. Sugumaran M. et al.: Oxidation of 3,4-dihydroxybenzyl alcohol: A sclerotizing precursor for the cockroach ootheca. Arch. Insect Biochem. Physiol. 16: 31–44 (1991)
240. Tabor C. W. and Tabor H.: Polyamines. Annual Rev. Biochem. 53: 749–790 (1984)
241. Tada M. et al.: Cloning and expression of a cDNA encoding mouse kidney D-amino acid oxidase. Gene 90: 293–297 (1990)
242. Takagi Y., Omura T. and Go M.: Evolutionary origin of thyroglobulin by duplication of esterase gene. FEBS Letters 282: 17–22 (1991)
243. Tang P. and Lo H. S.: Pentatrichomonas hominis: Purine salvage pathway. Comp. Biochem. Physiol. Pt. B 90: 419–425 (1988)
244. Teller J. K.: Purification and some properties of glutamate dehydrogenase from the mealworm fat body. Insect Biochem. 18: 101–106 (1988)
245. van Thoai N., Robin Y. and Guillou Y.: A new phosphagen, N'-phosphorylguanidinoethylphospho-O-(⁻-N,N-dimethyl)serine (phosphothalassemine). Biochemistry 11: 3890–95 (1972)
246. Thomas B. R. et al.: A trypsin-solubilized laccase from pharate pupal integument of the tobacco hornworm, Manduca sexta. Insect Biochem. 19: 611–622 (1989)
247. Thorpe D. S. and Garbers D. L.: The membrane form of guanylate cyclase. Homology with a subunit of the cytoplasmic form of the enzyme. J. Biol. Chem. 264: 6545–49 (1989)
248. Torruella M. et al.: Trypanosoma cruzi adenylate cyclase activity. Purification and characterization. Biochem. J. 234: 145–150 (1986)
249. Tsai C. S. et al.: Comparative studies of glutathione reductase and lipoamide dehydrogenase. Comp. Biochem. Physiol. Pt. B 90: 335–339 (1988)
250. Tsuji S. and Kanazawa S.: Chicken ornithine transcarbamylase: Its unexpected expression. Biochem. Genetics 25: 259–266 (1987)
251. Tutic M. et al.: Cloning and sequencing of mammalian glutathione reductase cDNA. Eur. J. Biochem. 188: 523–528 (1990)
252. Vanveldhoven P. P., Brees C. and Mannaerts G. P.: D-aspartate oxidase, a peroxisomal enzyme in liver of rat and man. Biochim. biophys. Acta 1073: 203–208 (1991)
253. van Waarde A.: Biochemistry of non-protein nitrogenous compounds in fish including the use of amino acids for anaerobic energy production. Comp. Biochem. Physiol. Pt. B 91: 207–228 (1988)
254. Waite J. H.: The phylogeny and chemical diversity of quinone-tanning glues and varnishes. Comp. Biochem. Physiol. Pt. B 97: 19–29 (1990)
255. Waraska J. C., Proos A. A. and Tillinghast E. K.: The utilization of carbohydrate reserves with the transition to ureotelism in the earthworm Lumbricus terrestris L. Comp. Biochem. Physiol. Pt. A 65: 255–257 (1980)
256. Watford M. and Smith E. M.: Distribution of hepatic glutaminase activity and mRNA in perivenous and periportal rat hepatocytes. Biochem. J. 267: 265–267 (1990)
257. Watts D. C.: Evolution of phosphagen kinases. In: Schoffeniels E. (ed.): Molecular evolution, Vol. 2, pp. 150–173. North-Holland, Amsterdam 1971
258. Watts S. A. et al.: Determination of polyamines in digestive and reproductive tissues of adult Asterias vulgaris (Echinodermata: Asteroidea). Comp. Biochem. Physiol. Pt. B 88: 309–312 (1987)
259. Webb K. L., Schimpf A. L. and Olmon J.: Free amino acid composition of scyphozoan polyps of Aurelia aurita, Chrysaora quinquecirrha and Cyanea capillata of various salinities. Comp. Biochem. Physiol. Pt. B 43: 653–663 (1972)
260. Weeda E., de Kort C. A. D. and Beenakkers A. M. T.: Oxidation of proline and pyruvate by flight muscle mitochondria of the Colorado beetle, Leptinotarsa decemlineata Say. Insect Biochem. 10: 305–311 (1980)
261. Weeda E. et al.: Proline synthesis in fat body of Leptinotarsa decemlineata. Insect Biochem. 10: 631–636 (1980)
262. Whitton P. S., Strang R. H. C. and Nicholson R. A.: The distribution of taurine in the tissues of some species of insects. Insect Biochem. 17: 573–577 (1987)
263. Wiginton D. A. et al.: Complete sequence and structure of the gene for human adenosine deaminase. Biochemistry 25: 8234–44 (1986)

264. Wilson R. P. and Poe W. E.: Nitrogen metabolism in channel catfish, Ictalurus punctatus – III. Relative pool sizes of free amino acids and related compounds in various tissues of the catfish. Comp. Biochem. Physiol. Pt. B 48: 545–556 (1974)
265. Wirz T. et al.: A unique chicken B-creatine kinase gene gives rise to two B-creatine kinase isoproteins with distinct N termini by alternative splicing. J. Biol. Chem. 265: 11656–66 (1990)
266. Wittich R. M. and Walter R. D.: A novel type of putrescine (diamine)-acetylating enzyme from the nematode Ascaris suum. Biochem. J. 260: 265–269 (1989)
267. Wolf J. and Kobel H. R.: Creatine kinase isozymes in pipid frogs – Their genetic bases, gene expressional differences, and evolutionary implications. J. exp. Zool. 234: 471–480 (1985)
268. Wong C. C., Chiu K. W. and Choy Y. M.: Metabolism of monoiodotyrosine in lower vertebrates: with emphasis on snakes. Comp. Biochem. Physiol. Pt. B 74: 739–742 (1983)
269. Wood C. M., Munger R. S. and Toews D. P.: Ammonia, urea and H-ion distribution and the evolution of ureotelism in amphibians. J. exp. Biol. 144: 215–233 (1989)
270. Wothe D. D., Charbonneau H. and Shapiro B. M.: The phosphocreatine shuttle of sea urchin sperm: Flagellar creatine kinase resulted from a gene triplication. Proc. Nat. Acad. Sci. USA 87: 5203–07 (1990)
271. Wright C. E., Tallan H. H. and Lin Y. Y.: Taurine: Biological update. Annual Rev. Biochem. 55: 427–453 (1986)
272. Wright M. S., Cook B. J. and Holman G. M.: Adenylate cyclase in the insect Leucophaea maderae: Biochemical properties of the enzyme from hindgut tissue. Comp. Biochem. Physiol. Pt. B 85: 413–418 (1986)
273. Wu C. et al.: Urate oxidase: Primary structure and evolutionary implications. Proc. Nat. Acad. Sci. USA 86: 9412–16 (1989)
274. Wyss M. et al.: Mitochondrial creatine kinase from chicken brain. Purification, biophysical characterization, and generation of heterodimeric and heterooctameric molecules with subunits of other creatine kinase isoenzymes. J. Biol. Chem. 265: 15900–08 (1990)
275. Xue G. P. and Snoswell A. M.: Comparative studies on the methionine synthesis in sheep and rat tissues. Comp. Biochem. Physiol. Pt. B 80: 489–494 (1985)
276. Yago M. and Kawasaki H.: The identification of five N-acyldopamine glucosides in the left colleterial gland of the praying mantid, Hierodula patellifera Serville. Insect Biochem. 14: 487–489 (1984)
277. Yamaoka L. H. and Skinner D. M.: Free amino acid pools in muscle and hemolymph during the molt cycle of the land crab, Gecarcinus lateralis. Comp. Biochem. Physiol. 55A: 129–134 (1976)
278. Yokota S. D. and Shoemaker V. H.: Xanthine excretion in a desert-scorpion, Paruroctonus mesaensis. J. comp. Physiol. B 142: 423–428 (1981)
279. Zakrzewska B. and Jezewska M. M.: Comparison of xanthine:NAD+ oxidoreductase from liver of toad Bufo viridis and other vertebrates. Comp. Biochem. Physiol. Pt. B 94: 361–365 (1989)
280. Zatta P.: The relationship between plasma proteins and intracellular free amino acids during osmotic regulation in Carcinus maenas. J. exp. Zool. 242: 131–136 (1987)
281. Zourgui L. et al.: Purification, immunological and biochemical characterization of an Ap4A binding protein from Xenopus laevis oocytes. Nucleic Acids Res. 16: 2913–29 (1988)

13 The Structure and Metabolism of Carbohydrates

13.1 Sugars and Sugar Derivatives

Compared with the variety of carbohydrates in plants, relatively few sugars or sugar derivatives are regularly found in animals either free or as components of more complex compounds. However, it is possible that sugars originating from plants in the diet are transiently retained in animals and distort the normal **sugar spectrum.** Approximately a dozen sugars and sugar derivatives are regularly found in animals: the pentoses D-ribose and 2-deoxy-D-ribose; the hexoses D-glucose, D-galactose, D-mannose, D-fructose and L-fucose; the uronic acids D-glucuronic acid and L-iduronic acid; and the hexosamines D-glucosamine and D-galactosamine. In addition, D-erythrose, D-ribulose, D-xylulose and D-sedoheptulose in the form of their phosphoric acid esters are intermediates of the pentose phosphate pathway.

Other sugars are found occasionally. In contrast to the occurrence in plants, D-galacturonic acid is very rare in animals; it has been detected, for example, in mammalian brain and in secretions of the elasmobranchs and gastropods. The pentoses L-arabinose and D-xylose and the deoxyhexose L-rhamnose are found in several invertebrate polysaccharides; D-xylose binds the polysaccharide chain to the protein component in various mucopolysaccharides. D-arabinose has been found in unusual nucleosides of the sponge *Thetya (Cryptothetya) crypta* (p. 741). Whereas D-galactose is ubiquitously distributed, e.g. as part of lactose and many glycoproteins, L-galactose has so far been found only in the galactans of several snails. 6-Desoxyglucose is present in the toxin of the holothurian *Actinopyga agassizi*, and free D-glucuronic acid or its lactone is found in the defence secretions of the cockroach *Eurycotis decipiens*. The distribution of ascarylose (3,6-dideoxy-L-mannose) (Fig. 13.1) is quite unique; apart from in the bacteria, it is found only in *Ascaris* and its relatives, and occurs in particular in a lipid material that was previously known as ascaryl alcohol. This material is in fact a mixture of so-called ascarosides in which the ascarylose is bound glycosidically to a mono- or divalent fatty alcohol with a chain length of 22–37 carbon atoms; the divalent alcohols may also be esterified with a short-chain fatty acid. Such ascarosides are found in all organs, but especially in the uterus and the ovary. The presence of ascarosides in the egg membrane is responsible for the unusual resistance of *Ascaris* eggs to aggressive

media. Of the **sugar alcohols** resulting from the reduction of sugars, D-ribitol is ubiquitously found as a component of flavin nucleotides; D-sorbitol is present at high concentrations in the seminal fluid of many mammals and, in addition to glycerol, it acts as an antifreeze agent in overwintering insect eggs and larvae. It could be shown in the cotton beetle *Anthonomus grandis* that the supply of sugar to starving animals resulted in the appearance of the corresponding sugar alcohols in the haemolymph, in particular D-ribitol, D-xylitol, D-arabitol and D-rhamnitol. In addition to myo-inositol, erythritol, threitol, ribitol and xylitol, the lens of the human eye contains hepitol and octitol [57, 244].

Inositols are polyalcohols that are comparable to alcohols (Fig. 13.1). Myo-inositol is universally found as a component of lipids; the free form is found only at high concentrations in the seminal fluid of the mammals. The pentakis-, tetrakis- and hexakisphosphates of myo-inositol in the erythrocytes of birds and reptiles have the same function of modulating haemoglobin oxygen affinity as do the other organophosphates found in other vertebrates. More recently, myo-1,4,5-trisphosphate has been recognized as a secondary messenger. Glycosidic bonds involving myo-inositol were originally known only in plants, but in fact also occur in animals, e.g. 6-O-β-D-galactosyl-myo-inositol in the milk glands of the rat. The isomer scyllo-inositol derives its name from the shark *Scyllium canicula* in which it was found; it is widely distributed in the animal kingdom but occurs only in a covalently bound form. All mammalian organs contain more free myo-inositol than scyllo-inositol; in contrast, scyllo-inositol predominates in the kidney and liver of the elasmobranchs but not in their nervous system. The two isomers can be interconverted via the common 2-carbonyl compound (myo-scyllo-inosose), which is found together with the two inositols in elasmobranch organs. 1-Methyl-scyllo-inositol was discovered in the edible mussel *Mytilus edulis* and is also known as "mytilitol" (Fig. 13.1); it is additionally found in the ascidian *Ciona intestinalis*. chiro-Inositol (Fig. 13.1) has been found, together with myo-inositol and scyllo-inositol, only in the fat bodies and haemolymph of the cockroach *Periplaneta americanus [3, 57]*.

Of particular interest for comparative biochemistry are the **sialic acids**, which arise as derivatives of neuraminic acid by condensation of mannosamine-6-phosphate with phosphoenolpyruvate. The sialic acids are N-acetyl- or N-glycolylneuraminic acids which can be substituted at C-4, -7, -8 or -9 with acetyl, glycolyl, lactyl, methyl, sulphate or phosphate residues (Fig. 13.2). Thus, the number of possible structural variants is very large, and so far there are more than 30 known sialic acids [198, 238]. A completely new form of sialic acid, a deaminosialic acid, was recently found in eggs of the rainbow trout *Salmo gairdneri* [129]. The sialic acids are mainly found in a bound form, especially in glycolipids (gangliosides) and glycoproteins, but also in proteoglycans and oligosaccharides from mammalian milk and urine. They commonly form the non-reducing ends of oligosaccharide structures in which they are bound α(2–3)-, α(2–4)- or α(2–6)- glycosidically to galactose or N-acetylgalactosamine, and less often to N-acetylglucosamine or glucose. Apart from in the bacteria, sialic acids are found at internal positions of oligosaccharide chains only in echinoderms, for example in the structure

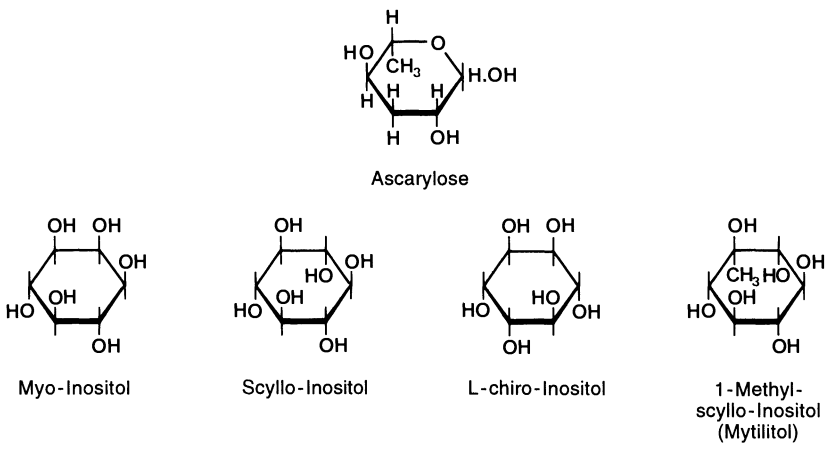
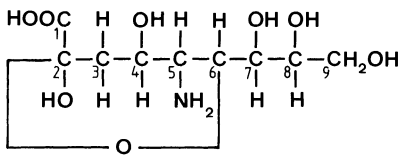


Fig. 13.1. Ascarylose and the animal inositols



Neuraminic acid

N-Acetyl-Neu	Trypanosoma, Echinoderms, Enteropneusta, Acrania, all vertebrates
N-Acetyl-4-0-acetyl-Neu	Ornithorhynchus milk and higher mammals
N-Acetyl-8-0-acetyl-Neu	Mammals
N-Acetyl-9-0-acetyl-Neu	Fish, birds, mammals
N-Acetyl-8,9-0-diacetyl-Neu	Mammals
N-Acetyl-7,8,9-0-triacetyl-Neu	Mammals
N-Acetyl-4-0-methyl-Neu	Mammals
N-Acetyl-9-0-lactyl-Neu	Mammals
N-Glycolyl-Neu	Trypanosoma, Echinoderms, all vertebrates
N-Glycolyl-4-0-acetyl-Neu	Mammals
N-Glycolyl-7-0-acetyl-Neu	Mammals
N-Glycolyl-9-0-acetyl-Neu	Echinoderms, Mammals
N-Glycolyl-7,9-0-diacetyl-Neu	Mammals
N-Glycolyl-8,9-0-diacetyl-Neu	Mammals
N-Glycolyl-7,8,9-0-triacetyl-Neu	Mammals
N-Glycolyl-8-0-methyl-Neu	Only echinoderms
N-Glycolyl-8-0-sulpho-Neu	Echinoderms, mammals

Fig. 13.2. Sialic acids

Fucα(1–4)NeuGc-. Series of several α(2–8)-linked sialic acid residues are widely found in glycoproteins and gangliosides. In correspondence with their presence in a wide variety of substances in the body, the sialic acids have a wide range of functions. They are responsible for the negative charges of cells and molecules in solution and thereby prevent their agglutination, aggregation and precipitation; they regulate the degradation of blood cells and serum proteins; they mask antigen sites and themselves form recognition sites [238].

The **distribution of the sialic acids amongst different organisms** is rather strange. They are undoubtedly present in bacteria but appear to be absent from fungi, algae and higher plants. They are found in *Trypanosoma cruzi* but not in amoebae or malarial agents. Only a few metazoans possess sialic acids. They have been detected in all the tissues of all echinoderms examined and in the enteropneustan *Dolichoglossus kowalevskii*. In the chordates, they are found in *Branchiostoma* and in all vertebrates from the agnathans to the mammals, but not in the tunicates. Sialic acids are lacking in the sponges and cnidarians, the annelids and arthropods, the molluscs, and many small phyla. Some exceptional positive results in crustaceans and cephalopods may be due to exogenous sialic acid or bacterial contamination [238, 239]. Little can be said for certain about the **evolution of the sialic acids**. Did they really appear initially with the echinoderms, as has been claimed [238]? There are in fact unique sialic acids in the echinoderms, e.g. 8-O-methyl derivatives (Fig. 13.2). Here, perhaps, there was some "experimentation" at the beginning of sialic acid evolution, as has been suggested for the evolution of other classes of substances, e.g. the fatty acids, sterols, bile acids and bile alcohols. However, the N-glycolyl derivatives are already present in the echinoderms; in fact, they predominate, despite the fact that they are synthesized via N-acetyl compounds. Consequently, the required monooxygenase must be present. The "invention" of sialic acids during evolution may be based upon the fact that these negatively charged molecules can be easily attached to the ends of carbohydrate chains by specific sialyltransferases, and just as easily detached by neuraminidases, and upon the advantages this brings to the very varied functions of the oligosaccharides bound to glycolipids or glycoproteins. The

Trehalose	Glcα(1-1)αGlc
Maltose	Glcα(1-4)Glc
Maltotriose	Glcα(1-4)Glcα(1-4)Glc
Lactose	Galβ(1-4)Glc

Fig. 13.3. Oligosaccharides found in animals (trivial names and simplified formulae)

O-substitution of the neuraminic acids modifies their susceptibility to attack by neuraminidases [238].

There are only a few **oligosaccharides** present at high concentrations in the tissues and body fluids of animals: trehalose, maltose and the corresponding tri- and tetrasaccharides, as well as the milk sugar lactose (Fig. 13.3). Trehalose was discovered in 1858 in the oriental pharmaceutical product "trehala", which consists of the larval cocoons of beetles from the genus *Larinus*. The animal origin of this sugar and its identity with the fungal sugar "mycose" was already clear at the time, but nevertheless, trehalose was for a long time considered to be predominantly a plant product. It was only in 1956–1957 that in several laboratories it was simultaneously recognized to be the typical blood sugar of insects. Soon afterwards, trehalose was identified in representatives of almost all phyla. As well as in the insects, high concentrations of trehalose are found in the tissues and fluids of the nematodes and acanthocephalans [13].

Glycosides in a restricted sense, i.e. sugars linked to non-sugar substances (aglycones), can be found in any number in almost all animals, where they occur as products of secondary metabolism. Unwanted substances can be converted to a hydrophilic, more easily excretable form by glycosidation. Glycosides also function as transport forms of somewhat insoluble precursors, e.g. in the sclerotization of insect cuticulae and in the synthesis of pigments and defence substances.

13.2 The Carbohydrate Spectrum of Cells and Extracellular Fluids

Carbohydrates are important **substrates for the functional and structural metabolism** of cells. Thus, in cells as well as in the extracellular fluids, which in fact make up the external environment of the majority of cells, carbohydrates constitute a considerable proportion of the low molecular weight organic compounds. Their concentrations are limited in particular by osmotic considerations. The seminal fluid (seminal plasma), as the nutritional medium for the spermatozoa, and nutritional solutions such as milk and honey are also especially rich in carbohydrates.

Carbohydrates and their derivatives also play an important role in the **cold tolerance** of insects and some anurans, and in the ability of the nematodes to withstand large water losses. The frost-tolerant insects accumulate various substances in their tissues and extracellular fluids: glycerol at concentrations of up to 3 mol/l [53, 164], sorbitol [71] and trehalose [13] very often predominate, but other sugar alcohols, such as erythritol, threitol or mannitol, sugars like glucose or fructose, and amino acids like proline and alanine contribute along with antifreeze proteins (p. 208) to the cold hardiness of the insects. It was first discovered in 1982 that some frog species can survive the freezing of 34–41 % of their total body water, i.e. all their extracellular fluids. The tree frog *Hyla versicolor* accumulates glycerol as a protective substance in the blood, up to a concentration of 423 mmol/l, and the wood frog *Rana sylvatica* makes use of glucose, which accumulates to a concentration of 185 mmol/l in the blood, 388 mmol/l in the liver, 121 mmol/l in the kidney, but only 27 mmol/l in the leg muscles [256]. Drought-resistant nematodes, such as *Anguina tritici, Aphelenchus avenae* and *A. dipsaci*, can increase their trehalose concentrations to values of more than 9 % of their dry weight.

The predominant **blood sugar** of most vertebrates is D-glucose. The average blood sugar level is correlated to the metabolic rate and is therefore higher in the mammals and birds than in the poikilothermic vertebrates (Table 13.1). The foetal blood of the ungulates contains free fructose at markedly high concentrations, whereas the glucose concentration is significantly reduced. The fructose content of foetal blood appears to be related more to the type of placenta than to the systematic position of the species. The prosimian *Galago senegalensis*, which, like the ungulates but in contrast to the other primates, has an epithelio-chorionic placenta, has more fructose than glucose in the foetal blood, amnion and allantoic fluid. In the **invertebrates**, the reduction methods used previously for the determination of the blood sugar level almost always gave too high values. Only chromatographically or enzymatically obtained results are of value here. The predominant sugar is again glucose but its concentration rarely exceeds 2 mmol/l (Table 13.1). The haemolymph of the decapod crustaceans contains a complex spectrum of low molecular weight carbohydrates. For example, in the crayfish *Orconectes limosus* glucose as well as galactose, fruc-

tose, trehalose, maltose, maltotriose and maltotetraose are found in proportions which vary greatly with the season and moulting stage [251]. The blood-sugar level of the insects is almost always higher than that of the other groups of invertebrates. Trehalose usually predominates and can reach concentrations as high as 70 g/l (e.g. in adult females of the aphid *Megoura viciae*). Although glucose in the haemolymph of insects rarely exceeds 10 % of the total carbohydrate, it is consistently higher in concentration than in other invertebrates (Table 13.1). A rather unusual sugar spectrum for an insect is found in the honey bee, where the haemolymph contains glucose, fructose and trehalose in more-or-less equal proportions. The blood sugar of the larvae of the botfly *Gasterophilus intestinalis*, which parasitizes the stomach of the horse, contains predominantly fructose.

Spermatozoa have no nutritional reserves of their own and are dependent for their survival and motility on energy-yielding substrates in the surrounding fluid. The **seminal fluid** of the higher mammals (Placentalia) contains, above all, fructose as a nutrient for spermatozoa; in addition sorbitol and myo-inositol are present. Glucose is almost entirely absent from the seminal plasma of most mammals. The exceptions are humans, the rat, mouse, guinea-pig and the field vole *Microtus agrestis*, which have both fructose and glucose. The main nutrient for the spermatozoa of the marsupials is N-acetylglucosamine; this sugar is also found in the seminal fluid of several eutherians, but is not used for example, by human sperm. In some marsupials, e.g. the Tasmanian devil *Sarcophilus harisii*, the dasyure *Dasyurus vivarrinus*, the wombat *Phascolomys* sp. and the North American opossum *Didelphis virginiana*, the main carbohydrate in the seminal fluid is glycogen [227].

The main nutritional constituents of the **milk** of the Placentalia are casein, milk fats and the sugar lactose; these are present in species-specific proportions (Table 13.2). The highest lactose values are found in the milk of humans and other

Table 13.1. The concentrations (mmol/l) of low molecular weight carbohydrates in blood plasma and haemolymph. The disaccharide and oligosaccharide concentrations are given as mmol/l glucose equivalents

Species [Ref.]	Glucose	Trehalose	Fructose
Oligochaeta			
Eisenia foetida [3]	1.3	–	–
Echiurida			
Urechis caupo [3]	2.4	–	–
Mollusca			
Helix pomatia [223]	1.2	–	–[a]
Octopus dofleini [3]	2.3	–	–
Crustacea			
Porcellio scaber [20]	Tr	Tr	–
Orconectes limosus [251]			
October	0.26	0.16	0[b]
March	0.63	0.33	0.45
Chelicerata			
Hexathele hochstetteri [20]	0.36	0.26	–
Urodacus abruptus [20]	0.22		–
Thysanura			
Lepisma saccharina [20]	0.24	0.52	–
Insecta			
Uropetala carovei [20]	0.47	4.2	–
Locusta migratoria [20]	Tr	43	–
Schistocerca gregaria [3]	1.3	39	–
Periplaneta americana [20]	3.1	38	–
Heimeidiens thoracica [20]	39	49	–
Gasterophilus intestinalis			
larva [81]	0.56	–	10–16
Phormia regina [3]	33	7–167	–
Chalcophora mariana [81]	0	260	–
Galleria melonella [81]	1.2	94	–
Hyalophora cecropia [81]	0	36–64	–
Apis mellifica [35]	26	48	12
[7]	34	50	36
Vertebrates			
Mammals	3–11		
Birds	7–12		
Reptiles	2– 9		
Amphibians	0.5–1.6		
Teleosts	1– 5		

[a] + 0.13 galactose

[b] + 0.03 galactose + 0.014 maltose + 0.027 maltose disaccharide + 0.033 maltose tetrasaccharide + 0.13 hexurinoc acid + 0.017 sugar phosphates

Tr, trace amount

Table 13.2. The sugar, fat and protein contents (%) of milk [81]

Species		Sugar	Fat	Protein
Ape	*Macaca mulatta*	7.0	4.0	1.1
Human	*Homo sapiens*	7.0	3.8	0.4
Kangaroo	*Mageleia rufica*	6.7	3.4	2.3
Horse	*Equus caballus*	6.2	1.9	1.3
Pig	*Sus scrofa*	5.5	6.8	2.8
Cow	*Bos taurus*	4.8	3.7	2.8
Cat	*Felis catus*	4.8	4.8	3.7
Elephant	*Elephas maximus*	4.7	11.6	1.9
Opossum	*Didelphis marsupialis*	4.1	7.0	2.8
Dog	*Canis familiaris*	3.1	12.9	5.8
Guinea-pig	*Cavia porcellus*	3.0	3.9	6.6
Rat	*Rattus norwegicus*	2.6	10.3	6.4
Blue whale	*Balaenopetera musculus*	1.3	42.3	7.2
Saddleback	*Pagophilus groenlandicus*	0.9	52.5	3.8
Echidna	*Tachyglossus aculeatus*	0.9	9.6	7.3
Grizzly bear	*Urus arctos*	0.6	22.3	6.7

primates, whereas the sugar contents of whale and seal milk are extremely low. The sea lion *Zalophus californianus* lacks α-lactalbumin and therefore no lactose synthesis is possible. Low milk-sugar concentrations are also found in bears, fish, the sea otter, beaver, coypu and the lesser white-toothed shrew *Neomys fodiens* [185, 209]. In addition to lactose, more than 100 other oligosaccharides that are derived from lactose and often contain sialic acids have been isolated from human milk [79, 101, 150] (Fig. 13.4); similar oligosaccharides are also found in the milk of the mouse and goat [48]. The carbohydrates of marsupial milk are very different in composition to those in the milk of the Placentalia. Lactose makes up only 3–5 % of the total carbohydrate in the milk of the giant kangaroo *Macropus giganteus*. In this case, the carbohydrate content of the milk increases in the first 6 months post partum to 13 % and then rapidly decreases to 2–3 %. Initially, a mixture of tri- and disaccharides predominates and later this changes to monosaccharides, with galactose being more abundant than glucose and glucosamine. Thus, the young in the pouch receive mainly oligosaccharides, whereas the mobile young animals receive monosaccharides [59, 61, 183]. The carbohydrate spectrum of monotreme milk is different again. The echidna *Tachyglossus aculeatus* has mainly sialyl- and fucosyllactose. In the duck-billed platypus *Ornithorhyncus anatinus* about one-half of the total carbohydrate is difucosyllactose, and the rest is a mixture of penta- and nonasaccharides with the composition $Glc_1Gal_{1-3}GlcNac_{1-2}Fuc_{1-3}$; lactose is completely absent [184].

Numerous di- and oligosaccharides have been detected in **bee honey** and in the so-called **honeydew of the aphids** (Fig. 13.5). The formation of honeydew stems from the curious mechanism of nutrient uptake by the phloem sap-sucking aphids. These tap into the phloem of the host plant; because of the internal pressure of the phloem, sap is forced through the proboscis and the gut of the insect and emerges as honeydew from the anus. Only a small fraction of the sap carbohydrate is absorbed by the aphid and, therefore, the honeydew is rich in sugars; these are exploited together with nectar for the production of honey by the honey bee. Honey is composed mainly of glucose and fructose arising from hydrolysis of the sucrose in plant sap. Several of the oligosaccharides detected in honeydew and honey arise directly from the sap without alteration, e.g. sucrose, raffinose and melezitose; others, e.g. maltose and maltose oligosaccharide, isomaltose and glucosucrose, arise by transglycosylation through the action of animal carbohydrases [40, 288].

13.3 Metabolism of Low Molecular Weight Carbohydrates

Usually the requirement of an animal for various sugars and sugar derivatives does not correspond to the availability of these sugars in the diet. Energy-yielding carbohydrate metabolism always proceeds via glucose and fructose phosphates; the biosynthesis of the body's own carbohydrate compounds and derivatives requires a range of sugars that are present in strict proportions. In contrast, the nutrition of herbivorous animals in particular results in the uptake of a complex, changing spectrum of sugars. Thus, many different processes must take place in each animal to bring about the interconversion of sugars and sugar derivatives.

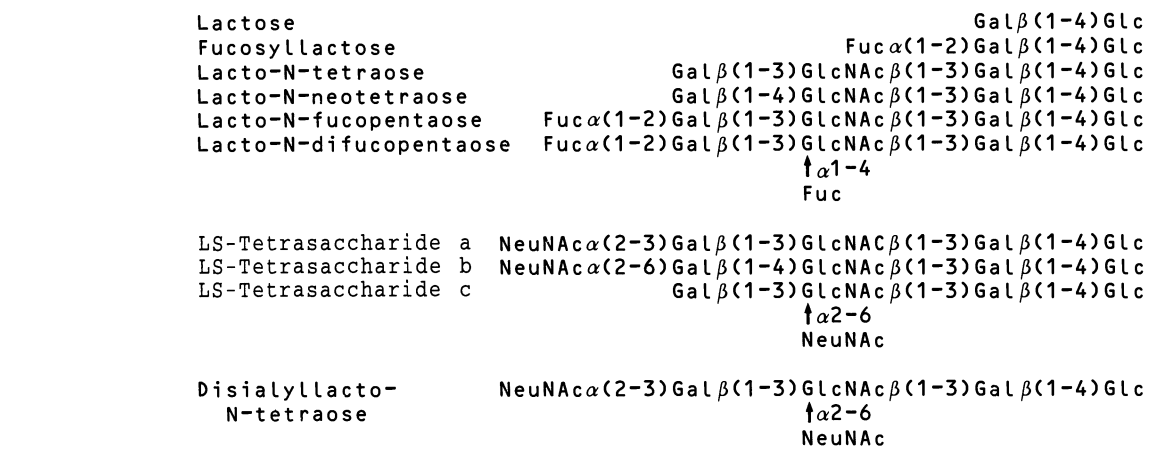


Fig. 13.4. Some of the oligosaccharides found in human milk [218]

Kojibiose (H)	Glcα(1-2)Glc
Nigerose (H)	Glcα(1-3)Glc
Maltose (H Hd)	Glcα(1-4)Glc
Isomaltose (H)	Glcα(1-6)Glc
Sucrose (H Hd)	Glcα(1-2)βFru
Turanose (H)	Glcα(1-3)Fru
Maltulose (H)	Glcα(1-4)Fru
Leucrose (H)	Glcα(1-5)Fru
Melibiose (H Hd)	Galα(1-6)Glc
Panose (H)	Glcα(1-6)Glcα(1-4)Glc
Dextrantriose (H)	Glcα(1-6)Glcα(1-6)Glc
Centose (H)	Glcα(1-4)Glc(2-1)αGlc
Glucosucrose (H Hd)	Glcα(1-4)Glcα(1-2)βFru
Raffinose (H)	Galα(1-6)Glcα(1-2)βFru
Kestose (H)	Glcα(1-2)βFru(6-2)βFru
Melezitose (H Hd)	Glcα(1-2)βFru(3-1)αGlc

Fig. 13.5. Di- and oligosaccharides from bee honey *(H)* and aphid honey dew *(Hd)*

These assumptions have been confirmed by experiments with insects on synthetic diets; different sugars and sugar alcohols were found to increase longevity or promote growth [222].

13.3.1 The Main Metabolic Reactions of Sugars

Most metabolic reactions of the carbohydrates do not involve free sugars but "activated sugars", mainly sugar phosphates and nucleoside diphosphate-bound sugars; sugars bound to dolichol pyrophosphate are used in glycoprotein biosynthesis. **Sugar phosphates** are formed by the transfer of a phosphate residue from ATP to a free sugar, or by transfer of a sugar residue from glycogen to inorganic phosphate:

$$\text{Glucose} + \text{ATP} \rightarrow \text{glucose-6-phosphate} + \text{ADP} \quad (13.1)$$

(hexokinase),

$$(\text{Glycogen})_n + P_i \rightarrow (\text{glycogen})_{n-1} + \text{glucose-6-phosphate} \quad (13.2)$$

(glycogen phosphorylase).

Comparative biochemical investigations of hexokinases and related enzymes and of phosphorylases are discussed in Chapter 14 in connection with glycolysis. For the formation of **nucleoside diphosphate sugars**, the uridyl residue of UTP or another nucleoside diphosphate sugar is transferred to a sugar-1-phosphate:

$$\text{UTP} + \text{glucose-1-phosphate} \rightarrow \text{UDP-glucose} + PP_i \quad (13.3)$$

(Glucose-1-phosphate uridyltransferase),

$$\text{UDP-glucose} + \text{galactose-1-phosphate} \leftrightarrows \text{glucose-1-phosphate} + \text{UDP-galactose} \quad (13.4)$$

(Hexose-1-phosphate uridyltransferase).

The tissues of vertebrates of all classes contain a larger or smaller variety of activated sugars according to their metabolic activities. For example, the oviduct of laying hens contains 14 nucleoside diphosphate sugars, 5 of which bear one or two sulphate groups; comparative investigations reveal specific patterns of such compounds in the oviducts of all avian species. The production of UDP-glucose, UDP-galactose, UDP-N-acetylglucosamine and UDP-N-acetylgalactosamine can be demonstrated in the salivary glands of various dipterans by use of radioactive uridine; apart from these results, there is little known from a comparative standpoint about the synthesis of nucleoside diphosphate sugars.

It can probably be assumed that reactions involved in the conversion of carbohydrates are, in principle, the same in all animals, even though the relevant data are rather limited. **Isomerases** catalyse the interconversion of aldoses and ketoses, and **epimerases** the reversal of the steric organization on a C-atom, e.g.:

$$\text{Glucose-6-phosphate} \leftrightarrows \text{fructose-6-phosphate} \quad (13.5)$$

(Glucose-6-phosphate isomerase),

$$\text{UDP-glucose} \leftrightarrows \text{UDP-galactose} \quad (13.6)$$

(UDP-glucose-4-epimerase).

Sugar alcohols are formed by reduction of the corresponding aldoses or ketoses (Fig. 13.11); 6-phosphoglucuronic acid arises by the oxidation of glucose-6-phosphate and may be decarboxylated to ribulose-5-phosphate (Fig. 13.9). Where the half-acetale-hydroxyl of the sugar is protected, **uronic acid** may be synthesized:

$$\text{UDP-glucose} + 2NAD^+ + H_2O \rightarrow \text{UDP-glucoronic acid} + 2NADH + 2H^+ \quad (13.7)$$

(UDP-glucose dehydrogenase).

Transketolases transfer C_2 fragments and **transaldolases** transfer C_3 fragments from ketoses to aldoses (Fig. 13.9). The formation of glycosidic

bonds requires the transfer of a sugar residue from a donor (usually UDP, other nucleoside diphosphates or dolichol pyrophosphate) to an acceptor through the action of a specific glycosyltransferase. Thus, numerous β-glucosides and β-glucuronides can be produced from UDP-glucose and UDP-glucuronic acid; the oligosaccharides and polysaccharides **arise by a similar process**. In the **biosynthesis of trehalose**, trehalose-6-phosphate is initially formed from UDP-glucose and glucose-6-phosphate, and the free sugar is released by the action of a specific phosphatase (Fig. 13.6a). The two enzymes involved have been purified in parallel from the fat bodies of several insects, and have been characterized. Trehalose phosphate synthase is regulated allosterically, with glucose-6-phosphate acting as the activator and trehalose as the inhibitor [83]. Trehalose phosphatase will be discussed elsewhere (p. 667).

The **biosynthesis of the milk sugar lactose** occurs in one step (Fig. 13.6b). Lactose synthase consists of two different subunits: the catalytic galactosyltransferase and the regulatory **α-lactalbumin**, which decreases the K_m of the transferase for glucose by several orders of magnitude to the millimole level, and increases the V_{max}. The α-lactalbumin is a calcium-binding protein, and its sequence of 122–123 amino acids has been determined partially or completely for a series of higher mammals (Placentalia) and several marsupials; there are 35 invariant positions [95, 195]. This protein belongs to the same super-family as lysozyme (p. 503), and this relationship is clearly illustrated by the close agreement in exons/intron structure between the human, guinea-pig, rat and sheep α-lactalbumin genes and the lysozyme gene of the chicken [213]. In contrast to the single human, rat or guinea-pig gene, the ruminants (cattle, sheep and goat) have a whole α-lactalbumin gene family [281]. Despite its clear homology with the other α-lactalbumins, that of the rat has some special features: it is extended by 17 amino acids (to 140) by mutation of the C-terminal stop codon and, in contrast to almost all other α-lactalbumins, it contains carbohydrate; three isoforms are known which differ in their carbohydrate content. The only other known carbohydrate-containing α-lactalbumin is that of the rabbit [165]. The mouse gene lacks the C-terminal extension. The milk glands of the Tammar wallaby *(Macropus eugenii)* contain two different galactosyltransferases: a β(1–4)-transferase, which corresponds to the catalytic subunit of the lactose synthase of other mammals, and a β(1–3)-transferase, which converts lactose to the oligosaccharide type typical of marsupial milk, Galβ(1–3)Galβ(1–4)Glc [186]. Comparative biochemical investigations of **galactose metabolism** have been carried out in particular in the gastropods, which synthesize galactan (galactogen) in the albumen glands and deposit it in the eggs, where it is used by the growing embryo (p. 488). It has been shown for various snail species that ^{14}C-glucose is incorporated into D-galactose and galactogen and that ^{14}C-galactogen is catabolized to $^{14}CO_2$ in the embryo [170]. The complete enzyme system required for the pathway of glucose to galactogen and further to glycolysis has been demonstrated in the albumen glands of *Helix pomatia* and other snails (Fig. 13.7); the activities of these enzymes are very low in other tissues. The UDP-galactose-4-epimerase of the gastropods, like the mammalian enzyme, is active only in the presence of NAD^+, and is inhibited by NADH [170]. The origin of the L-galactose found in some snail galactans is unclear. The sulphated polysaccharides of the ascidian tunica also contain L-galactose. Tracer experiments have shown that these arise from D-glucose by inversion of the configuration on C atoms 2, 3 and 5 [190]. The three enzymes of galactose metabolism have also been found in the rat tapeworm *Hymenolepsis diminuta*, albeit with low activity. In contrast, they are absent from axenically grown *Drosophila melanogaster* at all stages of development, with the exception of very low epimerase activity in the eggs. The ability of *Drosophila* to oxidize ^{14}C-galactose to $^{14}CO_2$, shown by tracer experiments, is therefore somewhat puzzling.

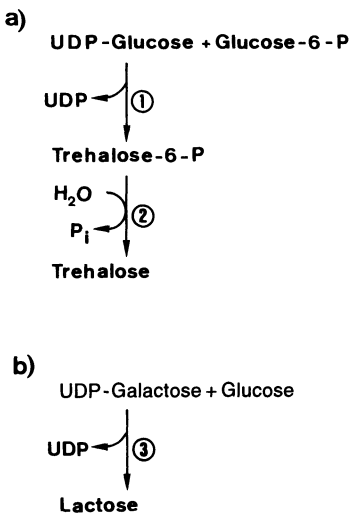


Fig. 13.6a, b. The biosynthesis of **a** trehalose and **b** lactose. *1*, Trehalose phosphate synthase; *2*, trehalose phosphatase; *3*, lactose synthase

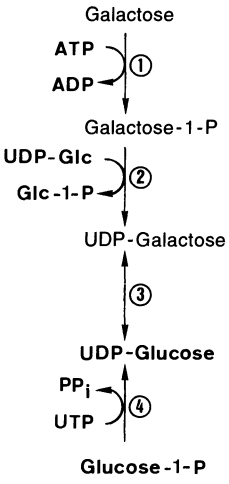


Fig. 13.7. Galactose metabolism. *1*, Galactokinase; *2*, hexose-1-phosphate uridilyltransferase; *3*, UDP-glucose-4-epimerase; *4*, glucose-1-phosphate-uridilyltransferase

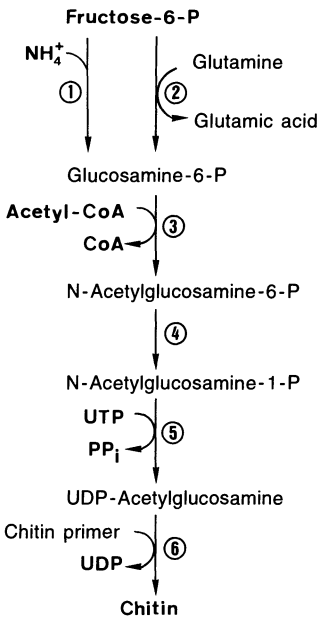


Fig. 13.8. Chitin biosynthesis. *1*, Glucosamine phosphate isomerase; *2*, glucosamine phosphate isomerase (glutamine forming); *3*, glucosamine phosphate acetyltransferase; *4*, acetylglucosamine phosphomutase; *5*, UDP-acetylglucosamine pyrophosphorylase; *6*, chitin synthase

At the time of moulting, the arthropods show particularly intensive **glucosamine metabolism**. In crustaceans and insects, a large part of the old chitin is hydrolysed by the chitinolytic system of the cuticula just prior to moulting. The released N-acetylglucosamine on the one hand satisfies the energy requirements during this period of abstinence, and on the other hand contributes to the formation of the new shell. For the production of the UDP-acetylglucosamine needed as a substrate for chitin synthase, the free N-acetylglucosamine must first be deacetylated, then phosphorylated and finally reacetylated; this is related to the fact that the hexokinases can only act upon glucosamine and not upon N-acetylglucosamine. The de novo synthesis of glucosamine from glucose occurs via fructose-6-phosphate (Fig. 13.8).

The enzymes involved have been found in various insects and crustaceans but have been only partly characterized [83, 253]. In the crayfish of the genus *Orconectes*, ^{14}C-glucosamine is incorporated into chitin prior to moulting, but labelled glucose is not. It would thus appear that during the time when N-acetylglucosamine derived from the old shell is available, the first step in the reaction chain, conversion of fructose-6-phosphate to glucosamine-6-phosphate, is inhibited [253]. There are two possible enzymes for the amination step; one requires glutamine as a cosubstrate, and the other requires ammonium ions (Fig. 13.8). The enzymes have been found together, for example, in mammalian liver, *Drosophila virilis* and *Orconectes limosus; Locusta migratoria* appears to have only the glutamine-specific enzyme, whilst *Musca domestica* has only the ammonium-specific type [83, 253]. There is some evidence that the glutamine-dependent transferase is mainly involved in the formation of glucosamine phosphate and that the ammonium-dependent isomerase brings about its deamination. Thus, the Michaelis constants for fructose-6-phosphate and the N-donor of the *Locusta* and *Drosophila* transferases are consistently less than 1 mmol/l whereas for the isomerases of *Musca* and *Drosophila* they are much higher (6–36 mmol/l). However, the Michaelis constants of the isomerases are reduced to about 3 mmol/l by physiological concentrations of glucose-6-phosphate. Although deamination is thermodynamically preferred in the isomerase reaction, amination is also possible when the glucosamine-6-phosphate produced is removed from the equilibrium situation by acetylation to N-acetylglucosamine-6-phosphate [83]. The corresponding acetyltransferase have been described from pigeon liver and from *Drosophila virilis;* in contrast, for the arthropods little is known about the subsequent two enzymes in the pathway leading to UDP-acetylglucosamine [83].

13.3.2 Formation of Pentoses and NADPH in the Pentose Phosphate Pathway

Ribose is required in large amounts for the synthesis of nucleic acids and free nucleotide phosphates. Pentoses can be produced in animals

either oxidatively or non-oxidatively from hexoses (Fig. 13.9). In the **oxidative pathway**, the two products are ribulose-5-phosphate, which can be isomerized to ribose-5-phosphate, and NADPH, which is required for many biosynthetic processes (e.g. production of fatty acids, cholesterol):

$$\text{Hexose phosphate} + 2\text{NADP}^+ \rightarrow \text{CO}_2 + \text{pentose phosphate} + 2\text{NADPH} + 2\text{H}^+. \quad (13.8)$$

The complicated **non-oxidative reaction** (Fig. 13.9) makes possible the conversion of pentoses to hexoses and vice versa according to the equilibrium equation:

$$3\ \text{Pentose phosphate} \leftrightarrows 2\ \text{hexose phosphate} + \text{triose phosphate}. \quad (13.9)$$

In organs with a particularly high NADPH requirement, the reactions (13.8) and (13.9) are linked in a cycle that produces no pentoses:

$$3\ \text{Hexose phosphate} + 6\text{NADP}^+ \rightarrow 3\text{CO}_2 + 2\ \text{hexose phosphate} \quad \text{triose phosphate} + 6\text{NADPH} + 6\text{H}^+. \quad (13.10)$$

After one sequence of the cycle (13.10), the C1 atoms of the three introduced hexose phosphate molecules should appear as CO_2, and the other C atoms should be found in the order 232456, 233456 and 456. Experimentally determined deviations from this distribution led in 1980 to the hypothesis that the cycle depicted in Fig. 13.9 occurs only in the fat tissue (F-type); in the liver and other tissues a different reaction (L-type) occurs, using the same enzymes in which sedoheptulose-1,7-bisphosphate as well as octulose-8-phosphate and octulose-1,8-bisphosphate appear. Further investigations, however, have not yet confirmed the existence of different types of the pentose phosphate pathway [166].

Direct oxidation, according to Eq. (13.8), and glycolysis compete for glucose-6-phosphate. In an intact animal or isolated organ, the exact proportion of the total glucose turnover which involves direct oxidation and that which results in complete catabolism in glycolysis and the citric acid cycle can be estimated by comparing $^{14}CO_2$, formation from 1-^{14}C-glucose and from 6-^{14}C-glucose. According to the results of such experiments, more than 50 % of the glucose turnover in organs with a high NADPH requirement, such as the milk glands and fat tissue in mammals or the fat bodies in insects, is via direct oxidation [83]. Organs with low activities of glucose-6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase can produce pentoses by reversal of the non-oxidative reaction (13.9).

Tracer experiments with variously labelled ^{14}C-glucose and direct enzyme assays show that the pentose phosphate pathway is almost universal from the protozoans onwards, although its contribution to the total glucose turnover can vary greatly with the species and cell type. It is even present in parasites with predominantly anaerobic respiration. The dysentery amoeba *Entamoeba histolytica* lacks both enzymes of the oxidative pathway and transaldolase, but has a transketolase. In this case, sedoheptulose-7-phosphate

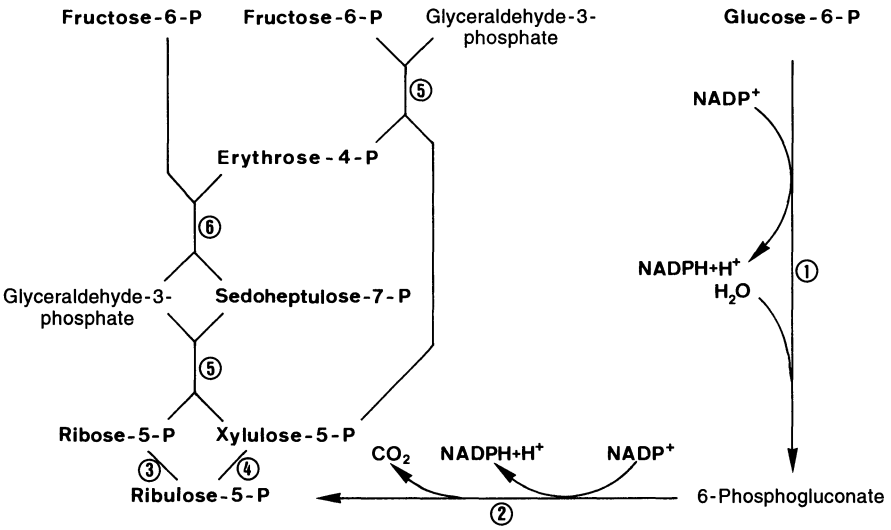


Fig. 13.9. Pentose phosphate pathway. *1*, Glucose-6-phosphate deyhdrogenase; *2*, 6-phosphogluconate dehydrogenase; *3*, ribose phosphate isomerase; *4*, ribulose phosphate-3-epimerase; *5*, transketolase; *6*, transaldolase

is phosphorylated by a pyrophosphate-specific phosphofructokinase to 1,7-bisphosphate, which is then cleaved by an aldolase into dihydroxyacetone phosphate and erythrose-4-phosphate (Fig. 13.10); all these reactions are reversible [261].

Of all the enzymes of the pentose phosphate pathway, only the dehydrogenases have been investigated comparatively. **Glucose-6-phosphate dehydrogenase** (G6PDH) has been isolated from mammalian organs and erythrocytes, from the liver of the teleost *Diocentrarchus labrax*, from the fat body of the silkworm *Bombyx mori*, and from the edible mussel *Mytilus edulis* [83, 180]. Human G6PDH exists as two variants with 515 and 531 amino acids; these differ only in the N-terminal sequence and probably arise by alternative splicing of the same transcript. An enzyme isolated from the rat corresponds to the shorter human variant [132]. The G6PDH cDNA and gene sequences of several *Drosophila* species have been examined. As in the mammals, the gene lies on the X chromosome. The enzyme of *Drosophila melanogaster* agrees in 63 % of its 523 amino acids with the human enzyme but shows no similarity whatsoever in N-terminal positions 1–42. The sequence agreement between *D. melanogaster* and *D. pseudoobscura* is 92 % [160]. The polymorphism of G6PDH has been the subject of intensive investigation and shows medium heterozygosity in invertebrates and vertebrates (see Table 4.9, p. 134). A series of investigations of human erythrocytes indicated G6PDH-deficiency mutations in about 400 million people [282]. NADPH from the G6PDH reaction is used in the formation of reduced glutathione, which protects the SH- groups of erythrocyte enzymes and haemoglobin from oxidation; mutations of the G6pd locus therefore often result in severe pathological symptoms. However, people with deficiency mutations of the G6pd gene show enhanced **resistance to malaria**. This is true for (heterozygous) females but not for males, who are always homozygous. Passage of the malarial agent *Plasmodium falciparum* through G6PDH-deficient erythrocytes results in the induction of an enzyme which compensates for the deficiency of the host cells. In males, in which all erythrocytes are defective, this induction is completely effective. In heterozygous females, where the defective X chromosome is active in only 50 % of the red blood cells, merozoites from normal erythrocytes have only a 50 % chance of reinfecting a normal cell; this leads to enhanced resistance [275]. Particularly high levels of G6PDH are found in cold-resistant insects, where they produce the reduction equivalents needed in the synthesis of the polyol cryoprotectants [257]. The enzyme from *Mytilus edulis*, like that of the rat, is inhibited by NADPH; this inhibition is relieved by oxidized glutathione. The G6PDH activity of the brine shrimp *Artemia salina* is the same with either NAD^+ or $NADP^+$ and both activities are increased by Mg^{2+}. This is in marked contrast to the situation in the milk gland, where the ratio of NAD^+- to $NADP^+$-dependent activity is always less than 1 and is further reduced in the presence of Mg^{2+}.

The G6PDH specific for glucose-6-phosphate and $NADP^+$ is apparently ubiquitous. However, a further dehydrogenase found in some animal groups has a broader substrate and coenzyme

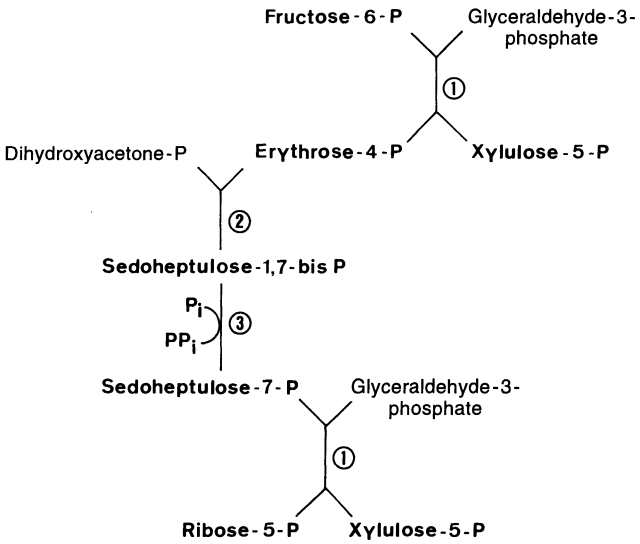


Fig. 13.10. The interconversion of hexoses and pentoses by *Entamoeba histolytica* [261]. *1*, Transketolase; *2*, aldolase; *3*, phosphofructokinase

specificity and, in addition to glucose-6-phosphate and galactose-6-phosphate, can oxidize deoxyglucose-6-phosphate and glucose with NAD^+ or $NADP^+$. The enzyme is designated **hexose-6-phosphate dehydrogenase** by some authors, but it is apparently identical to the **glucose dehydrogenase** (GDH) discovered in mammals in 1931. This enzyme has so far been found only in vertebrates (mammals and teleosts) and echinoderms. Of seven teleost species examined, six possessed several GDHs in addition to G6PDH; G6PDH was missing in *Bathygobius fuscus* [148]. Whereas the enzyme is present in all organs in the mammals, it is restricted to the liver of the teleosts. Its function here may be to provide NADPH for the mixed-function oxidases [24]. A unique reaction mechanism has been described for the GDH from the teleost *Fundulus heteroclitus:* in contrast to all other dehydrogenases, the glucose-6-phosphate here is bound before the coenzyme [229].

G6PDH and GDH are in many respects very different. In mammals, G6PDH is encoded in a gene on the X chromosome, whereas the GDH gene is autosomal. The G6PDH enzyme is localized in the cytoplasm and the vertebrate GDH is membrane bound (microsomes). All G6PDH enzymes from vertebrates and invertebrates are activated by Mg^{2+} and inhibited by the SH reagent *p*-chlormercuribenzoate (PCMB); the GDH of the starfish *Asterias amurensis*, in contrast, is inhibited by Mg^{2+} and not affected by PCMB. The liver GDHs from the rat and the common carp are dimers with subunits of 96–99 kDa, whilst the G6PDHs from mammalian and fish liver are dimers and tetramers, respectively, of 55-kDa subunits. Despite these differences, G6PDH and GDH appear to be homologous: the two enzymes in the starfish *Asterias amurensis* and those in the sea urchin *Strongylocentrotus* show clear relationships in immunoelectrophoresis. Although there is no discernible immunological relationship between the enzymes in mammals, a relationship does emerge when the amino acid sequences of the G6PDHs from rat and carp liver and *Strongylocentrotus* ovary are compared with the two liver GDHs [178]. The oxidation of non-phosphorylated glucose to gluconolactone or gluconic acid can, in principle, be catalysed by three different enzymes: the hydrogen peroxide-producing glucose oxidase (EC 1.1.3.4), glucose dehydrogenase (EC 1.1.1.47) and the FAD-glucose dehydrogenase (EC 1.1.99.10). Glucose oxidase has long been known as a component of bee honey. In *Drosophila melanogaster*, an enzyme present in pupae of both sexes and in the seminal vesicles of adult males is not glucose oxidase, as was previously assumed, but FAD-glucose dehydrogenase, which was not known previously from animals [45]. Two sugar dehydrogenases have been isolated from the hepatopancreas of the xiphosuran *Limulus polyphemus;* one reduces especially D-xylose with $NADP^+$ and the other reduces L-galactose and L-fucose with NAD^+ or $NADP^+$; both enzymes are found at low activity in muscle [96].

6-Phosphogluconate dehydrogenases (6PGDH) have been isolated from various organs and the erythrocytes of mammals, from *Drosophila melanogaster*, and from *Mytilus edulis* [175]. These enzymes all appear to be dimers of 100–105 kDa; the *Drosophila* isoenzyme A consists of two different subunits of 53 and 55 kDa [83]. Similarly, at least two different isoenzymes with somewhat different properties are found in *Musca domestica*. The many investigations of 6PGDH show a much lower level of polymorphism than that shown by G6PDH (see Table 4.9, p. 134).

13.3.3 Biosynthesis of Fructose, Sugar Alcohols and Glycerol

Fructose, which is found in mammals in the seminal fluid and in ungulates also in foetal blood at high concentrations, arises in two steps from glucose via **sorbitol** as the intermediate (Fig. 13.11). Glucose is reduced by the NADP-specific aldose reductase to sorbitol, which is reoxidized to fructose by the action of NAD-specific sorbitol dehydrogenase. High activities of **aldose reductase** are found in the seminal vesicles and the ungulate placenta, and also in the eye lens. The enzymes from the lens and placenta consist of 315 amino acids and are N-terminally acetylated. They belong, together with the aldehyde reductase of the liver and placenta, the prostaglandin F-synthase of the lung and the rho-crystallin of the frog eye lens, to a family of monomeric, NADPH-dependent oxidoreductases with a

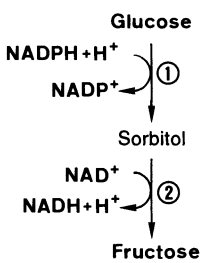


Fig. 13.11. The sorbitol pathway of fructose synthesis. *1*, Aldose reductase; *2*, sorbitol dehydrogenase

broad substrate specificity for carbonyl compounds [99, 237]. Sorbitol **dehydrogenase** (SDH) is found in the same organs as aldose reductase, although in fact the highest activities are found in the liver, in which there is no aldose reductase activity. Here again, the enzymes of many mammalian species have very similar properties: they contain Zn^{2+}, are strongly NAD specific, but can produce ketoses from a range of different sugar alcohols of five to seven C atoms. The native enzymes are homotetramers of about 140 kDa; the subunits from human liver and those from many other mammals show up to 25 % sequence agreement with the alcohol dehydrogenases (ADH) [146]. Comparison of the SDH sequence with the three-dimensional model of horse-liver ADH explains why the SDH is inactive against ethanol and the ADH is inactive against sorbitol: the zinc ligand in the SDH is a glutamate but in the ADH it is a cysteine [75]. Erythrocytes also contain SDH and human erythrocytes can therefore convert 3 % of the glucose taken up into sorbitol. The SDH activity of the blood cells varies in all vertebrate classes between 0.01 and 0.17 U/g haemoglobin [17].

NADP- and NAD-specific "sorbitol dehydrogenases" are also found in various insect but none has yet been isolated. Sorbitol is important mainly as a cryoprotectant in the insects, e.g. during diapause. From the fluctuations in activity of the two enzymes in *Bombyx mori*, it may be concluded that the sorbitol required at the beginning of diapause is synthesized from glucose or glycogen by the NADPH-specific aldose reductase, and that the NAD-specific sorbitol dehydrogenase is responsible for the opposite reaction, the resynthesis of glycogen at the conclusion of diapause. In fact, the NADPH-dependent activity is more-or-less constant at 2–3 mU/mg protein, and the enzyme activities of the NADPH-yielding pentose phosphate cycle increase considerably at the start of diapause. The NAD-specific sorbitol dehydrogenase is almost undetectable in diapause eggs but increases at the completion of diapause to values of over 6 mU/mg protein [83]. The sugar molecule **mannitol**, which as a storage product of fungi, algae and higher plants can reach concentrations of up to 10 % of the dry weight, is exploited by herbivorous animals, as shown for example by nutritional experiments with locusts. Initially, the NAD-specific sorbitol dehydrogenase was thought to be responsible. However, an alternative route is apparently taken in the pulmonate snail *Helix aspersa*. Here, the hepatopancreas and the gut wall contain an NAD(P)-dependent, H_2O_2-producing mannitol oxidase at high activities; this appears to be localized in the endoplasmic reticulum and not, as first assumed, in the mitochondria or peroxisomes [173].

The dihydroxyacetone phosphate produced during glycolysis can be reduced to sn-glycerol-3-phosphate by the cytoplasmic NAD-specific **glycerol-3-phosphate dehydrogenase** (α-glycerophosphate dehydrogenase, GPDH). Hydrolytic cleavage of the phosphate residue leads to glycerol. As a product of anaerobic energy-yielding metabolism, glycerol is not important for the Metazoa as its formation is energetically inefficient. Glycerol phosphate, however, is of great importance to all animals as the starting point for the synthesis of glycerol lipids. In particular, in insect flight muscles it is exploited as a transport vehicle for reducing equivalents across the mitochondrial membrane **(glycerol phosphate shuttle)**. In this process, the glycerol phosphate produced in the cytoplasm is reoxidized to dihydroxyacetone phosphate by glycerol-3-phosphate:coenzyme Q-oxidoreductase (glycerol phosphate oxidase), located on the outer surface of the inner mitochondrial membrane, and the electrons are passed on to the respiratory chain. Free glycerol serves as a cryoprotectant in some insects and anurans (p. 466). Exogenous glycerol can be phosphorylated by the mitochondrial glycerol kinase in mammals, bacteria and some insects [220].

Comparative biochemical data are available only for the GPDH of higher vertebrates and insects. The rabbit enzyme agrees in more than 90 % of its 348 amino acids with the mouse enzyme, and in almost 89 % of positions with that of the chicken. This gives a relatively low substitution rate of $0.19 \cdot 10^{-9}$ per nucleotide per year, similar to the rate for cytochrome c [14, 125]. Isoenzymes of GPDH have been found in all eukaryotes examined in detail. The mouse has two GPDH genes, of which the gdc-1 locus on chromosome 15 is expressed only in the adult animal, and locus gdc-2 on chromosome 9 only in the embryo. The genes consist of eight exons, whereby intron 2 lies on the border between the adenine- and nicotinamide-binding domains and intron 3 is on the border between the NAD-binding and catalytic domains. There would also appear to be two GPDH loci in man, the rat, rabbit and chicken [125].

Drosophila has only one GPDH gene with eight exons, from which three tissue- and stage-specific isoenzymes are produced by alternative

splicing. GPDH-1 is mainly found in adult flight muscle, where it is involved in the glycerol phosphate shuttle. GPDH-deficient mutants cannot fly. GPDH-2 and -3 are located in adults in the abdomen, where GPDH-3 in particular produces the starting material for the synthesis of phospholipids. GPDH-3 is encoded by exons 1–6 and has a length of 349 amino acids, GPDH-1 has 3 additional amino acids encoded by exon 8, and GPDH-2 has 10 more amino acids which are encoded by exon 7. The GPDH-3 of *Drosophila melanogaster* agrees with that of *D. virilis* at 98 % of positions and with the mammalian enzyme at 65–67 % of positions [27, 141]. Duplications and triplications of the GPDH locus have been discovered in Japanese populations of *D. melanogaster* [262]. Otherwise, the GPDH of the *Drosophila* genus shows unusually little polymorphism. Alleloenzymes of GPDH have been found in only three out of 175 *Drosophila* species examined, and in no case were differences in the in vitro catalytic properties or temperature sensitivity apparent [27].

13.3.4 Ethanol Metabolism

Three enzymes or enzyme systems may be involved in the **oxidation of ethanol to acetaldehyde** in animals: NAD-specific alcohol dehydrogenase, catalase, and the microsomal ethanol-oxidizing system. An NADP-dependent alcohol dehydrogenase like that of bacteria has been found in the animal kingdom only in the dysentery amoeba *Entamoeba histolytica* [171]. Despite the extensive similarity of their catalytic properties, the ADHs belong to two different protein families. The "short" ADHs have only about 250 amino acids and no zinc; these include the *Drosophila* ADH, which is significantly homologous to the protein P6 from the fat body, the 17β-hydroxysteroid dehydrogenase from human placenta, and the protein act III from the fungus *Streptomyces coelicolor* [12, 221]. The "long" ADHs have chains of about 350 amino acids and two zinc atoms; representatives of this family are known from such widely different organisms as mammals, maize and baker's yeast. Related to the long ADHs are the sorbitol dehydrogenases, threonine dehydrogenases and ζ-crystallin [140].

The **alcohol dehydrogenase of the vertebrates** is a cytoplasmic enzyme with the highest activity always in the liver and much lower activities in the kidney, heart, lung and gut. ADH molecules are homodimers of 40-kDa subunits, each chain of which carries two zinc atoms bound to cysteine; at least one of these zincs is involved catalytically in the reaction. ADH is NAD specific and catalyses the reversible conversion of numerous substances with primary or secondary hydroxyl groups to the corresponding aldehyde or ketone. The activity ratio NAD^+: $NADP^+$ of liver ADH in the horse has a value of 86.4, in the rat it is 17.5, and in an unidentified trout species it is 5.4 [271]. **Isoenzymes** with varying catalytic properties encoded at multiple loci have been found in all mammals examined, including humans. The numerous isoenzymes of human liver can be divided into three classes (I–III) [76, 245]. Whereas ADH–I and ADH–II are restricted to the liver, lung, kidney and alimentary canal, ADH–III is found in nearly all cells. The three ADH classes differ considerably in the structure of their catalytic centre, substrate specificity, affinity and inhibitor reactions. In the 11 amino acids of the substrate-binding pocket there are eight differences between I and II and ten differences between I and III [76]. Methanol is only oxidized by I, and the affinity for ethanol decreases markedly in the order I > II > III. The class-II and -III enzymes are particularly active with long-chain substrates. ADH–I is more strongly inhibited than ADH–II by 4-methylpyrazole whereas ADH–III is almost completely insensitive to this inhibitor [82]. Class I, which is by far the best investigated, includes homo- and heterodimers made up of the subunits α, β and γ; these are encoded by the three loci ADH_1, ADH_2 and ADH_3, which show about 95 % sequence agreement in pairwise comparisons. Adult liver contains all three types of subunit, whereas the lung contains only β, and kidney and gut contain mainly γ. Embryonal liver expresses only α, and β appears after the 3rd month; γ appears only after birth.

Allelic variants exist for ADH_2 (β_1, β_2 and β-Indianapolis) and ADH_3 (γ_1, γ_2). β_1 and β_2 differ at only one position, and γ_1 and γ_2 differ at two positions. The amino acid substitution 47-Arg to His between β_1 and β_2 is especially interesting [177]. Caucasians have predominantly β_1, whilst Asians have β_2. As the specific activity of the β_2 homodimer is considerably higher than that of β_1, the toxic acetaldehyde accumulates more easily in the liver of many Asians, leading to a higher sensitivity to alcohol [82]. Class-II ADH enzymes are homodimers of π chains encoded by the ADH_4 locus, and class III consist of χ subunits encoded in the ADH_5 locus. The subunits of classes I, II

and III differ by 35–40 % in their amino acid sequences.

The structural and functional variety in the protein family of the alcohol dehydrogenases may be explained by three successive gene duplications. The first duplication separated the sorbitol and alcohol dehydrogenases; the ADH classes arose in the second phase, and the third phase saw the appearance of the isoenzymes of class I [135].

Comparison of the human and baboon β sequences suggests that the α/βγ gene duplication occurred about 60 million years ago, at approximately the same time as the appearance of the first primates [270]. The rabbit ADH I also has a complicated pattern of homo- and heterodimers of three different subunits [174]. The tissues of the rat contain three ADH isoenzymes, ADH-1, -2 and -3, which are comparable with the human ADH II, III and I in their substrate specificities and pyrazole sensitivities. ADH-1 predominates in the eye, stomach and lungs; ADH-2 is present in all organs, and ADH-3 is found mainly in the liver [139]. Despite its functional similarity with ADH II, the ADH of rat stomach is so different in its sequence to all other rat ADHs (32–40 %) that it may be considered to represent a new class, IV, of mammalian ADHs. In this case, the number of gene duplications that occurred is larger than was previously assumed [204]. In the mouse, there are three ADH isoenzymes, A2, B2 and C2, encoded in the genes adh-1 to -3 on chromosome 3; allelic variants of adh-1 and -3 are found. The individual isoenzymes are structurally and functionally very different. A has a molecular mass of 43 kDa, B 39 kDa, and C 47 kDa. A and C contain two catalytic zinc atoms, but B has only one non-catalytic zinc. A2, which is present in particular in the liver, kidney and gut, is the only one that can convert ethanol under physiological conditions; the C2 from the stomach wall and lungs and the similarly distributed B2 have a very low affinity for ethanol. The complete gene sequence of the A subunit is known; the amino acid sequences of mouse A and horse E differ at 58 (15.5 %) out of 374 positions [74]. There are three dimeric ADH isoenzymes in the horse, and these are made up of the subunits E (ethanol active) and S (steroid and ethanol active). The complete sequence is known for both isoenzymes. It may be concluded from comparisons of the horse E/S and human β_1/γ_1 sequence that the two duplications occurred independently [205]. The liver ADHs of various avian species and the frog *Rana perezi* are more similar to mammalian ADH I than to ADH II or III [46, 140].

Almost the only **invertebrate ADH** investigated in any detail is that of *Drosophila;* among the many hundreds of publications, the majority deal with the particularly marked enzyme polymorphism, and leave open several important biochemical questions. Although there are always many ADH bands visible in electropherograms, there is only one ADH locus in *Drosophila melanogaster*, as in most *Drosophila* species. However, at least eight allelic variants of ADH have been described, of which the alleloenzymes F (fast) and S (slow) are present in all natural populations. This polymorphism is superimposed upon variability arising from post-translational modifications, which give three to five electrophoretic sub-bands for each of the alleloenzymes (p. 145).

The *Drosophila* ADH is very similar to the yeast and mammalian enzymes in functional properties but is structurally very different. The subunits have only 254 amino acids (27.4 kDa) compared with 347 (35.3 kDa) in yeast and 374 (39.8 kDa) in horse liver. The fly enzyme has no zinc and the amino acid sequence shows no homology to the sequences of the yeast and horse enzymes. Whereas the yeast and mammalian enzymes are clearly related to each other and to mammalian sorbitol dehydrogenase, the *Drosophila* enzyme belongs to a completely different protein super-family. The individual regions of the ADH gene show large differences in their rate of evolution due to varying restrictions to evolutionary change. This is already clear from comparisons of the allelic variants ADH-F and -S in *Drosophila melanogaster* (see Fig. 4.6b, p. 138), but is seen even better in comparisons of the ADH sequences of *D. melanogaster* with those of the closely related species *D. mauritana* and *D. orena*. Amino acid substitutions at only 72 out of 254 positions have so far been described in *Drosophila* ADH [280]. The ADH genes of almost all *Drosophila* species have two promoters which are active at different times in development. The gene transcripts encode the same amino acid sequence but differ in their 5′ non-translated (NT) regions. In *D. melanogaster* and *D. affinis*, the proximal promotor is used in the larvae and the distal promoter is used in the adult animal; in *D. formella*, the larval situation is also found in the imago [231, 232]. A 34-kDa protein that has been isolated activates the Adh distal promoter of *D. melanogaster;* it is known as the Adh distal factor 1 (adf-1), and also binds to the promoters of other genes [78]. The ADH gene of *D. lebanonensis* contains three TATA boxes [138].

As the result of a gene duplication, *D. mulleri, D. mojavensis* and other species of the *D. mulleri* subgroup have two loci, Adh-1 and -2, each with only one promoter. The polypeptides encoded by these genes form two homodimers and one heterodimer, which differ only slightly in their activities. Adh-1 is expressed only in the larvae of most of these species, but in *D. mojavensis* and its sibling species *D. arizonensis* it is also expressed in the ovary of adult females; adh-2 is expressed from larval stage 3 onwards [18, 286].

The ADHs of all *Drosophila* species are very similar to the mammalian enzyme in their catalytic properties; they are strictly NAD specific and have a broad substrate specificity, which includes not only primary and secondary alcohols but also, for example, terpenes. Secondary alcohols are metabolized faster than primary alcohols, and amongst the primary alcohols, short-chain forms up to butanol are metabolized more rapidly than longer-chain forms [290, 297]. The ADH of *D. melanogaster*, in contrast to the enzyme of the sibling species *D. simulans*, is able to oxidize acetaldehyde [191]. The alleloenzymes S and F of *D. melanogaster* show no significant differences in specific activity, substrate specificity or affinity; however, the SS homodimer is significantly more heat stable than the FF dimer (p. 145). The ADH activities in the genus *Drosophila* differ widely with the species, the organ and the developmental stage. *Drosophila* species that seek out fermenting fruits have relatively high ADH levels; acetaldehyde is either used as a substrate in metabolism or is tolerated. In contrast, species from rainforests and those that prefer *Hibiscus* or feed on fungi have lower ADH activities; in these cases, acetaldehyde is not further metabolized and is poorly tolerated. Compared with other *Drosophila* species, *D. melanogaster* is particularly ethanol tolerant and can make use of low concentrations of ethanol, n-propanol or n-butanol as the carbon source; ethanol is found in the habitats of these species and is chemoattractive to larvae and adults. The ADH in *D. melanogaster* is inducible by ethanol; the responsible DNA region lies upstream of the distal promoter between positions +604 and +634 [143]. Adults of *D. melanogaster* that are homozygous for the S allele have twice as much ADH activity and immunologically detectable ADH protein as the sibling species *D. simulans*, although the mRNA concentrations are similar. Thus, the differences in protein concentration involve differences in the rate of translation or protein stability. Gene transfer experiments have shown that the expression of the genes of the two species is very much dependent upon the genetic background, and is always greater in *D. melanogaster* than in *D. simulans* [168]. The ADH activity in *D. melanogaster* peaks in the third larval stage, and again in freshly hatched adults; these maxima correspond to variations in two mRNAs transcribed from the same gene. The fat body, gut, Malpighian tubules and male genital tract have the highest activities in *Drosophila*, and the ADH organ distribution is similar in other insects, e.g. in the house cricket *Acheta domestica*.

The acetaldehyde produced in the ADH reaction is much more toxic than ethanol itself and must be rapidly metabolized further. Four enzymes are involved in the **metabolism of acetaldehyde:** ADH itself, which can only oxidize acetaldehyde in some *Drosophila* species; the NAD-specific aldehyde dehydrogenase (EC 1.2.1.3); the flavin enzyme aldehyde oxidase (EC 1.2.3.1); and catalase. The NAD-dependent aldehyde dehydrogenases (ALDH) play a key role in ethanol metabolism in the mammals. In all mammals they are organized into organ-specific systems of isoenzymes. These isoenzymes may be localized in the cytosol or in the mitochondria and may have different substrate specificities, but they are always homotetramers with subunits of about 500 amino acids. Not all ALDH isoenzymes have been structurally or functionally characterized, even in the case of human ALDH. $ALDH_1$ (cytosolic) and $ALDH_2$ (mitochondrial) of the liver and γ-aminobutyrate dehydrogenase (γ-ABDH) have high affinities for acetaldehyde; the $ALDH_3$ of the stomach and lung, which preferentially uses benzaldehyde as a substrate, and $ALDH_4$, which is a glutamic γ-semialdehyde dehydrogenase, have a much lower affinity for acetaldehyde [122]. Comparisons of the sequences of ALDHs from the same cell compartment of different mammalian species in each case show about 95 % agreement, but exhibit only about 67 % agreement between cytoplasmic and mitochondrial enzymes of the same species [103]. After some initial difficulties, ALDH has also been detected in *Drosophila melanogaster.* However, 75–90 % of the acetaldehyde in tracer experiments is oxidized to acetic acid by alcohol dehydrogenase [108]. Nevertheless, it is clear that ALDH, like ADH, has a much higher activity in *D. melanogaster* than in the less ethanol-resistant sibling species *D. simulans*. The aldehyde oxidase of *D. melanogaster* has such a low activity for acetaldehyde (K_m of about 10 mmol/l) that it must be relatively unimportant for ethanol metabolism

[90]. Inhibition experiments with 3-amino-1,2,4-triazol indicate that the catalase of *D. melanogaster* is only marginally involved in ethanol metabolism but is important for the catabolism of methanol [277].

13.3.5 Biosynthesis of Ascorbic Acid

Ascorbic acid is a substrate involved in redox reactions such as the hydroxylation of proline in collagen synthesis or the formation of noradrenalin from dopamine by dopamine-β-monooxygenase. It is present in plant foodstuffs but can also be synthesized de novo in many vertebrates. Biosynthesis occurs via glucoronic acid to L-gulonolactone, which is oxidized to ascorbic acid by gulonolactone oxidase (Fig. 13.12). **Gulonolactone oxidase** has been isolated from rat and goat liver and from chicken kidney, and is a microsomal enzyme bearing covalently bound 8α(N^1-histidyl)flavin as a prosthetic group. The rat enzyme has been sequenced via the cDNA; the polypeptide chain consists of 440 amino acids, corresponding to a molecular mass of 50.6 kDa, and shows no significant homology to other flavin enzymes [157].

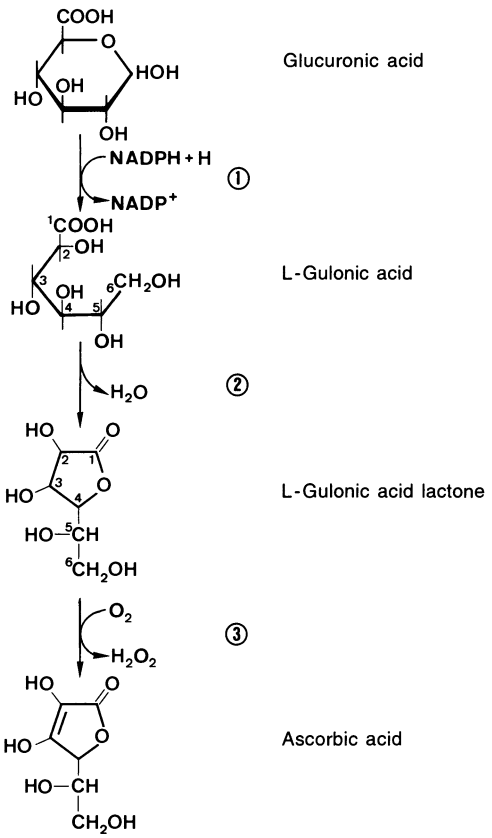


Fig. 13.12. The biosynthesis of ascorbic acid. *1*, Glucuronate reductase; *2*, aldonolactonase; *3*, gulonolactone oxidase

Two methods may be used to determine the **capacity for ascorbic acid synthesis**. The demonstration of the production of ascorbic acid from glucuronic acid includes all three enzymes of the biosynthetic pathway, but is relatively insensitive when non-radioactive substrates are used. The much more sensitive direct determination of gulonolactone oxidase is a reliable indication of the functioning of the whole pathway. The kidney and liver of the carp *Cyprinus carpio* have no detectable gulonolactone oxidase [62]; in various salmonids (e.g. trout and salmon) and ictalurids, and probably all other teleosts, the ability to synthesize ascorbic acid is insufficient or totally absent. Hence, in the latter cases, ascorbic acid is an essential nutrient. In the lungfish (Dipnoi), amphibians, reptiles, birds and egg-laying mammals (Monotremata), ascorbic acid production takes place in the kidney, and in the marsupials and higher mammals (Placentalia) it occurs in the liver. In some avian species and marsupials *(Perameles, Isoodon)*, synthesis takes place in both organs [29, 73]. Gulonolactone oxidase is expressed in various tissues of chicken embryos but disappears from all but the kidney.

Some birds and mammals are unable to synthesize ascorbic acid, e.g. humans and probably all other primates, the guinea-pig, marmot, flying fox *Pteropus* and the sparrow *Pycnonotus* (bulbul). Ascorbic acid (vitamin C) is an essential nutrient for these species, and a shortage results in deficiency symptoms (scurvy). As far as is known, the absence of biosynthesis lies in all cases in a deficiency of gulonolactone oxidase. Implantation into guinea-pigs of a dialysis sack containing this enzyme from rat liver or chicken kidney alleviates the deficiency symptoms caused by an ascorbic acid-free diet. The incidence of this metabolic "defect" in the evolution of these animals might be explained by the fact that their diet was at some stage so rich in ascorbic acid that the gulonolactone oxidase gene was no longer stabilized by selective pressure. Rat mutants have been reported which have no active gulonolactone oxidase and, conversely, three out of several thousand guinea-pigs were found to show no scurvy symptoms after an 8-month ascorbic acid-free diet [157].

The situation in the invertebrates is not at all clear. Ascorbic acid is essential for several phytophagous insects; in these cases, it is also involved

in the regulation of melanization and sclerotization of the cuticala. In crustaceans of the genus *Penaeus*, an ascorbic acid deficiency results in a lethal condition knows as "black death". The biosynthesis of ascorbic acid has been claimed to occur in, for example, various cockroach species but has never been unambiguously demonstrated. In contrast to those fed on a normal diet, caterpillars of the lepidopteran *Manduca sexta* fed on an ascorbic acid-free diet have no tissue ascorbic acid; gulonolactone oxidase is as undetectable in this species as it is in *Plodia interpunctella* or *Periplaneta americanus*. Searches for this enzyme were also unsuccessful in annelids, crustaceans, spiders, chilopods and pulmonate snails. The only reliable report of gulonolactone oxidase in an invertebrate is that from *Limulus polyphemus*. The heart and several other organs of this species showed specific activities of up to 5 mU/g, relatively low compared with the values for rat liver (195 mU/g) and chicken kidney (570 mU/g) [285].

13.4 Polysaccharides and Proteoglycans

Polysaccharides in living cells are always more-or-less tightly bound to proteins. The so-called protein-free sugar polymers are really artefacts, resulting in part from the use of drastic isolation procedures. This is also true for storage polysaccharides such as glycogen, which in the cell is very closely associated with the enzymes responsible for its synthesis and catabolism. It is especially true for the structural polysaccharides such as cellulose, chitin or the glycosaminoglycans. Covalent bonding of carbohydrates and proteins occurs, for example, by N-glycosidic binding to asparagine, or O-glycosidic binding to serine, theronine, hydroxylysine or other hydroyamino acids. However, structural details are known almost only for the proteoglycans and glycoproteins of the mammals.

Two types of carbohydrate-protein complexes can be clearly distinguished in the vertebrates. In the **proteoglycans**, the carbohydrate is the dominant component responsible for the properties of the complex. These are always linear glycosaminoglycans built up repetitively of disaccharide subunits. They all contain hexosamine and, except for keratosulphate, also uronic acids. With the exception of hyaluronic acid, they are all sulphated, and the sulphation gives the polysaccharides their polyanionic character. Less is known about the protein component. The large proteoglycan aggregate from rat cartilage is the only proteoglycan for which the complete amino acid sequence is known. In contrast to proteoglycans, the properties of the **glycoproteins** are determined by the protein component. In this case, the protein carries one or more oligosaccharides which may be identical or variable. The carbohydrate chains consist of up to 20 sugar residues and are often branched. They all contain glucosamine or galactosamine residues, the majority of which are acetylated, as well as mannose, galactose and fucose, but seldom glucose. Up to seven different sugars may be found in one carbohydrate chain. There is often a terminal sialic acid which gives the molecule its negative charge; sulphate residues are rare and there are no uronic acids [214].

The distinction between proteoglycans and glycoproteins is not so clear in the invertebrates. Sialic acid is not present in the glycoproteins of many invertebrate groups, but uronic acids are sometimes found. The proteoglycans are much more variable than the glycosaminoglycans of the vertebrates; their polysaccharide components are often made up only of sulphated or phosphorylated neutral sugars and lack hexosamines and uronic acids. In this case, as for other classes of substances in the organism, there is a clear tendency in the vertebrates towards specialization and simplification [123].

13.4.1 Glycosaminoglycans

The glycosaminoglycans or mucopolysaccharides are typical components of intercellular matrices. More recently, they have been recognized as components of the cell surface, where they probably play a role in cell-cell and cell-substrate interactions [115]. Several **basic types** are recognizable in the vertebrates (Fig. 13.13). The glycosaminoglycans are typically very variable in chain length, the number and position of the sulphate groups and, in derivatives of chondroitin and heparin, also in the ratio of glucuronic to iduronic acids; thus, the fractions isolated from tissues are also very heterogeneous. The relative proportions of individual glycosaminoglycans are species specific and dependent upon the developmental stage (Table 13.3); this is true for all classes of vertebrates [97].

The protozoans apparently have no glycosaminoglycans. In contrast, all examined metazoans have been found to have one or more mucopoly-

a)

```
Chondroitin sulphate
 β(1-4)GlcUAβ(1-3)GalNAcβ(1-4)GlcUAβ(1-3)Galβ(1-3)Galβ(1-4)XylβSer
                   4-ord 6-Sf
"Dermatan sulphate"
 β(1-4)IdUAα(1-3)GalNAcβ(1-4)GlcUAβ(1-3)Galβ(1-3)Galβ(1-4)XylβSer
       2-Sf       4-Sf
Heparin
 α(1-4)IdUAα(1-4)GlcNSfα(1-4)GlcUAβ(1-3)Galβ(1-3)Galβ(1-4)XylβSer
       2-Sf       6-Sf
Heparan sulphate
 α(1-4)GlcUAβ(1-4)GlcNAcα(1-4)GlcUAβ(1-3)Galβ(1-3)Galβ(1-4)XylβSer
                   6-Sf
Keratan sulphate I
 β(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Gal(GalNAc,Man)GlcNAcβAsn
                            6-Sf         6-Sf
Keratan sulphate II
 β(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Gal(1-6)GalNAcαThr(Ser)
                            6-Sf         6-Sf     Gal-NANA
Hyaluronic acid
 β(1-4)GlcUAβ(1-3)GlcNAcβ(1-4)GlcUAβ(1-3)GlcNAcβ(1-4)...
```

b)

```
Elasmobranchei    β(1-4)GlcUAβ(1-3)GalNAcβ(1-4)...
                        2-or 3-Sf  6-Sf
Myxine            β(1-4)GlcUAβ(1-3)GalNAcβ(1-4)...
                        2-or 3-Sf  4,6-BisSf
Xiphosura         β(1-4)GlcUAβ(1-3)GalNAcβ(1-4)...
                        2-or 3-Sf  4-Sf
Cephalopoda       β(1-4)GlcUAβ(1-3)GalNAcβ(1-4)GlcUAβ(1-3)GalNAcβ(1-4)...
                                   4,6-BisSf              6-Glc
```

Fig. 13.13a, b. The structure of the repetitive disaccharide subunits and the protein bonds of glycosaminoglycans. **a** Typical glycosaminoglycans of mammals [169]. **b** Disaccharide subunits of over-sulphated chondroitin sulphates [123]. *Sf*, Sulphate residue; *NANA*, N-acetylneuraminic acid

Table 13.3. Glycosaminoglycans in the organs of adult and newborn humans (after Buddecke)

Organ	MPS	Proportion in % MPS
Skin (adult)	0.5–1.0	DS 70, HA 30
Skin (newborn)	0.5–1.0	DS 25, HA 75
Aorta (adult)	1.0	CS 60, DS 10, HA 20, HS 10
Cartilage (adult)	15	CS 50, KS 50
Cartilage (newborn)	25	CS 95, KS 5
Cornea (adult)	1.7	CH 45, CS 5, KS 50

MPS, Total mucopolysaccharide as % dry weight. The proportion of each glycosaminoglycan is give as % MPS: CH, chondroitin; CS, chondroitin-4- and -6-sulphate; DS, dermatan sulphate; HA, hyaluronic acid; HS, heparin; KS, keratan sulphate

saccharides. The results of comparative studies with chromatographic and electrophoretic methods, subunit analysis and cleavage by specific enzymes suggest that many of the invertebrate glycosaminoglycans are comparable to one or other of the known vertebrates types [66]. However, in many cases more detailed studies have shown such fundamental differences between the proteoglycans of different animal groups that analogies must be treated with reservation [123].

With the exception of hyaluronic acid, all glycosaminoglycans of the vertebrates are bound to core proteins. Particularly large proteoglycan aggregates are found in the extracellular matrices of cartilage, aorta, skin, tendons, placenta and brain. Human fibroblasts produce a proteoglycan which, because of certain sequence similarities to the epidermal growth factor (EGF) and lectins, is known as **versican** (versatile proteoglycan). It consists of numerous chondroitin sulphate chains on a core protein of 2389 amino acids, and binds to hyaluronic acid. In addition, the fibroblasts produce a further proteoglycan with dermatan sulphate on two core proteins of 460 and 300 kDa which are encoded by different genes [37]. By far the best-known proteoglycans are the **aggrecans** from mammalian and avian cartilage. They have molecular masses of about 2.5 MDa and are made up of a core protein bearing about 100 chondroitin sulphate and 15–50 keratan sulphate molecules as well as numerous N- and O-bound oligosaccharides. Up to 100 such proteoglycan

molecules aggregate with a molecule of hyaluronic acid; the bindings is stabilized by so-called link proteins (Fig. 13.14). The hyaluronic acid interacts with the collagens of the cartilage. Amongst these collagens, collagen IX is in turn covalently bound to chondroitin sulphate or dermatan sulphate. Because of the high charge density on the aggregated glycosaminoglycans, cartilage is able to absorb large amounts of water, up to 30–50 times the dry weight. The consequent expansion is restricted by the network collagen fibres, and the resulting expansion pressure confers the characteristic mechanical properties of cartilage. Complete sequences are known for the **core proteins** of rat chondrosarcomes (2124 amino acids) and human cartilage (2316 amino acids), and partial sequences are known for those from bovine and chicken cartilage [68]. The rat protein is organized into eight domains of very different structure and function. N-terminal there are two globular structures of 37 and 22 kDa, the sequences of which appear to be complete and partial copies, respectively, of the link protein. The first N-terminal domain binds hyaluronic acid and the link protein. The middle part of the polypeptide chain consists of 1104 amino acids with 117 Ser-Gly sequences as binding sites for chondroitin sulphate; these are organized into three repetitive domains of 428, 503 and 173 amino acids. The C-terminal domain has a globular structure; it bears no glycosaminoglycan and has similarities to lectins. N- and O-glycosidic bound oligosaccharides are distributed over the whole chain [67]. The core protein of the human proteoglycans shows about 75 % sequence agreement with that of the rat. It has a highly repetitive, keratan sulphate-binding region behind the two globular domains; this is also found in the bovine gene but is missing from the rat [5, 68]. The cartilage of the shark fin appears to contain no hyaluronic acid and isolated proteoglycans lack the capacity to bind hyaluronic acid. Accordingly, the cartilage matrix in this case must have a completely different molecular architecture to that in the higher vertebrates [187].

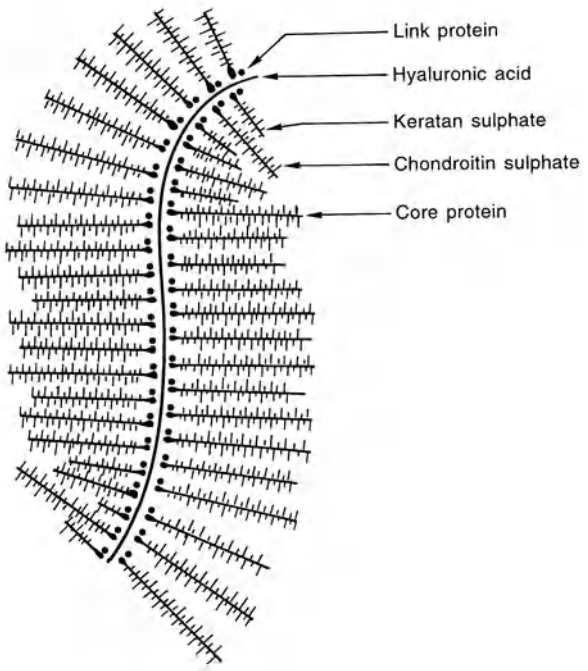


Fig. 13.14. The structure of the proteoglycan aggregate from bovine nasal cartilage [169]. On one molecule of hyaluronic acid of about 1 MDa there are 20–80 proteoglycan monomers, each of 2–3 MDa, bound non-covalently by means of smaller link proteins. Each proteoglycan monomer contains a central core protein of 200 kDa which carries about 100 covalently bound chondroitin sulphate chains and almost as many keratan sulphates. The chondroitin sulphate chains each consist of approximately 40 disaccharide subunits, whereas the keratan sulphate chains are shorter. The aggregate as a whole has a mass of 50–200 MDa

The sequences of the cartilage **link proteins** from several mammals and the chicken have been determined. The human polypeptide chain of 339 amino acids is organized into three domains. The first domain has the structure of an immunoglobulin subunit, although the sequence has little similarity to those of other members of the immunoglobulin super-family. The other two domains have a repetitive structure of hyaluronic acid and proteoglycan binding sites. The sequences of their approximately 100 amino acids have about 40 % similarity to each other, and apparently arose by a gene duplication [70, 211].

In addition to the proteoglycans (PGs) which form networks together with hyaluronic acid, link proteins and collagens, extracellular matrices include smaller **non-aggregating proteoglycans** of about 100 kDa with core proteins of about 40 kDa. PH I (biglycan) has two glycosaminoglycan chains, and PG II (decorin) has only one. The PGs in sclera, cartilage, skin and tendon contain dermatan sulphate, and in bone contain chondroitin sulphate. The amino acid sequence of PG I from bovine cartilage agrees in 55 % of its 331 amino acids with PG II from bovine bone. These two proteins belong to the same superfamily as that containing the leucine-rich α_2-glycoprotein (LRG) from serum, the glycoprotein Ib of blood platelets, and also the toll gene product and chaoptin from *Drosophila* [194, 236].

The synthesis of **chondroitin sulphate** in vertebrates begins with the transfer of a xylose residue from UDP-xylose to a particular serine residue of the protein component; finally, two galactose residues and glucuronic acid or, alternatively N-acetylgalactosamine are added [255]. Heparin and heparan sulphate are also bound to serine via the same trisaccharide as that which binds chondroitin sulphate (Fig. 13.13). A glycopeptide with galactose, xylose and serine has also been isolated from the glycosaminoglycan of *Loligo pealei* [123]; unfortunately, there is almost no information about the biosynthetic routes leading to the mucopolysaccharides of the invertebrates or about the structure of their protein bonds. Epimerization of part of the D-glucuronic acid residues to L-iduronic acid, and the O-sulphation on C-4 or C-6 of the galactosamine, or on C-2 or C-3 of iduronic acid, takes place after polymerization. There is usually one sulphate per disaccharide, but over- or under-sulphated chondroitin sulphate is also found. The mucopolysaccharides even of one tissue are heterogeneous in the number and position of sulphate and iduronic acid residues. According to the usual nomenclature, C-4 sulphated polysaccharides are referred to as chondroitin A, and C-6 sulphated polysaccharides as chondroitin C. However, there are frequently hybrid molecules with sulphate groups at both positions. Polysaccharides with a high proportion of iduronic acid are known as chondroitin B or **dermatan sulphate;** in this case, the remaining glucuronic acid residues are concentrated on the chain in one or two clusters [115]. As epimerization and sulphation occur after polymerization, proteoglycans with the same primary structure and the same protein components may bear different names. It would be preferable to refer to all as chondroitin sulphate, referring at the same time to the proportion of iduronic acid residues and C-4 or C-6 sulphate groups [115].

Chondroitin sulphate is considered as **oversulphated** when it includes bi- or even trisulphated disaccharide units, i.e. when there is, on average, more than one sulphate residue per disaccharide. The localization of the extra sulphate residues varies with the species. In chondroitin sulphate D from elasmobranch cartilage and in chondroitin sulphate K from the gill cartilage of the xiphosuran *Tachypleus tridentatus*, both the N-acetylgalactosamine and the glucuronic acid residue bear one sulphate residue. In the chondroitin E of the cranial cartilage of the cephalopods *Loligo* sp. and *Ommastrephes sloani*, about half the N-acetylgalactosamine residues are 4,6-bissulphated; the rest are C-4 sulphated and C-6 glucose substituted (Fig. 13.13) [124]. The N-acetylgalactosamine residues of the chondroitin sulphate from the chorda dorsalis of the agnathan *Myxine glutinosa* are 4,6,-bissulphated; the dermatan sulphate in the skin of *Myxine* and the related genus *Eptatretus* contains trissulphated subunits, in which the uronic acid is 2- or 3-sulphated and the galactosamine 2,6-bissulphated. Unlike in *Myxine*, the skin and chorda of the acranian *Branchiostoma lanceolatum* have normal dermatan sulphate, and not over-sulphated chondroitin sulphate. Cephalopod cartilage contains several proteoglycans with molecular masses between 0.6 and 1.3 MDa; these bear two to five over-sulphated chondroitin sulphate chains, and differ in both the content of hexuronic acid and sulphate groups and the size of the core proteins. These proteoglycans are also unable to integrate in vitro with hyaluronic acid [144, 283]. In addition to typical chondroitin sulphates, the body wall of some holothurians has been found to contain rather unusual fucose-containing glycosaminoglycans [145, 279]. In one such glycosaminoglycan from *Ludwigothurea grisea*, part of the N-acetylgalactosamine residues are 4,6-bissulphated; the 3-hydroxyl of the glucuronic acid residues bear either sulphate or a sulphated fucose disaccharide and they are, therefore, resistant to chondroitinase. The fucose chain has the structure Fuc(3,4-Sf_2)α(1–2)Fuc(4-Sf)α(1–3)GlcUA- [279]. However, over-sulphated chondroitin sulphate is found in the cartilage lying at the base of the tentacles of the polychaete *Eudistilia polymorpha*. The eggs of the gastropod *Pomacea* sp. contain the typical heparan sulphate and acidic galactans, as well as chondroitin sulphate residues which, however, are never C-6 sulphated and have no iduronic acid residues [134]. The skin of the cephalopod *Ommastrephes* contains sulphate-free chondroitin; in vertebrates this is found only in the cornea of the eye [123, 144].

Heparin is synthesized and stored in the mast cells of the connective tissues. Its clotting-inhibitory properties are well known; they arise from the fact that the inactivation of thrombin by antithrombin III is increased manyfold when thrombin and antithrombin are bound to the same heparin chain. Heparan sulphate, as a widely found component of the cell surface, is very different to heparin in function but the two are structurally very difficult to distinguish. Both substances show great similarity to chondroitin sulphate in their biosynthesis and the structure of

their protein bonds. These polysaccharides are also highly modified after their biosynthesis; the resulting variability of the heparins is even greater than that of the other glycosaminoglycans. The different heparin variants vary greatly in their clotting-inhibitory activity in proportion to the content of an oligosaccharide sequence necessary for antithrombin binding [118].

A typical heparin has 70–90 % iduronic acid; most glucosamine residues are N-sulphated and only a few carry N-acetyl groups. The C-6, and often the C-3, of the glucosamine is O-sulphated, as is sometimes the C-2 of many of the iduronic acids, but never glucuronic acid. Most of the heparins have a molecular mass of 5–15 kDa; individual chains carry galactose, xylose and serine as components of a protein bond. Thus, heparin is apparently synthesized as a proteoglycan and only later separated from the protein. A "macromolecular" heparin of 1 MDa, isolated from rat skin, has several polysaccharide chains of 60–100 kDa and these sit on a protein component of very unusual amino acid composition: this consists only of serine, glycosylserine and glycine and is protease resistant. **Heparan sulphate** appears to be a somewhat modified form of heparin but it is probably more correctly looked upon as an independent family of the glycosaminoglycans [89]; it has no clotting-inhibitory activity. Only a few glucosamine residues are deacetylated and can thus be N-sulphated; because of the specificity of the corresponding enzyme, this lower N-sulphation leads to a reduction in the epimerization of glucuronic acid to iduronic acid [273].

Heparin is thought to be absent from the lower vertebrates; it has been sought without success in the lamprey *Petromyzon marinus*, the shark *Squalus acanthias* and the teleosts *Sebastes marinus* and *Cyprinus carpio* [120]. As the quantity of clotting-active heparins is relatively low in mammalian organs, a practical source for this substance was sought amongst the invertebrates. In fact, the mussels *Spisula solidissima, Cypridina islandica, Anomalocardia brasiliana* and *Tivela mactroides* contain considerable amounts of heparin-like polysaccharides which have a large proportion of antithrombin-binding sequences and a corresponding clotting-inhibitory activity. The glycosaminoglycan from *C. islandica* is known as mactin [210]. An unusual heparin sulphate with many sulphate groups, little iduronic acid and low clotting-inhibitory activity has been isolated from the lobster; most of the disaccharides in this case consist of glucuronic acid and glucosamine-N,O-bissulphate, whilst the non-sulphated or mono-sulphated units of the typical vertebrate heparan sulphate are rare [120].

In vertebrates, there are two types of proteoglycan with the repetitive sequence of keratan sulphate (Fig. 13.13). Type I from the cornea has a molecular mass of about 80 kDa and contains one to three polysaccharide chains bound N-glycosidically to asparagine via N-acetylglucosamine. Such a bond has also been found in a keratan sulphate from the skin of the teleost *Scomber japonicus*. This type of carbohydrate-protein bonding is typical of glycoproteins which arise by the transfer of a dolichol-bound oligosaccharide to an asparagine residue. Type-II keratan sulphate is present in various connective tissues and is bound O-glycosidically via N-acetylgalactosamine to serine or threonine. The binding N-acetylgalactosamine residue is commonly substituted by N-acetylglucosamine or sialylgalactose [127]. Serine-bound keratan sulphate and chondroitin sulphate chains in cartilage sit on the same large, hyaluronic acid-linked core protein (Fig. 13.14). The presence of keratan sulphate in various invertebrates has been claimed but only on the basis of subunit analysis of the mucopolysaccharides and not on detailed structural analyses [123].

The approximately 1-MDa unbranched **hyaluronic acids** are found in vertebrates in, for example, connective tissue, the synovial fluids of joints and the vitreous humour of the eye. They occupy a central position in the macromolecular complexes of the cartilage matrix (Fig. 13.14) and have considerable water-binding capacity. Surprisingly, corresponding polysaccharides have been detected in bacteria. Amongst the invertebrates, the probable presence of hyaluronic acid has been shown in many species by histochemical methods but it has been isolated and chemically identified only in the insects, e.g. in the peritrophic membrane of the silkworm *Bombyx mori*, the midgut of the wax moth *Galleria melonella*, and the nervous system of the cockroach *Periplaneta americana* [123].

13.4.2 Glycan Sulphates and Glycan Phosphates

The radula (odontrophore) on the underside of the rasping tongue of the prosobranch snail *Busycon canaliculatum* has the macroscopic and histological character of a typical cartilage, but has **glucan sulphate** instead of glycosaminoglycans. Hydrolysis of the glucan sulphate gives glu-

cose and sulphate in the approximate ratio 1:1, but details of the structure are otherwise not known [123]. Glycan sulphates with obvious species-specific differences are found in the slime secretions of the molluscs. Whereas the properties of vertebrate mucus are determined by glycoproteins, in the molluscs the nature of the proteoglycan components is the decisive factor. As in other prosobranchs, the hypobranchial gland in the gill cavity of the Japanese species *Charonia lampas* produces a slime which binds foreign matter present in the respiratory water stream. This secretion was found to contain charonin sulphate, which is a mixture of glucan sulphates of varying sulphate content; the high-sulphate variety are β(1–4) linked, like cellulose, whereas the low-sulphate type have α(1–4) and α(1–6) bonds, as in glycogen. The hypobranchial gland of the common whelk *Buccinum undatum* also secretes a β(1–4) glucan sulphate in addition to a glycoprotein. On the other hand, the hypobranchial mucus of *Busycon canaliculatum*, whose odontophore contains glucan sulphate (as mentioned), has a sulphated glycan with galactose and glucosamine components. Even more complex sulphated polysaccharides built up of various hexosamines and hexoses have been found in the hypobranchial-gland secretions of *Neptunea lactea* and in the epidermal slime of the pulmonate slug *Agriolimax columbianus*. The mucopolysaccharide of the pulmonate snail *Otala lactea* actually contains uronic acid. In contrast, a simple glucan sulphate is again found in the mantle slime of the scallop *Pecten maximus* [123].

The eggs of many aquatic invertebrates are enveloped in a gelatinous membrane, the properties of which are often determined by the presence of mucopolysaccharides. Species-specific glycan sulphates have been found in the egg membranes of various sea urchins; these are referred to as **fertilisins** because of their probable involvement in sperm agglutination and fertilization. They mostly include fucose sulphates, which may be branched and in some species may also contain other hexoses as subcomponents. Other species have galactan sulphates, and at least in *Echinus esculentus* the galactose has predominantly the L- configuration. Finally, the egg jelly of the sand dollar *Echinarachnius parma* contains a fructosan sulphate. The only starfish to be investigated was found to contain a hetero-glycan sulphate with the components glucose, fucose and mannose [123].

Simple **glycan sulphates** have also been found in other animal groups, e.g. fucan sulphate in the skin of the holothurians or galactan sulphate in the mussels. More complex glycan sulphates with an even wider distribution contain various hexoses and hexosamines but no uronic acids. These include the horatin sulphate of the hepatopancreas in the snail *Charonia lampas;* polysaccharides from mussels (spisulan), the oligochaetes *Lumbricus* and *Tubifex*, and various ascidians; and lingulan from the stem of the brachiopod *Lingula* [120, 123, 208, 265]. Glycoproteins containing aminoethylphosphonic acid (AEPn) were discovered in the slime of sea anemones such as *Metridium senile* and *M. dianthus*. This sticky mucus is secreted by the ectodermal cells and serves mainly to trap plankton. It contains none of the typical acidic mucopolysaccharides but has glycoproteins with 20–30 % carbohydrate and 10–15 % AEPn ester-bound to the carbohydrate [39].

Glycan phosphates are seldom found in animals. One example of this class of substances is the acidic polysaccharide found in the tube of the polychaete *Hyalinoecia* (previously *Onuphis) tubicola*; this is composed of only glucose and phosphate and is known as onuphinic acid. The exact structure is not known but the sugar residues appear to have β(1–3) bonds and carry monophosphates in position 4 or 6. A phosphate-rich polysaccharide of mannose, glucose and xylose has been reported in the tube of the polychaete *Chaetopterus variopedatus* [123].

13.4.3 Chitin and Cellulose

The term chitin is used both for native proteoglycans and for derived polysaccharides. Chitin is widely distributed in fungi and animals; with an estimated annual production of 10^{10}–10^{11} t, it competes with cellulose for the position of the most abundant organic substance in the biosphere. In fungi, chitin is associated with other polysaccharides, whereas in animals it is probably exclusively bound to protein. The most well-known **example of chitin in animals** is the external and internal skeleton of the arthropods in which chitin also lines the anterior and posterior gut. In the annelids, chitin is found in the bristles, mouthparts and gut lining; in the molluscs it is present, for example, in the organic matrix of the shell and the radula; and in the cephalopods, it is also found in the cuttlebone, the suckers and the lining of the anterior gut. Examples of chitin-containing structures are the shells of the brachiopods; the outer skeleton of the bryozoans; the

egg shells of the platyhelminths, nematodes and acanthocephalans; the tubes of the pogonophorans; the perisarc of the hydrozoans; the internal skeleton of many hexacorallia; and the gemmules of many sponges [72, 123, 192]. Chitin appears to be completely absent from the group of phyla known as the Deuterostomia, i.e. echinoderms, hemichordates and chordates; there is, however, an unambiguous report of chitin in the gill arches of the lanzelet *Branchiostoma floridae*. Chitin has also been discovered in the arrow worms (chaetognathans), whose membership in the Deuterostomia is very doubtful [8].

Chitin is usually described as a linear β(1–4) polymer of N-acetylglucosamine but in many cases, even those involving less destructive isolation methods, some sugars turn out to be non-acetylated. This probably involves the activity of a chitin deacetylase, which is known in bacteria but has never been directly demonstrated in animals. The chitin of insect cuticulae usually contains about 1–20% deacetylated residues. The irregular abdominal expansion of the queen in the termite species *Macroterms estherae* involves **deacetylation:** the degree of deacetylation of the abdominal cuticula increases from zero in the winged sexually mature animal to over 90% during egg production [158]. The chitin chain has a conformation (ribbon conformation) similar to that of the cellulose chain and, thus, can form fibrils. The chitin chains are linked by H-bonds to produce two-dimensional structures (sheets) that are present in three different **crystal structures**. The adjacent chains of α-chitin run antiparallel, those of β-chitin run parallel, and in γ-chitin the chains are organized in groups of three (Fig. 13.15a). The α form is typical of the arthropod exoskeleton but is also detected in other chitin structures, e.g. in the perisarc of hydroid polyps, in the gastropod radula, in cephalopod "beaks" and in the bristle jaw of the chaetognathans. β-Chitin is found, for example, in the bristles of the annelids, in the cuttlebone of the cephalopod *Loligo*, in tubes inhabited by the pogonophorans, and in the cocoons of beetle species in the genera *Cionus, Cleopus* and *Murmidius*. γ-Chitin lines the stomach of the cephalopod *Loligo*, and is present in the cocoons of *Ptinus, Donacia* and several other beetles, as well as in the peritrophic membranes of the cockroach *Blaberus*, the migratory locust *Schistocera*, caterpillars of the butterfly *Antherea*, and the sawfly *Phymatocera*. Thus, *Loligo* has all three chitin forms: α in the beak, radula and oesophagus lining, β in the cuttlebone, and γ lining the stomach. The various chitin forms are clearly not specific to particular animal groups but, rather, are adapted to serve different functions. α-Chitin is typical of hard structures which are strengthened by sclerotization or calcification. In contrast to the β and γ forms, it is never found associated with collagen. In the evolution of the articulates, the α-chitin external skeleton of the arthropods replaced the collagen cuticula of the annelids [8, 123, 158, 192]. The polysaccharide chitin is always found **covalently bound to protein** and, for example, the protein content of the complex in the crustaceans used for industrial chitin extraction is 25–46%. In chitin preparations from crustaceans, insects or cephalopods, about one amino acid per 200–300 sugar residues remains even after several hours of hydrolysis in hot sodium hydroxide. These are often aspartic acids but, according to the origin of the chitin, may be other amino acids, e.g. glutamic acid, lysine or glycine. There are, however, non-covalent as well as covalent bonds between chitins and proteins. It may be that compounds derived from aromatic amino acids are responsible for chitin-protein bonds, like those which cross-link the cuticula proteins during sclerotization; and β-alanine may also be involved [158]. Additional proteins are often loosely bound to the chitin proteoglycan. The protein components of the chitin proteoglycan apparently vary considerably with the spe-

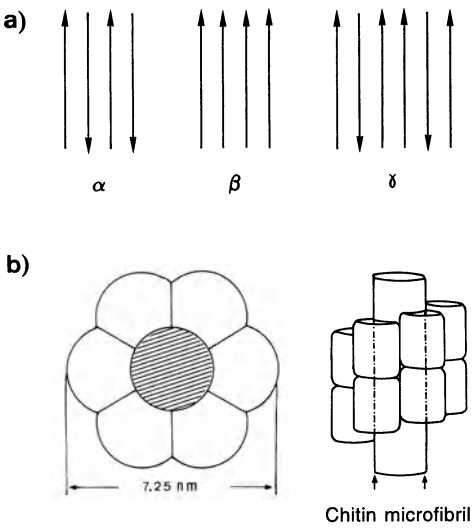


Fig. 13.15a, b. Chitin. **a** The three crystalline forms of chitin found in the animal kingdom [123]. **b** Chitin-protein fibrils from the ovipositor of the parasitic wasp *Megarhyssa* [30]. A central α-chitin microfibril is surrounded by six rows of globular protein molecules in a regular arrangement

cies, organ and chitin type. Even though several such proteins have received specific names, e.g. arthropodin or sclerotin in the arthropods, and conchiolin or conchagen in the molluscs, it is usually the case that little is known about their structural properties, except for their amino acid compositions [4, 23, 123, 192]. The arthropod cuticula is made up of chitin fibres embedded in a protein matrix. X-ray pictures show that both the chitin chains and the proteins in the ovipositor of the ichneumon *Megarhyssa* have a crystalline structure. The fibres are bundles of microfibrils, each made up of a central α-chitin fibril with a protein coat (Fig. 13.15b), packed in a regular hexagonal pattern. This is therefore a unique example of a fibre-matrix structure in which both components are strictly organized [30].

Chitin biosynthesis in the arthropods differs from that in the fungi by the involvement of dolichol phosphate (dol-PP). It has been shown in the brine shrimp *Artemia salina* that a specific transferase transfers a sugar phosphate residue from UDP-N-acetyl-D-glucosamine to dolichol phosphate to give dolichol pyrophosphate-N-acetyl-D-glucosamine (dol-PP-GlcNAc). As will be shown, this compound is the starting point for the production of asparagine-bound oligosaccharides in glycoproteins. In chitin biosynthesis, the carbohydrate components bound to the dol-PP are extended to form approximately octasaccharides by further N-acetylglucosamines, and are then transferred to a peptide acceptor. The resulting chitoprotein is the primer for **chitin synthase** [119]. Animal chitin synthase is a microsomal, integral membrane enzyme which is extremely labile and has so far proved impossible to isolate. However, cell-free systems have been prepared from *Artemia* and various insects and these catalyze transfer of N-acetylglucosamine residues from UDP to chitin oligosaccharide or an endogenous primer. The insecticide diflubenzuron, which in vivo inhibits the incorporation of glucosamine into chitin, is variable in its effects upon these in vitro systems [54, 158].

Cellulose was discovered as a component of animal tissues in 1858 in the mantle (tunica) of the ascidians, and was named tunicin. The cellulose of the tunicates corresponds in all its properties to plant cellulose. It is made up exclusively of linear β(1–4) linked glucose residues; the chains show the typical conformation where 250 chains are associated into fibrils that are 10 nm thick, and the fibrils are bundled into fibres of about 70 nm. The cellulose of the tunicates is bound covalently to protein and the proteoglycan is non-covalently linked to hyaluronic acid and further proteins. Fibres of protein-bound cellulose have also been found in the skin and aorta wall of humans and various mammals. The concept has been put forward that cellulose from the chordates corresponds to the chitin of other animal phyla, perhaps even with homologous enzymes for their synthesis [123, 162].

13.4.4 Glycogen, Galactan and Other Reserve Polysaccharides

Amongst the multicellular animals glycogens are universal as storage polysaccharides, their concentrations varying by orders of magnitude with the species, tissue and physiological state. Thus, the glycogen content of the liver of a starving mammal is 1–5 mg/g fresh weight, but on feeding increases by up to 100-fold. The highest values of about 50 % of the dry weight are found in the endoparasitic nematodes and cestodes, which have a particularly high requirement for carbohydrate as a result of their anaerobic metabolism. All the glycogens are branched glucans with α(1–4) and α(1–6) bonds and an average chain length of 10–13 sugar residues. Rabbit and porcine liver glycogen normally includes glucosamine in addition to glucose in the ratio 1:50 000 [149]. The degree of branching and the chain length of the glycogens vary with the species and the tissue, and may even be heterogenous in the same cell. Thus, for example, the muscle glycogens of the teleost species *Catla catla* have consistently longer chains than those of *Clarias batrachus*, and the glycogens of the white muscle of both species have longer chains than those of the red skeletal muscle and heart muscle. In the mussel *Mytilus galloprovincialis*, the average chain length in the hepatopancreas is shorter than in the mantle tissue [123, 161].

Values of over 100 MDa have been recorded for the **molecular mass**, and both species- and tissue-specific differences have been described; however, in view of the large molecular masses, artefacts of isolation are difficult to exclude. In most cells, glycogen forms self-limiting spherical particles that are about 15–30 nm in diameter and have a mass of 4–10 MDa (β-particles). In particularly glycogen-rich tissues, such as the liver and kidney of various vertebrates or the heart muscle of diving mammals, the β-particles may aggregate to form rosette-shaped structures with a diameter of 100 nm and a mass of 400 MDa (α-particles); molecular masses of up to 100 MDa

have also been reported in mammalian brain [49]. In certain active tissues, such as the eye muscle of some mammals or the heart muscle of mammals and lungfish (Dipnoi), the cisternae of the endoplasmic reticulum (ER) contain glycogen-membrane complexes that are about 35 nm in diameter, the so-called glycogen bodies. In such glycogen granulae, the polysaccharide is associated with proteins, especially the enzymes of glycogen synthesis and catabolism, and the corresponding regulatory enzyme cascade. Glycogen in the retina has glycosidic bonds to the phenolic hydroxyl group of tyrosine residues [6].

The primer for **glycogen biosynthesis** in muscle and liver is glycogenin, a protein of 332 amino acids which is released proteolytically from a high molecular weight precursor. In the absence of glycogen, glycogenin is bound to glycogen synthase. The first step in glycogen synthesis is the binding of a glucose to the 194-Tyr of glycogenin. Up to six further glucoses are added to the first, with the glycogenin itself functioning as a UDPG-specific glucosyltransferase. Further extension of the carbohydrate chain is brought about by the glycogen synthase, which for this purpose dissociates from glycogenin. Glycogen synthase transfers glucose residues from UDP-Glc to the primer polysaccharide with α(1–4) bonding. When the α(1–4)-linked chain reaches a length of 12–18 glucose residues, the "branching enzyme" creates an α(1–6)-bound sugar and, thus, a branch point. The primer protein remains bound and consequently each glycogen particle includes one molecule of glycogenin [42, 249]. The glycogen synthases from human muscle (737 amino acids) and rabbit muscle (734 amino acids) show 97% sequence agreement. The enzymes from mammalian muscle and liver have a total of up to ten sites in their N- and C-terminal regions, and these may be phosphorylated by one of seven different protein kinases. The enzyme of rat liver (703 amino acids) is about 80% similar to the mammalian muscle enzyme in its central region, but shows large terminal differences; the markedly shortened C-terminal region lacks two of the phosphorylation sites [9, 38, 298]. Mammalian muscle contains an almost phosphate-free I-form of the enzyme (I for independent) together with various D-forms (D for dependent) which arise by the phosphorylation of one or more serines and which are active only in the presence of glucose phosphate. The enzyme from rabbit muscle normally has 2.4–3.0 phosphates per subunit; the phosphate content is increased by adrenalin and reduced by insulin. The liver enzyme includes up to six alkali-labile phosphates. In mammalian kidney, the dephosphorylated I form is also activated by glucose-6-phosphate, and in muscle of the frog *Rana ridibunda* the activity of both forms is increased by glucose-6-phosphate. The I/D conversion of the vertebrate enzyme can occur in vitro through the action of various protein kinases, some of which are cAMP-dependent and some cAMP-independent [264]. In the invertebrates so far examined, I- and D-forms of glycogen synthase can be distinguished; however, these differ from each other and from the vertebrate enzymes in their allosteric properties [83, 106, 109].

Some **unusual reserve polysaccharides** have been found in the protozoans. *Tetrahymena* has a typical glycogen with an average chain length of 13 units. In contrast, the α(1–4), α(1–6) glucans from the rumen ciliates of the genera *Entodinium, Isotricha* and *Dasytricha* have chain lengths of 22–25 units; this is characteristic of the plant amylopectins. *Entodinium caudatum*, the gregarines and the sporozoon *Eimeria tenella* have intermediate chain lengths of about 19 sugar residues [284]. Homo- and heteroglycans with a different composition have been identified in other Protozoa; these are probably components of glycoproteins and fulfil functions other than nutrient storage. They include, for example, D-mannan and D-arabino-D-galactan in *Crithidia fasciculata*, and the β(1–2)-mannans and galactomannans found in several trypanosomes. *Trypanosoma mega* has no β(1–2)-mannan but, instead, a mixture of branched galactans with (1–2) and (1–3) bonds and several fucose, glucose and mannose residues [44, 261]. The main polysaccharide in the eggs of the migratory locust *Locusta migratoria* is a mannan, with about 12% glucosamine, which is bound covalently to protein; it makes up about 10% of the dry weight of freshly laid eggs but is rapidly metabolized during embryo development [293].

Galactans have been found in all pulmonate snails examined and in the prosobranch *Ampullarius* sp. This polysaccharide was discovered by Hammersten in 1885 and was initially known as "sinistrin"; in view of its structural and functional relationship to glycogen, it was later given the name "galactogen". The galactans are produced in the albumen glands and accumulate in the perivitelline fluid of the egg, where they constitute a nutrient reserve for the embryo. The galactan content of, for example, freshly laid eggs of *Helix* or *Lymnaea* is 30–36% of the dry weight, and the galactan is rapidly consumed during later embryo development *(Lymnaea)* or after hatching

(Helix). It is rather puzzling that the nutrient reserve for gastropod embryos is not glycogen, as it is in all other metazoans. The galactans of the gastropods are all polysaccharides of D-galactose residues with species-specific degrees of branching. Depending upon the origin of the galactans, L-galactose, L-fucose, L-glucose and phosphate residues may also be present. Most preparations contain amino acids but it is not clear whether in vivo the snail galactans really exist as protein complexes. Isolation with the help of lectins produces almost N-free polysaccharides. The galactans of different gastropods vary greatly in their reactivity with lectins and antibodies [152].

The D-galactose residues are predominantly β(1–3) or β(1–6) linked, although relatively few are found in linear regions and most are terminal or occur at branch points; some species also have β(1–2) and β(1–4) bonds. L-Galactoses probably always occur terminally [153]. Little is known about the galactan-synthesizing enzymes; β(1–3)- and β(1–6)-galactosyl-transferases have been isolated from the albumen glands of *Lymnaea* and *Helix* [98, 137]. The galactans of the aquatic Basommatophora (e.g. *Lymnaea, Biomphalaria)* are highly branched with only 5–8 % of the sugar residues in linear sections. L-galactose and β(1–2) bonds are lacking, and β(1–3) and β(1–6) bonds alternate more-or-less regularly. In the terrestrial Stylommatophora (e.g. *Helix, Arianta, Cepaea, Achatina)* up to 20 % of the sugar residues are linear β(1–3) bound. L-galactose and linear β(1–2) bonds are found only in the galactans of the family Helicidae. The galactans of *Ampullarius* sp., the only prosobranch examined so far, have an unusually large proportion of linearly arranged sugars, with 5 % β(1–3), 26 % β(1–6) and 10 % β(1–2). In contrast to those of the pulmonates, some of the D-galactose residues in *Ampullarius* sp. are replaced by fucose. Phosphate-substituted galactose residues are found in the galactans of individual species from various snail genera, e.g. in *Biomphalaria, Helix* and *Cepaea*, but are absent from, e.g., *Lymnaea* and *Arianta* [117]. In addition to heparin, the mammalian lung also contains a β(1–3), β(1–6)-D-galactan which, however, includes 2–3 % glucuronic acid residues [123].

13.5 Glycoproteins

Almost all proteins of the cell surface and the extracellular space bear carbohydrate chains; **glycosylation** is one of the most important mechanisms for the post-translational modification of proteins. This is achieved by the transfer of mono- or oligosaccharides to the amide-N of certain asparagine residues or to the -OH groups of serine, threonine, hydroxylysine or other hydroxy- amino acids. The biological importance of glycosylation lies on the one hand in the carbohydrates as recognition signals, and on the other hand in the resulting effect on the folding of the polypeptide chain, which influences its chemical and biological properties. The information content even of small oligosaccharides is very high; for example, the various possibilities for the anomeric configuration and position of the glycosidic bonding of three hexoses allows the formation of almost 1000 different trisaccharides, whereas in the case of three amino acids or nucleotides only 6 different combinations are possible. Several thousand carbohydrate sequences of oligosaccharides, polysaccharides and glycoconjugates have been determined and published to date [57]. The importance of carbohydrate structure for the **recognition of proteins** is clearly illustrated by the endocytosis of plasma glycoproteins by mammalian liver cells. Enzymatic cleavage of the terminal sialic acid of the carbohydrate chains of plasma proteins exposes galactose residues which are recognized by receptors on the cell surface. The asialoglycoprotein-receptor complex is taken up into the cell, the plasma protein is degraded in the lysosomes, and the receptor is recycled to the cell surface. Human hepatocytes have two receptor proteins (H1 and H2) of about 45 kDa whose cDNA-determined amino acid sequences are about 58 % identical. However, H1 shows high agreement with the corresponding sequence in rat, i.e. the gene duplication which produced the two different receptor sequences occurred before the separation of the human and rat evolutionary lines. The biological significance of the existence of two receptor proteins in hepatocytes is not clear; it is possible that they form dimers [252].

The oligosaccharide chains of the glycolipids and glycoproteins form a layer (the **glycocalyx)** on the surface of many cells and this is important in cell-cell interaction between cells of the same organism and with foreign cells [93]. Proteolytic cleavage of membrane proteins from tissue culture cells of all vertebrate classes releases about 60 glycopeptide fractions which are evolutionarily very conservative within the vertebrates in their size, lectin specificity, charge and chromatographic behaviour, but which differ greatly from, for example, those in the eggs of the sea urchin *Arbacia punctulata* [31].

There are many examples illustrating the **influence of the carbohydrate component on the conformation and physicochemical properties of glycoproteins**. The carbohydrate portion of the integral membrane proteins is critical to their arrangement in the membrane and to their stability [25, 93]. The repulsion between molecules of the negatively charged sialic acid-rich, 140-kDa glycoprotein podocalyxin on the podocytes of the kidney glomeruli ensures that the filtration slits between the interdigitating cell processes remain open [65]. The sialic acids of the slime substances (mucins) are responsible for their high viscosity and lubrication properties; the high density of charges forces the molecule into a stretched rod-shaped form and with many carbohydrate chains attached it resembles a bottle brush [2, 25]. The slime substances secreted by fish epidermis greatly reduce friction, even in dilute solutions with mucin concentrations below 0.1 %. Thus, swimming fish are constantly enveloped in a thin film of slime which helps them to maintain considerably higher velocities. Unfortunately, little is known about the molecular structure of fish mucin. Glycoproteins are important components of the jelly substances enveloping the eggs of many vertebrates from the fish to the birds [77, 131]. The gel-like consistency of the albumen of birds is due to the cross-linking of the glycoproteins ovomucin and lysozyme (p. 388). In the antifreeze glycoproteins in the blood plasma of antarctic fish, it is the disaccharides lying on the surface which, by an unknown mechanism, prevent the formation of ice crystals (see Fig. 5.3, p. 207). The mechanical properties of connective tissues are also determined by interaction of their protein and carbohydrate components with water and ions (p. 482).

There are **two types of glycoprotein** which differ fundamentally in their structure and biosynthesis: the mucin type with O-glycosidic bonds between N-acetylgalactosamine and serine or threonine (Fig. 13.16), and the asparagine-bound type with N-glycosidic bonds between N-acetylglucosamine and asparagine (Fig. 13.17). O-glycosidically bound sugars are usually clustered in certain regions of the polypeptide chain which are rich in serine and threonine. Interactions between these regions induces the polypeptide to assume a stiff, extended conformation. In the membrane glycoproteins epiglycanin and leucosialin, the whole of the extracellular domain is extended due to O-glycosylation. Other membrane proteins, such as sucrase-isomaltase or the LDL (low density lipoprotein) receptor, have O-glycosylated stems which hold the active part of the molecule high above the glycocalyx surface [133]. The best-examined O-glycosylated proteins are the **mucins** from the submaxillary, intestinal, tracheobronchial and cervical slimes of the mammals. Mucins are linear, unbranched molecules which contain about 75 % carbohydrate and have a relatively stiff, random-coil conformation. They are all polydisperse with molecular masses of 6–24 MDa, and apparently have a complicated subunit structure. Reduction of the disulphide bridges releases subunits of 0.5–2.5 MDa and so-called link proteins from the intestinal and tracheobronchial mucins; enzymatic deglycosylation produces apomucins of about 100–160 kDa. The cDNA sequences show that the glycosylated domains of the polypeptide chains are built up repetitively. The oligosaccharides bound to serines and threonines via N-acetylgalactosamine residues are heterogeneous and often have bran-

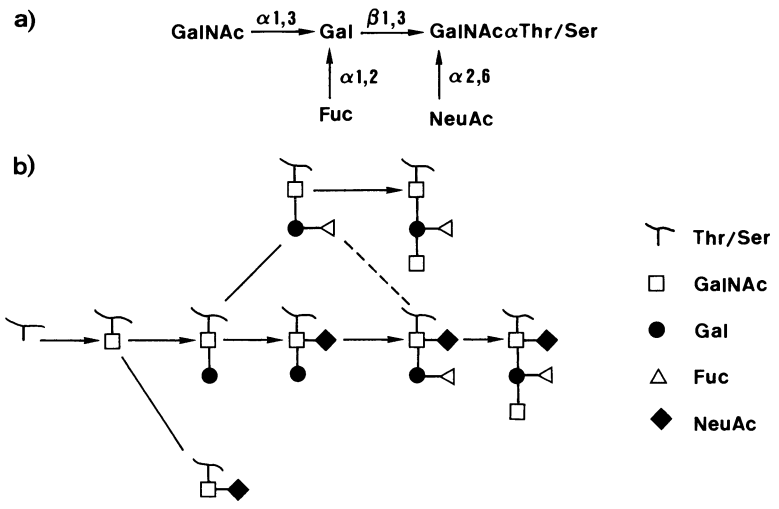


Fig. 13.16a, b. The glycoproteins of the mucin type [25]. **a** The structure of the most complex carbohydrate chain. **b** The biosynthetic pathway and the possible structural variants. Only the clearly indicated glycosyl reactions are possible; that indicated by the *dashed line* is very slow

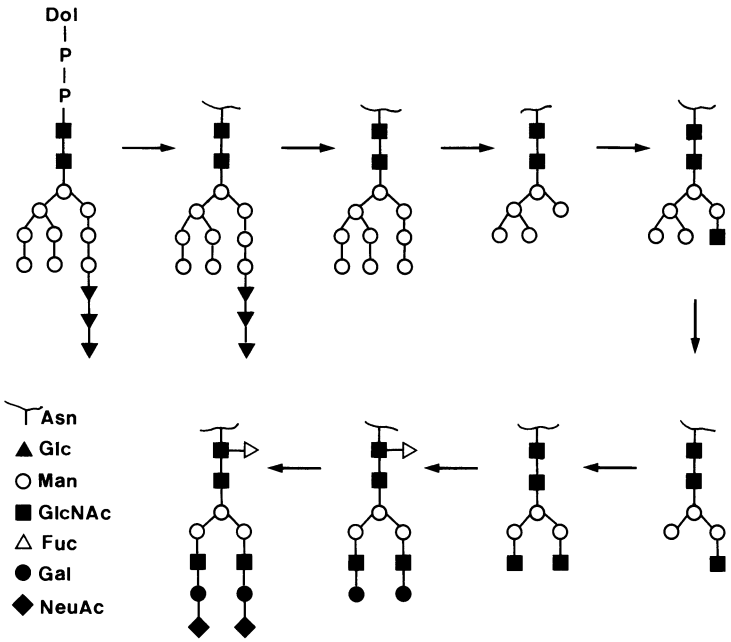


Fig. 13.17. The biosynthetic pathway of the asparagine-bound glycoproteins [25]. The characteristic oligosaccharide formed by sequential glycosyl transfer to dolichol phosphate is transferred from the lipid to a particular asparagine residue of the protein; examples are shown of how shortening and repeated extension of the carbohydrate chain can lead to multiple variants

ches of up to 20 sugars [104, 105, 133]. The mucins in the skin slime of the clawed frog *Xenopus laevis* are composed of subunits of about 150 kDa with about 50% carbohydrate. The polypeptide chain has a highly glycosylated central domain of threonine- and proline-rich repeats and two cysteine-rich terminal domains [219].

The **biosynthesis of glycoproteins** begins either during or immediately after polypeptide synthesis in the ER. Later alterations may also take place in the Golgi apparatus, from where the secretory or membrane proteins are transported to the cell surface. For glycoproteins of the **mucin type**, synthesis begins with the transfer of a UDP-bound N-acetylgalactosamine to a serine or a threonine. The carbohydrate is extended step wise, with a specific glycosyltransferase being responsible for each step ("one linkage one enzyme" hypothesis). The number of carbohydrate chains, their length, and the charge arising from their sialic acids and sulphates varies with the species and the tissue and is quite heterogeneous, even in the mucus of one organ. The synthesis of the carbohydrate chain may stop at an intermediate step ("incomplete chains"); at forks in the biosynthetic pathway, the relative proportions of the different enzyme activities determine the ratio of the various products (Fig. 13.16). For example, 95% of disaccharides in the submaxillary mucin of the sheep have the sequence NeuAc-GalNAc-Thr/Ser, which is found hardly at all in the pig. This may be explained by sialyltransferase being the predominant of the two competing enzymes at the first branch point in the sheep, whereas galactosyltransferase dominates in the pig [2, 25, 255].

The long-chain isoprenoid dolichol phosphate (p. 610) is involved in the **biosynthesis of the asparagine-bound glycoproteins** [156] (Fig. 13.17). As a rule, N-acetyl-glucosamine phosphate is transferred from UDP-GlcNAc to dolichol phosphate by the membrane-bound transferases of the ER; a further N-acetylglucosamine is then added, followed consecutively by nine mannose residues and three glucoses; UDP- and GDP-bound as well as dolichol pyrophosphate-bound sugars play a role. The oligosaccharide is transferred from the lipid to an asparagine residue; however, glycosylation is only possible with the sequence Asn-X-Ser/Thr, where X can be almost any other amino acid. The carbohydrate chains transferred to proteins in the ER are further processed in the Golgi apparatus. The terminal glucose and mannose residues are cleaved off and the shortened structure is extended by the addition of N-acetylglucosamine, galactose or sialic acid (Fig. 13.17). The responsible glycosyltransferases are anchored in the membranes of the ER and Golgi apparatus, but soluble enzymes of this type are also known in milk and other body fluids. It has been estimated that at least 100 different substrate-specific glycosyltransferases are required for the formation of the known carbohydrate structures of glycoproteins and glycolipids. Surprisingly, the six glycosyl-

transferases sequenced so far via the cDNA have a similar domain structure but absolutely no sequence similarity [207].

Such a biosynthetic process allows for an extraordinarily large **variety of carbohydrate structures** amongst the glycoproteins of the asparagine-bound type. Individual polypeptide chains can carry several oligosaccharides, e.g. two in the case of transferrin and immunoglobulin IgG, three in pro-thrombin, four in fibrinogen and antithrombin III, five in the acidic α_1-glycoprotein, and six in fibronectin. The carbohydrate chains may be variously branched (bi-, tri-, tetra-antennary) but in fact bi-antennary chains occur most frequently (Fig. 13.17). Different molecules of a protein can have different carbohydrate components; for example, 60% of human serum transferrin molecules have two bi-antennary chains and the rest have a tri- or even tetra-antennary chain in addition to a bi-antennary one [294].

The lower vertebrates already show deviations from this scheme. The eggs of the rainbow trout *Salmo gairdneri* contain **polysialoglycoproteins (PSGPs)** with a very unusual structure. The main component in the unfertilized egg is a PSGP of 200 kDa which contains about 85% carbohydrate, of which more than 50% is sialic acid. The sequences of the core proteins (apoPSGPs) consist predominantly of about 25 tridecapeptides with the sequence DDATSEAATGPSG; these are glycosylated on the underlined amino acids. These core proteins are encoded by a multi-gene family with more than 100 members which vary in the number of 39-bp repeats. After fertilization, the PSGPs enter the perivitelline space by exocytosis; here, they are proteolytically cleaved into 9-kDa fragments with only one repeat sequence. The oligosaccharides bound to a threonine or serine of the core protein via a GalNAc residue contain galactose, N-acetylgalactosamine and fucose as well as stretches of α(2–8)-linked N-glycolylneuraminic acids, which may be capped by a deaminoneuraminic acid (KDN), or chains of α(2–8)-linked KDN residues; they are resistant to sialidases [142, 250]. The name "hyosophorins" has been proposed for these and similar proteins in the eggs of the carp *Cyprinus carpio* and other teleosts. A completely new type of hyosophorin has been isolated from the flounder *Paralichthys olivaceus*. Its core protein is made up of repetitive decapeptides, each of which carries a sialic acid-free, penta-antennary oligosaccharide bound to an asparagine [243]. The glycoproteins from the eggs of the char *Salvelinus leucomaenis* have been found to contain 6-deoxyaltrose; this is the first report of this unusual deoxyhexose in animals [128].

The **glycoproteins of the invertebrates** show similarities, but also extensive differences, to those of the vertebrates both in biosynthesis and in the resulting structures. Sialic acid is not present in most phyla. It has been mentioned already that the difference between proteoglycans and glycoproteins is much less marked in the invertebrates than in the vertebrates, and that glycosaminoglycans or glycan sulphates are also found in invertebrates as significant components of mucous secretions and gels. However, in addition to proteoglycans, the representatives of various phyla have been shown to possess typical glycoproteins; such phyla include flagellates, cnidarians, cestodes, nematodes, annelids, insects, molluscs, echinoderms and tunicates [77, 123, 181, 200]. Detailed investigations of the biosynthesis, structure and function of glycoproteins have been carried out only for insects [41, 121] and Kinetoplastida [44, 182, 292].

The biosynthesis of asparagine-bound glycoproteins in tissue culture cells of the mosquito *Aedes albopictus* is very similar to the process in the higher vertebrates. The cells synthesize $Glc_3Man_9GlcNAc_2$ bound to dolichol pyrophosphate and transfer this to a protein, after cleavage of the glucose residues. Six further mannose residues are subsequently removed, leaving a chain of only three mannose residues, as is found in the mammals; complex chains are not produced. None of the N-bound carbohydrates contains sialic acid, galactose or fucose [121]. An interesting glycoprotein has been isolated from the basement membranes of *Drosophila* Kc tissue culture cells, and has been given the name papilin. This is a glycoprotein of 900 kDa with 64% carbohydrate, including hexosamine, uronic acid and neutral sugars. The protein component has a mass of about 400 kDa and contains 25% aspartate/glutamate, 22% serine/threonine and carries about 80 O-linked sulphated oligosaccharide chains. The molecule is a thread with a length of 225 nm and a diameter of 15 nm and is held in the form of a loop by disulphide bridges. Several such protomers may be bound together by further disulphide bridges to give a cloverleaf-shaped oligomer. This proteoglycan-like glycoprotein is very different to the heparan sulphate proteoglycan in the basement membranes of vertebrate cells. However, the other components of the *Drosophila* basement membranes, laminin and collagen IV, are very similar to those of the vertebrates [41].

A glycoprotein isolated from the egg jelly of the starfish *Asterias amurensis* has a tri-antennary carbohydrate chain with the formula $Man_{9-12}GlcNAc_2$ [77]. A glycoprotein from the mucus of the coral *Acropora formosa* has no similarity to the mucus of higher vertebrates; it consists of a polypeptide chain with 77 % serine/threonine bearing O-linked carbohydrate chains of varying, consistently unusual structures which involve mainly D-arabinose, D-mannose and N-acetylglucosamine but have no sialic acid or uronic acid [181]. Many of the haemocyanins contain N-linked carbohydrates, whereby the structures of the oligosaccharides of the arthropods and molluscs are very different (see Fig. 7.10, p. 278). There is probably only one glycosylated haemoglobin and that occurs in the pulmonate snail *Biomphalaria;* the oligosaccharide structure has not yet been determined.

The variable surface glycoproteins (VSGs) of the African *Trypanosoma* species were discussed in Chapter 6 in connection with their characteristic immunological functions. The particularly well-investigated VSG variant MITat 1.4 of *T. brucei* bears two asparagine-bound oligosaccharides. The oligosaccharide on 432-Asn is always mannose rich ($Man_{4-7}GlcNAc_2$), whilst the initial $Man_3GlcNAc_2$ on 419-Asn is later processed to the bi-antennary structure $(GalGlcNAcMan)_2ManGlcNAc_2$. In all Trypanosomatidae examined, formation of the asparagine-bound oligosaccharides of the VSGs involves dolichol; however, the dolichol-PP intermediates never contain glucose residues but have only six, seven or nine mannose residues and two N-acetylglucosamines in species-specific proportions. After transfer to the VSG, a single glucose residue may be added to the asparagine-bound oligosaccharide but this is then removed again by a specific glucosidase [15, 34]. Thus, the Trypanosomatidae differ from the mammals, insects, higher plants and fungi, where $Glc_3Man_9GlcNAc_2$ is produced on dolichol and then transferred to the acceptor protein (Fig. 13.17). The South American *T. cruzi*, the agent of Chagas disease, has no VSGs but possesses invariable glycoproteins of another type, which appear in stage-specific forms of 25–90 kDa (p. 241). The trypomastigotes which penetrate the cells of the mammalian host are protected by these glycoproteins from recognition and destruction by lysosomes. The 90-kDa glycogen (GP90) of the trypomastigotes carries asparagine-bound oligosaccharides that are similar to the mannose-rich type of the mammals; dolichol is present in this case. Antibodies against GP72 prevent the conversion of epimastigotes from the insect host to the trypomastigotes of the mammalian host. In addition to the typical N-linked oligosaccharides of mannose and N-acetylglucosamine, GP72 carries O-linked oligosaccharides which contain galactose, fucose, ribose and xylose and are highly phosphorylated. It is the latter oligosaccharides which are recognized by the transformation inhibitory antibodies. The smaller structures GP37, GP31 and GP25, present on the epimastigotes, have not been investigated in detail. As well as glycoproteins, the plasma membrane of *T. cruzi* includes an unusual lipopeptidophosphoglycan which will be described in Chapter 15.

Crithidia fasciculata, which also belongs to the Kinetoplastida, contains the oligosaccharide $Man_7GlcNAc_2$, both bound to dolichol pyrophosphate and as carbohydrate chains on glycoproteins. Thus in this case the carbohydrate chains transferred from lipid to the protein differ from the typical vertebrate chains by the absence of glucose. On the other hand, a $Glc_3Man_9GlcNAc_2$ bound to dolichol pyrophosphate is found in the flagellate *Euglena gracilis* [43]. Mature glycoproteins of *C. fasciculata* also contain small amounts of $Gal_1Man_6GlcNAc_2$, in which the galactose residue surprisingly all have the furanose configuration [182]. The **dolichol** of *C. fasciculata* consists of only 11 isoprene units, whereas that of *T. cruzi* has 13 units; thus, both are significantly smaller than the dolichols of yeast, higher plants and vertebrates, which have 17–20 units [44]. The surface of the kinetoplastid parasite *(Leishmania tarentolae)* of the lizards has a mixture of glycoproteins which consists mainly of linearly bound galactose residues [292].

13.6 Carbohydrases

Enzymes that hydrolyse complex carbohydrates were traditionally termed carbohydrases and assigned to either the polysaccharidases or the glycosidases, according to the molecular weights of their substrates. However, the International Union of Biochemistry (IUB) enzyme commission uses the term glycosidase in a broad sense for all enzymes that hydrolyse glycosidic bonds, e.g. in glycosides, oligo- or polysaccharides, glycopeptides, glycolipids or nucleosides. Glycosidases are specific for the sugar released and the anomeric configuration of the glycosidic bond

(Table 13.4); the shorter the chain, the greater are their activities with oligosaccharides, with maximum activities found for disaccharides. The polysaccharidases can be divided into endo- and exohydrolases. Endohydrolases hydrolyse internal glycosidic bonds of carbohydrate chains and produce oligosaccharides or polysaccharides with a lower molecular weight; their activities decrease with decreasing chain length. Exohydrolases cleave disaccharides from the non-reducing ends of longer chains. Examples of this latter type are the β-amylase of plants, which releases maltose from α(1–4)-glucans, and cellobiohydrolase, which cleaves cellobiose from cellulose. Some glycosidases also function as glycosyltransferases, transferring sugar residues to various acceptors; hydrolysis in this case appears as a special case of glycosyl transfer with water as the acceptor. Transferase activity is found, for example, with many α-glucosidases and β-fructosidases but not with trehalases. Thus, for example, the complicated oligosaccharide spectrum of bee honey or the honeydew of the aphids is not completely developed in the sap, but is largely the result of the activity of transferases from the animal itself or from the plant (p. 468).

Digestive secretions of gut extracts from many animals have the capacity to hydrolyse a wide range of oligo- and polysaccharides (Table 13.4). If **glucosidase specificity** were related only to the nature and anomeric configuration of the sugar and its glycosidic bond, then relatively few glycosidases would suffice for all these hydrolytic activities. In fact, the specificity of the glycosidases is much narrower than this, as is illustrated by the trehalases and the cellobiases. Consequently the total number of different glycosidases present in an organism can only be deduced by their isolation and characterization. Such data are available for humans and a few mammals. In general, the **variety of carbohydrases** is greater in herbivores than in other nutritional types, but evolutionary adaptation of the carbohydrases to the type of nutrition is by no means clear in all cases. For example, carnivores possess enzyme activities for which there is no immediately obvious biological function. There is clearly a danger of drawing false conclusions about apparently superfluous enzyme activities; for example, it is not immediately obvious why the blood-sucking bug *Rhodnius prolixus* possesses a complex spectrum of different glycosidases, but the physiological role is probably related to digestion of the endosymbiontic bacteria (*Nocardia rhodnii)* whose mass may reach 5 % of the bug body weight after blood intake [224].

Table 13.4. Glycosidases

a) The glycosidic bonds cleaved by common glycosidases, and typical disaccharides in which they occur

α-glucosidic:	sucrose	= Glcα (1–2)βFru
	maltose	= Glcα(1–4)Glc
	isomaltose	= Glcα(1–6)Glc
	turanose	= Glcα(1–3)Fru
	trehalose	= Glcα(1–1)αGlc
β-glucosidic:	cellobiose	= Glcβ(1–4)Glc
	gentiobiose	= Glcβ(1–6)Glc
α-galactosidic:	melibiose	= Galα(1–6)Glc
	raffinose	= Galα(1–6)Glcα(1–2)βFru
β-galactosidic:	lactose	= Galβ(1–4)Glc
β-fructo-furan-osidic	sucrose	= Fruβ(2–1)αGlc
	gentianose	= Fruβ(2–1)αGlc(6–1)αGlc
	raffinose	= Fruβ(2–1)αGlc(6–1)αGal

b) Cleavage of carbohydrates in the digestive organs of several invertebrates [163, 193]

Carbohydrate	Hp	Tf	Cc	Ar
Sucrose	+	+	+	+
Lactose	+	–	+	+
Maltose	+	+	+	+
Starch = α(1–4),α(1–6)-glucan	+	+	+	+
Glycogen = α(1–4),α(1–6)-glucan	+	+	+	+
Yeast glucan = β(1–2), β(1–3)-glucan	+			
Laminarin = β(1–3)-glucan	+	+	–	–
Carboxymethylcellulose	+	+	+	+
Chitin = β(1–4)-poly-N-acetyl-glucosamine	+		+	+
Xylan = β(1–4)-xylan	+		+	+
Caragenin = α(1–3)-galactan		–	–	–
Hp galactan = β(1–3),β(1–6)	+			
Agar = β(1–3)-galactan		–	+	+
Mannan = β(1–4)-mannan	+			
Alginic acid = β(1–4)-polymannuronic acid		+	+	+

Hp, *Helix pomatia*; Tf, *Tegula funebralis*; Cc, *Crangon crangon*; Ar, *Asterias rubens*; +, is hydrolysed; –, is not hydrolysed

Specific nutritional adaptation of carbohydrase activity by enzyme induction is often found in bacteria and mammals, and in the honey bee the provision of cane sugar increases the trehalase and sucrase activities in the haemolymph [36]. The complete digestion of a polysaccharide requires at least one endo- and one exoglycosidase which usually appear together, e.g. amylase and maltase, chitinase and β-N-acetylglucosaminidase (chitobiase), cellulase and cellobiase. In the vertebrates, the components of such an enzyme system are often restricted to different consecutive sections of the gut, corresponding to the order of their activities. Thus, human salivary amylase functions in the mouth cavity and the stomach, the pancreas amylase functions in the gut cavity, and maltase occurs in the microvillus layer of the small intestine.

Intracellular glycosidases are found in all tissues. Their biological role is probably related to the conversion and degradation of bound carbohydrates and glycoconjugates. In the case of lysosome enzymes, the pH optima are characteristically 3–5; however, enzymes with such pH optima are also found extracellularly, e.g. in mammalian blood plasma. There have been few comparative investigations of glycosidases with functions other than digestion, but such enzymes are known in all animals down to the Protozoa. For example, at least nine different glycosidase activities have been detected in the blood forms of *Trypanosoma brucei* and these are probably involved in the turnover of the variable surface glycoproteins [44].

The literature is full of information about the hydrolytic effects of digestive secretions, other body fluids, and tissue extracts of many different vertebrates and invertebrates on a wide range of glycosides, oligosaccharides and polysaccharides. In many cases, not only have the activities been demonstrated but functional properties of the enzymes, such as substrate affinity and specificity, cofactor dependence, pH optimum and temperature dependence, have been investigated. However, there is as yet little information about the molecular structures of the carbohydrases and little can be said, therefore, about their evolution.

13.6.1 Glycosidases

The human and mammalian guts contain several **α-glucosidases** which are referred to as maltase, isomaltase and sucrase, according to their substrate specificities; there is also a trehalase (Table 13.4). The enzyme activities sucrase-isomaltase and maltase-glucoamylase are involved in the last steps of starch digestion, during which they cleave glucose residues from dextrins, maltose, isomaltose and linear glucans. Both disaccharidases, together with various peptidases, are bound to the microvillous membrane of intestinal cells. Their substrates are cleaved by "membrane-contact digestion" and the products are transferred into the cells by an Na^+-dependent, hydrolase-linked transport system. All enzymes bound to the microvillous membrane are anchored via a non-polar, N-terminal domain and form homodimeric complexes. The maltase-glucoamylase is synthesized as a single polypeptide of about 240 kDa but is cleaved at one position by a pancreatic protease after insertion into the membrane [199]. Maltase-glucoamylase and trehalase have also been isolated from the brush border of kidney tubules of various mammals, and sucrase-isomaltase has been isolated from the kidney of the goose *Anser anser;* apparently, the kidney trehalase is anchored to the membrane via phosphatidylinositol [87]. α-Glucosidases are also found in the liver and other organs, where they occur as either acidic lysosomal or neutral cytosolic enzymes. Human lysosomal α-glucosidase agrees in 36–41 % of its 951 amino acids with the isomaltases and sucrases [114].

In invertebrates, the digestive glycosidases may be freely solubilized in the gut or more-or-less firmly bound to the gut cells. In caterpillars of the moth *Erinnyis ello*, these enzymes are localized in the glycocalyx of the gut cells; in larvae of the fly *Rhynchosciara americana* they are, like the aminopeptidases, integral membrane proteins. The soluble enzymes are usually monomers and the bound forms dimers or oligomers [267]. The α-glucosidases present in the midgut and haemolymph of the cockroach *Periplaneta americana* differ considerably in pH optimum, affinity for the *p*-nitrophenyl substrate, and substrate specificity; in contrast to the haemolymph enzyme, that of the gut also cleaves trehalose.

The α-glucosidase in the midgut of *Erinnyis ello* caterpillars (Lepidoptera: Sphingidae) is found as a tetramer of 151 kDa and as monomers, whereas the enzyme from the imago has a mass of 86 kDa and is immunologically unrelated to the larval enzyme; thus *E. ello* possesses at least two α-glucosidase genes [235]. In contrast to most other α-glucosidases, the haemolymph enzyme of the honey bee shows the typical kinetics of an allosteric enzyme with cooperativity. Interestingly, the Hill constant increases from 0.65 to 1.21 during pupa development. Cooperativity has also been demonstrated for other insect α-glucosidases.

Trehalases have been found in many organisms from the bacteria, yeasts and slime moulds to the mammals. Of the animal trehalases, those from mammalian kidney and gut [225, 295], from the crustaceans *Artemia* and *Macrobrachium*, and from various insects [22, 202, 259] have been isolated and characterized. The trehalases from different organisms differ in many properties: the pH optima vary between 3.5 and 7.5, and the K_m values vary between 0.4 and 20 mmol/l; Mg^{2+} inhibits the yeast trehalase but not that of the beetle *Harmonia axyridis*; and the mammalian enzyme is inhibited by phloridzin but the cockchafer enzyme is not. However, all these enzymes show agreement in that their only natural sub-

strate is trehalose; the claim that purified trehalase from the brine shrimp *Artemia salina* also cleaves cellobiose and lactose should probably be re-examined. The oligosaccharides and glucosides cleaved by the non-specific α-glucosidases are not attacked by trehalases. In experiments with artificial trehalose analogues, the enzymes of bacteria, insects and mammals all behaved similarly in that substrates with intact α-glucopyranoside residues, such as 6′-deoxy-α,α-trehalose or α-galactosyl-α-glucoside, were cleaved but α,β- or β,β-trehalose were not. Despite these results, the mechanism of action of the insect and mammalian enzymes appear to differ: the trehalase from the blowfly *Sarcophaga barbata* produces equal proportions of α- and β-glucose, whereby the oxygen from the water is incorporated into β-glucose; the trehalase from pig kidney, on the other hand, produces only α-glucose. The trehalases of rabbit kidney, *Periplaneta* and *Bombyx* and also from the slime mould *Dictyostelium* are glycoproteins [202, 259].

Mammalian gut trehalase, like the sucrase-isomaltase and the maltase-glucoamylase, is one of the disaccharidases that is bound to the microvillous layer but has been little investigated. The enzymes of rat and rabbit gut are dimers with subunits of 66–75 kDa together with a small membrane domain of about 5 kDa [225, 295]. As trehalose is the typical blood sugar of insects, it is rather amazing that the highest trehalase activities are found in the haemolymph. However, this situation is explained by the localization of trehalase completely or partially in blood cells. Furthermore, the haemolymph of the cockroach *Periplaneta americana* contains a soluble trehalase inhibitor of 86 kDa [107]. In addition to the easily extractable, soluble trehalase, insect tissues contain a particulate form. In the flight muscles of the cockroach *Blaberus discoidalis*, the moth *Hyalophora cecropia* and the locust *Locusta migratoria* the enzyme is bound to the microsomes, whereas in the flies *Phormia regina, Calliphora erythrocephala* and *Sarcophaga barbata* it is localized on the inner mitochondrial membrane [22]. The presence of soluble trehalases in flight muscles often results from the presence there of haemolymph; however, in *Phormia regina* and *Locusta migratoria* the soluble enzymes in flight muscle and haemolymph differ electrophoretically [276]. Trehalose serves as an energy source in flight muscles, at least for short flight periods; the regulation of trehalase in the flight muscles is therefore an interesting, but as yet unsolved, question [276].

Soluble and particulate trehalases differ not only in intracellular localization but also in kinetic properties and organ distribution. In *Bombyx mori*, the predominant midgut enzyme is soluble with a pH optimum of 5.5 and a K_m of 1 mmol/l, whereas the flight muscle enzyme is particulate, has a somewhat higher pH optimum and a lower substrate affinity. It may be assumed that the two forms have different, tissue-specific functions; however, both forms are sometimes found in the same tissue, e.g. in the gut of *Bombyx mori* and *Blaberus*, or in the flight muscles of *Apis, Locusta* and *Periplaneta*. In the flight muscle of *Periplaneta americana*, the cytosolic form is an α-glucosidase with a broad specificity, but the particulate form is highly specific for trehalose [130]. In *Bombyx* gut, the particulate form increases at the expense of the soluble form during pupal moulting; intermediate forms which appear during this period are particle bound but have the kinetics of the soluble form; the particulate forms can be solubilized with detergents [259]. Thus, the two forms are apparently products of post-translational modification, and in *Drosophila* the soluble and particulate trehalases are encoded by the same gene. The soluble *Bombyx* trehalase is a dimer with identical 67-kDa subunits, whereas the particulate form is monomeric; thus, conversion of the particulate to the soluble form probably involves dimerization [259].

Specific **β-glucosidases (cellobiases)** mainly serve to digest cellulose, and **β-N-acetylglucosaminidases (chitinases)** are involved in chitin metabolism. They will be discussed in relation to the corresponding polysaccharidases. Multiple lysosomal β-N-acetylglucosaminidases have been detected in many mammalian tissues and also in the carp [274]. The cleavage of cane sugar (invertase activity) may involve either α-glucosidase or β-fructosidase activity. Genuine **β-fructofuranosidases**, which hydrolyse sucrose, raffinose and inulin but not maltose, are found in plants but only rarely in animals, e.g. in the beetle *Dermestes maculatus* and the moths *Heliothis zea* and *Erinnyis ello* [235].

β-Galactosidases are involved, together with other enzymes, in the metabolism of proteoglycans, glycoproteins and glycolipids, and are, therefore, ubiquitous. The corresponding bacterial enzymes are among the most investigated of glycosidases. Animal tissues contain multiple β-galactosidases with acidic pH optima in the lysosomes and cytoplasm. Apart from those of mammalian organs, only the enzymes of *Drosophila*

and the hepatopancreas of various molluscs have been examined in any detail [84]. The β-galactosidase 1 of *Drosophila melanogaster* is a homodimeric glycoprotein of 160 kDa; null mutants are known for the β-gal locus [84]. The human and murine lysosomal β-galactosidases form a high molecular mass complex of 600–700 kDa together with neuraminidase (sialidase) and a "protective protein" [88]. They have attracted particular attention due to their involvement in glycosphingolipid metabolism and the demonstration that genetic defects of this enzyme lead to GM1 gangliosidose, a classical lysosomal storage disease. The microvillous layer in the small intestine of the higher mammals contains a β-galactosidase (lactase) with a pH optimum of 6 and whose natural substrate is lactose [174]. The microvillous lactase is absent from various kangaroo species of the genus *Macropus* and from the monotreme *Tachyglossus aculeatus* (echidna), in which lactose plays no important role in the milk. In contrast to that of the higher mammals, the lactose of these species is not hydrolysed either during or before resorption, but is cleaved intracellularly by a lysosomal, acidic β-galactosidase [58, 254].

α-Mannosidases and **α-galactosidases** are widely found in plants and animals and are probably involved mainly in the metabolism of the organism's own substrates. The α-galactosidase and α-mannosidase of the hepatopancreas of the edible mussel *Mytilis edulis*, however, may also serve in the digestion of plankton [33, 212]. Similarly, the ubiquitous lysosomal **α-L-fucosidases** are required for the metabolism of the organism's own fucose-containing glycoconjugates [80, 201]. The liver enzymes of various vertebrates are very similar both immunologically and in their catalytic properties, but show large differences in their heat stability; the latter differences appear to have no adaptive significance as, for example, the activities of the human, rhesus monkey and armadillo enzymes are completely destroyed after 3 hours at 60 °C but those of the zebra and the fish *Racophilus toxotes* are reduced by only 14 %. The activity in human blood, which is high relative to that in various other mammals, is not even stable at 37 °C. Considerable differences in heat stability are also found between the different α-fucosidase isozymes of the same organ, e.g. for the enzymes isolated from the hepatopancreas of the snail *Turbo cornutus* [55]. The widely found enzymes referred to as **β-fucosidases** are β-glucosidases with a broader specificity and can also hydrolyse β-glucosides and β-galactosides. It could be shown with the enzymes from the snail *Achatina balteata* and the fly *Rhynchosiciara americana* that all three types of substrate bind to the same active centre [217].

The digestive organs of many marine invertebrates contain **β-glucuronidases**, together with arylsulphatases, which are probably involved in the digestion of sulphated polysaccharides from algae. The richest sources of these enzymes are again found amongst the gastropods [272]. The very first glycosidase to be crystallized (in 1957) was the β-glucuronidase of the edible snail. In mammals, the intracellular localization of the β-glucuronidases is organ specific: in the rat liver, 79 % is found in the lysosomes and 13 % in the microsomes; in the small intestine 62 % is in the soluble fraction, 30 % in the lysosomes and only 4 % in the microsomes. The gut enzymes are responsible for the so-called enterohepatic cycle: β-glucuronides produced in the liver are hydrolysed in the gut so that the aglycone can be reabsorbed and returned to the liver [63]. Two glucuronidase forms are also found in *Drosophila melanogaster* and in the hepatopancreas of the marine snail *Littorina littorea;* these, like the mouse-liver enzymes and other glycosidases, are glycoproteins [167]. The glycosidic bonds of nucleosides in human tissues and those of other mammals are cleaved phosphorolytically rather than hydrolytically. However, almost nothing is known about **nucleoside phosphorylases and hydrolases (nucleosidases)** from a comparative biochemical standpoint. Phosphorylases specific for purine and pyrimidine ribonucleosides are known from the flagellate *Trypanosoma cruzi*. However, both here and in other Kinetoplastida there are also several deoxyribonucleosidases and ribonucleosidases of different specificity [189].

13.6.2 α-Amylases

The **α-amylases** are amongst the best-investigated carbohydrases. They are endoamylases that cleave α(1–4) glucosidic bonds in the chain and break down α(1–4), α(1–6)-glucans, such as amylopectin or glycogen, stepwise, at first into smaller fragments and finally into maltose, isomaltose and glucose. Starch hydrolysis is completed by maltases and isomaltases, e.g. the maltase-glucoamylase and sucrase-isomaltase in the mammalian small intestine. The α-amylase reaction begins with an attack on any one bond within the chain; one of the two resulting fragments remains bound to the enzyme and is shor-

tened at the end by one to three glucose residues at a time during several cleavage reactions. The chain is then released and a new reaction initiated. α-Amylase also catalyses the transfer of maltose and glucose residues to appropriate receptors. Exoamylases, such as the **β-amylases** of higher plants and the lysosomal **exo-α(1–4)-glucosidases**, attack the chain at the non-reducing end and each time remove one molecule of β-maltose or glucose. β-Amylases have been detected in many bacteria, fungi and higher plants. In view of their localization in lysosomes and their acid pH optima, the starch-hydrolysing enzymes involved in intracellular digestion in the Protozoa, Porifera and Enidaria are probably exo-α(1–4)-glucosidases; in contrast, all animals with extracellular digestion processes probably possess genuine α-amylases. These enzymes are usually secreted into the gut lumen by associated glands. A remarkable specialization is the gelatinous crystalline style of the mussels and some gastropods; amongst other digestive enzymes, this contains mainly α-amylase and is gradually solubilized while it is introduced into the stomach. The induction of α-amylase by a starch-rich diet has been demonstrated in the mammals and also in **Drosophila** [110].

In several vertebrates, α-amylases with rather different properties are produced in the pancreas and in the salivary glands. The pancreas type is probably present in all vertebrates, whereas the salivary gland amylase is found only in the anurans, the monotreme *Tachyglossus aculeatus*, rodents and primates. α-Amylase is found in the blood and urine of many mammals; just how the enzyme comes to be present in these fluids and what function, if any, it has remains a mystery. The production of α-amylases in other tissues, particularly in mammalian liver, was for a long time controversial. In the meantime, α-amylase mRNA was identified with certainty in mouse liver [240].

α-Amylases have been isolated and characterized not only from various mammals but also from a series of invertebrates, especially insects, spiders, crabs and molluscs [11, 269]. The molecular masses of the animal α-amylases are all about 52 kDa but may be somewhat higher as the result of glycosylation. The α-amylases are metalloenzymes which are stabilized and activated by calcium ions. Many are glycoproteins; the carbohydrate content of the human salivary gland enzyme is higher than that of the pancreatic enzymes of humans and other mammals; the chicken pancreatic enzyme is carbohydrate free. The α-amylases of mammals and molluscs have neutral pH optima, whereas the optima in insects are more acidic (pH 5–6) and in the spiders are more alkaline (pH 7.2–7.8). Systematic differences in pH optima are found in the crustaceans: pH 6.3 and lower in the Isopoda, Amphipoda and Astacura, and pH 6.8 or higher in the Natantia and Brachyura. All α-amylases are activated by Cl^- ions; the effectiveness of the halogen ions decreases in the order $Cl^- > Br^- > I^- > F^-$. The optimal Cl^- concentration varies greatly with the species and there is often no clear relationship with physiological ion concentrations: the optimal NaCl concentration is about 0.1 mmol/l for the marine crustaceans *Homarus americanus* and *Carcinus maenas* and for the crayfish *Orconectes virilis;* 0.5 mmol/l for the marine mussel *Mya arenaria;* and 1.0 mmol/l for the freshwater isopod *Asellus aquaticus*. The beetle *Callosobruchus sinensis* is the only known exception where chloride is inhibitory [11, 269]. The catalytic properties of the amylases of various *Drosophila* species vary widely with the habitat [188]. The α-amylase of the porcine pancreas is a polypeptide of 496 amino acids and has five internal disulphide bridges. The sequence shows differences of 14 % to the rat and mouse enzymes and 17 % to the human enzyme. The difference between sequences in the rat and mouse is 6 % and between those in the rodents and humans is 17.3 %. The complete sequence is also known for the 475 amino acids of *Drosophila melanogaster* and for 228 amino acids of the C-terminus of the enzyme from the beetle *Tribolium castaneum*. The difference between these two insect species is about 40 %, which is only slightly less than the 46 % difference between the insects and humans [111, 196, 206].

Isoenzymes of the α-amylases have been found in almost all investigated vertebrates and invertebrates. This is partly the result of post-translational modifications (glycosylation, deamidation) and partly determined by post-secretory changes (deglycosylation, deamidation). Genetic analysis has already shown that many animals possess several α-amylase loci. In the mammals, there are separate loci for the pancreas and salivary gland enzymes, and *Drosophila* and *Musca* also have at least two loci. A more detailed picture of the number of **α-amylase genes** has been obtained by mRNA and DNA analysis. In the mouse, the α-amylase genes lie within a region of only 78 kb on chromosome 3; only one gene is present for the salivary gland enzyme, which is also expressed in the liver, but at least four genes exist for the pancreas amylase, although only two of these are apparently tran-

scribed. The gene for the salivary gland/liver enzyme contains two tissue-specific leader sequences that are separated from the coding sequence by an intron of almost 10 kb; the genes for the pancreas enzyme lack this long intron. In contrast, the other introns of the transcribed genes are all found at homologous positions. The rat possesses nine α-amylase genes, each with at least seven introns [289]. *Drosophila melanogaster* and other species of the *melanogaster* subgroup have two amylase genes. As the result of frequent gene conversion, the coding sequences for the two genes are identical, but the flanking DNA sequences vary markedly [112]. Considerable **polymorphism** of α-amylase has been observed in almost all species examined, e.g. in *Drosophila, Bombyx*, and the salivary gland and pancreatic enzymes of man and mammals. In the domestic pig there are one to three isoenzymes of the pancreatic amylase (PPA) with identical catalytic properties but differing frequencies. In one population of 50 animals, 10 had only PPA-2, and 39 had PPA-1 and PPA-2; only 1 animal had all three isoenzymes [258].

13.6.3 Digestion of Cellulose, Lichenin and Laminarin

Cellulose is the most abundant organic substance in the biosphere (even more abundant than chitin) and represents an enormous food reservoir. Many animals can exploit cellulose directly in metabolism, although this occurs mostly with the cooperation of protozoans, bacteria or fungi. The vertebrates have no cellulolytic enzyme systems, and cellulose can only be used indirectly for nutrition; this occurs via anaerobic degradation by symbiontic bacteria in particular regions of the alimentary tract. The products of bacterial metabolism, e.g. volatile fatty acids, as well as vitamins and essential amino acids are then taken up by the animal. **Cellulose fermentation** and metabolism of the volatile fatty acids formed is widely found in the mammals as well as, for example, in the tortoise *Chelonia mydas* and some herbivorous birds. In most mammals, the cellulolytic bacteria inhabit the large intestine and the appendix. However, the bacterial products are exploited much more efficiently where the cellulose fermentation occurs in a spacious rumen, as is found in the ruminants, hippopotamuses, rock hyraxes, sloths, certain ape species and kangaroos [16, 26]. The ciliates present in the stomach of the ruminants are not important for cellulose digestion. The genera *Isotricha* and *Dasytricha* from the Holotricha have no capacity whatsoever for cellulose hydrolysis; several genera of the Entodiniomorpha, e.g. *Eudiplodinium* and *Eremoplastron*, can hydrolyse cellulose but others, e.g. *Entodinium, Metadinium* and *Ophryoscolex*, cannot [56]. Bacterial cellulose fermentation and utilization of the resulting volatile fatty acids has also been demonstrated in beetles of the family Scarabaeidae [19].

There are several different indications for **cellulose utilization** in animals: comparison of the cellulose content of the diet and the faeces; the incorporation of carbon from ^{14}C-labelled cellulose into carbon dioxide and tissue components; the capacity to survive on an exclusively cellulose diet; and the in vitro demonstration of the hydrolysis of native (microcrystalline) cellulose [176]. Based upon these parameters, cellulose utilization has been shown in earthworms, pulmonate snails, insects and mammals. Amongst the insects, cellulose digestion is found especially in the wood-eating species, whereas it is often absent from the leaf-eaters [266]. It has been shown in thysanurans *(Thermobia, Ctenolepisma)*, cockroaches, termites, beetles (Anobiidae, Cerambycidae, Coccinellidae, Scarabeidae), hymenopterans (Siricidae), and the floor mites of the family Oribatidae. The efficiency of cellulose utilization is especially high in the thysanurans and the termites [176, 299]. The cellulose fibres of wood are embedded in a matrix of the phenolic polymer lignin, which is covalently bound to hemicellulose. Unlike the fungi, animals have no enzymes for the hydrolysis of lignin and can remove it only by mechanical means. However, not all wood-eating insects make use of the cellulose; for example, the scoletids and buprestids use the low molecular weight sugars, starch, pectins and xylans of wood [47].

The complete **digestion of native cellulose** requires at least three types of enzyme (the cellulase complex):

1. Endo-β(1–4)-glucanase or Cx enzyme (EC 3.2.1.4) hydrolyses internal β(1–4)-glucosidic bonds in carboxymethylcellulose or other soluble celluloses. The activity against corresponding oligosaccharides (cellodextrins) declines with decreasing chain length. It is inactive against crystalline cellulose.
2. Cellobiohydrolase or C1 enzyme (EC 3.2.1.91) cleaves cellobiose from the non-reducing ends of linear β(1–4)-glucan chains.
3. Cellobiase (EC 3.2.1.21) cleaves cellobiose and cellodextrins. The activity declines markedly with an increase in the number of sugar residues.

The digestion of native cellulose is initiated by the hydrolysis of bonds in less strictly organized regions of the cellulose chain; this is achieved by endoglucanase, which creates sites for cellobiohydrolase activity. This cleaves off cellobiose at these sites, thereby destroying the strict organization of the crystalline regions and revealing new sites for attack by the endoglucanase. By the combined activities of the Cx and C1 enzymes, the cellulose fibres are finally completely degraded to cellobiose and the corresponding oligosaccharides; these are then hydrolysed by the joint action of Cx and cellobiase to give free glucose residues [176].

The classical method for the **determination of cellulase** activity using preparations of soluble cellulose measures only the activity of Cx; the C1 enzyme can be assayed only indirectly by the release of reducing sugars from crystalline cellulose [176]. Endoglucanase (Cx) and cellobiase activities have been detected in numerous species, especially amongst the molluscs and the arthropods, including species whose capacity for cellulose utilization is not particularly great or is completely absent. In the latter cases, it is important to know whether C1 activity is present and, thus, whether the hydrolysis of native cellulose is possible. Appropriate investigations have shown, quite surprisingly, that the eight carnivorous species of molluscs (five cephalopods and three prosobranchia) have a complete cellulolytic system which is probably never active in the living animal [1, 176]. Where the complete cellulase complex is found, it needs to be ascertained whether the individual enzyme components are all produced by the animal itself or whether they originate exogenously. Many animals digest cellulose either with the help of C1-producing protozoans or fungi, or through the involvement of bacteria which can hydrolyse cellulose without the secretion of cellulolytic enzymes.

It is now quite certain that the thysanurans *Thermobia domestica* and *Ctenolepisma lineata* and the floor mite *Steganacarus magnus* (Oribatidae) do not require bacteria or other symbionts for cellulose utilization [176, 299]. However, the situation is more complicated for the termites and some other insects. In the posterior gut of the lower termites (Mastotermitidae, Kalotermitidae, Hodotermitidae and Rhinotermitidae) and the wood cockroach *Cryptocerus punctulatus* there are anaerobic protozoans which ferment cellulose to acetate and are indispensable for cellulose digestion. These protozoans belong to certain genera of the flagellate group of the Polymastigina and are found only in this habitat. The total cellulase activity and more than 90 % of the cellobiase activity in the termite posterior gut are generated by the anaerobic protozoans present. Feeding these insects with starch results in the disappearance of the protozoans and loss of the capacity for cellulose utilization by the host; reinoculation with the protozoans restores the cellulose utilization capacity. The bacteria present in the termite gut are not involved in cellulose digestion. Termites also produce their own cellulolytic enzymes; the activities of endoglucanase and cellobiase in the salivary glands and in the anterior gut and midgut are not affected by removal of the protozoans [176, 278]. For the fungus-farming termites, such as *Macrotermes natalensis*, and the wood wasps (Siricidae), the greater part of the cellulolytic enzymes found in the gut originate from the fungal diet. The amphipod crustaceans of the genus *Gammarus* also contain cellulolytic enzymes derived from fungi [230].

The degree to which symbiontic gut bacteria are important for the utilization of cellulose by invertebrates is in most cases not known. A critical role may be assumed when the extent of cellulose degradation correlates with the extent of the bacterial flora, e.g. in the beetle *Oryctes*, the diplopod *Glomeris* and the isopod *Tracheoniscus* [176]. The isolation of animal cellulases has seldom been attempted. The endo-β(1–4)-glucanase of the edible snail *Helix pomatia* has all the typical properties of this enzyme type: the molecular mass is 51 kDa; it cleaves soluble β(1–4)-glucans down to cellotetraose with a pH optimum of 6; and it is inactivated by EDTA and can be reactivated by calcium ions. Three cellulolytic enzymes have been isolated from the long-horned beetle *Ergates faber*. Enzyme A has a mass of 25 kDa and is a typical cellobiohydrolase (C1 enzyme); enzymes B and C have masses of 57 and 70 kDa, respectively, and are exoglucanases of broad specificity which cleave off terminal glucose residues from longer or shorter β(1–4)-glucans. Many β-glucosidases of the invertebrates are active against cellobiose; one such enzyme from a beetle larva has been purified and characterized [47].

Most endo-β(1–4)-glucanases are active against the lichenin found in Irish moss, which is essentially composed of β(1–3)-linked cellotriose sequences. The hydrolysis of β(1–3)-glucans found, for example, in the cell wall of fungi, in the brown algae (laminarin) and in the phloem of higher plants, requires specific **β(1–3)-glucanases (laminarinases)**. Such enzymes have been found

in many invertebrates. They are widely distributed amongst the molluscs and crustaceans which feed on brown algae or phytoplankton. They are present in almost all inhabitants of the forest floor and leaf litter, such as oligochaetes, isopods, chilopods, diplopods and insects. Amongst the laminarinases, a distinction is drawn between the endo-β(1–3)-glucanases, which cleave internal bonds, and the exo-β(1–3)-glucanases, which change the anomeric configuration and release α-glucose residues from the non-reducing ends of β-(1–3)-glucans. The marine copepod *Acartia clausi* possesses both types of enzyme with the exo-activity predominating; this is bound to a protein of 21 kDa and has a pH optimum of 5.6 [226]. Animals with high laminarinase activity often have only the endohydrolase but, in contrast to the corresponding enzymes of the fungi and plants, the animal enzymes are still active against tri- and disaccharides and therefore can compensate for the lack of exo-activity [51]. Exo-β(1–3)-glucosidases have been detected, for example, in *Helix pomatia*, other pulmonate snails and sea urchin eggs [263]. Several of these enzymes have been isolated and characterized. The endohydrolases usually have lower molecular masses (20-22 kDa) than do the exohydrolases, which are usually more than 100 kDa. The exceptions to this rule are the endo-β(1–3)-glucanase isolated from larvae of the beetle *Rhagium inquisitor*, which is about 100 kDa, and the small exohydrolase from the copepod *Acartia*, mentioned above [51, 226, 263]. The crystalline style of the mussel *Chlamys abbidus* contains both a β(1–3)- and a β(1–6)-specific endoglucanase [233].

Further polysaccharide-hydrolysing enzyme activities are found in the tissues, body fluids and poisonous secretions of animals, but there are few comparative biochemical data available. On the one hand, there are enzymes that attack the cell wall substances present in herbivore diets, and on the other hand, there are the enzymes that can hydrolyse animal proteoglycans. The cell walls of higher plants consist of cellulose fibres embedded in a matrix of lignin, hemicelluloses and pectins. Neither lignin nor the mixture of branched heteroglycans known as hemicellulose are effectively digested by animal enzymes. The pectins are made up of α(1–4)-linked galacturonic acid residues which are, in part, methylated. Pectinases are widely found in fungi and higher plants and have also been detected in molluscs, coleopterans and dipterans [102]. Alginases from sea urchins and marine mussels cleave the alginic acids found, for example, in the brown algae; these consist of β(1–4)-linked mannuronic acid and single residues of its 5-epimer L-guluronic acid. The hyaluronidases and chondroitinases of the vertebrates are involved in the metabolism of the extracellular matrix; the hyaluronidases present at high activity in the venoms of snakes, bees, spiders and scorpions serve in the dissemination of the toxin in the tissues. The saliva of the leech *Hirudo medicinalis* contains an endo-β-glucuronidase which is highly specific for hyaluronic acid and which, in contrast to the endo-β-hexosaminidases from mammalian connective tissues, does not attack chondroitin. This property makes the enzyme particularly suitable for cell separation and it has been marketed under the name "Orgelase".

13.6.4 Chitinolysis

At least two types of enzyme are required for the complete **hydrolysis of chitin:** a chitinase, which reduces chitin to oligosaccharides, and a β-N-acetylglucosaminidase or a -N-acetylhexosaminidase, which releases N-acetylglucosamine. Chitinases with a digestive function are found in the entoderm of the Cnidaria and in the gut and associated glands of many invertebrates. Contrary to early assumptions, these enzymes are probably not of bacterial origin but are produced by the animals themselves [100, 151, 172, 192, 248]. The vertebrates show an evolutionary trend towards loss of chitinases; they are still widely found in the fish, amphibians and reptiles, but are absent from most species of birds and mammals that exist on a herbivorous or otherwise chitin-free diet. Thus, chitinase is present in the gut of rats, mice, dogs and foxes, but not in the gut of tortoises, pigeons, rabbits, sheep, cats or humans [192]. β-N-Acetylglucosaminidases and N-N-acetylhexosaminidases, which are also required for the metabolism of proteoglycans and glycoproteins, are probably ubiquitous. It is somewhat extraordinary to find a high activity of a genuine chitinase in calf serum. Nematode larvae have a chitinase of ectodermal origin and this is probably involved in hatching [192]. The similar ectodermal chitinases that appear at certain times in the integument of arthropods certainly have a function in moulting. In insects, spiders and crustaceans, during the period between separation of the cuticula from the epidermis (apolysis) and moulting itself (ecdysis), up to 90% of the material of the old cuticula is released, reabsorbed and reused [50, 247].

The best-investigated **chitinolytic system** is that of the moth *Manduca sexta*. This consists of six enzymes (E-I to E-VI), all of which have been isolated and characterized in detail. E-IV, E-V and E-VI from the moulting fluid and the integument are chitinases which can also degrade chitin oligosaccharides to di- and trisaccharides at a pH optimum of 6; their activity increases with chain length. They have different molecular masses (75, 62 and 50 kDa, respectively) and amino acid compositions, but are immunologically related. The glycosidase E-I in the haemolymph of larvae and pupae as well as in the moulting fluid and integument cleaves di- to hexasaccharides to about the same extent as it cleaves *p*-nitrophenyl-β-N-acetylglucosamine; E-III from pupal haemolymph has similar properties. In contrast, E-II from larval and pupal haemolymph is active almost exclusively against glycosides. E-I is without doubt a moulting enzyme; the functions of E-II and E-III probably lie more in the metabolism of proteoglycans, glycoproteins and glycolipids. Whereas E-I and E-III are immunologically similar, E-II again occupies a special position. All three glycosidases have the same molecular mass of 61 kDa, although E-I consists of one polypeptide chain and E-II and E-III have two chains [154]. The isolated chitinases (E-IV to E-VI) are monomers; nevertheless, the chitinase activity of the moulting fluid shows sigmoid kinetics (Hill constant n = 1.95), which suggests a homotropic allosteric interaction. The reconstituted chitinolytic system produces more free N-acetylglucosamine than predicted from the activities of the isolated enzymes; the largest turnover is achieved when chitinases and β-N-acetylglucosaminidases are mixed in a ratio of 6 : 1. Thus, the chitinases apparently form non-productive enzyme-substrate complexes with low molecular weight substrates and these can be degraded by the β-N-acetylglucosaminidases [50, 85]. The chitinolytic systems involved in the moulting of other insects, e.g. *Platysamia, Bombyx, Drosophila, Musca* and *Tribolium*, have also been investigated [50, 158, 159]. A 215-kDa protein identified immunologically in *Bombyx mori* is apparently a chitinase zymogen from which the active 65-kDa enzyme is released the day before moulting. The silkworm also appears to contain inactive precursors of the β-N-acetylglucosaminidase [155].

The majority of insect **chitinases** have molecular masses between 45 and 75 kDa; similar values have been reported for other chitinases, e.g. from calf serum (47 kDa), from the digestive juices of the spider *Cupiennius salei* (48 kDa), and from the hepatopancreas and the crystalline style of the edible mussel (60 and 76 kDa, respectively) [28, 50]. However, the enzyme from the brine shrimp *Artemia* is reported to be a glycoprotein of only 32 kDa [86]. The catalytic activities of the chitinases vary greatly with the species. The three chitinases from *Manduca sexta* produce oligomers of N-acetylglucosamine; in contrast, the chitinases of the silkmoth *Bombyx mori*, the fly *Stomoxys calcitrans* and the spider *Cupiennius salei* primarily release dimers. The spider chitinases degrade chitin to chitobiose (88 %). chitotriose (5 %) and free N-acetylglucosamine (7 %), whereas, for example, the moulting enzyme of *Bombyx mori* produces no trisaccharides. Almost all chitinases produce free N-acetylglucosamine, probably by the cleavage of chitotriose. Even the chitinases from *Manduca* have at least a low hydrolytic activity with chitotriose. The vertebrate chitinases produce no free N-acetylglucosamine [50]. One problem with the investigations of chitinases is the difficulty of avoiding contamination by bacterial and fungal enzymes. A chitinase which is certainly not contaminated is secreted into the medium of *Drosophila* cell cultures [32].

The **N-acetylglucosaminidases** of the insects present a very heterogeneous picture. The enzymes of *Manduca sexta* have a mass of about 61 kDa, whilst enzymes from other sources are over 100 kDa. The enzyme of *Bombyx mori* is a dimer with subunits of equal size, whilst that of *Stomoxys calcitrans* is a heterodimer with subunits of 67 and 43 kDa [50]. The N-acetylglucosaminidases or N-acetylhexos aminidases of animals have in some cases digestive functions but are mainly involved in the metabolism of proteoglycans, glycoproteins and glycolipids. Apart from chitobiose, they cleave other β-glycosidic derivatives of N-acetylglucosamine and N-acetylgalactosamine, whereby their relative activity against the individual substrate varies greatly with the origin of the enzyme. The molecular masses of these enzymes span a wide range: 55 kDa in *Tetrahymena*, 125 kDa in *Entamoeba*, 100–200 kDa in various mussel species, 112–124 kDa in the nematode *Turbatrix* and 150–160 kDa for the enzyme from rat kidney [21, 234, 268]. Even in closely related species there are large differences in temperature stability and these can not be explained as adaptations; for example, the enzyme activity of the mussel *Spisula solidissima* is not affected at all after 1 h at 45 °C, but the activity is reduced in *Mya*

arenaria by about 47 %, and in *Mercenaria mercenaria* by as much as 91 % [234]. Various mammalian organs contain two β-N-acetylhexosaminidases, A and B, which have many very similar properties but vary in their activities against different gangliosides. These enzymes are tetramers, made up of two different subunits according to the formulae A = $\alpha_2\beta_2$ and B = β_4. The human genes for α and β are located on different chromosomes. It has been shown for the liver enzymes of various primates that the two genes separated a long time ago, and their evolution is particularly slow. The immunological distance, determined by microcomplement fixation, between A and B of the same species is much greater than the species-specific differences within A and B. Isoenzymes of β-N-acetyl-hexosaminidase have also been recorded in protozoans and molluscs [203].

13.6.5 Lysozymes

Based on the observation that bacteria were lysed by chicken albumen, Fleming in 1922 postulated the presence of an enzyme which he called lysozyme. It is now known that lysozymes cleave β(1–4) bonds between N-acetylmuraminic acids and N-acetylglucosamine in the cell wall material of Gram-positive bacteria. The susceptibility of *Micrococcus lysodeikticus* makes it particularly suitable for lysozyme detection. According to their origin, lysozymes may have a higher or lower chitinase activity; the ratio of chitinase to muraminidase activity for the enzymes from the chicken egg or the edible mussel *Mytilus edulis* are about 1 : 100, whereas for the lysozyme from the eggs of the Mediterranean fruit fly *Ceratitis capitata* the ratio is much higher. Relevant here is the fact that lysozymes are inhibited by free N-acetylglucosamine and even more so by chitin oligosaccharides. The primary function of the albumen lysozymes is apparently not defence against microorganisms; many bacteria on the surface of the egg are not attacked and the lysis products are highly toxic to embryos. Lysozymes appear in fact to be responsible for the gel structure of the albumen in freshly laid eggs. This is brought about by lysozyme and ovomucin, which together produce a cross-linked structure; the positively charged lysine residues of the lysozyme interact electrostatically with the negatively charged terminal sialic acid residues on the carbohydrate chains of ovomucin. The temperature increase during brooding causes a partial degradation of these carbohydrate chains and an increase in the fluidity of the egg white.

Lysozymes have been found in bacteriophages, bacteria, lower fungi, higher plants, and both invertebrates and vertebrates. Many of them have been isolated and characterized and, consequently, they are amongst the most investigated enzymes of comparative biochemistry. In the mammals, they are found in tissues, blood and urine, and also occur in secretions such as milk and tears. In the ruminants and certain species of ape, lysozymes function as digestive enzymes. The lysozyme content of avian albumen varies from 3–4 % of the total protein in chickens and geese to only trace amounts in penguins [81]. Reptilian albumen also contains lysozymes. No fewer than eight isozymes of lysozyme from various organs have been detected in the frog *Rana pipiens*. Detailed investigations have been carried out on the lysozymes of the plaice *Pleuronectes platessa*, the polychaete *Nephthys hombergi*, the starfish *Asterias rubens*, the edible mussel *Mytilus edulis* and many insects [64, 116, 179, 242, 296]. Haemolymph lysozymes can be induced in insects by inoculation with bacteria; however, other haemolymph proteins, such as cecropins and attacins, have a much higher bactericidal activity than the lysozymes.

Two different types of lysozyme found in **avian albumen** are designated, according to the first examined species, as c type (chicken) and g type (goose). The two **lysozyme types** are so different in structural and functional properties that they must be considered as two completely different proteins. From the soluble peptidoglycan of *Micrococcus lysodeikticus*, the chicken egg-white lysozyme (HEWL) produces: (1) the disaccharide GlcNAcβ(1–4)MurNAc; (2) a tetrasaccharide made up of two such sequences; and (3) a disaccharide-hexapeptide; the goose egg-white lysozyme (GEWL) produces only (1) and (3), and the lysozyme from phage T4 produces only (3) [241]. Compared with the g type, the c type has greater activity against chitin oligosaccharides and is more strongly inhibited by N-acetylglucosamine. The pH optimum of the c type is weakly alkaline, and that of the g type is strongly acidic [10]. HEWL has a molecular mass of 14.5 kDa (129 amino acids) and GEWL 20 kDa (185 amino acids). Apart from certain similarities in the active centre, there is no significant sequence agreement between the lysozymes of chicken and goose; however, the spatial structures derived from the sequences are very similar.

In the gallinaceous birds (order Galliformes) and the pigeon, the albumen contains only type-c lysozymes, and in eight further avian orders only type g occurs (Table 13.5). In the goose orders (Anseriformes), some species have only type g and others only type c; however, the black swan *Cygnus atratus* and many other species have both types g and c [10, 69]. The albumens of many bird orders do not react with antisera raised against the lysozymes of chicken, goose or duck, suggesting the existence in these cases of greatly altered forms of types g and c or of other, as yet undefined, lysozymes [215]. Whereas only one of the two lysozyme types is usually expressed in the cells of the oviduct, the polymorphically nucleated leukocytes of the chicken and goose contain both types [113].

The complete **amino acid sequences** are known for at least 17 albumen lysozymes, 14 of type c and 3 of type g. With the exception of the lysozyme from the pigeon, the various lysozymes show differences of 3–21 % to the HEWL of the chicken, and the g lysozymes show a difference of 3 % to the GEWL of the black swan and of 16 % to that of the ostrich [92, 246]. The albumen lysozyme of the pigeon is clearly of type c but shows about 50 % difference in its polypeptide chain of 127 amino acids to the other known c types; including this highly deviant example, there appear to be only 24 invariant positions in the c-type lysozymes [228]. A 586-bp pre-lysozyme mRNA has been isolated from chicken oviduct. The corresponding gene has about 4000 bp and contains three large introns. The signal sequence of the HEWL and other pre-lysozymes of the c type is made up of 18 amino acids (see Fig. 3.4, p. 85). The pre-lysozyme of the ringed-neck pheasant *Phasianus colchicus* includes proline at position 17 and is cleaved between positions 17 and 18; thus, the signal peptide is reduced to 17 amino acids and the lysozyme itself is extended to 130 amino acids by an N-terminal glycine. The signal peptidase in this case can apparently only cleave bonds within a particular spatial arrangement (β turn), and proline is not acceptable as the penultimate amino acid of the signal peptide [287].

The lysozymes of the **mammals** are homologous to avian-egg c type. The human enzyme agrees in 61 % of its 130 amino acids with the chicken HEWL, and the mouse lysozyme agrees in 57 % of positions [52, 60]. The α-lactalbumins are also significantly homologous to HEWL (p. 470); in fact, the horse lysozyme includes the complete Ca^{2+}-binding site of α-lactalbumin, and

Table 13.5. The distribution of lysozymes of types c and g in the albumen of birds from various orders and species [10, 91, 215]

Order-species	c-type	g-type
Galliformes (fowl)	+	–
Columbiformes (pigeon)	+	–
Anseriformes (geese)		
in 3 of 25 species e.g. *(Mergus serrator)*	+	–
in 16 of 25 species e.g. *(Cygnus atratus, Anser indicus)*	+	+
in 6 of 25 species e.g. *(Anser anser, Anser canadensis)*	–	+
Struthioniformes (ostrich)	–	+
Rheiformes (American rhea)	–	+
Casuariiformes (cassowaries)	–	+
Apertygiformes (kiwis)	–	+
Tinamiformes (tinamous)	–	+
Podicipediformes (grebes)	–	+
Sphenisciformes (penguins)	–	+
Charadriiformes (gulls)	–	+

the lysozymes of the monotreme *Echidna* and the pigeon bind calcium [94, 197]. Lysozymes have assumed digestive roles in two groups of mammals; in the ruminants and in the East African leaf-eating apes of the family Colobidae. Both have multichambered stomachs and rumen fermentation of cellulose. Lysozymes are secreted into the last stomach chamber, the real stomach (abomasum), where they digest the bacteria coming from the rumen. Sequence comparisons indicate that the stomach lysozymes of the two groups developed independently at different times from the antibacterial enzymes present in all mammals. Initially, further lysozyme genes arose by gene duplication in the apes and the ruminants; by alterations in gene regulation some of these were expressed in the stomach and assumed digestive functions, whereas others retained their original functions. After this change in function, the rate of evolution of the lysozyme genes in both groups of mammals increased transiently. There are three bovine lysozymes which are specific to the stomach and are encoded by four of the ten lysozyme genes. Sheep and goats, but not cattle, have lysozymes in their tears. The multiple lysozymes of tears and milk apparently arose by gene duplication from stomach lysozymes [126, 136]. One of the two lysozymes from the albumen of the turtle *Trionyx gangeticus*, the only reptile system to have been examined in any detail, is clearly homologous to

HEWL, although it has 16 substitutions and 1 deletion in the 48 N-terminal amino acids. Distant relationships with HEWL are also discernible in the lysozymes of the frog *Rana pipiens*, the trout *Onkorhynchus mykiss* and the plaice *Pleuronectes platessa* [64]. Although it is generally agreed that **α-lactalbumin** originated from a lysozyme, there is some disagreement as to when this occurred. According to one model of evolution, the gene duplication between lysozyme and α-lactalbumin occurred before, or shortly after, the separation of the mammals and the birds. This model avoids the assumption of accelerated evolution of the α-lactalbumin line, and explains why α-lactalbumin and the mammalian lysozymes resemble the c-type lysozymes of the birds and reptiles to a similar degree. According to an alternative model, the α-lactalbumin gene arose at the beginning of mammalian evolution, at the same time as the appearance of the milk gland. This model assumes a higher rate of evolution for α-lactalbumins than for lysozymes [197, 216].

Because of a shortage of sequence data, little can be said definitively either about the **evolutionary relationships** between the lysozymes of vertebrates and invertebrates, or about the total number of lysozyme types to be found in the animal kingdom. The lysozyme of the moth *Hyalophora cecropia* agrees in about 40 % of its 120 amino acids with HEWL. The lysozymes of the moths *Bombyx mori, Galleria melonella* and *Spodoptera littoralis* are significantly homologous in the 33 N-terminal amino acids to HEWL and human lysozyme, and have about the same size of 15–16 kDa [260]. Results obtained with the enzyme-linked immunosorbent assay (ELISA) method, which allows the detection of distant immunological relationships, suggest that the lysozymes of the migratory locust *Locusta migratoria* and various avian c types are quite similar, although the sequence similarities show only about 60 % agreement [116]. In contrast, the lysozyme of the Mediterranean fruit fly *Ceratitis capitata* with a mass of 23 kDa (210 amino acids) is much larger than the other known lysozymes, and with its chitinase activity shows different enzymatic properties. The lysozyme of the starfish *Asterias rubens* differs consistently from all other lysozymes in its length of 142 amino acids, N-terminal sequence and functional properties. On the other hand, the lysozyme of the polychaete *Nephthys hombergi* is similar to type-c vertebrate enzymes, at least in its enzymatic activity; however, in this case, there are as yet no sequence data.

References

1. Agnisola C., Salvadore S. and Scardi V.: On the occurrence of cellulolytic activity in the digestive gland of some marine carnivorous molluscs. Comp. Biochem. Physiol. Pt. B 70: 521–525 (1981)
2. Allen A.: Mucus – A protective secretion of complexity. Trends biochem. Sci. 8: 169–173 (1983)
3. Altman P. L. and Dittmer D. S. (eds.): Biological data book, 2. ed., Vol. III, pp. 1936–42. Fed. Amer. Soc. Exp. Biol., Bethesda 1974
4. Andersen S. O.: Sclerotization and tanning of the cuticle. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 3, pp. 59–75. Pergamon Press, Oxford 1985
5. Antonsson P., Heinegård D. and Oldberg Å.: The keratan sulfate-enriched region of bovine cartilage proteoglycan consists of a consecutively repeated hexapeptide motif. J. Biol. Chem. 264: 16170–73 (1989)
6. Aon M. A. and Curtino J. A.: Protein-bound glycogen is linked to tyrosine residues. Biochem. J. 229: 269–272 (1985)
7. Arslan A., Standifer L. N. and Don H.: Carbohydrates in honey bee hemolymph. Comp. Biochem. Physiol. Pt B 84: 363–367 (1986)
8. Atkins E. D. T., Dlugosz J. and Foord S.: Electron diffraction and electron microscopy of crystalline α-chitin from the grasping spines of the marine worm Sagitta. Int. J. biol. Macromol. 1: 29 (1979)
9. Bai G. et al.: The primary structure of rat liver glycogen synthase deduced by cDNA cloning. Absence of phosphorylation sites 1a and 1b. J. Biol. Chem. 265: 7843–48 (1990)
10. Bailey C. J., Geoffroy P. and Mills K. H. G.: An examination of the distribution of hen- and goose-type lysozymes in Anseriformes. Comp. Biochem. Physiol. Pt. B 55: 429–433 (1976)
11. Baker J. E.: Properties of amylases from midguts of larvae of Sitophilus zeamais and Sitophilus granarius. Insect Biochem. 13: 421–428 (1983)
12. Baker M. E.: A common ancestor for human placental 17ß-hydroxysteroid dehydrogenase, Streptomyces coelicolor act III protein, and Drosophila melanogaster alcohol dehydrogenase. Faseb J. 4: 222–226 (1990)
13. Bakken H.: Cold hardiness in the alpine beetles Patrobus septentrionis and Cathus melanocephalus. J. Insect Physiol. 31: 447–453 (1985)
14. Ballas R. A., Garavelli J. S. and White III H. B.: Estimation of the rate of glycerol-3-phosphate dehydrogenase evolution in higher vertebrates. Evolution 38: 658–664 (1984)
15. Bangs J. D. et al.: Biosynthesis of a variant surface glycoprotein of Trypanosoma brucei. Processing of the glycolipid anchor and N-linked oligosaccharides. J. Biol. Chem. 263: 17697–705 (1988)
16. Barboza P. S. and Hume I. D.: Hindgut fermentation in the wombats: two marsupial grazers. J. comp. Physiol. B 162: 561–566 (1992)
17. Barretto O. C. O. et al.: Erythrocyte sorbitol dehydrogenase of selected nonmammalian species. Comp. Biochem. Physiol. Pt. B 82: 317–319 (1985)
18. Batterham P. et al.: Origin and expression of an alcohol dehydrogenase gene duplication in the genus Drosophila. Evolution 38: 644–657 (1984)

19. Bayon C. and Mathelin J.: Carbohydrate fermentation and by-product absorption studied with labelled cellulose in Oryctes nasicornis larvae (Coleoptera: Scarabaeidae). J. Insect Physiol. 26: 833–840 (1980)
20. Bedford J. J.: The carbohydrate levels of insect hemolymph. Comp. Biochem. Physiol. Pt. A 57: 83–86 (1977)
21. Bedi G. S., Shah R. H. and Bahl O. P.: Studies on Turbatrix aceti β-N-acetylglucosaminidase. 1. Purification and physicochemical characterization. Arch. Biochem. Biophys. 233: 237–250 (1984)
22. Beenakkers A. T., van der Horst D. J. and van Marrewijk W. J. A.: Biochemical processes directed to flight muscle metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 451–486. Pergamon Press, Oxford 1985
23. Bereiter-Hahn J., Matolsky A. G. and Richards K. S. (eds.): Biology of the integument, 2 vol. set. Springer, Berlin 1986
24. Berg W. J. and Buth D. G.: Glucose dehydrogenase in teleosts: tissue distribution and proposed function. Comp. Biochem. Physiol. Pt. B 77: 285–288 (1984)
25. Berger E. G. et al.: Structure, biosynthesis and functions of glycoprotein glycans. Experientia 38: 1129–1162 (1982)
26. Bergman E. N.: Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. Phsiol. Rev. 70: 567–589 (1990)
27. Bewley G. C. et al.: Sequence, structure and evolution of the gene coding for sn-glycerol-3-phosphate dehydrogenase in Drosophila melanogaster. Nucleic Acids Res. 17: 8553–67 (1989)
28. Birkbeck T. H. and McHenry J. G.: Chitinase in the mussel, Mytilus edulis (L.). Comp. Biochem. Physiol. Pt. B 77: 861–865 (1984)
29. Birney E. C., Jenness R. and Hume I. D.: Evolution of an enzyme system – Ascorbic acid biosynthesis in monotremes and marsupials. Evolution 34: 230–239 (1980)
30. Blackwell J. and Weih M. A.: Structure of chitin-protein complexes: Ovipositor of the ichneumon fly Megarhyssa. J. mol. Biol. 137: 49–60 (1980)
31. Blithe D. L., Clark H. F. and Warren L.: Similarity in the bound carbohydrate groups of glycoproteins from the cells of several vertebrate classes. Biochim. biophys. Acta 719: 190–198 (1982)
32. Boden N., Sommer U. and Spindler K. D.: Demonstration and characterization of chitinases in the Drosophila Kc cell line. Insect Biochem. 15: 19–23 (1985)
33. Bonay P. and Hughes R. C.: Purification and characterization of a novel broad specificity (α1–2, α1–3 and α1–6) mannosidase from rat liver. Eur. J. Biochem. 197: 229–238 (1991)
34. Bosch M. et al.: Characterization of dolichol diphosphate oligosaccharide, protein oligosaccharide transferase and glycoprotein-processing glucosidases occuring in trypanosomatid protozoa. J. Biol. Chem. 263: 17360–65 (1988)
35. Bounias M.: A comparison of haemolymph levels and inter-relations of trehalose, glucose and fructose in worker bees from different races and hybrids. Comp. Biochem. Physiol. Pt. B 69: 471–477 (1981)
36. Bounias M. and Morgan M. R. J.: Induction of honeybee haemolymph sucrase activity by higher levels of dietary sucrose. Comp. Biochem. Physiol. Pt. B 79: 75–80 (1984)
37. Breuer B. et al.: A novel large dermatan sulfate proteoglykan from human fibroblasts. J. Biol. Chem. 266: 13224–232 (1991)
38. Browner M. F. et al.: Human muscle glycogen synthase cDNA sequence: A negatively charged protein with an asymmetric charge distribution. Proc. Nat. Acad. Sci. USA 86: 1443–47 (1989)
39. Bunde T. A., Dearlove G. E. and Bishop S. H.: Aminoethylphosphonic acid-containing glycoproteins: The acid mucopolysaccharide-like components in mucus from Metridium senile (L.). J. exp. Zool. 206: 215–221 (1978)
40. Byrne D. N. and Miller W. B.: Carbohydrate and amino acid composition of phloem sap and honeydew produced by Bemisia tabaci. J. Insect Physiol. 36: 433–439 (1990)
41. Campbell A. G. et al.: Papilin: a Drosophila proteoglycan-like sulfated glycoprotein from basement membranes. J. biol. Chem. 262: 17605–12 (1987)
42. Campbell D. G. and Cohen P.: The amino acid sequence of rabbit skeletal muscle glycogenin. Eur. J. Biochem. 185: 119–125 (1989)
43. de la Canal L. and Parodi A. J.: Glycosylation of proteins in the protozoan Euglena gracilis. Comp. Biochem. Physiol. Pt. B 81: 803–805 (1985)
44. Carroll M. and McCrorie P.: Glycoproteins of trypanosomes: their biosynthesis and biological significance (Mini-review). Comp. Biochem. Physiol. Pt. B 88: 7–12 (1987)
45. Cavener D. R. and Macintyre R. J.: Biphasic expression and function of glucose dehydrogenase in Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 80: 6286–88 (1983)
46. Cederlund E. et al.: Amphibian alcohol dehydrogenase, the major frog liver enzyme. Relationships to other forms and assessment of an early gene duplication separating vertebrate class-I and class-III alcohol dehydrogenases. Biochemistry 30: 2811–16 (1991)
47. Chararas C. et al.: Purification of three cellulases from the xylophageous larvae of Ergastes faber (Coleoptera: Cerambycidae). Insect Biochem. 13: 213–218 (1983)
48. Chaturvedi P. and Sharma C. B.: Goat milk oligosaccharides: purification and characterization by HPLC and high-field proton-NMR spectroscopy. Biochim. biophys. Acta 967: 115–121 (1988)
49. Chee N. P., Geddes R. and Wills P. R.: Metabolic heterogeneity in rabbit brain glycogen. Biochim. biophys. Acta 756: 9–12 (1983)
50. Chen A. C.: Chitin metabolism. Arch. Insect Biochem. Physiol. 6: 267–277 (1987)
51. Chipoulet J. M. and Chararas C.: Purification and partial characterization of a laminarinase from the larvae of Rhagium inquisitor. Comp. Biochem. Physiol. Pt. B 77: 699–706 (1984)
52. Chung L. P., Keshav S. and Gordon S.: Cloning the human lysozyme cDNA. Proc. Nat. Acad. Sci. USA 85: 6227–31 (1988)
53. Churchill T. A. and Storey K. B.: Metabolic correlates to glycerol biosynthesis in freeze-avoiding insect, Epiblema scudderiana. J. comp. Physiol. B 159: 461–472 (1989)
54. Cohen E.: Chitin synthetase activity and inhibition in different insect microsomal preparations. Experientia 41: 470–472 (1985)
55. Cohenford M. A., Urbanowski J. C. and Dain J. A.: Purification and properties of two forms of α-L-fucosidase from Turbo cornutus. Comp. Biochem. Physiol. Pt. B 72: 695–703 (1982)

56. Coleman G. S.: Rumen ciliate protozoa. In: Lewandowsky M. and Hutner S. H. (eds.): Biochemistry and physiology of protozoa, 2. ed., Vol 2, pp. 381–408. Acad. Press, New York 1979
57. Collins P. M. and Munasinghe R. (eds.): Carbohydrates. Chapman & Hall, London 1987
58. Crisp E. A., Czolij R. and Messer M.: Absence of β-galactosidase (lactase) activity from intestinal brush borders of suckling macropods: Implications for mechanism of lactose absorption. Comp. Biochem. Physiol. Pt. B 88: 923–927 (1987)
59. Crisp E. A., Messer M. and Vandeberg J. L.: Changes in milk carbohydrates during lactation in a didelphid marsupial, Monodelphis domestica. Physiol. Zool. 62: 1117–25 (1989)
60. Cross M. et al.: Mouse lysozyme M gene: Isolation, characterization, and expression studies. Proc. Nat. Acad. Sci. USA 85: 6232–36 (1988)
61. Crowley H. M., Woodward D. R. and Rose R. W.: Changes in milk composition during lactation in the potoroo, Potorus tridactylus (marsupialia: Potoroinae). Aust. J. biol. Sci. 41: 289–296 (1988)
62. Dabrowski K.: Gulonolactone oxidase is missing in teleost fish, the direct photometric assay. Biol. Chem. Hoppe-Seyler 371: 207–214 (1990)
63. D'Amore M. A. et al.: Complete sequence and organization of the murine β-glucuronidase gene. Biochemistry 27: 7131–40 (1988)
64. Dautigny A. et al.: cDNA and amino acid sequences of rainbow trout (Onkorhynchus mykiss) lysozymes and their implications for the evolution of lysozyme and lactalbumin. J. mol. Evol. 32: 187–198 (1991)
65. Dekan G., Gabel C. and Farquhar M. G.: Sulfate contributes to the negative charge of podocalyxin, the major sialoglycoprotein of the glomerular filtration slits. Proc. Nat. Acad. Sci. USA 88: 5398–5402 (1991)
66. Dietrich C. P. et al.: Characteristic distribution of heparan sulfates and chondroitin sulfates in tissues and organs of the Ampullaridae Pomacea sp. Comp. Biochem. Physiol. Pt. B 76: 695–698 (1983)
67. Doege K. J. et al.: Complete primary structure of the rat cartilage proteoglycan core protein deduced from cDNA clones. J. biol. Chem. 262: 17757–67 (1987), Correction 263: 10040 (1988)
68. Doege K. J. et al.: Complete coding sequence and deduced primary structure of the human cartilage large aggregating proteoglycan aggrecan: Human-specific repeats, and additional alternatively spliced forms. J. Biol. Chem. 266: 894–902 (1991)
69. D'Surney S. J. and DeKloet S. R.: Biochemical characterization of lysozymes present in egg white of selected species of anatid birds. Comp. Biochem. Physiol. Pt. B 82: 555–558 (1985)
70. Dughia J. and Hardingham T. E.: The primary structure of human cartilage link protein. Nucleic Acids Res. 18: 1292 (1990)
71. Duman J. G. et al.: Freeze-tolerance adaptations, including haemolymph protein and lipoprotein nucleators, in the larvae of the cranefly Tipula trivittata. J. Insect Physiol. 31: 1–8 (1985)
72. Dunn D. F. and Liberman M. H.: Chitin in sea anemone shells. Science 221: 157–159 (1983)
73. Dykhuizen D. E., Harrison K. M. and Richardson B. J.: Evolutionary implications of ascorbic acid production in the Australian lungfish. Experientia 36: 945–946 (1980)
74. Edenberg H. J. et al.: Cloning and sequencing of cDNA encoding the complete mouse liver alcohol dehydrogenase. Proc. Nat. Acad. Sci. USA 82: 2262–66 (1985)
75. Eklund H. et al.: Molecular aspects of functional differences between alcohol and sorbitol dehydrogenases. Biochemistry 24: 8005–12 (1985)
76. Eklund H. et al.: Comparison of three classes of human liver alcohol dehydrogenase: Emphasis on different substrate binding pockets. Eur. J. Biochem. 193: 303–310 (1990)
77. Endo T. et al.: Structures of the sugar chains of a major glycoprotein present in the egg jelly coat of a starfish, Asterias amurensis. Arch. Biochem. Biophys. 252: 105–112 (1987)
78. England B. P., Heberlein U. and Tjian R.: Purified Drosophila transcription factor, Adh distal factor-1 (Adf-1), binds to sites in several Drosophila promoters and activates transcription. J. Biol. Chem. 265: 5086–94 (1990)
79. Fievre S. et al.: Primary structure of a trisialylated oligosaccharide from human milk. Biochem. biophys. Res. Commun. 177: 720–725 (1991)
80. Fisher K. J. and Aronson N. N. jr.: Isolation and sequence analysis of a cDNA encoding rat liver α-L-fucosidase. Biochem. J. 264: 695–701 (1989)
81. Florkin M. and Stotz E. H.: Comprehensive biochemistry, Vol. 6. Elsevier, Amsterdam 1965
82. Fong W. P. and Keung W. M.: Substrate specificity of human class I alcohol dehydrogenase homo- and heterodimers containing the β2 (oriental) subunit. Biochemistry 26: 5726–32 (1987)
83. Friedman S.: Carbohydrate metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 43–76. Pergamon Press, Oxford 1985
84. Fuerst T. R., Knipple D. C. and Macintyre R. J.: Purification and characterization of β-galactosidase-1 from Drosophila melanogaster. Insect Biochem. 17: 1163–71 (1987)
85. Fukamizo T. and Kramer K. J.: Mechanism of chitin hydrolysis by the binary chitinase system in insect moulting fluid. Insect Biochem. 15: 141–145 (1985)
86. Funke B. and Spindler K. D.: Characterization of chitinase from the brine shrimp Artemia. Comp. Biochem. Physiol. Pt. B 94: 691–695 (1989)
87. Galand G.: First purification and characterization of a sucrase-isomaltase from goose kidney microvillous membrane. Biochim. biophys. Acta 1033: 35–40 (1990)
88. Galjart N. J. et al.: Mouse "protective protein": cDNA cloning, sequence comparison, and expression. J. Biol. Chem. 265: 4678–84 (1990)
89. Gallagher J. T. and Walker A.: Molecular distinctions between heparan sulphate and heparin. Analysis of sulphation patterns indicates that heparan sulphate and heparin are separate families of N-sulphated polysaccharides. Biochem. J. 230: 665–674 (1985)
90. Garcin F. et al.: Acetaldehyde oxidation in Drosophila melanogaster and Drosophila simulans: evidence for the presence of an NAD-dependent dehydrogenase. Comp. Biochem. Physiol. Pt. B 75: 205–210 (1983)
91. Gavilanes J. G. et al.: Pigeon egg white lysozyme. Purification, structural and enzymic characterization. Int. J. Peptide Protein Res. 20: 238–245 (1982)

92. Gavilanes J. G., Menendez-Arias L. and Rodriguez R.: Comparative study on the secondary structure of lysozymes from different sources. Comp. Biochem. Physiol. Pt. B 77: 83–88 (1984)
93. George J. N., Nurden A. T. and Phillips D. R. (eds.): Platelet membrane glycoproteins. Plenum, New York 1985
94. Godovav-Zimmermann J., Conti A. and Napolitano I.: The primary structure of donkey (Equus asinus) lysozyme contains the Ca(II)-binding site of α-lactalbumin. Biol. Chem. Hoppe-Seyler 369: 1109–15 (1988)
95. Godovav-Zimmermann J., Conti A. and Napolitano I.: The complete primary structure of α-lactalbumin isolated from pig (Sus scrofa) milk. Biol. Chem. Hoppe-Seyler 371: 649–653 (1990)
96. Goldsmith P. K. and Stetten M. R.: Hexokinases and sugar dehydrogenases of Limulus polyphemus. Comp. Biochem. Physiol. Pt. B 77: 41–50 (1984)
97. Gomes P. B. and Dietrich C. P.: Distribution of heparin and other sulfated glycosamino-glycans in vertebrates. Comp. Biochem. Physiol. Pt. B 73: 857–863 (1982)
98. Goudsmit E. M. et al.: Biosynthesis of galactogen: Identification of a β-(1–6)-D-galactosyltransferase in Helix pomatia albumen glands. Biochim. biophys. Acta 992: 289–297 (1989)
99. Graham A. et al.: Structure of the human aldose reductase gene. J. Biol. Chem. 266: 6872–77 (1991)
100. Grisley M. S. and Boyle P. R.: Chitinase, a new enzyme in octopus saliva. Comp. Biochem. Physiol. Pt. B 95: 311–316 (1990)
101. Groenberg G. et al.: Isolation and structural analysis of three new disialylated oligosaccharides from human milk. Arch. Biochem. Biophys. 278: 297–311 (1990)
102. Grover P. B. jr., Ross D. R. and Shukle R. H.: Identification and partial characterization of digestive carbohydrases in larvae of the hessian fly, Mayeticola destuctor (Say) (Diptera: Cediomyidae). Arch. Insect Biochem. Physiol. 8: 59–72 (1988)
103. Guan K. and Weiner H.: Sequence of the precursor of bovine liver mitochondrial aldehyde dehydrogenase as determined from its cDNA, its gene, and its functionality. Arch. Biochem. Biophys. 277: 351–360 (1990)
104. Gum J. R. et al.: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evidence for genetic polymorphism. J. Biol. Chem. 264: 6480–87 (1989)
105. Gupta R. and Jentoft N.: Subunit structure of porcine submaxillary mucin. Biochemistry 28: 6114–21 (1989)
106. Hannigan L. L., Donahue M. J. and Masaracchia R. A.: Comparative purification and characterization of invertebrate muscle glycogen synthase from the porcine parasite Ascaris suum. J. biol. Chem. 260: 16099–105 (1985)
107. Hayakawa Y. et al.: Purification and characterization of trehalase from the American cockroach, Periplaneta americana. J. Biol. Chem. 264: 16165–69 (1989)
108. Heinstra P. W. H. et al.: The metabolism of ethanol-derived acetaldehyde by alcohol dehydrogenase (EC 1.1.1.1) and aldehyde dehydrogenase (EC 1.2.1.3) in Drosophila melanogaster larvae. Biochem. J. 259: 791–797 (1989)
109. Heminga M. A., Whittle M. A. and Gabbott P. A.: Glycogen synthase in the mantle tissue of the freshwater snail Lymnaea stagnalis: interconversion of kinetic forms and evidence for activity regulation by a concerted push-pull mechanism. Comp. Biochem. Physiol. Pt. B 82: 535–543 (1985)
110. Hickey D. A. and Benkel B.: Regulation of amylase activity in Drosophila melanogaster: effects of dietary carbohydrate. Biochem. Genetics 20: 1117–29 (1982)
111. Hickey D. A. et al.: Enzyme-coding genes as molecular clocks: The molecular evolution of animal alpha-amylases. J. mol. Evol. 26: 252–256 (1987)
112. Hickey D. A. et al.: Concerted evolution of duplicated protein-coding genes in Drosophila. Proc. Nat. Acad. Sci. USA 88: 1611–15 (1991)
113. Hindenburg A., Spitzragel J. and Aynheim N.: Isozymes of lysozyme in leukocytes and egg white: Evidence for the species-specific control of egg-white lysozyme synthesis. Proc. Nat. Acad. Sci. USA 71: 1653–57 (1974)
114. Hoefsloot L. H.: Primary structure and processing of lysosomal α-glucosidase; homology with the intestinal sucrase-isomaltase complex. Embo J. 7: 1697–1704 (1988)
115. Hoeoek M. et al.: Cell-surface glycosaminoglycans. Annual Rev. Biochem. 53: 847–869 (1984)
116. Hoffmann D. and van Regenmortel M. H. V.: Detection of distant antigenic relationships between insect and bird lysozymes by ELISA. J. mol. Evol. 21: 14–18 (1984)
117. Holst O. et al.: Structural studies on the galactan from the albumin gland of Achatina fulica. Z. Naturforsch. Sect. C 39: 1063–65 (1984)
118. Horner A. A.: Rat heparins. A study of the relative sizes and antithrombin-binding characteristics of heparin proteoglycans, chains and depolymerization products from rat adipose tissue, heart, lungs, peritoneal cavity and skin. Biochem. J. 240: 171–179 (1986)
119. Horst M. N.: Isolation of a crustacean N-acetyl-D-glucosamine-1-phosphate transferase and its activation by phospholipids. J. comp. Physiol. B 159: 777–788 (1990)
120. Hovingh P. and Peter A.: An unusual heparan sulfate isolated from lobsters (Homarus americanus). J. biol. Chem. 257: 9840–44 (1982)
121. Hsieh P. and Robbins P. W.: Regulation of asparagine-linked oligosaccharide processing. Oligosaccharide processing in Aedes albopictus mosquito cells. J. biol. Chem. 259: 2375–82 (1984)
122. Hsu L. C. and Chang W. C.: Cloning and characterization of a new functional human aldehyde dehydrogenase gene. J. Biol. Chem. 266: 12257–65 (1991)
123. Hunt S.: Polysaccharide-protein complexes in invertebrates. Acad. Press, New York 1970
124. Inoue H. et al.: Difference between N-acetylgalactosamine 4-sulfate-6-O-sulfotransferases from human serum and squid cartilage in specificity toward the terminal and interior portion of chondroitin sulfate. J. biol. Chem. 261: 4470–75 (1986)
125. Ireland R. C. et al.: Primary structure of the mouse glycerol-3-phosphate dehydrogenase gene. J. biol. Chem. 261: 11779–85 (1986)
126. Irwin D. M. and Wilson A. C.: Concerted evolution of ruminant stomach lysozymes. Characterization of lysozyme cDNA clones from sheep and deer. J. Biol. Chem. 265: 4944–52 (1990)
127. Ito M. and Yamagata T.: The linkage of teleost skin keratan sulfate to protein. Biochim. biophys. Acta 801: 381–387 (1984)

128. Iwasaki M., Inoue S. and Inoue Y.: Identification and determination of absolute and anomeric configurations of the 6-deoxyaltrose residue found in polysialoglycoprotein of Salvelinus leucomaenis pluvius eggs – The first demonstration of a 6-deoxyhexose other than fucose in glycoprotein. Eur. J. Biochem. 168: 185–192 (1987)
129. Iwasaki M., Inoue S. and Troy F. A.: A new sialic acid analogue, 9-O-acetyl-deaminated neuraminic acid, and α-2,8-linked O-acetylated poly(N-glycolylneuraminyl) chains in a novel polysialoglycoprotein from salmon eggs. J. Biol. Chem. 265: 2596–2602 (1990)
130. Jahagirdar A. P., Downer R. G. H. and Viswanatha T.: Purification and characterization of trehalose-hydrolyzing enzymes from thoracic musculature of the American cockroach, Periplaneta americana. Insect Biochem. 20: 511–516 (1990)
131. James T. C. et al.: Xenopus egg jelly coat proteins – 2. Characterization of messenger RNAs of the oviduct and cloning of complementary DNAs to poly(A)-containing RNA. Comp. Biochem. Physiol. Pt. B 80: 89–97 (1985)
132. Jeffery J. et al.: Molecular diversity of glucose-6-phosphate dehydrogenase. Rat enzyme structure identifies NH2-terminal segment, shows initiation from sites nonequivalent in different organisms, and established otherwise extensive sequence conservation. Proc. Nat. Acad. Sci. USA 85: 7840–43 (1988)
133. Jentoff N.: Why are proteins O-glycosylated? Trends biochem. Sci. 15: 291–294 (1990)
134. Jeronimo S. M. B., Dietrich C. P. and Nader H. B.: Structure of sulfated glycoaminoglycans synthesied during the ontogeny of the mollusc Pomacea sp. Comp. Biochem. Physiol. Pt. B 93: 899–903 (1989)
135. Jörnvall H. et al.: Mammalian alcohol dehydrogenase of separate classes: Intermediates between different enzymes and intraclass enzymes. Proc. Nat. Acad. Sci. USA 84: 2580–84 (1987)
136. Jollès et al.: Amino acid sequences of stomach and nonstomach lysozymes of ruminants. J. mol. Evol. 30: 370–382 (1990)
137. Joziasse D. H. et al.: Identification of a UDP:ß-galactoside β1–3-galactosyltransferase in the albumen gland of the snail Lymnaea stagnalis. FEBS Letters 221: 139–144 (1987)
138. Juan E., Papaceit M. and Quintana A.: Nucleotide sequence of the Adh gene of Drosophila lebanonensis. Nucleic Acids Res. 18: 6420 (1990)
139. Julià P, Parés X. and Joernvall H.: Rat liver alcohol dehydrogenase of class-III: Primary structure, functional consequences and relationships to other alcohol dehydrogenases. Eur. J. Biochem. 172: 73–83 (1988)
140. Kaiser R. et al.: Avian alcohol dehydrogenase. Characterization of the quail enzyme, functional interpretations, and relationships to the different classes of mammalian alcohol dehydrogenases. Biochemistry 29: 8365–71 (1990)
141. von Kalm L. et al.: Structural characterization of the α-glycerol-3-phosphate dehydrogenase-encoding gene of Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 86: 5020–24 (1989)
142. Kanamori A. et al.: Deaminated neuraminic acid-rich glycoprotein of rainbow trout egg vitelline envelope. Occurrence of a novel alpha-2,8-linked oligo(deaminated neuraminic acid) structure in O-linked glycan chains. J. Biol. Chem. 265: 21811–19 (1990)
143. Kapoun A. M. et al.: Molecular control of the induction of alcohol dehydrogenase by ethanol in Drosophila melanogaster larvae. Genetics 124: 881–888 (1990)
144. Karamanos N. K. et al.: Chondroitin proteoglycans from squid skin. Isolation, characterization and immunological studies. Eur. J. Biochem. 192: 33–38 (1990)
145. Kariya Y. et al.: Occurrence of chondroitin sulfate E in glycosaminoglycan isolated from the body wall of sea cucumber Stichopus japonicus. J. Biol. Chem. 265: 5081–85 (1990)
146. Karlsson C., Joernvall H. and Hoog J. O.: Sorbitol dehydrogenase-cDNA coding for the rat enzyme. Variations within the alcohol dehydrogenase family independent of quaternary structure and metal content. Eur. J. Biochem. 198: 761–765 (1991)
147. Keung W. M. and Yip P. K.: Rabbit liver alcohol dehydrogenase. Isolation and characterization of class I isozymes. Biochem. biophys. Res. Commun. 158: 445–453 (1989)
148. Kidder G. M.: Glucose-6-phosphate dehydrogenase isozymes in fish – A comparative study. J. exp. Zool. 226: 385–390 (1983)
149. Kirkman B. R. and Whelan W. J.: Glucosamine is a normal component of liver glycogen. FEBS Letters 194: 6–11 (1986)
150. Kitagawa H. et al.: Novel oligosaccharides with the sialyl-Lea structure in human milk. Biochemistry 30: 2869–76 (1991)
151. Klug M. et al.: Presence and localization of chitinase in Hydra and Podocoryne. J. exp. Zool. 229: 69–72 (1984)
152. Knels U. and Bretting H.: Comparative structural analysis of snail galactans by a radioimmunoassay to elucidate species-specific determinants. J. comp. Physiol. B 159: 629–639 (1989)
153. Knels U. and Bretting H.: Structural studies on the galactan from the snail Helix pomatia. Comp. Biochem. Physiol. Pt. B 96: 147–155 (1990)
154. Koga D. et al.: Immunological relationships between β-N-acetylglucosaminidases from the tobacco hornworm, Manduca sexta L. Insect Biochem. 13: 407–410 (1983)
155. Koga D. et al.: Appearance of chitinolytic enzymes in integument of Bombyx mori during larval-pupal transformation. Evidence for zymogenic forms. Insect Biochem. 19: 123–128 (1989)
156. Kornfeld R. and Kornfeld S.: Assembly of asparagine-linked oligosaccharides. Annual Rev. Biochem. 54: 631–664 (1985)
157. Koshizaka T. et al.: Isolation and sequence analysis of a complementary DNA encoding rat liver L-gulono-γ-lactone oxidase, a key enzyme for L-ascorbic acid biosynthesis. J. biol. Chem. 263: 1619–21 (1988)
158. Kramer K. J. and Koga D.: Insect chitin. Physical state, synthesis, degradation and metabolic regulation (Minireview). Insect Biochem. 16: 851–877 (1986)
159. Kramer K. J. and Aoki H.: Chitinolytic enzymes from the pupae of the red flour beetle, Tribolium castaneum. Comp. Biochem. Physiol. Pt. B 86: 613–621 (1987)
160. Krasney P. A., Carr C. and Cavener D. R.: Evolution of the glucose dehydrogenase gene in Drosophila. Mol. Biol. Evol. 7: 155–177 (1990)

161. Krishnamoorthy R. V. and Begum B. J.: Muscle specific structural differences in piscine muscle glycogens. Experientia 34: 832–833 (1978)
162. Krishnan G. and Ramamurthi R.: On the polysaccharide-protein complexes of the tunic of Polyclinum madrasensis Sebastian. Indian J. exp. Biol. 14: 113–116 (1976)
163. Kristensen J. G.: Carbohydrases of some marine invertebrates with notes on their food and on the natural occurrence of the carbohydrates studied. Mar. Biol. 14: 130–142 (1972)
164. Kukal O., Serianni A. S. and Duman J. G.: Glycerol metabolism in a freeze-tolerant arctic insect: an in vivo 13C-NMR study. J. comp. Physiol. B 158: 175–183 (1988)
165. Laird J. E. et al.: Structure and expression of the guinea-pig α-lactalbumin gene. Biochem. J. 254: 85–94 (1988)
166. Landau B. R. and Wood H. G.: The pentose cycle in animal tissues: evidence for the classical and against the „L-type" pathway. Trends biochem. Sci. 8: 292–296 (1983)
167. Langley S. D. et al.: A genetic variant of β-glucuronidase in Drosophila melanogaster. J. biol. Chem. 258: 7416–24 (1983)
168. Laurie C. C. et al.: Genetic basis of the difference in alcohol dehydrogenase expression between Drosophila melanogaster and Drosophila simulans. Proc. Nat. Acad. Sci. USA 87: 9674–78 (1990)
169. Lennarz W. J. (ed.): Biochemistry of glycoproteins and proteoglycans. Plenum, New York 1980
170. Livingstone D. R. and de Zwaan A.: Carbohydrate metabolism of gastropods. In: Wilbur K. M. (ed.): The Mollusca, Vol. 1, pp. 177–272. Acad. Press, New York 1983
171. Lo H. S. and Chang C. J.: Purification and properties of NADP-linked alcohol dehydrogenase from Entamoeba histolytica. J. Parasitol. 68: 372–377 (1982)
172. Lynn K. R.: Chitinases and chitobiases from the American lobster (Homarus americanus). Comp. Biochem. Physiol. Pt. B 96: 761–766 (1990)
173. Malik Z., Jones C. J. P. and Connock M. J.: Assay and subcellular localization of hydrogen peroxide-generating mannitol oxidase in the terrestrial slug Arion ater. J. exp. Zool. 242: 9–15 (1987)
174. Mantei N. et al.: Complete primary structure of human and rabbit lactase-phlorizin hydrolase: implications for biosynthesis, membrane anchoring and evolution of the enzyme. Embo J. 7: 2701–13 (1988)
175. Martin L. O. G. et al.: Purification and properties of 6-phosphogluconate dehydrogenase from Mytilus galloprovincialis digestive gland. Comp. Biochem. Physiol. Pt. B 79: 599–606 (1984)
176. Martin M. M.: Cellulose digestion in insects. Comp. Biochem. Physiol. Pt. A 75: 313–324 (1983)
177. Matsuo Y., Yokoyama R. and Yokoyama S.: The genes for human alcohol dehydrogenase β1 and β2 differ by only one nucleotide. Eur. J. Biochem. 183: 317–320 (1989)
178. Matsuoka N. et al.: Evidence for the homology of hexose 6-phosphate dehydrogenase and glucose 6-phosphate dehydrogenase: Comparison of the amino acid compositions. Comp. Biochem. Physiol. Pt. B 76: 811–816 (1983)
179. McHenery J. G., Allen J. A. and Birkbeck T. H.: Distribution of lysozyme-like activity in 30 bivalve species. Comp. Biochem. Physiol. Pt. B 85: 581–584 (1986)
180. Medina-Puerta M. and Garrido-Perriera A.: Purification and properties of the enzyme 6-phosphogluconate dehydrogenase from Dicentrarchus labrax L. liver. Comp. Biochem. Physiol. Pt. B 83: 215–220 (1986)
181. Meikle P., Richards G. N. and Yellowlees D.: Structural determination of the oligosaccharide side chains from a glycoprotein isolated from the mucus of the coral Acropora formosa. J. biol. Chem. 262: 16941–47 (1987)
182. Mendelzon D. H. and Parodi A. J.: N-linked high mannose-type oligosaccharides in the Protozoa Crithidia fasciculata and Crithidia harmosa contain galactofuranose residues. J. biol. Chem. 261: 2129–33 (1986)
183. Merchant J. C. and Libke J. A.: Milk composition in the northern brown bandicoot, Isoodon macrourus (Peramelidae, Marsupialia). Aust. J. biol. Sci. 41: 495–505 (1988)
184. Messer M. et al.: Carbohydrates in the milk of the platypus. Aust. J. biol. Sci. 36: 129–137 (1983)
185. Messer M., Crisp E. A. and Newgrain K.: Studies on the carbohydrate content of milk of the crabeater seal (Lobodon carcinophagus). Comp. Biochem. Physiol. Pt. B 90: 367–370 (1988)
186. Messer M. and Nicholas K. R.: Biosynthesis of marsupial milk oligosaccharides: Characterization and developmental changes of two galactosyltransferases in lactating mammary glands of the tammar wallaby, Macropus eugenii. Biochim. biophys. Acta 1077: 79–85 (1991)
187. Michelacci Y. M. and Horton D. S. P. Q.: Proteoglycans from the cartilage of young hammerhead shark, Sphyrna lewini. Comp. Biochem. Physiol. Pt. B 92: 651–658 (1989)
188. Milanovic M., Andjelkovic M. and Stamenkovic-Bojic G.: Adaptive significance of amylase polymorphism in Drosophila. IV. A comparative study of biochemical properties of the alpha-amylase in Drosophila melanogaster, D. hydei, D. subobscura and D. busckii. Comp. Biochem. Physiol. Pt. B 93: 629–634 (1989)
189. Miller R. et al.: Nucleoside hydrolases from Trypanosoma cruzi. J. biol. Chem. 259: 5073–77 (1984)
190. Mourao P. A. S.: Epimerization of D-glucose to L-galactose during the biosynthesis of a sulfated L-galactan in the ascidian tunic. Biochemistry 30: 3458–64 (1991)
191. Moxon L. N. et al.: Purification and molecular properties of alcohol dehydrogenase from Drosophila melanogaster: Evidence from NMR and kinetic studies for function as an aldehyde dehydrogenase. Comp. Biochem. Physiol. Pt. B 80: 525–535 (1985)
192. Muzzarelli R. A. A.: Chitin. Pergamon Press, Oxford 1977
193. Myers F. L. and Northcote D. H.: A survey of enzymes from the gastrointestinal tract of Helix pomatia. J. exp. Biol. 35: 639–648 (1958)
194. Neame P. J., Choi H. U. and Rosenberg L. C.: The primary structure of the core protein of the small, leucine-rich proteoglycan (PG I) from bovine articular cartilage. J. Biol. Chem. 264: 8653–61 (1989)
195. Nicholas K. et al.: Isolation, partial sequence and asynchronous appearance during lactation of lysozyme and α-lactalbumin in the milk of a marsupial, the

common ringtail possum (Pseudocheirus peregrinus). Comp. Biochem. Physiol. Pt. B 94: 775–778 (1989)
196. Nishide T. et al.: Primary structure of human salivary α-amylase gene. Gene 41: 299–304 (1986)
197. Nitta K. and Sugai S.: The evolution of lysozyme and α-lactalbumin. Eur. J. Biochem. 182: 111–118 (1989)
198. Nöhle U. et al.: Structural parameters and natural occurrence of 2-deoxy-2,3-didehydro-N-glycoloylneuraminic acid. Eur. J. Biochem. 152: 459–463 (1985)
199. Norén O. et al.: Pig intestinal microvillar maltase-glucoamylase. Structure and membrane insertion. J. biol. Chem. 261: 12306–09 (1986)
200. Nyame K., Cummings R. D. and Damian R. T.: Schistosoma mansoni synthesizes glycoproteins containing terminal O-linked N-acetylglucosamine residues. J. biol. Chem. 262: 7990–95 (1987)
201. Occhiodoro T. et al.: Human α-L-fucosidase: Complete coding sequence from cDNA clones. Biochem. biophys. Res. Commun. 164: 439–445 (1989)
202. Ogiso M. et al.: Further purification and characterization of trehalase from the American cockroach, Periplaneta americana. J. comp. Physiol. B 155: 553–560 (1985)
203. Orlacchio A. et al.: Isozymes separation by chromatofocusing and kinetic properties of β-hexosaminidase from Spirographis spallanzanii, Allolobophora caliginosa and Hirudo medicinalis. Comp. Biochem. Physiol. Pt. B 80: 923–926 (1985)
204. Parès X et al.: Class-IV mammalian alcohol dehydrogenase. Structural data of the rat stomach enzyme reveal a new class well separated from those already characterized. FEBS Letters 277: 115–118 (1990)
205. Park D. H. and Plapp B. V.: Isoenzymes of horse liver alcohol dehydrogenase active on ethanol and steroids: cDNA cloning, expression, and comparison of active sites. J. Biol. Chem. 266: 13296–302 (1991)
206. Pasero L. et al.: Complete amino acid sequence and location of the five disufide bridges in porcine pancreatic α-amylase. Biochim. biophys. Acta 869: 147–157 (1986)
207. Paulson J. C. and Colley K. J.: Glycosyltransferases. Structure, localization, and control of cell type-specific glycosylation. J. Biol. Chem. 264: 17615–18 (1989)
208. Pavàò M. S. G. et al.: Structural heterogeneity among unique sulfated L-galactans from different species of ascidians. J. Biol. Chem. 264: 9972–79 (1989)
209. Peddemors V. M., de Muelenaere H. J. H. and Devchand K.: Comparative milk composition of the bottlenosed dolphin (Tursiops truncatus), humpback dolphin (Sousa plumbea) and common dolphin (Delphinus delphis) from southern African waters. Comp. Biochem. Physiol. Pt. A 94: 639–641 (1989)
210. Pejler G. et al.: Structure and antithrombin-binding properties of heparin isolated from the clams Anomalocardia brasiliana and Tivela mactroides. J. biol. Chem. 262: 11413–21 (1987)
211. Perkins S. J. et al.: Immunoglobulin fold and tandem repeat structure in proteoglycan N-terminal domains and link protein. J. mol. Biol. 206: 737–753 (1989)
212. Pernas R. V., Martinez I. R. and Rama M. F.: Purification and properties of Zn-dependent α-mannosidase from lysosomes of Mytilus edulis L. hepatopancreas. Comp. Biochem. Physiol. Pt. B 70: 125–131 (1981)
213. Peters C. W. B. et al.: The human lysozyme gene: Sequence organization and chromosomal localization. Eur. J. Biochem. 182: 507–516 (1989)
214. Poole A. R.: Proteoglycans in health and disease: Structures and functions. Biochem. J. 236: 1–14 (1986)
215. Prager E. M., Wilson A. C. and Arnheim N.: Widespread distribution of lysozyme g in egg white of birds. J. biol. Chem. 249: 7295–97 (1974)
216. Prager E. M. and Wilson A. C.: Ancient origin of lactalbumin from lysozyme: Analysis of DNA and amino acid sequences. J. mol. Evol. 27: 326–335 (1988)
217. Pratviel-Sosa F. et al.: Studies on glycosidases and glucanases in Taumetopoea pityocampa larvae – II. Purification and some properties of a broad specificity β-D-glucosidase. Comp. Biochem. Physiol. Pt. B 86: 173–178 (1987)
218. Prieto P. A. and Smith D. F.: A new sialyloligosaccharide from human milk: isolation and characterization using anti-oligosaccharide antibodies. Arch. Biochem. Biophys. 229: 650–656 (1984)
219. Probst J. C., Gertzen E. M. and Hoffmann W.: An integumentary mucin (FIM-B.1) from Xenopus laevis homologous with von Willebrand factor. Biochemistry 29: 6240–44 (1990)
220. Ramabrahmam P. and Subrahmanyam D.: Mitochondrial glycerol kinase of Culex pipiens. Insect Biochem. 13: 523–528 (1983)
221. Rat L., Veuille M. and Lepesant J. A.: Drosophila fat body protein P6 and alcohol dehydrogenase are derived from a common ancestral protein. J. mol. Evol. 33: 194–203 (1991)
222. Reinecke J. P.: Nutrition: artificial diets. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 4, pp. 394–419. Pergamon Press, Oxford 1985
223. Renwrantz L. et al.: A quantitative determination of the hemolymph carbohydrates of Helix pomatia. J. comp. Physiol. B 105: 185–188 (1976)
224. Ribeiro J. M. C. and Pereira M. E. A.: Midgut glycosidases of Rhodnius prolixus. Insect Biochem. 14: 103–108 (1984)
225. Riby J. and Galand G.: Rat intestinal brush border membrane trehalase: some Properties of the purified enzyme. Comp. Biochem. Physiol. Pt. B 82: 821–827 (1985)
226. Roche-Mayzaud O and Mayzaud P.: Purification of endo- and exolaminarinase and partial characterization of the exoacting form from the copepod Acartia clausi (Giesbrecht, 1889). Comp. Biochem. Physiol. Pt. B 88: 105–110 (1987)
227. Rodger J. C. and White I. G.: Glycogen not N-acetylglucosamine the prostatic carbohydrate of three Australian and American marsupials, and patterns of these sugars in Marsupialia. Comp. Biochem. Physiol. Pt. B 67: 109–113 (1980)
228. Rodriguez R. et al.: Structure of the pigeon lysozyme and its relationship with other type c lysozymes. Comp. Biochem. Physiol. Pt. B 88: 791–796 (1987)
229. Ropson I. J. and Powers D. A.: A novel dehydrogenase reaction mechanism for hexose-6-phosphate dehydrogenase isolated from the teleost Fundulus heteroclitus. J. Biol. Chem. 263: 11697–11703 (1988)
230. Rouland C. et al.: Synergistic activities of the enzymes involved in cellulose degradation, purified from Macrotermes mulleri and from its symbiontic fungus Termitomyces sp. Comp. Biochem. Physiol. Pt. B 91: 459–465 (1988)

231. Rowan R. G. and Dickinson W. J.: Two alternative transcripts coding for alcohol dehydrogenase accumulate with different species of picture-winged Drosophila. Genetics 114: 435–452 (1986)
232. Rowan R. G. and Dickinson W. J.: Nucleotide sequence of the genomic region encoding alcohol dehydrogenase in Drosophila affinidisjuncta. J. mol. Evol. 28: 43–54 (1989)
233. Rudakova V. Y., Shevchenko N. M. and Elyakova L. A.: Isolation and some properties of endo-ß-1,6-glucanase from marine bivalve Chlamys abbidus. Comp. Biochem. Physiol. Pt. B 81: 677–682 (1985)
234. Santoro P. F. and Dain J. A.: A comparative study of β-N-acetylglucosaminidase from Mercenaria mercenaria, Mya arenaria and Spisula solidissima. Comp. Biochem. Physiol. Pt. B 69: 337–344 (1981)
235. Santos C. D. and Terra W. R.: Midgut α-glucosidase and β-fructosidase from Erinnyis ello larvae and imagoes. Physical and kinetic properties. Insect Biochem. 16: 819–824 (1986)
236. Sawhney R. S., Hering T. M. and Sandell L. J.: Biosynthesis of small proteoglycan-II (decorin) by chondrocytes and evidence for a procore protein. J. Biol. Chem. 266: 9231–40 (1991)
237. Schade S. Z. et al.: Sequence analysis of bovine lens aldose reductase. J. Biol. Chem. 265: 3628–35 (1990)
238. Schauer R. (ed.): Sialic acids. Springer, Wien 1982
239. Schauer R., Wember M. and Howard R. J.: Malaria parasites do not contain or synthesize sialic acids. Hoppe Seyler's Z. physiol. Chem. 365: 185–194 (1984)
240. Schibler U. et al.: The mouse α-amylase multigene family. Sequence organization of members expressed in the pancreas, salivary gland and liver. J. mol. Biol. 155: 247–266 (1982)
241. Schindler M., Mirelman D. and Sharon N.: Substrate-induced evolution of the lysozymes. Biochim. biophys. Acta 482: 386–392 (1977)
242. Schneider P. M.: Purification and properties of three lysozymes from hemolymph of the cricket, Gryllus bimaculatus (De Geer). Insect Biochem. 15: 463–470 (1985)
243. Seko A. et al.: Structural studies of fertilization-associated carbohydrate-rich glycoproteins (hyosophorins) isolated from the fertilized and unfertilized eggs of flounder, Paralichthys olivaceus. J. Biol. Chem. 264: 15922–29 (1989)
244. Settine R. L. and Prchal J.: Gas chromatographic/mass spectrometric evidence for the identification of a heptitol and an octitol in human and Octodon degu eye lenses. Biochem. biophys. Res. Commun. 116: 988–993 (1983)
245. Sharma C. P. et al.: cDNA sequence of human class III alcohol dehydrogenase. Biochem. biophys. Res. Commun. 164: 631–637 (1989)
246. Simpson R. J. and Morgan F. J.: Complete amino acid sequence of embden goose (Anser anser) egg-white lysozyme. Biochim. biophys. Acta 744: 349–351 (1983)
247. Sing G. J. P. and Vardanis A.: Chitinases in the house fly, Musca domestica. Pattern of activity in the life cycle and preliminary characterization. Insect Biochem. 14: 215–218 (1984)
248. Smucker R. A. and Wright D. A.: Characteristics of Crassostrea virginica crystalline style chitin digestion. Comp. Biochem. Physiol. Pt. A 83: 489–493 (1986)
249. Smythe C., Villar-Palasi C. and Cohen P.: Structural and functional studies on rabbit liver glycogenin. Eur. J. Biochem. 183: 205–209 (1989)
250. Sorimachi H. et al.: Organization and primary sequence of multiple genes coding for the apopolysialoglycoproteins of rainbow trout. J. mol. Biol. 211: 35–48 (1990)
251. Speck U.: Das Kohlenhydratspektrum in den Organen des Flußkrebses Orconectes limosus. Z. vergl. Physiol. 65: 51–69 (1969)
252. Spiess M. and Lodish H. F.: Sequence of a second human asialoglycoprotein receptor: conservation of two receptor genes during evolution. Proc. Nat. Acad. Sci. USA 82: 6465–69 (1985)
253. Stevenson J. R.: Dynamics of the integument. In: Bliss D. E. (ed.): The biology of Crustacea, Vol. 9, pp. 1–42. Acad. Press., New York 1985
254. Stewart I. M. et al.: Intestinal glycosidase activities in one adult and two suckling echidnas: absence of a neutral lactase (ß-D-galactosidase). Aust. J. biol. Sci. 36: 139–146 (1983)
255. Stoddart R. W.: The biosynthesis of polysaccharides. Croom Helm, London 1984
256. Storey K. B.: Glycolysis and the regulation of crypoprotectant synthesis in liver of the freeze tolerant wood frog. J. comp. Physiol. B 157: 373–380 (1987)
257. Storey K. B. et al.: Glucose-6-phosphate dehydrogenase in cold hardy insects: Kinetic properties, freezing stabilization, and control of hexose monophosphate shunt activity. Insect Biochem. 21: 157–164 (1991)
258. Strumeyer D. H. et al.: Isozymes of α-amylase in porcine pancreas: population distribution. Comp. Biochem. Physiol. Pt. B 91: 351–357 (1988)
259. Sumida M. and Yamashita O.: Purification and some properties of soluble trehalase from midgut of pharate adult of the silkworm, Bombyx mori. Insect Biochem. 13: 257–265 (1983)
260. Sun S. C., Asling B. and Faye I.: Organization and expression of the immunoresponsive lysozyme gene in the giant silk moth, Hyalophora cecropia. J. Biol. Chem. 266: 6644–49 (1991)
261. Susskind B. M., Warren L. and Reeves R. E.: A pathway for the interconversion of hexose and pentose in the parasitic amoeba Entamoeba histolytica. Biochem. J. 204: 191–196 (1982)
262. Takano T. et al.: Polymorphism for the number of tandemly multiplicated glycerol-3-phosphate dehydrogenase genes in Drosophila melanogaster. Proc. Nat. Acad. Sci. USA 86: 5000–5004 (1989)
263. Takeuchi K.: Purification and characterization of exo-?-1,3-glucanase from a hatching supernatant of Strongylocentrotus intermedius. Can. J. Biochem. Cell Biol. 61: 54–62 (1983)
264. Tan A. W. H. and Nuttall F. Q.: In vivo phosphorylation of liver glycogen synthase. J. biol. Chem. 260: 4751–57 (1985)
265. Tanaka S. and Seno N.: A novel sulfated glycosaminoglycan, lingulan sulfate, composed of galactose and N-acetylgalactosamine from Lingula unguis. Biochim. biophys. Acta 704: 549–551 (1982)
266. Taylor E. C.: Cellulose digestion in a leaf eating insect, the Mexican bean beetle, Epilachna varivestis. Insect Biochem. 15: 315–320 (1985)
267. Terra W. R., Ferreira C. and Bastos F.: Phylogenetic considerations of insect digestion. Disaccharidases and the spatial organization of digestion in the Tenebrio molitor larvae. Insect Biochem. 15: 443–449 (1985)

268. Tiedtke A.: Purification and properties of secreted N-actyl-ß-D-hexosaminidase of Tetrahymena thermophila. Comp. Biochem. Physiol. Pt. B 75: 239–243 (1983)
269. Trainer D. G. and Tillinghast E. K.: Amylolytic activity of the crystalline style of Mya arenaria (Bivalvia, Mollusca). Comp. Biochem. Physiol. Pt. A 72: 99–103 (1982)
270. Trezise A. E. O. et al.: Cloning and sequencing of cDNA encoding baboon liver alcohol dehydrogenase: Evidence for a common ancestral lineage with the human alcohol dehydrogenase β subunit and for class I ADH gene duplications. Proc. Nat. Acad. Sci. USA 86: 5454–58 (1989)
271. Tsai C. S. et al.: Purification and comparative studies of alcohol dehydrogenases. Comp. Biochem. Physiol. Pt. B 87: 79–85 (1987)
272. Tsukada T. and Yoshino M.: β-Glucuronidase from Ampullaria. Purification and kinetic properties. Comp. Biochem. Physiol. Pt. B 86: 565–569 (1987)
273. Turnbull J. E. and Gallagher J. T.: Sequence analysis of heparan sulphate indicates defined location of N-sulphated glucosamine and iduronate 2-sulphate residues proximal to the protein-linkage region. Biochem. J. 277: 297–303 (1991)
274. Ueno R. and Yuan C. S.: Purification and properties of neutral β-N-acetylglucosaminidase from carp blood. Biochim. biophys. Acta 1074: 79–84 (1991)
275. Usanga E. A. and Luzzatto L.: Adaptation of Plasmodium falciparum to glucose-6-phosphate dehydrogenase-deficient host red cells by production of parasite-encoded enzyme. Nature 313: 793–795 (1985)
276. Vaandrager S. H. et al.: Fractionation and kinetic properties of trehalase from flight muscles and haemolymph of the locust, Locusta migratoria. Insect Biochem. 19: 89–94 (1989)
277. Vanderzel A. et al.: The involvement of catalase in alcohol metabolism in Drosophila melanogaster larvae. Arch. Biochem. Biophys. 287: 121–127 (1991)
278. Veivers P. C., O'Brien R. W. and Slaytor M.: Selective defaunation of Mastotermes darwiniensis and its effect on cellulose and starch digestion. Insect Biochem. 13: 95–101 (1983)
279. Vieira R. P., Mulloy B. and Mourao P. A. S.: Structure of a fucose-branched chondroitin sulfate from sea cucumber. Evidence for the presence of 3-O-sulfo-ß-D-glucuronyl residues. J. Biol. Chem. 266: 13530–36 (1991)
280. Villarroya A.: The primary structure of alcohol dehydrogenase from Drosophila lebanonensis. Extensive variation within insect „short chain" alcohol dehydrogenase lacking zinc. Eur. J. Biochem. 180: 191–197 (1989)
281. Vilotte J. L. et al.: Sequence of the goat α-lactalbumin-encoding gene: Comparison with the bovine gene and evidence of related sequences in the goat genome. Gene 98: 271–276 (1991)
282. Vulliamy T. J. et al.: Diverse point mutations in the human glucose-6-phosphate dehydrogenase gene cause enzyme deficiency and mild or severe hemolytic anemia. Proc. Nat. Acad. Sci. USA 85: 5171–75 (1988)
283. Vynios D. H. and Tsiganos C. P.: Squid proteoglycans: Isolation and characterization of three populations from cranial cartilage. Biochim. biophys. Acta 1033: 139–147 (1990)
284. Wakita M. and Hoshino S.: Physicochemical properties of a reserve polysaccharide from sheep rumen ciliates genus Entodinium. Comp. Biochem. Physiol. Pt. B 65: 571–574 (1980)
285. Wallace C. A. et al.: L-Gulonolactone oxidase is present in the invertebrate, Limulus polyphemus. Experientia 41: 485–486 (1985)
286. Weaver J. R., Andrews J. M. and Sullivan D. T.: Nucleotide sequence of the Adh-1 gene of Drosophila navojoa. Nucleic Acids Res. 17: 7524 (1989)
287. Weisman L. S., Krummel B. M. and Wilson A. C.: Evolutionary shift in the site of cleavage of prelysozyme. J. biol. Chem. 261: 2309–13 (1986)
288. White J. W.: Honey. Adv. Food Res. 24: 287–374 (1978)
289. Wiebauer K. et al.: A 78-kilobase region of mouse chromosome 3 contains salivary and pancreatic amylase genes and a pseudogene. Proc. Nat. Acad. Sci. USA 82: 5446–49 (1985)
290. Winberg J. O. et al.: Biochemical properties of alcohol dehydrogenase from Drosophila lebanonensis. Biochem. J. 235: 481–490 (1986)
291. Wolfenson C. and de Lederkremer R. M.: Purification and composition of glycoconjugates from Crithidia oncopelti. Comp. Biochem. Physiol. Pt. B 77: 555–560 (1984)
292. Xavier M. T. et al.: Chemical structures of a galactose-rich glycoprotein of Leishmania tarentolae. Comp. Biochem. Physiol. Pt. B 88: 101–104 (1987)
293. Yamasaki K.: Characterization and partial purification of a mannanlike polysaccharide in the eggs of Locusta migratoria. Insect Biochem. 3: 79–90 (1973)
294. Yet M. G. and Wold F.: The distribution of glycan structures in individual N-glycosylation sites in animal and plant glycoproteins. Arch. Biochem. Biophys. 278: 356–364 (1990)
295. Yokota K., Nishi Y. and Takesue Y.: Purification and characterization of amphiphilic trehalase from rabbit small intestine. Biochim. biophys. Acta 881: 405–414 (1986)
296. Zachary D. and Hoffmann D.: Lysozyme is stored in the granules of certain haemocyte types in Locusta. J. Insect Physiol. 30: 405–411 (1984)
297. Zera A. J., Koehn R. K. and Hall J. G.: Allozymes and biological adaptation. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 633–674. Pergamon Press, Oxford 1985
298. Zhang W. et al.: Primary structure of rabbit skeletal muscle glycogen synthase deduced from cDNA clones. Faseb J. 3: 2532–36 (1989)
299. Zinkler D., Götze M. and Fabian K.: Cellulose digestion in „primitive insects" (Apterygota) and oribatid mites. Zool. Beitr. N. F. 30: 17–28 (1986)

14 Glycolysis

14.1 The Pathway from Hexose to Pyruvate

The **enzyme system of glycolysis** (Fig. 14.1) is present in all animal cells. This pathway, together with the citric acid cycle and respiratory chain phosphorylation, is involved in the provision of energy under aerobic conditions. In contrast to other substrates, however, carbohydrates can also serve as a source of energy under anaerobic conditions. The reactions and end products of anaerobic carbohydrate catabolism vary greatly within the animal kingdom, and thus provide an especially interesting subject for comparative biochemistry. However, glycolysis should be considered not only as a source of energy but also as the starting point for the biosynthesis of important structural components, e.g. hexose phosphate for the biosynthesis of carbohydrates, dihydroxyacetone phosphate and glycerol phosphate for lipid biosynthesis, and 3-phosphoglycerate and pyruvate for the biosynthesis of various amino acids. The synthetic metabolism may constitute a considerable proportion of the total glycolytic turnover. Thus, in vertebrate liver and other lipogenic tissues, the greater part of the pyruvate arising from glycolysis and the subsequent acetyl-CoA is used in fatty acid synthesis.

The enzymes of glycolysis are a textbook example of **cytoplasmic localization**. However, they are not all found in soluble form in the cytosol; they may also be more or less tightly bound to cytoplasmic structures such as membranes or the cytoskeleton. The binding of enzymes to cell structures has the effect of a metabolic compartmentalization and can change the kinetic-regulatory properties of the enzymes. This balance between free and bound enzymes is influenced by the concentration of low molecular weight effectors. In vertebrate skeletal muscle, triose phosphate isomerase binds to myofibrills only when these already carry bound aldolase or glyceraldehyde-3-phosphate dehydrogenase; thus, the living muscle contains a mini-enzyme complex which catalyses the reaction step from fructose-1,6-bisphosphate to 3-phosphoglycerate [234]. In glioma cells, glycolytic enzymes are found to be associated with the inner side of the plasma membrane [44]. The regulatory importance of the reversible binding of glycolytic enzymes to cell structures has often been postulated, but this has also been contradicted [20, 193]. Very convincing examples are found in the sea snails: binding to structures in the muscles of *Busycotypus canaliculatum* is reduced during oxygen deficiency, and the rate of glycolysis diminishes; the

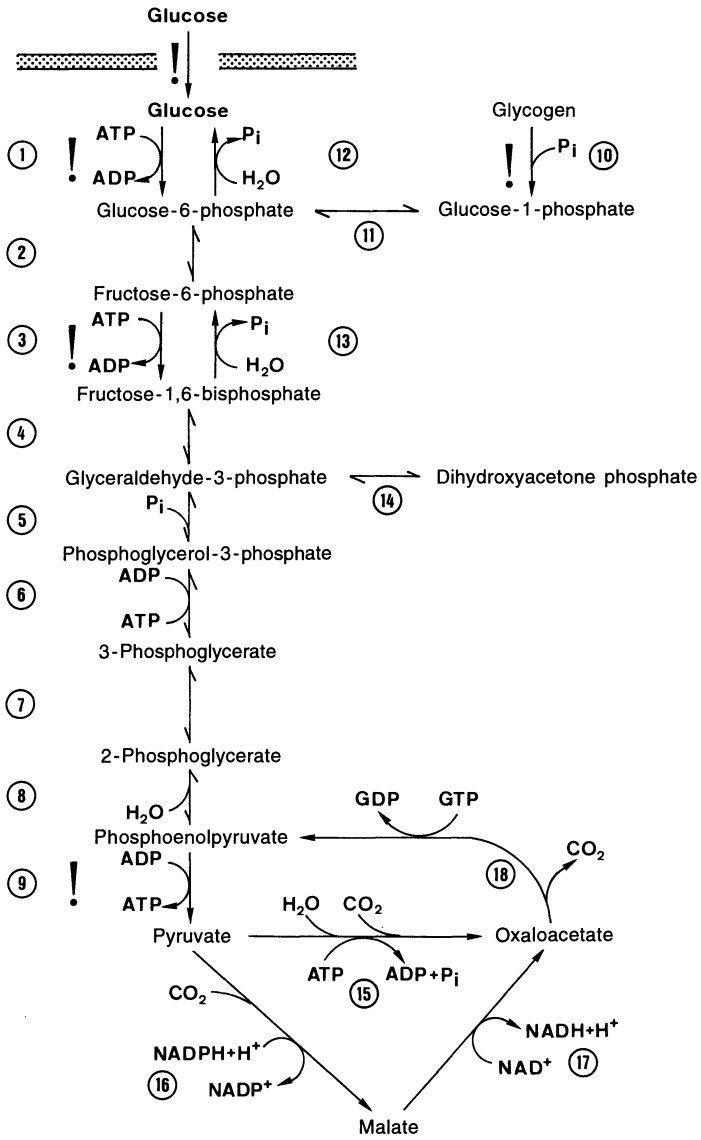


Fig. 14.1. The pathways of glycolysis and gluconeogenesis. The regulatory reactions of glucose and glycogen catabolism, including membrane transport of glucose, are marked with *!*. In principle, gluconeogenesis involves the reverse of the degradation reaction, with the hydrolytic cleavage of phosphate replacing the reverse of the kinase reactions. The formation of enol pyruvate from pyruvate occurs via oxaloacetate in the liver, and via malate and oxaloacetate in skeletal muscle. The following enzymes are involved: *1*, hexokinase; *2*, hexose phosphate isomerase; *3*, phosphofructokinase; *4*, fructose bisphosphate aldolase; *5*, glyceraldehyde phosphate dehydrogenase; *6*, phosphoglycerate kinase; *7*, phosphoglycerate mutase; *8*, enolase; *9*, pyruvate kinase; *10*, phosphorylase; *11*, phosphoglucomutase; *12*, glucose-6-phosphatase; *13*, hexose bisphosphatase; *14*, triose phosphate isomerase; *15*, pyruvate carboxylase; *16*, malic enzyme; *17*, malate dehydrogenase; *18*, phosphoenolpyruvate carboxykinase

structural binding of five glycolytic enzymes increases in parallel to the rate of glycolysis in stimulated muscles of *Patella caerulea* [158].

In the ciliate *Tetrahymena*, phosphofructokinase and lactate dehydrogenase are completely bound to mitochondria, as is 50% of the glyceraldehyde-3-phosphate dehydrogenase [228]. The structural organization of the glycolytic enzymes in the Kinetoplastida is rather unique. Here, the mitochondria are defective and glucose is not fully catabolized to carbon dioxide and water during energy-yielding metabolism; this reduced efficiency results in approximately five times more intensive glucose degradation than occurs in mammalian cells. The first seven enzymes of glycolysis and two enzymes of glycerol metabolism are enclosed in microbody-like cell organelles **(glycosomes):** these enzymes are hexokinase, hexose phosphate isomerase, phosphofructokinase (PFK), aldolase, triose phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), phosphoglycerate kinase (PGK), glycerol-3-phosphate and glycerol kinase. These nine enzymes make up only 3.3 ± 1.6% of the total cell protein in *Trypanosoma brucei*. In the glycosomes the accumulation of the enzyme proteins up to 320–360 mg/ml shortens the diffusion distances and allows for high glycolytic flux rates. The individual enzymes may also lie in an appropriate spatial arrangement within the glycosome. It is not at all certain that the evolution of the glycosomes can be fully explained by shortened

diffusion pathways and interactions between the enzymes [10]. Most glycolytic enzymes are found exclusively in the glycosomes; only PGK and GAPDH have additional cytosomal isoenzymes. In addition to the glycolytic enzymes, the glycosomes contain further specific proteins [124]. All glycosomal proteins are formed on free ribosomes in the cytoplasm and are then introduced into the glycosomes. This has led to the question of the topogenic signals that direct this intracellular transport. Answers have been sought by comparing the glycosomal enzymes with the homologous cytoplasmic enzymes of mammals and, where appropriate, with those of the Kinetoplastida. Only two of the glycosomal enzymes of *T. brucei* show any differences to those of other organisms. The hexokinase of *T. brucei* is a hexamer of 51-kDa subunits, whereas the mammalian hexokinase is a monomer of 80–90 kDa. PFK is tetrameric as in the mammals, but the subunits have a mass of only 51 instead of 96–100 kDa. The other glycolytic enzymes of *T. brucei* have the same quaternary structure as those from mammalian cells but, with the exception of triose phosphate isomerase, have consistently 1–5-kDa larger subunits. All the enzymes of the glycosomes (again with the exception of triose phosphate isomerase) are strongly basic proteins with pI values between 8.8 and 10.2, i.e. 1–4 pH units higher than the corresponding mammalian enzymes [2, 25, 93, 162]. In contrast to the chloroplast, mitochondrial and secretory proteins, the glycosomal enzymes are not subjected to proteolytic processing when they reach the target organelle; the topogenic signals are thus located in the internal structure of the chain. According to the hypotheses of the Weirenga group, the signal consists of one or two closely situated, positively charged amino acid residues ("hot spots"). Recent results, especially with *Crithidia fasciculata*, have cast some doubt on this hot-spot theory [117, 181]. In addition to the glycosomal phosphoglycerate kinase (gPGK), *T. brucei* and *C. fasciculata* have a cytoplasmic PGK (cPGK) which is encoded by its own gene. The gPGK of *T. brucei* has a C-terminal extension of 20 predominantly basic amino acids but agrees 93 % with the cPGK. The C-terminal extension of the gPGK *C. fasciculata* has a length of 38 amino acids and consists mostly of hydrophobic residues and hydroxyamino acids; the N-terminal sequence has only two amino acid substitutions compared with the N-terminal sequence of cPGK. The aldolase and triose phosphate isomerase of *T. brucei* have no C-terminal extensions. Thus, there is apparently no universal topogenic signal for import into the glycosomes [181]. A polycistronic transcript is produced from the *T. brucei* PGK locus and this is then processed to give the mRNAs of the individual enzymes. In mammalian blood, the mRNA and protein of gPGK predominate, whereas in culture (which corresponds to the forms from the tsetse fly vector) the cPGK forms predominate [77].

A ubiquitous and intensively investigated enzyme system such as that of glycolysis provides an excellent opportunity for posing the following **basic questions about enzyme evolution:** Are enzymes which catalyse the consecutive steps of a reaction chain homologous, perhaps by derivation from a common gene? Can one assume high sequence agreement between enzymes of similar reaction specificity or which have similar ligands? The amino acid sequences and spatial structures are known for almost all glycolysis enzymes, and the rates of evolution of the enzymes appear to be relatively low (see Table 4.12, p. 161). Of the seven glycolytic enzymes that bind ATP or NAD^+, GAPDH, lactate dehydrogenase (LDH) and alcohol dehydrogenase (ADH) have similar spatial structures with an NAD-binding and a catalytic domain on either side of the active centre. The coenzyme-binding domains all consist of two mononucleotide-binding folds with very similar conformations (Rossmann folds). PGK also includes two such folds, whereas pyruvate kinase (PK) has only one, and in hexokinase and PFK there are none. Despite the similarities of spatial structure, the mononucleotide-binding folds of the different enzymes show no sequence similarity. In the glycolysis chain, there are several groups of enzymes with similar reaction specificities: the kinases, mutases, isomerases and dehydrogenases. However, there is no significant sequence similarity within such functional types; in fact, even enzymes with the same specificity in different organisms are not homologous: the ADH of *Drosophila* is not homologous to the enzymes from yeast or mammals; among all organisms, there are two types of aldolase; and in the animal kingdom there are two types of phosphoglycerate mutase (PGM) with very different properties. Thus, the enzyme system of glycolysis apparently arose early in evolution by the combination of independently derived enzymes; the present structural similarities are examples of convergence [59].

The ubiquitous reaction pathways of primary cellular metabolism are adapted to the different metabolic requirements of different animal groups and cell types. This occurs on the one

hand by variation in enzymes activities, and on the other hand through differences in the kinetic-regulatory properties of the enzymes. The **activities of the glycolytic enzymes** have been investigated in numerous tissues (Table 14.1). During comparisons of such enzyme spectra, it may be assumed (with some reservation) that differences in in vitro enzyme activities under more-or-less optimal conditions allow conclusions to be drawn about the flux rates in vivo. In this way, for example, the ratio of hexokinase:phosphorylase would indicate the proportions in which free or glycogen-bound glucose is involved in glycolysis, the ratio lactate or opine dehydrogenase:phosphofructokinase suggests the relative importance of anaerobic glycolysis, and the ratio of pyruvate kinase:phosphoenolpyruvate carboxykinase indicates the relative activities of glycolysis and gluconeogenesis or the relative importance of anaerobic energy release in the succinate pathway. Consecutive enzymes in those parts of the glycolytic reaction chain that have no branch points always occur in the same relative proportions, irrespective of the absolute activities; for this, Pette in 1965 coined the term "constant proportions group".

Table 14.1. The activities (U/g fr. wt.) at 25 °C of glycolysis enzymes in different cell types

	1	2	3	4	5
Phosphorylase	18	8		0.69	3.34
Hexokinase	2.4	12		0.05	0.13
Phosphoglucomutase	21			6.0	12.2
Phosphoglucose isomerase	112	20		18.9	43
Phosphofructokinase	1.5	17	11.7	2.3	12.0
Aldolase	5.0	13		2.7	15.7
Triose phosphate isomerase	476	2200		7.9	
Glyceraldehyde phosphate dehydrogenase	85	330	124	27	42
Phosphoglycerate kinase	75	340		11.9	14
Phosphoglycerate mutase	36	425		8.8	7.4
Enolase	8.0	42		9.9	16.3
Pyruvate kinase	25	90	6.9	7.5	40
Lactate dehydrogenase	115	1.5	28	8.1	4.0
Octopine dehydrogenase			283	17.5	117
Strombine dehydrogenase				2.7	18.1
Alanopine dehydrogenase			97	4.9	21.5
Phosphoenolpyruvate carboxykinase			11.3	0.38	0.08

1, Rat liver [237]; 2, flight muscle of the migratory locust *Locusta migratoria* [237]; 3, heart of the snail *Busycon contrarium* [51]; 4, adductor of the edible mussel *Mytilus edulis* [68]; 5, foot of the mussel *Cardium tuberculatum* [68]

Comparative biochemical analysis of the **regulatory mechanisms** is of critical importance for understanding biochemical adaptation. Although there are more relevant observations available for glycolysis than for any other reaction pathway, the presentation of a complete picture of all regulatory variants in the animal kingdom is not yet possible. Phosphofructokinase occupies the most important regulatory position, followed by pyruvate kinase, the membrane transport of glucose when this is the substrate, and hexokinase; where glycogen is involved, the phosphorylase is regulatory (Fig. 14.1). All animal cells have a membrane glucose transporter which allows facilitated diffusion, and the cells of the mammalian small intestine also have a Na^+/glucose cotransporter (symporter). In humans there are at least five different tissue-specific facilitative glucose transporters (GLUT1 to GLUT5), the genes for which lie on different chromosomes; their sequences of 492–509 amino acids differ by as much as 60% [81, 236, 281]. Enzymes whose activities are influenced by factors other than substrate concentration can be described as regulatory. Regulatory alterations of enzyme properties may occur via two different molecular mechanisms: the binding of allosteric effectors and the post-translational modification of the enzyme protein, e.g. phosphorylation/dephosphorylation.

For a reaction which lies far outside thermodynamic equilibrium, the catalysis in opposing directions by two enzymes existing side-by-side in the same compartment leads to a **substrate cycle** (futile cycle). Two possible regulatory effects on the enzymes of such a cycle can be distinguished. Effects that influence both enzymes to the same degree (e_c effects) may change the cycle rate but not the net flux. Effects that influence only one of the two enzymes (e_r effects) change the net flux and, correspondingly, the ratio cycle rate:net flux rate. The regulatory effect in this case is amplified. Whilst allosteric regulation of a single enzyme can only lead to relatively minor changes in the flux rate, there is in theory no limit to the amplification by allosteric effects in a substrate cycle [166]. However, each passage of the cycle results in energy losses, and in special cases such a substrate cycle serves for the release of the chemical energy of ATP as heat energy. Comparative biochemical data on the substrate cycle

between glucose and Glc-6-P as well as between Fru-6-P and Fru-1,6-P_2 will be presented in the sections dealing with the enzymes involved, namely hexokinase and phosphofructokinase.

The **regulation of glycolysis** in most animal cells appears to conform to the scheme determined from the many investigations of mammalian muscle (Fig. 14.1). However, the simple concept of a universal regulatory role for phosphofructokinase must be qualified even for the mammals. Whilst the ratio Fru-6-P:Fru-1,6-P_2 is greater than 1 for most mammalian cells, it is less than 0.1 in ascites tumour cells and in rat spermatids. Because in such cases the product of phosphofructokinase accumulates, this cannot be the rate-limiting enzyme [16]. Extreme deviations from the scheme of glycolysis regulation that occurs in yeast and mammalian cells are to be found in the Protozoa. For example, neither hexokinase nor phosphofructokinase have typical allosteric properties in *Trypanosoma cruzi* or *T. brucei;* the only known glycolysis regulatory mechanism in this case is the activation of pyruvate kinase by Fru-1,6-P_2 [25].

In general, glycolysis usually begins with free glucose or the glucose residues of glycogen, but other sugars in the diet can be introduced into the glycolysis pathway (Fig. 14.2). Three specific enzymes are required for **fructose catabolism**. Fructokinase produces Fru-1-P, which is then cleaved by an aldolase. Dihydroxyacetone phosphate and glyceraldehyde, after phosphorylation by triose kinase, are normal intermediates of glycolysis. Because of the low affinity of hexokinase for fructose, the direct introduction of fructose into the normal glycolysis pathway by phosphorylation to Fru-6-P is not significant. **Galactose** is phosphorylated by a galactokinase to Gal-1-P, which is converted to UDP-galactose by uridyl transfer; the action of UDP-galactose-4-epimerase produces UDP-glucose. The uridyl donor may be UTP, e.g. in adult humans, or it may be UDP-glucose, e.g. in infants and in the cestode *Hymenolepis diminuta* (see Fig. 13.7, p. 471). The **mannose** present in large amounts in vegetable diets is phosphorylated to Man-6-P by hexokinase and then converted to Fru-6-P by a phosphomannose isomerase. Although so far these reactions have been mainly determined in mammals, they are probably quite widely distributed.

14.2 Glycogen Phosphorylases

The (glycogen) phosphorylase catalyses the transfer of a glucose residue from glycogen to inorganic phosphate. Two further enzymes are required for **the complete catabolism of glycogen**, 4-α-glucan transferase and amylo-1,6-glucosidase; in yeast and the vertebrates these are associated into a single "debranching enzyme" (Fig. 14.3). This enzyme system has been isolated from elasmobranchs and from mammals; although it has not been investigated in detail in

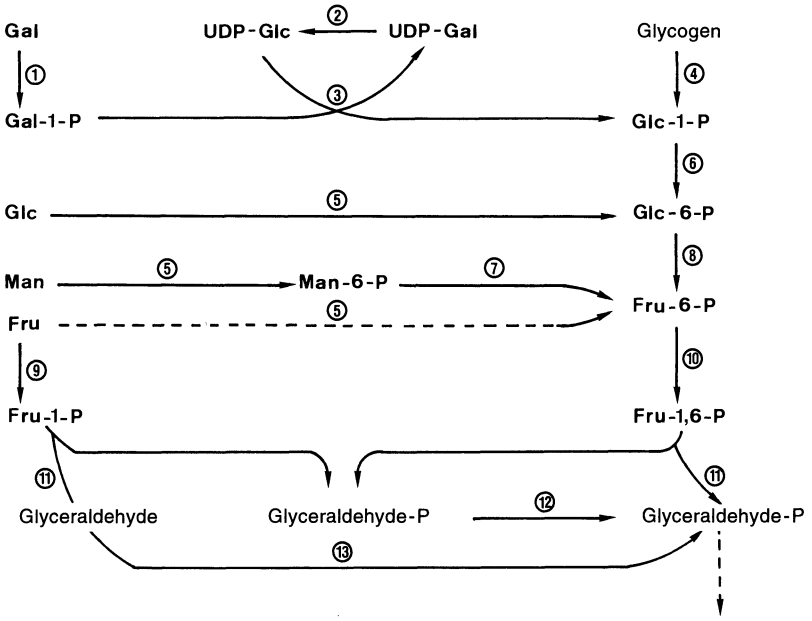


Fig. 14.2. The introduction of different sugars into glycolysis. *1*, Galactokinase; *2*, UDP-galactose-4-epimerase; *3*, UTP: galactose-1-phosphate uridilyl transferase or hexose-1-phosphate uridilyl transferase; *4*, phosphorylase; *5*, hexokinase; *6*, phosphoglucomutase; *7*, mannose phosphate isomerase; *8*, hexose phosphate isomerase; *9*, ketohexokinase; *10*, phosphofructokinase; *11*, aldolase; *12*, triose phosphate isomerase; *13*, triose kinase

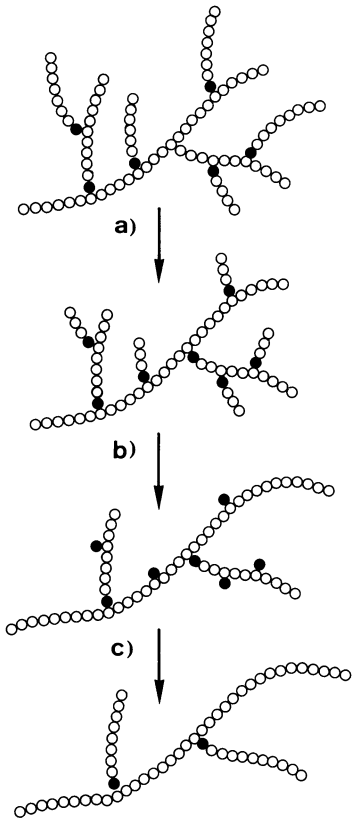


Fig. 14.3a–c. Degradation of glycogen by phosphorylase and the "debranching system". **a** The phosphorylase shortens the 1,4-linked outer chains stepwise down to three glucose residues. **b** Oligo-1,4-1,4-glucan transferase shifts maltotriose units between the outer chains. **c** Amylo-1,6-glucosidase cleaves off 1,6-linked glucose residues *(black)*. The "debranching system" of mammals and yeast is made up of a transferase and a glucosidase

the invertebrates, an amylo-1,6-glucosidase has been found in the cestode *Hymenolepis diminuta* [9]. It is reported that an exo-α(1–4)-glucosidase plays a significant role in the intracellular catabolism of glycogen by carp liver, and in fact the phosphorylase activity of this organ (9 U/g fr. wt.) is much lower than, for example, that of goldfish liver (45 U/g fr. wt.); it corresponds approximately to the activity of the exo-α(1–4)-glucosidase (5–6 U/g fr. wt.) [189]. **Glycogen phosphorylases** have mainly been isolated from vertebrates, e.g. from the muscle, liver, brain, fat tissue and leukocytes of various mammals, from the muscle and liver of teleosts [223], and from the muscles of the spiny dogfish *Squalus acanthias* and the agnathans *Entosphenus japonicus* and *Paramyxine atami* [286]. In the invertebrates, only the phosphorylases of molluscs, crustaceans and insects have been examined closely, e.g. from the adductor of the scallop *Patinopecten yessoensis* [94], from the claw muscle and various organs of the crayfish *Orconectes limosus* [98], from the flight muscles of the fly *Phormia regina*, from the fat bodies of *Bombyx mori*, from the flight muscles and fat bodies of the migratory locust *Locusta migratoria*, and from *Drosophila melanogaster* [46, 73, 150, 167, 233, 263]. In the vertebrates, the phosphorylases vary specifically with the organ; from the elasmobranchs onwards a muscle and a liver type may be distinguished, and from the teleosts onwards there is also a brain type [286]. In humans, the muscle phosphorylase (841 amino acids) is more similar to the liver type (846 amino acids) than to the brain type (862 amino acids) [176]. Two agnathans, the lamprey *Entosphenus japonicus* and the common hagfish *Paramyxine atami*, possess only one form of the phosphorylase. Thus, two duplications of the phosphorylase gene apparently occurred in the evolution of the vertebrates [286]. No isoenzymes of the phosphorylase have so far been found in the invertebrates. In some cases, a microheterogeneity of unknown origin has been detected: for example in the liver type, but not in the brain or muscle types of the rat, and in all the tissues examined of *Orconectes* [98].

The phosphorylases of yeast, green plants, vertebrates and insects all have subunits of 90–100 kDa, which may be combined as **monomers, dimers or tetramers**. Most vertebrate phosphorylases are dimeric but, in particular, those found in skeletal muscle are often tetrameric; the dimers of phosphorylase-a have higher activities than do the tetramers. The muscle phosphorylases of the shark *Squalus acanthias* exist only as dimers, and the brain type of the rat exist only as monomers. The crustaceans mostly possess dimers but may also have monomers and tetramers [98]. Monomers and dimers are found together in the edible mussel *Mytilus edulis;* only dimers exist in the scallop *Patinopecten yessoensis* [94], *Drosophila melanogaster* and *Locusta migratoria;* and only monomers occur in *Phormia regina* and *Bombyx mori* [46, 167, 233]. As well as the dimeric and tetrameric phosphorylases, the monomeric enzymes from insect flight muscle and fat bodies also show, under certain conditions, sigmoidal kinetics with Hill coefficients (n) of over 1.5 [167]. The complete **sequence** of the phosphorylase from rabbit muscle consists of 843 amino acids and has no significant similarities to any other proteins, i.e. these enzymes belong to a unique super-family. The region of positions 560–713 resembles in **spatial structure** the nucleotide-binding fold of the dehydrogenases,

but is not similar in sequence. Amongst the special features of the structure are the occurrence of a prosthetic group pyridoxal phosphate on 679-Lys and an acetylated N-terminal serine [58, 175]. The *Drosophila melanogaster* phosphorylase also includes one pyridoxal phosphate per subunit [46].

In all animals, the **regulation of glycogen phosphorylase** involves allosteric effects on the one hand, and enzyme conversion through phosphorylation/dephosphorylation on the other hand. Both regulatory mechanisms are active only for animal enzymes; the potato phosphorylase has no allosteric control and is always active, and the yeast enzyme is activated by phosphorylation but not by AMP. With the exception of 50–80 N-terminal amino acids, the sequences of rabbit muscle, potato, bakers' yeast and *Escherichia coli* agree at more than 50 % of positions; the N-terminal regions, which in the case of the rabbit enzyme includes the phosphorylation site (14-Ser) and the AMP-binding site, are quite different. This region was apparently added quite late in the evolution of the phosphorylases [110, 184]. Otherwise, glycogen phosphorylase is a very conservative enzyme: the human, rabbit and rat muscle enzymes agree at 96 % of positions, and 14 of the 18 N-terminal amino acids of the rabbit and of *Squalus acanthias* are identical. Monomers from rabbit muscle can combine to give active dimers not only with monomers from rabbit liver but also with monomers from frog muscle [58, 110]. Even the *E. coli* enzyme has 49 % of its 790 amino acids in common with the mammalian enzyme [35].

The phosphorylase exists in at least two forms in all animals: the phosphorylated "a" and the dephosphorylated "b" forms. The simultaneous determination of both forms is made possible by the fact that the b form is usually active only in the presence of AMP, whereas the a form is active without AMP. However, phosphorylase b of mammalian liver shows some activity in the absence of AMP, particularly in the rat and mouse. The affinity of phosphorylase b for allosteric activators varies with the species; for example, the K_a (AMP) of the b forms of fat-body phosphorylases in insects ranges from 0.1 to 2 mmol/l. The affinity for AMP of the b form is increased by P_i, Glc-1-P and Ca^{2+}; ATP and ADP hinder the binding of AMP to the enzyme and therefore reduce activity [233]. In the determination of the relative proportions of the a and b forms, it must be borne in mind that the a form is also stimulated by AMP. This AMP effect involves an increase in the substrate affinity without any change in V_{max}. ATP inhibits phosphorylase a only in so far as it prevents the enhancement of activity by AMP [58]. Stimulation of phosphorylase a by AMP is found in mammals and in various fish species, and is particularly marked in the insects. The activity of phosphorylase a in the flight muscles of the blowfly *Phormia regina* at low substrate concentrations is stimulated 10-fold by as little as 1 μmol/l AMP, whereas a 1000-fold higher AMP concentration is required for the b form; thus, as in the case of the mammals, phosphorylase a is most likely saturated with AMP [233]. The limiting factor for glycogenolysis in muscle is the concentration of free phosphate, P_i, which is simultaneously the substrate of the phosphorylase and the allosteric activator of phosphofructokinase. At the P_i concentration of 1–3 μmol/g found in resting muscle, the conversion of phosphorylase a to b would have no effect. However, cleavage of phosphagen results in a 10- to 20-fold increase of P_i in active muscle [85, 274].

The conversion of phosphorylase b to a involves the phosphorylation of 14-Ser by phosphorylase b kinase; dephosphorylation is brought about by a relatively unspecific protein phosphatase. **Phosphorylase b kinase** from rabbit muscle is a protein of 1.3 MDa built up of 16 subunits according to the formula $(\alpha\beta\gamma\delta)_4$ and present as tissue-specific isoenzymes. The sequence of the catalytic γ chain (41–45 kDa) is similar to that of other protein kinases. The δ chain is identical to calmodulin and confers calcium sensitivity to the enzyme. The α and β chains are homologous and have regulatory functions. In the rabbit and other mammals, the phosphorylase b kinases of white and red skeletal muscle are activated by exogenous Ca^{2+}-calmodulin. The heart muscle of various mammals, together with the white and red muscles of the chicken, contain the same enzyme type as mammalian red muscle contains [3]. The phosphorylation of single serine residues in the regulatory chains activates phosphorylase b kinase. Seven phosphorylation sites have been identified in the α chain (118–145 kDa) and three in the β chain (108–128 kDa) [120]. The cAMP-dependent protein kinase responsible for the phosphorylation of these sites is identical to the enzyme that converts active glycogen synthase I by phosphorylation to the in vivo inactive form D.

The cAMP dependence of this protein kinase is the basis for **the hormonal regulation of the phosphorylases**. Thus, binding of adrenalin to

liver-cell receptors initiates the chain of activation of adenylate cyclase, increase in the level of intracellular cAMP, activation of protein kinase, phosphorylation and, thereby, activation of phosphorylase b kinase, phosphorylation and activation of the glycogen phosphorylase. At the same time, glycogen synthesis is inactivated by the cAMP-activated protein kinase; amongst other effects, this results in an increase in the blood sugar level. The very rapid activation of glycolysis in active muscle is due to the fact that the increase in sarcoplasmic calcium concentration, triggered by the nerve impulse, not only stimulates muscle contraction but also activates the phosphorylase b kinase. Even the non-phosphorylated enzyme can be activated by binding of Ca^{2+} to the calmodulin subunit; the calcium sensitivity is further increased by phosphorylation.

Modification by phosphorylation and dephosphorylation appears to be a common mechanism in the animal phosphorylases. Phosphorylases and phosphorylase kinases of quite different origin can be freely combined, e.g. the kinase from rabbit muscle can combine with the phosphorylase of the agnathan *Entosphenus japonicus* [286], and even the phosphorylase b of rabbit muscle can combine with the kinase from *Bombyx mori* fat bodies. Fat-body extracts from the lepidopterans *Hyalophora cecropia* or *Manduca sexta* dephosphorylate rabbit phosphorylase a [233]. The phosphorylase of *Drosophila melanogaster* is phosphorylated and dephosphorylated in vitro, not only by the species' own enzymes but also by those of rabbit muscle [46]. On the other hand, the different systems of the phosphorylation/dephosphorylation of glycogen phosphorylase show considerable species-specific differences. For example, the phosphorylase b kinase from the muscle of the spiny dogfish *Squalus acanthias* shows greater specificity than does that of the rabbit; this enzyme cannot utilize GTP as the phosphate donor nor can it phosphorylate troponin. The *Squalus* enzyme is absolutely calcium dependent but is neither phosphorylated nor dephosphorylated; it is not clear whether it is hormonally regulated [194]. The idea that the conversion of the b form to the a form is central to the regulation of animal glycogen phosphorylases has been brought into question by investigations of the lugworm *Arenicola marina*. Normally one-third of the enzyme is found as the a form; anoxia has no effect on the proportion of the a form or on the total phosphorylase activity. The isolated b form is phosphorylated only 40 % by the rabbit-muscle phosphorylase kinase. The increase in the rate of glycolysis during anoxia is apparently not related to phosphorylation of the b form but to the AMP activation of the a form present [115].

The **insect phosphorylases** are stimulated by octopamine and by adipokinetic or hyperglycaemic neuropeptides (AKHs). There is some suggestion that cAMP-dependent protein kinase and phosphorylase b kinase are involved in this hormonal regulation in a way that is similar to that in higher vertebrates, but there is no direct evidence for this. Hormone-induced increases in the cAMP level and activation of the phosphorylase by cAMP have already been demonstrated in several insect species, but it is still possible that cAMP functions indirectly via stimulation of calcium uptake into the cells [46, 233]. In the cockroach *Periplaneta americana*, the conversion of phosphorylase b into the a form, brought about by the hyperglycaemic peptide hormone, is not accompanied by an increase in the cAMP level [182].

Cold activates the fat-body phosphorylase in many insects; as a result, the glycogen stored in the fat body is hydrolysed and large amounts of trehalose or polyvalent alcohols, such as glycerol and/or sorbitol, are produced and accumulated. For example, in the diapause pupae of the silkworm *Hyalophora cecropia* the glycerol concentration in the haemolymph increases within 30 days at 4 °C from 250 mmol/l to over 600 mmol/l; coincident with the decrease in temperature, the proportion of phosphorylase a increases within 1 h from about 10 to more than 50 %. This also occurs in isolated fat bodies and can be explained by the fact that at temperatures of around 0 °C the phosphorylase kinase is still weakly active but the phosphatase is completely inactive [33, 95, 233]. Cold activation of glycogen phosphorylase has also been found in the liver of the terrestrial hibernating frog *Rana sylvatica*. This amphibian survives freezing of its extracellular fluids because its tissues are protected by glucose concentrations of over 200 μmol/g wet weight [42].

The in vitro phosphorylation of phosphorylase b may result in a partially phosphorylated a/b form with intermediate catalytic properties. In some animals, three variously phosphorylated phosphorylase forms have also been found in living cells, e.g. in the insects *Locusta migratoria* and *Periplaneta americana*, in the lobster, in the scallop *Pecten maximus*, and even in the muscles of the crucian carp *Carassius carassius* [73, 150, 223].

14.3 Hexokinases

The typical hexokinases of the animal kingdom transfer a phosphate residue from ATP to the C-6 of different hexoses; they are active, for example, with glucose, fructose, mannose, glucosamine and deoxyglucose, but hardly at all with galactose. C-6-phosphorylating enzymes with a narrower **specificity** have so far been found in several invertebrates; an N-acetylglucosamine kinase is found in the liver and other tissues of the mammals [203]. C-1-phosphorylating fructokinases and galactokinases appear to be more common, but have been relatively little investigated. A galactokinase isolated from the ciliate *Tetrahymena thermophila* was found to be very similar to the corresponding enzymes of yeast and mammalian liver [131].

Almost all vertebrates possess multiple hexokinases. The model for the **hexokinase spectrum of the mammals** is the rat liver (Table 14.2a). The hepatocytes contain four isoenzymes which are referred to as A–D, according to the order of their elution from DEAE cellulose, or as I–IV in the order of their increasing anodic electrophoretic mobility [203]. A(I), B(II) and C(III) are all monomers of 100 kDa with a high affinity for glucose. At sufficiently high substrate concentrations they metabolize fructose as well as glucose, and they are inhibited by relatively low concentrations of Glc-6-P; P_i counteracts the inhibition by Glc-6-P. Only hexokinase C is also inhibited by glucose. The hexokinases A–C are widely distributed in the body of the rat, with A occurring especially in the brain and the erythrocytes, and B occurring in muscle. In addition to the product inhibition by Glc-6-P, an important regulatory mechanism of hexokinase activity in the brain, and thus of the cerebral rate of glycolysis, is the reversible binding of the enzyme to the ATP-producing mitochondria. The hexokinase interacts with the pores of the outer mitochondrial membrane and, at the same times, adopts a tetrameric structure [279]. Hexokinase D(IV) is found only in the liver and the pancreatic β cells. This differs from the other

Table 14.2. Hexokinases

a The four hexokinases of rat liver [26]

	A	B	C	D
K_m (Glc) in mmol/l	0.044	0.130	0.020	7.5
K_m (Fru) in mmol/l	3.1	3.0	1.3	420
V_{max} (Fru)/V_{max} (Glc)	1.1	1.2	1.3	2.4

b Kinetic properties of the hexokinases from various invertebrates. The apparent Michaelis constants are given for D-glucose, D-mannose and D-fructose in mmol/l. Values in parentheses are turnover rates relative to glucose (= 100)

Origin and type of enzyme		Glc	Man		Fru		References
Trypanosoma cruzi		0.09	–		0.35		[25]
Dirofilaria immitis		0.32	–		0.86		[109]
Angiostrongylus cantonensis		0.22	0.07	(35)	85	(72)	[179]
Mytilus galloprovincialis	I	0.036	–		14		[214]
	II	0.08	–		13		
Limulus polyphemus							[80]
Hepatopancreas	I	0.09	0.05	(47)	25	(128)	
	II	8	0.04	(6500)	1.6	(1350)	
	III	0.11	–				
Muscle		0.15	0.14	(70)	33	(130)	
Homarus americanus	I	0.08	0.13		6.7		[235]
	II	6	0.07		1.2		
Drosophila melanogaster	A	0.30	0.19	(100)	50	(50)	[169]
	B-1	0.18	0.29	(98)	50	(41)	
	C-2	0.09	0.33	(44)	2.5	(134)	
Bombyx mori larvae	I	0.024	–		1.2		[280]
	II	3.4	–		2.4		
	III	4.9	–		0.21		
	IV	8.6	–		0.88		
Asterias amurensis		0.045	–		4		[164]

three isoenzymes by its smaller size (50 kDa), its much lower affinity for glucose and other substrates, and its reduced sensitivity to the inhibitory effect of Glc-6-P. At substrate concentrations of about 100 mmol/l, hexokinase D metabolizes fructose with much lower efficiency than it metabolizes glucose; however, reference to this enzyme as a glucokinase is misleading, as may be seen from the values of V_{max} (Table 14.2a). The liver contains a 60-kDa inhibitory protein that forms an inactive complex with glucokinase in the presence of Fru-6-P; this is inhibited by Fru-1-P [264]. The hexokinase D of the pancreatic β cells differs from the liver enzyme only at the 15 N-terminal positions, and is encoded by the same gene. The glucokinase gene has two promoters and is regulated by insulin in the liver and by glucose in the β cells [146].

The **hexokinase spectrum** found in the rat may not be taken as a model even for the mammals, and certainly not for the vertebrates as a whole. The D type with low substrate affinity is absent from the horse, llama, cow, goat, sheep and chinchilla, and the glucose-inhibited C type is not found in the domestic dog and many rodents; both are lacking in the domestic cat. In investigations of 97 vertebrate species from all classes, the D type was detected only in the turtles and amphibians in addition to the mammals; and the C type was detected only in several amphibians that lacked hexokinase A, and perhaps also in several teleosts. Neither the D nor the C type was found in any of 13 avian or 13 reptilian species (excluding the turtles); the hexokinases of the A and B types found in these latter species differed significantly from those of the other vertebrates in having a lower glucose affinity and a higher V_{max} with fructose [262]. The C-type hexokinase found to have high activity in the liver of the frog *Calyptocephalella caudiverba* shows sigmoidal ATP kinetics and other features that are missing in the corresponding enzyme from *Rana catesbeiana*. The hexokinases of the vertebrates are, in general, ATP specific, although the enzyme from mammalian brain can also use ITP. A C-like hexokinase isolated from the liver of the rainbow trout shows an unusually wide nucleotide specificity; in addition to the different nucleotide triphosphates, ADP can be used as the phosphate donor, with the production of AMP [174]. Although multiple hexokinases are found in all other vertebrates, the lamprey *Entosphenus japonicus* has only one hexokinase of 90 kDa that is similar to the A type of mammals [165].

The hexokinase spectrum is dependent upon the developmental stage. In developing rat liver, A appears first, followed by B and C, and finally D. Each hepatocyte of the adult liver contains all four isoenzymes, whereby type D accounts for about 95 % of the glucose-phosphorylating activity. On the 10th day of brooding in the chicken there are four hexokinases present, of which one disappears on day 17 and a further one during hatching. In the frog *Calyptocephalella*, type D predominates in the liver of the tadpoles, but during metamorphosis C gradually completely replaces D. Type D reappears later and constitutes about 30 % of the total activity in the adult liver. B appears after metamorphosis but remains at a relatively low level, and A is absent from the liver at all stages [203, 262].

The 100-kDa hexokinases A–C are polypeptides of 917 amino acids, the C- and N-terminal halves of which agree in more than 50 % of their sequences. The binding sites for glucose and ATP in the C-terminal catalytic domain are present in the N-terminal regulatory domain as allosteric binding sites [256]. The 50-kDa hexokinase D from the rat liver agrees in 53 % of its 465 amino acids with the C-terminal catalytic domain of hexokinase A from the rat brain [4]. The 100-kDa hexokinases apparently arose by gene duplication from 50-kDa enzymes such as those found in yeast and higher plants. The 50-kDa enzymes in turn may have arisen from 25-kDa enzymes such as those found in the bacteria. The rat hexokinase D shows immunological cross-reactivity not only with type-D enzymes of other mammals but also with those of the turtles, although not with those of the amphibians. In contrast, antisera raised against rat type-A react only to a minor extent with the corresponding enzymes from other rodents, and not at all with those of the cow, sheep or pig. The rate of evolution of type A appears to far exceed that of type D [132, 262].

Multiple hexokinases are also found in many **invertebrate animals** (Table 14.2b). However, only one hexokinase was found in 11 out of 25 insect species, some echinoderms and the ascidian *Halocynthia roretzi* [165, 250]. Tissue-specific isoenzymes have been demonstrated in *Bombyx mori* and *Drosophila melanogaster*. In the silkworm, isoenzyme I is found in all organs including muscle; II occurs in the midgut, testis and Malpighian tubules; III is found only in the fat body; and IV is present in the midgut and fat body [280]. Particulate hexokinases that are bound to cell structures and are like those found in the vertebrates have also been described in

various invertebrates, e.g. in nematodes and cestodes [9, 179].

Most of the characterized hexokinases of the invertebrates show the typical wide substrate specificity, although with large differences in substrate affinities and relative turnover rates for different sugars. As shown in Table 14.2b, several arthropod enzymes have been found to have low glucose affinities, similar to that of vertebrate hexokinase D. On the other hand, many invertebrate enzymes have much greater affinities for fructose than do any of the four rat liver hexokinases. The predominating isoenzyme II of *Limulus* hepatopancreas may be classified as a fructomannokinase, whereas the muscle enzyme of this species shows the typical wide specificity of a hexokinase. The trematode *Schistosoma mansoni* possesses specific kinases for glucose, fructose, mannose and glucosamine, all of which phosphorylate in the 6-position. Kinases for specific hexoses have also been described for *Echinococcus granulosus, Ascaridia galli* and *Entamoeba histolytica*, whereas other parasitic worms have hexokinases with a wide specificity [9, 179]. As a rule, the invertebrate enzymes are also composed of 50- or 100-kDa subunits, corresponding to the scheme for hexokinase evolution described above. For example, the enzymes of the echinoderms *Asterias amurensis, Ophioplocus japonicus* and *Cucumaria frondosa* have subunits of about 50 kDa, although the latter species has dimers of 2 × 50 kDa [165]. The isoenzymes of the lobster *Homarus*, the silkworm *Bombyx mori* and the mosquito *Culex pipiens* are also about 50 kDa [235, 280]. The enzymes from the nematode *Angiostrongylus cantonensis* and from the fat body of the silkworm *Hyalophora cecropia* were found to be 60 and 64 kDa, respectively [63, 179]. The monomeric hexokinases from the roundworm *Ascaris suum* and from the flight muscles of the migratory locust *Schistocerca gregaria*, and isoenzyme I of the edible mussel *Mytilus galloprovincialis* have a mass of about 100 kDa. The isoenzyme II of the edible mussel species is in fact reported to be 165 kDa [214].

The majority of the invertebrate hexokinases examined are inhibited, like the vertebrate enzymes A–C, by Glc-6-P. Only a few of the invertebrate hexokinases show the absence of product inhibition that is typical for the rat enzyme D; examples include the hexokinases of the flagellates *Trypanosoma cruzi* and *Leishmania donovani*, the enzyme of the nematode *Angiostrongylus cantonensis*, and isoenzyme I from the hepatopancreas of the lobster *Homarus americanus* [25, 179, 235]. The isoenzyme III of *Bombyx mori* is the only known hexokinase to show substrate inhibition by ATP [280]. The hexokinase of *Apis mellifera* is inhibited by Glc-6-P but not by Man-6-P; thus, mannose is toxic for the honey bee because the absence of product inhibition results, after mannose feeding, in the excessive consumption of ATP, reduction in intracellular ATP and accumulation of Man-6-P. The **physiological regulation of hexokinase** has been examined in some detail in the flight muscles of the migratory locust *Schistocera gregaria*. The rate of glycolysis increases by about 50-fold at the beginning of flight. The muscle glycogen is used up in the first 10 s, and in the following minutes the trehalose and glucose of the haemolymph serve as a fuel; triacylglycerol gradually assumes this role after 10–20 min. The hexokinase must be at its maximal activity in order for the blood sugar to be introduced into glycolysis at the required rate. Although the Glc-6-P concentration of 200 mmol/l is far above the K_i value for the hexokinase, the substrate inhibition is completely counteracted by P_i from arginine-P, the alanine derived from pyruvate and the glycerol-3-P released [238].

A **substrate cycle** between glucose and Glc-6-P is active in mammalian liver, with a cycle rate of about 1 μmol/min per g fresh weight. At normal blood-sugar levels, the turnover rates of the hexokinase and glucose-6-phosphatase are about equal and the net flux is zero. Even small changes in the glucose or Glc-6-P concentration leads to a directed flux, and thus to the excretion or uptake of glucose by the liver cells. This regulation mechanism makes up for the lack of product inhibition of the predominant hexokinase D of the liver [100]. Comparative studies of the flight muscles of various insects have shown a significant correlation between the activities of hexokinase and glucose-6-phosphatase; this indicates that a substrate cycle between glucose and Glc-6-P is widespread in this tissue. In the moth *Acherontia atropos*, the cycle rate in the flight muscles at the beginning of flight increases from 0.06 to 3.9 μmol/min per g fresh weight; this allows for better regulation of the hexokinase reaction and, for example, more efficient adaptation to sudden changes in flight velocity [248].

14.4 Phosphofructokinases

The phosphofructokinases (PFKs) transfer a phosphate from ATP to fructose-6-phosphate. With some PFKs, e.g. from mammals, the tapeworm *Moniezia expansa* and the flagellate *Trypanosoma cruzi*, ITP and GTP also function as phosphate donors but, unlike ATP, are not effective as allosteric inhibitors [25]. The PFKs are mostly present as tetramers with subunits of 35 kDa in the bacteria, 49 kDa in *Trypanosoma brucei*, and between 80 and 95 kDa in the metazoans [43, 49]. The yeast enzyme is an octamer with 100-kDa subunits; in the mammals, there is aggregation to octamers and higher polymers, but also dissociation to dimers [90]. The PFK of the carrot has a size of 5 MDa but its 60-kDa subunits are smaller than those of the yeast and mammals [277]. Membrane-bound PFKs are known from various mammals and the ciliate *Tetrahymena;* these differ in their kinetic and regulatory properties from the soluble form [209]. The subunits of PFK from rabbit muscle and mouse liver both consist of 780 amino acids and agree 68 % in their sequences [75]. The N- and C-terminal halves show significant homology to each other and to the 35-kDa bacterial enzyme. This supports the concept that the mammalian PFK arose by duplication and fusion of an ancestral PFK of about half the size [196]. However, subdivision of the rabbit PFK gene into 22 exons of 15–63 codons reveals no internal symmetry [135].

Tissue-specific PFK isoenzymes have been detected in all vertebrates examined. Four types of PFK subunit are known from humans and various other mammals, and these have different chemical, immunological and regulatory characteristics. Heterotetrameric forms may arise where different types are present in the same cell. Muscle contains only subunit A (or M), which is also predominant in other glycolytically active tissues such as the heart. Subunit B (or L) predominates in gluconeogenic organs such as the liver or kidney cortex. The third type, known as C or F or P, is found in the brain, kidney, placenta, fibroblasts, erythrocytes and blood platelets. Embryonal cells initially express all three genes. The tissue-specific pattern that arises later differs between species or even between breeding lines. Thus, human and rabbit brain contain all three subunits, with A predominating in humans and C predominating in the rabbit. Rat brain contains mainly, or exclusively, type A, B or C, depending upon the population examined [49, 208]. It was first discovered in 1983 that the gut mucosa has a type-D enzyme with unique regulatory properties [119].

Most comparative studies of PFK have concentrated mainly on the **regulatory characters**. The results of in vitro experiments on mammalian PFKs have given rise to a long list of low molecular weight substances that can function as activators, inhibitors or de-inhibitors. The binding of calmodulin to rabbit muscle PFK confers calcium sensitivity. In fact, PFK be looked upon as the prototype of a multiply modulated enzyme. Mammalian PFKs display a series of common features which are apparently valid for all vertebrates and, in principle, also for many invertebrates [34, 125, 217, 243, 244]. Positive cooperativity results in a sigmoidal curve describing the dependence of enzyme activity on the fructose-6-phosphate concentration; high ATP concentrations bring about an allosteric inhibition. Citrate, phosphocreatine or phosphoarginine, phosphoenolpyruvate and 6-phosphogluconate, for example, are inhibitory. AMP, cAMP, ADP, P_i, Fru-1,6-P_2 and NH_4^+ act as positive modulators, reduce the cooperativity with fructose-6-phosphate, and counteract the inhibition by ATP and citrate [100].

The activation of mammalian PFK by Fru-2,6-P_2 was first discovered in 1980 [107], and has since been demonstrated in other PFKs, e.g. those in amphibians, insects, molluscs, cestodes, nematodes, higher plants and fungi [34, 100, 188, 271, 272]. In contrast, the PFK of *Trypanosoma cruzi* is not affected by Fru-2,6-P_2 and that of the carrot is affected only to a minor degree [254, 277]. Fru-2,6-P_2 is the most effective known activator, being 1000-fold more active than Fru-1,6-P_2 with the PFK of mammalian liver. At high concentrations Fru-1,6-P_2 prevents the stimulation by Fru-2,6-P_2, and in this case has an inhibitory effect. Fru-2,6-P_2 has not yet been clearly shown to have a significant regulatory function in mammalian muscle; however, the Fru-2,6-P_2 concentration and glycolytic flux are closely correlated in muscle of the frog *Rana temporaria* [273]. The different activators can have synergistic effects. The four activators of the PFK present in the flight muscles of the cockroach *Periplaneta americana* at the physiological concentrations of 0.4 mmol/l AMP, 1 μmol/l Fru-2,6-P_2, 3 mmol/l NH_4^+, and 10 mmol/l P_i at pH 7 enhance the substrate affinity by 640-fold and reduce the affinity for the inhibitor ATP by 10-fold. In this way, the half-maximal PFK activity is achieved at 0.025 mmol/l, the in vivo concentration in muscle, instead of at 16 mmol/l [244].

Activation by Fru-2,6-P_2 may explain some effects that previously had been attributed to stable modifications of the enzyme. The PFK from mammalian liver and muscle can in fact be phosphorylated by a cAMP-dependent phosphokinase, but the catalytic or regulatory properties of the enzyme are not significantly changed. Of greater regulatory importance is the **phosphorylation** of the PFK from mammalian fat tissues, and especially of the PFKs from the molluscs, nematodes and trematodes [14, 147]. The PFK of the roundworm *Ascaris suum* is very similar to the mammalian enzyme in its regulatory characteristics, despite the somewhat larger size of the subunits. However, the phosphorylation of a particular serine residue leads to a three- to fourfold increase in activity under physiological conditions; this mainly involves a reduction in sensitivity to inhibitory ATP. The phosphorylation occurs in vitro with a cAMP-dependent protein kinase from bovine heart; *Ascaris* itself has two quite different protein kinases that are active on PFK, neither of which is influenced by cAMP or cGMP [125]. Similar results have been obtained with the PFK of the nematode *Dirofilaria immitis* [232]. The PFK of the liver-fluke *Fasciola hepatica* can also be phosphorylated by the cAMP-dependent mammalian enzyme and, as a consequence, has a higher affinity for Fru-6-P and reduced inhibition by ATP [147].

Fru-2,6-P_2 is produced from Fru-6-P by a specific 6-phosphofructo-2-kinase, designated PFK-2 to distinguish it from the classical PFK (PFK-1). The Fru-2,6-P_2-forming enzyme of the mammals cannot be separated from the Fru-2,6-P_2-cleaving bisphosphatase (FBPase-2), and in fact sequence data show that the **6-phosphofructo-2-kinase and the fructose-2,6-bisphosphatase** exist as domains of the same polypeptide chain of 470 amino acids; two such subunits make up a homodimer of 110 kDa. The kinase domains resemble PFK-1 in their spatial structure, and amino acids 250–349 of the bisphosphatase are significantly homologous to phosphoglycerate mutase; thus, the bifunctional enzyme apparently arose by gene fusion [11, 139]. Fru-2,6-P_2-synthesizing and -cleaving enzyme activities are also found in the same bifunctional enzyme protein in spinach leaves and in the flagellate *Euglena gracilis;* yeast, however, has two separate proteins [54, 122]. Mammalian liver, muscle, heart and brain have tissue-specific isoenzymes of the PFK-2/FBPase-2 with different functional and structural properties [265]. The mammalian PFK-2, like the PFK-1, is activated by AMP and inhibited by citrate and phosphoenolpyruvate but, in contrast to PFK-1, it is not inhibited by ATP. The phosphorylation of a particular serine residue of the PFK-2/FBPase-2 of mammalian liver by a cAMP-dependent protein kinase leads to inhibition of the kinase and activation of the bisphosphatase [121, 130]; the phosphorylation is triggered, for example, by glucagon or adrenalin. Phosphorylation of the mammalian heart enzyme has exactly the opposite effect [121].

The complicated regulatory properties of the animal PFKs evolved quite gradually. The enzymes of bacteria and the slime mould *Dictyostelium* show Michaelis-Menten kinetics with the substrate Fru-6-P and no allosteric effects. Sigmoidal Fru-6-P kinetics already exist for the PFKs of certain protozoans, e.g. *Leishmania histolytica, Crithidia fasciculata* and *Entamoeba histolytica*, but, with the exception of AMP, none of the usual modulators are effective. The PFK of *Trypanosoma brucei* shows little improved regulation and, like other glycolytic enzymes, is bound to the glycosome. This enzyme is tetrameric, like other PFKs, but is the "odd-man-out" with small subunits of only 49 kDa. AMP and ADP are stimulatory only at low concentrations, and are inhibitory at high concentrations. Citrate, Fru-1,6-P_2, Fru-2,6-P_2 and P_i have no effect [43]. The PFK of *T. cruzi* has similar characteristics [25]. The tissues of some metazoans also contain PFKs on which the typical effectors of the mammalian enzyme, such as ADP, Fru-1,6-P_2 and citrate, have no effect; this is true, for example, in the roundworm *Ascaris suum*, the shell-adductor muscle of the oyster *Crassostrea virginica*, and the flight muscle of the cockroach *Periplaneta americana* [244].

PFK as the key enzyme of glycolysis is adapted to the specific internal and external conditions of individual cells by **species- and tissue-specific differences in regulatory properties**. This becomes apparent when one compares the heart muscle of the rainbow trout *Salmo gairdneri* and the turtle *Pseudemys scripta*. In the turtle heart, as in mammalian heart, hypoxia results in an increase in the rate of glycolysis, and thus provision of ATP (Pasteur effect). This effect involves changes in the concentration of allosteric modulators of PFK, whereby the positive effectors AMP, ADP and P_i become increased and the negative effectors ATP and citrate are decreased. In trout heart, on the other hand, the modulator sensitivity of PFK, and thus the extent of the Pasteur effect, is much lower; this organ is to a great extent adapted to aerobic conditions [113]. A further example of the

adaptation of PFK to specific conditions is found in the production of antifreeze substances in overwintering larvae of the gallfly *Eurosta solidaginis*. As the ambient temperature gradually decreases, the haemolymph accumulates first glycerol and then sorbitol. This switch in polyol synthesis is due to the inactivation of PFK at low temperatures. The specific activity of PFK declines rapidly between +10 and 0 °C, the Q_{10} having the rather high value of 3.64. At the same time, the substrate affinity and AMP sensitivity decrease, and the sensitivity to inhibitors increases [240]. At the beginning of flight, the concentration of Fru-2,6-P_2 in the flight muscles of the migratory locust *Locusta migratoria* sinks dramatically. The consequently low PFK activity facilitates the conservation of carbohydrate reserves for prolonged flight. Deterioration of the energy status of the flight muscles results in an increase in AMP and P_i and activation of PFK [272]. The PFK of the honey-bee flight muscle is rather unusual in that it is inhibited by Glc-1,6-P_2 and Fru-1,6-P_2. This inhibition is relieved by Fru-6-P and, in particular, by Fru-2,6-P_2, but not by AMP. The minor regulatory significance of AMP coincides with the relatively low activity of adenylate kinase in these muscles [271].

The opposing reactions of PFK-1 and fructose-1,6-bisphosphatase give rise to a **substrate cycle** that allows for fine regulation of Fru-1,6-P_2 synthesis. AMP and Fru-2,6-P_2 activate PFK-1 and simultaneously inhibit the phosphatase; this regulation of the substrate cycle between Fru-6-P and Fru-1,6-P_2 is probably the most important biological function of Fru-2,6-P_2. Two extreme situations can be distinguished in the cells of the liver. In the hungry animal, the Fru-2,6-P_2 concentration is low, PFK-1 is inactive, and only gluconeogenesis occurs. In the well-fed animal, the Fru-2,6-P_2 concentration is high, PFK-1 is active, the phosphatase is more-or-less inhibited, and thus glycolysis predominates. Because the fructose-1,6-bisphosphatase in the presence of Fru-2,6-P_2 shows sigmoidal substrate kinetics, it can function simultaneously as an emergency brake that prevents the occurrence of excessive Fru-1,6-P_2 concentrations [100]. In the bumble bee *Bombus affinis*, ATP catabolism resulting from the Fru-6-P/Fru-1,6-P_2 substrate cycle serves to increase the temperature and, consequently, the capacity of the flight muscles. Accordingly, the cycle rate increases with decreasing temperature in the resting muscle and reaches 10.4 μmol/g per min at 5 °C, whereas the rate of glycolysis is only 5.8 μmol/g per min. At the onset of flight, the cycling is stopped and the rate of glycolysis increases to 20.4 μmol/g per min [248]. Fru-2,6-P_2 and Ca^{2+} are involved in the regulation of the substrate cycle in the flight muscles [82, 136].

14.5 Aldolases

The aldolases cleave Fru-1,6-P_2, and most of them also cleave Fru-1-P, to the corresponding trioses. In the vertebrates, the invertebrate metazoans and many unicellular organisms, aldolases are tetramers of 40-kDa subunits [62, 63, 163]. At least three different **types of subunit** are found in the vertebrates, and each is encoded by a different gene. Electrophoresis at pH 8.6 of mammalian aldolase preparations separates a weakly cathodic band A from the strongly cathodic band B and the anodic band C. A fourth subunit, D, is encoded in the tetraploid genome of the salmonids and lies between A and C on electropherograms. The subunits form heterotetramers both in vivo and in vitro. Type A_4 from skeletal muscle metabolizes Fru-1,6 P_2 many times more efficiently than it metabolizes Fru-1-P, and is therefore especially well adapted for glycolytic sugar catabolism. Type A is also the predominant form in mammalian embryos. Type B_4 from the liver, kidney cortex and small intestine shows high activity with Fru-1-P and is therefore especially well adapted to fructose metabolism and gluconeogenesis. Type B predominates in avian embryos. Type C_4 is found in addition to heterotetramers of C, A and B in the brain and the spinal cord; the catalytic properties of this type are intermediate between those of A_4 and B_4 [74].

The aldolases of many mammals, birds, amphibians and fish have been isolated and characterized [1], as have several from nematodes, insects and the edible snail [22, 63, 148, 239]. Most invertebrates appear to possess only one aldolase; an exception is the fly *Phormia regina*, in which different isoenzymes have been detected in the muscle and the fat body [63]. These enzymes are all very similar in their catalytic properties; the turnover rate of Fru-1,6-P_2 and the affinity for this substrate is usually much higher than for Fru-1-P; an exception is aldolase B from vertebrate liver (Table 14.3). In general, the aldolases are not considered to be important for the regulation of glycolytic flux. However, the enzyme from the flight muscles of the locust *Schistocerca gregaria* is specifically inhibited by citrate and acylcarnitine; this inhibition of aldolase may be important

Table 14.3. Maximal activities of various aldolases on fructose-1,6-bisphosphate compared with fructose-1-phosphate [22, 163, 239]

Aldolase	Ratio
Rabbit	
Type A (muscle)	50
Type B (liver)	1
Type C (brain)	10
Rat muscle	100
Trypanosoma sp.	100
Eimeria stiedai	21
Helix pomatia	20
Drosophila melanogaster pupae	10–15
Schistocerca gregaria flight muscle	37

in vivo for the switch in energy-yielding metabolism from carbohydrates to fatty acid oxidation [239].

The similarity in functional properties between these enzymes is based on a surprisingly high degree of structural similarity. The aldolase subunits are comprised of 360–363 amino acids. The **sequences** are known for all three isoenzymes of the rat and for aldolases of man and other mammals, chicken, *Drosophila melanogaster, Trypanosoma brucei* and maize [148, 172, 267]. The aldolases of mammalian muscle, with only 2% substitutions per 100 million years, are amongst the most conservative known enzymes (see Table 4.12, p. 161); the invertebrate aldolases, with 4% substitutions per 100 million years, have about the same **rate of evolution** as have the other glycolysis enzymes [62]. The sequence similarity between the aldolases of *Drosophila melanogaster* and the rabbit is 71%, and between the aldolase of *Trypanosoma brucei* and the vertebrate and insect enzymes about 45–48% [148, 267]. The close structural relationship between all aldolases is confirmed in a most impressive way by the combination of aldolase subunits from very different origins into functioning **hybrid tetramers**, e.g. type-C subunits from chicken brain will combine with subunits from lobster muscle or wheatgerm, but not with those from *Ascaris* or *Loligo*. The B subunits from rat liver do not form functional hybrid molecules with subunits from lobster or wheatgerm [249].

Sequence comparisons between A and B subunits of aldolase indicate that the separation of the isoenzymes is phylogenetically very old; the human B sequence agrees with that of the rabbit at 94% of positions, with the chicken subunit at 81% of positions, but with rabbit A at only 76% of positions [23, 185]. Aldolase C is closer to A than to B [186]. **Spatial structures**, albeit with low resolution, have been published for the enzymes from human and rat muscle and from *Drosophila* [62]. The secondary structure derived from the amino acid sequence is quite complicated with 12 helical regions and 19 folded-sheet regions [258]. Only one gene for chicken B is present per haploid genome; this consists of eight exons, seven of which are in the coding region, and one very long (4.6-kb) intron in the 5′-NT (non-translated) region [23]. It has been demonstrated for some **aldolase genes** that several mRNAs with the same coding sequence, but with different lengths of the NT sequence, are produced from the same gene by alternative splicing. The mRNA for rat A is longer in brain than in muscle, and in human liver there are at least four different B mRNAs [171, 213, 259].

Two **classes of aldolases** can be distinguished from among all organisms. The class-I aldolases, typical of animals and higher plants, produce a Schiff base from the substrate; the inhibitor sodium borohydride, which reacts with all such compounds, is effective on all enzymes of this class. The aldolases of class II differ from those of class I in size (70 compared with 160 kDa); they are metalloenzymes and are inhibited by ethylenediaminetetraacetic acid (EDTA). Most bacteria, lower fungi and algae contain only enzymes of class II, although some bacteria and algae possess both classes. The distribution of the two classes amongst the protozoans appears to follow no particular rule; *Tetrahymena, Leishmania, Trypanosoma* and *Eimeria* have class-I aldolases, whereas *Trichomonas* and *Entamoeba* have class II [163].

14.6 Glyceraldehyde Phosphate Dehydrogenases

The glyceraldehyde phosphate dehydrogenases (GAPDHs) catalyse a complex reaction in which glyceraldehyde-3-phosphate is oxidized to 3-phosphoglycerate, NAD^+ is reduced, and ATP is formed from P_i and ADP. This enzyme makes up 5 to 10% of the total protein in muscle. The GAPDHs from bacteria to man are all homotetramers. The **subunits** of eukaryotic GAPDH usually consist of 328–334 amino acids, corresponding to a mass of 37–38 kDa; the GAPDH of the nematode *Caenorhabditis elegans* is somewhat longer (340 amino acids). The polypeptide chain consists of an NAD-binding domain, which is

very similar in its spatial structure to the nucleotide-binding domains of other dehydrogenases (Rossmann fold), and a catalytic domain. Yeast GAPDH shows positive cooperativity, whereas the animal GAPDHs are negatively cooperative, with Hill coefficients of around 0.8. According to investigations of the muscle GAPDHs of the sturgeon species *Huso huso* and *Acipenser transmontanus*, this results from the fact that the first two NAD^+ are bound to non-neighbouring binding sites of high affinity, and the next two are bound to neighbouring interacting sites [266].

The **amino acid sequences** of the GAPDHs from *Drosophila*, lobster, chicken, pig, rat and humans have 63 % identical amino acids; the GAPDH of *Caenorhabditis* is 67 % similar to that of *Drosophila* and the chicken. The GAPDHs of bakers' yeast and two bacterial species have 32 % of amino acids in common with the animal enzymes [48, 159, 257, 283]. Thus, GAPDH is a protein with a relatively low **rate of evolution** (see Table 4.12, p. 161). In agreement with this conclusion is the fact that the GAPDH subunits of rabbit and yeast can form functional heterotetramers. At least two-thirds of the amino acids that are invariant in all organisms are found in the catalytic domain; the nucleotide-binding domain is highly conserved in its spatial structure but not in its amino acid sequence [48, 257].

The chicken **GAPDH gene** has 11 introns, 3 of which lie on the boundaries between the nucleotide-binding domain, the catalytic domain and the helical tail region. The human gene has only eight introns, and the equivalent of chicken exons VIII–XI are fused together; the remaining introns occupy similar positions to those in the chicken gene [55]. The GAPDH gene of *Caenorhabditis elegans* has only two introns [283]. Although only one enzyme protein is produced, the human genome contains 100 copies of the gene, and in the rat there are as many as 300–400 copies; many pseudogenes are found without introns and with many nucleotide substitutions. The presence of transcription products with differing sequences suggests that some of these pseudogenes are transcribed [210]. There is apparently only one functional human gene, and isoenzymes are in general not found in the vertebrates. In contrast, the nematode *Caenorhabditis elegans* has two tissue-specific isoenzymes encoded by two pairs of genes. The almost identical genes, gdp-1 and gdp-4, give rise to GAPDH-1, which is found at low activity in all cells; genes gdp-2 and gdp.3 produce the more active, muscle-specific GAPDH-2. The two isoenzymes differ mostly in a cluster of amino acids at the N-terminus [106, 282]. *Drosophila* also has two GAPDH isoenzymes encoded by different genes [247], and *Trypanosoma brucei* possesses two GAPDH isoenzymes with different intracellular localizations: one is found in the glycosomes and makes up 80 % of the total activity, whilst the other is soluble in the cytosol. The glycosomal enzyme (358 amino acids) has only 55 % of its sequence in common with the cytosolic enzyme (330 amino acids), and is much more basic with a charge excess of +11 compared with +2 for the cytosolic enzyme. The glycosomal enzyme is about 4 kDa larger than the cytosolic enzyme as the result of seven insertions of one to eight amino acids and a C-terminal extension of seven amino acids [160].

14.7 Pyruvate Kinases

The pyruvate kinases (PK) of the vertebrates and invertebrates are usually tetramers of about 60-kDa subunits. The liver of the rat and the flounder *Platichthys flesus* also contain higher polymers of several megadaltons [36, 216, 245]. Tissue-specific isoenzymes of PK are found in all vertebrates (with the exception of the cartilaginous fish), and these differ markedly in their structural and kinetic-regulatory properties. All PKs are apparently homologous; those from bakers' yeast and the chicken agree at 55 % of positions [142].

The PKs constitute the last potential point of regulation of glycolysis; however, there are radical differences in kinetic-regulatory properties between the tissue-specific isoenzymes of the vertebrates and the PKs of various invertebrates. The necessity for **PK regulation** in certain metabolic situations is obvious. During gluconeogenesis, for example, the PK reaction cannot be reversed for thermodynamic reasons; the conversion of pyruvate to phosphoenolpyruvate (PEP) occurs rather by the enzymes pyruvate carboxylase (PC) and phosphoenolpyruvate carboxykinase (PEPCK). However, even in tissues with a relatively high capacity for gluconeogenesis, e.g. mammalian liver, the maximal PK activity is significantly higher than the PC and PEPCK activities. Thus, PK must be down-regulated during gluconeogenesis. The inhibition of PK is also necessary during hypoxia, when pyruvate is not reduced to lactate or opines and other routes for

PEP metabolism must be introduced; this is especially the case in many invertebrates [173].

14.7.1 Pyruvate Kinase Isoenzymes of Mammals

Mammals and birds have four **tissue-specific pyruvate kinase isoenzymes**. All the activity in skeletal muscle and most of the activity in heart muscle and the brain stems from PK-M_1 (also known as PK-M or -A). The PK-M_2 isoenzyme (also known as PK-K or -C) is widely distributed in the body and is the most important isoenzyme of the kidney medulla, lungs, gut, fat tissue and leukocytes. PK-M_1 and -M_2 arise by alternative splicing of transcripts from the same gene [252]. PK-L (also known as PK-B) is the predominant form in the liver; it is present only in the hepatocytes, and the other cells of the liver contain PK-M. The proximal kidney tubules also contain PK-L. The fourth isoenzyme, PK-R, is the only form found in erythrocytes. Embryonal muscle initially has PK-M_2, but this is gradually replaced by PK-M_1. The occurrence of M_2 and M_1 or L in the same cell leads to the formation of M_1M_2 or LM_2 heterotetramers [12].

The **PK-L of mammalian liver** is regulated allosterically. It is activated by PEP, Fru-1,6-P_2 and H^+, and inhibited by ATP, OH^-, alanine, phenylalanine or other amino acids; when inhibited it shows sigmoidal PEP kinetics. Thus, apart from the PK substrate itself, the product of the previous regulatory enzyme, PFK, is also stimulatory. At physiological PEP concentrations, PK-L has very low activity, but it is strongly stimulated by only micromolar concentrations of Fru-1,6-P_2. Alanine and other glucogenic amino acids have an inhibitory effect on PK and therefore favour gluconeogenesis. A cAMP-dependent protein kinase can phosphorylate PK-L on the N-terminal serine residues of the four subunits. The phosphorylated form has a lower affinity for the activators PEP and Fru-1.6-P_2 but a higher affinity for the inhibitors ATP and alanine. Conversely, the activators can hinder the phosphorylation. Because both activators and inhibitors are present under physiological conditions, the enzyme activity in vivo is strongly reduced by phosphorylation; V_{max}, however, remains unchanged. The hormonal stimulation of gluconeogenesis involves, in particular, the cAMP-dependent phosphorylation and inactivation of the liver PK. In contrast, the PE/pyruvate substrate cycle appears to be relatively unimportant for the regulation of glycolysis and gluconeogenesis. This substrate cycle results from the production of pyruvate from PEP by the cytosolic PK and the simultaneous conversion of pyruvate into PEP by the combined activities of the mitochondrial PC and the cytosolic-mitochondrial PEPCK (see Fig. 14.1) The dephosphorylation of PK is brought about by a non-specific protein phosphatase which is not the same as the phosphorylase dephosphorylating enzyme. The PK-M_2 and the PK-R of the erythrocytes are similar to the liver enzyme in their allosteric regulation; however, the homo- and heterotropic allosteric interactions are less marked in PK-M_2 than in PK-L and PK-R [100, 253].

In contrast to the other isoenzymes, the **PK-M_1 of muscle** always shows hyperbolic kinetics with PEP as the substrate, and it is not regulated allosterically under physiological conditions. PK-M is activated only by Fru-1,6-P_2 at non-physiological, lower magnesium concentrations or in the presence of the inhibitor phenylalanine [173, 178]. The seal *Phoca hispida* and other diving marine mammals differ from the other mammals in that their skeletal muscle and heart contain different PK isoenzymes. The heart form corresponds to the M-type of other species, whereas the muscle enzyme appears to be particularly well adapted to wide temperature fluctuations and to the alternation of long periods of anaerobic glycolysis with short phases of fatty acid oxidation [86].

The **sequence** of 543 amino acids of the rat PK-L agrees by 67–68 % with the two rat PK-M forms, by 68 % with the PK-M_1 of the chicken, and still by 48 % with the yeast PK [111, 143, 178]. The sequences of rat PK-M_1 and -M_2 (531 amino acids) are identical apart from a partial sequence of 45 amino acids in the C-terminal domain (amino acids 388–432). The corresponding mRNAs are apparently produced from the primary transcript of the same gene by alternative splicing. In the variable partial sequence, M_2 is more similar to the L type than to M_1, and this region thus appears to be responsible for the allosteric regulation of PK [178]. PK-R is very similar to PK-L in amino acid composition, cleavage peptide pattern and immunological characteristics [111]. The erythrocyte PKs of different mammals show immunological cross-reactivity but vary greatly in their kinetics, e.g. the Hill coefficients of PEP kinetics vary under the same conditions from a value of 0.94 in the opossum *Didelphis marsupialis*, through values of 1.11 in humans and 1.53 in rabbits to a value of 2.41 in cattle; the opossum shows no ATP inhibition.

14.7.2 Pyruvate Kinase Isoenzymes of Other Vertebrates

The **pattern of PK isoenzymes** from the mammals to the amphibians is very similar. It is, of course, questionable whether PK forms from lower vertebrates or even invertebrates can be related to particular isoenzymes of the mammals solely on the basis of electrophoretic data; at least their kinetic-regulatory properties should also be taken into consideration. The L type is not detected electrophoretically in the birds or the urodelans, and although it is recognized immunologically in the chicken, the avian liver PK does not react with glucagon or cAMP. In the clawed frog *Xenopus laevis*, no less than nine PK isoenzymes may be separated by isoelectric focusing; these have been identified as L, M_2 and M_1 homotetramers and as M_2L and M_1M_2 hybrids. The initial subunits in the embryo are M_2 and L, with M_1 appearing later; the PK of oocytes and embryos correspondingly shows sigmoidal PEP kinetics [50]. The so-called L type has a relatively high PEP affinity and low PEP cooperativity, and it is not affected by cAMP-dependent phosphorylation; thus, this PK corresponds more to mammalian PK-M_2 than to PK-L [60].

The situation in the **fish** is especially complicated. The cartilaginous fish have only the M_2 type; in the teleosts there is almost always an additional E type, with M_2E hybrid molecules occurring quite frequently in the heart. The liver and erythrocytes of many species contain a third type "L" which, however, cannot be compared directly with the PK-L of the mammals [204]. In complete contrast to the mammals, the skeletal muscle PKs of many fish have distinct allosteric characteristics, whereas the liver PK is only weakly regulated. The regulation of the muscle PK varies widely with the species. For example, the muscle PK of two out of four Amazonian fish species responds strongly to the usual allosteric effectors, whereas the others respond only weakly, and this is not related to whether the species respire in water or in air.

The tissue-specific PK isoenzymes of the fish often show only minor differences in regulatory characters. Of the two PK isoenzymes in the pink salmon *Oncorhynchus kisutch*, the predominant type found in the muscle and heart is PK-M_1, whereas the form that is more generally distributed is PK-M_2; both show allosteric properties, although PK-M_2 appears to be more easily regulated than PK-M_1. PK-M_2 shows immunological cross-reactivity not only with PK-M_1 of the same species but also with PK-M_1 and PK-L of the mammals. The muscle enzyme of the sturgeon *Acipenser fulvescens*, which shows strong allosteric effects, exhibits immunological cross-reactivity with PK-M_1 of the frog and the dog [87]. Thus, the far-reaching differences in kinetic-regulatory properties appear to be the result of relatively minor differences in sequence [91]. The red and white muscles of the American eel *Anguilla rostrata* contain different PKs but both types are regulated allosterically. It is likely that the red muscles of the migrating eel, in which the liver degenerates, have gluconeogenic functions. The liver PKs of *A. rostrata* and of the rainbow trout *Salmo gairdneri* both have no allosteric properties [206]. On the other hand, the Amazonian fish *Arapaima gigas* has different PKs in the liver and kidneys; both can be regulated and appear to allow gluconeogenesis in both organs. The liver of the flounder *Platichthys flesus* has two PKs, one of which is much more strongly inhibited by phenylalanine than the other; it is therefore specifically adapted to gluconeogenic metabolic conditions [216].

14.7.3 Pyruvate Kinases of Invertebrates

With the exception of several parasitic worms such as *Ascaris lumbricoides*, which because of the peculiarities of its carbohydrate metabolism has negligible PK activity, all invertebrates have one or more PKs. **Tissue-specific isoenzymes** have been reported, for example, in the mussels *Crassostrea* and *Mytilus*, the sea snail *Busycotypus*, the cephalopod Octopus, the crayfish *Orconectes*, and the insects *Periplaneta* and *Schistocerca* [36, 63, 192, 245]. *Busycotypus canaliculatum* has three PK isoenzymes distributed between red muscle, e.g. of the radula retractor and heart ventricle, white muscle of the foot and mantle, and non-muscle tissues such as the gill, kidney and hepatopancreas [192].

Most of the invertebrate PKs can be **allosterically modulated;** in some cases, covalent modification by **phosphorylation/dephosphorylation** has also been reported. The PKs from tissues with a high potential for gluconeogenesis are especially well regulated; examples include the hepatopancreas of various molluscs [29, 158, 215], the hepatopancreas and hypodermis of crustaceans, and the fat bodies of insects [63, 245]. Most facultative anaerobic muscles of the invertebrates also contain PKs with complex allosteric properties [158, 246]. In contrast, the PKs of predominantly

aerobic functioning muscles show little or no regulation. PKs similar to PK-M_1 of the mammals are found, for example, in the skeletal muscles of crabs, the flight muscles of insects and the mantle muscles of the cephalopods [36, 53, 63, 173, 245]. However, the PK from the muscles of the crayfish *Orconectes limosus* exists in two forms with different substrate affinities, the high-affinity form being stabilized by Fru-1,6-P_2. The whole range of PKs is found in the molluscs, from the hardly regulated PK of cephalopod muscles, which resemble most closely mammalian PK-M_1, through the M_2-like PKs to those which have levels of allosteric modulation and covalent modification just as complex as those of mammalian liver PK-L [36]. Allosterically regulated muscle PKs have been investigated in detail in, for example, the mussels *Crassostrea, Cardium, Mytilus, Scapharca* and *Venus*, and the South American purple snail *Concholepas* [29, 36, 40, 99].

The **allosteric effectors** of the invertebrate PKs are, in particular, Fru-1,6-P_2, ATP and alanine or other amino acids [173]. The frequently observed inhibitory effects of phosphagens is probably an artefact produced by crude extracts in which the phosphagen kinase that is always present uses up the PK substrate, ADP. A reduction in pH brings about the inhibition of PK and the stimulation of PEPCK in marine mussels, and thus a switch from pyruvate production to the more efficient succinate pathway of anaerobic energy production. The pH effect is much less marked in the freshwater mussel *Anodonta cygnea* [105, 224]. Bicarbonate inhibits PK and activates PEPCK, for example, in the oligochaete *Tubifex* and the cestode *Hymenolepis diminuta*. There have been reports of alanine inhibition of PK from cnidarians, bivalves and crustaceans but not from *Tubifex* or the frog.

The regulation of PK by allosteric effectors and pH is complemented in many molluscs by covalent modification. In particular, in anoxia-tolerant species, such as *Mytilus edulis*, *Patella caerulea* and *Busycotypus (Busycon) canaliculatum*, but also in the drought-resistant terrestrial snail *Otala lactea* the muscle PK is subject to phosphorylation [157, 158, 275]. A PK-specific protein kinase has been isolated from *B. canaliculatum* [19]. The substrate constant for PEP is increased by **phosphorylation**, and the inhibitor constant for ATP and L-alanine is reduced; the end result is thus a reduction in PK activity. However, the converse has been observed for the PK from adductor muscles of the bivalves *Venus gallina* and *Scapharca inaequivalis;* in this case, PEP affinity and enzyme activity are increased after in vitro phosphorylation by added cAMP-dependent protein kinase [88].

Some **pyruvate kinases of the Protozoa** have rather unusual properties. For example, the PK from *Leishmania donovani* shows sigmoidal PEP kinetics, and thus homotropic interaction, but it exhibits no allosteric effects. In contrast, Fru-1,6-P_2 and in particular Fru-2,6-P_2 reduce cooperativity and increase the affinity of PK from *Trypanosoma cruzi* and *T. brucei;* ATP counteracts this effect [25, 222]. *Entamoeba histoloytica* possesses a diphosphotransferase instead of a typical PK, a situation otherwise found only in plants. In contrast to PK, this enzyme catalyses a reversible reaction:

$$\text{PEP} + \text{AMP} + \text{PP}_i \leftrightarrows = \text{pyruvate} + \text{ATP} + \text{P}_i. \quad (14.1)$$

14.8 Further Glycolysis Enzymes

Compared with the six enzymes discussed above, the remaining enzymes of the glycolysis pathway have received much less attention from comparative biochemistry. **Phosphoglucomutase** has been isolated from mammals, teleosts and elasmobranchs, as well as from higher plants and lower fungi. All examples have a molecular mass of 62–63 kDa, and the great similarity between the amino acid compositions of the enzymes from various fish, the rabbit and the potato is indicative of a relatively slow evolution. The sequence is known only for the rabbit enzyme [200]. **Glucose phosphate isomerase** (GPI) from the trypanosomes to the mammals has very similar properties. Whereas the tetrapod vertebrates (amphibians to mammals) have only one GPI locus, the teleosts (except for the Clupeomorpha) have two isoenzymes that differ in their tissue location and catalytic properties; GPI-A predominates in muscle, and GPI-B predominates in the liver and other organs [13]. Two GPI loci are also found in the agnathans of the order Petromyzontoidea, but only one locus is found in the order Myxinoidea, the cartilaginous fish and the ancient actinoptergygians of the genera *Amia* and *Lepisosteus*. *Branchiostoma* and the ascidian *Ciona* also have only one GPI. The explanation for this distribution probably lies in independent duplications of the GPI gene in the development of the lampreys and the teleosts, and loss of one of the genes in the Clupeomorpha [38, 97].

Triose phosphate isomerase (TIM) is the most active glycolysis enzyme in the insects and vertebrates (Table 14.1). This enzyme not only introduces dihydroxyacetone phosphate to glycolytic degradation via glyceraldehyde phosphate but can also function in the opposite direction as the starting point for the formation of glycerol phosphate, which is required for lipid and glycerol metabolism as well as for the α-glycerophosphate cycle. TIM is always a homodimer with subunits of 247–249 amino acids and 27–28 kDa. Sequence comparisons point to a relatively slow rate of evolution; compared with the human sequence, the rabbit sequence shows 16 differences, the chicken sequence 38, and that of *Latimeria* 80, but the bacteria sequence shows 170 differences [145]. The TIM of the Kinetoplastida is located in the glycosomes together with eight other enzymes of glycolysis and glycerol metabolism. The enzyme of *Trypanosoma brucei* differs at 48–52 % of positions from the vertebrate enzyme but is not significantly different in its kinetic-regulatory properties [128, 159]. **Phosphoglycerate kinase** (PGK) has mainly been investigated in the mammals, although the enzyme isolated from the halibut *Hippoglossus stenolepis* has been the subject of studies of the reaction mechanism [108]. The mammals possess two isoenzymes of PGK. PGK-1, which is found in all cells, is encoded by one gene on the X chromosome; the human sequence (417 amino acids) differs from that of the horse (416 amino acid) at only 14 positions. The human and murine PGK-2, as in some other mammals, is restricted to the testis and sperm; it is encoded by an intronless gene that arose as a retroposon from a transcript of the PGK-1 gene. There is a further human PGK-1 retroposon that has become nonfunctional due to several nonsense mutations [152]. In the dog, fox and kangaroo PGK-2 is found in other tissues in addition to the testis. The kinetoplastids *Trypanosoma brucei* and *Crithidia fasciculata* have three PGK genes, one of which encodes the glycosomal PGK and a second encodes the cytosolic enzyme; the protein product of the third gene has not yet been identified. The two enzymes of *T. brucei* agree at 93 % of positions, but the glycosomal enzyme has a more positive charge and a C-terminal extension of 20 amino acids [133].

The **phosphoglycerate mutases** and **2,3-bisphosphoglycerate mutases** belong to a family of multifunctional enzymes which catalyse various reactions between phosphoglycerate (PG) and bisphosphoglycerate (BPG):

1. Mutase reaction: 3-PG + 2,3-BPG ⇆ 2,3-BPG + 2-BG.
2. Synthase reaction: 1,3-BPG + 3-PG ⇆ 3-PG + 2,3-BPG.
3. Phosphatase reaction: 1,3-BPG → 3-PG + P_i.

The phosphoglycerate mutases (PGM; EC 5.4.2.1) are enzymes with predominantly mutase activity, and the bisphosphoglycerate mutases (BPGM; EC 2.7.5.4) are those which, above all, catalyse the synthesis and cleavage of 2,3-BPG. The agnathans have only one PGM, whereas all other vertebrates, with the exception of the birds, have several homo- or heterodimeric PGMs with two types of subunit known as muscle type (M) and brain type (B). One of the PGM genes has been silenced during the evolution of the birds. Skeletal muscle always contains MM; the liver, kidney and brain contain mainly BB; and the heart contains MM, MB and BB. B always appears before M in ontogeny. The BPGM subunit E is expressed only in erythroid cells. The activity of the homodimer EE in erythrocytes leads to high concentrations of 2,3-BPG, which reduces the oxygen affinity of haemoglobin to physiological values. The human subunits B, M and E agree in about 46 % of their amino acids [30, 66, 134].

The **enolases** are dimers of non-covalently linked subunits of 47 kDa. Three different types of subunit (α, β, γ) are found in the mammals and birds; these show about 80 % sequence agreement with each other and form heterodimers. The α and β subunits are found in many different tissues, whereas γ is expressed only in neurons. In embryonal muscle, αα is formed initially but is replaced during myogenesis by ββ. Whereas in adult skeletal muscle almost all the enolase activity is due to ββ, the smooth muscle and heart muscle retain larger proportions of αα [75, 127, 151, 180]. The frog *Xenopus laevis* also has three types of enolase subunit but these show a tissue distribution different to that of the higher vertebrates [227]. The enolase from *Drosophila melanogaster*, which is the only invertebrate enolase to have been sequenced so far, is similar in sequence to the mammalian subunits α and γ [15].

14.9 Gluconeogenesis

The glucose transported to cells via the blood is important for the energy metabolism of all cells and for some, e.g. erythrocytes, cells of the mammalian brain and the retinal cells of birds, it is indispensable. Glucose is also required in the structural metabolism of all cells for the biosynthesis of glycolipids and glycoproteins. In the vertebrates, the blood-sugar level is maintained constant by the activity of the liver, usually by glycogen formation and degradation, and during periods of starvation it is maintained by gluconeogenesis from amino acids, lactate or glycerol. Gluconeogenesis is the reverse of glycolysis, minus the non-reversible reactions of PK and PFK, for which other reactions are substituted. In the vertebrate liver and kidney, during gluconeogenesis from pyruvate, lactate or alanine, the reverse of the PK reaction is achieved by the combined actions of **pyruvate carboxylase** (PC) and **phosphoenolpyruvate carboxykinase** (PEPCK), and the reverse of the PFK reaction is achieved by the hydrolytic activity of fructose-1,6-bisphosphatase (see Figs. 14.1 and 18.2). The PC is a purely mitochondrial enzyme, whereas PEPCK shows wide species-specific differences in its compartmentalization; in the rat and mouse it is almost exclusively cytoplasmic but in the other vertebrates it is predominantly mitochondrial (p. 692). Thus, whilst glycolysis is restricted to the cytoplasm, gluconeogenesis requires the interaction of two cell compartments [100]. **Serine** is important amongst the glucose-forming (glucoplastic) amino acids, and there are two metabolic pathways leading from serine to 2-phosphoglycerate and then to glucose (Fig. 14.4). The pathway via serine dehydratase, PC and PEPCK predominates in the liver of the adult rat, and the pathway via serine:pyruvate aminotransferase, glycerate dehydrogenase and glycerate kinase predominates in the liver of newborn rats. Whereas both pathways are usually active in mammalian liver, the aminotransferase is lacking in amphibian liver, and the dehydratase is absent from the liver of the rainbow trout *Salmo gairdneri* [270].

Gluconeogenesis is especially important for ruminants and other mammals with cellulose-fermenting stomach symbionts, where more than 90% of the glucose present in the diet is converted to acetate, propionate and butyrate before absorption. The importance of gluconeogenesis for eels, salmon and other fish that do not feed during long migratory periods is equally obvious.

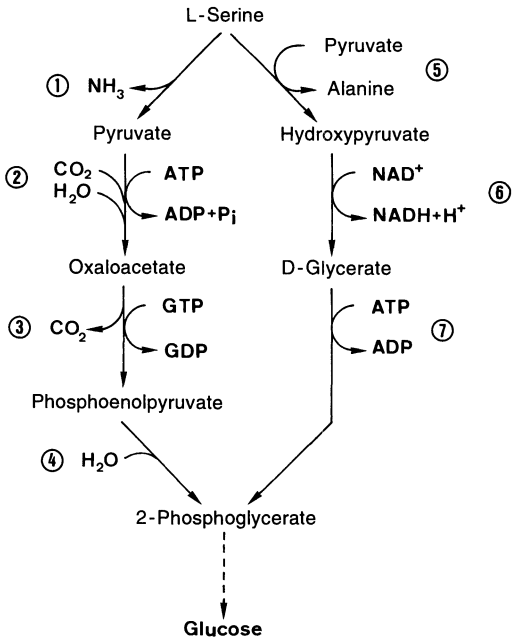


Fig. 14.4. Gluconeogenesis from serine. *1*, L-Serine dehydratase; *2*, pyruvate carboxylase; *3*, phosphoenolpyruvate carboxykinase; *4*, enolase; *5*, serine:pyruvate aminotransferase; *6*, glycerate dehydrogenase; *7*, glycerate kinase

Lactate accumulating in muscles and other tissues is also processed by gluconeogenesis, the task being divided between the muscles and the liver. The lactate arising from glycogen during anaerobic energy production in muscles is transported in the blood to the liver, where it is partly degraded oxidatively and partly reconverted to glycogen. The glucose stored as liver glycogen is transported as blood sugar to the muscle, where it is used in the resynthesis of muscle glycogen. The whole cycle is named after its discoverer, Cori [100]. Gluconeogenesis from lactate has been demonstrated by tracer experiments and enzyme determinations in the liver of all classes of vertebrates. Only in the case of the cyclostome liver, which is very small and contains negligible glycogen, is gluconeogenesis of no significance [220].

In the mammals, the kidneys and the liver are capable of gluconeogenesis. In the toad *Bufo marinus*, the stomach mucosa has the appropriate complement of enzymes and can produce glucose from pyruvate, alanine and other precursors [57]. Whereas, in general, the nervous system has a low capacity for gluconeogenesis, in the retina of the frog *Rana catesbeiana* it is unusually high [79]. The **muscles** of most vertebrates and invertebrates contain no PC and their capacity for gluconeogenesis is usually denied. However, tracer exper-

iments have demonstrated the formation of glucose from lactate in the muscles of fish, amphibians and mammals. Here, the PC reaction is replaced by a combination of malic enzyme and malate dehydrogenase (Fig. 14.1). Muscle gluconeogenesis in the vertebrates is restricted to certain types of fibre, and in fish, amphibians and reptiles it is probably only significant in the red muscles [78, 170]. PC has been detected in the muscles of some teleosts, the only case known in the vertebrates, and otherwise is found at high activities only in insect flight muscles [170].

Lactate also reaches high levels in the haemolymph of malacostracan **crustaceans** during brisk muscle activity but, in contrast to the situation in the mammals, it is removed only very slowly from the blood. For example, in the lobster *Homarus gammarus* and in the crayfish *Cherax destructor* it requires about 6 h, compared with 30 min in humans. The hepatopancreas, which as a central organ for metabolism is often compared with the liver, has no capacity for gluconeogenesis from lactate.

14.10 Anaerobiosis

14.10.1 Activity- and Habitat-Dependent Anaerobiosis

In the presence of an adequate supply of oxygen, carbohydrates are completely catabolized to CO_2 and H_2O with the production of energy, the theoretical yield being 38 ATPs per glucose residue of glycogen. However, in their natural habitats many animals experience a transient or long-term deficiency of oxygen (hypoxia), or the complete absence of oxygen (anoxia), and must therefore be capable of anaerobic energy production (habitat-dependent anaerobiosis). An anaerobic energy-yielding process is also necessary when the energy requirement of the muscles during active locomotory effort exceeds the capacity of the aerobic process (activity-dependent anaerobiosis). The various anaerobic mechanisms for obtaining energy in the animal kingdom are interesting examples of biochemical adaptation. They have, therefore, been widely investigated and extensively reviewed [21, 83, 102].

Activity-dependent anaerobiosis occurs only as a consequence of sudden, transient enhanced muscle activity, e.g. in fleeing from predators or catching prey. In contrast, muscles that operate continuously at high activity, e.g. the flight muscles of insects or birds, are dependent upon aerobic energy production. Activity-dependent anaerobiosis is widely found in the animal kingdom. A typical example is the white muscle of vertebrates, especially of amphibians and fish, and also of some invertebrates. The cephalopods achieve high swim-velocities by the rapid ejection of water from the mantle cavity by the contraction of mantle muscles; however, the musculature is usually exhausted after 10–20 such contractions. The scallops (Pectinoidea) can escape attack from starfish by swimming; they clap their shells together by contracting the shell-closing muscles (adductors), thereby propelling themselves. In this case, the muscles are also exhausted after 30–50 contractions. The cockle *Cardium tuberculatum* executes jumping movements to escape starfish. The closely related species *C. edule*, like other mussels, only uses the foot musculature to bury itself. This occurs particularly rapidly in the case of the razor shell *Ensis directus*. The common whelk *Buccinum undatum* escapes from predators by executing rotating movements brought about by the columellar muscle. The worms *Sipunculus nudus* and *Arenicola marina* bury themselves rapidly in the mud; in experiments, they perform this action several times until the muscles of the body wall become exhausted. In the characteristic defence reaction of the crayfish, the posterior body section is thrust forward by contraction of the abdominal muscles, and the animal springs backwards. Several types of insect muscle, e.g. the jumping muscles of the locust, also show activity-dependent anaerobiosis. In most vertebrates, anaerobic metabolism in muscles is only necessary at the commencement of locomotory activity, and continues until the oxygen supply to the muscle is adequately reinforced. However, in many terrestrial and all aquatic amphibian species the energy requirements for locomotion are always predominantly derived from anaerobic processes [161, 198].

At the onset of sudden muscle activity, the rate of ATP cleavage increases manyfold without a concomitant increase in the oxygen supply. **Rephosphorylation of ADP** at first makes use of the energy-rich phosphate stored in phosphagen. Only when this is exhausted does the muscle revert to degradation of glycogen to cover its energy requirements. This occurs anaerobically only to the extent to which the aerobic process is insufficient. For example, in the swimming scallop *Placopecten magellanicus* glycogen catabolism is about 3 % aerobic, but because of its far greater effectiveness this covers about one-third

of the energy requirement [68]. The maximal rate of ATP production in muscle is reached during utilization of the energy-rich phosphates of ATP and phosphagen; the second highest rate is achieved during anaerobic glycogen catabolism; and the lowest rate occurs during the aerobic process. However, the reserves of energy-rich phosphate are used up in a relatively short time, and the anaerobic supply of energy, because of its lower effectiveness, leads more rapidly to exhaustion than does the aerobic process (Table 14.4).

Extreme examples of **habitat-dependent anaerobiosis** are supplied by the endoparasitic nematodes, trematodes, cestodes and ancanthocephalans that infest the intestines and bile ducts of vertebrates; their specific adaptation to this extreme environment caught the attention of biochemists quite early on. It has become obvious only relatively recently that many free-living animals also possess highly specialized mechanisms for anaerobic energy production. This is true, for example, for the free-living nematode *Panagrellus redivivus*, the African swamp oligochaete *Alma emini*, and the oligochaetes and insect larvae from the low-oxygen mud zone of stagnant water. Goldfish can survive for several days at low temperature without oxygen. Sea anemones, mussels, snails, crabs and worms of the tidal zones may experience hypoxia when, at low tide, the water level sinks below the habitat of sessile species, or water exchange in the tube inhabited by burying species is no longer possible. However, it must not be assumed that all species in the tidal zone experience hypoxia during the ebb. In fact, oxygen uptake by the limpet *Patella vulgata* does not change at all when the tide is out. With its shell open, the cockle *Cardium edule* can take up from the air 28–78 % of the oxygen that it normally obtains from water; the edible mussel *Mytilus edulis* manages only 4–15 % [18].

Table 14.4. Energy-yielding substrates in human skeletal muscle [102]: concentration, effective energy and the maximal rate of energy production

	Concentration (μmol/g fr. wt.)	Effective energy (μmol ATP/g fr. wt.)	Rate (μmol ATP/ g fr. wt. per min)
ATP	25	10	96–360
Creatine phosphate	75	60	
Glycogen	370	aerob. 14 200 anaerob. 1120	30 60
Triacyl glycerol	50	24 520	20.4

The species *Mytilus galloprovincialis*, living in the sublittoral zone of the Mediterranean coast, never experiences hypoxia but is, nevertheless, able to obtain energy anaerobically, although less than *M. edulis* obtains [269]. This example shows that the capacity for habitat-dependent anaerobiosis is primarily an adaptation to hypoxia, but is also found in animals that never experience oxygen deficiency in their natural habitats. This is therefore a biochemical character, selected for during evolution, which has been retained despite the disappearance of the selective pressure. There are, of course, further examples of transient or persistent oxygen shortage. Air-breathing animals have, in principle, no problem with oxygen supply, but species living underground, e.g. the earthworm, may be subjected to hypoxia after heavy rain. Most cells of the insects are directly supplied with oxygen by terminal branches of the trachea, independent of the blood supply, and therefore have a purely aerobic metabolism. A notable exception are the haemocytes that are suspended in the low-oxygen haemolymph; these are completely adapted to hypoxia and can survive in tissue culture without oxygen, where they produce succinate, acetate and alanine, the typical end products of highly specialized anaerobic metabolism [129].

Very few animal species experience simultaneously **activity-dependent and habitat-dependent anaerobiosis;** the few that do include marine annelids *(Arenicola marina)*, sipunculids *(Sipunculus nudus)* and bivalves *(Ensis directus)*. For the other marine mussels, there is either activity-dependent anaerobiosis following fleeing reactions (as seen in the pectinids, *Lima hians* and *Cardium tuberculatum)* or habitat-dependent anaerobiosis (e.g. in *Mytilus edulis* and freshwater mussels) [288]. It is, of course, possible to generate experimentally hypoxia or anoxia with such species, i.e. to produce conditions that they would not normally encounter in their natural environments. All the results indicate that the two forms of anaerobiosis are fundamentally different processes. In the activity-dependent process, a transient high energy demand is satisfied by anaerobic glycolysis, which is uneconomic but rapidly available. In contrast, the habitat-dependent process persists for hours or days, rather than for minutes; it may even continue for the lifetime of the animal. This requires the highest possible economy in both the utilization and production of ATP [83]. In fact, it is observed

in many molluscs that the **intensity of energy-yielding metabolism** is drastically reduced during habitat-dependent anaerobiosis; for example, in *Mytilus edulis* it is reduced to 5 % and in isolated adductor muscles of various species it is reduced to 12.5–20 % [68, 83, 155, 289]. Whilst there is a 10-fold increase in energy turnover during activity-dependent anaerobiosis in *Cardium tuberculatum*, this is reduced 20-fold during the habitat-dependent process [156]; the contributions of the two processes to energy-yielding metabolism in the same tissue may differ by as much as 500:1 [68]. The goldfish also reduces its energy turnover to 20 % during persistent anoxia [102]. The mechanisms regulating the reduction in metabolism in hypoxic and anoxic conditions are not completely understood. The crustaceans appear to have a much less efficient regulatory mechanism; for example, ATP production in the isopod *Cirolana borealis* is reduced by only 25 % during experimental anoxia. The basic difference between activity- and habitat-dependent anaerobiosis also shows itself in the metabolic pathways and end products involved.

14.10.2 End Products of Anaerobic Metabolism

The end product of anaerobic glycolysis in vertebrate muscle is L-lactic acid, which Berzelius first discovered in the meat of hunted deer. Lactate is the main product of **activity-dependent anaerobiosis** in the vertebrates and the arthropods, but is also produced in the body-wall muscles of the leech *Hirudo medicinalis* and other annelids, and following experimental irritation of the foot muscle of the cockle *Cardium edule* [70, 71, 83, 155, 288]. L-Lactate or D-lactate is produced according to the stereo-specificity of the lactate dehydrogenase involved (see Table 14.6, p. 542). The reduction of pyruvate to lactate by lactate dehydrogenase results in the reoxidation of NADH from the GAPDH reaction, and restoration of the redox balance of glycolysis; three ATPs are produced per glucose residue of glycogen. Invertebrates, however, possess other enzymes that reduce pyruvate and thereby reoxidize NADH. These catalyze the reductive condensation of pyruvate with arginine, glycine, alanine, taurine or β-alanine to so-called opines, and they are, therefore, known as opine dehydrogenases [70, 84, 140]. To date, five opines are known together with their corresponding dehydrogenases (Fig. 14.5; Table 14.5). The main product of activity-dependent anaerobiosis from pyruvate and arginine is octopine; this was the first-discovered opine, detected in 1927 in the cephalopod *Octopus* sp. In addition to its occurrence in the cephalopods *(Loligo, Octopus, Nautilus)*, it is formed by various mussels (Pectinidae, *Cardium tuberculatum, Solen marginatus, Ensis* sp.), the common whelk *Buccinum undatum*, and the sipunculid *Sipunculus nudus*. Some scallops (Pectinidae) produce octopine during swimming, whilst others *(Chlamys opercularis)* produce it only during the recovery period. In the lugworm *Arenicola marina*, the end product of activity-dependent anaerobiosis is alanopine. The prosobranch snails of the genus *Strombus* produce mainly octopine during muscle activity, and then strombine during the recovery period [68, 70, 71, 84, 102, 140, 288]. Activity-dependent anaerobiosis and a general oxygen deficiency lead to the production of tauropine in the adductor muscle of the sea snail *Haliotis lamellosa*, whereas the anoxic foot muscle produces D-lactate [72].

Table 14.5. Activities (U/g fr. wt.) in various animals of D- and L-lactate dehydrogenase (LDH), strombine dehydrogenase (StrDH), alanopine dehydrogenase (AloDH) and octopine dehydrogenase (ODH) [84]

	LDH	StrDH	AloDH	ODH
Porifera				
Halichondria panicae (total)	0	0.1	0	0
Nemertina				
Cerebratulus lacteus (proximal end)	0.5	0	3.0	50
Mollusca				
Littorina littorea (foot)	14.2	0	2.4	0
Nassa mutabilis (foot)	10.7	2.4	29	186
Buccinum undatum (various muscles)	3.1	0	9.3	69
Mytilus edulis (adductor)	13.3	4.9	2.7	17.5
Cardium edule (adductor)	9.0	9.7	7.5	30.6
Cardium edule (foot)	11.2	3.3	2.5	15.6
Cardium tuberculatum (foot)	4.0	18.1	22	117
Lima hians (adductor)	<0.1	4.1	4.6	60
Ensis siliqua (foot)	0.9	6.1	13.6	167
Sepia officinalis (mantle)	0.3	0	0	97
Annelida, Sipunculida				
Nephthys hombergi (body wall)	0	44	50	0
Nereis diversicolor (total)	49	0	0	0
Arenicola marina (body wall)	0.2	4.0	45	0
Glycera convoluta (total)	–	52	171	–
Sipunculus nudus (body wall)	<0.05	9.2	49	445

–, not determined

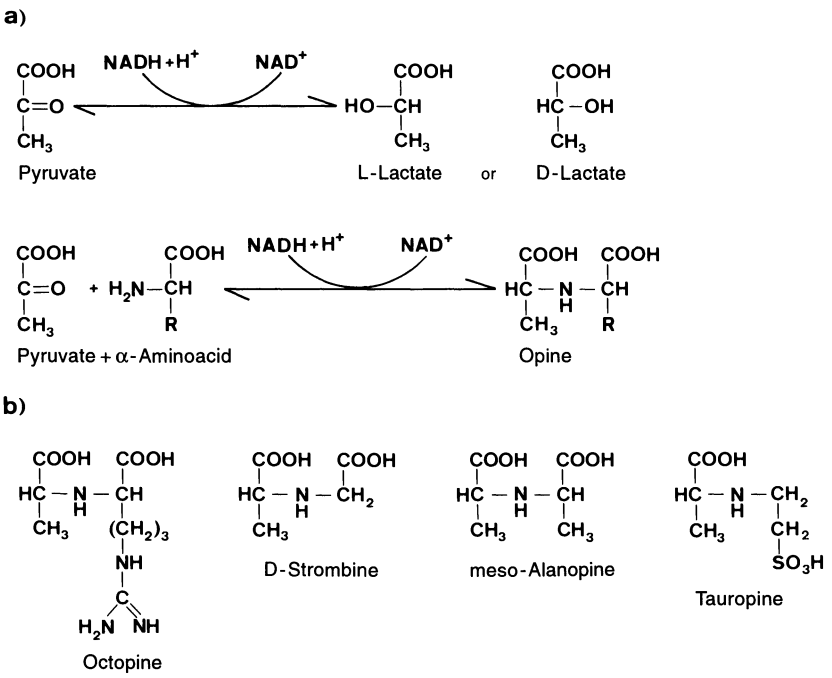


Fig. 14.5a, b. Opines. **a** The reactions of lactate and opine production. **b** The structural formulae of the most important opines. A further opine, β-alanopine, has since been discovered and this is synthesized from β-alanine and pyruvate [263]

It is not quite clear why many invertebrates produce opines instead of lactate during anaerobic metabolism in the muscles; the ATP yield of anaerobic glycolysis is not increased in any way (three ATPs per glucose residue). Octopine formation binds the arginine released by cleavage of the phosphagen arginine phosphate, and octopines disturb the enzyme functions of cells much less than does arginine. However, in *Sipunculus nudus* octopine production during activity-dependent anaerobiosis cannot compensate completely for the increasing free arginine concentration in the muscle [83]. One important advantage of the synthesis of octopine and other opines is that lower values of the redox coefficient NADH/NAD^+ are attained at the reaction equilibrium of pyruvate reduction [56, 84].

During the **recovery** period after muscle activity, the end products of anaerobic glycolysis must be either further metabolized oxidatively or used in gluconeogenesis. These processes can take place in the same muscle in which the end products are formed. This has been shown for the lactate of vertebrates and crustaceans and for the opines of the snails and mussels. On the other hand, there may also be a division of labour between organs, in which the end products formed in the muscle are transported in the blood to other organs where they are then further metabolized; at least part of the sugar produced by gluconeogenesis is then returned to the muscle via the bloodstream. In such cases, organ-specific dehydrogenases are found; these are adapted to the relative activities of the forward and reverse reactions. The best-known example of such a division of labour between organs is the oxidation of lactate in vertebrate heart, lung and brain and its utilization for gluconeogenesis in liver and kidney. An analogous situation is found in the cephalopods with octopine, which is released from the muscle into the haemolymph and is further metabolized in the heart, brain and gills [52, 68, 70]. However, the proportion of anaerobic end products transported via the blood to other organs is only 20–45 % in the rat and 15–20 % in the cephalopod *Sepia* [102].

In many animals, energy extraction during **habitat-dependent anaerobiosis** also occurs via the Embden-Meyerhof pathway; pyruvate is reduced to balance the redox coefficient and three ATPs are formed per glycogen glucose. The end products are D- or L-lactate or opines, and in some cases also ethanol (Fig. 14.6). Lactate as an end product of habitat-dependent anaerobiosis is

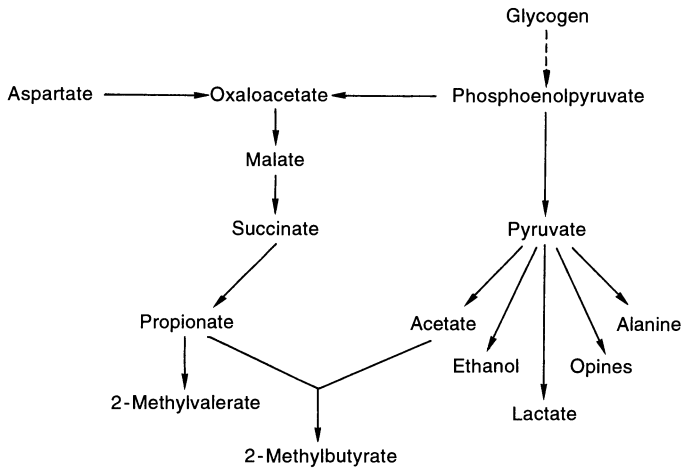


Fig. 14.6. The end products of anaerobic metabolism

found, for example, in all terrestrial and freshwater gastropods, in several marine snails *(Littorina littorea)* and mussels *(Cardium tuberculatum)*, in all crustaceans, in *Limulus* and in spiders, in some annelids *(Hirudo medicinalis)*, and in fish [37, 68, 70, 84, 168, 288]. Octopine is the end product of both anaerobic processes in the cuttlefish *Sepia officinalis*, but otherwise does not play an important role in habitat-dependent anaerobiosis [70, 71]. Strombine, often together with small amounts of alanopine, is found at least during the initial phases of habitat-dependent anaerobiosis in the sponge *Halichondria panicea*, in *Mytilus edulis* and other mussels *(Crassostrea, Nucula, Mercenaria, Ensis, Solen)*, in *Arenicola marina* and other polychaetes *(Nephthys, Pectinaria)*, and in *Sipunculus nudus*. Alanopine is the predominant product in several molluscs *(Littorina littorea, Nucella lapillus)* and polychaetes of the genus *Glycera* [68, 70, 84, 231, 288, 289]. Investigations of 25 marine invertebrates have revealed the reason why many species produce different opines during activity- and habitat-dependent anaerobiosis. During the habitat-dependent process, all opine dehydrogenases function close to the reaction equilibrium, and the relative amounts of the different end products are determined by the concentrations of the corresponding amino acids. During the activity-dependent process, only those opine dehydrogenases with activities high enough to keep pace with the increased glycolysis flux attain the reaction equilibrium [173]. Ethanol as the main end product of habitat-dependent anaerobiosis is found in the common carp and goldfish, the bitterling *Rhodeus amarus*, larvae of the nonbiting midge *Chironomus thummi*, the free-living nematode *Panagrellus redivivus* and certain Protozoa [24, 177, 201, 276].

In many animals, particular metabolic pathways invoked during habitat-dependent anaerobiosis have a significantly better ATP yield (seven ATPs per glucose) than does anaerobic glycolysis. The main end products here are succinate, propionate and/or acetate, and in free-living species almost always alanine (Fig. 14.6). These processes have been demonstrated and investigated in nematodes, trematodes, cestodes, acanthocephalans, annelids, echiurids, sipunculids, lamellibranchs and gastropods [71, 103, 112, 158, 197, 225, 288]. Branched fatty acids are also produced in the nematodes *Ascaris lumbricoides* and *Parascaris equorum*, the trematode *Paragonimus westermanni* and the oligochaete *Alma emini* [205, 221]. Most of the results have been obtained under experimental conditions of anoxia in the laboratory, but data from free-living *Mytilus edulis* and the polychaetes *Arenicola marina* and *Nereis diversicolor* are essentially in agreement [225, 289].

The **main substrate of energy-yielding metabolism** in habitat-dependent anaerobiosis is glycogen (Fig. 14.6). Thus, animals adapted to a chronic or temporary oxygen deficiency usually have large reserves of glycogen. Glycogen as a proportion of dry weight is usually 10–20 % in parasitic worms, over 40 % in some cestode larvae, and up to 50 % in the hepatopancreas of the edible mussel *Mytilus edulis*. Anoxia-tolerant vertebrates, like the goldfish and turtles, have glycogen reserves in the liver of about 16 % of the fresh weight (1000 μmol glucose equivalents per g fresh weight); in the trout or in mammals it is only about 3 % (200 μmol per g fresh weight) [102]. Especially during the initial phases of habitat-dependent anaerobiosis, aspartate may be converted to succinate to yield ATP. This metabolic

route was first discovered in 1977 in the heart of the oyster *Crassostrea gigas* and is, in fact, of relatively great importance for the mussel heart; it occurs in other tissues of the molluscs but not at all in the arthropods [68, 199]. Succinate production from aspartate has also been detected during hypoxia in heart and skeletal muscle of mammals and fish. Malate, which is an intermediate of succinate synthesis from glycogen or aspartate, is found at such high concentrations in the tissues of several oligochaetes, leeches and the larvae of non-biting midges that it is a significant substrate at least during the initial phase of habitat-dependent anaerobiosis; the concentration is 14–20 μmol/g in *Eisenia foetida*, 30–40 μmol/g in *Hirudo medicinalis*, about 50 μmol/g in the tubificid *Lumbriculus variegatus*, and 40-50 μmol/g in *Chaoborus flavicans* [288].

Activity-dependent anaerobiosis and (natural or experimental) anoxia can lead to different **end products** in the same animal, or even in the same muscle tissue. The products of the activity-dependent process are mainly those of pyruvate reduction, such as lactate or opine, whereas in the habitat-dependent process succinate, volatile fatty acids and alanine are also produced. There are some well-investigated examples: The mussel *Ensis directus* synthesizes octopine during rapid digging movements but produces mainly strombine, together with succinate, propionate, acetate and alanine, under anoxic conditions. *Spinunculus nudus* also produces octopine during activity-dependent anaerobiosis but synthesizes the products of the succinate pathway together with strombine during the habitat-dependent process. The burying action of *Arenicola marinus* results in alanopine but the initial phases of the habitat-dependent process produce strombine. *Cardium tuberculatum* accumulates octopine in its foot during fleeing jumps, but produces lactate under conditions of experimental anoxia [71, 84, 156, 225, 231, 288]. It is of particular interest that different pyruvate reductases are quite often involved in the two types of anaerobiosis (p. 537).

The accumulation of the acidic end products of anaerobic glycogenolysis leads to alterations in the **proton balance** of the cells and of the whole animal. One proton arises per carboxyl group produced, i.e. 1 mol H^+ per mol end product, irrespective of its chemical nature (lactate, opine, alanine, acetate, propionate); succinate synthesis yields two protons per mole. This has led to the hypothesis that the proton stoichiometry of anaerobiosis is independent of the prevailing metabolic pathway [102]. In fact, the situation is complicated by further metabolic processes, and the proton balance of anaerobiosis may vary greatly depending upon the substrate consumed and the end product formed. For example, the number of protons produced in the succinate-propionate pathway is lower when aspartate is the substrate than when glycogen is used. The cleavage of phosphagen consumes protons, whereas dephosphorylation of $MgATP^{2-}$ generates protons; furthermore, the P_i released from phosphagen or ATP acts as an intracellular buffer. Analysis of the proton balance of a tissue or of a whole organism must also take into account the excretion of H^+ and NH^+_4 ions [195]. Anoxia occurring in the presence of enhanced CO_2 in the environment (hypercapnia) leads to increased acidosis [116]. The degree of acidosis is reduced in some animals by excretion of part of the end products. In this case, the possibilities for exploiting fully the remaining energy content of the end products by oxidation to CO_2 and water, or for using them in gluconeogenesis, are forfeited after the return to normal conditions. Excretion of lactate is found, for example, in the crustaceans and also occurs to a limited extent in the tadpoles of *Rana catesbeiana* [198]. Many molluscs and annelids secrete propionate, acetate and sometimes succinate [103, 105, 187, 225]. Another way of reducing anaerobic acidosis, namely the further metabolism of the acidic end products to ethanol, has been adopted only by the larvae of the non-biting midge *Chironomus thummi*, the free-living nematode *Panagrellus redivivus*, and the goldfish. Ethanol is particularly easily disposed of by secretion into the aqueous environment [24, 201, 276].

In many free-living animal species several **phases of anaerobiosis** may be distinguished during persistent anoxia. This is to be expected because substrates such as aspartate are gradually used up during the time needed for regulation of the switch from aerobic to anaerobic energy production, and because the accumulation of various metabolites alters the conditions for further cell metabolism. In the initial phase, phosphagen is consumed and aspartate is converted to succinate; during the subsequent reactions, propionate (and sometimes also acetate) is produced as the main product [83, 225, 288]. These changes with time have been especially well investigated in *Arenicola marina*. During the first 3 h, the phosphagen phosphotaurocyamine is used up, aspartate is metabolized to succinate and volatile fatty acids, and pyruvate is metabolized to strombine and L-alanine; the latter is racemized to a mixture of L- and D-alanine. After 3 h, glycogen catabol-

ism at the PEP branch point gradually switches from PK to PEPCK and the metabolic rate decreases. The degree of phosphorylation of adenosine (energy charge) declines drastically during the first 9 h and then becomes stable. It would appear that a new stable metabolic state is reached only with some difficulty. From 12 h, the succinate and alanine concentrations in the animal show no further increase; only propionate and acetate are produced and these are mainly excreted into the surroundings. A very similar metabolic process is also observed in the errant polychaetes *Nephthys hombergii* and *Anaitides mucosa*, and in the sedentary forms *Scoloplos armiger* and *Scololepis foliosa* [225, 288].

Different metabolic reactions are found in the various species of the polychaete genus *Nereis;* during experimental anoxia, these produce D-lactate as well as succinate, propionate, acetate and alanine, but no strombine. *Nereis diversicolor*, which as an inhabitant of the higher eulittoral may also experience hypoxic conditions in nature and is particularly anoxia resistant, on balance produce more succinate and propionate than lactate during long-lasting anaerobiosis. In contrast, lactate predominates in *N. pelagica* from the sublittoral, and this species tolerates only short periods of anoxia. *N. virens*, which extends from the sublittoral to the lower eulittoral, exhibits intermediate behaviour [288]. Apart from the changes in metabolic pathways and end products with time during long-lasting habitat-dependent anaerobiosis in free-living animals, there are other complications that render more difficult the scientific analysis of anaerobic energy metabolism. Distinct seasonal differences have been observed in the edible mussel *Mytilus edulis*, which in nature is exposed to air during low tides. There are, of course, also organ-specific differences in enzyme complement and the end products of anaerobic metabolism. For example, the body-wall muscles, the nervous system and the scales of the polychaete *Aphrodite aculeata* contain an alanopine dehydrogenase of broad specificity, whereas the gut has a highly specific strombine dehydrogenase [242]. In *Cardium edule*, more L-lactate than succinate is produced in the muscles during habitat-dependent anaerobiosis, and the reverse is true for the gills [155]. Organ-specific differences in anaerobic end products have also been reported in *Mytilus edulis*.

The far-reaching changes in energy-supplying cell metabolism during habitat- and activity-dependent anaerobiosis point to complex regulatory processes and interactions which are as yet only poorly understood. The targets of the **regulatory mechanisms** are probably mainly PFK, hexokinase and PK, perhaps also PEPCK, and in some animals lactate dehydrogenase. The most important regulatory factors are the relative concentrations of ATP, ADP and AMP (and also the energy charge) as well as arginine phosphate [53, 68, 71, 84]. Octopine is thought to have important regulatory functions in the cephalopod *Sepia*, as has pH in the crustaceans and bivalves [53, 224]. The activities of the various dehydrogenases depend very much, of course, on the NADH:NAD^+ ratio [71]. Information about regulatory mechanisms is provided by the relationships between the kinetic-regulatory properties of the enzymes and changes in the metabolic rate and metabolite concentrations during anaerobiosis; such data are available for various bivalves [68, 70, 71, 105, 155, 289], gastropods [51, 68, 190], cephalopods, annelids [103, 225, 288] and crustaceans [71]. The mechanism contributing to the reduction in metabolism during habitat-dependent anaerobiosis is still something of a mystery.

14.11 Pyruvate Reductases

The NADH produced by the GAPDH reaction can be reoxidized in various reactions with pyruvate as the hydrogen acceptor (Fig. 14.5a). There are several types of pyruvate reductase in the animal kingdom whose phylogenetic relationships have not yet been defined. The lactate dehydrogenases (LDHs) catalyse the reversible reduction of pyruvate to lactate with NADH as the coenzyme. The LDHs are more-or-less ubiquitous; they are absent from, for example, some protozoans of the genera *Trypanosoma* and *Leishmania*. They are especially active in muscles with a marked capacity for activity-dependent anaerobiosis. The highest recorded LDH activity of 5000 U/g at 25 °C was for the white musculature of the tuna fish *Euthynnus pelamis*, which is one of the fastest swimmers in the animal kingdom [102]. Animals have either L- or D-specific LDH according to their systematic positions (Table 14.6). The opine dehydrogenases catalyse the reductive condensation of pyruvate or other keto acids with the α-amino group of an amino acid, using NADH as the coenzyme. Such enzymes have been detected in various groups of marine invertebrates.

Table 14.6. The distribution of L- and D-specific lactate dehydrogenases in animals [9, 67, 144, 224]

L-specific LDH	D-specific LDH
Cnidaria	Rhizopoda
Nematoda	Acanthocephala
Echinodermata	
Trematoda	
Cestoda	Gastropoda
Cephalopoda	Bivalvia
Polychaeta e.g. *(Arenicola marina)*	Polychaeta e.g. *(Nereis virens)*
Clitellata	
Crustacea Malacostraca	Crustacea (excl. Malacostraca)
Insecta (inc. Tracheata)	Chelicerata
Tunicata	
Vertrebrata	

14.11.1 Lactate Dehydrogenases of Vertebrates

The vertebrate LDHs are tetramers of about 140 kDa which appear as **tissue-specific** isoenzymes with different catalytic properties. In the mammals (whose LDHs are a well-known model for isoenzymes) there are always five isoenzymes that exist in organ-specific proportions, designated in order of decreasing anodic mobility as LDH-1 to -5. These arise by all possible combinations to homo- and heterotetramers of two different **subunits** known as A and B or, according to the tissue showing the highest expression, as M (muscle) and H (heart) (Fig. 14.7). Subunits A and B vary characteristically throughout the whole of the vertebrates. In all mammals, almost all birds and most reptiles and amphibians B is more negatively charged than A. In the fish, there is a much greater variation in net charge, even to the point that A is more negative than B [38]. In the mammals and birds, A and B can associate freely to form heterotetramers; in many fish, and probably also in amphibians and reptiles, the possibilities for isoenzyme formation by association are restricted. The absence of heterotetramers may be due to synthesis of the subunits in different cells or at different times, or to properties of the polypeptides that prevent association. Only in the former case is association possible in vitro. Most elasmobranchs, as well as several teleosts from the orders Cypriniformes, Siluriformes and Clupeiformes, possess the complete set of five isoenzymes, but most fish have only the two homotetramers A_4 and B_4 or three isoenzymes including the heterotetramer A_2B_2 [38, 61, 126, 149]. All orders of the lampreys (Petromyzontoidea) have a single LDH subunit; in the more primitive common hagfish *Myxine glutinosa* and *Eptatretus cirrhatus* both subunits are present but do not form heterotetramers [8]. Various salamanders of the genus *Triturus* have only A_4 and B_4 [149]. In the Pleuronectiformes, Tetraodontiformes and the family Cottidae from the order Perciformes, almost all tissues have only A_4; B is suppressed apart from traces in the eye and brain [149]. In contrast, A is partially repressed in *Gambusia affinis* and *Oryzias latipes*.

The two subunits and the derived homotetramers show characteristic differences in their **enzymatic properties**. The predominant LDH-5 of skeletal muscle (A_4) has the activity of a pyruvate reductase, and the LDH-1 (B_4) of the heart is a lactate oxidase. A_4 has a lower affinity for pyruvate but a higher turnover than B_4; the latter is much more strongly inhibited by pyruvate than is A_4 due to the formation of an abortive complex enzyme-NAD^+-pyruvate. This clear difference between the two subunits and their organ-specific distribution may be interpreted as functional adaptations to the specific requirements of different tissues. However, comparative observations that reveal many irregularities in the organ distribution of the two LDH types throw some doubts

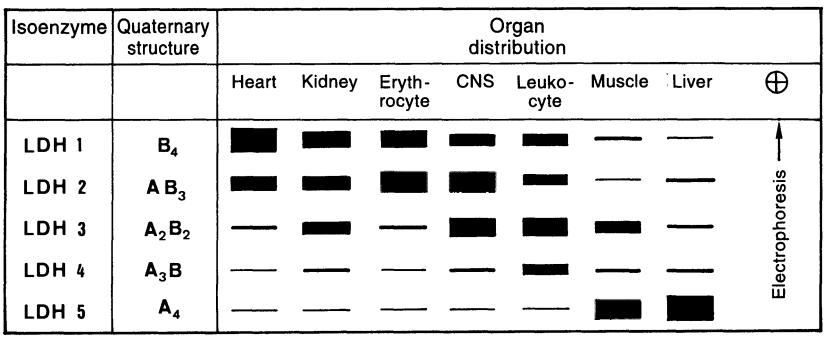


Fig. 14.7. Isoenzymes of mammalian lactate dehydrogenase

on this interpretation. It is in many cases not possible to explain why one tissue preferentially contains A and another tissue contains B. For example, the ratio of A:B in mammalian liver varies with the species from 10:90 to 90:10. B predominates in bovine eye lens but A is the predominant type in both rat and rabbit lenses. Embryonal mammalian tissues contain almost exclusively A_4, although lactate is preferentially oxidized [102].

Like other enzymes, the LDHs may also be adapted to particular internal or external conditions. The homotetrameric LDHS A_4 and B_4 of the common hagfish *Myxine glutinosa* and *Eptatretus cirrhatus* have almost identical kinetic properties, corresponding to those of A_4 from the skeletal muscles of other vertebrates. The isoenzyme B_4, present in the heart, is apparently specifically adapted to the necessity for anaerobic glycolysis in the heart muscle [8]. In turtles of the genus *Pseudemys*, the B_4 of the heart has all the characteristics of muscle type A_4; this may be interpreted as an adaptation to the diving habit of this animal. But, despite similar behaviour, the species *Caretta caretta* has a typical B_4 in its heart muscle [5]. In marine elasmobranchs that retain urea in order to regulate their osmolality, A_4 shows only the low pyruvate affinity and inhibition typical of the muscle enzyme in the presence of physiological urea concentrations of about 400 µmol/l; the LDHs from the freshwater ray *Potamotrygon* sp. or marine teleosts do not have this property. The muscle LDH of teleosts is very suitable for investigating adaptation to a high hydrostatic pressure. Compared with the muscle enzyme of the coastal species *Sebastodes alascanus*, the A_4 of the deep-sea species *S. altivelis* has a much lower pressure sensitivity for substrate and cosubstrate binding. This is most likely due to the fact that the LDH of the coastal species has a histidine at position 115, the ionization of which is pressure sensitive, whereas the LDH of the deep-sea species has an asparagine in this position [230].

The primary spermatocytes of mammals and birds contain LDH-X, which is made up tetramerically from a third type (C) of LDH subunit. C_4 makes up more than 80 % of LDH in mammalian sperm; it is present in much lower amounts in avian sperm, and is completely absent from turkey sperm. The sperms of the monotrome mammals also lack C_4 [211, 278]. **LDH-C_4** differs in many specific ways from A_4 and B_4; it is localized in the mitochondria and involved in the transport of reduction equivalents through the mitochondrial membrane. It can also process α-keto acids or α-hydroxy acids with more than three C atoms. A similar broad specificity is found in only one B_4, that of the lizard *Gecko gecko*. C_4 shows large species-specific differences in its catalytic properties: bovine, guinea-pig and rabbit C_4 are only active with α-ketobutyrate; the human, pig, goat and pigeon enzyme also metabolizes acids with five or six C atoms, but only with a low affinity; the C_4 from the rat and mouse has a high affinity for C_5 and C_6 acids and also a high affinity for branched α-keto acids. The pyruvate inhibition of C_4 of the rat, mouse and pigeon exceeds that of B_4, is about the same in humans, guinea-pigs and rabbits, and is absent in the cow, goat and pig [39].

An LDH-C subunit is found in addition to A and B in all Actinopterygii, i.e. the Chondrostei, Polypteri, Holostei and Teleostei, but is absent from the agnathans, elasmobranchs, dipnoans and crossopterygians. It appears also that the C subunit is lacking in the reptiles and amphibians [149]. The C subunit is present in many different tissues of the ancient Actinopterygii, i.e. the Chondrostei, Polypteri and Holostei, and in primitive teleosts such as the Elopiformes, Anguilliformes and Osteoglossiformes; in the remaining teleosts it is found only in neuronal tissues such as the retina and brain and less often in the liver (Table 14.7) [104, 118, 149, 202]. The C_4 from cod liver, like that of mammalian sperm, is also active with mono- and bicarbonic acids with five C atoms [202]. Fish liver C_4 always has a net positive charge and lies cathodic of A_4 and B_4. The retina–brain enzyme is usually the most negatively charged but may also be found between A_4 and B_4 or even cathodic to them [38, 149]. Fish C can form heterotetramers with A and B; the frequently observed restriction of A and B hybridization in fish is sometimes loosened in the presence of the C subunit. Thus, the eye and mesencephalon of *Poecilia latipinna* and *P. velifera* contain CA, CB and CAB heterotetramers in addition to A_4, B_4 and C_4, but they contain no AB hybrids.

The **amino acid sequences** are known for a series of vertebrate LDHs: *Squalus acanthias* A, chicken A and B, and all three subunits of various mammals. These sequences have identical amino acids at 60–90 % of their 330 positions, are thus clearly homologous, and probably arose via duplications of the same ancestral gene. The greater similarity between subunits of the same type in different species than between different subunits in the same species indicates that the isoenzymes arose very early in the evolution of

Table 14.7. The localization of type-C LDH subunits in the bony fish [149]. Only those organs that contain medium or high activities of LDH-C are listed. Where no species is given, the results were similar in all species examined

LDH-C in various organs		
Chondrostei	Acipenseriformes	
	Scaphirhynchus albus	Liver > gill
	Polyodon spathula	Kidney > spleen
Holostei	Amiiformes	
	Amia calva	Brain > eye
Teleostei	Elopiformes	
	Albula vulpes	Kidney > spleen
	Anguilliformes	Kidney, spleen, gill
	Osteoglossiformes	
	Pantodon buchholzi	Kidney > heart, stomach
	Osteoglossum bicirrhosum	Brain, eye
	Xenomystus nigri	Liver
LDH-C in the eye and brain		
Teleostei	Clupeiformes	
	Salmoniformes	
	Myctophiformes	
	Percopsiformes	
	Atheriniformes	
	Beryciformes	
	Perciformes	
	Pleuronectiformes	
	Tetraodontiformes	
LDH-C in the liver		
Teleostei	Cypriniformes	
	Gadiformes	

the vertebrates. Contrary to earlier conclusions, the agreement between A and B is greater than between either of them and C, and it appears, therefore, that C separated earlier than A and B [41, 101]. In agreement with this conclusion is the fact that the only LDH of the lampreys (Petromyzontoidea) and the B_4 from the heart of the common hagfish *Myxine glutinosa* is immunologically more similar to the C_4 of the teleost *Pseudophycis barbata* than to the B_4 of the latter [6, 45]. Similarly, the LDH of the ascidian *Pyura stolonifera* corresponds to the vertebrate A_4 in its kinetic properties, and shows more similarity to the LDH-C_4 of the teleosts in its amino acid composition and immunological characteristics than to their A_4 or B_4 [7]. The **spatial structure** of all vertebrate LDHs is very much the same; in addition to the substrate-binding domain there is a nucleotide-binding domain, as found in all dehydrogenases, and this has a typical Rossmann fold structure (Fig. 14.8). The genes of the human and murine LDH subunits A, B and C have the same intron/exon structure, with six introns lying in random-coil regions or near the ends of secondary structures on the surface of the molecule [65, 212, 251]. The products of additional LDH genes are found in the tetraploid salmonids and cyprinids, although they do not show the expected twofold increase; this is probably the result of silencing of duplicated genes or a failure to detect similar gene products with minor differences in electrophoretic mobility [96, 149]. **Polymorphism** of LDH subunits has been observed in many vertebrates, leading to very complicated electropherograms in, for example, fish and amphibians [96, 268].

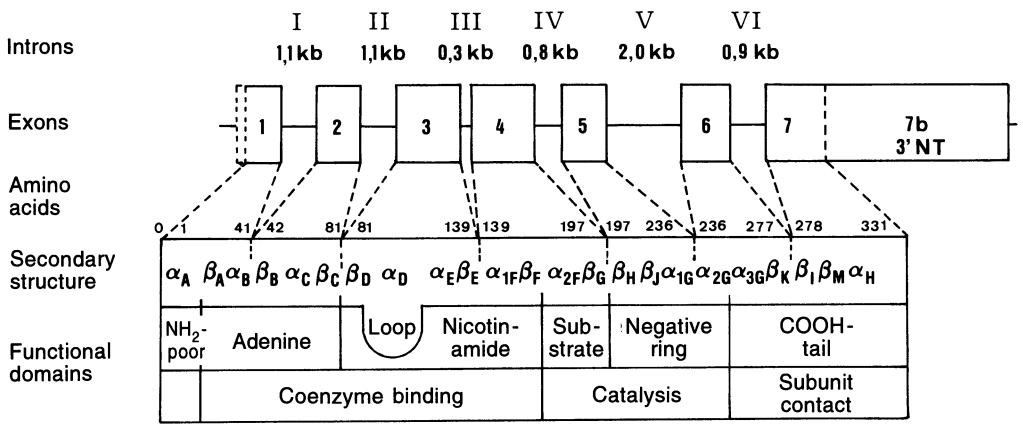


Fig. 14.8. The LDH-A of the mouse [138]. The exon/intron structure of the gene and the domain structure of the polypeptide

14.11.2 Lactate Dehydrogenases of Invertebrates

In comparison with the vertebrate enzymes, the **L-specific LDHs** of the invertebrates have no special features. They are almost all tetramers of about 140 kDa [63, 70]. A dimeric L-specific LDH of 70 kDa has been isolated from a squid *(Loligo vulgaris)* which, due to a genetic defect, had no octopine dehydrogenase in the mantle but showed high LDH activity [67]. The assumption that all these enzymes are homologous to vertebrate LDH awaits confirmation by sequence comparisons. In substrate affinity and pyruvate inhibition, most invertebrate LDHs are intermediate between the muscle and heart types of the vertebrates [63, 67]. **Isoenzymes** of L-specific LDH have been reported for only a few invertebrates. The LDH from the brain of the lobster *Homarus americanus* gives a five-banded electrophoretic pattern; a B-like type predominates in the tail muscle, and an A-like type predominates in the leg muscle. The crayfish *Orconectes limosus*, in contrast, has no LDH isoenzymes [70]. In the cephalopod *Sepia officinalis*, B-like forms with marked pyruvate inhibition are found in the heart, gill, brain, nidamental gland and ovary, and an A-like form with little substrate inhibition is found in the mantle muscle, hepatopancreas and skin [70]. The five-banded pattern found in the mosquito *Aedes aegyti* probably arises by random combination of two different subunits, one of which predominates in the head and the other in the thorax and abdomen. Two LDHs are present in the tapeworm *Hymenolepsis diminuta*, one being expressed in all tissues and the other only in the ripe ovary and eggs. The small liver-fluke *Dicrocoelium dendriticum* possesses six LDH isoenzymes and the larger *Fasciola hepatica* has four [137]. The allosteric activation by fructose bisphosphate shown by the L-specific LDH of the turbellarian *Polycelis nigra* is a peculiarity which otherwise probably occurs only in some parasitic platyhelminths.

The first D-specific LDH of the animal kingdom was discovered in 1968 in the xiphosuran *Limulus polyphemus* [27]. Since then, many systematic studies have shown that D- or L-specific enzymes are present in all animal groups. The **distribution of the two enzyme types** is extraordinarily sporadic (Table 14.6). The conclusions reached are that L- and D-specific LDHs are homologous, that few amino acid substitutions are required to change the stereo-specificity, and that an L/D switch has occurred many times in LDH evolution [144]. However, comparisons of the tryptic digest patterns of D-specific *Limulus* LDH and L-specific vertebrate LDH show such large differences that only a comparison of the amino acid sequences will determine whether or not the two LDH types are homologous [229]. In any case, D-lactate can be metabolized in some circumstances even in animals that possess only L-specific LDH. Mammalian liver, in fact, can turn over D-lactate 15-fold faster than L-lactate. This is of particular importance in the ruminants, where the D-lactate produced in the rumen by symbiontic bacteria is oxidized with the help of mitochondrial **D-2-hydroxy-acid dehydrogenases**, which are found especially in the brain cortex but also in the heart and liver [92].

D-Specific LDHs have been isolated from various animals and have been characterized: from the xiphosuran *Limulus polyphemus*, the barnacle *Balanus nubilis*, the polychaete *Nereis virens*, the cockle *Cardium edule* and the snails *Helix pomatia, H. aspersa, Haliotis cracherodii* and *Achatina achatina* [27, 70, 261]. The majority of D-specific LDHs are dimers of 70–80 kDa. The enzyme from *Helix aspersa* was described earlier as a tetramer but, more recently, was described as a dimer. The D-specific LDHs of the barnacle *Balanus nubilis* and the snail *Achatina achatina* are, however, tetrameric [261]. In their general catalytic properties, most of the D-specific invertebrate LDHs lie intermediate between the heart and muscle types of the vertebrates. Some, e.g. from *Limulus, Haliotis* and *Achatina*, are inhibited by pyruvate but others, e.g. from *Cardium* and *Helix aspersa*, are not [27, 261]. At least the enzymes from *Balanus, Haliotis* and *Helix aspersa* show relatively high activity against 2-keto- and 2-hydroxybutyrate.

Limulus polyphemus has a complicated **system of tissue-specific isoenzymes** which must be considered in relation to processing of the end products arising from activity-dependent anaerobiosis. Altogether, there are four different LDH subunits, A–D, which form six dimeric isoenzymes: a "heart-type triad" (AA, AB, BB) and a "muscle-type triad" (CC, CD, DD). The muscle-type triad predominates in the skeletal muscles, and the heart type predominates in the hepatopancreas; the heart contains more of the muscle-type LDH than of the heart type. The lactate affinity and the pyruvate sensitivity decrease in the order AB > BB > CD > DD. Under physiological conditions, lactate production is clearly preferred by the muscle type, whereas the predominant heart type of the hepatopancreas is active in both direc-

tions and can easily switch from lactate production to lactate oxidation. This distribution of isoenzymes of D-specific LDH allows a division of labour between the musculature and the other tissues just like that found for the L-specific LDHs of the vertebrates and the isoenzymes of octopine dehydrogenase in the cephalopods [27, 70]. Isoenzymes of D-specific LDH are also found in various gastropods but probably not as part of a complex adapted system like that of *Limulus* [183].

14.11.3 Opine Dehydrogenases

To date, five different opine dehydrogenases have been described (Fig. 14.5). The first to be characterized, in 1959, was the opine dehydrogenase from various molluscs and *Sipunculus nudus;* this was followed in 1976 by the alanopine/strombine dehydrogenase of the oyster *Crassostrea gigas*, in 1985 by the tauropine dehydrogenase of the snail *Haliotis lamellosa*, and in 1987 by the β-alanopine dehydrogenase of the mussel *Scapharca broughtonii*. It is quite possible that further enzymes of this type exist. As well as in the molluscs, octopine dehydrogenases (ODH) have since been reported in sea anemones, nemertines, sipunculids, and brachiopods and have, in part, been characterized; in the annelids, they have as yet been detected only in the polychaete *Glycera lapidum*. The distribution of ODH amongst the molluscs is by far the best documented. They are probably present in all cephalopods. In the gastropods, they are widely found in the prosobranch neogastropods as well as in several aracheo- and mesogastropods, but not in the opisthobranchs or pulmonates. ODHs are widely distributed in both marine and freshwater mussels with the exceptions of the families Ostreidae and Myidae. Alanopine and strombine dehydrogenases are usually found together in the poriferans, cnidarians, nemertines, sipunculids, brachiopods, annelids and mussels; in the molluscs, they are absent from the cephalopods, gastropods, nudibranchs (opisthobranchs) and all freshwater species [68, 84, 140]. Tauropine dehydrogenase not only is found in all species of the genus *Haliotis* but has also recently been isolated from the stem of the brachiopod *Glottidia pyramidata* [47, 69, 218]. β-Alanopine and the corresponding dehydrogenase are only known so far in the mussel *Scapharca broughtonii* [219].

All known opine dehydrogenases are monomeric proteins of about 35–40 kDa [47, 69, 70, 190, 242]. The similarities of the molecular masses and catalytic properties make it likely that **all opine dehydrogenases are homologous**, although this awaits confirmation from sequence data. It is already clear that there is no immunological relationship between ODH and other pyruvate dehydrogenases, and that ODH apparently has a very high rate of evolution; antibodies against ODH of the eight-armed cephalopod *Haplochlaena maculosa* cross-react with the ODH of all octopods but react only weakly with the enzyme of all other cephalopods and not at all with the enzyme of mussels and snails [70].

The **reactions** of the ODHs from the marine molluscs *Pecten maximus* and *Concholepas concholepas* have been examined in detail. In the forward reaction, NADH is bound first, followed randomly by arginine and pyruvate, whereas in the reverse reaction first NAD^+ is bound and then octopine [28]. In the **substrate specificities**, the opine dehydrogenases present a rather confusing picture. All can metabolize 2-oxobutyrate as well as pyruvate, but they metabolize keto acids with five or six C atoms much less, and only a few of them can react with oxaloacetate. All possibilities of amino acid specificity are found, from the high specificity for glycine as the only substrate, shown by strombine dehydrogenase from the sponge *Halichondria panicea*, to the broad specificity of the enzyme from the sea anemone *Bunodosoma cavernata*, which apart from arginine and lysine metabolizes alanine, glycine and other neutral amino acids; the latter enzyme is classifiable as a real ODH only from the fact that the Michaelis constants for all amino acids other than arginine have high, non-physiological values.

The **octopine dehydrogenases** can be ordered according to their activities against ω-amino substrates such as lysine or ornithine. First comes the enzyme from the sea anemone *Calliactis parasitica*, which reacts with lysine as well as with arginine, and with lysinopine as well as with octopine; the activity with lysine among various species declines in the order *Mytilus edulis* > *Glycimeris glycimeris* > *Sipunculus nudus* > *Sepia* brain > *Sepia* mantle. At the end come the ODHs that are absolutely specific for ω-guanidine substrates such as arginine, homoarginine or canavanine, e.g. from the mussels *Arctica, Pecten*, and *Anodonta* [241]. A similar situation exists for the **alanopine and strombine dehydrogenases**. The alanopine dehydrogenases of *Littorina littorea* and *Busycotypus canaliculatum* and the strombine dehydrogenase of *Halichondria panicea* are

highly specific. Between these two are enzymes that metabolize alanine and glycine, and these are better described as alanopine/strombine dehydrogenases, e.g. from the mussels *Crassostrea gigas, Mytilus edulis* and *Mercenaria mercenaria* and from the polychaete *Arenicola marina* [190, 242]. The tauropine dehydrogenases are also active with alanine [69]. All opine dehydrogenases are stereo-specific. Of all possible octopine isomers, only N-(S)-1-carboxy-4-guanidinobutyl-(R)-alanine occurs naturally; mesoalanopine arises from alanine. The enzymes from *Halichondria panicea* and *Crassostrea gigas* are only active against N-carboxymenthyl-(R)-alanine and not against the (S)- isomer [70].

Critical for the **regulation of opine dehydrogenases** is the fact that the affinity for one of the substrates is positively influenced by the other two. Thus, for example, the affinity of the ODHs of sea anemones and cephalopods for arginine and pyruvate is increased by NADH. Similarly, alanine enhances the pyruvate affinity of the alanopine dehydrogenase of *Littorina littorea* and vice versa. The affinity of the *Littorina* enzyme for alanine and pyruvate is also increased by a decrease in pH, and in this way alanopine production is stimulated by the anoxia-related acidosis [190]. A marked product inhibition is typical of most opine dehydrogenases, e.g. the effect of octopine on the ODH of sea anemones and cephalopods, and the effect of alanopine on the enzyme of the polychaete *Aphrodite aculeata* [70, 242]. Allosteric inhibition by ATP, ADP and AMP has been reported for the alanopine dehydrogenase of the oyster *Crassostrea gigas*.

The **biological role** of the different pyruvate reductases is by no means always obvious. For example, both the sea anemones and the cockle *Cardium edule* possess an ODH but do not produce octopine [155]. Opine dehydrogenases are usually found together with LDH, and a not insignificant number of species contain more than one type of opine dehydrogenase; the relative proportions of the individual pyruvate reductases vary from tissue to tissue [70]. The species- and organ-specific spectrum of the pyruvate reductases and the differences in their substrate specificity and kinetic-regulatory properties can probably be looked upon as adaptations to the species-specific spectrum of free amino acids or the different physiological functions of the enzymes. In the polychaete *Aphrodite aculeata*, each tissue contains only one opine dehydrogenase: longitudinal muscles, nervous tissue and scales have a broad-specificity alanopine dehydrogenase and the gut has a highly specific strombine dehydrogenase; LDH and ODH are both absent [242]. The snail *Busycotypus canaliculatum* possesses alanopine dehydrogenase, ODH and LDH. Alanopine dehydrogenases are to be found in all the organs and are important especially during habitat-dependent anaerobiosis; ODH occurs mainly in the muscle, and links phosphagen cleavage and glycolysis during activity-dependent anaerobiosis; LDH is restricted to the heart ventricle and the continuously active proboscis muscle [191]. In the edible mussel *Mytilus edulis*, the opine dehydrogenase of the gut is alanine specific, whereas the muscle enzyme, which has a particularly high glycine concentration, can also efficiently metabolize glycine [190].

Many animals possess different pyruvate reductases that are active in different physiological situations [84] (Table 14.5). This is of biological advantage when both hypoxic conditions and the need to escape predators occur in the life of an animal. The presence of multiple pyruvate reductases ensures that the products formed in one reaction do not inhibit the enzymes of the other reactions [70]. *Cardium tuberculatum* has altogether a D-specific LDH, two ODHs, four alanopine dehydrogenases and a strombine dehydrogenase. However, activity-dependent anaerobiosis gives rise exclusively to octopine, whereas habitat-dependent anaerobiosis produces mainly D-lactate (in addition to alanine) but only a little octopine, very little alanopine and no strombine. This ratio of end products is in accordance with the kinetic-regulatory properties of the enzymes and the intracellular concentrations of substrates and cosubstrates [156].

For the alanopine/strombine dehydrogenase of *Arenicola marina*, the ratio of the activities against alanine and glycine shifts with decreasing NADH concentration in favour of the latter, so that during activity-dependent anaerobiosis with a high rate of glycolysis mainly alanopine is produced, and in habitat-dependent anaerobiosis with reduced metabolism mainly strombine is produced, in addition to succinate, propionate and alanine [231]. *Sipunculus nudus* produces mainly octopine during activity-dependent anaerobiosis but strombine during anoxia. However, the octopine/strombine dualism is not fully explained by the kinetic properties of ODH and strombine/alanopine dehydrogenase and by the concentrations of the substrates and cosubstrates. There are further cases in which the adjustments in metabolism are not completely understood: The activity of the pedal retractor muscle of the

snail *Nassarius coronatus*, as in *Cardium tuberculatum*, produces octopine, and during anoxia produces D-lactate. The razor shell *Ensis directus* produces octopine during activity-dependent anaerobiosis and strombine during the habitat-dependent process [288]. The snail *Strombus luhuanus* produces octopine during excessive muscle activity, mainly strombine during the subsequent recovery period, but alanopine during anoxia [68, 102].

The cephalopods *Sepia officinalis* and *Loligo vulgaris* possess **tissue-specific isoenzymes of octopine dehydrogenase** which resemble the vertebrate LDH isoenzymes in the differences between their properties and their biological roles. The enzyme from the brain has higher affinities for octopine and NAD^+ than does the enzyme from the mantle musculature, and it is more strongly inhibited by pyruvate and octopine. According to this, the muscle isoenzyme is preferentially involved in octopine formation and that from the brain in octopine oxidation. In agreement with this is the fact that arginine is released from ^{14}C-labelled octopine in the brain and other organs but not in muscle. The octopine produced in the mantle muscle during activity-dependent anaerobiosis is transported in the haemolymph to other tissues, where it is oxidized [70]. The bivalves and the gastropod *Strombus luhuanus* have only one ODH locus; however, this is highly polymorphic, e.g. there are eight alleles in *Cardium edule* [70]. The two alleloenzymes of ODH in the sea anemone *Metridium senile* not only differ in their kinetic properties to the same extent as the cephalopod isoenzymes, but also show large frequency differences between populations. This has been interpreted as adaptation to different living conditions. Three tissue-specific isoenzymes of alanopine dehydrogenase can be distinguished in the snail *Busycotypus canaliculatum*. The enzyme from the hepatopancreas has a high affinity for alanopine; there is marked product inhibition of alanopine formation but not of its oxidation. In contrast, alanopine production is clearly favoured in the case of the muscle enzyme, and the enzyme of the gill is intermediate between these two extremes [191].

14.12 Special Pathways of Anaerobic Energy Production

14.12.1 Synthesis of Ethanol and Acetic Acid

Ethanol production has been described for acanthocephalans, cestodes, nematodes, some insect larvae and carp-like fish. In the cyprinids, however, it is found in the crucian carp *Carassius carassius*, the goldfish *C. auratus* and the bitterling *Rhodeus amarus* but not in the common carp *Cyprinus carpio* [276]. None of these species possesses a cytoplasmic pyruvate carboxylase as occurs in yeast. According to investigations of the goldfish, the larvae of the non-biting midge *Chironomus thummi* and the nematode *Panagrellus redivivus*, the pyruvate dehydrogenase complex is responsible for the decarboxylation of pyruvate to acetaldehyde. The aldehyde is then reduced to ethanol by NADH-dependent alcohol dehydrogenase. In *C. carassius* this has its highest activity in the skeletal muscles and not in the liver, as in the carp, the rainbow trout *Salmo gairdneri* and the rat. On the other hand, the NAD-specific aldehyde dehydrogenase of the common carp has a much lower activity in the skeletal muscle than in the gill, kidney or heart. Probably the most important function of this enzyme is the oxidation of aldehyde produced from biogenic amines by a monoaminooxidase. The musculature of the crucian carp is highly specialized for ethanol production by the high activity of alcohol dehydrogenase and its separation from the aldehyde dehydrogenase [177]. Other pathways are utilized in the Protozoa. The flagellate *Crithidia fasciculata*, like yeast, possesses an ethanol-forming system that comprises a cytosolic pyruvate decarboxylase and an alcohol dehydrogenase which mainly reduces aldehyde [31]. In contrast, *Entamoeba histolytica* lacks a pyruvate decarboxylase; in this case, acetyl-CoA is reduced by an acylating aldehyde dehydrogenase to acetaldehyde, and further to ethanol by the NADPH- or NADH-specific alcohol dehydrogenase [141]. **Glycerol** has been shown to be the main metabolite of *Trichomonas vaginalis* [32].

The rumen ciliates produce acetate, butyrate, lactate, CO_2 and $\mathbf{H_2}$. The metabolism of pyruvate to acetate and H_2 takes place in special cell organelles that are about 0.5 μm in diameter, the so-called hydrogenosomes, which were found initially in *Tritrichomonas foetus* and related flagellates. *Isotricha* species have been found to contain not only all four enzymes involved in the

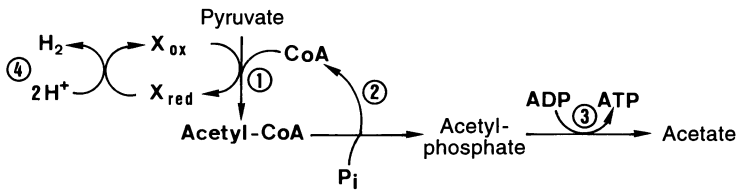


Fig. 14.9. The production of acetate and hydrogen by rumen ciliates [284]. *1*, Pyruvate synthase; *2*, phosphate acetyltransferase; *3*, acetate kinase; *4*, hydrogenase

formation of acetate and H_2 (Fig. 14.9) but also the enzyme activities required in the formation of butyrate from acetyl-CoA [114, 284, 285].

14.12.2 Synthesis of Succinate, Volatile Fatty Acids and Alanine

The highly specialized energy-extraction pathways which lead to succinate, propionate, acetate, methyl fatty acids and alanine as end products were discovered in the 1960s in *Ascaris* and other endoparasitic worms. Their presence in free-living animals, in particular mussels and annelids, was confirmed in the 1970s. The **synthesis of succinate and volatile fatty acids** occurs via oxaloacetate as intermediate. This is reduced by malate dehydrogenase (MDH) to malate, which is then transported from the cytoplasm into the mitochondria. There follows a complex dismutation reaction which produces succinate, propionate or branched fatty acids on the one hand, and pyruvate or acetate on the other hand (Fig. 14.10). The oxaloacetate is produced either from the glycolysis intermediate PEP by phosphoenolpyruvate carboxykinase (PEPCK) or from aspartate by the action of aspartate aminotransferase (GOT). The cleaved amino groups are transferred to pyruvate by alanine aminotransferase (GPT); the GOT and GPT reactions are linked by the cosubstrate 2-ketoglutarate/glutamate [289]. Alanine is an important end product of anaerobic metabolism, particularly during the initial stages of the process. It was shown in the polychaete *Arenicola* that more alanine is formed than aspartate is used. This suggests the existence of other pathways of alanine production in addition to the GOT/GPT link, both in *Arenicola* and in the edible mussel *Mytilus edulis*, although the origin of the amino nitrogen remains unclear [224, 287, 289]. Strangely enough, about equimolar amounts of D- and L-alanine arise in, e.g. *Arenicola marina, Sipunculus nudus* and the razor shell *Ensis directus* [225, 288]. With the help of the specific enzyme inhibitors 3-mercaptopicolinate for PEPCK and aminooxyacetate for the transaminase, it is possible to estimate what proportion of the anaerobically formed succinate originates from glycogen or from aspartate. In the adductor muscle of *Mytilus edulis*, succinate arises exclusively from aspartate but in other organs it is derived partly from glycogen [289]. The aspartate pathway is especially important in the hearts of marine mussels, where in fact it was discovered, but it is also important in the heart of the snail *Busycon contrarium* [51, 68, 70].

The formation of the C_4 compound succinate from the C_3 intermediate of glycogen catabolism requires a **carboxylation reaction**. There are, in principle, three relevant enzymes (see Fig. 18.2; p. 689): PEPCK, malic enzyme (ME) and pyruvate carboxylase (PC). **PEPCK** is always active in the succinate pathway of both endoparasitic and free-living anaerobes. The enzyme isolated from *Ascaris suum* is a monomer of about 80 kDa; the polypeptide has bound Mn^{2+} but not Mg^{2+}, although the Mg^{2+} complexes of GDP or IDP are as efficient as substrates as are the Mn^{2+} complexes [207]. PC is absent from the polychaete *Arenicola* and *Nereis* and most endoparasitic worms. The inhibition of PEPCK with 3-mercaptopicolinate stops anaerobic succinate formation in both polychaete species; lactate is formed as the alternative, with L-lactate occurring in *Arenicola* and D-lactate in *Nereis* [225]. PEPCK competes with pyruvate kinase (PK) for their common substrate, and the ratio of the PEPCK and PK activities is a good indicator of the overall capacity for anaerobic energy production in the succinate pathway. This is illustrated very nicely by a series of *Haliotis* species whose habitats range from the tidal zone *(H. rufescens, H. cracherodii)* to purely aerobic conditions at depths of 20 m *(H. corrugata)* (Table 14.8).

The mechanism that switches from PK (aerobic) to PEPCK (anaerobic) at the PEP branch point is not yet completely understood. A decrease in pH, an increase in HCO_3^- or a rise in the alanine concentration are possible triggering factors. In marine mussels, a decrease in pH simultaneously inhibits PK and stimulates PEPCK [105, 224]; the same effects are brought about by bicarbonate ions in the oligochaete

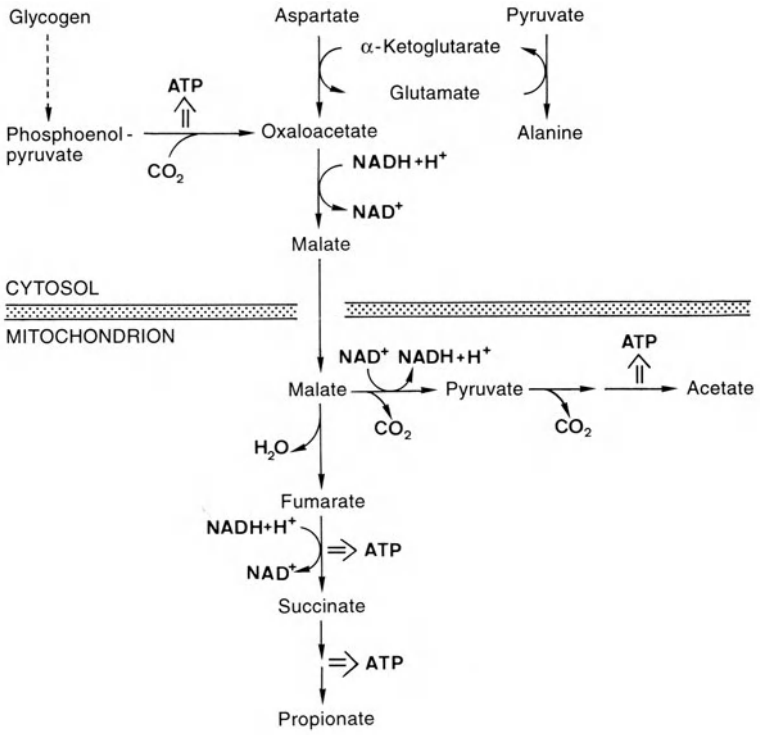


Fig. 14.10. The production of alanine, succinate, propionate and acetate in anaerobic metabolism [287]. The oxaloacetate produced from glycogen or aspartate is reduced to malate by the cytosolic malate dehydrogenase (MDH) and is transported into the mitochondrion; in a complicated dismutation reaction it is then either reduced to succinate via fumarate or oxidized to pyruvate and then converted to acetate. See the text for further explanation. The production of propionate from succinate is shown in Fig. 14.12

Table 14.8. The relative activities of pyruvate kinase (PK) and phosphoenolpyruvate carboxykinase (PEPCK) in various animals

	PK/PEPCK	References
Ascaris lumbricoides	0.04	[9]
Dicrocoelium dendriticum	0.05	[9]
Moniezia expansa	0.1	[9]
Strongyluris brevicaudata	0.19	[260]
Hymenolepis diminuta	0.20	[9]
Trichinella spiralis	0.3	[9]
Busycon contrarium: heart	0.6	[51]
Cardium edule: foot muscle	4.6	[155]
Schistosoma mansoni	5–10	[9]
Dirofilaria immitis	10.6	[9]
Mytilus edulis: adductor	19.7	[68]
adductor	3.6	[105]
heart	8.7	[105]
Crassostrea virginica:		
tonic adductor	8.8	[289]
phasic adductor	24.4	[289]
Anodonta cygnea: adductor	38	[105]
heart	90	[105]
Cardium tuberculatum: foot	500	[68]
Haliotis rufescens	5.9	[17]
H. fulgens	10.3	[17]
H. cracherodii	15.8	[17]
H. corrugata	24.3	[17]

Tubifex and the cestode *Hymenolepsis diminuta* [103]. Alanine inhibition of PK has been observed in various cnidarians, bivalves and crustaceans, but does not occur in *Tubifex* or the frog. Alanine relieves the inhibition of PEPCK by the PEPCK product ITP in jellyfish and mussels, but is without effect in crustaceans and amphibians. This fits the concept that PEPCK in the cnidarians and bivalves is adapted for the conversion of PEP to oxaloacetate in anaerobic energy production, whilst in the crustaceans and amphibians it functions in the reverse reaction for gluconeogenesis [105, 224]. The PEPCK of endoparasites like the roundworm *Ascaris lumbricoides* or the tapeworm *Hymenolepsis diminuta* can only be active in the decarboxylation mode because the almost complete absence of PK results in high concentrations of PEP, whereas oxaloacetate is rapidly reduced by the high activity of MDH [221]. The fact that the PK/PEPCK switch in *Arenicola* and other animals requires several hours is probably related to modifications of PK by such processes as phosphorylation and dephosphorylation [224].

The oxaloacetate produced from PEP or aspartate is reduced to malate by the cytoplasmic MDH, and the NADH from the GAPDH reaction is reoxidized in the process. The malate is introduced into the mitochondria where, on the one hand, it is reduced to succinate and, on the other hand, it is oxidized to pyruvate and acetate

(Fig. 14.10). In the reduction pathway, malate is converted to fumarate by fumarase, and further to succinate with the formation of ATP. These processes were discovered in 1961 in the roundworm *Ascaris lumbricoides* and studied initially in these and other endoparasitic worms [9, 221]; since 1977 they have been characterized in free-living animals such as *Arenicola marina* and *Mytilus edulis* [224]. The **fumarate reductase** includes part of the electron transport system of the inner mitochondrial membrane. Inhibitor experiments have shown that the enzyme complex consists of an NADH:coenzyme Q oxidoreductase coupled to phosphorylation site I (Fig. 14.11). The fumarate reductases of endoparasitic as well as of free-living animals are NADH specific, although the enzyme from the cestode *Spirometra mansonoides* has been reported to be active with NADPH. The chemical nature of coenzyme Q is not yet known, and the relationship between the succinate:coenzyme Q oxidoreductase from fumarate reductase and the succinate dehydrogenase acting in the reverse direction is not understood. There is some indication that these are two different enzymes, at least in the case of the nematode *Nippostrongylus brasiliensis* [64, 224].

A large part of the reducing equivalents needed for fumarate reduction arises from the oxidative decarboxylation of malate to pyruvate by **malic enzyme**. In contrast to the NADP-specific cytoplasmic ME of mammalian liver, the ME of *Ascaris* is mitochondrial and NAD specific. One particular problem is that 80 % of the fumarase and 60 % of the ME activity of *Ascaris* mitochondria is located in the intermembrane space. How then do fumarate and NADH reach the fumarate reductase in the inner mitochondrial membrane? The NADH:NAD^+ transhydrogenase of the mitochondria is probably responsible for the NADH transport [9, 205, 221]. The ME of the tapeworm *Spirometra mansonoides* is more active with NAD^+ than with $NADP^+$, whereas ME, for example from the cestode *Hymenolepsis diminuta* and the annelids *Arenicola marina* and *Tubifex* sp., is NADP specific. The production of NADH for the fumarate reductase involves an NADPH:NAD^+ transhydrogenase [103, 153, 224]. As well as the ME reaction, both endoparasitic and free-living animals have in the mitochondria further reactions of anaerobic metabolism which can yield NADH for fumarate reduction. The most important of these is the oxidative decarboxylation of pyruvate to acetyl-CoA, which is catalysed by the **pyruvate dehydrogenase complex**. In *Ascaris*, this complex produces NADH as in mammals, but in contrast to the mammalian enzyme it is not inhibited by very low NAD^+:NADH ratios. The ratio in *Ascaris* mitochondria, which are specialized for anaerobic energy production by the almost complete absence of cytochrome b, cytochrome c and cytochrome oxidase, and therefore lack a functional respiratory chain, is only 0.07–0.7 compared with 10 in mammalian mitochondria [9]. The release of acetic acid from acetyl-CoA is apparently the responsibility of CoA transferase which transfers CoA to succinate. Free succinate is recovered from the resulting succinyl-CoA by the action of succinyl-CoA synthetase, in the course of which GDP is phosphorylated to GTP [225].

In both endoparasitic and free-living animals, the mitochondrial production of NADH also involves the oxidation of malate by MDH, whereby the resulting oxaloacetate is decarboxylated to pyruvate either non-enzymatically or through the weak decarboxylating activity of ME [153]. Finally, it has been shown in *Arenicola* that acetyl-CoA is condensed with oxaloacetate to citrate under anaerobic conditions, and is introduced into the citric acid cycle; this produces succinate as well as two NADH for fumarate reduction. Stoichiometrically, five–sixths of the anaerobically produced succinate should be formed here by fumarate reductase and one–sixth should come from the citrate cycle [225]. In *Ascaris* mitochondria there is no citrate formation [9, 221].

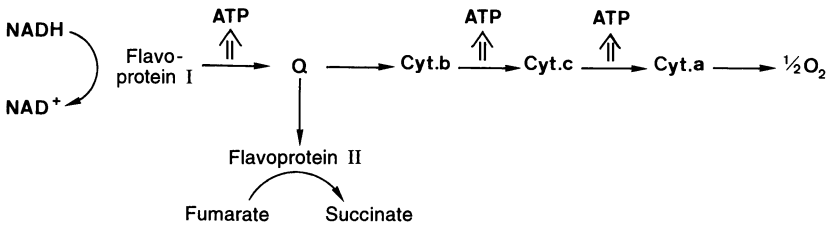


Fig. 14.11. Fumarate reductase is a part of the electron transport system. The reaction includes NAD:ubiquinone-oxidoreductase and succinate:ubiquinone-oxidoreductase; in the latter reaction the hydrogen is transferred anaerobically from coenzyme Q to succinate. The reaction produces one ATP per NADH [287]

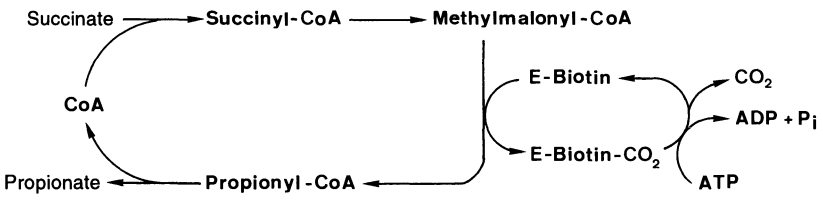


Fig. 14.12. The production of propionate from succinate [226]

Particularly during prolonged anaerobiosis, succinate is further converted to **propionate**, which is more easily excreted. This reaction chain, which has been examined in some detail in, for example *Ascaris lumbricoides, Fasciola hepatica, Arenicola marina* and *Mytilus edulis,* apparently corresponds to that in mammals by which propionate is introduced into the citrate cycle, albeit in the opposite direction (Fig. 14.12). The substrate for propionate synthesis is succinyl-CoA, which is converted via L-malonyl-CoA to D-malonyl-CoA by methylmalonyl-CoA mutase and methylmalonyl-CoA racemase; it is further decarboxylated to propionyl-CoA by propionyl-CoA carboxylase [9, 221, 225]. Of the enzymes involved, only the methylmalonyl-CoA mutase from *Ascaris* mitochondria has, as yet, been isolated. Like the mammalian liver enzyme, it is a homodimer of 147 kDa with tightly bound adenosylcobalamin [89]. Because of the vitamin dependency of the mutase activity, cestodes like *Spirometra mansonoides*, which can reabsorb cobalamin, are able to produce propionate, whilst cobalamin-deficient species, such as *Hymenolepsis diminuta*, are not [221]. It has not yet proved possible to isolate the *Ascaris* enzyme required for the synthesis of succinyl-CoA. A dimeric, 2 × 50-kDa acyl-CoA transferase purified from muscle mitochondria was found to be specific for short-chain substrates and inactive with succinate. At the most, it could be involved in the formation of propionate from succinate if supplied with catalytic amounts of propionyl-CoA [154].

In *Ascaris*, 70–80 % of the secreted volatile fatty acids consists of a mixture of racemic methylbutyrate and methylvalerate. These **branched-chain fatty acids** arise by condensation of a propionyl-CoA with either an acetyl-CoA or a second propionyl-CoA with the formation of NADH; the course of the reaction at the moment can only be presented theoretically (Fig. 14.13), but is apparently coupled to a site-I phosphorylation. As the production of methylvalerate and the corresponding phosphorylation, unlike in succinate synthesis, are not inhibited by malonate, the methylvalerate synthesizing system and the fumarate reductase are apparently linked to phosphorylation site I by various terminal reductases [205].

The metabolic pathways from PEP or aspartate to succinate or volatile fatty acids include five

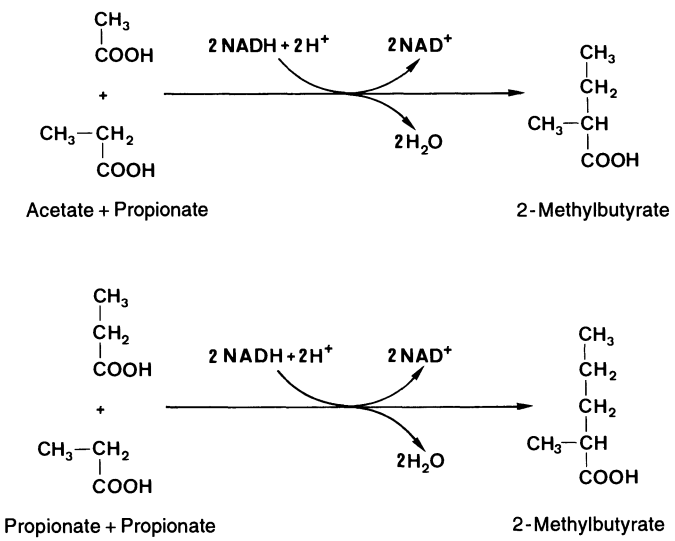


Fig. 14.13. A hypothetical scheme for the production of branched-chain fatty acids in the roundworm *Ascaris* [9]. For clarity, the formulae of the free acids are shown; however, the reactions probably start from acetyl- and propionyl-CoA [205]

ATP-forming steps. PEPCK primarily produces ITP or GTP, the terminal phosphate residue of which, however, is transferred to ADP by a nucleoside diphosphate kinase [9]. The fumarate reductase reaction is accompanied by a phosphorylation in all invertebrates, but not in the goldfish. One further ATP arises from each step of the conversion of succinate to propionate and of acetyl-CoA to acetate, and from the formation of branched-chain fatty acids by condensation of two acyl-CoA residues [83, 102, 225, 226] (Figs. 14.11–14.13). Thus, the complete conversion of a glycogen-bound glucosyl residue to volatile fatty acids yields seven ATP, i.e. more than twice the number (three ATP) arising during the production of lactate or opines. The synthesis of methyl fatty acids, instead of unbranched volatile fatty acids, in *Ascaris* and some other animals does not increase the ATP yield and must, therefore, have some other biological importance [205]. One ATP per mole amino acids arises during succinate formation from aspartate [102].

References

1. Abraham M, Horvath L. and Szajani B.: Isolation and characterization of pig muscle aldolase. A comparative study. Comp. Biochem. Physiol. Pt. B 80: 847–852 (1985)
2. Aman R. A., Kenyon G. L. and Wang C. C.: Cross-linking of the enzymes in the glycosome of Trypanosoma brucei. J. biol. Chem. 260: 6966–73 (1985)
3. Andreeva I. E. et al.: Phosphorylase kinase from chicken skeletal muscle – Quaternary structure, regulatory properties and partial proteolysis. Eur. J. Biochem. 158: 99–106 (1986)
4. Andreone T. L. et al: The amino acid sequence of rat liver glucokinase deduced from cloned cDNA. J. Biol. Chem. 264: 363–369 (1989)
5. Baldwin J. and Gyuris E.: Loggerhead turtle lactate dehydrogenases. How general is the apparent adaptation to prolonged anaerobiosis displayed by the lactate dehydrogenase isoenzymes from turtles of the genus Pseudemys? Comp. Biochem. Physiol. Pt. B 76: 191–195 (1983)
6. Baldwin J. and Lake P. S.: Lactate dehydrogenase homopolymer of hagfish heart and the single lactate dehydrogenase of lampreys display greater immunochemical similarity to LDH-C4 than to LDH-B4 of teleost fish. J. exp. Zool. 242: 99–102 (1987)
7. Baldwin J., Mortimer K. and Patak A.: Do ascidians possess the ancestral subunit type of vertebrate lactate dehydrogenase? J. exp. Zool. 246: 109–114 (1988)
8. Baldwin J., Davison W. and Forster M. E.: Properties of the muscle and heart lactate dehydrogenases of the New Zealand hagfish, Eptatretus cirrhatus: Functional and evolutionary implications. J. exp. Zool. 250: 135–139 (1989)
9. Barrett J.: Biochemistry of parasitic helminths. Macmillan, London 1981
10. Batke J.: Remarks on the supramolecular organization of the glycolytic system in vivo. FEBS Letters 251: 13–16 (1989)
11. Bazan J. F., Fletterick R. J. and Pilkis S. J.: Evolution of a bifunctional enzyme: 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. Proc. Nat. Acad. Sci. USA 86: 9642–46 (1989)
12. Becker K. J. et al.: Purification of pyruvate kinase isoenzymes type M1 and M2 from dog (Canis familiaris) and comparison of their properties with those from chicken and rat. Comp. Biochem. Physiol. Pt. B 83: 823–829 (1986)
13. van Beneden R. J. and Powers D. A.: The isozymes of glucose-phosphate isomerase (GPI-A2 and GPI-B2) from the teleost fish Fundulus heteroclitus (L.). J. biol. Chem. 260: 14596–603 (1985)
14. Biethinger M., Hoffmann R. and Hofer H. W.: Phosphofructokinase from mollusc muscle is activated by phosphorylation. Arch. Biochem. Biophys. 287: 263–267 (1991)
15. Bishop J. G. and Corces V. G.: The nucleotide sequence of a Drosophila melanogaster enolase gene. Nucleic Acids Res. 18: 191 (1990)
16. Boscá L. and Corredor C.: Is phosphofructokinase the rate-limiting step of glycolysis? Trends biochem. Sci. 9: 372–373 (1984)
17. Bowen C. E.: Facultative anaerobiosis in Haliotis (ormer/abalone): Pyruvate kinase and phosphoenolpyruvate carboxykinase activities. Comp. Biochem. Physiol. Pt. B 77: 197–200 (1984)
18. Brinkhoff W., Stöckmann K. and Grieshaber M.: Natural occurrence of anaerobiosis in molluscs from intertidal inhabitats. Oecologia 57: 151–155 (1983)
19. Brooks S. P. J. and Storey K. B.: The role of protein kinases in anoxia tolerance in facultative anaerobes: Purification and characterization of a protein kinase that phosphorylates pyruvate kinase. Biochim. biophys. Acta 1073: 253–259 (1991)
20. Brooks S. P. J. and Storey K. B.: Where is the glycolytic complex? A critical evaluation of present data from muscle tissue (Review Letter). FEBS Letters 278: 135–138 (1991)
21. Bryant C. (ed.): Metazoan life without oxygen. Chapman & Hall, New York 1991
22. Buczylko J., Hargrave P. A. and Kochman M.: Fructose-bisphosphate aldolase from Helix pomatia (I-II). Comp. Biochem. Physiol. Pt. B 67: 225–238 (1980)
23. Burgess D. G. and Penhoet E. E.: Characterization of the chicken aldolase B gene. J. biol. Chem. 260: 4604–14 (1985)
24. Butterworth P. E. and Barrett J.: Anaerobic metabolism in the free-living nematode Panagrellus redivivus. Physiol. Zool. 58: 9–17 (1985)
25. Cannata J. J. B. and Cazzulo J. J.: The aerobic fermentation of glucose by Trypanosoma cruzi. Comp. Biochem. Physiol. Pt. B 79: 297–308 (1984)
26. Cardenas M. L., Rabajille E. and Niemeyer H.: Fructose is a good substrate for rat liver glucokinase (hexokinase D). Biochem. J. 222: 363–370 (1984)
27. Carlsson K. H. and Gäde G.: Isolation and characterization of D-lactate dehydrogenase from muscle and hepatopancreas of Limulus polyphemus. J. comp. Physiol. B 155: 723–731 (1985)

28. Carvajal N. and Kessi E.: Kinetic mechanism of octopine dehydrogenase from the muscle of the sea mollusc, Concholepas concholepas. Biochim. biophys. Acta 953: 14–19 (1988)
29. Carvajal N., Gonzalez R. and Kessi E.: Aspartate activation of pyruvate kinase from the kidney of Concholepas concholepas (Gastropoda: Muricidae). Comp. Biochem. Physiol. Pt. B 95: 85–89 (1990)
30. Castellà-Escolà J. et al.: Sequence of rat skeletal muscle phosphoglycerate mutase cDNA. Biochem. biophys. Res. Commun. 165: 1345–51 (1989)
31. Cazzulo J. J. and Franke de Cazzulo B. M.: Pyruvate decarboxylase and alcohol dehydrogenase in the insect trypanosomatid, Crithidia fasciculata. Comp. Biochem. Physiol. Pt. B 81: 1019–22 (1985)
32. Chapman A. et al.: C13-NMR reveals glycerol as an unexpected major metabolite of the protozoan parasite Trichomonas vaginalis. FEBS Letters 191: 287–292 (1985)
33. Chen C. P. and Denlinger D. L.: Activation of phosphorylase in response to cold and heat stress in the flesh fly, Sarcophaga crassipalpis. J. Insect Physiol. 36: 549–553 (1990)
34. Chih C. P. and Ellington W. R.: Control of phosphofructokinase from the phasic adductor muscle of the bay scallop, Agropecten irradians concentricus. Comp. Biochem. Physiol. Pt. B 87: 767–772 (1987)
35. Choi Y. L. et al.: Molecular cloning and sequencing of the glycogen phosphorylase gene from Escherichia coli. FEBS Letters 243: 193–198 (1989)
36. Chrispeels J. and Gäde G.: Purification and characterization of pyruvate kinase from the foot muscle of the cockle, Cardium tuberculatum. Comp. Biochem. Physiol. Pt. B 82: 163–172 (1985)
37. Churchill T. A. and Storey K. B.: Intermediary energy metabolism during dormancy and anoxia in the land snail Otala lactea. Physiol. Zool. 62: 1015–30 (1989)
38. Coppes Z. L. et al.: Multilocus systems in fishes. Comp. Biochem. Physiol. Pt. B 96: 1–13 (1990)
39. Coronel C. E. et al.: Catalytic properties of the sperm-specific lactate dehydrogenase (LDH X or C4) from different species. J. exp. Zool. 225: 379–385 (1983)
40. Cortesi P. et al.: Characterization of pyruvate kinase isolated from the adult adductor muscle of the mollusc Scapharca inaequivalis. Comp. Biochem. Physiol. Pt. B 81: 123–129 (1985)
41. Crawford D. L., Constantino H. R. and Powers D. A.: Lactate dehydrogenase-B cDNA from the teleost Fundulus heteroclitus: Evolutionary implications. Mol. Biol. Evol. 6: 369–383 (1989)
42. Crerar M. M., David E. S. and Storey K. B.: Electrophoretic analysis of liver glycogen phosphorylase activation in the freeze-tolerant wood frog. Biochim. biophys. Acta 971: 72–84 (1988)
43. Cronin C. N. and Tipton K. F.: Purification and regulatory properties of phosphofructokinase from Trypanosoma (Trypanozoon) brucei. Biochem J. 227: 113–124 (1985)
44. Daum G., Keller K. and Lange K.: Association of glycolytic enzymes with the cytoplasmic side of the plasma membrane of glioma cells. Biochim. biophys. Acta 939: 277–281 (1988)
45. Dell'Agata M. et al.: Lactate dehydrogenase from Lampetra planeri is composed of chains of unique type which show intermediate properties between the heart and muscle isozymes of vertebrates. Comp. Biochem. Physiol. Pt. B 89: 323–327 (1988)
46. Dombrádi V. et al.: Regulation of phosphorylase kinase in Drosophila melanogaster. Insect Biochem. 17: 579–585 (1987)
47. Doemen C. and Ellington W. R.: Isolation and characterization of a taurine-specific opine dehydrogenase from the pedicles of the brachiopod, Glottidea pyramidata. J. exp. Zool. 243: 25–31 (1987)
48. Dugaiczyk A. et al.: Cloning and sequencing of a deoxyribonucleic acid copy of glyceraldehyd-3-phosphate dehydrogenase messenger ribonucleic acid isolated from chicken muscle. Biochemistry 22: 1605–13 (1983)
49. Dunaway G. A. and Kasten T. P.: Nature of the rat brain 6-phosphofructo-1-kinase isozymes. J. biol. Chem. 260: 4180–85 (1985)
50. Dworkin M. B., Segil N. and Dworkin-Rastl E.: Pyruvate kinase isozymes in oocytes and embryos from the frog Xenopus laevis. Comp. Biochem. Physiol. Pt. B 88: 743–749 (1987)
51. Ellington W. R.: Energy metabolism during hypoxia in the isolated, perfused ventricle of the whelk, Busycon contrarium Conrad. J. comp. Physiol. B 142: 457–464 (1981)
52. Ellington W. R.: The recovery from anaerobic metabolism in invertebrates. J. exp. Zool. 228: 431–444 (1983)
53. England W. R. and Baldwin J.: Anaerobic energy metabolism in the tail musculature of the Australian yabby Cherax destructor (Crustacea, Decapoda, Parastidae). Regulation of anaerobic glycolysis. Comp. Biochem. Physiol. Pt. B 80: 327–335 (1985)
54. Enomoto T. et al.: Occurrence and characterization of fructose-6-phosphate-2-kinase and fructose-2,6-bisphosphatase in Euglena gracilis. Comp. Biochem. Physiol. Pt. B 92: 477–480 (1989)
55. Ercolani L. et al.: Isolation and complete sequence of a functional human glyceraldehyde-3-phosphate dehydrogenase gene. J. Biol. Chem. 263: 15335–41 (1988)
56. Fields J. H. A.: Alternatives to lactic acid: possible advantages. J. exp. Zool. 228: 445–457 (1983)
57. Finol J. and Chin J.: Gluconeogenesis in toad gastric mucosa. Comp. Biochem. Physiol. Pt. B 65: 651–656 (1980)
58. Fletterick R. J. and Madsen N. B.: The structures and related functions of phosphorylase a. Annual Rev. Biochem. 49: 31–61 (1980)
59. Fothergill-Gilmore L.: The evolution of the glycolytic pathway. Trends biochem. Sci. 11: 47–51 (1986)
60. Fournier P. and Guderley H.: Evolution of the functional properties of pyruvate kinase isozymes: Pyruvate kinase L from Rana pipiens. J. comp. Physiol. B 156: 691–699 (1986)
61. Frankel J. S.: Lactate dehydrogenase isozymes in the island barb, Barbus oligolepis (Cypriniformes, Teleostei): Their characterization and ontogeny. Comp. Biochem. Physiol. Pt. B 87: 581–585 (1987)
62. Freemont P. S., Dunbar B. and Fothergill-Gilmore L. A.: The complete amino acid sequence of human skeletal-muscle fructose-biphosphate aldolase. Biochem. J. 249: 779–788 (1988)
63. Friedman K. H.: Carbohydrate metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 43–76. Pergamon Press, Oxford 1985
64. Fry M and Brazeley E. P.: NADH-fumarate reductase and succinate dehydrogenase activities in mitochondria of Ascaridia galli and Nippostrongylus brasiliensis. Comp. Biochem. Physiol. Pt. B 77: 143–150 (1984)

65. Fukasawa K. M. and Li S. S. L.: Complete nucleotide sequence of mouse lactate dehydrogenase-A functional gene: Comparison of the exon-intron organization of dehydrogenase genes. Genetics 116: 99–105 (1987)
66. Fundele R., Jagerbauer E. and Jarms G.: Phosphoglycerate mutase isozymes in the lower vertebrates. Experientia 41: 1471–72 (1985)
67. Gaede G.: L-Lactate-specific, dimeric lactate dehydrogenase from the mantle muscle of the squid, Loligo vulgaris: purification and catalytic properties. Comp. Biochem. Physiol. Pt. B 63: 387–393 (1979)
68. Gaede G.: Energy metabolism of arthropods and molluscs during environmental and functional anaerobiosis. J. exp. Zool. 228: 415–429 (1983)
69. Gaede G.: Purification and properties of tauropine dehydrogenase from the shell adductor muscle of the ormer, Haliotis lamellosa. Eur. J. Biochem. 160: 311–318 (1986)
70. Gaede G. and Gieshaber M. K.: Pyruvate reductases catalyze the formation of lactate and opines in anaerobic invertebrates (Review). Comp. Biochem. Physiol. Pt. B 83: 255–272 (1986)
71. Gaede G. and Meinardus-Hager G.: Anaerobic energy metabolism in Crustacea, Xiphosura and Mollusca: lactate fermentation versus multiple fermentation products. Zool. Beitr. N. F. 30: 187–203 (1986)
72. Gaede G.: Energy metabolism during anoxia and recovery in shell adductor and foot muscle of the gastropod mollusc Haliotis lamellosa: Formation of the novel anaerobic end product tauropine. Biol. Bull. 175: 122–131 (1988)
73. Gaede G.: Glycogen phosphorylase in the fat body of the cockroach species, Periplaneta americana and Nauphoeta cinerea: Isolation, partial characterization of 3 forms and activation by hypertrehalosaemic hormone. Z. Naturforsch. Sect. C 46: 149–162 (1991)
74. Gamblin S. J. et al.: Activity and specificity of human aldolases. J. mol. Biol. 219: 573–576 (1991)
75. Gehnrich S. C., Gegakis N. and Sul H. S.: Liver (B-type) phosphofructokinase mRNA. Cloning, structure, and expression. J. Biol. Chem. 263: 11755–59 (1988)
76. Giallongo A. et al.: Structure of the human gene for α-enolase. Eur. J. Biochem. 190: 567–573 (1990)
77. Gibson W. C., Swinkels B. W. and Borst P.: Post-transcriptional control of the differential expression of phosphoglycerate kinase genes in Trypanosoma brucei. J. mol. Biol. 201: 315–325 (1988)
78. Gleeson T. T.: Glycogen synthesis from lactate in skeletal muscle of the lizard Dipsosaurus dorsalis. J. comp. Physiol. B 156: 277–283 (1985)
79. Goldman S. S.: Gluconeogenesis in the amphibian retina. Lactate is preferred to glutamate as the gluconeogenic precursor. Biochem. J. 254: 359–365 (1988)
80. Goldsmith P. K. and Stetten M. R.: Hexokinases and sugar dehydrogenases of Limulus polyphemus. Comp. Biochem. Physiol. Pt. B 77: 41–50 (1984)
81. Gould G. W. and Bell G. I.: Facilitative glucose transporters, an expanding family. Trends biochem. Sci. 15: 18–23 (1990)
82. Greive H. and Surholt B.: Dependence of fructose-bisphosphatase from flight muscles of the bumblebee (Bombus terrestris L.) on calcium ions. Comp. Biochem. Physiol. Pt. B 97: 197–200 (1990)
83. Grieshaber M. K.: Metabolic regulation of energy metabolism. In: Addink A. D. F. and Spronk N. (eds.): Exogenous and endogenous influences on metabolic and neural control, pp. 225–242. Pergamon Press, Oxford 1982
84. Grieshaber M. K. and Kreutzer U.: Opine formation in marine invertebrates. Zool. Beitr. N. F. 30: 205–229 (1986)
85. Gruschczyk B. and Kamp G.: The shift from glycogenolysis to glycogen resynthesis after escape swimming: Studies on the abdominal muscle of the shrimp, Crangon crangon. J. comp. Physiol. B 159: 753–760 (1990)
86. Guderley H. and Bouchard C.: Isozymic differentiation of cardiac and muscle pyruvate kinase in the ringed seal Phoca hispida. Physiol. Zool. 55: 190–198 (1982)
87. Guderley H., Hamel L. and Lafond J.: Close resemblance between muscle pyruvate kinase from a primitive vertebrate, the river sturgeon Acipenser fulvescens, and the ancestral type K isozyme. J. comp. Physiol. B 153: 247–256 (1983)
88. Hakim G. et al.: Regulation by phosphorylation-dephosphorylation of pyruvate kinase in Venus gallina and Scapharca inaequivalis. Comp. Biochem. Physiol. Pt. B 80: 109–112 (1985)
89. Han Y. S., Bratt J. M. and Hogenkamp H. P. C.: Purification and characterization of methylmalonyl-CoA mutase from Ascaris lumbricoides. Comp. Biochem. Physiol. Pt. B 78: 41–45 (1984)
90. Hand S. C. and Somero G. N.: Phosphofructokinase of the hibernator Citellus beecheyi: Temperature and pH regulation of the activity via influences of the tetramer-dimer equilibrium. Physiol. Zool. 56: 380–388 (1983)
91. Harkins R. N. et al.: A comparison of the structure and activity of cat and trout muscle pyruvate kinases. Eur. J. Biochem. 136: 341–346 (1983)
92. Harmon D. L., Britton R. A. and Prior R. L.: In vitro rates of oxidation and gluconeogenesis from L(+)- and D(-)-lactate in bovine tissues. Comp. Biochem. Physiol. Pt. B 77: 365–368 (1984)
93. Hart D. T. et al.: Biogenesis of the glycosome in Trypanosoma brucei: The synthesis, translocation and turnover of glycosomal polypeptides. Embo J. 6: 1403–11 (1987)
94. Hata K. et al.: Purification and properties of glycogen phosphorylase from the adductor muscle of the scallop, Patinopecten yessoensis. Comp. Biochem. Physiol. Pt. B 87: 747–753 (1987)
95. Hayakawa Y.: Activation mechanism of insect fat body phosphorylase by cold: phosphorylase kinase, phosphatase and ATP level. Insect Biochem. 15: 123–128 (1985)
96. Henry T. and Ferguson A.: Kinetic studies on the lactate dehydrogenase isozymes of the brown trout, Salmo trutta L. Comp. Biochem. Physiol. Pt. B 85: 491–496 (1986)
97. Henry T. and Ferguson A.: Phosphoglucose isomerase isozymes and allozymes of the brown trout, Salmo trutta L. Comp. Biochem. Physiol. Pt. B 88: 751–756 (1987)
98. Hergenhahn H. G.: Purification and characterization of glycogen phosphorylase from abdominal muscle, heart and integument of the crayfish, Orconectes limosus. Comp. Biochem. Physiol. Pt. B 74: 473–479 (1983)
99. Herranz J. M. J. and Ruiz-Amil M.: Pyruvate kinase of mussel foot (Mytilus edulis L.) – kinetic properties, effectors. Comp. Biochem. Physiol. Pt. B 80: 361–364 (1985)

100. Hers H. G. and Hue L.: Gluconeogenesis and related aspects of glycolysis. Annual Rev. Biochem. 52: 617–653 (1983)
101. Hiraoka B. Y. et al.: The cDNA and protein sequences of mouse lactate dehydrogenase B: Molecular evolution of vertebrate lactate dehydrogenase genes A (muscle), B (heart) and C (testis). Eur. J. Biochem. 189: 215–220 (1990)
102. Hochachka P. W. and Somero G. N.: Biochemical adaptation. Princeton Univ. Press, Princeton 1984
103. Hoffmann K. H. et al.: Aerobic and anaerobic metabolism in Tubifex, a freshwater oligochaete. Zool. Beitr. N. F. 30: 153–170 (1986)
104. Holt R. W. and Leibel W. S.: Coexpression of distinct eye- and liver-specific LDH isozymes in chichlid fish. J. exp. Zool. 244: 337–343 (1987)
105. Holwerda D. A. and Veenhof P. R.: Aspects of anaerobic metabolism in Anodonta. Comp. Biochem. Physiol. Pt. B 78: 707–711 (1984)
106. Huang X. Y. et al.: Genomic organization of the glyceraldehyde-3-phosphate dehydrogenase gene family of Caenorhabditis elegans. J. mol. Biol. 206: 411–424 (1989)
107. Hue L. and Rider M. H.: Role of fructose-2,6-biphosphate in the control of glycolysis in mammalian tissues (Review article). Biochem. J. 245: 313–324 (1987)
108. Huskins K. R., Bernhard S. A. and Dahlquist F. W.: Halibut muscle 3-phosphoglycerate kinase: Chemical and physical properties of the enzyme and its substrate complexes. Biochemistry 21: 4180–88 (1982)
109. Hutchison W. F., Turner A. C. and Oelschlegel F. J. jr.: Hexokinase of the adult dog heartworm, Dirofilaria immitis. Comp. Biochem. Physiol. Pt. B 58: 131–134 (1977)
110. Hwang P. K. and Fletterick R. J.: Convergent and divergent evolution of regulatory sites in eukaryotic phosphorylases. Nature 324: 80–84 (1986)
111. Inoue H., Noguchi T. and Tanaka T.: Complete amino acid sequence of rat L-type pyruvate kinase deduced from the cDNA sequence. Eur. J. Biochem. 154: 465–469 (1986)
112. Isani G. et al.: Energy metabolism during anaerobiosis and recovery in the posterior adductor muscle of the bivalve Scapharca inequivalis (Bruguière). Comp. Biochem. Physiol. Pt. B 93: 193–200 (1989)
113. Jensen G. S.: A comparison of some properties of myocardial phosphofructokinase from the rainbow trout (Salmo gairdneri) and the diving turtle (Pseudemys scripta). Comp. Biochem. Physiol. Pt. B 70: 161–164 (1981)
114. Johnson P. J. et al.: Molecular analysis of the hydrogenosomal ferredoxin of the anaerobic protist Trichomonas vaginalis. Proc. Nat. Acad. Sci. USA 87: 6097–6101 (1990)
115. Kamp G.: Feature of glycogen phosphorylase from the body wall musculature of the lugworm Arenicola marina and the mode of activation during anoxia. Biol. Chem. Hoppe-Seyler 367: 109–117 (1986)
116. Kamp G. and Juretschke H. P.: Hypercapnic and hypocapnic hypoxia in the lugworm Arenicola marina: A P31-NMR study. J. exp. Zool. 252: 219–227 (1989)
117. Kendall G. et al.: Trypanosoma cruzi glycosomal glyceraldehyde-3-phosphate dehydrogenase does not conform to the „hotspot" topogenic signal model. Embo J. 9: 2751–58 (1990)
118. Kettler M. K. and Whitt G. S.: An apparent progressive and recurrent evolutionary restriction in tissue expression of a gene, the lactate dehydrogenase-C gene, within a family of bony fish (Salmoniformes: Umbridae). J. mol. Evol. 23: 95–107 (1986)
119. Khoja S. M.: Phosphofructokinase from the epithelial cells of rat small intestine. Comparison of regulatory properties with those of skeletal muscle, liver and brain phosphofructokinase. Comp. Biochem. Physiol. Pt. B 85: 337–341 (1986)
120. Kilimann M. W. et al.: The α and β subunits of phosphorylase kinase are homologous: cDNA cloning and primary structure of the β subunit. Proc. Nat. Acad. Sci. USA 85: 9381–85 (1988)
121. Kitamura K. et al.: Purification and characterization of rat skeletal muscle fructose-6-phosphate-2-kinase/fructose-2,6-bisphosphatase. J. Biol. Chem. 264: 9799–9806 (1989)
122. Kretschmer M. et al.: Fructose-2,6-biphosphatase and 6-phosphofructo-2-kinase are separable in yeast. Biochem. J. 246: 755–759 (1987)
123. Kreutzer U., Siegmund B. R. and Grieshaber M. K.: Parameters controlling opine formation during muscular activity and environmental hypoxia. J. comp. Physiol. B 159: 617–628 (1989)
124. Kueng V. et al.: A glycosomal protein (p60) which is predominantly expressed in procyclic Trypnosoma brucei. Characterization and DNA sequence. J. Biol. Chem. 264: 5203–09 (1989)
125. Kulkarni G. et al.: Ascaris suum phosphofructokinase. Phosphorylation by protein kinase and sequence of the phosphopeptide. J. biol. Chem. 262: 32–34 (1987)
126. Ladewig de Panepucci L L., Schwantes M. L. B. and Schwantes A. R.: Loci that encode the lactate dehydrogenase in 23 species of fish belonging to the orders Cypriniformes, Siluriformes and Perciformes: adaptive features. Comp. Biochem. Physiol. Pt. B 77: 867–876 (1984)
127. Lamandé N. et al.: Murine muscle-specific enolase: cDNA cloning, sequence, and developmental expression. Proc. Nat. Acad. Sci. USA 86: 4445–49 (1989)
128. Lambeir A. M., Opperdoes F. R. and Wierenga R. K.: Kinetic properties of triose-phosphate isomerase from Trypanosoma brucei brucei – A comparison with the rabbit muscle and yeast enzyme. Eur. J. Biochem. 168: 69–74 (1987)
129. Landureau J. C. and Toulmond A.: Insect hemocytes – Cells adapted to anaerobiosis. Experientia 36: 966–968 (1980)
130. Lange A. J. and Pilkis S. J.: Sequence of human liver 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase. Nucleic Acids Res. 18: 3652 (1990)
131. Lavine J. E. et al.: Purification and properties of galactokinase from Tetrahymena thermophila. Biochim. biophys. Acta 717: 76–85 (1982)
132. Lawrence G. M. and Trayer I. P.: Hexokinase isoenzymes: Antigenic cross-reactivities and amino acid compositional relatedness. Comp. Biochem. Physiol. Pt. B 79: 233–238 (1984)
133. LeBlancq S. M. et al.: Evidence for gene conversion between the phosphoglycerate kinase genes of Trypanosoma brucei. J. mol. Biol. 200: 439–447 (1988)
134. LeBoulch P. et al.: Molecular cloning and nucleotide sequence of murine 2,3-bisphosphoglycerate mutase cDNA. Biochem. biophys. Res. Commun. 156: 874–881 (1988)

135. Lee C. P., Kao M. C. and French B. A.: The rabbit muscle phosphofructokinase gene. Implications for protein structure, function, and tissues specificity. J. biol. Chem. 262: 4195–99 (1987)
136. Leite A. et al.: Phosphofructokinase from bumblebee flight muscle. Molecular and catalytic properties and role of the enzyme in regulation of the fructose 6-phosphate/fructose 1,6-bisphosphate cycle. J. Biol. Chem. 263: 17527–33 (1988)
137. León P., Hermoso R. and Monteoliva M.: Isoenzymes of lactate dehydrogenase (EC 1.1.1.27) in Dicrocoelium dendriticum and Fasciola hepatica (Trematoda). Comp. Biochem. Physiol. Pt. B 83: 159–161 (1986)
138. Li S. S. L. et al.: Protein structure and gene organization of mouse lactate dehydrogenase-A isozyme. Eur. J. Biochem. 149: 215–225 (1985)
139. Lively M. O. et al.: Complete amino acid sequence of rat liver 6-phosphofructo-2-kinase/fructose-2,6-biphosphatase. J. biol. Chem. 263: 839–849 (1988)
140. Livingstone D. R. et al.: Studies on the phylogenetic distribution of pyruvate oxidoreductase. Biochem. Syst. Ecol. 11: 415–425 (1983)
141. Lo H. S. and Reeves R. E.: Pyruvate-to-ethanol pathways in Entamoeba histolytica. Biochem. J. 171: 225–230 (1978)
142. Lonberg N. and Gilbert W.: Intron/exon structure of the chicken pyruvate kinase gene. Cell 40: 81–90 (1985)
143. Lone Y. C. et al.: Complete nucleotide and deduced amino acid sequences of rat L-type pyruvate kinase. FEBS Letters 195: 97–100 (1986)
144. Long G. L.: The stereospecific distribution and evolutionary significance of invertebrate lactate dehydrogenases. Comp. Biochem. Physiol. Pt. B 55: 77–83 (1976)
145. Lu H. S., Yuan P. and Gracy R. W.: Primary structure of human triosephosphate isomerase. J. biol. Chem. 259: 11958–68 (1984)
146. Magnuson M. A. and Shelton K. D.: An alternate promoter in the glucokinase gene is active in the pancreatic β cell. J. Biol. Chem. 264: 15936–42 (1989)
147. Mahrenholz A. M., Hefta S. A. and Mansour T. E.: Phosphofructokinase from Fasciola hepatica. Sequence of the cAMP-dependent protein kinase phosphorylation site. Arch. Biochem. Biophys. 288: 463–467 (1991)
148. Malek A. A. et al.: Fructose-1,6-bisphosphate aldolase from Drosophila melanogaster: Primary structure analysis, secondary structure prediction, and comparison with vertebrate aldolases. Arch. Biochem. Biophys. 266: 10–31 (1988)
149. Markert C. L., Shaklee J. B. and Whitt G. S.: Evolution of a gene. Multiple genes for LDH isozymes provide a model of the evolution of gene structure, function and regulation. Science 189: 102–114 (1975)
150. van Marrevijk W. J. A.: Occurrence of three forms of glycogen phosphorylase and flight-induced enzyme activation in fat body of the American cockroach, Periplaneta americana. Comp. Biochem. Physiol. Pt. B 94: 165–169 (1989)
151. McAleese S. M. et al.: Complete amino acid sequence of the neuron-specific gamma isozyme of enolase (NSE) from human brain and comparison with the non-neuronal alpha form (NNE). Eur. J. Biochem. 178: 413–417 (1988)
152. McCarrey J. R.: Molecular evolution of the human Pgk-2 retroposon. Nucleic Acids Res. 18: 949–955 (1990)
153. McKelvey J. R. and Fioravanti C. F.: Coupling of „malic" enzyme and NADPH:NAD transhydrogenase in the energetics of Hymenolepis diminuta (Cestoda). Comp. Biochem. Physiol. Pt. B 77: 737–742 (1984)
154. McLaughlin G. L., Saz H. J. and de Bruyn B. S.: Purification and properties of an acyl CoA transferase from Ascaris suum muscle mitochondria. Comp. Biochem. Physiol. Pt. B 83: 523–527 (1986)
155. Meinardus G. and Gäde G.: Anaerobic metabolism of the common cockle, Cardium edule – IV. Time dependent changes of metabolites in the foot and gill tissue induced by anoxia and electrical stimulation. Comp. Biochem. Physiol. Pt. B 70: 271–277 (1981)
156. Meinardus-Hager G. and Gäde G.: The separate function of D-lactate-, octopine-, and alanopine-dehydrogenase in the foot muscle of the jumping cockle Cardium tuberculatum during anaerobiosis. J. comp. Physiol. B 156: 873–881 (1986)
157. Michaelidis B. and Storey K. B.: Influence of pH on the regulatory properties of aerobic and anoxic forms of pyruvate kinase in a marine whelk. J. exp. Zool. 253: 245–251 (1990)
158. Michaelidis B. and Beis I.: Studies on the anaerobic energy metabolism in the foot muscle of marine gastropod Patella caerulea (L.). Comp. Biochem. Physiol. Pt. B 95: 493–500 (1990)
159. Michels P. A. M.: Evolutionary aspects of trypanosomes: analysis of genes. J. mol. Evol. 24: 45–52 (1986)
160. Michels P. A. M. et al.: The cytosolic and glycosomal isoenzymes of glyceraldehyde-3-phosphate dehydrogenase in Trypanosoma brucei have a distant evolutionary relationship. Eur. J. Biochem. 198: 421–428 (1991)
161. Miller K.: The role of lactate production in the metabolic support of locomotion by clawed frogs Xenopus laevis. Physiol. Zool. 56: 580–584 (1983)
162. Misset O., Bos O. J. M. and Opperdoes F. R.: Glycolytic enzymes of Trypanosoma brucei – Simultaneous purification, intraglycosomal concentration and physical properties. Eur. J. Biochem. 157: 441–453 (1986)
163. Mitchell J. M., Daron H. H. and Frandsen J. C.: Purification and properties of aldolase from the parasitic protozoan Eimeria stidae (Coccidia). Comp. Biochem. Physiol. Pt. B 73: 221–229 (1982)
164. Mochizuki Y and Hori S. H.: Purification and properties of hexokinase from the starfish, Asterias amurensis. J. Biochem. Tokyo 81: 1849–55 (1977)
165. Mochizuki Y.: Evolutionary aspects of lower deuterostomian hexokinases. Comp. Biochem. Physiol. Pt. B 70: 745–751 (1981)
166. Moore S.: Substrate cycles. Trends biochem. Sci. 10: 386–387 (1985)
167. Morishima I. and Ueno T.: Structural and kinetic properties of glycogen phosphorylase a from fat body of the silkworm, Bombyx mori. Comp. Biochem. Physiol. Pt. B 96: 591–595 (1990)
168. Morris S. and Greenaway P.: Adaptations to a terrestrial existence in the robber crab, Birgus latro L. IV. L-Lactate dehydrogenase function and L-lactate accumulation during exercise. Comp. Biochem. Physiol. Pt. B 94: 59–64 (1989)
169. Moser D., Johnson L. and Lee C. Y.: Multiple forms of Drosophila hexokinase. Purification, biochemical and immunological characterizations. J. biol. Chem. 255: 4673–79 (1980)

170. Mosse P. R. L.: An investigation of gluconeogenesis in marine teleosts, and the effect of long-term exercise on hepatic gluconeogenesis. Comp. Biochem. Physiol. Pt. B 67: 583–592 (1980)
171. Mukai T. et al.: Tissue-specific expression of rat aldolase A mRNAs. Three molecular species differing only in the 5'-terminal sequences. J. biol. Chem. 261: 3347–54 (1986)
172. Mukai T. et al.: The structure of the brain-specific rat aldolase-C gene and its regional expression. Biochem. biophys. Res. Commun. 174: 1035–42 (1991)
173. Munday K. A., Giles I. G. and Post P. C.: Review of the comparative biochemistry of pyruvate kinase. Comp. Biochem. Physiol. Pt. B 67: 403–411 (1980)
174. Nagayama F. et al.: A hexokinase from fish liver with wide specificity for nucleotides as phosphoryl donor. Biochim. biophys. Acta 615: 85–93 (1980)
175. Nakano K., Hwang P. K. and Fletterick R. J.: Complete cDNA sequence for rabbit muscle glycogen phosphorylase. FEBS Letters 204: 283–287 (1986)
176. Newgard C. et al.: Human brain glycogen phosphorylase. Cloning, sequence analysis, chromosomal mapping, tissue expression, and comparison with the human liver and muscle isozymes. J. Biol. Chem. 263: 3850–57 (1988)
177. Nilsson G. E.: A comparative study of aldehyde dehydrogenase and alcohol dehydrogenase activities in crucian carp and three other vertebrates: apparent adaptation to ethanol production. J. comp. Physiol. B 158: 479–485 (1988)
178. Noguchi T., Inoue H. and Tanaka T.: The M1- and M2-isozymes of rat pyruvate kinase are produced from the same gene by alternative RNA splicing. J. biol. Chem. 261: 13807–12 (1986)
179. Oguchi M., Kanda T. and Akamatsu N.: Hexokinase of Angiostrongylus cantonensis: Presence of a glucokinase. Comp. Biochem. Physiol. Pt. B 63: 335–340 (1979)
180. Ohshima Y. et al.: cDNA cloning and nucleotide sequence of rat muscle-specific enolase (ßß enolase). FEBS Letters 242: 425–430 (1989)
181. Opperdoes F. R.: Glycosomes may provide clues to the import of peroxisomal proteins. Trends biochem. Sci. 13: 255–260 (1988)
182. Orr G. L. et al.: Cyclic AMP does not mediate the action of synthetic hypertrehalosemic peptides from the corpus cardiacum of Periplaneta americana. Insect Biochem. 15: 703–709 (1985)
183. Ottavani E. and Ferrari R.: Isoenzymes of lactate dehydrogenase in four species of muricid gastropods (Mollusca, Prosobranchia). Comp. Biochem. Physiol. Pt. B 73: 581–583 (1982)
184. Palm D., Goerl R. and Burger K. J.: Evolution of catalytic and regulatory sites in phosphorylases. Nature 313: 500–502 (1985)
185. Paolella G. et al.: Isolation and nucleotide sequence of a full-length cDNA coding for aldolase B from human liver. Nucleic Acids Res. 12: 7401–10 (1984)
186. Paolella G. et al.: Structure and expression of mouse aldolase genes – Brain specific aldolase C amino acid sequence is closely related to aldolase A. Eur. J. Biochem. 156: 229–235 (1986)
187. Patience R. L., Thomas J. D. and Sterry P. R.: Production and release of carboxylic acids during oxic and anoxic metabolism by the pulmonate snail Biomphalaria glabrata. Comp. Biochem. Physiol. Pt. B 76: 253–262 (1983)
188. Payne M. A. et al.: Fructose 2,6-bisphosphate and AMP increase the affinity of the Ascaris suum phosphofructokinase for fructose-6-phosphate in a process separate from the relief of ATP inhibition. J. Biol. Chem. 266: 8891–96 (1991)
189. Picukans I. and Umminger B. L.: Comparative activities of glycogen phosphorylase and γ-amylase in livers of carp (Cyprinus carpio) and goldfish (Carassius auratus). Comp. Biochem. Physiol. Pt. B 62: 455–457 (1979)
190. Plaxton W. C. and Storey K. B.: Alanopine dehydrogenase: Purification and characterization of the enzyme from Littorina littorea foot muscle. J. comp. Physiol. B 149: 57–65 (1982)
191. Plaxton W. C. and Storey K. B.: Purification and properties of alanopine dehydrogenase isozymes from the channeled whelk, Busycotypus canaliculatum. Comp. Biochem. Physiol. Pt. B 76: 321–326 (1983)
192. Plaxton W. C. and Storey K. B.: Tissue specific isozymes of pyruvate kinase in the channeled whelk Busycotypus canaliculatum: Enzyme modification in response to environmental anoxia. J. comp. Physiol. B 155: 291–296 (1985)
193. Plaxton W. C. and Storey K. B.: Glycolytic enzyme binding and metabolic control in anaerobiosis. J. comp. Physiol. B 156: 635–640 (1986)
194. Pocinwong S. et al.: Phosphorylase kinase from dogfish skeletal muscle. Purification and properties. Biochemistry 20: 7219–26 (1981)
195. Poertner H. O.: Contributions of anaerobic metabolism to pH regulation in animal tissues: Theory. J. exp. Biol. 131: 69–87 (1987)
196. Poorman R. A. et al.: Evolution of phosphofructokinase – Gene duplication and creation of new effector sites. Nature 309: 467–469 (1984)
197. Putzer V. M., de Zwaan A. and Wieser W.: Anaerobic energy metabolism in the oligochaete Lumbriculus variegatus Mueller. J. comp. Physiol. B 159: 707–715 (1990)
198. Quinn D. and Burggren W.: Lactate production, tissue distribution and elimination following exhaustive exercise in larval and adult bullfrogs Rana catesbeiana. Physiol. Zool. 56: 597–613 (1983)
199. Rao Y. P. and Rao D. G. V. P.: End products of anaerobic metabolism in Cerithidea (Cerithideopsylla) cingulata (Gmelin 1790) and Cerithium coralium Kiener 1841. Can. J. Zool. 61: 1304–10 (1983)
200. Ray W. J. et al.: The complete amino acid sequence of rabbit muscle phosphoglucomutase. J. biol. Chem. 258: 9166–74 (1983)
201. Redecker B. and Zebe E.: Anaerobic metabolism in aquatic insect larvae. Studies on Chironomus thummi and Culex pipiens. J. comp. Physiol. B 158: 307–315 (1988)
202. Rehse P. H. and Davidson W. S.: Purification and properties of a C-type isozyme of lactate dehydrogenase from the liver of the Atlantic cod (Gadus morhua). Comp. Biochem. Physiol. Pt. B 84: 145–150 (1986)
203. Reyes A. and Cárdenas M. L.: All hexokinase isoenzymes coexist in rat hepatocytes intermediary metabolism and metabolic control. Biochem. J. 221: 303–309 (1984)
204. Rigaut J. P. and Chalumneau M. T.: Pyruvate kinase isozyme patters of fish. Comp. Biochem. Physiol. Pt. B 77: 451–458 (1984)

205. Rioux A. and Komuniecki R.: 2-Methylvalerate formation in mitochondria of Ascaris suum and its relationship to anaerobic energy generation. J. comp. Physiol. B 154: 349–354 (1984)
206. Roberts B. and Anderson P. J.: The purification and kinetic characterization of eel white muscle pyruvate kinase. Comp. Biochem. Physiol. Pt. B 80: 51–56 (1985)
207. Rohrer S. P., Saz H. J. and Nowak T.: Purification and characterization of phosphoenolpyruvate carboxykinase from the parasitic helminth Ascaris suum. J. biol. Chem. 261: 13049–55 (1986)
208. Rongnoparut P. et al.: Isolation and characterization of the transcriptionally regulated mouse liver (B-type) phosphofructokinase gene and its promoter. J. Biol. Chem. 266: 8086–91 (1991)
209. Royt P. W.: Partial characterization and inactivation of membrane-bound phosphofructokinase from Tetrahymena pyriformis. Biochim. biophys. Acta 657: 138–152 (1981)
210. Sabrouty S. R. E. et al.: The muridae glyceraldehyde-3-phosphate dehydrogenase family. J. mol. Evol. 29: 212–222 (1989)
211. Sakai I., Sharief F. S. and Li S. S.: Molecular cloning and nucleotide sequence of the cDNA for sperm-specific lactate dehydrogenase-C from mouse. Biochem. J. 242: 619–622 (1987)
212. Sakai I. et al.: The cDNA and protein sequences of human lactate dehydrogenase B. Biochem. J. 248: 933–936 (1987)
213. Sakakibara M. et al.: Human aldolase isozyme gene: The structure of multispecies aldolase B mRNAs. Nucleic Acids Res. 13: 5055–69 (1985)
214. Sanchez J. L. et al.: Hepatopancreas hexokinase isozymes from the mussel Mytilus galloprovincialis L. Comp. Biochem. Physiol. Pt. B 74: 807–814 (1983)
215. Sanchez J. L. et al.: Pyruvate kinase from the hepatopancreas of Mytilus galloprovincialis: interaction with its allosteric effectors. Comp. Biochem. Physiol. Pt. B 80: 577–582 (1985)
216. Sand O.: The effect of fructose 1,6-diphosphate on pyruvate kinase from the liver of the flounder (Platichthys flesus L.). Comp. Biochem. Physiol. Pt. B 90: 401–407 (1988)
217. Sangster N. C. and Mettrick D. F.: Effects of 5-hydroxytryptamine, cyclic AMP, AMP, and fructose 2,6-bisphosphate on phosphofructokinase activity in Hymenolepis diminuta. Comp. Biochem. Physiol. Pt. B 88: 317–321 (1987)
218. Sato M. and Gäde G.: Rhodoic acid dehydrogenase: a novel amino acid-linked dehydrogenase from muscle tissue of Haliotis species. Naturwissenschaften 73: 207–209 (1986)
219. Sato M. et al.: Isolation of a new opine, β-alanopine, from the extracts of the muscle of the marine bivalve mollusc, Scapharca broughtonii. Comp. Biochem. Physiol. Pt. B 88: 803–806 (1987)
220. Savina M. V. and Derkachev E. F.: Switch on and switch off phenomenon of liver glyconeogenic function in the lamprey (Lampetra fluviatilis L.) under the influence of season and temperature. Comp. Biochem. Physiol. Pt. B 75: 531–539 (1983)
221. Saz H. J.: Energy metabolism of parasitic helminths: Adaptation to parasitism. Annual Rev. Physiol. 43: 323–341 (1981)
222. van Schaftingen E., Opperdoes F. R. and Hers H. G.: Stimulation of Trypanosoma brucei pyruvate kinase by fructose 2,6-bisphosphate. Eur. J. Biochem. 153: 403–406 (1985)
223. Schmidt H. and Wegener G.: Glycogen phosphorylase in fish muscle: Demonstration of three interconvertible forms. Amer. J. Physiol. 258: C344–C351 (1990)
224. Schöttler U.: Energy metabolism in environmental anaerobiosis. Studies on annelids (In German). Verh. Dtsch. Zool. Ges. 73: 228–240 (1980)
225. Schöttler U.: Further studies on anaerobic energy metabolism in the polychaete, Arenicola marina L. (In German). Zool. Beitr. N. F. 30: 141–152 (1986)
226. Schroff G. and Zebe E.: The anaerobic formation of propionic acid in the mitochondria of the lugworm Arenicola marina. J. comp. Physiol. B 138: 35–41 (1980)
227. Segil N. et al.: Enolase isoenzymes in adult and developing Xenopus laevis and characterization of a cloned enolase sequence. Biochem. J. 251: 31–39 (1988)
228. Shrago E. and Elson C.: Intermediary metabolism of Tetrahymena. In: Levandowsky M. and Hutner S. H. (eds.): Biochemistry and physiology of Protozoa, 2.ed., Vol. 3, pp. 287–312. Acad. Press, New York 1980
229. Siebenaller J. F. et al.: Comparison of the D-lactate stereospecific dehydrogenases of Limulus polyphemus with the active-site regions of L-lactate dehydrogenases. Biochim. biophys. Acta 749: 153–162 (1983)
230. Siebenaller J. F.: Structural comparison of lactate dehydrogenase homologs differing in sensitivity to hydrostatic pressure. Biochim. biophys. Acta 786: 161–169 (1984)
231. Siegmund B. et al.: Alanopine and strombine are endproducts of anaerobic glycolysis in the lugworm Arenicola marina L. (Annelida, Polychaeta). Comp. Biochem. Physiol. Pt. B 82: 337–345 (1985)
232. Srinivasan N. G. et al.: Phosphofructokinase from Dirofilaria immitis. Stimulation of activity by phosphorylation with cyclic AMP-dependent protein kinase. J. biol. Chem. 363: 3482–85 (1988)
233. Glycogen phosphorylase in insects (Review). Insect Biochem. 12: 131–147 (1982)
234. Stephan P., Clarke F. and Morton D.: The indirect binding of triose-phosphate isomerase to myofibrils to form a glycolytic enzyme mini-complex. Biochim. biophys. Acta 873: 127–135 (1986)
235. Stetten M. R. and Goldsmith P. K.: Two hexokinases of Homarus americanus (lobster), one having great affinity for mannose and fructose and low affinity for glucose. Biochim. biophys. Acta 657: 468–481 (1981)
236. Stevens B. R. et al.: Intestinal brush border membrane Na+/glucose cotransporter functions in situ as a homoteramer. Proc. Nat. Acad. Sci. USA 87: 1456–60 (1990)
237. Storey K. B. and Bailey E.: The intracellular distribution of enzymes of carbohydrate degradation in the fat body of the adult male cockroach. Insect Biochem. 8: 73–79 (1978)
238. Storey K. B.: Regulatory properties of hexokinase from flight muscle of Schistocerca americana gregaria. Role of the enzyme in control of glycolysis during rest-to-flight transition. Insect Biochem. 10: 637–645 (1980)
239. Storey K. B.: Kinetic properties of purified aldolase from flight muscle of Schistocerca americana gregaria. Role of the enzyme in the transition from carbohydrate to lipid-fueled flight. Insect Biochem. 10: 647–655 (1980)

240. Storey K. B.: Phosphofructokinase from the overwintering gall fly larva, Eurosta solidaginis: Control of cryoprotectant polyol synthesis. Insect Biochem. 12: 501–505 (1982)
241. Storey K. B. and Dando P. R.: Substrate specificities of octopine dehydrogenases from marine invertebrates. Comp. Biochem. Physiol. Pt. B 73: 521–528 (1982)
242. Storey K. B.: Tissue-specific alanopine dehydrogenase and strombine dehydrogenase from the sea mouse, Aphrodite aculeata (Polychaeta). J. exp. Zool. 225: 369–378 (1983)
243. Storey K. B.: Phosphofructokinase from the foot muscle of the whelk, Busycotypus canaliculatum: Evidence for the covalent modification of the enzyme during anaerobiosis. Arch. Biochem. Biophys. 235: 665–672 (1984)
244. Storey K. B.: Phosphofructokinase from flight muscle of the cockroach, Periplaneta americana. Control of the enzyme activation during flight. Insect Biochem. 15: 663–666 (1985)
245. Storey K. B.: Kinetic and regulatory properties of pyruvate kinase isozymes from flight muscle and fat body of the cockroach Periplaneta americana. J. comp. Physiol. B 155: 339–345 (1985)
246. Storey K. B.: Aspartate activation of pyruvate kinase in anoxia tolerant molluscs. Comp. Biochem. Physiol. Pt. B 83: 807–812 (1986)
247. Sullivan D. T. et al.: Glyceraldehyd-3-phosphate dehydrogenase from Drosophila melanogaster – Identification of two isozymic forms encoded by separate genes. J. biol. Chem. 260: 4345–50 (1985)
248. Surholt B., van der Horst J. and Newsholme E. A.: Energy metabolism in insect flight muscles – Regulation of substrate transport and intermediate metabolism. Zool. Beitr. N. F. 30: 85–102 (1986)
249. Swain M. S. and Lebherz H. G.: Hybridization between fructose diphosphate aldolase subunits derived from diverse biological systems: Anomalous hybridization behavior of some aldolase subunit types. Arch. Biochem. Biophys. 244: 35–41 (1986)
250. Tabachnick W. J. and Howard D. J.: Genetic control of hexokinase variation in insects. Biochem. Genetics 20: 47–57 (1982)
251. Takano T. and Li S. S. L.: Human testicular lactate dehydrogenase-C gene is interrupted by six introns at positions homologous to those of LDH-A (muscle) and LDH-B (heart) genes. Biochem. biophys. Res. Commun. 159: 579–583 (1989)
252. Takenaka M. et al.: Isolation and characterization of the human pyruvate kinase M-gene. Eur. J. Biochem. 198: 101–106 (1991)
253. Tani K. et al.: Human liver type pyruvate kinase: Complete amino acid sequence and the expression in mammalian cells. Proc. Nat. Acad. Sci. USA 85: 1792–95 (1988)
254. Taylor M. and Gutteridge W. E.: The regulation of phosphofructokinase in epimastigote Trypanosoma cruzi. FEBS Letters 201: 262–266 (1986)
255. Thalhofer H. P. et al.: Identification of two different phosphofructokinase-phosphorylating protein kinases from Ascaris suum muscle. J. biol. Chem. 263: 952–957 (1988)
256. Thelen A. P. and Wilson J. E.: Complete amino acid sequence of the type-II isozyme of rat hexokinase. Arch. Biochem. Biophys. 286: 645–651 (1991)
257. Tso J. Y. et al.: Isolation and characterization of rat and human glyceraldehyd-3-phosphate dehydrogenase cDNAs: Genomic complexity and molecular evolution of the gene. Nucleic Acids Res. 13: 2485–2502 (1985)
258. Tsutsumi K. et al.: Nucleotide sequence of rat liver aldolase B messenger RNA. J. biol. Chem. 259: 14572–75 (1984)
259. Tsutsumi R. et al.: Two different aldolase A mRNA species in rat tissue. Eur. J. Biochem. 142: 161–164 (1984)
260. Umezurike G. M. and Anya A. O.: Studies on energy metabolism in the endoparasite Strongyluris brevicaudata (Nematoda). Comp. Biochem. Physiol. Pt. B 59: 147–151 (1978)
261. Umezurike G. M. and Eke L. O.: Some properties of lactate dehydrogenase from the foot muscle of the giant African snail (Achatina achatina). Comp. Biochem. Physiol. Pt. B 75: 263–268 (1983)
262. Ureta T.: The comparative isozymology of vertebrate hexokinases (Minireview). Comp. Biochem. Physiol. Pt. B 71: 549–555 (1982)
263. Vaandrager S. H., van Marrewijk W. J. A. and Beenakkers A. M. T.: Kinetic properties of glycogen phosphorylase a, ab and b from flight muscles of the locust, Locusta migratoria. Insect Biochem. 17: 695–700 (1987)
264. Vandercammen A. and van Schaftingen E.: The mechanism by which rat liver glucokinase is inhibited by the regulatory protein. Eur. J. Biochem. 191: 483–489 (1990)
265. Ventura F. et al.: 6-Phosphofructo-2-kinase/fructose-2,6-bisphosphatase in rat brain. Biochem. J. 276: 455–460 (1991)
266. Vieira M. M., Veiga L. A. and Nakano M.: Muscle D-glyceraldehyde-3-phosphate dehydrogenase from Caiman sp. – II. New data for enzyme characterization. Comp. Biochem. Physiol. Pt. B 79: 427–433 (1984)
267. Vijayasarathy S. et al.: The genes encoding fructose bisphosphate aldolase in Trypanosoma brucei are interspersed with unrelated genes. Nucleic Acids Res. 18: 2967–75 (1990)
268. Vonwyl E.: Lactate dehydrogenase isozymes in the genus Xenopus: Analyses of complex isozyme patterns. Comp. Biochem. Physiol. Pt. B 74: 725–733 (1983)
269. de Vooys C. G. N.: Anaerobic metabolism in sublittoral living Mytilus galloprovincialis Lam. in the Mediterranean – II. Partial adaptation of pyruvate kinase and phosphoenolpyruvate carboxykinase. Comp. Biochem. Physiol. Pt. B 65: 513–518 (1980)
270. Walton M. J. and Cowey C. B.: Gluconeogenesis from serine in rainbow trout (Salmo gairdneri) liver. Comp. Biochem. Physiol. Pt. B 62: 497–499 (1979)
271. Wegener G. et al.: Antagonistic effects of hexose 1,6-bisphosphates and fructose 2,6-bisphosphate on the activity of 6-phosphofructokinase purified from honey bee flight muscle. Biochem. J. 236: 925–928 (1986)
272. Wegener G. et al.: Properties of locust muscle 6-phosphofructokinase and their importance in the regulation of glycolytic flux during prolonged flight. J. comp. Physiol. B 157: 315–326 (1987)
273. Wegener G., Krause U. and Thuy M.: Fructose 2,6-bisphosphate and glycolytic flux in skeletal muscle of swimming frog. FEBS Letters 267: 257–260 (1990)

274. Wegener G., Bolas N. M. and Thomas A. A. G.: Locust flight metabolism studied in vivo by 31P NMR spectroscopy. J. comp. Physiol. B 161: 247–256 (1991)
275. Whitwam R. E. and Storey K. B.: Pyruvate kinase from the land snail Otala lactea. Regulation by reversible phosphorylation during estivation and anoxia. J. exp. Biol. 154: 321–337 (1990)
276. Wissing J. and Zebe E.: The anaerobic metabolism of the bitterling Rhodeus amarus (Cyprinidae, Teleostei). Comp. Biochem. Physiol. Pt. B 89: 299–303 (1988)
277. Wong J. H., Yee B. C. and Buchanan B. B.: A novel type of phosphofructokinase from plants. J. biol. Chem. 262: 3185–91 (1987)
278. Wu K. C. et al.: Molecular isolation and sequence determination of the cDNA for the mouse sperm-specific lactate dehydrogenase-X-gene. Biochem. biophys. Res. Commun. 146: 964–970 (1987)
279. Xie G. and Wilson J. E.: Tetrameric structure of mitochondrially bound rat brain hexokinase: A crosslinking study. Arch. Biochem. Biophys. 276: 285–293 (1990)
280. Yanagawa H. A.: Tissue distribution, purifications, and properties of multiple forms of hexokinase in the silkworm, Bombyx mori. Insect Biochem. 8: 293–305 (1978)
281. Yano H.: Tissue distribution and species difference of the brain type glucose transporter (GLUT3). Biochem. biophys. Res. Commun. 174: 470–477 (1991)
282. Yarbrough P. O. and Hecht R. M.: Two isoenzymes of glyceraldehyde-3-phosphate dehydrogenase in Caenorhabditis elegans. Isolation, properties, and immunochemical characterization. J. biol. Chem. 259: 14711–20 (1984)
283. Yarbrough P. O. et al.: The glyceraldehyde-3-phosphate dehydrogenase gene family in the nematode, Caenorhabditis elegans: Isolation and characterization of one of the genes. Biochim. biophys. Acta 908: 21–33 (1987)
284. Yarlett N. et al.: Hydrogenosomes in a mixed isolate of Isotricha prostoma and Isotricha intestinalis from ovine rumen contents. Comp. Biochem. Physiol. Pt. B 74B: 357–364 (1983)
285. Yarlett N., Lloyd D. and Williams A. G.: Butyrate formation from glucose by the rumen protozoan Dasytricha ruminantium. Biochem. J. 228: 187–192 (1985)
286. Yonezawa S.: Properties of muscle phosphorylase b from a hagfish, Paramyxine atami. J. exp. Zool. 238: 279–285 (1986)
287. Zebe E., Grieshaber M. and Schöttler U.: Environment- and activity-dependent anaerobiosis: The metabolism of invertebrate animals in hypoxia (In German). Biologie in unserer Zeit 10: 175–182 (1980)
288. Zebe E. and Schöttler U.: Comparative studies on environmental anaerobiosis (In German). Zool. Beitr. N. F. 30: 125–140 (1986)
289. de Zwaan A., de Bont A. M. T. and Hemelraad J.: The role of phosphoenolpyruvate carboxykinase in the anaerobic metabolism of the sea mussel Mytilus edulis L. J. comp. Physiol. B 153: 267–274 (1983)

15 Lipids

The term lipid is applied to a multitude of substances of differing chemical structure and biological function whose only common feature is that they are insoluble in water but readily soluble in organic solvents such as chloroform, benzene, diethyl ether or hot ethanol. A large proportion of the lipids of living organisms are bound to proteins and these are normally extracted using a chloroform:methanol mixture; the alcohol component takes up water from the tissue, denatures the proteins and releases the lipids, and the chloroform solubilizes the lipids. The hydrophobic nature of the lipids is the basis for their **main biological functions** as energy reserves, as a critical part of the water-repellent surfaces of terrestrial organisms, and as a component of cellular membranes. Some individual lipids or lipid-like substances have special functions, e.g. as pheromones or hormones (juvenile hormone, steroid hormones), as regulators of hormone action (prostaglandins), as precursors of secondary messengers (phosphatidylinositol), or as recognition signals on the cell surface (glycolipids) [76].

The **water insolubility** of the lipids presents the organism with serious problems of lipid digestion and absorption in the gut, of intra- and extracellular lipid transport, and of the mobilization of lipid reserves. The lipases only express their hydrolytic activity at the interphase between lipid and water; thus, dietary fats are finely emulsified prior to digestion, in the vertebrates by bile acids and alcohols and in the invertebrates by emulsifiers with a rather different chemical structure (p. 646). The resulting hydrolysis products, ionized fatty acids (soaps), 2-monoacylglycerol and lysophosphatides, amplify the emulsification. The end products of fat digestion in the mammalian gut are liquid-crystalline vesicles, which are 80–120 mm in diameter, and smaller micelles. Just how these vesicles and micelles come into contact with the microvilli of the gut mucosa and how their components penetrate the membranes of the gut cells is not yet completely known, even for the vertebrates. In the mammals, a large part of the absorbed fatty acids are incorporated back into triacylglycerol in the gut cells, combined with cholesterol, phospholipids and proteins into "chylomicrons", and released into the lymph [34, 142]. Lipid absorption in the insects appears to follow a completely different scheme. The fatty acids in the larvae of the butterfly *Pieris brassicae* are incorporated into polar lipids before being absorbed by the gut [266].

Lipid transport in vertebrate blood or lymph and in arthropod haemolymph is carried out by specific lipoproteins (as described in Chapter 5). In vertebrate blood, triacyglycerol, phospholipids and cholesterol are bound to plasma lipoproteins, and free fatty acids are bound to serum albumin. The lipophorins of insects carry diacylglycerols as

well as sterols, hydrocarbons and carotinoids. The precursors of the yolk proteins in vertebrates and invertebrates (vitellogenins) aid in lipid transport. There are also specific transport proteins, e.g. for steroid hormones and vitamin D, in the blood of vertebrates and arthropods. Cellular lipid-binding proteins have so far mostly been investigated in the mammals, but are certainly ubiquitous in their distribution. The **fatty acid-binding proteins** (FABPs) of the mammals are small proteins of 12–15 kDa that bind fatty acids and acyl-CoA thioester with high affinity. They are found in all tissues that are able to metabolize fatty acids, e.g. the liver, gut wall, fat tissue, milk glands, and heart and skeletal muscle. There are tissue-specific isoforms of FABPs which are encoded by different genes. I-FABP occurs only in the gut wall, L-FABP occurs in the liver and in the gut wall, and H-FABP (or M-FABP) has its highest concentration in heart muscle but also occurs in skeletal muscle, brain and other tissues. The protein super-family of the FABPs further includes the adipocyte-FABP, the cellular retinol- and retinoic acid-binding proteins, the mammary-derived growth factor, and the myelin protein P2. Thus, this protein family not only is responsible for intracellular fatty acid transport but also is involved, via the binding of prostaglandins and retinoids, in the regulation of cell growth and differentiation. However, other known lipid-binding proteins do not belong to this super-family, e.g. an FABP that occurs together with H-FABP in the kidney, the acyl-CoA-binding protein (ACBP) of the liver, and membrane-bound FABPs that are probably involved in fatty acid transport across cell membranes [96, 117, 268]. FABPs similar to those of the mammals have also been found in the chicken and various teleosts [47, 231, 255]. The flight muscles of the migratory locus *Schistocerca gregaria* contain an FABP which is comparable with mammalian H-FABP in many properties but is present at a much higher concentration in accordance with the extremely high rate of fatty acid oxidation during extended flight [87].

The lipids are extraordinarily variable in their structural and functional characters, and this leads to problems in their **classification and nomenclature**. Different lipid classes are distinguished by the type and number of their components; these are listed in Table 15.1. Representatives of these classes are found in all animal species but vary widely in their relative proportions according to the species, the developmental stage and the physiological state of the animal. Even

Table 15.1. The structural variety of lipids

a Structural variability of the fatty acids
- Number of C atoms
  - Chain length
  - Even or odd number
- Double bonds
  - Number (up to 6)
  - Location in the chain
  - Number of $-CH_2-$ between the double bonds
  - Steric configuration (almost always cis- in animals)
- Substitution by methylation
  - Number
  - Location
  - on the last-but-one C atom (iso-) (Fig. 15.1a)
  - on the last-but-two C atom (anteiso-) (Fig. 15.1b)
  - on internal C atoms
- Substitution with ethyl, propyl or butyl groups
- Cyclopropane fatty acids (cis-methyl substituted) (Fig. 15.1 d)
- Hydroxy fatty acids (hydroxy substituted)
  - Location of the hydroxy group
- Keto acids (oxo- substituted) (Fig. 15.16b)
- Dicarbonic acid

b The main classes of lipids

Non-hydrolysable lipids
- Fatty acids (variable in structure, as above)
- Long-chain aldehydes (arise from plasmalogens)
- Long-chain alkanols (fatty alcohols)
  - Number of C atoms, double bonds and methyl branches vary as in the fatty acids
- Long-chain diols with different locations of the two hydroxyl groups
- Aliphatic hydrocarbons
  - Number of C atoms, double bonds and methyl branches vary as in the fatty acids
- Squalene and pristene (Fig. 15.12), dolichols (Fig. 15.28), sterols (Chapter 16), and carotinoids (Chapter 19)

Neutral fats
- Mono-, di-, and triester waxes (Fig. 15.20a–j)
- Mono-, di-, and triacylglycerols
- 1-Alkyl-2,3-acylglycerol
- (Sterol ester)

Phospholipids
- Glycerophospholipids (Fig. 15.20a–j) and glycerophosphonolipids (Fig. 15.23a–b)
- Sphingophospholipids (sphingomyelins) (Fig. 15.20 k–l) and sphingophosphonolipids (Fig. 15.23 c–d)

Glycolipids
- Glyceroglycolipids, e.g. seminolipid (Fig. 15.25b)
- 1-Alkyl-2-acyl-3-(2,3-diacylglycerol)-glycerol
- Neutral glycosphingolipids = glycosyl ceramides (cerebrosides) (Fig. 15.25a)
- Acidic glycosphinoglipids
  - Sulphatides (sulphoglycosyl ceramides)
  - Gangliosides (contain sialic acid)
- Glycosyldolichols (Fig. 13.17)

the different organs of an animal may have widely different lipid compositions. The aliphatic lipid components, i.e. the fatty acids and derived long-chain alcohols and hydrocarbons, also exist in various forms with differences in chain length, the number and position of double bonds, branches and substituents. Because the components of lipids can occur in almost any combination, each class of lipids contains an almost astronomical number of molecular species. This large variety of possible lipid structures allows each organism to adapt its lipid spectrum to specific conditions. However, as only a part of the necessary components can be produced by the animal itself, the composition of lipids in the body is strongly influenced by the diet.

One or more **lipid classes** is involved in each of the three **main functions** of the lipids referred to above. The fat stores that serve as an energy reserve consist in most animals of triacylglycerols, although in some marine species they may also include large amounts of ether glycerides, wax esters or the hydrocarbon squalene. The lipid-containing secretions used to nourish young animals, e.g. mammalian milk, the crop milk of the pigeons or the corresponding secretions of some insect species, are, in principle, similar in composition, although they may differ quite specifically in detail. Wax esters and other hydrocarbons are typical of the water-repellent lipid layers of mammal, bird and arthropod body surfaces. In addition to sterols, the construction of cellular membranes involves, above all, glycerophospholipids, sphingophospolipids and glycolipids.

Thus, the subdivisions of this Chapter are based upon these three main functions without losing sight of the structural classification. However, we will consider only lipids that are directly related metabolically to long-chain fatty acids, i.e. the fatty acids themselves, the eicosonoates, fatty alcohols and hydrocarbons derived during metabolism, and the lipids with fatty acids as components. The sterols and sterol esters will be discussed in Chapter 16 and the non-polar substances that act as pheromones, the carotinoids and the terpenes will be considered in Chapter 19.

15.1 Chemistry and Metabolism of the Fatty Acids

15.1.1 Structure and Nomenclature of the Fatty Acids

The fatty acids found in the body are in part taken up in the diet, either unaltered or mostly changed by metabolism, and in part synthesized de novo. Thus, the structure of the fatty acids becomes clear by considering their metabolism. The biosynthesis normally begins with acetyl-CoA as the "primer"; acyl-CoA is extended at each cycle by a C_2 unit and is usually released from the synthase complex only on reaching a length of 14–18C atoms. The fatty acids obtained in the diet or by de novo synthesis can be further extended in an elongation system by one or more C_2 units. This process explains why typical animal fatty acids have chain lengths of C_{14} to C_{22} and an even number of C atoms (Table 15.2).

Fatty acids with an odd number of C atoms arise from propionyl-CoA as primer. Methyl-branched fatty acids (Table 15.1 and Fig. 15.1a–c) are formed by fatty acid synthase using methyl-substituted primers of elongation substrates. Isovaleryl-CoA (from leucine) as the primer gives rise to iso- fatty acids with an odd number of C atoms; iso-butyryl-CoA (from valine) produces iso- fatty acids with an even number of C atoms; and 2-methyl-butyryl-CoA (from isoleucine)

Table 15.2. Numerical symbols, systematic names and trivial names of some important fatty acids

Symbol	Systematic name	Trivial name
4:0	Butanoic	Butyric
5:0	Pentanoic	Valeric
6:0	Hexanoic	Caproic
8:0	Octanoic	Caprylic
10:0	Decanoic	Capric
12:0	Dodecanoic	Lauric
14:0	Tetradecanoic	Myristic
16:0	Hexadecanoic	Palmitic
16:1(n-7)	9-Hexadecanoic	Palmitoleic
18:0	Octodecanoic	Stearic
18:1(n-7)	1-Octadecenoic	Vaccenic
18:1(n-9)	9-Octadecenoic	Oleic
18:2(n-6)	9,12-Octadecadienoic	Linoleic
18:3(n-3)	9,12,15-Octadecatrienoic	α-Linolenic
18:3(n-6)	6,9,12-Octadecatrienoic	γ-Linolenic
20:0	Eicosanoic	Arachidic
20:4(n-6)	5,8,11,14-eicosatetraeneic	Arachidonic
22:0	Docosanoic	Behenic
24:0	Tetracosanoic	Lignoceric
24:1	15-Tetracosenoic	Nervonic

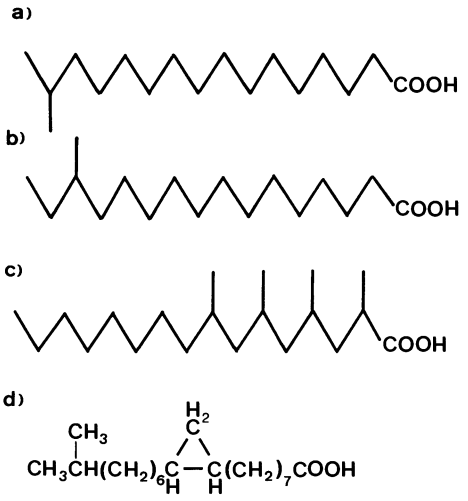


Fig. 15.1a–d. The structure of unusual fatty acids in animals. **a** Iso-heptadecanoic acid; **b** anteiso-heptadecanoic acid; **c** 2,4,6,8-tetramethyl-hexadecanoic acid from the uropygial gland of the goose [108]; **d** cyclopropane fatty acid (17-methyl-cis-9,10-methylene octadecanoic acid) from the parasitic flagellate *Herpetomonas megaseliae* [97]

results in anteiso-fatty acids with an odd number of C atoms (Fig. 15.6). Methylmalonyl-CoA, instead of malonyl-CoA, as the elongation substrate gives rise to internal methyl branches; methyl transfer from methionine to fatty acids, as known from the bacteria, apparently does not occur in animals. Ethyl-, propyl-, and butyl-substituted fatty acids, as found in lipids of the uropygial gland, are derived from the corresponding substituted elongation substrates [108]. Double bonds can be introduced into the fatty acid chain by specific desaturases (Fig. 15.7). These always have the cis configuration; trans isomers are only produced in bacteria and plants but may find their way into animal lipids via the diet or as products of gut symbionts. The house cricket *Gryllus domesticus* is reported to contain 6% elaidic acid, the trans isomer of oleic acid [144]. Each cis double bond results in a kink in the hydrocarbon chain.

The fatty acids have **systematic names** and the most common also have **trivial names**. Many tables and texts use a **digital code** in which the first digit gives the number of C atoms and the second digit gives the number of double bonds (Table 15.2). The localization of the double bonds is also indicated for unsaturated fatty acids, e.g. 18:3(6,9,12) for γ-linolenic acid, where (6,9,12) represents the numbers of C atoms relative to 1-carboxyl from which double bonds extend. Because the elongation of the fatty acids occurs at the carboxyl end, the elongation of 18:3(6,9,12) gives rise to 20:3(8,11,14). This complicated change in the numbering is avoided if the position of the double bonds is expressed relative to the methyl end. Because the double bonds in the majority of unsaturated animal fatty acids are almost always separated by a CH_2 group, notation of the double bond closest to the methyl end is sufficient. Thus, γ-linolenic acid and its elongation product are described by 18:3(n-6) and 20:3(n-6); these are sometimes written as 18:3ω6 and 20:3ω6. This nomenclature system is especially useful in that mutual conversion of the (n-3), (n-6) and (n-9) families is not possible in animal metabolism.

Typical animal fatty acids are unbranched, even-numbered with 14–22 C atoms, and saturated or unsaturated with one to six double bonds. Most frequent are the C_{16} and C_{18} fatty acids, with palmitic acid 16:0 being the predominant C_{16} form and oleic acid 18:1(n-9) and polyunsaturated fatty acids (PUFA) constituting the predominant C_{18} species. Also commonly found is vaccenic acid 18:1(n-7), which arises by elongation of palmitoleic acid 16:1(n-7) and in some invertebrates is equivalent to, or exceeds, oleic acid in concentration [101]. Species-specific differences in the substrate specificity of the Δ^9-desaturase probably determine whether stearic acid 18:0 is desaturated to oleic acid 18:1(n-9) rather than palmitic acid 16:0 to palmitoleic acid 16:1(n-7). Noteworthy in this context is the large proportion of the unusual fatty acid 20:1(n-11) in the triacylglycerols of the cockchafer *Melolontha vulgaris* [251].

The **proportion of unsaturated fatty acids**, which determines the melting point of body fat, is dependent upon the diet but also varies greatly with the species (Table 15.3). The storage fat of the ruminants is high in saturated fatty acids, especially stearic acid (18:0), because the unsaturated fatty acids in the diet are completely saturated by the rumen symbionts. Compared with the mammals, the teleosts have a significantly higher proportion of polyunsaturated fatty acids, in particular 18:2(n-6), 20:4(n-6), 20:5(n-3), 22:5(n-3) and 22:6(n-3) [14]. The body fats of the amphibians, reptiles and birds are intermediate in this respect. Large amounts of polyunsaturated fatty acids are also found in many invertebrate animals, e.g. the molluscs, echinoderms, ascidians and free-living nematodes (Table 15.3) [101, 103]. There are many interesting connections between habitat and fatty acid spectrum, probably based upon differences in the composition of the diet. For example, marine mussels always have more (n-3) than (n-6) fatty acids,

Table 15.3. Fatty acid composition (%) of the total lipids of some animals

Fatty acid	1	2	3	4	5	6	7
14:0	1.9	6.8	2.1	16	1.0		
14:1	0.4						
15:0	0.3	0.4	0.7		0.1		
16:0	6.8	22	14.8	23	2.4	7.9	8.0
16:1	11	5.6	5.0	29	4.6		1.4
16:2	0.2	0.5	0.8				
17:0	0.5	0.2			0.6		
17:1	0.4	0.6	0.2				
18:0	0.9	1.5	3.5	3.4	8.1	13.0	11.8
18:1	30	23.2	3.6	26	32.4	16.0	29.0
18:2	0.9	6.0	1.7	3.2	8.5	22.9	46.7
18:3	0.1	3.5	1.5		0.2		3.1
18:4	0.5	2.0	1.8				
20:0					0.7	2.3	
20:1	13.8	0.7	12.3		0.7	0.3	
20:2	0.2	0.6	0.8		2.5	1.0	
20:3			0.8		8.4	0.7	
20:4	0.8	1.4	3.9		11.5	3.3	
20:5	6.5	9.7	14.5		8.7		
22:0						10.9	
22:1	2.7	0.2				0.3	
22:2			4.6			0.8	
22:4	0.7	0.8	0.1				
22:5	7.8	0.2	1.8			0.6	
22:6	12.6	12	23			14.1	
24:1	0.4	0.2				2.1	
Branched		0.9			8.5		
Odd	1.2	1.2	0.9		0.7		
Saturated	10	32	21	42	21	34	20
Monoenic	59	31	21	55	38	19	30
Polyenic	30	37	55	3	40	44	50

1, Seal *Phoca vitullina*, subcutaneous fat tissue [280]; 2, herring *Clupea harengus* [152]; 3, edible mussel *Mytilus edulis* [292]; 4, *Drosophila melanogaster*, 38 h old [146]; 5, *Caenorhabditis elegans* [101]; 6, larvae of the tapeworm *Taenia crassipes* [174]; 7, *Trypanosoma cruzi* [262]

whereas the (n-6) family predominates by two- to fourfold in freshwater mussels and by six- to ninefold in the pulmonate snail *Cepaea nemoralis* [78]. In the insects, the saturated, monoenic and dienic acids are by far the most common [146], but higher unsaturated C_{18}, C_{20} and C_{22} fatty acids are also present. Their accumulation in specific organ lipids indicates some special functions. Thus, for example, the phosphatidylethanolamine from the retina of the moth *Deilephila elpenor* contains 40% 20:5(n-3), and the phosphatidylcholine from the spermatophores of the cricket *Teleogryllus commodus* contains 25% 20:4(n-3) [252].

In addition to, or instead of, the typical acids, many animal lipids contain fatty acids of unusual structure; this is a particularly interesting subject for comparative biochemistry. **Fatty acids with an odd number of C atoms** are usually found in only trace amounts in animal fats (Table 15.3), and probably originate in bacteria. However, the storage fat of the teleost *Mugil celphalus* contains no less than 25 C_{15} and C_{17} acids. The fatty acid 23:1(n-9) is found in various echinoderms in amounts of up to 5% of the total fatty acids, and may arise from the α-oxidation of 24:1(n-9) [119]. **Short-chain fatty acids** of C_8 to C_{12} are found in the milk fat of all terrestrial mammals, and butyric acid C_4 occurs only in the ruminants (Table 15.4). **Very long-chain fatty acids (VLCFAs)** are widely found in animals. In the marine invertebrates, they originate in the algae of the diet. In beeswax, 25% of the fatty acids have more than 30 C atoms. VLCFAs are also present in the skin fats of the mammals, and very long-chain polyunsaturated fatty acids (VLCPUFAs) are found in the retina of mammals and teleosts. For example, the cod retina contains 15% 32:6(n-3) in addition to 65% 22:6(n-3) [220]. Unusually large saturated or monounsaturated fatty acids are found in the glycosphingolipids of mammalian brain: 22:0 (behenic acid), 24:0 (lignoceric acid), 24:1(n-9) (nervonic acid) and the hydroxy acid 2-OH-24:0 (cerebronic acid). Human testis tissue and spermatozoa predominantly contain two- to fourfold unsaturated fatty acids with up to 32 C atoms, and those of bulls, boars and rams have five or six double bonds with as many as 34 C atoms; lipid analysis of ram spermatozoa indicates that these

Table 15.4. The fatty acid spectra of various milk fats [66, 73, 159]. The data are expressed as percent by weight of the total fatty acids

Fatty acid	Echidna	Cow	Horse	Mouse	Seal[a]
Saturated					
4:0	–	3–4	Tr	–	–
6:0	–	1–2	1	–	–
8:0	–	1	3	0.6	–
10:0	–	2–3	5	5	–
12:0	–	2–4	6	8	–
14:0	1	8–13	7	11	3
16:0	26	25–32	16	23	16
18:0	7	8–13	3	1.8	3
Unsaturated					
14:1	–	1	2	0.2	2
16:1	9	2–5	7	1.8	13
18:1	38	27–34	18	17	}37
18:>1	9	3–4	24	28	
20:≥1	–	} 1	} 5	2.5	14
22:≥1	–			<0.1	13

[a] Horsehead seal, *Halichoerus grypus*
Tr, trace

are located exclusively in the sphingomyelins [207]. The glycerophospholipids of the photoreceptor membranes in the vertebrate retina, in contrast to those of other cell membranes, mostly contain two highly unsaturated fatty acids. These are similar in structure from the mammals to the fish and have up to 36 C atoms and six double bonds [227]. Mammalian brain contains unusual molecular species of phosphatidylcholine which carry fatty acids with 32–38 C atoms and four to six double bonds at position sn-1, and may play some as yet undefined role in brain physiology [226]. The fatty acid spectrum of the Porifera appears to be fundamentally different to that of all eumetazoans. The Demospongia examined predominantly contain long-chain C_{24} to C_{30} fatty acids which are all 5,9-unsaturated and sometimes branched. It could be shown in *Jaspis stellifera* that 26:2(5,9), iso-27:2(5,9) and anteiso-27:2(5,9) arise from 16:0, iso-15:0 and anteiso-15:0, respectively, by elongation. The unusual desaturation apparently occurs only after elongation [33]. *Cliona celata* also contains 30:4(n-6) and 30:5(n-3), i.e. fatty acids with double bonds at normal locations. The fatty acids of several sponge species carry unusual substituents such as cyclopropyl, methoxy, acetoxy or cyclic peroxide. The marine species *Dictyonella incisa* has furan fatty acids esterified with sterols which release histamines; their inflammatory effect probably serves to deter predators [46].

Animals also occasionally have PUFAs in which neighbouring double bonds are not interspersed with CH_2 groups, as is usually the case. Fatty acids with conjugated double bonds are found only in plants, whereas fatty acids with widely separated double bonds, known as **NMI fatty acids** (non-methylene interrupted), have been detected in various animals. In marine gastropods and bivalves, fatty acids such as 20:2(5,11), 20:2(5,13), 22:2(7,13) and 22:2(7,15) make up more than 15% of the total [102, 292]. These include the 5,9-unsaturated fatty acids of the sponges, as mentioned above [33]. Whole series of 5-unsaturated fatty acids in proportions of up to 6.5% are present in starfish and other echinoderms, e.g. 20:1(5), 20:2(5,11), 20:2(5,13), 20:3(5,11,14), 20:4(5,11,14,17) and 20:5(5,11,14,17,20), together with 7-unsaturated fatty acids like 22:2(7,13) and 22:2(7,15) [134]. The fatty acid 20:3(5,11,14), found in the domestic cat and related species, the echidna and some insects, arises from linoleic acid 18:2(n-6) when reduced activity of the Δ^6-desaturase prohibits the normal pathway via 20:3(n-6) to arachidonic acid 20:4(n-6) (Fig. 15.7). Linoleic acid 18:2(9,12) is in these animals elongated to 20:2(11,14), and Δ^5-desaturated to 20:3(5,11,14) [187]. The fatty acid spectra of the marine invertebrates are all particularly complex. For example, 56 different fatty acids have been identified in the amphipod *Themisto gaudichaudii* and 55 have been identified in a holothurian [68, 119].

Fatty acids with branched chains (Fig. 15.1a–c) are produced by many bacteria and are taken up by animals in their diet. Quite considerable quantities of iso- and anteiso- fatty acids have been found in some animal fats, e.g. over 6% in the body fat of the teleost *Fundulus heteroclitus* [50], 3–7% in different phospholipids of the nematode *Turbatrix aceti*, and more than 20% in the lipids of rat faeces. The endogenous formation of branched fatty acids may be detected in some animal tissues. This is true, for example, for the tissues of the forehead projection (melon) of the toothed whale, the avian uropygial gland and the sebaceous glands of the mammals. The melon functions as a "sound lens" in echo navigation in that differences in concentration and composition of the lipid components result in a higher velocity of sound transmission in the inner layers of the melon compared with the outer layers (Table 15.5). The triacylglycerols and wax esters of the melon in the dolphins (Delphinidae), bottlenosed dolphins (Phocaenidae) and narwhales (Monodontidae) contain considerable quantities of isovaleric acid iso-5:0, but this is not the case for the river dolphins (Platanistidae), sperm whales (Physeteridae) and beaked whales (Ziphiidae) [180]. The avian uropygial glands and the skin fats of mammals contain an extraordinarily complex mixture of branched fatty acids and derived long-chain alcohols. In addition to iso- and anteiso-methyl-branched fatty acids and fatty alcohols,

Table 15.5. The content and composition of the lipids in the forehead projection ("melon") and subcutaneous tissue ("blubber") of the dolphin *Inia geoffrensis* [153]

	Inner melon	Outer melon	Blubber
Lipid content (% fresh weight)	48–95	24–66	13–43
Composition (%)			
triacylglycerols	36–68	80–95	96–99
wax esters	32–64	5–20	1– 4
Number of C atoms			
triacylglycerols	40–42	42–44	46–48
wax esters	28–29	30–31	32–34
Velocity of sound transmission		declining	

there are also molecules which carry up to four or five methyl groups on internal C atoms of the chain; titmice (Paridae) have fatty acids with one or two ethyl substitutions, and owls possess 2-propyl- and 2-butyl- substituted fatty acids [108].

Cyclopropane fatty acids (Fig. 15.1 d) are formed in bacteria by the attachment of an activated methyl group from S-adenosylmethionine to a fatty acid double bond, and they are taken up into animal lipids in the diet. This explains the presence of cyclopropane fatty acids in ruminants, and the detection of 0.2 % cis-9,10-methylene-16:0 and -18:0 in the body fat of the teleost *Fundulus heteroclitus* [50]. In contrast, endogenous production of cyclopropane fatty acids has been demonstrated in several Kinetoplastida, e.g. a cyclopropane- and methyl- substituted stearic acid (Fig. 15.1 d) in *Herpetomonas megaseliae* [97]. Remarkable in this respect is the large amount of cyclopropane fatty acids with 17, 18 or 19 C atoms in millipedes of the order Spirostreptida. They constitute as much as 30 % of the total fatty acids in female animals of the species *Graphidostreptus tumuliporus*, as high as 35 % in the eggs, and about 20 % in the related species *Orthopterus ornatus*. Males of both species, on the other hand, show no trace of these unusual fatty acids and their significance in the female is unknown.

Hydroxy fatty acids are found only occasionally in the animal kingdom. Cerebronic acid (2-OH-24:0) from the glycosphingolipids of mammalian brain was mentioned earlier. 2-Hydroxy and 3-hydroxy fatty acids (earlier known as α- and β-hydroxy acids) are constituents of diester waxes in the uropygial gland lipids [108]. C_{30} to C_{34} fatty acids with terminal hydroxy groups (ω-hydroxy fatty acids) have been detected in the epidermial fats of various mammals [278]. 3,7,13-Trihydroxy- and 2,3,7,13-tetrahydroxystearic acid are components of the unusual taurolipids (Fig. 15.19a) from the ciliate *Tetrahymena* [125].

The **fatty acid spectrum** of all tissues includes a very large number of different molecular species, the complete description of which is an almost impossible task even with present-day highly sophisticated analytical methods. The spectrum varies widely not only between animal groups and species and between the organs of any one animal, but also between the different lipid classes of the same organ. For example, phosphatidylcholines are often rich in C_{16} acids, whereas arachidonic acids, which are important as precursors of prostaglandins, are mainly bound to phosphatidylinositol [14]. Differences in the fatty acid spectrum are even found between the **molecular positions** within the same class of glycerolipids. A description of these relationships is given by the stereospecific (sn-) numbering system of the glycerol C atoms. For this purpose, the molecule is orientated such that the secondary OH- group is on the left, and the uppermost C atom is labelled sn-1 (Fig. 15.2). As a rule, the primary sn-1-OH of the glycerophospholipids is esterified with a saturated fatty acid, and the secondary sn-2-OH is esterified with an unsaturated fatty acid [63, 89]. In the teleosts, with their high content of unsaturated fatty acids, glycerophospholipids are found with two unsaturated acids [14]. The triacylglycerols may show differences not only between the primary and secondary hydroxyls but also between sn-1 and sn-3. For example, the triacylglycerols of the larvae and adults of the Mediterranean fruit fly *Ceratitis capitata* are markedly asymmetrical with saturated fatty acids on sn-1, unsaturated fatty acids on sn-2 and both saturated and unsaturated fatty acids on sn-3. In contrast, the triacylglycerols of the eggs are more often symmetrical with unsaturated fatty acids also present on sn-1.

Random distribution amongst the lipids of the many fatty acids present in each cell would result in an astronomical number of different molecular species. For example, n^2 different molecules of the glycerophospholipids with their two ester bonds would result from n different fatty acids, and, correspondingly, n^3 triacylglycerols would be produced. The restrictions described above limit the **variety of lipid molecular species** but the range is still enormous. Each species, organ and cell type has a unique spectrum of lipid molecules. Here again – as always when one considers the confusing molecular variety of different organisms – the question arises whether the observed differences are adaptations to specific internal and external conditions or are simply the result of chance combinations of factors, e.g. diet, the availability of structural components, or variation

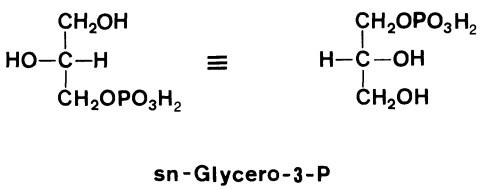


Fig. 15.2. For the stereo-specific (sn-) numbering of the C atoms of glycerol the molecule is orientated so that the secondary hydroxyl group is on the left; the first C atom above then receives the number sn-1. The natural phosphoric acid ester (formerly α-glycerophosphate) should accordingly be referred to as glycerol-sn-3-phosphate

in the enzymes of lipid metabolism. In some cases, the lipid molecules must satisfy very specific functions with corresponding structural specifications, whereas in other cases their biological importance only involves simple physicochemical properties found in molecules of very different structure.

15.1.2 Biosynthesis of Fatty Acids

The de novo synthesis of fatty acids can take place in various organs of the mammals: in particular, in the fat tissues of the cow, sheep, goat, pig, dog, cat and guinea-pig, in the human liver, and in both tissues in the rat, mouse and rabbit. Acetate is more important than glucose as a starting substance in the ruminants, as it is also in the pig, dog and cat [223, 250]. Biosynthesis follows more-or-less the same **scheme** in all organisms (Fig. 15.3). In the primer reaction, an acyl residue (usually acetyl) is transferred from acyl-CoA to the acyl-binding protein (ACP) and further to the SH group of β-ketoacyl-CoA synthetase (condensing enzyme) (reactions 1 and 2). This acyl group forms the methyl end of the finished fatty acid. In reactions 3 and 4, the CoA-bound elongation substrate (usually malonyl-CoA) binds to ACP and is linked to the primer acyl with cleavage of a CO_2. The resulting β-ketoacyl residue is converted by two reduction steps and a dehydration (reactions 5–7) to an acyl, which is now two C atoms longer than the primer. Transfer of the extended acyl residue to the condensing enzyme (reaction 2) initiates the next synthesis cycle. The long-chain acyl residue that results from several passages of the cycle is released by the action of thioesterase (reaction 8). ACP and the enzyme activities described are found in *E. coli* and plants as single, monofunctional subunits. In animals, the **fatty acid synthase** (FAS) is a multi-enzyme complex in which ACP and all seven enzyme activities are localized on one polypeptide. The FAS of yeast is also complex but is composed, according to the formula $\alpha_6\beta_6$, of two different subunits encoded by different genes. The α subunit (1894 amino acids) consists of ACP and two enzyme activities in the order ACP-KR-KS (see Fig. 15.4 for the abbreviations), and the β subunit (1980 amino acids) is made up of the remaining five enzyme activities in the order ER-AT-DH-MT-TE [42, 177].

The animal FASs are **homodimers** with molecular masses of 450–500 kDa in the vertebrates and invertebrates [195, 219 272]. The complete sequences are known for the FAS subunits from rat milk glands (2505 amino acids) and chicken liver (2446 amino acids), and partial sequences are known for the FAS from the mouse and goose [98, 118, 203, 284]. The sequence agreement between the chicken and rat is 67 % and between

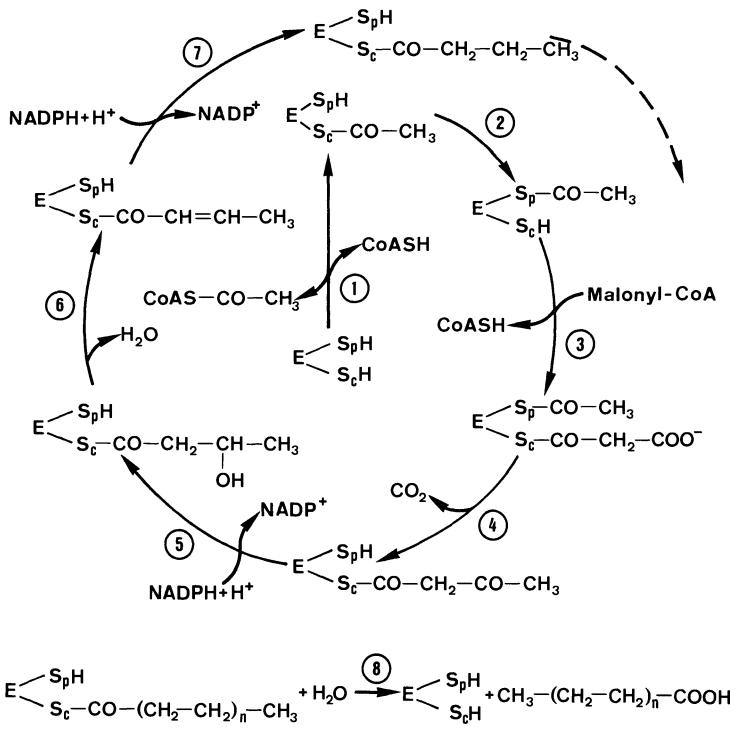


Fig. 15.3. Biosynthesis of the fatty acids [195]. The fatty acid synthase complex *E* carries two reactive SH- groups: -S_cH on the acyl-binding protein (ACP), and -S_pH on the condensing enzyme. The individual reaction steps are: *1* and *2*, acyl transfer; *3*, malonyl transfer; *4*, condensation; *5*, keto reduction; *6*, dehydration; *7*, enoyl reduction; *8*, thioester cleavage

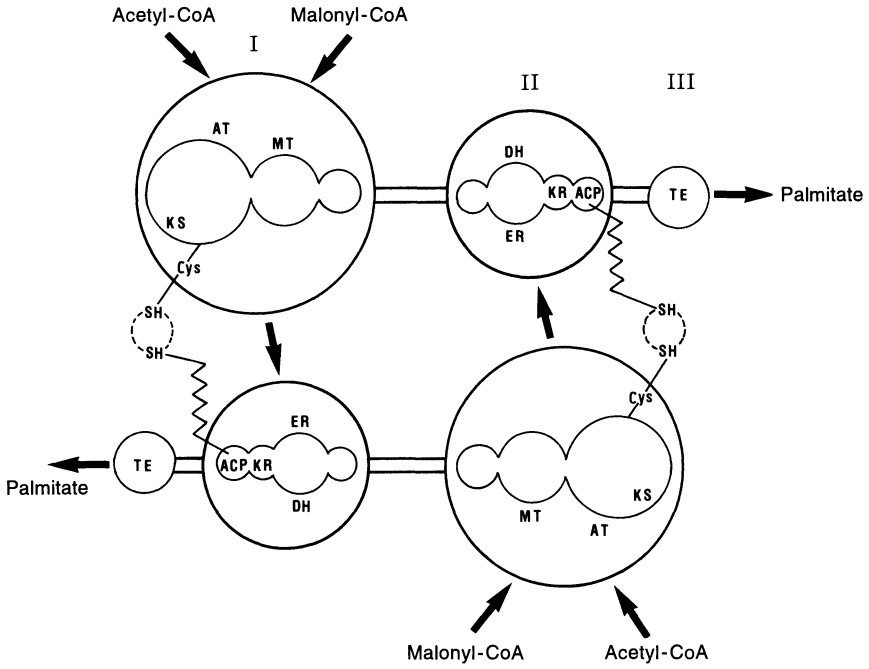


Fig. 15.4. The fatty acid synthase complex of the chicken is a dimer of two identical, oppositely orientated (head-to-tail) subunits. The seven enzyme functions and the acyl-binding protein function (ACP) are distributed amongst three domains: domain *I* (elongation) of each subunit interacts functionally with domains *II* (reduction) and *III* (palmitate release) of the other subunits [264]. *AT*, Acetyl transacylase; *MT*, malonyl transacylase; *KS*, β-ketoacyl synthase; *KR*, β-ketoacyl reductase; *DH*, dehydratase; *ER*, enoyl reductase; *TE*, thioesterase; *ACP*, acyl carrier protein

different FAS subdomains of chicken and yeast it is 14–31 % [36]. The polypeptide chain is subdivided into three **domains** (Fig. 15.4). Domain I (127 kDa) includes the activities of β-ketoacyl synthase (KS) and the acetyl and malonyl transacylases (AT and MT). This is the attachment point for the primer and elongation substrate, and is responsible for the elongation reactions. Domain II (107 kDa) includes the β-ketoacyl and enoyl reductases (KR and ER), the dehydratase (DH) and ACP; the acyl residue remains bound to ACP during the reduction reactions on this domain. Binding is via the same 4′-phosphopantetheine structure found in CoA (Fig. 15.5). When the growing chain reaches a length of 16–18 C atoms, it is released by the thioesterase activity (TE) localized on domain III (33 kDa). The two subunits are arranged head-to-tail such that the pantetheine SH of one subunit lies opposite the cysteine SH of KS on the other subunit. Thus, the functional unit of fatty acid biosynthesis is composed of domain I of one subunit and domains II and III of the other subunit [264]. Dissociation to monomers inactivates the whole reaction but this is reversible [195, 246]. Immunological comparisons show that FASs from liver, milk gland and fat tissues of the same species are identical, that the FASs of different mammals are parti-

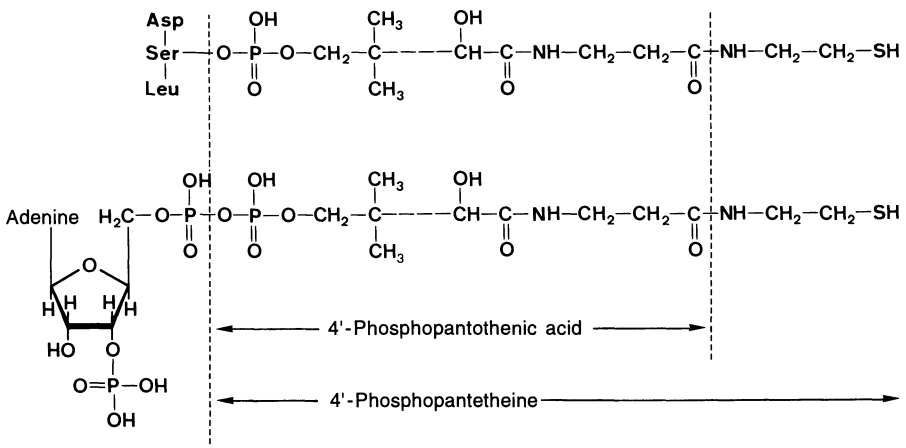


Fig. 15.5. The acyl-binding proteins (ACP) of fatty acid synthase *(above)* and CoA *(below)* contain the same 4′-phosphopantetheine structure [195]

ally identical, and that little cross-reactivity exists between FASs from birds and mammals [195].

FAS is localized in the cytosol; the necessary **acetyl-CoA**, however, arises predominantly in the mitochondria, the inner membranes of which are impermeable to acetyl-CoA. Transport of acetyl residues out of the mitochondria into the cytosol in the rat is about 80 % in the form of citrate, which is produced in the mitochondria by citrate synthase and cleaved in the cytoplasm by citrate lyase. This mechanism has also been demonstrated in the fat bodies of the locust *Locusta migratoria*. Acetyl carnitine as a transport form is of relatively minor importance [63]. There are characteristic species-specific differences in the origin of the acetyl-CoA and NADPH required for fatty acid synthesis in the milk glands of various mammals. In contrast to other species, ruminants do not incorporate ^{14}C-glucose into fatty acids; free acetate that can be turned over by acetyl-CoA synthetase is apparently more important in this case. Furthermore, the NADPH-dependent isocitrate dehydrogenase here is cytosolic and not mitochondrial as in other mammals. In termites, acetate delivered by symbionts can be converted to acetyl-CoA, but pyruvate dehydrogenase, by means of which carbohydrates can be introduced into fatty acid synthesis, is missing [196].

The elongation substrate **malonyl-CoA** is provided by **acetyl-CoA carboxylase**, which has been isolated from fatty acid-synthesizing tissues of the mammals and birds as well as from higher plants, yeast and bacteria. The enzyme is localized in the cytoplasm and contains one biotin per monor
The monomeric form is in
form is filamentous wi
enzyme from chi
acids, corres
nificar
h

the pigeon and goose, to 60:40 in the chicken. Insect FAS also produces shorter fatty acids, e.g. in the Mediterranean fruit fly *Ceratitis capitata* there are fatty acids of 10–18 C atoms, and in *Drosophila melanogaster* 14:0 occurs as well as 16:0 and 18:0 [219]. Pulse-labelling experiments in *Leishmania donovani* have shown that the primary products are 12:0, 14:0 and 16:0 [112]. In the species-specific very variable fatty acid spectrum of milk (Table 15.4), the rat, rabbit and goat have considerable quantities of intermediate-length fatty acids (8:0–12:0). However, isolated milk-gland FASs of the rat and rabbit only produce short- and long-chain fatty acids; those of intermediate length arise by premature release from the FAS complex through the action of a soluble, 29-kDa thioesterase II. This soluble enzyme is absent from the milk glands of the goat and other ruminants, and in this case the formation of intermediate-length fatty acids involves deviations in the specificity of the FAS thioesterase domain itself [195].

Soluble thioesterases that react only with acyl-CoA are widely found in animals and plants. In contrast, a second type of enzyme that can also cleave FAS-bound fatty acids is found in the milk glands of some mammals and, in particular, in the uropygial glands of many avian species. Here, they are also responsible for the release of shorter fatty acids from the FAS complex [40, 195]. Thus, the uropygial gland of the pigeon contains no fewer than four TE activities: the relevant FAS domain, an acyl-FAS specific enzyme and two soluble acyl-CoA hydrolases [40]. The sequence
ble acyl-FAS TE from the mallard *Anas*
at least in the region follow-
about 60 % with the TE
ble TE has also been
thosiphon pisum,
mation of myr-
es up 70 % of
glycerol-bound

the primer or
S accepts ace-
qually as pri-
CoA are also
y of the TE
ict of even-
15:0 and 17:0
yryl-CoA is
AS of chicken
tty acids that
e, isoleucine
anched fatty

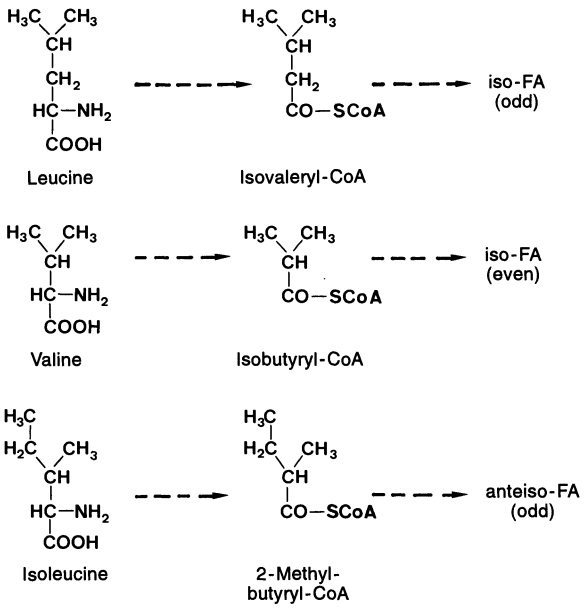


Fig. 15.6. Leucine, valine and isoleucine provide the primers for the biosynthesis of iso- and anteiso- fatty acids with an even or odd number of C atoms

acids (Fig. 15.6); this has been demonstrated in tracer experiments with tissue slices of the melon from the toothed whale *Stenella caeruleo-alba* [180]. Internal methyl branching is produced with methylmalonyl-CoA as the elongation substrate; this has been shown by tracer experiments on the Harderian glands of the guinea-pig and on the uropygial glands of the goose and duck [108, 132, 195]. Methylmalonyl-CoA is formed from propionyl-CoA by acetyl-CoA carboxylase. The ethyl-, propyl- or butyl-branched fatty acids found in the uropygial glands of some birds probably arise in a similar way with the corresponding substituted elongation substrates [108]. The fatty acid precursors of hydrocarbons with internal methyl branching, widely found in the cuticular lipids of insects, are produced in a comparable process. For example, tracer experiments with the termite *Zootermopsis angusticollis* and the cockroach *Periplaneta americana* demonstrated the incorporation of propionate and succinate into branched hydrocarbons, probably via methylmalonyl-CoA as the intermediate [81]. *Drosophila* FAS can utilize acetyl-CoA or propionyl-CoA, but neither butyryl-CoA, isobutyryl-CoA nor isovaleryl-CoA, as primer. The enzyme is absolutely specific for the elongation substrate malonyl-CoA and is inactive with methylmalonyl-CoA. However, the presence of both elongation substrates simultaneously results in methyl-branched fatty acids; methylmalonyl-CoA acts as a competitive inhibitor to malonyl-CoA [219].

The uropygial glands of the goose *Anser anser* and of the duck *Anas platyrhynchos* produce large quantities of multiply methyl-branched fatty acids, despite the fact that the affinity of the FAS for malonyl-CoA is significantly higher than its affinity for methylmalonyl-CoA. The explanation lies in the reduction of the concentration of malonyl-CoA by a very active **malonyl-CoA decarboxylase** which does not attack methylmalonyl-CoA. This enzyme is present in mammalian mitochondria, where it prevents inhibition of methylmalonyl-CoA mutase and pyruvate decarboxylase by malonyl-CoA. The enzyme of the uropygial gland is located in the cytoplasm, where fatty acid synthesis occurs. The goose enzyme (462 amino acids) has been sequenced via the cDNA [113]. When the enzyme is absent, as from the uropygial gland of the chicken or from the liver of the goose, no methyl-branched fatty acids are produced [108, 195].

15.1.3 Conversion of Fatty Acids

Fatty acids that are synthesized de novo in animals or are taken in with the diet can be altered in many ways, in particular by elongation and shortening, desaturation and hydrogenation of the double bonds, substitution to, for example, hydroxy acids, reduction to long-chained alcohols (fatty alcohols), decarboxylation to hydrocarbons and incorporation into lipids. Most of these processes, like the oxidative catabolism of the fatty acids, require activation by CoA. Depending upon the order in which the individual conversion processes occur, very different intermediates and end products can arise from the same fatty acid. As the **reactions of fatty acid metabolism take place in different cell compartments**, the production of any particular end product may require several transport events through different permeability barriers.

In the vertebrates, the binding of long-chain fatty acids (C_{12} to C_{24}) to CoA and their incorporation into phospholipids or triacylglycerols occurs mainly in the cytoplasm; their elongation and desaturation occurs in the endoplasmic reticulum; their shortening by partial β-oxidation takes place in the peroxisomes; and their complete oxidation and ketone-body formation occurs in the mitochondria. Short-chain fatty acids (up to C_{10}) are activated by acyl-CoA synthetases, which are located mainly in the mitochondrial matrix [195]. The carnitine-dependent

transport of acyl residues through the inner mitochondrial membrane (Fig. 15.9) also plays a role in the conversion process, but will be discussed only in connection with β-oxidative degradation. The processes of lipid metabolism in invertebrates appear to be distributed in a similar fashion amongst various cell compartments, although desaturation in insects has been demonstrated in both microsomes *(Locusta)*, and mitochondria *(Drosophila)*, and even at different rates in both cell organelles *(Ceratitis)* [63]. The enzymes involved in the formation and hydrolysis of acyl-CoA have, up to now, been examined in detail only in the mammals, where there are acyl-CoA synthetases and hydrolases with varying chain-length specificity and intracellular localization [195].

Chain elongation is the most important reaction of fatty acid metabolism after desaturation. Thus, the palmitic acids initially formed by the FAS of all animals are mostly converted into unsaturated fatty acids with 18–22 C atoms. The mammalian elongation systems are localized in the microsomes or mitochondria, and catalyse a reaction sequence analogous to that of FAS, except that cytochrome b_5 is involved together with NADH as hydrogen donor. The rate-limiting condensing enzyme of this process exists in at least three different forms in the rat liver, with specificity for saturated, mono- or polyunsaturated fatty acids [195, 209]. Tracer experiments have shown a capacity for elongation in all investigated invertebrates, but nothing is known about the enzymes involved. Of particular interest is the formation of extremely long fatty acids from palmitic acid, as seen in the sponge *Jaspis stellifera* [33], and as is assumed to be the case for the long-chain fatty alcohols and acids in wax esters. The **shortening** of long-chain fatty acids occurs by partial β-oxidation in the peroxisomes.

The first double bond is introduced into position 9 in animals and most other eukaryotes, i.e. the **desaturase** begins counting from the carboxyl C. Further double bonds usually arise between existing bonds and the carboxyl end in animals, and the methyl end in plants; the flagellate *Euglena* carries out both processes. The desaturases are both position- and chain-length specific. The stearoyl-CoA-Δ^9-desaturase from the microsomes of mammalian liver is the only such enzyme to have been characterized in detail so far. This is, in principle, an electron transport chain which utilizes O_2 and electrons from reduced cytochrome b_5 to introduce a double bond into stearoyl-CoA or another acyl-CoA substrate. The system consists of an NADH:cytochrome b_5 reductase, cytochrome b_5 and the actual desaturase; the sequence of the latter, determined from the cDNA, consists of 358 mostly non-polar amino acids. Phospholipid stabilizes the structure and is indispensable for the enzyme activity [195, 260]. The importance of the phospholipid component has also been demonstrated for the microsomal desaturase of insects, e.g. for the palmitoyl-CoA-specific enzyme from the Mediterranean fruit fly *Ceratitis capitata* [168]. Mammalian liver apparently contains two Δ^9-desaturases which attack acyl-CoAs of different chain length. In mammals, further double bonds can be introduced at positions 6, 5 and 4 (Fig. 15.7), but much less is known about the corresponding specific desaturases than about the Δ^9-desaturases [124, 195].

The stearoyl-CoA-Δ^9-desaturase of rat liver microsomes is specific for acyl-CoA and inactive against phosphatidylcholine-bound acids. In contrast, the Δ^5- and Δ^6-desaturases of the mammals can process both acyl compounds, although it is not clear which is the natural substrate. The plant desaturases are specific for phospholipid-bound fatty acids. In the ciliate *Tetrahymena pyriformis*, palmitic acid can only be desaturated in the form of a CoA derivative; the desaturation of oleic acid to linoleic acid, which is not found in higher animals, occurs in this case with phosphatidylcholine or oleoyl-CoA [137]. Bacteria desaturate their fatty acids already during de novo synthesis and cannot produce polyunsaturated fatty acids (PUFAs). Desaturation of fatty acids still attached to the FAS was described some time ago for the dipterans *Calliphora erythrocephala* and *Drosophila melanogaster* but has not been further investigated [63].

Polyunsaturated fatty acids (PUFAs) are essential components of membrane lipids and the precursors of prostaglandins. It is claimed in many textbooks that the specificity of their desaturases prevents all animals from producing linoleic acid 18:2(n-6) and α-linolenic acid 18:3(n-3); these are therefore essential in the diet. On the other hand, as the starting substance of the (n-6) and (n-3) families, these can be further desaturated and elongated in animals, just like the (n-9) fatty acids can be (Fig. 15.7). However, the situation is not quite so straightforward. The requirement for both fatty acids has been shown clearly in the rat. The absence of both fatty acids from the diet results in deficiency symptoms such as dermatitis, slower growth and kidney damage. However, rats can be normally reared through two generations with linoleic acid alone, and linolenic acid alone only partially prevents the sym-

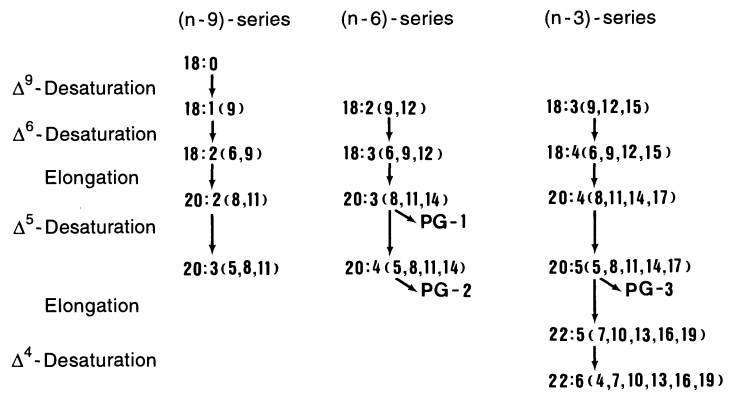


Fig. 15.7. Desaturation and elongation of animal fatty acids. *PG*, prostaglandin

ptoms. In humans, only a requirement for linoleic acid has been demonstrated. Arachidonic acid 20:4(n-6), important as a precursor of prostaglandins, can be produced by most mammals by desaturation and elongation of linoleic acid 18:2(n-6) (Fig. 15.7). The domestic cat and other felines lack the required desaturase, and for these animals arachidonic acid is essential.

The requirements of **fish** for essential fatty acids varies greatly with the species. The trout *(Salmo gairdneri)* requires only α-linolenic acid 18:3(n-3) or another (n-3) fatty acid. In contrast, carp *(Cyprinus carpio)*, eels *(Anguilla japonica)* and salmon *(Onchorhynchus kisutch)* also require linoleic acid 18:2(n-6). In marine fish *(Chrysophrys major, Pleuronectes platessa*, etc.), the PUFA requirement is only satisfied by long-chain highly unsaturated fatty acids such as 20:5(n-3) or 22:6(n-3). The difference in requirement for (n-3) fatty acids probably results from the fact that the trout rapidly converts 18:3(n-3) into 22:6(n-3), whilst the marine species are unable to do so. Thus, the trout is very suitable for the study of animal desaturases. In fish farming, the problem of species-specific differences in PUFA requirement is avoided by the use of feedstuffs that are rich in 20:5(n-3) and 22:6(n-3) [82]. The requirement for essential fatty acids demonstrated in the prawn *Penaeus japonicus* is satisfied more effectively by linolenic acid than by linoleic acid.

PUFAs also serve as precursors for prostaglandins and other eicosanoids in the **insects** (p. 612). Insect species can be assigned to one of four groups according to their requirements for PUFAs. Many species, in particular the herbivores, require linoleic acid 18:2(n-6) or α-linolenic acid 18:3(n-3) in their diet. Tracer experiments show that the wax moth *Galleria melonella* can elongate and desaturate linoleic acid to 20:4(n-6) and α-linolenic acid to 20:5(n-3), i.e. their PUFA metabolism corresponds to that of many vertebrates. According to nutritional experiments, the growth of *G. melonella* is optimal with α-linolenic acid; similarly effective are 20:3(n-3) or 22:3(n-3), which can be shortened to 18:3(n-3). Linoleic acid 18:2(n-6) is an order of magnitude less effective than 18:3(n-3). In tracer experiments with linoleic acid, labelled 20:5(n-3) appeared in phospholipids, probably through the activity of gut symbionts. The insect species of the second group require in their diet arachidonic acid 20:4(n-6) or eicosapentaenoic acid 20:5(n-3) in addition to linoleic acid and α-linolenic acid because they are unable to elongate C_{18} fatty acids. This group includes *Culex pipiens* and other mosquito species, and the moth *Platynota idaeusalis* [23].

The species of the third group have the unique ability in the animal kingdom to synthesize linoleic acid de novo with further elongation and desaturation to 20:3(n-6) and 20:4(n-6). They therefore have no requirement for fatty acids in their diet. This ability to synthesize linoleic acid has been demonstrated in tracer experiments with 15 species, including cockroaches *(Periplaneta americana)*, orthopterans *(Acheta domestica, Teleogryllus domesticus)*, homopterans *(Myzus persicae)*, termites *(Zootermopsis angusticollis)* and neuropterans *(Chrysopa carnea)*. The biosynthesis of 18:2(n-6) by these species does not involve the activity of symbionts but is entirely due to the animal's own metabolism. These insects possess a Δ^{12}-desaturase, which is otherwise unknown in the animal kingdom, and with this enzyme they can convert oleic acid 18:1(n-9) to 18:2(n-6). The Δ^{12}-desaturases of *A. domestica* and *P. americana* have been examined in some detail and compared with the analogous enzymes of plants. They are localized in the microsomes and have their highest activity in the fat bodies. The plant enzyme uses 18:1 fatty acid esterified

with phospholipid as the substrate and prefers NADH as the cosubstrate. In contrast, the insect Δ^{12}-desaturase uses acyl-CoA as the substrate, like all animal desaturases, and is more active with NADPH than with NADH [23, 52]. *Drosophila melanogaster* appears to represent a fourth group of insects which neither synthesize nor require PUFAs. Fruit flies reared for ten generations on a PUFA-free diet still contained polyunsaturated C_{20} fatty acids [23].

Just how **sponges** and **molluscs** produce the PUFAs described above, whose double bonds are separated by more than one CH_2 group (nonmethylene interrupted), has not been defined in detail [293]. The marine sponge *Microciona prolifera* synthesizes 26:3(5,9,19) from 16:1(9) by elongation and desaturation at positions 5 and 9, and the freshwater sponge *Ephydatia fluviatilis* can convert 26:2(5,9) to 26:3(5,9,19) by means of a Δ^{19}-desaturase. This is the first evidence for a Δ^{19}-desaturase in a living organism [80]. The metabolism of unsaturated fatty acids by the **Protozoa** shows even greater deviation from the metabolic patterns of the Eumetazoa. However, there is also large variation between the different groups, indicating once again how the so-called protozoans are an extremely heterogeneous collection from a comparative biochemistry standpoint. Oleic acid is essential for the ciliate *Paramecium aurelia* but can be further metabolized to 18:2(n-6), 18:3(n-6) and 20:4(n-6) [221]. The ciliate *Tetrahymena* can not only desaturate stearic acid to oleic acid and further to linoleic acid 18:(n-6) and γ-linolenic acid 18:3(n-6) but can also elongate palmitoleic acid 16:1(n-7) to vaccenic acid 18:1(n-7); *Acanthamoeba* and other amoeba species have a similar capacity [137]. Many of the Kinetoplastida, e.g. *Crithidia*, produce mainly 18:0 and 18:1 and the (n-6) fatty acids 18:2, 18:3 and 22:5 from acetate. Other Kinetoplastida, like *Trypanosoma lewisi*, synthesize mainly the highly unsaturated, long-chain fatty acids 20:5, 22:5 and 22:6; the predominant C_{16} and C_{18} fatty acids of their spectrum are apparently taken up from the blood of the host.

In contrast to desaturation of the fatty acids, the hydrogenation of double bonds has been relatively little investigated. The utilization of fatty acids as hydrogen acceptors by rumen symbionts, and the resulting complete saturation of all unsaturated fatty acids was referred to above. In this way, for example, oleic and linoleic acids are quantitatively converted to stearic acid. The corresponding metabolism in animals has been demonstrated in the mitochondria of rat liver, where the double bond at C-4 of 22:4(n-9) is hydrogenated by a specific 4-enoyl reductase.

The reduction of free fatty acids to long-chain alcohols requires CoA and NADPH. The corresponding reductase activity has been investigated in, for example, the uropygial glands of the bunting *Zonotrichia leucophrys* and the ovary of the gourami *Trichogaster cosby*, where intensive synthesis of wax esters occurs [108]. Traces of the presumed intermediate aldehyde have also been detected; the alcohols produced are simultaneously incorporated into the wax esters. Because of differences in the specificity of the reductases and acyltransferases, the fatty acids and alcohols of the wax esters differ in their chain structure. The uropygial gland of *Zonotrichia* can also reduce 2-hydroxyacyl-CoA to the corresponding alkane-1,2-diols; 2-hydroxylase activity has been reported. As has been shown for the pheasant *Phasianus colchicus*, the alkane-2,3-diols in the uropygial gland secretions of the fowls arise from acyloins (Fig. 15.8), which themselves are produced by the condensation of an aldehyde with hydroxyethylthiamine pyrophosphate [108]. The oxidation of long-chain alcohols to fatty acids has been demonstrated in the gut mucosa and other tissues of fish species that accumulate wax esters. In vitro preparations from the gourami catalyse the reaction with either $NADP^+$ or NAD^+ [94].

Decarboxylation of fatty acids gives rise to the corresponding hydrocarbons with one less C atom. This process has been examined in insects, where hydrocarbons are an important component of cuticular lipids. It has been shown in the termite *Zootermopsis angusticollis* that 2-hydroxy-24:0 is more rapidly decarboxylated than 24:0; thus, a 2-hydroxylation is apparently the first step of hydrocarbon synthesis. The 2-hydroxy fatty acids are probably oxidatively decarboxylated to aldehydes with one less C atom; these are then reduced to the alcohols, which are in turn converted to the alkenes by cleavage of water, and further reduced to the alkanes by the addition of hydrogen. The enzyme system is localized in the

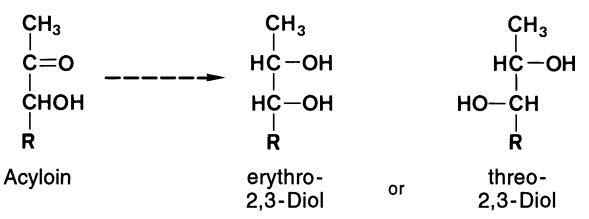


Fig. 15.8. Acyloins are precursors of erythro- and threo-2,3-diols in the diester waxes of fowl uropygial glands [108]

endoplasmic reticulum and is stimulated in vitro by ascorbic acid, like the plant and mammalian enzymes, but is not stimulated by ATP, CoA, NADH or NADPH [44].

15.1.4 Oxidation of Fatty Acids

β-Oxidation cleaves C_2 fragments (acetyl-CoA) from acyl-CoA and these can then function as, for example, the starting material for the de novo synthesis of fatty acids and sterols, or energy-yielding substrates. The complete degradation of fatty acids occurs in the mitochondria, and partial β-oxidation in the peroxisomes. The order of the reactions is the same for both systems but the enzymes involved differ significantly. The peroxisomal system brings about the shortening of long-chain fatty acids and also the β-oxidation of other substances from the body's own metabolism or of xenobiotics. The two different β-oxidation systems are found only in animals; plants, lower fungi and yeast have only one type of peroxisomal system in the microbodies [235]. In addition to β-oxidation of acyl-CoA, free fatty acids are subject to α-oxidation in the mitochondria and ω-oxidation in the cytoplasm. Because of its relatively large mass, the musculature contributes significantly to the β-oxidation of fatty acids and to the metabolism of ketone bodies; the heart is also very active in this respect. In addition to β-oxidation, the liver and kidney show considerable ketone-body formation, and the fat tissue also synthesizes triacylglycerols [195].

Prior to transport into the mitochondria, fatty acids must be bound to CoA by the activity of acyl-CoA synthetases of the cytosol or the outer mitochondrial membrane. An acyl-CoA synthetase from the rat shows a surprising 36 % agreement with firefly luciferase in part (amino acids 458–592) of its sequence of 699 amino acids [257]. A further acyl-CoA synthetase that is specific for fatty acids with more than 22 C atoms is present in the peroxisomes of rat liver but not in the mitochondria [243]. Because the CoA-bound acyl residues cannot cross the inner mitochondrial membrane, they must first be transferred to carnitine by the action of the external **carnitine acyltransferase**. The acylcarnitine is transported to the mitochondrial matrix by a specific translocase, and in the matrix is converted to acyl-CoA by the internal carnitine acyltransferase; acyl-CoA is then completely degraded by β-oxidation (Fig. 15.9). Investigations of the mitochondria of mammalian heart suggest that outer and inner carnitine acyltransferases are identical; however, two enzymes exist with different specificities for short- and long-chain acyl residues [17, 195]. The involvement of carnitine in mitochondrial fatty acid transport has also been shown in many insects. In the migratory locust *Locusta migratoria*, which consumes mainly lipids during sustained flight, the activity of carnitine acyltransferase is much higher than in the carbohydrate-utilizing Hymenoptera and Diptera. The flight-muscle mitochondria of the moth *Prodenia eridania* are specialized for fatty acid degradation and can oxidize palmitate in vitro without the addition of carnitine, whereas those of the related species *Manduca sexta* require carnitine. The carnitine acyltransferase of the lepidopteran mitochondria may act to prevent the increase of acetyl-CoA to inhibitory concentrations when there is insufficient oxaloacetate present for its introduction into the citrate cycle [63].

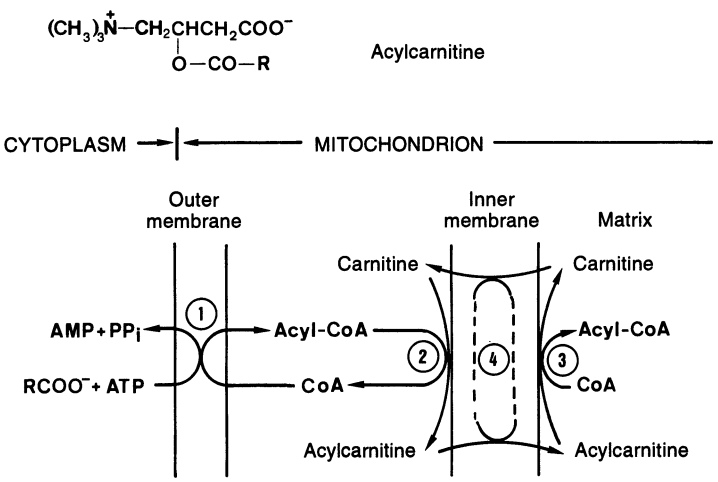


Fig. 15.9. Carnitine-dependent fatty acid transport across the inner mitochondrial membrane [195]. *1*, Acyl-CoA synthetase of the outer mitochondrial membrane; *2*, outer carnitine acyltransferase; *3*, inner carnitine acyltransferase; *4*, carnitine translocase

The β-oxidation of saturated fatty acids involves four **reaction steps** per cycle (Fig. 15.10). Intermediate products are not released. Although individual active enzymes can be isolated, the system is apparently highly organized. It appears in fact that, in contrast to glycolysis or the citrate cycle, none of the intermediates are the starting points of other metabolic pathways. The **enzymes** of β-oxidation have only been examined in detail in the mammals. In most cases, several isoenzymes of different substrate specificity are present. Thus, in the rat liver there are no less than five different acyl-CoA dehydrogenases, of which SCAD, MCAD and LCAD bring about the β-oxidation of short-, medium- and long-chain fatty acids respectively, whereas isovaleryl-CoA dehydrogenase (IVD) is involved in the metabolism of leucine, and the 2-methyl-branched acyl-CoA dehydrogenase is involved in the metabolism of isoleucine and valine. Pairwise comparisons of LCAD (400 amino acids), SCAD (388 amino acids) and IVD (394 amino acids) show 31–35 % sequence agreement [165]. There are at least two different enoyl-CoA hydratases (crotonases), one each for short and long chains. Both are stereo-specific and convert the trans-2-enoyl-CoA resulting from acyl-CoA activity into L-3-hydroxyacyl-CoA. The cis isomer results in D-3-hydroxyacyl-CoA, which cannot be further processed by the next enzyme, L-3-hydroxyacyl dehydrogenase. The mitochondrial enoyl-CoA hydratase has a molecular mass of 161 kDa and consists of six identical subunits of 290 amino acids. The hydration reaction in the peroxisomes is catalysed by a monomeric, trifunctional enzyme that also has the activities of 3-hydroxyacyl-CoA dehydrogenase and the Δ^3,Δ^2-enoyl-CoA isomerase involved in the degradation of unsaturated fatty acids. The hydratase and dehydrogenase domains of the trifunctional polypeptide show significant sequence agreement with the corresponding monomeric enzymes of the mitochondria [175, 235]. There are also two isoforms of acetyl-CoA acyltransferase (thiolase), the last enzyme of the series; one isoform is relatively unspecific and the other quite specific for acetoacetyl-CoA [195].

Problems arise in the **β-oxidation of unsaturated fatty acids**. Because the double bonds in natural fatty acids always have the cis configuration, stepwise degradation during β-oxidation gives rise to either Δ^3-cis or Δ^2-trans-Δ^4-cis compounds that cannot be effectively processed further by the stereo-specific enoyl-CoA hydratase (Fig. 15.10). The usual pathway involves the isomerization of Δ^3-cis-enoyl-CoA and the reduction of Δ^2-trans-Δ^4-cis-dienoyl-CoA to Δ^2-trans-enoyl-CoA; the β-hydroxyacylepimerase, which converts the otherwise unmetabolizable D-3-hydroxyacyl-CoA to the L-isomer, seems to be absent, at least in the rat. On the other hand, there are at least three isoenzymes of Δ^3,Δ^2-enoyl-CoA isomerase, two in the mitochondria for short- and long-chain fatty acids, and a peroxisomal form that occurs as part of a multifunctional protein [130]. **Propionyl-CoA** is produced by the β-oxidation of fatty acids with an odd number of C atoms; this compound cannot be metabolized by all cells, e.g. in mammals it can be metabolized by cells of the liver but not by those of the heart. Propionyl-CoA and 3-hydroxy-3-methylglutaryl-CoA also arise from the β-oxidation of methyl-substituted fatty acids [195]. In the vertebrates, propionyl-CoA is car-

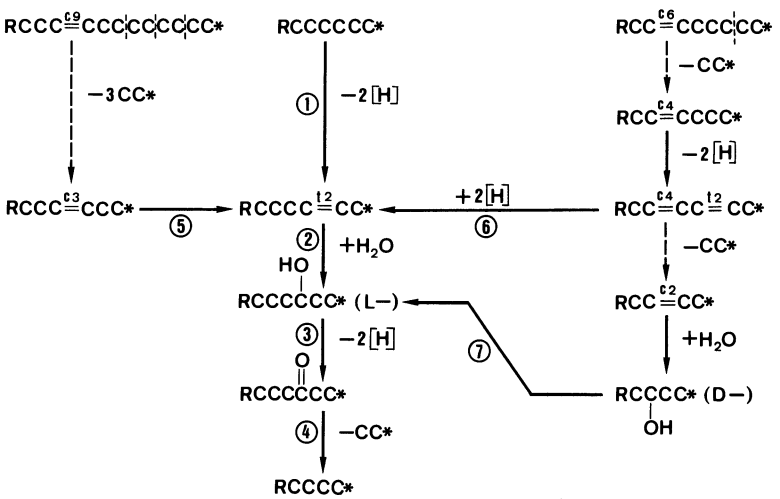


Fig. 15.10. β-Oxidation of saturated and unsaturated fatty acids [195]. The first C atoms of the chains are designated *C* and the remainder *R; C** denotes the group -C-SCoA; *c* or *t* followed by a number indicates the configuration and position of a double bond, e.g. *c9* = cis- Δ^9, *t2* = trans- Δ^2. *1*, Acyl-CoA dehydrogenase; *2*, enyol-CoA hydratase; *3*, 3-hydroxyacyl-CoA dehydrogenase; *4*, 3-ketoacyl-CoA thiolase; *5*, Δ^3,Δ^2-enyol-CoA isomerase converts cis- or trans-3-enoyl-CoA into trans-2-enoyl-CoA; *6*, 2,4-dienoyl reductase (NADPH) converts 2,4-dienoyl-CoA into 2-enoyl-CoA; *7*, 2-hydroxyacyl-CoA epimerase converts D-2-hydroxyacyl-CoA into L-2-hydroxyacyl-CoA

boxylated to D-methylmalonyl-CoA, isomerized to L-methylmalonyl-CoA, and then converted by the cobalamin-dependent methylmalonyl-CoA mutase to succinyl-CoA, which then enters the citrate cycle. The same reactions but operating in the opposite direction are used by some invertebrates to produce propionate from succinate in anaerobic metabolism, so long as vitamin B_{12} is available (p. 552). Various invertebrates are relatively rich in cobalamin and catabolize propionate, just as the vertebrates do. In contrast, all insects, regardless of whether they have a lot of vitamin B_{12} (e.g. the termite *Zootermopsis nevadensis)*, a little (e.g. the cockroach *Periplaneta americana)* or none at all (e.g. the housefly *Musca domestica)*, degrade propionate by the same cobalamin-independent pathway: this involves the formation of 3-hydroxypropionate by modified β-oxidation and its decarboxylation to acetate. In the process, the C-2 of the propionate becomes the carboxyl group of acetate, and the C-3 becomes the methyl group [81].

The existence of a β-oxidation system in peroxisomes was first discovered in the endosperm of beans and in the ciliate *Tetrahymena;* it was later found in the rat liver and other vertebrate tissues and is probably also present in all invertebrates [65]. The β-oxidation of fatty acids in *Tetrahymena* is restricted entirely to the peroxisomes. Although the reactions in the peroxisomes are the same as those which occur in the mitochondria, the enzymes are completely different. Whereas the first mitochondrial enzyme is a dehydrogenase, which introduces electrons into the respiration chain and yields two ATPs, the first peroxisomal enzyme is an FAD-containing acyl-CoA oxidase which produces H_2O_2 and yields no ATP. The peroxide is removed by the catalase in the peroxisomes. Because acyl-CoA oxidase is practically inactive against acyl residues with fewer than ten C atoms, the animal peroxisomes, in contrast to plant microbodies, cannot completely catabolize fatty acids [235]. Enoyl-CoA hydratase and 3-hydroxylacyl-CoA dehydrogenase are combined with Δ^3,Δ^2-enoyl-CoA isomerase to give a trifunctional enzyme. In contrast to the homotetrameric enzyme of the mitochondria, the 3-ketoacyl-CoA thiolase is a homodimer. How fatty acids reach the peroxisomes is still not known. Carnitine is not required; the carnitine acyltransferase in the peroxisomes is possibly involved in the transport of short-chain fatty acids into the mitochondria [235].

The concept that by shortening long-chain fatty acids the peroxisomes prepare them for complete degradation in the mitochondria apparently does not hold for all vertebrates. Certain copepod species, which form the principal diet of herring, mackerel and other pelagic fish, are relatively rich in wax esters with the alcohol 22:1(n-11); this is oxidized to the corresponding fatty acids in the gut mucosa. However, investigations of the rainbow trout *Salmo gairdneri* show that the fatty acids are not processed in the peroxisomes, the activity of which decreases with increasing chain length from C_{12} to C_{18}. The mitochondria are in fact adapted to the particular fatty acid composition of the diet in that they can oxidize a wide fatty acid spectrum of 12:0 to 22:6 [94].

The **hydroxylation** of lignoceric acid 24:0 to cerebronic acid 2-OH-24:0, an important constituent of brain glycolipids, occurs by the action of a mixed-function oxygenase [195]. In the uropygial glands of many birds, fatty acids are hydroxylated at C-2, C-3 or terminal C atoms and are then metabolized further; for example, in the bunting *Zonotrichia leucophrys* the 2-hydroxyacyl-CoA that is formed is immediately reduced to alkane-1,2-diol. 3-Hydroxy fatty acids probably arise by hydration of a 2-enoyl-CoA, and ω-hydroxy fatty acids arise by the action of a microsomal cytochrome P-450 monooxygenase [108, 134]. **α-Oxidation** is the name given to the process whereby the C-1 of a fatty acid is cleaved off as CO_2, and the C-2 is oxidized to a carboxyl group. For example, in the uropygial gland, shortening of 4-methyl fatty acids leads to 3-methyl fatty acids, the synthesis of which is otherwise difficult to imagine. This pathway allows the degradation of 3-methyl-branched fatty acids in mammalian liver and kidney; these fatty acids cannot be directly β-oxidized. This is important, for example, in the degradation of phytanic acid (3,7,11,15-tetramethylpalmitinic acid) that arises from phytol in plant foodstuffs [195]. Oxidation of the last or penultimate C atom **[ω- or (ω-1)-oxidation]** leads to the formation of ω- or (ω-1)-hydroxy fatty acids and then to dicarbonic acid, which may be shortened still further by β-oxidation. These reactions involve the participation of cytochrome P-450, require molecular oxygen and NADPH and, in addition to their occurrence in the mammalian liver and kidney and in the uropygial gland, have been described in th liver of the frog *Rana catesbeiana* [134, 1' 188, 195]. PUFAs can be oxidized to hyd- acids by lipogenases and reduced hydroxy acids with internal substi' ent types of mammalian blood cell

mes that vary in their specificity for the C atom carrying the hydroperoxy or hydroxy group [188]; corresponding enzyme activities have also been reported in the gills of the trout *Salmo trutta* [69].

15.1.5 Ketone Body Formation and Degradation

Hunger brings about within a few hours a switch in the mammalian liver from carbohydrate consumption and fatty acid biosynthesis to fatty acid oxidation and ketone body formation, the latter process predominating in diabetes. The term **ketone bodies** refers to acetoacetate, D-3-hydroxybutyrate and acetone. The ketone bodies produced by the liver can be oxidized by many extra-hepatic tissues. Because of their high solubility in water, ketone bodies are preferentially consumed as respiratory substrates by cells during periods of hunger and in diabetes. The **formation of ketone bodies** occurs via 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) (Fig. 15.11). In principle, the enzymes of the HMG-CoA pathway are present in all organs; however, the rate-limiting enzyme HMG-CoA synthase only has a high activity in the liver, which is therefore the main site of ketone body production. The rate-limiting enzyme of **ketone body degradation** is 3-ketoacyl-CoA-transferase. The liver lacks this enzyme and can produce, but cannot degrade, ketone bodies. Acetone arises from acetoacetate by spontaneous decarboxylation. Contrary to earlier assumptions, only part of the acetone produced in humans during hunger or diabetes is excreted in exhaled air, urine and sweat; as much as two-thirds is metabolized to CO_2 and glucose by an unknown process [5]. Ketone bodies in the liver and blood do not only increase in hungry mammals; high levels are also found in the ground-squirrel *Spermophilus undulatus* during hibernation (Table 15.6). This suggests their importance as an energy source [217]. Ketone body formation is found in the liver of all vertebrates from the agnathans to the mammals. A hunger-associated increase in blood levels of acetoacetate and 3-hydroxybutyrate has also been reported for the shark *Scyliorhinus canicula* (Table 15.6). In contrast, there is no change in the acetoacetate content of the blood in the sunfish *Dicentrarchus labrax*, and 3-hydroxybutyrate is completely absent from this and other teleost species. In this case, free fatty acids apparently play a large role during hunger [290]. A large-scale study of the muscles of many vertebrates and invertebrates showed the presence of 3-ketoacyl-CoA transferase and acetoacetyl-CoA thiolase in all species examined. On the other hand, little 3-hydroxybutyrate dehydrogenase was found in the muscles of marine teleosts, the snail *Buccinum*

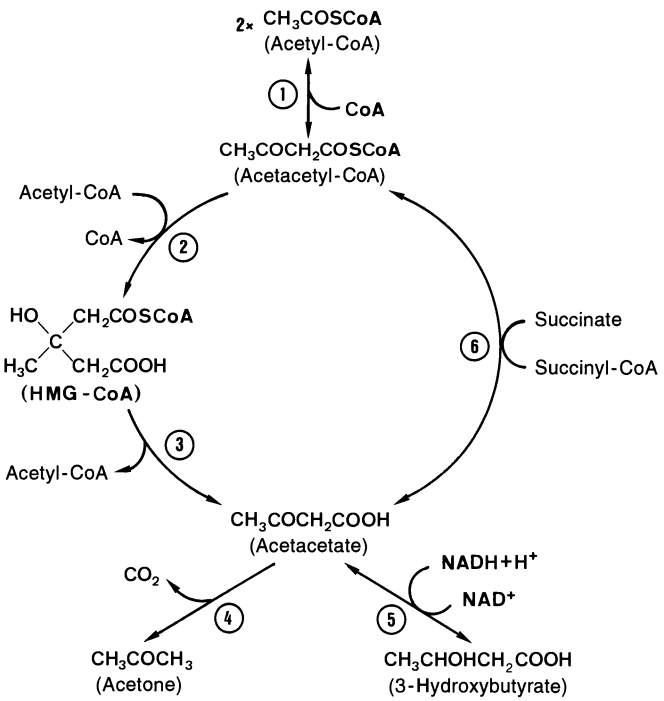


Fig. 15.11. Ketone body formation (order of reactions 1-2-3-4/5) and degradation (order of reactions 5-6-1). *1*, Acetoacetyl-CoA thiolase; *2*, 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthetase; *3*, HMG-CoA lyase; *4*, spontaneous decarboxylation; *5*, 3-hydroxybutyrate dehydrogenase; *6*, 3-ketoacyl transferase

Table 15.6. Changes in the ketone body concentration (μmol/l or μmol/kg fresh weight) during hunger in the rat [71], the ground-squirrel *Spermophilus undulatus* [217], the dogfish *Scyliorhinus canicula* [290], the cockroach *Periplaneta americana* [236], and the snail *Biomphalaria glabrata* [170]

		Acetoacetate		3-Hydroxybutyrate	
		Fed	Hungry	Fed	Hungry
Rat	Blood	49	283	40	426
	Liver	86	244	107	957
Squirrel	Blood	29	248	54	1232
	Liver	39	247	116	906
Dogfisch	Blood	40	290	30	1610
Cockroach	Haemolymph	470	3000	1330	1100
Snail	Haemolymph	146	42	571	40

Fed, fed animal; hungry, hungry animal (hibernating in the case of the squirrel)

undatum, and various crustaceans and insects; however, it was consistently present in elasmobranchs, amphibians, birds and mammals. Muscles may be divided into three groups on the basis of the activity of 3-ketoacyl-CoA transferase. Type 1, with activities of less than 0.01 U/g fresh weight, probably covers its energy requirements mainly by glycolysis; examples of such muscles are the abdominal muscles of the lobster and the white muscles of several teleosts. Type 2, with activities of up to 5 U/g fresh weight, only turns over ketone bodies during periods of hunger. Type 3, with activities over 5 U/g fresh weight, probably continuously oxidizes ketone bodies so long as they are present; for example, the brain is able to utilize glucose or ketone bodies depending upon their availability. Type 3 in particular includes continuously active muscles such as the red muscle of the teleosts, the heart muscle of all vertebrate classes, the muscular stomach of the chicken and the diaphragm of the mammals [13].

Little is known about the **ketone body metabolism of the invertebrates**. Only the crayfish *Orconectes limosus*, the snail *Biomphalaria glabrata* and several insects have been investigated in any detail [63, 170]. All the organs of *Orconectes* can produce, as well as degrade, ketone bodies; the clear specialization associated with mammalian organs is not known. 3-Hydroxybutyrate dehydrogenase is not present in any organ, and thus there is no 3-hydroxybutyrate in the haemolymph. The acetoacetate content of the haemolymph increases only after prolonged periods of hunger. In *Biomphalaria*, the complete set of enzymes for ketone body synthesis is found only in the hepatopancreas; all other organs lack HMG-CoA synthase. 3-Hydroxybutyrate dehydrogenase also has its highest activity in the hepatopancreas but is present in all other organs. The concentrations of acetoacetate and 3-hydroxybutyrate are very high in the haemolymph of well-fed animals but drop considerably during hunger (Table 15.6). It is possible that during hunger ketogenesis is prevented by the lack of substrates [170]. All enzymes necessary for ketone body production and degradation have been detected in one insect species or another, although ketone body metabolism has not really been closely examined in any species. The blood of the desert locust *Schistocera gregaria* contains mainly acetoacetate and the fat bodies contain 3-hydroxybutyrate; in the cockroach *Periplaneta americana* both are present in the haemolymph at very high concentrations and the acetoacetate concentration increases still further during hunger (Table 15.6).

15.2 Reserve Lipids and Lipid-Rich Secretions

Neutral lipids, such as triacylglycerols (fats), wax esters and hydrocarbons, have several advantages over glycogen as **energy reserves**. They not only have a higher energy potential (about 39 kJ/g compared with 17 kJ/g for glycogen) but also are not hydrated; thus, a fat reserve can deliver eight to nine times more energy than can a glycogen reserve of the same mass. The consumption of lipids is also advantageous for the water balance. The oxidative catabolism of 1 g trioleoylglycerol produces 1.06 ml metabolic water, compared with 0.67 ml from 1 g glycogen. Thus, high fat contents are mainly found in animals that need large energy reserves, e.g. in hibernating species, in migratory fish species such as eels and salmon, and, in particular, in long-distance fliers such as migratory birds, locusts and certain lepidopterans, for which the reduced weight and yields of water are of critical importance. The "fat index" ratio of fat:fat-free dry weight is usually about 0.3 for non-migratory birds but reaches values of over 3.0 in migratory birds at the beginning of their journey.

The advantages of fats as energy reserve substances must be set against some disadvantages. Their water insolubility brings problems of transport from the storage tissue to the site of utilization and, in contrast to glycogen, they can only be

catabolized in the presence of sufficient oxygen. As well as their major importance as energy reserves, fat stores can have other biological functions, e.g. **heat insulation** in seals and whales, or the **improvement of buoyancy** in the elasmobranchs, which have no swim bladders [270], and also in some teleosts, whales and aquatic invertebrates. The lipid-rich forehead projection (melon) of dolphins and toothed whales presents a very unusual function for lipids; it forms an **ultrasound lens** with which to focus the sounds emitted in echo navigation (p. 567; Table 15.5). Fats also serve as energy supplies in the secretions intended for the nutrition of young animals. The fat content of mammalian milk is particularly high in species in which the young have especially high energy requirements (see Table 13.2, p. 467); this is the case, for example, for young mice because of their small body size, and for young whales because of their unusual lifestyle. The nutritional secretions discussed in the following sections include the crop milk of pigeons, the stomach oils of petrels and the secretion from the Dufour glands of several hymenopterans.

15.2.1 Triacylglycerols (Triglycerides)

The most common reserve fats, the triacylglycerols (TAGs), have fatty acid compositions which vary widely with the species and diet. In mammals, the main site of TAG production and storage is the fat-tissue cells (adipocytes), but the gut mucosa, liver and milk glands are also capable of intensive TAG synthesis. In the insects, the gut wall and the fat-body cells are the main sites of TAG synthesis and storage. The **biosynthesis of TAG** can occur via one of three pathways. The glycerol-3-phosphate (GP) pathway and the dihydroxyacetone phosphate (DHAP) pathway proceed via phosphatidic acids, which are also the starting point for phospholipid synthesis (Fig. 15.22, p. 603). The phosphatidic acids are dephosphorylated by an acidic phosphatase to 1,2-diacylglycerol (DAG), which can then be converted to TAG by acyl transfer. The third possibility, the monoacylglycerol (MAG) pathway, leads from the product of lipase activity 2-monoacylglycerol via DAG to TAG. The DHAP pathway is of great importance in just a few mammalian organs, e.g. the brain. The MAG pathway has been mainly investigated in the cells of the small intestine, in which the MAG and fatty acids extracted from the gut lumen are resynthesized to give TAG and are given up into the lymph as chylomicrons after combination with phospholipids, cholesterol and proteins (p. 194).

The fat tissues of the rat contain all three pathways of TAG synthesis in the order GP > DHAP > MAG, but the existence of the MAG pathway in the milk glands is in some doubt. The acylation of 2-monoacylglycerol is not an unambiguous indication of this pathway, because in fact MAG can enter the GP pathway after hydrolysis to glycerol and phosphorylation to glycerol-3-phosphate. Even the low acylation of 2-monoacylglycerol without prior hydrolysis, observed in the milk glands of the goat, is no indication for the existence of the MAG pathway, as this can be explained by acyl transfer by lipases [85]. The GP pathway of TAG synthesis is widely found in plants and animals, and the MAG pathway has been clearly demonstrated in insects. In the microsomes of the Mediterranean fruit fly *Ceratitis capitata*, the acylation of glycerol-3-phosphate (G-3-P) is inhibited by MAG, whereas, conversely, G-3-P has no effect on MAG acylation. In *Periplaneta americana*, the MAG pathway appears to play a larger role in the midgut than in the fat bodies; in *Locusta migratoria*, the MAG pathway serves primarily in the formation of DAG as a transport form, and the GP pathway serves for TAG storage [63].

15.2.2 Energy Substrates in Muscles and Other Tissues

The greatest specialization in favour of particular energy-yielding substrates is found in the **flight muscles of the insects**. Because of their tracheal system, they have a purely aerobic metabolism in which different species show preference for the utilization of carbohydrate, fat or proline. The muscles display corresponding characteristic differences in their complement of enzymes for energy-yielding metabolism (Table 15.7a). The flight muscles of hymenopterans and dipterans, which usually fly only short distances, and the thorax muscles of the non-flying cockroaches are specialized to use **carbohydrates**. The glycogen reserves of the flight muscles last for only a limited time; the haemolymph trehalose, which is made in the fat bodies from glycogen, is then respired. In contrast, the flight muscles of long-distance fliers, such as many lepidopterans and orthopterans can utilize fats. Several of these species also use mainly carbohydrates at the beginning of flight, but then gradually switch over to the oxidation of fatty acids; this has been inves-

Table 15.7. Enzyme activities of energy-yielding metabolism in various muscles

a Flight muscles of various insects
Enzyme activities in μmol/mg muscle protein h^{-1}

Representatives of different insect orders [12]:
1, *Calliphora erythrocephala* (uses carbohydrate);
2, *Locusta migratoria* (uses carbohydrate and fat);
3, *Philosamia cynthia* (uses fat);
4, *Leptinotarsa decemlineata* (uses proline).
Ants with different feeding habits [162]:
5, *Formica ulkei* (sucks honeydew, uses carbohydrate);
6, *Pogonomyrmex californicus* (seed eating, uses fat);
7, *Atta colombica* (fungus eating, uses carbohydrate and fat)

	1	2	3	4	5	6	7
GAPDH	194	69	18	30	43	0.8	25
LDH	<0.05	0.3	0.6	0.1	–	–	–
GPDH	48	33	13	17	21	2.1	2.7
HOAD	<0.05	66	98	11	<0.4	3.1	3.4
CS	45	57	80	11	32	15.3	10.5
GluDH	–	1	–	58	–	–	–
GPT	–	4	–	80	–	–	–

b Red and white musculature of the tuna fish *Euthynnus pelamis* [75] (data in U/g fresh weight)

	Red muscle	White muscle
Phosphorylase	22	106
Aldolase	36	269
Lactate dehydrogenase	510	5500
Glycerol-sn-3-phosphate dehydrogenase	22	105
Citrate synthase	21	2.1
Glutamate dehydrogenase	5.9	3.0
Aspartate aminotransferase	102	43
Alanine aminotransferase	7.7	2.0

GAPDH, Glyceraldehyde phosphate dehydrogenase; LDH, lactate dehydrogenase; GPDH, glycerol-sn-3-phosphate dehydrogenase; HOAD, β-hydroxyacyl dehydrogenase; CS, citrate synthase; GluDH, glutamate dehydrogenase; GPT, alanine aminotransferase

tigated in detail in the migratory locust *Locusta migratoria* [12]. In most insect species, the fatty acids oxidized in flight muscles originate from the DAG of the haemolymph, where they are bound to special transport proteins (lipophorins; p. 195). The bug species *Rhodnius prolixus* and *Triatoma infestans* use mainly fatty acids in flight; in this case these come not only from haemolymph DAG but also from TAG in the flight muscles [273]. Based on the enzyme spectra, different ant species are specialized for different energy substrates (Table 15.7a): *Formica ulkei*, which feeds on honeydew, is specialized to use carbohydrates; the fungus-eating species *Atta colombica* can apparently use both carbohydrates and lipids; and, somewhat surprisingly for a hymenopteran species, the grain-eating *Pogonomyrmex californicus* has an enzyme system completely adapted to fatty acid oxidation [162].

In the gut cells of insects, lipids are absorbed as free fatty acids, MAG and DAG in species-specific proportions, and are resynthesized in the cells to give TAG. In contrast to the gut epithelium cells of mammals, fat droplets are produced freely suspended in the cytoplasm; and unlike the chlyomicrons of mammals, the fat droplets are not subjected to exocytosis. Instead, fatty acids and DAG are released into the cytoplasm as individual molecules [135]. In insect haemolymph DAGs are apparently sn-1,2 in structure [158]. There are two possible mechanisms for their formation in fat bodies and the relative importance of these mechanisms is controversial: (1) hydrolysis of TAG by an sn-3-specific lipase can occur or (2) there may be hydrolysis of TAG as far as 2-MAG, and then acylation by MAG acyltransferase [12, 158]. DAG release from fat bodies is hormonally regulated. Adipokinetic hormones from the corpus cardiaca have been found in many insects and have been isolated and sequenced from several orthopterans and lepidopterans (see Fig. 8.6 g, p. 306). They stimulate the fat-body lipase, although this does not happen in all species; during the process a cAMP-dependent phosphorylation apparently occurs. The DAGs in haemolymph are bound to transport proteins, the lipophorins, which are removed again upon uptake of DAG into the flight muscle. This is significantly different to the situation in the vertebrates, where unaltered lipoproteins are taken up by endocytosis into different sorts of cells [12].

The brains of several insect species, e.g. *Drosophila melanogaster* and *Bombyx mori*, can also use fatty acids as energy substrates. They therefore constitute an interesting exception to the other nervous systems examined, as these can only cover their energy requirements with glucose or, under certain conditions, partly with ketone bodies. This is true not only for the mammals but also for all other vertebrates, the edible snail *Helix pomatia*, the earthworm *Lumbricus terrestris*, the crayfish *Orconectes limosus* and, amongst the insects, the bees and the fly *Calliphora erythrocephala* [275]. **Proline** also participates in the flight muscle metabolism of many insects; it provides oxaloacetate for the introduction into the citrate cycle of CoA derived from pyruvate or fatty acids. Proline is the main sub-

strate of flight muscle metabolism in the stable fly *Glossina morsitans* and the potato beetle *Leptinotarsa decemlineata*. The proline metabolism of these species was described on p. 421. Proline additionally plays a significant role as an energy substrate in the gut wall of the migratory locust and in the mantle muscles of the cephalopods [35].

Fkatty acids are also important energy substrates in many vertebrate muscles, including those of humans. The various energy substrates of human muscles differ, on the one hand, in the amounts of energy stored and, on the other hand, in the maximal possible rates of energy production (see Table 14.4, p. 536). An adult male has an energy reserve of about 340 000 kJ in body fat and only about 7500 kJ in the liver glycogen and muscle glycogen. Thus, the 12 000 kJ of energy required for a marathon run can only be obtained by oxidation of fatty acids in the muscles. The proportion of total energy used that is derived from fatty acids is limited by the rate at which these can be made available. In contrast to some birds and fish, human muscles contain only small amounts of fatty acids in the form of TAG; the muscle therefore oxidizes mainly exogenous fatty acids obtained from the fat tissues and transported in the blood as non-esterified fatty acids (NEFA) bound to albumin. The blood NEFA concentration is regulated between 0.3 and 2.0 mmol/l, with the highest values being found during persistent muscle activity, during hunger and in the diabetic condition. The release of fatty acids from the fat tissues (lipolysis) is hormone controlled with cAMP acting as the second messenger to activate a protein kinase and bring about the conversion of lipase b to active lipase a. The fact that fatty acids can be so rapidly turned over, despite their low concentration in the plasma, is due to their short residence time in the plasma; they have a half-life of 2 min.

The intensive utilization of fatty acids in muscle metabolism is a feature of the breast muscles of the pigeon and other long-distance fliers amongst the birds, and of certain fish muscles. Vertebrate muscles are made up of different types of fibres which differ not only in their functional and contractile characteristics but also in their complement of enzymes for energy-yielding metabolism. The extremes are represented by the "fast-twitch white" type, which is specialized for anaerobic glycolysis production of energy for sudden, short-lived activity, and the "slow-twitch red" type, which is adapted for persistent activity, has a high capacity for fatty acid and pyruvate oxidation, and owes its red colour to its myoglobin content (Table 15.7b). Most vertebrates possess intermediate types with many muscles composed of several types of fibre. However, most of the body musculature of the primitive fish is made up exclusively of white fibres which are used only during fleeing or hunting movements; the thin red lateral muscles serve during continuous swimming. The red muscles of the salmon use mainly fatty acids at the start of spawning migration, and later use amino acids, with alanine being much more important than proline [178].

15.2.3 Wax Esters, Hydrocarbons and "Ether Glycerides"

Certain animal groups contain 1-alkyl-2,3-diacylglycerols (ether glycerides), wax esters and hydrocarbons as storage lipids, either in addition to or instead of triacylglycerols. Glycerolipids with an alcohol (alkyl residue) instead of a fatty acid (acyl residue) at position sn-1 are widely found in the phospholipids but occur much less in the neutral fats. Small quantities of **1-alkyl-2,3-diacylglycerols** are found in the liver of mammals and amphibians; larger quantities are found in the preputial glands of the mouse and rat; and the largest quantities amongst the vertebrates are present in the elasmobranch. In fact, the trivial names chimyl, batyl and selachyl alcohols for 1-hexadecyl-, 1-octadecyl- and 1-octadec-9-enylglycerol come from the elasmobranch Chimaeridae, Batoidei and Selachii, in which they were first discovered. Ether glycerides with alkyl residues with 8–19 C atoms make up about 20–69 % of the total lipids in the hepatopancreas of several Japanese brachyurans. Almost 6 % of the body lipids of the cephalopod *Octopus dofleini* consist of ether glycerides, with the highest concentrations again occurring in the hepatopancreas. 1(2′-Hydroxyalkyl)- and 1(2′-methoxyalkyl)-2,3-diacylglycerols have also been found in some marine animals [104].

Traces of wax esters arising from the diet are to be found in probably all animals. Larger amounts, constituting approximately 10 to over 90 % of the total lipids, have been reported in more than 120 species of marine animals from seven phyla [133]; a selection of the data is shown in Table 15.8. Wax esters are particularly widespread in the crustacean groups Copepoda, Euphausiacea and Mysidacea, in the cartilaginous and bony fish, including the crossopterygian *Latimeria chalumnae*, and in the toothed whales.

Table 15.8. Wax esters of marine animals (% total lipid) [133]

Cnidaria	*Conylactis gigantea*	40
Ctenophora	*Beroe cucumis*	12
Chaetognatha	*Eukronia hamata*	12
Mollusca	Cranchiidae squids	32
Ann. Polychaeta	Alciopidae	76
Crustacea	*Cyphocaris chalengeri*	69
	Calanus helgolandicus	33
	C. hyperboreus	92
	Euchaeta japonica	54
	Gnathophausia ingens	69
	Euphausia crystallorophias	30
	Hymenodora glacialis	62
Elasmobranchii	*Chlamydoselachus anguineus:* liver	58
Teleostei	*Cyclothone signata*	33
	Merluccius capensis: ovary	25
	Laemonema morosum: muscle	50
	liver	60
	Stenobrachius leucopsarus: muscle	90
	Ruvettus pretiosus	92
	Mugil cephalus: ovary	67
Crossopterygii	*Latimeria chalumnae:* muscle	93
	fat tissue	97
Odontoceti	*Tursiops truncatus:* jaw	35
(Toothed whale)	*Physeter catodon:* blubber	66
	spermaceti organ	73
	Hyperoodon ampullatur: blubber	70

Several of these species occur as enormous numbers of individuals, e.g. the Antarctic krill *Euphausia crystallorophias* and the "deep-sea minnow" *Cyclothone signata* from the family Gonostomatidae. Thus, the wax esters belong to the most important organic compounds of the biosphere, with an estimated yearly production of 300 million t [133]. As a group characteristic, wax esters are found only in the copepods Calanoidea, where all 40 species from the families so far examined contain such lipids. In general, the wax ester content of crustaceans and fish appears to be more dependent on the habitat and nutrition than on the systematic position. It is noteworthy that the plankton inhabiting the deeper layers of the ocean (bathypelagic) are all rich in wax esters. Hence, species of both the copepods Calanoidea and the Gonostomatidae that inhabit the deeper, colder depths contain more wax esters than do those in warmer surface layers [133, 193]. Fish species living close to the surface (epipelagic) contain wax esters almost exclusively in the ovary, whereas in bathypelagic species they are also present in muscles and other tissues. In contrast to the waxes of terrestrial species, the marine wax esters always consist of even-numbered, unbranched fatty acids and fatty alcohols with the usual numbers of C atoms. The chain lengths lie between 26 and 42 C atoms with most in the region 32–38. The principal fatty acids are 12:0 and 14:0 in the sperm whale, 18:1 in fish muscle, and higher unsaturated fatty acids in fish ovary. The structural differences between the wax esters of muscle and ovary suggest various functions; it may be that the egg wax esters are a reserve for phospholipid synthesis and energy metabolism, and those of the muscle are more involved in buoyancy regulation. Amongst the fatty alcohols, 16:0 predominates in bathypelagic plankton, and 20:1 and 22:1 predominate in epipelagic types; the krill *Euphausia crystallorophias* has 14:0 fatty alcohols. There are few reports dealing with the **metabolism of the wax esters** of marine animals. The fatty acids may come partially from de novo synthesis but most are derived from the diet. The reduction of fatty acids to alcohols is brought about by a special reductase system that prefers NADH to NADPH, requires ATP and CoA, and is apparently only present in significant amounts in species producing wax esters. However, acyl transferases, which can produce wax esters from acyl-CoA and fatty alcohols, appear to be widely distributed [133, 193]. Just as little is known about the degradation of wax esters. The lipases of the examined copepods and fish cleave wax esters much more slowly than TAG and, in addition, are inhibited by fatty alcohols. Thus, it is conceivable that the released fatty alcohol is immediately oxidized by an NAD-dependent alcohol dehydrogenase, as found, for example, in the liver of the shark *Squalus acanthias*. Fish that themselves have no wax esters can, however, metabolize them; the fatty alcohols released in the gut lumen are oxidized to fatty acids during absorption [133, 193].

It is only possible to speculate about the **functions** of the wax esters and their advantages over TAG. They can probably function as energy reserves, although there is no direct evidence for this. Like all lipids, they are lighter than seawater and can therefore help to adjust the density of the animal to that of the surrounding water. In this respect they are superior to TAG: wax esters have a density of 0.904 g/cm^3 at 5 °C compared with 0.942 for TAG; therefore, the esters provide 45 % more buoyancy. To attain the density of seawater, an animal must contain about 29 % wax esters but 37 % TAG [133]. In the toothed whales, wax esters form the main component of the blubber and are involved in heat insulation. The blubber is a fibre-rich subcutaneous connective tissue that can make up 40 % of the body weight; it contains 50–80 % lipid, which in the sperm whales (Physeteridae) and the beaked whales (Ziphiidae) con-

sists mainly of wax esters. An unexpectedly high TAG content (50 %) found in the blubber of a very young beaked whale *Mesoplodon bidens* is probably related to the changeover from TAG-rich milk to a wax-ester-rich diet [157]. Enormous cavities in the head of the sperm whale contain several tonnes of a fat substance known as spermaceti. This gives rise to spermaceti wax which consists of esters of cetyl alcohol 12:0 with different saturated fatty acids of C_{12} to C_{18}. The large commercial interest in this wax has led to the near extinction of the sperm whale [133]. An extremely long-chain fatty alcohol with chains of 38 to more than 44 C atoms and one to three methyl branches has been detected in the internal, but not the surface, lipids of the pupae of the lepidopterans *Trichoplusia ni* and *Manduca sexta*. The internal fatty alcohols differ from the branched hydrocarbons of the surface lipids in the localization of the methyl branches. Tracer experiments with ^{14}C-acetate and ^{14}C-propionate showed these fatty alcohols to be the most actively synthesized lipids of the pupa. In *M. sexta* they reach a maximum of 550 µg/pupa 9 days after pupation; they are completely absent from the larvae and adults [192].

Hydrocarbons are found in all marine organisms but usually comprise no more than 1 % of the total lipids. Marine phytoplankton contains large quantities of highly unsaturated hydrocarbon compounds, especially heneicosahexane (21:6), which arises by the decarboxylation of the fatty acid 22:6(n-3). Alkanes, alkenes and alkadienes with chain lengths of 17 to 25 C atoms are also found in the tissues of various shark species (Table 15.9). The branched hydrocarbon squalene (Fig. 15.12) was discovered in 1916 in the liver oil of the shark genus *Squalus* and is also present in the tissues of other shark species. The widespread presence of trace amounts of squalene is explained by the fact that it is an intermediate of steroid biosynthesis. The branched hydrocarbon pristane (Fig. 15.12) arises from the chlorophyll component phytol in plant feedstuffs and is therefore also widely found in trace amounts. Copepods of the genus *Calanus* accumulate pristane to unusual levels and then pass it on in the marine food-chain, e.g. to the herring-type teleosts and various shark species [133, 270]. Comparative studies of 11 different shark species showed a surprising variety of storage lipids composed, according to the species, of varying proportions of TAG, ether glycerides, wax esters and hydrocarbon compounds (Table 15.9). The main storage site is the liver, which forms up to 37 % (average of 12 %) of the body weight and can hold more than 90 % of the total lipids [270].

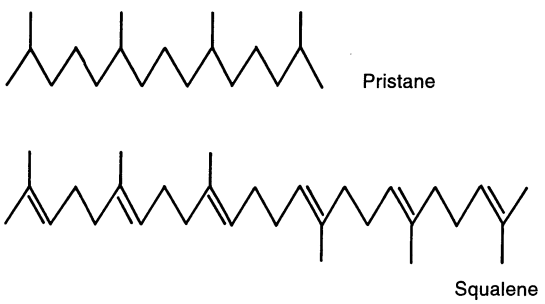


Fig. 15.12. Pristane and squalene

Table 15.9. Percentage composition of storage lipids in the liver and muscle of various Caribbean sharks [270]

	Lip	WE	TAG	Hc	mainly
Dalatias licha	>99	35	19	3	Sq
Centrophorus granulosus	>99	41	14	2	17:2 + Sq
Hexanchus vitulus	>99	0	83	12	17:2 + 18:0
Etmopterus hillianus	94	2	40	34	17:2 + Pr + 17:1
Carcharhinus perezii	93	39	18	11	25:0
Sphyrna lewini (liver)	51	14	20	14	18:0
(muscle)	49	4	2	30	25:0 + Sq

Lip, Organ lipid (% total body lipid); WE, wax ester; TAG, triacylglycerol; Hc, hydrocarbon (the latter three as % organ lipid); Sq, squalene; Pr, pristane

15.2.4 Fat-Rich Secretions

The fat content of **mammalian milk** mostly lies between 2 and something over 10 % of dry matter (Table 13.2, p. 467). The extreme values of more than 50 % found in seals, whales and the grizzly bear are probably related to the large amounts of energy that the young animals require in order to maintain their body temperatures. Milk fat is composed entirely of TAG. Of the more than 150 different fatty acids that have been identified in cows' milk, which, next to human milk [232], is the best investigated, only the even-numbered unbranched fatty acids with 4–18 C atoms are present in proportions of over 1 % (Table 15.4). The fatty acids that are present in trace amounts, including the odd-numbered and branched types, as well as propane-, hydroxy- and oxo-fatty acids, are derived from the diet or, in particular, from the metabolism of rumen bacteria. Medium- to

long-chain fatty acids (C_8 to C_{12}) are found in the milk of all terrestrial mammals, whereas butyric acid (C_4) is present only in the milk of ruminants. The fatty acid spectrum in the milk of marine mammals is similar to that of the storage fat [159].

The **crop milk of pigeons** (Columbiformes) is in many ways comparable with mammalian milk. Male and female pigeons feed the young with a cheese-like white mass from the crop, the secretion of which, like the secretion of mammalian milk, is regulated by prolactin from the hypophysis. Crop milk consists of fat-laden epithelial cells which contain little carbohydrate but much protein and lipid. In addition to about 80 % TAG, the lipids also include phospholipids and cholesterol. The fatty acid spectrum corresponds to that of the body fat and consists of more than 90 % of the usual C_{16} and C_{18} fatty acids [239]. The lipid-rich oesophagal fluids of the flamingo *Phoenicopteris ruber* (total fat content 18 %) and of the penguin *Aptenodytes forsteri* (total fat content 29 %) also serve to feed the young. In contrast, the **stomach oils** found in the proventricules of the albatrosses (Diomedeiae), petrels (Procellariidae) and stormy petrels (Hydrobatidae), but not the diving petrels (Pelecanoididae), which form the fourth family of the order of the tube-nosed swimmers (Procelariiformes), have a variety of functions. These are yellow to brown oils that occur in volumes of up to 200 ml in the proventriculus in both sexes, even outside of the breeding season, and also in young animals. They are fed not only to the young but also to the sexual partner during courtship feeding. Finally, the stomach oils are used in the defence reaction of this avian order; this involves so-called oil spitting, which is also seen in young animals. The composition of the oil is very variable and differs between individuals of the same species. It can therefore probably be assumed that the oils have their origin in foodstuffs, and their variability reflects that of the diet. The most important components, which are found in very different proportions, are TAG and then alkyldiacylglycerols, wax esters and hydrocarbons. The TAG includes large amounts of the fatty acids 20:5 and 22:6, just like in marine fats; this further suggests that the origin of the oils lies in nutrition. The components of the wax esters are predominantly monounsaturated fatty acids and fatty alcohols of C_{14} to C_{22}. Hydrocarbons are seldom found in large amounts but include, above all, squalene, and also pristane (Fig. 15.12) [109].

Lipid-rich nutritional secretions are also found in the insects. The cockroach *Diploptera punctata*, which bears living young, releases a milky secretion into the breeding sac and this brings about a 60-fold increase in the weight of the first larval stage compared with the weight of the egg. The secretion contains proteins as 45 % of the dry weight; 25 % is carbohydrate and 16–22 % lipids, mainly phospholipids and cholesterol. A lipid-rich secretion is produced in the Dufour glands of some hymenopterans; these are associated with the sting apparatus. This secretion serves as food for the larvae of the solitary bees of the genus *Centris*. It consists of odd-numbered alkanes with 19–29 C atoms, and in *C. analis* also contains wax esters with C_8 to C_{14} fatty alcohols and C_2 to C_{14} fatty acids [32]. *Anthophora abrupta* secretes from the Dufour gland a complex mixture of fluid TAGs that always contains a palmitic acid residue and two short-chain acyl residues (C_2, C_4 or C_6). The breeding cavities are lined with DAG, mainly dipalmitoylglycerol. The easily transportable TAGs produced in the Dufour gland are presumably hydrolysed and re-esterified for production of the honeycomb, perhaps by enzymes from the digestive glands. The pheromones produced in the Dufour glands of the ants will be discussed in Chapter 19.

15.3 Lipids of the Body Surface

The surfaces of terrestrial plants and animals are covered with a lipid layer whose general function is to protect the organism from unfavourable environmental factors. In some groups of organisms, the **main function** is protection against dehydration (plants, terrestrial arthropods), whilst in others it is protection against water penetration (birds, mammals). In addition, the lipid layer hinders the entry of bacteria and other foreign organisms, and may play a role in "chemical communication" between members of a species. The term "wax" is often used for the surface lipids, although they include not only wax esters but usually complex mixtures of very different neutral lipids. Despite the structural variety, surface lipids share certain common **properties** that can be considered as functional adaptations. Independent of whether they are liquid or solid, they always have very long hydrocarbon chains and are thus relatively non-volatile. They contain a relatively low number of double bonds, which can be oxidized by atmospheric oxygen, and are particularly resistant to hydrolytic and other metabolic processes.

Based upon their specific adaptations, the surface lipids have unique **structural characters** and differ fundamentally from the internal lipids. Thus, branched hydrocarbon chains are common in the surface lipids but rare in the internal lipids. The hydrocarbon compounds and the mono- and diester waxes of the surface waxes play only a minor role in the body fat of most organisms. Surface lipid preparations may be contaminated with internal lipids during isolation, leading to considerable errors in analytical results. Extensive analytical results are available for the skin fats of mammals, the uropygial-gland lipids of birds and the cuticular lipids of insects, but there are fewer results for the surface lipids of the reptiles and the terrestrial arthropods other than the insects. There have been almost no experimental investigations of the **metabolism** of most surface lipids. Conclusions about the synthesis of the most important components, such as hydrocarbons, fatty alcohols and alkane diols, have been mainly drawn by analogy. This is particularly regrettable as the unique structure of many of these substances makes their synthesis especially interesting. It must be remembered, however, that most surface lipids are produced close to their functional sites in epidermal cells or small skin glands, which are difficult to isolate. An exception is the uropygial glands of birds, with the result that they have become a model for lipid biosynthesis [133].

15.3.1 Epidermal Lipids of Terrestrial Vertebrates

In the epidermis of mammals, birds and reptiles, the innermost layer, e.g. the stratum germinativum of mammalian epidermis, is made up of living cells (keratinocytes) which are similar to other living cells in the complexity of their lipid content. During keratinization, they are transformed into the flattened, dead cells of the outer layer, e.g. the stratum corneum of mammalian epidermis. In the process, lipids are released from the epidermal cells into the intracellular space and become an important barrier to water diffusion. These epidermals lipids include mainly shingolipids of the type O-acyl-glucosylceramide (Fig. 15.13 d) and cholesterol, as well as smaller amounts of free fatty acids and polar composite lipids [133, 276, 277]. To date, little is known about the epidermal lipids of the other terrestrial vertebrates, i.e. birds and reptiles, but they appear to be relatively similar to those of the mammals. Acyl-glucosylsterols were discovered as a component of chicken epidermal lipids; they are probably also present in reptiles and amphibians but not in mammalian epidermis [277]. A relatively uniform spectrum of glycosylceramides, cholesterols, free fatty acids and various phospholipids is found in the discarded skin of various snakes [30, 163].

15.3.2 Skin Fats of the Mammals

The fat substances that can be isolated from human skin or from the fur of mammals arise mainly as secretions of the sebaceous glands. In contrast to the epidermal lipids, the **sebaceous gland lipids** show large species-specific differences in both the content of the various lipid classes (Table 15.10) and the structure of the aliphatic components, i.e. fatty acids, fatty alcohols and diols. Some mammalian families have uniform lipid spectra, e.g. in the dog (Canidae), cattle (Bovidae) and horses (Equidae), but others have widely different species-specific spectra, e.g. raccoons (Procyonidae), squirrels (Sciuridae) and martens (Mustelidae). The water-repellent properties of the sebaceous lipids are uniform and cannot explain the structural variety. However, if, as is now generally accepted, the skin fats play an important role in chemical communication, even closely related species should show distinct differences [48]. The lipids of mammalian hair make up 0.6–1.6 % of the fresh weight and consist mainly of ceramides, cholesterol sulphates and glycosophingolipids [286].

In contrast to their role in the surface lipids of the arthropods and plants, **aliphatic hydrocarbons** are not especially important in the surface lipids of the mammals. A characteristic of human-skin fats is the presence of large amounts of squalene and triacylglycerols, which are otherwise seldom found (Table 15.10). The skin fat of the otter *Lutra canadensis* in fact contains 44 % squalene, which has especially effective water-repellent properties [49]. **Mono- and diester waxes** (Fig. 15.13 a–c) are common components of mammalian skin fats. Two types of diester wax can be distinguished: type I, which contains a hydroxy fatty acid esterified with an unsubstituted fatty acid and a fatty alcohol, and type II, which is a 1,2-diol carrying two unsubstituted fatty acids. Most mammals, however, have only one of the two types; the presence of both is seldom seen (Table 15.11). The skin fats of horses (Equidae)

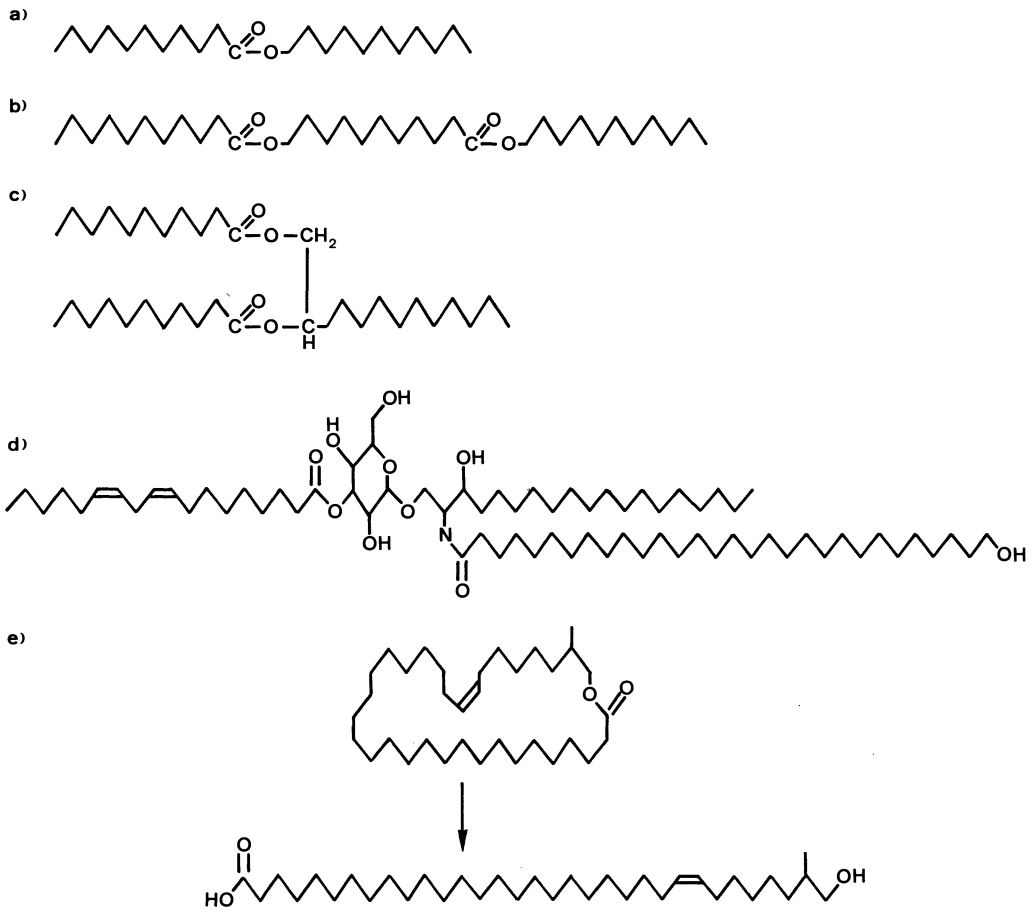


Fig. 15.13a–e. Skin fats of mammals, In **a–c**, the hydrocarbon chains have been shortened and shown without the frequent methyl branches. **a** Monoester wax; **b** diester wax type I (fatty acid + ω-hydroxy fatty acid + alcohol); **c** diester wax type II (fatty acid + 1,2-diol + fatty acid); **d** acylglucosylceramide from pig epidermis with the systematic name 1-(3′-O-acyl)-β-D-glucopyranosyl-N-(ω-hydroxyacyl) sphinganine [276]; **e** giant ring lactone from the skin fat of the horse, and the ω-hydroxy fatty acid released from it by hydrolysis [276]

have a unique composition, and all species contain fats with 47–67% of so-called **equolids** (Fig. 15.13e); these are lactones that arise by cyclization of ω-hydroxy fatty acids with 32–36 C atoms. These substances are probably involved in chemical communication; whatever the case;

Table 15.10. Percentage composition of different lipid classes in the skin fats of humans and other mammals [48, 133, 289]

	TAG	EG	fFA	fFAl	ME	DE	Sq	St	StE	Lac
Human	41		16		25		12	1	3	
Sheep				11	10	21		12	46	
Cow	4					76			3	
Horse								14	38	48
Rat		8	1		17	21	5	6	26	
Mouse	6				5	65		13	10	
Gerbil *(Meriones unguliculatus)*	26		5	9		36		8	10	

TAG, Triacylglycerol; EG, 1-alkyl-2,3-diacylglycerol (ether glyceride); fFA, free fatty acids; fFAl, free fatty alcohols; ME, monoester waxes; DE, di- and triester waxes; Sq, squalene; St, cholesterol and other sterols; StE, sterol ester; Lac, lactone

Table 15.11. Percentage of diester waxes types I and II in the skin fat of various mammals [133, 289]. Type I has a 2-hydroxy acid esterified with a fatty acid and a fatty alcohol, and type II a 1,2-diol with two fatty acids

Species	Type I	Type II
Rabbit	66	–
Domestic cat	66	–
Cow	35	–
Rat	12	14
Sheep	9	7
Mouse	–	67
Dog	–	50
Guinea-pig	–	41
Gerbil	–	36
Baboon	–	21
Human (newborn)	–	3

they show distinct species-specific differences. In the domestic horse *Equus caballus*, the basic ω-hydroxy fatty acids are all branched, but in the donkey *E. asinus* and Grevy's zebra *E. grevyi* they are almost exclusively unbranched; the Mongolian wild horse *E. przewalski*, the mule and the onager *E. hemionus onager* have both branched and unbranched forms [48].

The **aliphatic components** of the skin fats of all mammals examined in detail show an extraordinary variety and large species-specific differences. The wool fat (lanolin) of the sheep contains many more than 100 such compounds, including unsubstituted fatty acids, ω-hydroxy fatty acids, α-hydroxy fatty acids, aliphatic alcohols and 1,2-diols, the chains of which have odd or even numbers of C atoms between 12 and 34, and are unbranched or mostly iso- or anteiso-methyl branched. All these compounds are saturated [133]. A completely different composition of the skin fats is found in domestic cattle *(Bos taurus)*; here, diesters of type I (with hydroxy fatty acids) are found alongside type II (with 1,2-diols) and triesters (with hydroxy fatty acids and 1,2-diols), which contain almost exclusively unbranched even-numbered saturated fatty acids and hydroxy fatty acids with 14 to 18 C atoms. A similar situation is found in the buffalo *Bubalis bubalis*, although the relative proportions of the different fatty acids differ from the proportions in *Bos taurus* [64].

Exclusively unbranched aliphatic components are also found in the skin fats of the gerbil *Meriones unguliculatus*. This animal is unique in containing no monoesters but free fatty alcohols (mainly 16:0 and 18:0) and type-II diesters of a C_{15}- or C_{17}-diol and two saturated or unsaturated even-numbered fatty acids [289]. An immensely complicated spectrum of fatty acids and fatty alcohols is also found in human skin fats. These components have even or odd numbers of C atoms between 12 and 34; they are unbranched or branched with methyl substituents on the last-but-one (iso-), last-but-two (anteiso-) or an internal C atom. They can be saturated or monounsaturated but are seldom hydroxy substituted. The spectrum of unsaturated fatty acids suggests that various 14:0 to 24:0 fatty acids are desaturated (mainly 16:0), that Δ^6 desaturation is more frequent than the otherwise more usual Δ^9 desaturation, and that the desaturated fatty acids are then elongated [133]. In contrast, Δ^9 desaturation predominates in the rat and mouse, which have similarly complex aliphatic skin-fat components [133, 198].

15.3.3 Lipids of the Uropygial Gland

All avian species have uropygial glands. Their secretions confer water-repellent properties on the feather coat and maintain the suppleness of the feathers, which as dead, non-vascularized keratin structures are only replaced after one or more years. However, not all the lipids of the plumage originate in the uropygial gland; a proportion, the details of which are not known, comes from the epidermis. There is a wealth of comparative data on the **composition of the uropygial gland lipids**, and more than 450 of the 9000 known bird species have been examined to date. The main substituents are always monoester waxes. Their component alipathic alcohols and fatty acids have chains with even or odd numbers of C atoms between 8 and 24, often 16 or 18; they are unbranched or branched in various ways, but are always without double bonds. The property of unsaturated fatty acids that they harden in the presence of atmospheric oxygen, as known, for example, from varnish, would be a disadvantage for the lipids of the plumage. Unsaturated fatty acids are present only in the uropygial gland secretion of the non-flying kiwis *(Apteryx sp.)* of New Zealand. In contrast to the monoester waxes, the five types of diester waxes and the triester waxes are restricted to certain orders (Fig. 15.14). The TAGs of the plumage originate mostly from epidermal cells; only in the case of the stork and several of its relatives are TAGs with relatively short-chain fatty acids (C_8 to C_{12}) the main products of the uropygial gland. Steroids are also only present in minor quantities. Cholestanol is the main component of the uropy-

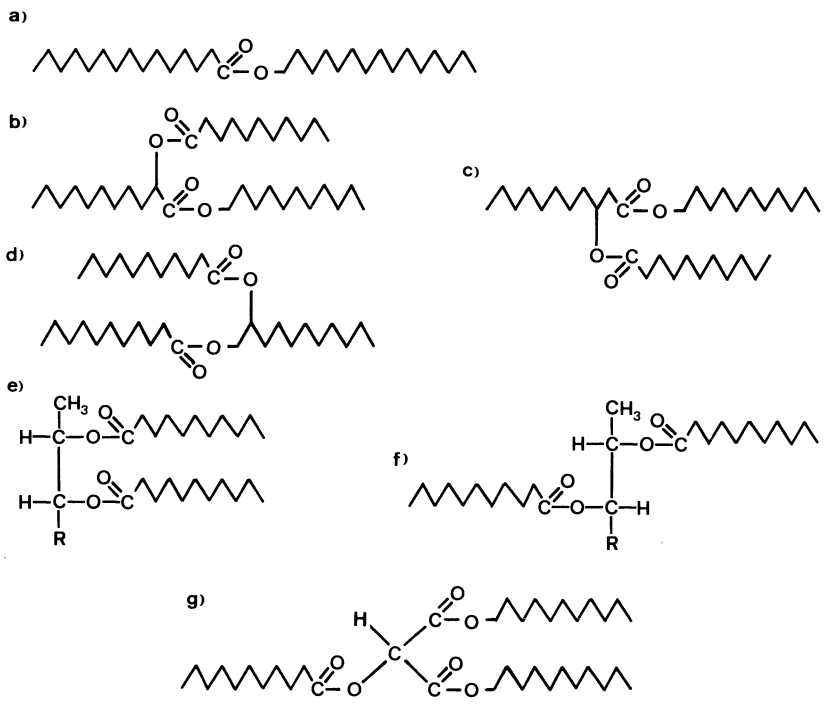


Fig. 15.14a–g. The structure of wax esters from the uropygial gland fats, and examples of their occurrence [108]. The hydrocarbon chains have all been shortened and shown without the commonly present branches. **a** Monoester wax (common); **b** diester wax with 2-hydroxy acid (stork, Ciconiiformes); **c** diester wax with 3-hydroxy acid (pigeon, Columbiformes); **d** diester wax with 1,2-diol (sparrow, Passeriformes; woodpecker, Piciformes; chicken, Galliformes); **e** diester wax with erythro-2,3-diol (chicken, Galliformes; ostrich, Apterygiformes; tinamous, Tinamiformes); **f** diester wax with threo-2,3-diol (chicken, Galliformes), **g** triester wax with hydroxymalonic acid (goose, Anseriformes; crane, Gruiformes; woodpecker, Piciformes; sparrow, Passeriformes)

gial gland lipids from the mallard *(Anas platyrhynchos)*, the wood pigeon *(Columba palumbus)* and (in addition to cholesterol) the marabou *(Leptoptilos crumeniferus)*. Hydrocarbons seldom figure largely; squalene predominates in the Muscovy duck *(Cairina moschata)* and peculiar magpie goose *(Anseranus semipalmata)*, and n-alkanes, n-alkenes and n-alkadienes as well as monomethyl- and dimethyl-branched alkanes occur in some grebes [108].

Aliphatic components with branched chains have a confusingly large number of different structures. They are mostly methyl branched; monomethyl and diethyl-substituted fatty acids are found in the tits (Paridae) and several other birds, and 2-propyl- and 2-butyl-substituted fatty acids are present only in the owls (Strigiformes). The degree of branching of the alcohols is usually less than that of the fatty acids. Fatty alcohols may carry up to four methyl branches, and fatty acids may have as many as five. The simply substituted compounds carry the methyl group mostly on C-2, C-3 or C-4, but also on the last-but-one or last-but-two C atoms (iso- or anteiso- compounds). Methyl groups may be adjacent (e.g. 2,4-, 2,4,6-, 2,4,6,8-, 2,4,6,8,10-) or not (2,6-, 2,6,10-). The unique ability of the uropygial gland to synthesize branched fatty acids and fatty alcohols has, of course, aroused considerable interest. Tracer experiments carried out in vivo and in vitro, especially in the domestic goose, have shown that even- and odd-numbered unbranched and methyl-branched fatty acids are formed with acetyl-CoA and propionyl-CoA as the primer and malonyl-CoA and methylmalonyl-CoA as elongation substrates. The reduction of fatty acids to alcohols as well as the biosynthesis of alkane-1,2-diols and alkane-2,3-diols has also been demonstrated in the uropygial gland. Several of the enzymes involved have been isolated and studied in detail (p. 575). However, many details of the biosynthesis of uropygial gland lipids remain unclear [108].

The majority of uropygial gland lipids are built of complicated mixtures of various aliphatic components; in the admittedly extreme case of the coot, 54 different fatty acids and 44 fatty alcohols have been described. As the components can apparently be more-or-less freely combined, an enormous number of different wax esters is possible. It is seldom that one finds the simple situation that pertains in the kingfisher *(Alcedo atthis)*, where 90% of the uropygial gland lipid consists of 2,4,6,8-tetramethylundecyl-2,4,6,8-tetramethyl-undecanoate. It seems reasonable to examine how far the chemical properties of the uropygial gland lipids are **taxonomic characters**, in particular the structure of the fatty alcohols and fatty acids in the monoester waxes, the presence and type of diester waxes, and the presence

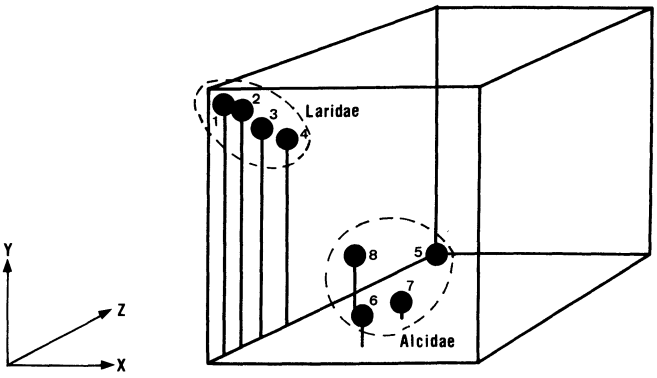


Fig. 15.15. Taxonomic interpretation of the fatty acid spectrum of uropygial gland lipids [110]. Each *circle* corresponds to a seagull species (Laridae) or guillemot species (Alcidae). X-axis: proportion of 4-, 4,x- and 4,x,y-methyl-substituted fatty acids. Y-axis: proportion of 2-, 2,x- and 2,x,y-methyl-substituted fatty acids. Z-axis: proportion of unbranched fatty acids

of other classes of lipid. These characters are, in general, typical of the order, and sometimes of the family; subspecies, and usually also species, cannot be distinguished. For chemotaxonomic purposes these characters can be arranged in a three-dimensional system, the axes of which represent the proportions of the differently branched fatty acids (Fig. 15.15). Several difficult problems of bird taxonomy have been solved in this way. Thus, the predominance of unbranched fatty alcohols and fatty acids suggests a closer relationship of the condor *(Vultus gryphus)* to the storks (Ciconiiformes) than to the diurnal birds of prey (Falconiformes); this conclusion is supported by certain anatomical and behavioural characters. The joint possession of the unusual erythro-alkane-2,3-diol places the kiwis (Apterygiformes) and the tinamous (Tinamiformes) closer to the fowls (Galliiformes) than to the ostriches (Struthioformes) [108, 110, 111]. However, the large individual variability observed in several species suggests caution in the use of uropygial gland lipids in taxonomy.

15.3.4 Cuticular Lipids of the Insects

The insect **cuticula** has a complicated layered structure (p. 391). The outermost layer is the thin epicuticula with four sublayers, referred to from the inside to outside as the inner and outer epicuticulae, wax layer and cement layer. Little is known about the molecular structure of the epicuticula as a whole. On the other hand, there is a wealth of information about the cuticular lipids, which not only are found in the wax layer but impregnate the whole epicuticula. These lipids are produced in special cells, oenocytes, which arise from epithelial cells but can also lie deeper in the body cavity. Lipophorin (p. 195) is involved in the transport of cuticular lipids from the site of synthesis to the site of deposition. Movement of lipids within the cuticula is within pores that lead from the epidermis to the epicuticula [21, 22, 154]. Unfortunately, the composition of the cuticular lipids of terrestrial arthropods other than the insects has been little investigated but appears to be basically similar.

The **main function of the cuticular lipids** is the prevention of dehydration. The removal of the lipids with organic solvents or mechanically with abrasive powders drastically increases the transpiration rate. In addition, the wax layer also protects against penetration by microorganisms. Finally, cuticular hydrocarbons may serve in chemical communication, as described in detail in Chapter 19. Thus certain alkanes or alkenes produced by the females of some Diptera, Hymenoptera and Coleoptera function as sexual pheromones, e.g. (Z)-9-tricosene in the housefly *Musca domestica*, (Z)-7-heneicosene and (Z)-7-tricosene in the beetle *Aleochara curtula*, various (Z,Z)-9-19-alkadienes in the ichneumon *Pikonema alaskensis* or 15,19,23-trimethyl-heptatriacontane in flies of the genus *Glossina* [10, 154, 191, 204]. (Z)-10-Heneicosene is produced by the males of *Drosophila virilis* and functions as an aggregation pheromone [11]. Cuticular lipids in termites serve in species and caste recognition, and the cuticular lipids of the rove beetle *Trichopsenius frosti*, which inhabits the nests of *Reticulotermes flavipes*, are at least very similar to those of the host [154]. The large amounts of cuticular waxes produced by two insect groups have become commercially important: in the honeybee and relatives they serve as the building material for the combs, and in the mealy bugs (Coccinea) the sedentary female is protected by a shield made out of wax and discarded larval skins.

The cuticular lipids of the insects comprise very different classes of polar and non-polar

lipids in species- and group-specific proportions (Table 15.12). Extreme variability, however, is found within the individual lipid classes, in particular the hydrocarbons and the aliphatic components of the wax esters. As in the case of the surface fats of the mammals and birds, this confusing **variety of cuticular lipids** is not explained by their biological function. A specific role in chemical communication can be assigned to only a few compounds, and as substances preventing transpiration the individual hydrocarbons or wax esters are not so different that each species must have a different spectrum. Thus, the variety of surface lipids remains an intriguing puzzle. In contrast to the role of the skin fats of the mammals or the uropygial gland lipids of the birds, hydrocarbons form an important, often dominant component of the insect cuticular lipids; they quite often amount to more than 90% of the total. Only exceptionally are they a minor component, e.g. in the pupae of the moth *Heliothis virescens* they constitute 0.5% [21, 22, 154]. Because of their importance, and also probably because of the ease with which they can be analysed, the hydrocarbons of insect cuticulae have been more extensively investigated than have the polar lipids, the ester waxes, free fatty acids, primary and secondary alcohols, ketones, epoxides, sterols, sterol esters and triacylglycerols. The cuticular hydrocarbons are now known from more than 100 insect species [154].

The **hydrocarbons of the insect cuticula** include unbranched alkanes, alkenes, alkadienes and monomethyl- to tetramethyl-substituted alkanes. A complex mixture of such substances is found in almost every insect species, and with appropriate methods these can usually be separated into more than 50 different compounds [19, 154, 156]. The relative proportions of the different molecular types differ not only between the insect species but also between males and females, larvae and adults, and even between adults of different ages. The unbranched chains have lengths of C_{21} to C_{37} and almost always an odd number of C atoms. They apparently arise by elongation and decarboxylation from common even-numbered fatty acids. **n-Alkanes** are present but seldom predominate, e.g. in the beetles, where 97–98% of the cuticular hydrocarbons are of this type [154]. In many insects, there are both n-alkanes and **n-alkenes**, the double bonds of which are usually from C-7, C-9 or C-11 but can also be at other positions. Occasionally, one finds complete homologous series of alkenes, e.g. a series in males of the ichneumon *Pikonema alaskensis* of 23:1 to 33:1, and in various beetles of the genus *Sitophilus* of 25:1 to 47:1 [10, 154, 160, 189]. Long-chain **n-alkadienes** and **n-alkatrienes** are only minor components in most insects. However, various *Sitophilus* species have up to 57% alkadienes of 31:2 to 39:2, with the first double bond at C-7 or C-9 and the second double bond, depending upon the chain length, in the region of C-17 to C-25 [189]. 6,9-Alkadienes are also major components in *Periplaneta americana* and *P. japonica*, whereas methyl alkanes predominate in other cockroach species [21, 22, 154]. *Drosophila pseudoobscura* contains 27% alkadienes and 9% alkatrienes with double bonds mainly in the 5,9- and 5,9,17- or 5,9,19- positions, respectively [19].

2-Methylalkanes (iso-) and **3-methylalkanes** (anteiso-) are the most common cuticular hydrocarbons. They mostly occur in complicated mixtures, and often in complete homologous series [154, 156]. The 2-methylalkanes have an even or odd number of C atoms in the chain, and the 3-methylalkanes have mostly an odd number. The chain occasionally has a double bond, e.g. in 2-methyl-24-hexatriacontene of the house-cricket *Acheta domestica*. **Internally branched mono-**

Table 15.12. Percentage of different lipid classes in the cuticular lipids of various insects[a] [107, 114, 133]

	1	2	3	4	5	6	7	8	9
Hydrocarbon	12	90	31	53	60	21	10	5	62
Wax ester	4		65	10	28	10	13	16	
Triacylglycerol	7	7			1	21*			16
Free fatty acids	49	2		17	6	7	5		7
Free fatty alcohols			4	3	2	38	58	70	
Sterol and sterol esters	18	1			1	3	1		
Other lipids	10			17	2		13	9	15

* Including di- and monoacylglycerols

[a] 1, Plecoptera: *Pteronarcys californica; 2, Blattodea: Periplaneta australasia, P. brunnea* and *P. fulginosa*; 3, *Arenivaga investigata*; 4, Orthoptera: *Anabrus simplex*; 5, Orthoptera: *Melanoplus sanguinipes*; 6, Hemiptera: *Triatoma infestans*; 7, Coleoptera: *Tenebrio molitor*; 8, Lepidoptera: *Cynthia ricini*; 9, Diptera: *Lucilia cuprina*.

methylalkanes and dimethylalkanes are also widely found, although they are completely absent from, for example, *Drosphila melanogaster* and *Apis mellifera*. Complete series may often be found based upon the localization of the substitutions, e.g. 5- to 15-methyl-nonacosane and 4- to 14-methyl-triacontane in the beetle *Physadesmia globosa* [154, 156]. In the di-substituted hydrocarbons, the substitutions are often separated by three C atoms, e.g. they occupy positions 9,13- to 17,21-. There are other possibilities, e.g. 11,19- in the moth *Heliothis* or 11,12- in the cigarette beetle *Attagenus megatome*. To date, **trimethylalkanes** have only been detected in about 15 insect species, e.g. the tobacco hornworm *Manduca*, the orthopterans *Schistocerca* and *Hemideina*, the stable fly *Glossina* and several ant species [77, 155, 156]. **Tetramethylalkanes** are present in *Glossina* sp. and the potato beetle *Leptinotarsa decemlineata* [161, 190].

The variety and species-specificity of the cuticular hydrocarbons make them very useful for **taxonomic purposes.** For example, investigations of 22 beetle species from the family Tenebrionidae, 6 species of the termite genus *Odontotermes* and 3 species of the ant genus *Iridomyrmex* have clearly demonstrated the suitability of the hydrocarbon spectrum for species diagnosis; a single gas chromatogram is in many cases sufficient [116, 154, 156]. However, cuticular hydrocarbons are not appropriate for the analysis of phylogenetic relationships; for example, cluster analysis of hydrocarbon spectra from 11 species of the *Drosophila virilis* group often show great similarity between closely related species but no really satisfactory agreement with the already well-charted family relationships [11]. As with other low molecular weight biological substances, the causal connections with the genetic information are so long and complicated that no unambiguous correlation with phylogenetic relationships can really be expected.

Wax esters are only minor components of the cuticular lipids of most insects (Table 15.12). As a rule, they are monoesters of unbranched, mostly saturated fatty acids with similar primary alcohols, both of which have unusually long chains with up to 34 C atoms. A novel type of wax ester, as yet unique to the grasshoppers of the genus *Melanoplus*, has been described (Fig. 15.16a). These are esters of secondary alcohols with 21 to 27 C atoms and hydroxyl groups that are mostly close to the middle of the chain [21, 22, 154]. The wax used to build the comb of the honeybee *Apis mellifera* is produced in wax glands on the ventral side of the abdomen. In addition to hydrocarbons, the comb wax mainly contains wax esters with an extraordinary variety of types and individual compounds made up in part of methyl-branched fatty acids, ω-hydroxy fatty acids, primary alcohols and diols (Fig. 15.17). The lipids found in other regions of the cuticula contain the same components but with a greater proportion of hydrocarbons (Table 15.13). Thus, the wax glands are specialized for the production of wax esters and the remaining epidermis is specialized for the synthesis of hydrocarbons. Examination of the incorporation of radioactive acetate into wax esters indicates that the wax glands have their highest activity in 12- to 18-day-old bees [18, 21, 22, 154].

Other species of the family Apidae have much simpler waxes. Thus, bifunctional components such as hydroxy fatty acids and diols are present in other species of the subfamily Apinae but are lacking in the bumble bees (Bombinae), e.g. *Bombus rufocinctus*. The waxes of the stingless Meliponiae are extremely simple in composition (Table 15.13), consisting mostly of saturated n-alkanes with 25–33 C atoms [20, 173]. *Cerogenes auricoma* and other cicadas from the family Fulgoridae, the aphid *Prociphilus tesselatus* and the soft scales of the genus *Dactylopius* develop feather-like wax structures on the abdomen as

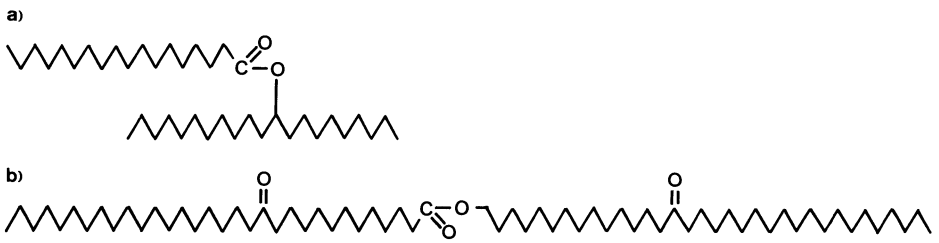


Fig. 15.16a, b. Unusual cuticular lipids of insects. **a** Monoester wax with a secondary alcohol carrying a hydroxy group, in approximately the middle of the chain, and a fatty acid, from the grasshopper *Melanoplus* sp. [21]. **b** Monoester wax with an oxo-fatty acid and an oxo-alcohol, from the aphid *Prociphilus tesselatus* [61]

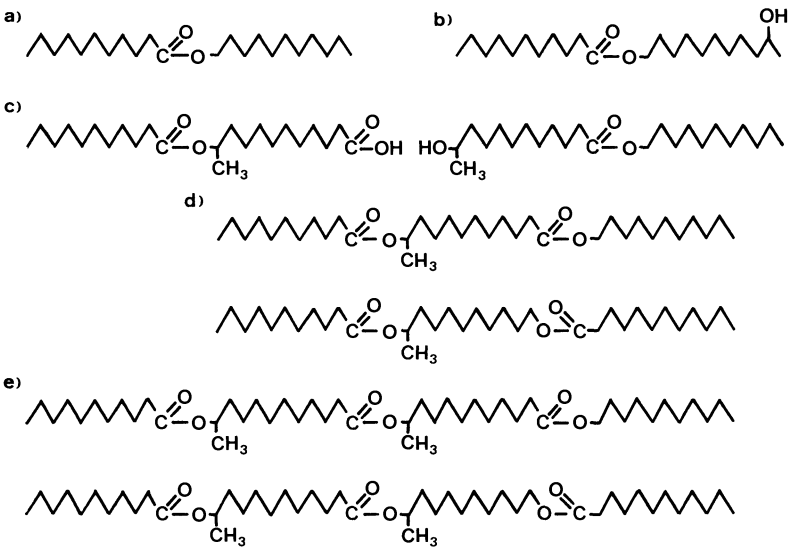


Fig. 15.17a–e. Wax esters from the honeycomb wax of the honeybee [133]. (The components are: S, fatty acid; A, alcohol; HS, (ω-1)-hydroxy fatty acid; D, 1,(ω-1)-diol.) **a** Monoester wax of the form S-A; **b** hydroxy monoester wax of the form S-D or HS-A; **c** acidic monoester of the form S-HS; **d** diester wax of the form S-HS-A or S-D-S; **e** triester wax of the form S-HS-HS-A or S-HS-D-S

protection against predators and parasites. The wax material in all these species consists of C_{30} or C_{32} oxo-acids and C_{34} or C_{36} oxo-alcohols with central keto groups (Fig. 15.16b) [164]. Oxo-alcohols with 26–28 C atoms and oxo-groups in positions 11 or 12 have also been found in pupae of the moth *Manduca sexta;* these are mostly esterified with acetoacetic acid, acetic acid or 3-hydroxybutyric acid; the corresponding oxo-aldehydes are also found [28].

The endogenous synthesis of cuticular hydrocarbons and wax esters has been demonstrated in many insect species by tracer experiments with ^{14}C-acetate. However, little is known about the mechanisms of biosynthesis or the enzymes involved [21, 22, 154]. As in many insects, acetate is also rapidly incorporated into hydrocarbons in the millipede *Graphidostreptus tumuliporus*. Corresponding experiments with the spider *Aphonepelma* sp. and the scorpions *Centruroides sculpturatus* and *Paruroctonus mesaensis*, however, gave negative results. In these cases, it is not known whether the hydrocarbons are of exogenous origin or simply have very low turnover rates. The **formation of extremely-long-chain hydrocarbons** apparently occurs by the elongation and ensuing decarboxylation of fatty acids and not, as in bacteria, by head-to-head condensation and reduction. Thus, these insect hydrocarbons usually have an odd number of C atoms. The decarboxylation of long-chain fatty acids to alkanes has been shown directly in cockroaches and termites by tracer experiments. However, whereas extremely-long-chain fatty alcohols and hydrocarbons are detected in insects, the corresponding fatty acids are not found; this suggests that reduction or decarboxylation occurs immediately after elongation [267]. The main site of hydrocarbon biosynthesis is the oenocytes derived from integument cells, although the integument cells themselves also have the capacity to synthesize long-chain fatty acids and hydrocarbons. The surface

Table 15.13. Percentage composition of the cuticular lipids and comb wax of the honeybee *Apis mellifera* and of the comb wax of the stingless bees *Trigona atomaria* and *T. australis* [18, 21, 173]

	Cuticula	Comb		
	Apis	*Apis*	*T. atom.*	*T. austr.*
Hydrocarbons	58	13–17	71	90
Monoester	23	31–35	26	6
Diester	9	10–14	–	–
Triester	2	3	–	–
Strongly polar lipids[a]	9	34	3	4

[a] Free fatty acids and fatty alcohols, and in *Apis* also diols, hydroxymonoesters, hydroxypolyesters and acid monoesters

lipids are lost with the exuviae and the integument produces a new generation of lipid-synthesizing oenocytes after every moult [115, 156].

The **biosynthesis of branched hydrocarbons** has also been investigated during tracer experiments with insects [39, 74]. The fatty acid precursors of the 2-methylalkanes (iso-) with an even number of C atoms apparently arise with isobutyryl-CoA (from valine) as the primer, and those with an odd number of C atoms have isovaleryl-CoA (from leucine) as the primer (Fig. 15.16). The 3-methylalkanes of insects, however, are not synthesized, as in plants and microorganisms, with 2-methylbutyryl-CoA (from isoleucine) but by the use of methylmalonyl-CoA as the elongation substrate in the second step of fatty acid synthesis (Fig. 15.18). Internal methyl substitutions are also introduced with methylmalonyl-CoA as the elongation substrate [39]. Methylmalonyl-CoA is formed in various ways in the insects. In the termite *Zootermopsis angusticollis* the methyl branches of the alkanes originate from succinate, which is converted to methylmalonyl-CoA via succinyl-CoA. Because of the poor protein diet of the termites, they have insufficient amino acids with branched chains; on the other hand, *Zootermopsis* possesses large amounts of vitamin B_{12}, which is necessary as the prosthetic group for methylmalony-CoA mutase in the conversion of succinyl-CoA to methylmalony-CoA. The housefly *Musca domestica*, the cockroach *Periplaneta americana* and other insects have a protein-rich diet, they have relatively little cobalamin, and in these species methylmalonyl-CoA is produced by valine metabolism [21, 22, 81].

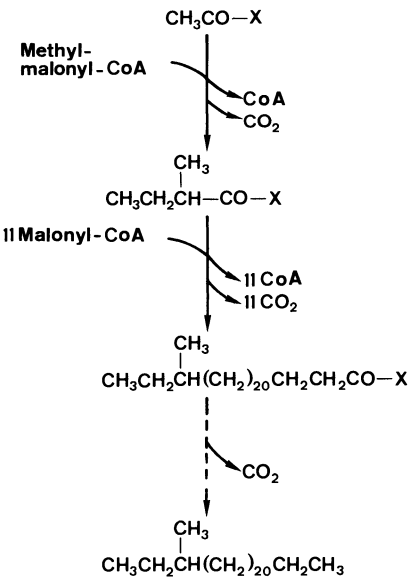


Fig. 15.18. Biosynthesis of 3-methylalkanes in insects with the use of methylmalonyl-CoA as elongation substrate [21]. *X*, Enzyme

Little is known about the **biosynthesis of the polar cuticular lipids**. The production of wax esters has mainly been examined in *Apis mellifera*, where it occurs predominantly in the microsomes. The reaction requires acyl-CoA and free alcohol; alcohols with 16 or 18 C atoms are metabolized more quickly than those with 24 or 28 C atoms. Thus, the predominance of long-chain alcohols in beeswax is due more to the specificity of the fatty acid reductase than of the acyltransferase. ^{14}C from acetate or palmitate appears in *Apis mellifera* in the fatty acids and alcohols of the wax esters; conversely, however, long-chain alcohols are not oxidized to acids. In the grasshopper *Melanoplus sanquinipes*, n-alkanes are oxidized, apparently by a mixed-function oxygenase, to the secondary alcohols that are characteristic of this animal [21, 22, 154].

15.4 Membrane Lipids

Lipids with structural functions are found especially as components of cellular membranes. The most important membrane lipids are the glycerophospholipids, sphingophospholipids (sphingomyelins), glycolipids and sterols. According to the **fluid-mosaic model**, developed by Singer and Nicolson in 1972, lipids in the membrane form a double layer in which membrane proteins can move freely in two dimensions. However, the lipids have functions other than the formation of a matrix for membrane proteins. Only in this way is it possible to explain why the lipids of some membranes have such a complex composition with often more than 100 different molecular species, and why the differences in the lipid spectrum between animal species, cell types and even membrane types are so great (Tables 15.14–15.16). The physicochemical properties that are important for the biological functions of membranes are not only determined by their chemical structure but also influenced in a complex manner by pH, temperature, hydration, the binding of divalent metal ions, and interactions with other molecules such as sterols, proteins and drugs. The ratio of protein to lipid can vary in different membranes from approximately 1:1 to 4:1 (Table 15.14) [124].

Table 15.14. The lipid composition of cellular membranes from the rat liver [55]. The proportions of the individual phospholipid classes are given as a percentage of the total phospholipid

	Mit	Mic	Lyo	PM	NM	GM
Proportion (mol %)						
Cardiolipin	18	1	1	1	4	1
Phosphatidyl-ethanolamine	35	22	14	23	13	20
choline	40	58	40	39	55	50
inositol	5	10	5	8	10	12
serine	1	2	2	9	3	6
Phosphatidic acid	–	1	1	1	2	1
Lyso-glycerophospholipid	1	11	7	2	3	3
Spingomyelin	1	1	20	16	3	8
Amount (μg/mg protein)						
Phospholipids	175	374	156	672	500	825
Cholesterol	3	14	38	128	38	78

Mit, Mitochondria; Mic, microsomes (endoplasmic reticulum); Lyo, lysosomes; PM, plasma membrane; NM, nuclear membrane; GM, Golgi membrane

The critical **structural characters of membrane lipids** are the "backbone" glycerol or sphingosine, the polar head groups and the hydrocarbon chains. The head groups can be zwitterionic ("neutral" phospholipids), anionic or uncharged. The hydrocarbon chains can have lengths of 12–24 C atoms and be either saturated or unsaturated with up to six double bonds. The rigid double bonds with the cis configuration bring about angular molecular structures which take up more space in the membrane. Fatty acids with branched chains are not found in animal cell membranes. The asymmetry of the lipid composition in the two membrane layers is particularly marked in the outer plasma membrane, but is also found in intracellular membranes and is of great importance, e.g. for the insertion of membrane proteins, the connection to the cytoskeleton and interactions with other cells [57]. Even though membrane structure and function may appear to be constant under a certain set of conditions, the membrane lipids are constantly in motion. Each lipid molecule changes place with its neighbour in the same monolayer approximately 10^7 times per second; exchanges with the other monolayer are rare. The diffusion coefficients of the membrane components are of the order of 10^{-8} cm^2/s, corresponding to a velocity of about 2 μm/s. Transport processes between different membranes or intracellular compartments are required especially to bring the membrane lipids built in the endoplasmic reticulum to the site of their incorporation. The transport of lipids in the cytoplasm involves either their attachment to specific proteins or their inclusion in vesicles [124].

Table 15.15. The lipid composition of different membranes of the ciliate *Tetrahymena pyriformis* WH-14 [124]. The proportions of the individual lipid classes are given as a percentage of the total phospholipids, and the concentration of tetrahymanol is given as mol/100 mol phospholipid

	Cil	Mit	Mic	Pell	NM
Proportion (mol %)					
Cardiolipin	1	10	1	2	3
Phosphatidyl-ethanolamine (PE)	11	35	34	34	26
choline (PC)	28	35	35	25	31
Aminoethylphosphonolipid (AEPL)	47	18	23	30	23
Lyso-PC	1	2	1	5	6
Lyso-PE and -AEPL	9	–	3	3	6
Glycerylether	53	25	18	33	–
Concentration of tetrahymanol	30	4.8	4.1	8.4	3.6

Cil, Cilia; Mit, mitochondria; Mic, microsomes; Pell, pellicula; NM, nuclear membrane

The **membranes of vertebrate erythrocytes** (see Fig. 10.9, p. 360) are particularly stable and can be easily isolated after haemolysis (erythrocyte ghosts). Like most membranes of animal cells, their main lipid components are the zwitterionic phospholipids phosphatidylcholine, phosphatidylethanolamine and sphingomyelin, and in addition include anionic glycerophospholipids, such as phosphatidylserine and phosphatidylinositol, glycolipids and sterol. The lipid spectrum of vertebrate erythrocytes from the agnathans to the mammals is, in principle, very similar, although the fatty acid spectrum of marine fish differs from that of the terrestrial vertebrates by having a high proportion of highly unsaturated C_{20} and C_{22} fatty acids. Haemoglobin-containing blood cells are also found in the annelids, sipunculids and molluscs; however, their membrane lipids have only been examined in the marine mussel *Scapharca broughtoni*. Unlike the erythrocytes of the mammals, those of the mussel contain considerable quantities of ceramide aminoethylphosphonate; they are further interesting in that they do not interact with phospholipase C, streptolysin, Staphylococcus toxin or encephalitis virus, as do the mammalian erythrocytes [41].

The easily bred **ciliates** of the genus *Tetrahymena* offer a useful single-cell system for the study of membranes; they have been used in particular for studies of adaptation to temperature.

Table 15.16. The lipid composition (mol %) of the mitochondria from different organs in the rat (R) and sheep (S) and from the flight muscles of the housefly *Musca domestica* [55]

	Liver		Heart		Kidney		Brain	Muscle
	R	S	R	S	R	S	R	Fly
Cardiolipin	12–17	14	8	10–15	9–20	13	2	25
Phosphatidyl-								
ethanolamine	22–36	19	44	16–35	30–37	22	32	59
choline	34–55	48	41	43–46	36–41	41	55	8
inositol	5– 8	4	3	1– 5	3	4	5	9
serine	1	1	1	1	1	2	4	
glycerol	1			4	3			
Lyso-glycero-phospholipid	1– 2	27	1		1			
Sphingomyelin	1– 3		1	2	1	5		
Plasmalogen	2	3		36		9		

However, the cell surface (pellicula) of *Tetrahymena* is atypical of animal cells: beneath the plasma membrane, which also covers the cilia, are two further membranes, the inner and outer alveolar membranes. The membranes of the cell surface and the cell organelles are significantly different in their lipid composition but all have phosphatidylcholine and phosphatidylethanolamine, as well as the unusual aminoethylphosphonolipid (Table 15.15). The usual sterols are missing, but the rather odd pentacyclic tetrahymanol is present, especially in the cilia membrane (see Fig. 16.3, p. 626) [124]. The taurolipids of *Tetrahymena* are also rather unique. In these, a tri-, tetra- or pentahydroxystearic acid is bound to the amino group of taurine by an amide bond, and itself is esterified with one or two non-hydroxylated fatty acids (Fig. 15.19a). *T. mimbres* (formerly *T. pyriformis NT-1)* contains taurolipid A and 7-acyl-taurolipid A, in which the trihydroxystearic acid in positions 3 and 7 is esterified. *T. thermophila* possesses both taurolipid B with tetrahydroxystearic acid and taurolipid C with pentahydroxystearic acid. Tracer experiments have shown that taurine and stearic acid initially give rise to lipotaurine with dihydroxystearic acid, and that further hydroxylation yields taurolipids A, B and C [126, 127]. Unusual lipids have also been found in other protozoans, e.g. sulphoquinovosyl-diacylglycerol in the marine luminous *Noctiluca miliaris* (Fig. 15.19b); this lipid is otherwise mainly known from the chloroplasts of algae and higher plants [59].

Amongst the intracellular membranes of animals cells, the **mitochondrion**, with its special properties and its importance for electron transport and oxidative phosphorylation, has been thoroughly investigated. The lipid spectrum in the mitochondria of plants and animals is characterized by a scarcity of sterols and a high content of the acidic diphosphatidylglycerol (cardiolipin) (Tables 15.14–15.16). Cardiolipin is localized above all in the inner mitochondrial membrane but is also found in the outer membrane; the inner membrane is practically free of phosphatidylinositol and phosphatidylserine [100]. Numerous mitochondrial proteins are bound to lipids or are activated by lipids. Despite all the fine-structural and functional similarities between the

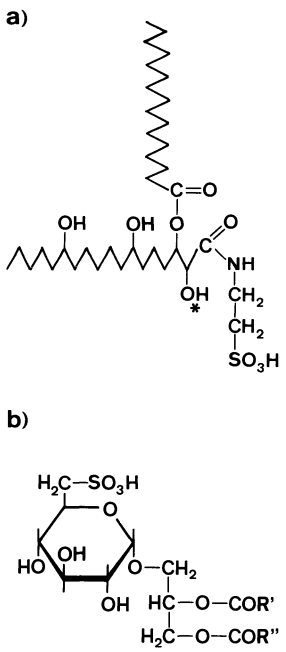


Fig. 15.19a, b. Unusual lipids of protozoans. **a** Taurolipids A and B from *Tetrahymena thermophilia* [125]; taurolipid A differs from B only by the absence of the hydroxyl group marked *. **b** Sulphoquinovosyl-diacylglycerol from *Noctiluca miliaris* [59]: quinovose is 6-deoxy-D-glucose; the molecule contains a carbohydrate-sulphur bond(!)

mitochondria, there are considerable organ- and species-specific differences [55, 100].

The mobility of membrane components, which is apparently important for the various functions, is described by terms like **membrane fluidity**, membrane dynamics or the converse, membrane viscosity. Membrane fluidity is determined by the structural arrangement of the lipid components. The relationships between lipid composition, fluidity and biological function form one of the most active research areas of cell biology [124, 287]. The fluidity of each cell membrane is temperature dependent. At the so-called transition temperature (T_c) the membrane lipids suddenly change to higher structural order (crystalline) in which the mobility of the phospholipid molecules is drastically reduced. T_c is dependent upon the relative proportions of the different lipid classes and directly proportional to the length and degree of saturation of the fatty acids.

The homeothermic mammals and birds have a constant high body temperature (except during hibernation and conditions of lethargy). The cells of all other organisms must be capable of adapting to changes in body temperature. In the process, the lipid composition of the membranes is so modified that membrane fluidity remains constant. This **homeoviscosity adaptation** was discovered in 1974 by Sinensky in *E. coli*, and was demonstrated in animals on the synaptosomes of the goldfish brain in 1977. The process is apparently similar in all organisms but has been studied in particular in bacteria, the ciliate *Tetrahymena* and fish. Adaptation to temperature by cellular membranes involves changes in the relative amounts of the individual lipid classes as well as changes in the degree of saturation and in the chain length of the fatty acids, especially those of the phospholipids. During cooling, phosphatidylcholine, including the related 1-alkenyl compound (plasmalogen), is reduced in favour of phosphatidylethanolamine [37, 43, 92, 287]. In fish and crustaceans, phosphatidylcholine is deacylated to lyso- compounds during cold periods [37]. In the same conditions, both phosphatidylcholine and phosphatidylethanolamine are reduced in *Tetrahymena*, whereas aminoethylphosphonolipid (AEPL), lyso-AEPL and lysophosphatidylethanolamine are increased. Investigations with *Tetrahymena* have shown that the alterations in the lipid spectrum are regulated independently of the changes in desaturation of fatty acids and, furthermore, occur to very different extents in different membranes; for example, they are much more drastic in the cilia membranes than in the microsomes [124, 216]. The rapid reduction of temperature from 39.5 to 15 °C, which was typical in the *Tetrahymena* experiments, also appears to effect a reduction in the tetrahymanol content [124]; this corresponds to the lowering of cholesterol observed in the carp [124, 287]. A decrease in the degree of saturation of the fatty acids in membrane lipids has been recorded in cold-adapted bacteria, higher plants and various animals, and thus appears to be widespread (Table 15.17). This phenomenon has been particularly well investigated in the ciliate *Tetrahymena* [122, 124], in the carp, goldfish and other teleosts [43, 148, 287], and also, for example, in the molluscs, crustaceans and beetles [16, 261]. A particularly high potential for adaptive desaturation is found in the teleosts and many aquatic invertebrates, where the membrane lipids are rich in fatty acids of the (n-3) series; in contrast, mammals mainly have linoleic acid 18:2(n-6) and arachidonic acid 20:4(n-6). The (n-3) fatty acids can of course carry more double bonds than can the (n-6) series [37]. The temperature-dependent differences in the degree of saturation are usually most marked in the phospholipids; however, comparative investigations of the brains of 39 teleost species showed that differences in fatty

Table 15.17. Homeoviscous adaptation of the fatty acid spectra of membrane lipids. The percentage fatty acid composition of the total phospholipids of the ciliate *Tetrahymena pyriformis* [124] and of the liver microsomes of the short sunfish *Lepomis cyanellus* [43] after adaptation to different temperatures

Fatty acid	*Tetrahymena*		*Lepomis*	
	15 °C	39.5 °C	15 °C	25 °C
12:0	0.6	0.9	–	–
14:0	6.9	6.5	–	–
15:0	1.5	4.6	–	–
16:0	8.9	12.6	15.4	19.8
16:1	8.7	8.7	11.4	8.9
17:0	2.2	4.4	–	–
17:1	0.8	1.4	–	–
18:0	0.6	2.1	6.4	7.2
18:1	9.6	10.6	35.8	32.0
18:2	27.2	17.9	8.5	8.1
18:3	31.3	24.5	2.6	2.8
19:0	Tr	2.6	–	–
20:4	–	–	1.9	2.0
20:5	–	–	0.9	0.9
22:6	–	–	13.0	13.7
Saturated	21	34	23	28
Monoenic	19	21	47	42
Polyenic	59	43	29	29
Sat./unsat.	0.26	0.53	0.31	0.40

Tr, Trace

acids between cold-water and warm-water species were even higher in the gangliosides [6].

As position sn-2 of the glycerophospholipids is in any case mostly occupied by an unsaturated fatty acid, this must be exchanged for an even more unsaturated fatty acid; in position sn-1, there is an increase in the proportion of monoenic acids. Desaturation often coincides with an increase in chain length; this occurs, for example, in *Tetrahymena* and the rainbow trout [92, 124]. Two mechanisms are proposed for the adaptive desaturation of fatty acids in membrane lipids. The direct stimulation of desaturases by a cold-dependent decrease in fluidity has been postulated for bacteria and *Tetrahymena* but has not yet been demonstrated. In contrast, the induction of desaturases as a result of a decrease in temperature has been clearly shown in bacteria, *Tetrahymena* and the carp [66, 122, 287]. The cells of the homeothermic mammals are generally not exposed to large fluctuations in temperature; they become dysfunctional at temperatures below 10–15 °C. During hibernation, the body temperature may fall to a few degrees above zero, and adaptations of cell membranes are to be expected in these conditions. In fact, only minor changes that are indicative of homeoviscous adaptation of the membranes of hibernating animals are discernible; the issue undoubtedly requires further study [2].

Because cooling lowers the pH of the body fluids (Δ pH per °C ranges from -0.015 to -0.013) there are interacting effects of pH and temperature on membrane fluidity [93]. Compared with the multitude of investigations on the effects of temperature, **the reactions of cell membranes to other environmental changes** have received little attention. In deep-sea fish, the desaturation of fatty acids in phosphatidycholine and phosphatidylethanolamine of the liver increases with depth under water, presumably to compensate for the increasing pressure [51]. It has been shown in marine crustaceans that dilution of the aqueous environment brings about an increase both in the contents of cardiolipin, phosphatidylethanolamine and phosphatidylserine in the posterior osmoregulation-active gills, and in the degree of desaturation of phospholipids. However, in the absence of investigations on isolated membranes or cell organelles, these results are difficult to interpret [37].

15.4.1 Phospholipids

Phospholipids are the main lipid components of cellular membranes and are also found in cells in soluble transport forms. Non-structurally bound phospholipids are found in the vertebrates in plasma lipoproteins, in the bile and in the lining of the inner wall of lung alveoli (lung surfactant). In crustaceans, such as the shore crab *Carcinus meanas*, the lobster *Homarus americanus* and the crayfish *Orconectes virilis*, the protein-bound transported lipid of the haemolymph is not diacylglycerol, as in the insects, but consists of two-thirds phospholipids, with more phosphatidylcholine (66–88 %) than phosphatidylethanolamine (12–20 %) [37]. The spermatozoa of the sea urchin *Hemicentrotus pulcherrimus* utilize the fatty acids of endogenous phosphatidylcholine as an energy source for swimming [176].

In **the classification of the phospholipids** a distinction is made, according to the nature of the backbone, between glycerophospholipids and sphingophospholipids (sphingomyelins) (Table 15.1). The glycerophospholipids are further subdivided on the basis of their head groups; the basic structure common to all, 1,2-diacyl-sn-glycero-3-phosphate (Fig. 15.20), is known as phosphatidic acid. The different glycerophospholipids are derived by substitution of the phosphate residue with choline, ethanolamine, L-serine, inositol or glycerol; in cardiolipin two phosphatidic acids are bound by glycerol (Fig. 15.20 g). **Phosphatidylcholine and phosphatidylethanolamine** are the predominant phospholipids of cellular membranes. In the case of these two compounds, the negative charge of the phosphate residue is compensated by the positive charge of the nitrogen under physiological conditions. The other phospholipids, however, generate a negative surface charge on the membrane, and this may be the reason why they are present in lower amounts. Phosphatidylcholine is usually more prevalent than phosphatidylethanolamine, e.g. in the vertebrates, crustaceans and molluscs (Tables 15.14 and 15.16). The situation in the insects is more confusing. Phosphatidylcholine predominates in the Hymenoptera, Lepidoptera and Orthoptera, and phosphatidylethanolamine in the aphids and Diptera. Finally, in the Coleoptera the ratio of the two phospholipid classes is approximately 1:1 [37, 55, 63, 89], and a similar situation pertains in the ciliate *Tetrahymena* (Table 15.15). The temperature dependence of the phospholipid spectrum must also be taken into account: in cold-adapted cells, the proportion of

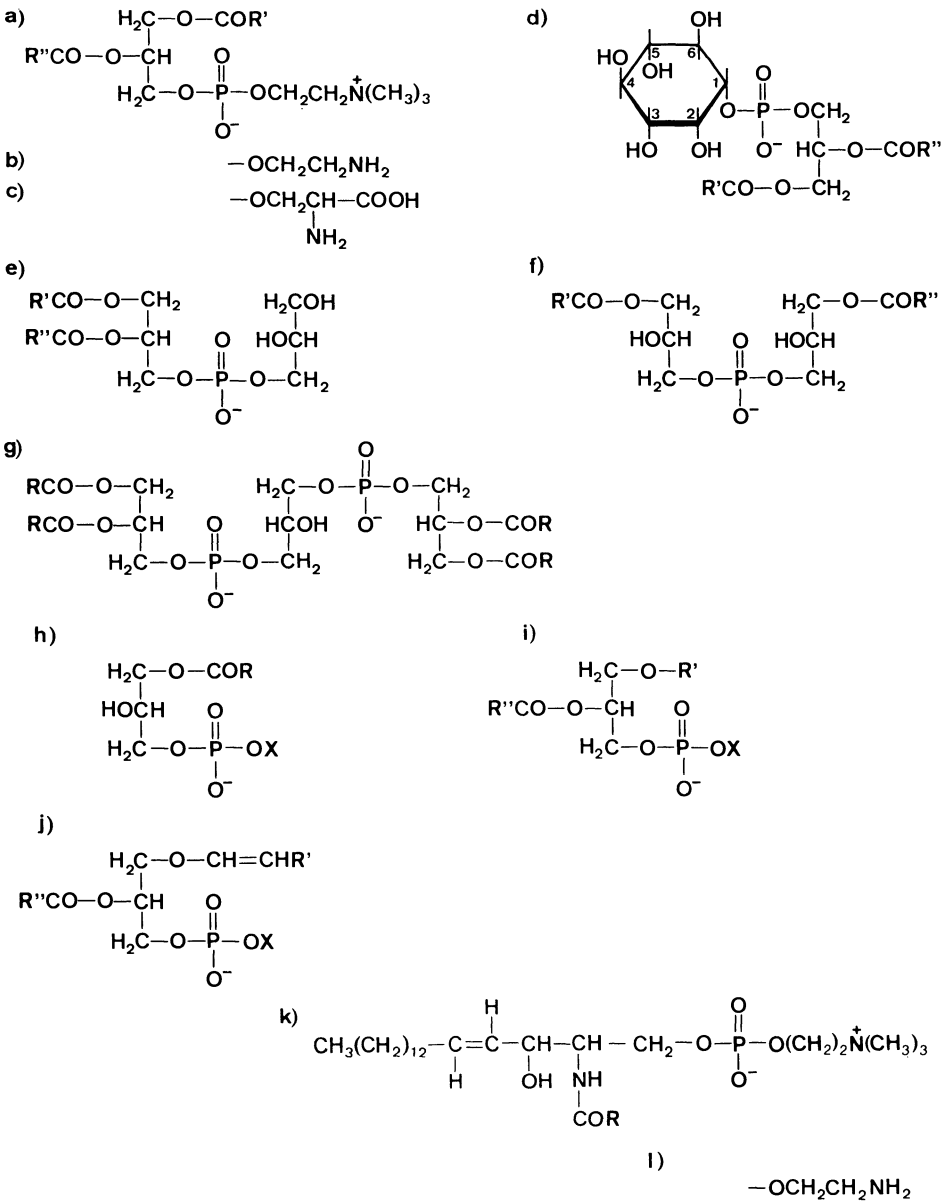


Fig. 15.20a–l. Phospholipids. **a** Phosphatidylcholine; **b** phosphatidylethanolamine; **c** phosphatidylserine; **d** phosphatidylinositol; **e** phosphatidylglycerol; **f** bis-(monoacylglycero)-phosphate; **g** bisphosphatidylglycerol (diphosphatidylglycerol) = cardiolipin; **h** lyso-glycerophospholipid; **i** 1-alkyl-2-acyl-glycerophospolipid; **j** 1-(1′-alkenyl)-2-acyl-glycerophospholipid = plasmalogen; **k** ceramide phosphocholine = spingomyelin; **l** ceramide phosphoethanolamine. In **h–j**, *X* represents choline or ethanolamine

phosphatidylcholine is reduced in favour of phosphatidylethanolamine.

In the phospholipids of the mammals and many other animals, saturated fatty acids are usually found in position sn-1 and unsaturated fatty acids occur in position sn-2. In contrast, the phospholipids from the brain and retina of the rainbow trout *Salmo gairdneri* have highly unsaturated fatty acids such as 22:6(n-3) in both positions; only phosphatidylinositol has 18:0 in position sn-1 and 20:4 or 20:5 in position sn-2, as in all other vertebrates [15]. In various Porifera, both positions in phosphatidylethanolamine are occupied by the same C_{26} to C_{30} polyene fatty acids [143]. The combination of different fatty acids results in a **multitude of molecular species** in each phospholipid class; the number of different phospholipid molecules in the erythrocyte membrane is estimated to be 150-200. This variety may be very important for the different physicochemical and

functional properties of cellular membranes. Phospholipids with saturated fatty acids result in dense packing with reduced fluidity and permeability, whereas those with polyunsaturated fatty acids are much less dense. It is, however, difficult to imagine that the exchange of fatty acid 20:4(n-6) with 22:5(n-3) or 22:6(n-3) would have a significant effect on membrane properties; thus, the high proportion of such fatty acids in the cell membranes of marine fish must be related to some other function [14].

In addition to phosphatidylcholine, a further choline-containing phospholipid, **sphingomyelin** or ceramide phosphocholine (Fig. 15.20k), plays an important role in some cell membranes. However, the proportion of this phospholipid in the brain and erythrocytes of mammals differs widely with the species: 10% in the red blood cells of the rat, dog and horse; 20–25% in the rabbit, pig, and humans; and as much as 45–50% in sheep, cattle and goats [89, 138]. Amongst the insects, the dipterans are particularly poor in choline-containing phospholipids; in these animals ceramide phosphoethanolamine is found instead of sphingomyelins [56]. The long-chain bases of the sphingolipids are derived from sphinganine, which is produced from palmitoyl-CoA and serine (Fig. 15.21a). Vertebrate sphingolipids mainly contain sphing-4-enine or sphingosine (Fig. 15.21b), whilst invertebrates have bases with different chain lengths, more double bonds and methyl branching (Fig. 15.21c).

Other phospholipids are found in significant amounts only in certain cell membranes. Serine phospholipids and inositol phospholipids amount to not more than 10%; however, the phosphatidylinositol polyphosphates are very important as precursors of the secondary messengers involved in the regulation of cellular processes [211] (p. 664). Bisphosphatidylglycerol or cardiolipin (Fig. 15.20g) has already been referred to as a typical component of the mitochondrial inner membrane. **Phosphatidylglycerol** (Fig. 15.20e) is the only phospholipid of the cyanobacteria, and makes up 20–30% of the total lipid phosphorus in higher plants. It is found at only one site in animals, viz. as the surface secretion lining the alveoli of vertebrates **(lung surfactant)**. This secretion is synthesized in type-II alveolar cells and stored in the form of layered granules (lamella bodies). After its extrusion, it forms a thin film on the cell surface; this reduces the surface tension by more than 50% and thereby facilitates lung ventilation [225]. The lung surfactant consists of about 90% lipid, 10% protein and some carbohydrate. In mammals, and also in the frog *Rana pipiens*, the lipid contains about 10–12% phosphatidylglycerol in addition to 70–75% phosphatidylcholine and its 1-alky-2-acyl- analogues [70]. Several of the protein components of the mammalian secretion have been characterized in detail. Surfactant protein A (SP-A) is a mixture of several lectin-like glycoproteins which are present as monomers of 28–36 kDa or as dimers of 60–65 kDa. The N-

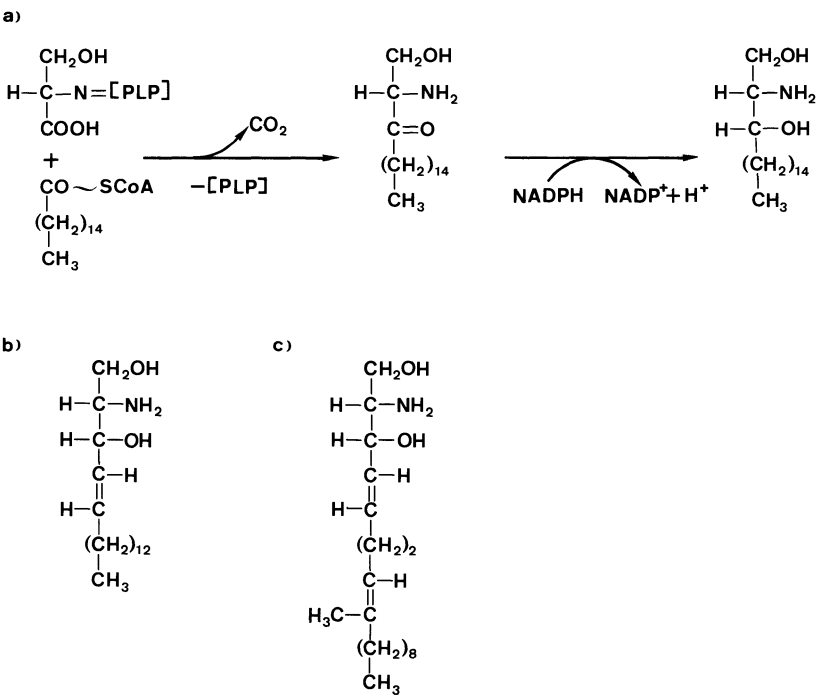


Fig. 15.21a–c. Long-chain bases in sphingolipids. **a** Biosynthesis of sphinganine. Pyridoxal phosphate-(PLP)-dependent condensation of L-serine with palmitoyl-CoA leads by decarboxylation to a 3-oxo compound which is reduced by an NADPH-dependent dehydrogenase to sphinganine (D-erythro-2-amino-1,3-octadecandiol); **b** Sphingosine (trans-4-sphingenine); **c** methylated octadecadiene base from the sea anemone *Metridium senile* [121]

terminal third of the sequence is made up, like collagen, of two dozen tripeptides with the structure Gly-X-Y, where hydroxyproline is common in position Y. An SP-A-like protein of 43 kDa was recently found in the rat and designated SP-D. The surfactant proteins SP-B and SP-C are extremely hydrophobic. Rat SP-B has 376 amino acids and agrees in 69 % of its sequence with the corresponding dog and human protein; SP-C has only 34–35 amino acids. SP-B and SP-C arise from larger precursors, whereby the precursor of SP-C has 197 amino acids and is an integral membrane protein with its N-terminus retained in the cytoplasm, i.e. it shows type-I orientation. The transmembrane segment corresponds to the mature SP-C [128, 274].

The bis-(monoacylglycero)phosphates are composed of two or three glycerol molecules linked by phosphoric acid residues and esterified with two to five fatty acids (Fig. 15.20 f). These rather unusual lipids are observed only in the alveolar macrophages of humans and other mammals and during a rare pathological lipid accumulation, the Niemann-Pick disease [89]. Equally uncommon in its distribution is N-acyl-phosphatidylethanolamine, which carries one of the usual C_{16} or C_{18} fatty acids bound in an amide-like fashion to the ethanolamine residue; this occurs in amounts of up to 1 % of the total lipid in the brains of pike and carp, but has otherwise been found only in bacteria and plant seeds [186].

Hydrolytic cleavage of the fatty acid bound at position sn-2 gives rise to **lysoglycerophospholipids**, whose name derives from their lytic activity on erythrocytes. They usually make up only a small proportion of the structural lipids (Tables 15.14–15.16) but reach high levels in the blood plasma of some mammals, e.g. 23 % in the rat [89]. Haemolytic effects have also been attributed to the **ether phospholipids**, which carry an alkyl or 1′-alkenyl residue instead of an acyl bound ether-like at position sn-1 (Fig. 15.20 i,j). Hydrolytic cleavage of the 1′-alkenyl residue and enol-keto conversion gives rise to a long-chain aldehyde. From the term "plasmal" to describe this aldehyde come the names **"plasmalogens"** for 1-(1′-alkenyl)-2-acyl-glycerophospholipids and "plasmal reaction" for their histochemical colour detection. Ether phospholipids are ubiquitous in the animal kingdom [63, 194, 265]. The most common are the plasmalogens that are produced by dehydrogenation of the corresponding 1-alkyl-lyso compound. In the elasmobranch and marine invertebrates, 1-alkyl-2-acyl-glycerophospholipids give rise to the common 1-alkyl-2,3-diacylglycerols, mentioned in Section 15.2.3 as storage lipids. The proportions of alkyl phospholipids and plasmalogens vary greatly with the species, organ and phospholipid class. In the mammals, the proportion of plasmalogens in the phospholipids varies from 1 % in the liver to 23–25 % in the brain and muscle; thus, in humans 19 % of the total body phospholipids fall into this category. From the elasmobranchs to the mammals, the plasmalogens make up between 46 and 68 % of the ethanolamine phospholipids in the brain [58, 89]. A large part of the phospholipids in many marine invertebrates (cnidarians, annelids, molluscs, arthropods, echinoderms and tunicates) consists of plasmalogens and 1-alkyl-2-acyl-glycerol compounds [38, 89]. Alkyl compounds also form a significant part of the membrane lipids of the ciliates *Tetrahymena* and *Paramecium* (Table 15.15).

15.4.2 Biosynthesis of the Phospholipids

The general **scheme of phospholipid biosynthesis** in vertebrates has been known for more than two decades (Fig. 15.22), although the nature of the enzymes and the reaction mechanisms have by no means been fully defined. In particular, it is not known how the proportions of the different types are regulated. This gap in our knowledge is due to the fact that the enzymes of lipid biosynthesis are bound tightly to the microsomes or mitochondrial membranes and are difficult to approach experimentally [89]. For this reason, comparative biochemical investigations of lipid biosynthesis are restricted almost entirely to tracer studies, which, as it happens, are in general agreement with the current scheme for the vertebrates. Both mitochondria and microsomes are capable of lipid biosynthesis but differ in the enzymes involved. Thus, the methylation of phosphatidylethanolamine to phosphatidylcholine occurs only in the endoplasmic reticulum, and cardiolipin synthesis and the decarboxylation of phosphatidylserine to phosphatidylethanolamine occur only in the mitochondria. Both cell compartments contain different isoenzymes of glycerophosphate acyltransferase (Fig. 15.22, reaction 1) and acyl-CoA synthetase. We can only speculate about why lipid biosynthesis is distributed between two cell compartments in such a complicated fashion. Prokaryotes produce mainly phosphatidylethanolamine, phosphatidylglycerol and cardiolipin; hence, phosphatidylethanolamine and cardiolipin are synthesized in the mitochondria, which,

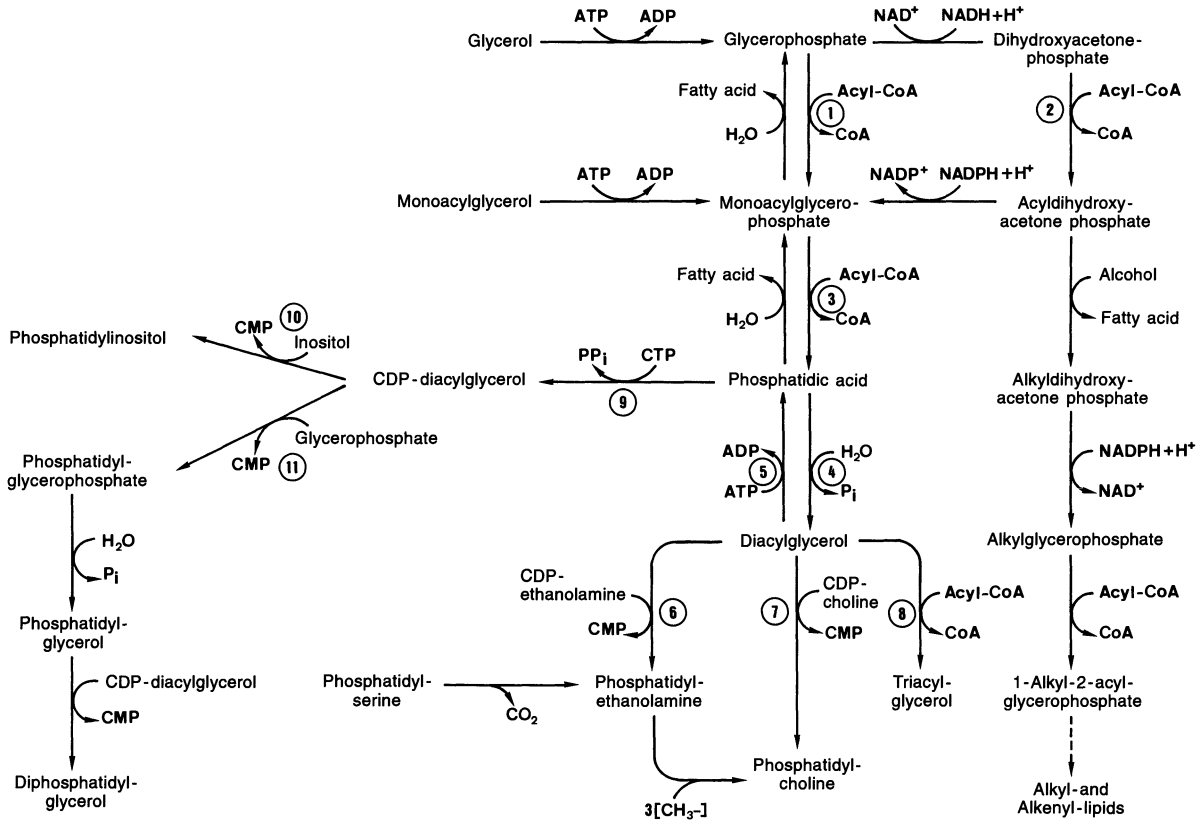


Fig. 15.22. Biosynthetic pathways of the phospholipids. *1*, Glycerophosphate acyltransferase; *2*, dihydroxyacetone phosphate acyltransferase; *3*, monoacylglycerophosphate acyltransferase; *4*, phosphatid phosphohydrolase; *5*, diacyglycerol kinase; *6*, ethanolamine phosphotransferase; *7*, choline phosphotransferase; *8*, diacyglycerol acyltransferase; *9*, phosphatidate cytidyltransferase; *10*, CDP diacylglycerol:inositol-3-phosphatidyl transferase; *11*, glycerophosphate phosphatidyl transferase

according to the symbiont theory, are derived from symbiontic prokaryotes. Most bacteria cannot produce phosphatidylcholine, phosphatidylinositol, sphingolipids or sterols and that is why the endoplasmic reticulum is responsible for these processes in the eukaryotes. However, this separation has partly disappeared during the evolution of the eukaryotes and the enzymes phosphatidylinositol synthase and diacylglycerol cholinetransferase are present in both compartments [55]. Despite its complicated nature, the scheme depicted in Fig. 15.22 does not give all the possible synthesis pathways for phospholipids. For example, base exchange between different phospholipids, which forms the main pathway for the synthesis of phosphatidylserine in animals, is not shown. The production of phosphatidylcholine and phosphatidylethanolamine can also take place by acylation of the appropriate lyso- compound; this is a pathway that plays a special role in linoleic and arachidonic acid metabolism. The synthesis of phosphatidic acid, one of the most important intermediates of lipid biosynthesis, can occur by various routes. Phosphorylation of diacylglycerol to phosphatidic acid (Fig. 15.22, reaction 5) is important in, for example, the brain. The biosynthesis of phosphatidic acid can also begin with dihydroxyacetone phosphate (the DHAP pathway). This route is obligatory for the production of the so-called ether lipids, but it is not clear what role it plays in the formation of diacyl- and triacylglycerol lipids. DHAP acyltransferase (Fig. 15.22, reaction 2) is localized in the peroxisomes of rat liver and probably of other organs [89].

The biosynthesis of phosphatidylethanolamine and phosphatidylcholine by transfer of CDP-bound bases to diacylglycerol has also been demonstrated in various **insects**. The corresponding kinases have been isolated from the dipterans *Phormis regina* and *Culex pipiens*. Thus, both here and in the mammals, there is apparently a choline-specific enzyme in addition to the one that metabolizes both ethanolamine and choline

[63]. Choline is essential for the dipterans but it can be replaced in the diet by various analogues which are also incorporated into phospholipids. It is still not clear whether the three methylation reactions from phosphatidylethanolamine to phosphatidylcholine are catalysed in the mammals by one or two enzymes. In lower fungi, such as yeast and *Neurospora*, there is genetic evidence for the existence of separate enzymes for the first and the two following methylation steps. A protein complex of 140–150 kDa has been isolated from *Drosophila melanogaster* and this has all three methylation activities [249]. Tracer experiments in the crustaceans, whose tissues are especially rich in phospholipids, have confirmed the presence of the vertebrate scheme of phospholipid biosynthesis (Fig. 15.22). Phosphatidylethanolamine and phospholipid arise from diacylglycerol (DAG) and a CDP base, and phosphatidylinositol and cardiolipin arise via CDP-diacylglycerol; phosphatidylserine can be decarboxylated to phosphatidylethanolamine [37]. The phospholipid metabolism of the **Porifera** coincides with that of the higher animals. In *Microciona prolifera* phosphatidylethanolamine and phosphatidylcholine are mainly synthesized from CDP bases (Fig. 15.22, reactions 6 and 7); the methylation of phosphatidylethanolamine plays only a minor role [143]. The reactions of the vertebrate scheme have also been demonstrated in some **Protozoa**, e.g. the formation of phosphatidylethanolamine and phosphatidylcholine with DAG and CDP bases in *Tetrahymena*, the production of phosphatidylinositol from myo-inositol and CDP-diacylglycerol in *Crithidia fasciculata*, and the pathway from phosphatidylserine via phosphatidylethanolamine to phosphatidylcholine in the malarial parasite *Plasmodium knowlesi*.

15.4.3 Phosphonolipids

In 1959, Horiguchi discovered a novel lipid component in rumen ciliates, **2-aminoethylphosphonic acid** (AEPn or ciliatin), the first known substance with a phosphorus-hydrocarbon bond (Fig. 15.23). Further such bonds have since been found in lipids, e.g. 2-amino-3-phosphonopropionic acid (phosphonoserine, Fig. 15.24) and N-methylated AEPn. In addition to free and lipid-bound forms, AEPn also occurs as a substituent of glycoproteins (p. 485). Lipid-bound AEPn is known only from the molluscs and cnidarians, and amongst the protozoans from ciliates, amoebae and flagellates [172, 182, 185]. There are no AEPn lipids in *Locusta migratoria*, but the haemolymph contains free AEPn at a high concentration; this is not derived from the symbionts [129]. Glycerolipids and sphingolipids may contain AEPn (Fig. 15.23). In the cilia membrane and the outer pellicular layer of *Tetrahymena*, AEPn lipids make up about half of the

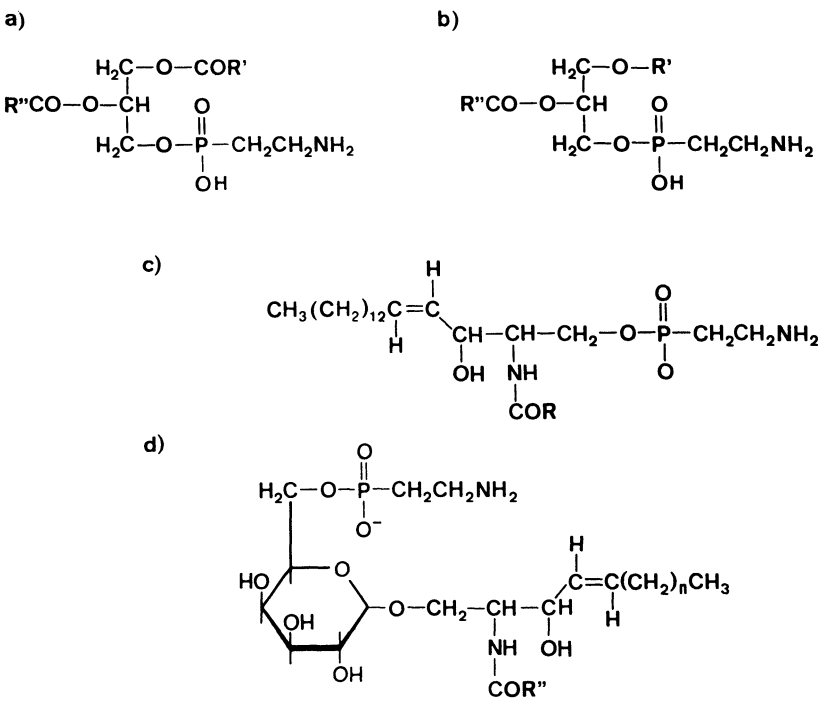


Fig 15.23.a–d. Phosphonolipds [89]. **a** 1,2-diacylglycerophosphonolipid; **b** 1-akyl-2-acylglycerophosphonolipid; **c** sphingophosphonolipid; **d** sphingophosphonoglycolipid

total phospholipids, and in other membranes they constitute about one-quarter (Table 15.15). The most frequently found are glycero-AEPn lipids (Fig. 15.23a,b), in particular an ether lipid with hexadecyl alcohol in position sn-1. Position sn-2 carries fatty acids of the type 18:2 or 18:3 in *Tetrahymena*, but four- to fivefold an amount of polyunsaturated fatty acids in *Paramecium*. Under certain culture conditions, the glycero-AEPn lipids of *T. pyriformis* strain WH-14 contain branched 2-hydroxy fatty acids [89]. Like phosphatidylcholine and phosphatidylethanolamine, the AEPn lipids are also involved in membrane adaptation to temperature.

The first **AEPn-sphingolipid** (Fig. 15.23 c) was discovered in 1963 in the sea anemone *Antopleura elegantissima*. Since then, such lipids have been found in large quantities in various cnidarians and molluscs; for example, they make up 45 % of the total sphingolipids in the shell-adductor muscle of the oyster. In contrast, the ciliate *Tetrahymena* contains only small amounts of AEPn-sphingolipids. In the molluscs, up to one-third of these lipids are N-methylated, and in the sphingophosphonolipids of the sea anemone *Metridium senile* about one-third of the fatty acids are branched. In addition to the special case of *T. pyriformis* mentioned above, this is one of the very few examples of branched fatty acids in animal membrane lipids. Two relatively simple sphingophosphonoglycolipids with a terminal AEPn have been isolated from the gut of the sea snail *Turbo cornutus* (Fig. 15.23 d), and several more complicated glycolipids with one to three terminal AEPn have been isolated from the nervous system of the snail *Aplysia kurodai*, together with a pyruvylated galactose residue found here for the first time as a component of animal sphingolipids [4, 90]. More recently, sphingophosphonoglycolipids have been detected in the crustaceans: namely, AEPn-glucosylceramide has been found in the Antarctic krill *Euphausia superba* [106]. The sphingosines of the phosphonolipids in *Tetrahymena* are mainly iso- branched with 16–19 C atoms and one double bond. In mollusc species, they are complex mixtures of predominantly unbranched bases with 16–22 C atoms and one or two double bonds [89]. **AEPn-glycoproteins** were discovered in the slime of sea anemones such as *Metridium senile* and *M. dianthus* (p. 485). A glycolipoprotein with a 22.5-kDa polypeptide, 30 % carbohydrate and 8 % AEPn has been reported in the cell membrane of *Tetrahymena* [245].

It is essentially a mystery why phosphonocompounds are formed and incorporated into

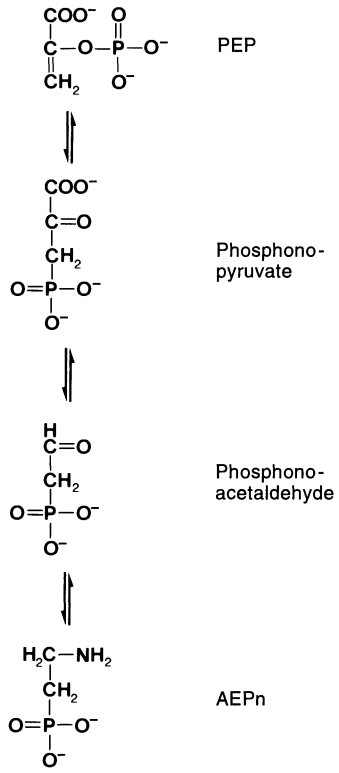


Fig. 15.24. The AEPn biosynthetic pathway of *Tetrahymena pyriformis* [24]

various lipids and proteins exclusively in the molluscs, cnidarians and some protozoans. An interesting explanation was recently put forward for the presence of AEPn compounds in the outer pellicular layer of *Tetrahymena*. The AEPn group inhibits phospholipase C and the 1-alkyl group inhibits phospholipase A_1. Both phospholipases are, however, secreted by *Tetrahymena*, presumably as a defence against amoebae and other enemies, or to lyse the cells of aquatic insects in which *Tetrahymena* is an endoparasite. The phosphonolipids protect *Tetrahymena* from its own lipases. It is not known, however, whether cnidarians and molluscs also secrete lipases into their environments [67]. About 98 % of the total phosphorus in freshly laid eggs of the aquatic pulmonate snail *Helisoma* sp. is present as AEPn, which is apparently bound to galactans. The AEPn is largely consumed during embryo development and represents only 29 % of the total phosphorus in the adult. Thus, AEPn stored in the eggs of this and other water snails (Planorbidae, Physidae, Lymnaediae) apparently serves as a phosphate store for use during development [172].

Analogous to the synthesis of the glycerophospholipids, the glycerophosphonolipids in the cili-

ate can be derived from free AEPn by AEPn transfer from CMP-AEPn to diacylglycerol; however, this appears to be simply a mechanism for the recycling of free AEPn (salvage pathway). **AEPn biosynthesis** starts with phosphoenolpyruvate (PEP). The main route involves the conversion of phosphonopyruvate by a PEP phosphomutase, decarboxylation to phosphonoacetaldehyde and direct amination to 2-aminoethyl phosphonate (Fig. 15.24). The phosphomutase of *T. pyriformis* has been isolated [9, 24].

15.4.4 Glycolipids

Three **main types of glycolipid** can be distinguished on the basis of structure and function: the glycosphingolipids, including the sialic acid-containing gangliosides and the glycoglycerolipids, are important membrane components; the glycosylphosphopolyprenols (dolichols) are cofactors for the N-glycosylation of proteins; and sulphated glycoglycerolipids are important components (in addition to glycoproteins) of the mucosal lining of the stomach wall, protecting it from mechanical and chemical injury. The characteristic ascaroside from the ovary and eggs of the roundworm *Ascaris* is built up from the sugar ascarylose, a long-chain, monovalent or divalent fatty alcohol, and sometimes also a fatty acid; this has already been discussed (p. 463).

The **glycosphingolipids** (GSL; Fig. 15.25a) are mainly located in the outer layer of the cell membrane. The ceramide portion, consisting of a long-chain base and an amide-like bound fatty acid, is integrated in the lipid layer, whereas the carbohydrate part protrudes over the cell surface.

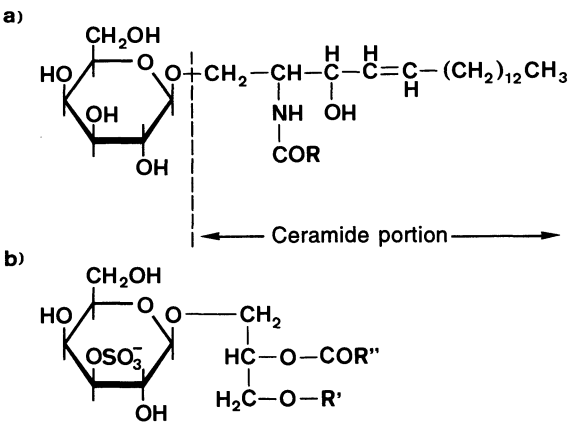


Fig. 15.25a, b. Glycolipids [281]. **a** Galactosylceramide as an example of a glycosphingolipid; **b** 1-akyl-2-acyl-3-(3′-sulphogalactosyl)-sn-glycerol (seminolipid) as an example of a glycoglycerolipid

Experiments with reagents that act from without have shown, however, that a considerable part of the GSL molecule is inaccessible ("masked"). GSLs also contribute to the mechanical strength of the cell membrane, but their main functions lie in the area of cell recognition, cell-cell interaction, as immunological antigen determinants, differentiation markers, and receptors for hormones and, inevitably, toxins and viruses [281]. The surface of the epimastigotes of *Trypanosoma cruzi* includes a lipopeptidophosphoglycan (LPPG) that is important for host-cell penetration and has immunological properties. The glycan is bound to inositol phosphorylceramide via a non-acylated glucosamine residue. The latter carries a molecule of AEPn [210]. The LPPG of *T. cruzi* is structurally and functionally comparable with the glycosylphosphatidylinositols, described below, which serve as anchors for membrane proteins (p. 610). This is also true for a glycophosphosphingolipid found recently in *Tritrichomonas foetus* [242].

The **classification of GSLs** is based upon the structure of the carbohydrate portion. The number of monosaccharide residues in the molecule is described by the suffixes, -biose, -triaose, -tetraose, etc.; the more correct -triose is deliberately modified to avoid confusion with the name of the C_3 sugars. Most GSLs can be assigned to one of six series according to the sequence of the sugars (Fig. 15.26); the mollu- and arthro- series are found only in the corresponding invertebrate groups. Binding variants are indicated by the addition of "iso-" or "neo-". Not all GSLs can be so easily classified. For example, a fucose-containing GSL isolated from the teleost *Parophrys vetulus* does not fit into any of the six series [199]. GSLs acquire a negative charge through the linking of the oligosaccharide chains with acidic components; in sulphated GSLs, referred to earlier as sulphatides, the acidic components are sulphuric acid residues esterified with sugar OH groups, and in the gangliosides they are sialic acids. In insects there are GSLs that have terminal glucoronic acid residues as the negatively charged component [279]. Sulphated GSLs are found in all cell membranes where they constitute specific binding sites for adhesive glycoproteins such as laminin, thrombospondin and the von Willebrand factor [224]. Individual cells may contain numerous different glycolipids; for example, it is estimated that human erythrocytes contain about 25 neutral GSLs and just as many gangliosides, only half of which have, as yet, been described. Probably the most complicated GSLs

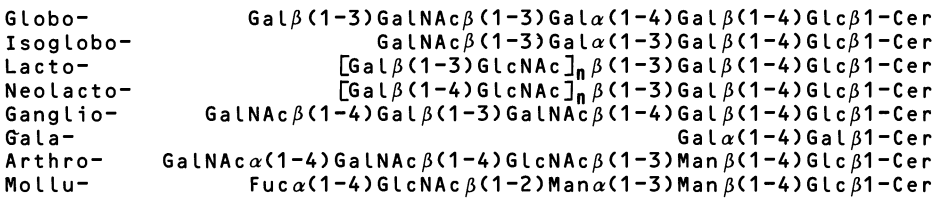


Fig. 15.26. The most important series of the glycosphingolipids [281]

are the polyglycosylceramides or macroglycolipids which can contain up to 20 sugar residues; they have so far been reported only in the mammals, in particular in the erythrocytes [281].

The mammalian **ceramide components** include the very common sphingosine (4-sphingenine), the saturated sphinganine (Fig. 15.21a,b), 4-hydroxysphinganine and the longer-chain eicosasphingenine. Further GSL components have been discovered in the invertrebrates, e.g. unusually short-chain (C_{14}, C_{15}, C_{16}) sphinganines and sphingenines in glucosylceramides of the nervous system of the prawn *Penaeus aztecus*, and a methylated octadecadiene base (Fig. 15.21 c) in glucosylceramides of the sea anemone *Metridium senile* [121, 240]. The fatty acids of GSLs mostly have chains of more than 20 C atoms and are less unsaturated than those of the phospholipids. The GSLs of the brain, kidney and small intestine contain large amounts of 2-hydroxy fatty acids [281].

The simplest GSLs of the mammals are **galactosyl- and glucosylceramides** (Fig. 15.25a). Galactosylceramide was discovered in 1884 in the human brain as the first GSL, and is also found in various organs of lower vertebrates. Glusosylceramide was discovered first in the spleen of patients with Gaucher disease but, in fact, next to lactosylceramide and globotetraosylceramide (globoside), it is one of the main GSLs of the extraneuronal tissues of healthy humans and mammals. Galactosyl- and glucosylceramides are also widely found in the invertebrates where, in contrast to the mammals, there are also other monoglycosylceramides, e.g. mannosylceramide in the freshwater mussel *Hyriopsis schlegelii* [105]. The main components of vertebrate **myelin** are galactose-containing GSLs which are partly sulphated [179]. In mammals, the myelin of the central nervous system contains proteolipid protein (PLP) and myelin basic protein (MBP) as protein components, whilst the myelin of the peripheral nervous system has protein P_o and MBP. The P_o and MBP of the shark *Heterodontus francisci* and mammals have about 44–46% sequence similarity [230]. Invertebrates have no myelin but do have a loose membrane surrounding the axons of the nerve cells. Systematic investigations of 23 species showed that the nervous tissues of the echinoderms, acranians and vertebrates have mainly galactose-containing GSLs, which have hydroxy fatty acids and are in part sulphated; in contrast, the nervous tissues of the molluscs, annelids and arthropods (Xiphosura, Crustacea, Insecta) have only glucose-containing GSLs without hydroxy fatty acids or sulphate residues [197].

The L-fucose-containing GSLs **(fucolipids)** carry the antigenic characters of the human ABH (formerly ABO) and Lewis blood group systems. At least 36 different fucolipids have so far been recognized in mammalian cells, with especially large amounts occurring in the stomach wall, gut and pancreas, but little in the erythrocytes. In order to define the structure of the ABH antigenic carhohydrates, glycoproteins and oligosaccharides from the saliva and other secretions were initially investigated. In the meantime, however, the ABH antigens of the erythrocytes have been identified as GSLs. These are structurally very heterogeneous, with the terminal structure of the carbohydrate chain being the determinant of antigenicity [281]. Fucose- and methylfucose-containing GSLs have also been found in mussels and snails [281]. Completely novel GSL structures have emerged from the less-detailed investigations of **invertebrates**, i.e. the bivalves, gastropods and insects, and these led to the formulation of the mollu- and arthro- GSL series (Fig. 15.26). The GSL spectra of the invertebrates appear to be just as complex as those of the vertebrates [3, 53, 91, 105, 281]. Systematic investigations of further invertebrates will presumably reveal other GSL series. Almost nothing is known about GSL metabolism in the invertebrates. Most noteworthy is the detection of a ceramide glycanase in the earthworm *Lumbricus terrestris* and the leech *Macrobdella decora;* this can cleave the bonds between carbohydrates and ceramides in various GSLs [151, 291].

In agreement with the characteristic distribution of sialic acids in the animal kingdom (p. 465),

sialic acid-containing GSLs **(gangliosides)** have, as yet, been found only in the vertebrates and the echinoderms. The structural variety of the gangliosides is almost impossible to summarize. To the variability of the carbohydrate portion, the long-chain base and the fatty acid of all GSLs, may be added the variable nature, number and position of the sialic acids. The terminology is just as complicated. The system of abbreviations put forward by Svennerholm (1963), and which is still in frequent use, designates the individual gangliosides with a G together with a capital letter subscript, indicating the number of sialic acids (G_M for monosialo- to G_H for hexasialo-), and Arabic numbers (and lower-case letters) for the relative mobility in thin-layer chromatograms (Fig. 15.27) [281].

Gangliosides may contain up to six sialic acid residues which are almost always found at the ends of the carbohydrate chains, or form side-branches. Gangliosides with internal sialic acids

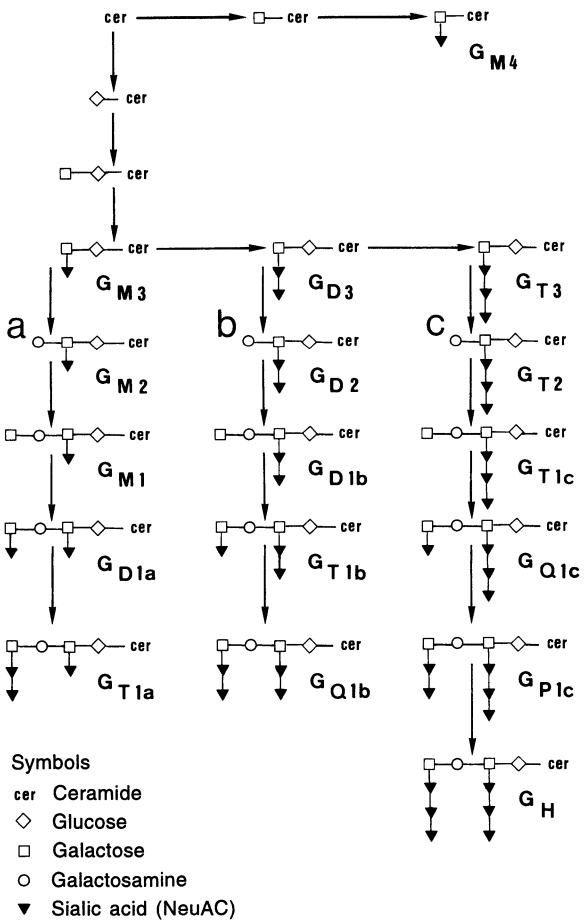


Fig. 15.27. The three biosynthetic pathways of the gangliosides [213]. The individual gangliosides are labelled according to the abbreviations of Svennerholm (1963)

are known only in the echinoderms. The relative proportions of N-acetyl and N-glycolyl neuraminic acid can vary in the same cell type (erythrocytes) between different species. Sugar- and O-methyl-substituted sialic acids are present in the gangliosides of the echinoderms. The **carhohydrate** protein of the gangliosides mostly belongs to the ganglio-, lacto- or globo- series (Fig. 15.26); however, gangliosides from cows' milk and echinoderms were found to contain a novel branched oligosaccharide [244, 259]. The simplest gangliosides with one galactose or glucose residue and a terminal sialic acid are present in, for example, sea urchin gametes, and the galactose compound is also a major component of myelin. Gangliosides with lactose and one to three sialic acids (haematosides) predominate in extraneuronal tissues. The more complicated gangliosides can be subdivided according to three **synthetic pathways**, *a–c*, which all start with the lactose compound having one, two or three terminal sialic acids (Fig. 15.27). A trisialoganglioside isolated from the brain of the dogfish *Squalus acanthias* does not, however, fit this scheme: its three sialic acid residues are separately linked to the gangliotetraosyl backbone [184]. The ceramide portion of vertebrate gangliosides differs from the neutral GSLs by having a higher content of eicosasphing-4-enine. Amongst the gangliosides of the **echinoderms**, those of the sea urchins always have the same simple carbohydrate component 6-sialylglucose. In contrast, the starfish have species-specific gangliosides which differ in the type and localization of the sialic acid residues and contain a complicated spectrum of partly branched C_{16}, C_{17} and C_{18} bases; up to 90 % of the fatty acids are hydroxyl substituted [244, 281].

Species comparisons are only available for the gangliosides of mammalian erythrocytes and vertebrate brain. The oligosaccharides of **mammalian erythrocyte gangliosides** vary with the species, with the ganglio- series present in the rat, the lacto- series in cattle, the globo- series in humans, lactose in other species, or just galactose in the mouse. The lactose-containing gangliosides that predominate in many extraneuronal cells carry one to three sialic acids. Those with one sialic acid are found in, for example, the erythrocytes of humans, cattle, the horse, giant panda, rabbit and dog; those with two occur in the erythrocytes of the cat and other Felidae, and also in the retina of many mammals. The relative proportions of N-acetyl- and N-glycolyl-neuraminic acid in erythrocytes also differ with the species: only NeuAc in man; predominantly NeuAc in Euro-

pean dogs but NeuGc in Oriental dogs; exclusively NeuGc in the horse, giant panda and domestic cat; and both NeuAc and NeuGc in Persian cats [281].

The gangliosides of the brain show the highest concentrations and complexity; in the synaptosome membranes they constitute up to 15 % of the total lipids. Critical to their function is the fact that they are the only lipids to change in properties after binding calcium ions, and they are therefore probably modulators of synaptic signal transmission. Intensive comparative investigations of representatives of all the vertebrates groups from the elasmobranchs to the mammals have shown the following [7, 95, 213, 214, 215]:

1. The brain gangliosides of the vertebrates belong almost exclusively to the ganglioseries, whereby lactose and gangliotriaose predominate in the elasmobranchs, and ganglio-tetraose predominates in all other vertebrates up to the mammals.
2. Despite large species-specific differences, the **ganglioside concentrations in the brain** clearly increase in the vertebrate order, with values of 0.5–1.0 nmol ganglioside-bound sialic acid per g fresh weight in the fish and amphibians, more than 1.3–1.7 nmol/g in the reptiles, and up to 2–3 nmol/g in the birds and mammals. It is noteworthy that all elasmobranchs have relatively high values, and highly developed forms, such as *Myliobatis australis*, show higher values (1.88 nmol/g) than do primitive species such as *Squalus acanthias* (0.87 nmol/g).
3. The **spectrum of brain gangliosides** differs between the different sections of the brain. However, there is a clear trend towards simplicity with increasing brain organization. In the lower poikilothermic vertebrates there is a large number of complex gangliosides that have many sialic groups; in contrast, in the birds and mammals there is a lower number of less strongly polar gangliosides. Similarly, in the mammals the ganglioside spectrum decreases in complexity from the monotremes to the marsupials to the Placentalia [213, 214]. Simplification of the spectrum with increasing degree of organization is also observed for other metabolites, e.g. the sterols. The phenomenon can be explained by the assumption that nature began with experiments involving several different substances, and with time selected those that were most suitable. The agnathan *Lampetra fluviatilis* has one-third each of di-, tri- and tetrasialogangliosides, 80 % of which are produced by the *b* pathway (Fig. 15.27). In the elasmobranchs, pentasialo- and hexasialo- compounds constitute 18–25 % of the gangliosides. The brain gangliosides of the many teleosts to have been investigated differ greatly in polarity. This is probably related to the average temperature of their habitats. Strongly polar gangliosides at low temperatures have the same calcium-binding capacity as do less polar forms at higher temperatures. Thus, the tropical tapir fish *Gnathonemus petersi* has 16 % monosialoganglioside and 45 % disialoganglioside, whereas the Antarctic ice fish *Trematomus hansoni* has the most polar ganglioside spectrum of all vertebrates with 40 % pentasialo- and hexasialo-compounds. In the mammals, on the other hand, 68–84 % of the gangliosides have only one or two sialic acids, 8–22 % have three sialic acids, and only 4–12 % have four or more sialic acids. The proportion of gangliosides with a high number of sialic acids increases during hibernation. The relationship between calcium binding, polarity and temperature explains why changes in the spectrum of brain gangliosides is among the mechanisms of adaptation to temperature found in the vertebrates [213, 214].
4. Alteration of the ganglioside spectrum is accompanied by a **change in the synthetic pathway** (Fig. 15.27). The *b* pathway predominates in the agnathans; the Chondrichthyes and Osteichthyeses have in addition the *c* pathway, which is clearly the dominant pathway in the higher teleosts, such as the pike *Esox lucius*. Both the *b* and *c* pathways are also found in the amphibians, whereas reptiles and birds use the *a* and *b* pathways. Finally, the mammals have only the *a* pathway [213, 214]. During ontogenesis in the chicken, there is a "recapitulation of the phylogeny" in that the embryo makes use of the more primitive combination of the *b* and *c* pathways, whereas the adult utilizes *a* and *b*.

Glycoglycerolipids have been found only so far in vertebrate sperm and myelin. The main representative of this class in mammalian sperm is the 1-alkyl-2-acyl compound known as seminolipid (Fig. 15.25b). This is also found in the rat brain, together with the corresponding diacyl compound; in contrast, the brain of the cod *Gadus morrhua* contains only the diacylated form [281]. Glycoglycerolipids with a rather peculiar structure have been isolated from the cockroach

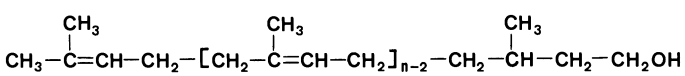
Blaberus colosseus. These "blaberosides" contain 6′-3-hydroxy-11-eicosenoyl-gentiobiose as the carbohydrate component and are unusual in that this carbohydrate is located in position sn-2 of a 1-alkyl-3-acyl-glycerol [256]. 1-Alkyl-2-acyl-3-(2′,3′-diacylglycerol)-glycerol, discovered in Harderian gland tumours of mice, may be considered the representative of a new class of lipids [123]. The **glycosyl phosphatidylinositols** (GPIs) of the plasma membrane form the anchor for the different proteins of the cell surface, e.g. thy-1 or acetylcholinesterase. The membrane-bound GPIs of unicellular eukaryotes have similar functions; the GPIs found in the cytoplasm are precursors of the membrane components. The attachment of the variable surface glycoproteins (VSGs) of *Trypanosoma brucei* has already been described (p. 240). The parasitic protozoans of the genus *Leishmania* inhabit their sandfly (Phlebotominae) vector as extracellular promastigotes, and mutiply in the mammalian host as intracellular amastigotes in the phagolysosomal compartment of the macrophages. The GPIs of the promastigotes are membrane anchors for the surface proteins and the heterogeneous lipophosphoglycan, which plays a central role during the infection of mammalian macrophages and during survival in the phagolysosomes. The promastigote surface protease (PSP) of *L. major* is a zinc endopeptidase. The glycan of its GPI anchor is very similar to the VSG of *T. brucei* and to the thy-1 of the rat brain. The lipid component, in contrast to that of VSG and thy-1, is a completely saturated 1-alkyl-2-acyl-phosphatidylinositol [233]. The lipophosphoglycans (LPGs) of *L. major* are made up of repetitive phosphoglycans of 5–40 kDa, a phosphorylated hexasaccharide core and a lysoalkylphosphatidylinositol anchor. The LPGs of the related species *L. donovani* are similar, but less complex in structure [166, 167]. The **glycosylphosphopolyprenols** are the only glycolipids in which the carbohydrate and lipid is linked by a phosphoric acid; they function in the N-glycosylation of animal proteins (see Fig. 13.17, p. 491). The lipid component is provided by the dolichols (Fig. 15.28), which are polyprenols with 16–23 prenyl residues. The number of prenyl residues is species- and organ specific: human liver contains 20 > 21 > 19 in decreasing amounts; rat liver contains 18 > 19 > 17; and rat testis contains 18 > 20 > 19. Dolichols with 19 and 20 prenyl residues predominate in marine invertebrates [281]. Dolichol phosphorylation in vertebrates, crustaceans, nematodes and sea urchins requires CTP, whereas ATP is apparently the phosphoryl donor in the insects [99].

15.5 Lipids with Regulatory Functions

It has become increasingly obvious during recent years that lipids not only form important energy reserves and structure-forming membrane components but also act in very specific ways in the regulation of many cellular processes. Two such mechanisms have already been discussed, i.e. the effects of the phospholipid-dependent protein kinases (Chapter 3) and the receptor-stimulated hydrolysis of phosphatidylinositol-4,5-bisphosphate into the two secondary messengers inositol-1,4-5-trisphosphate and diacylglycerol. The biologically highly active lipids also include the platelet-activating factor (PAF), which is produced by leukocytes and other cells and, even at the extremely low concentration of 10^{-11} mol/l, stimulates platelets to release vasoactive amines; it also triggers other cellular reactions. This is the ether lipid 1-alkyl-2-acetyl-sn-glycerol-3-phosphocholine, which carries an aliphatic alcohol with 14–18 C atoms at sn-1 and an acetic acid residue at sn-2 [84]. The following section examines in some detail two types of bioactive lipids which are particularly interesting for comparative biochemistry: the prostaglandins, and the juvenile hormones. The steroids, of course, also belong to the class of lipids with regulatory functions, but their variety and unique structural features will be dealt with in Chapter 16.

15.5.1 Prostaglandins and Other Eicosanoids

The oxidation of polyunsaturated C_{20} fatty acids, in particular arachidonic acid 20:4(n-6), gives rise to a variety of products, the so-called **eicosanoids**. The best-known of the eicosanoids are the **prostaglandins**, which were first discovered in



Dolichol

Fig. 15.28. Dolichol. *n*, Number of prenyl residues

mammals in 1933, and were named after their origin in the prostate. However, their structure and biosynthesis have only been described during the last two decades. They have been assigned a confusing number of **effects** in mammals, in some cases with different prostaglandins acting on the same cell or with the same prostaglandin acting on different tissues and having very different (even opposing) effects. In general, the prostaglandins are probably best described as a modulation system with mainly homeostatic functions; they are involved, for example, in reproduction, blood clotting, the circulation, kidney function, pain and stress reactions, protection of the gut from self-digestion, and regulation of immunological responses [188, 201]. Several classes of specific prostaglandin receptors have been described [183]. Prostaglandins and other eicosanoids are also known from various invertebrates, where they apparently have regulatory functions in reproduction, ion balance, temperature control and parasite-host relationships [253].

The rate-limiting step in the **biosynthetic pathway** of the prostaglandins is the release of arachidonic acid, the nervous and hormonal regulation of which has therefore been studied to some depth in man and other mammals. The arachidonic acid at position sn-2 of phosphatidylcholine and phosphatidylethanolamine is released by phospholipase A_2. However, in many tissues the arachidonic acid is bound mainly to phosphatidylinositol and, after the removal of inositol phosphate by phospholipase C, is released from the resulting diacylglycerol (DAG) by a DAG lipase. During prostaglandin synthesis from arachidonic acid, a cyclooxygenase reaction gives rise to 15-hydroperoxy-9,11-endoperoxide (PGG_2) which is reduced in a subsequent peroxidase reaction to the corresponding 15-hydroxy compound, PGH_2 (Fig. 15.30a). Both reactions are catalysed by the same protein, known as prostaglandin synthase [169, 285]. The classical prostaglandins (PG) E, D and F arise by enzymatic or non-enzymatic reactions, in which the endoperoxide is either isomerized to β-hydroxyketones (PGE and PGD) or reduced to a 1,3-diol (PGF) (Fig. 15.29b). The NAD^+-linked 15-hydroxyprostaglandin dehydrogenase of the human placenta is a dimer of 266-amino-acid subunits which have a surprising 54 % sequence similarity to the short alcohol dehydrogenase of *Drosophila melanogaster* [141]. The basis of the prostaglandins is prostanic acid (Fig. 15.29a), which arises from a cyclopentane ring with two side-chains of seven and eight C atoms. In **the classification and nomenclature of the prostaglandins**, the structure and substituents of the cyclopentane ring are designated by the letters A to I, and the number of double bonds in the side-chain by the numbers 1 to 3 (Fig. 15.29c). Prostaglandins of series 1 are derived from 13-trans-prostenic acid, of series 2 from 5-cis,13-trans-prostadienic acid, and of series 3 from 5-cis,13-trans,17-cis-prostatrienic acid. The common prostaglandins $PGF_{1\alpha}$ and $PGF_{2\alpha}$ (Fig. 15.29c) carry a hydroxyl group on C-9; the position of this group below the plane of the molecule is denoted by α [201].

The first indication of **prostaglandins in the invertebrates** was obtained in 1969 when prostaglandins (15R)- and (15S)-PGA_2 (unknown in the

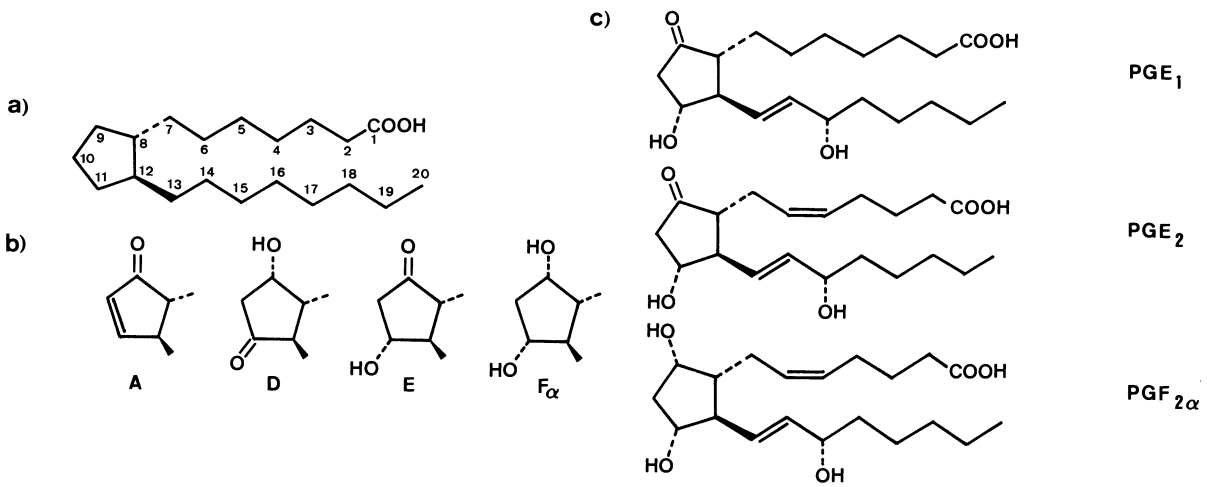


Fig. 15.29a–c. Prostaglandins [201]. **a** Prostanic acid; **b** ring portion of various prostaglandins; **c** structural formulae of three common prostaglandins. See text for an explanation of the classification and nomenclature

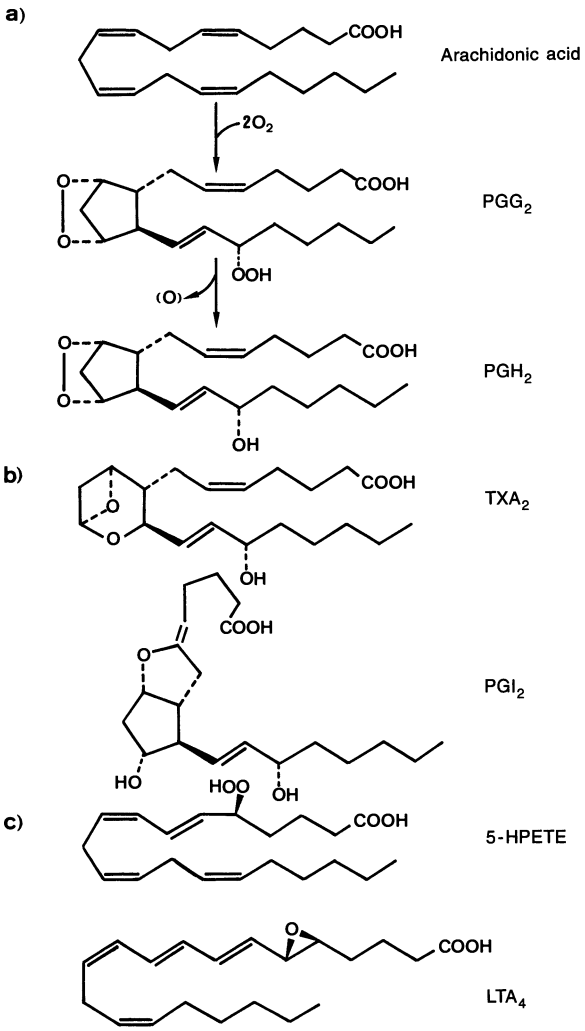


Fig. 15.30a–c. Eicosanoids [201]. **a** Biosynthesis of prostaglandins PGG_2 and PGH_2; **b** thromboxane *(TXA_2)* and prostacyclin *(PGI_2)*; **c** hydroperoxy-eicosatetraenic acid *(HPETE)* and leucotrine *(LT)*

mammals) were found in relatively large amounts in the gorgonian *Plexaura homomalla*. PGA_2 acts to deter fish and is apparently accumulated by the coral as a defence substance. Large amounts of prostaglandins are otherwise found only in the seminal vesicles of some mammals and insects; most tissues do not accumulate prostaglandins but, on the contrary, subject them to rapid turnover. Prostaglandin-like substances are found in many, if not all, gorgonians (Gorgonacea): PGF occurs in *Lobophyton depressum*, claviridenone occurs in *Clavularia viridis*, and 1-chloro,12-hydroxyprostanoids occur in *Telesto riisei;* in contrast, no prostaglandins are found in, for example, *Pseudoplexaura porosa* [29, 45]. Prostaglandins have been found in many molluscs, e.g. in the freshwater snail *Heliosoma durgi* and the mussels *Mytilus californianus* and *Ligunia subrostrata*. The gastropod *Tethys fimbria* (Opisthobranchia) synthesizes prostaglandin-1,15-lactones of the E and F series, and on the basis of their tissue distribution these have been assigned a variety of biological functions in defensive behaviour, smooth muscle contraction and egg production or fertilization control [60].

The first indication of prostaglandins in the **insects** came in 1974 from the orthopteran *Acheta domestica*. Since then, PGE and PGF have been identified in a whole series of insects from different orders, with the variation in concentration ranging over many orders of magnitude. The biosynthesis of prostaglandins has been shown in various insects both in vitro and in vivo [25, 271]. As expected, prostaglandins of series 1 are produced from (n-6) fatty acids 20:3, and series 2 are produced from arachidonic acid 20:4 (Fig. 15.7). PG synthase activity is found in the microsomes of *Musca domestica*, as in the mammals, but in *Bombyx mori* it is present in the 100 000-*g* supernatant, and in *Acheta* and *Teleogryllus commodus* it is present in the 300-*g* pellet. It is not clear how far these results really indicate differences in the intracellular localization of the enzyme complex [271]. PG synthase is mainly found in certain sections of the male genital tract. In the orthopterans *Locusta migratoria, Teleogryllus commodus* and *Acheta domestica*, and in the bug *Triatoma infestans*, during copulation PG synthase activity is transferred together with the contents of the spermatophore to the female spermatheca, where it catalyses prostaglandin synthesis from the available arachidonic acid. In the silkworm *Bombyx mori*, however, PGE_2 and $PGF_2\alpha$ are transferred directly from the male to the female [26, 145]. Prostaglandins in insects appear to play an important role in reproduction, similar to that played in humans and mammals, in that the prostaglandins are involved in a variety of ways in all phases of reproduction from conception to birth. For example, stimulation of egg laying by PGE_2 could be demonstrated in *Acheta* and *Teleogryllus* but not in *Locusta migratoria* or *Trichoplusia ni*. Inhibitors of prostaglandin synthesis, such as steroidal inflammation inhibitors, which block the release of arachidonic acid, or aspirin, indomethacin and phenylbutazone, which affect the PG synthase complex, disturb the whole of development in the mosquito *Culex pipiens;* they also affect the flight capacity of the adults [58]. Prostaglandins are apparently

involved in the triggering of immune reactions in the moth *Manduca sexta* [254].

From PGH_2, the first stable intermediate of prostaglandin synthesis, further **metabolic pathways lead to the other eicosanoids**, thromboxanes and prostacyclins (Fig. 15.30b), about which, however, there is little comparative information. Arachidonic acid is oxidized not only by the cyclooxygenase activity of PG synthase but also by lipoxygenases, without ring closure, to hydroperoxy-5,8,10,14- eicosatetraenic acids (HPETEs); it is further converted to leucotrienes (Fig. 15.30c) and hydroxy-eicosatetraenic acids (HETEs) [201, 247]. The specific lipoxygenases show large differences in their sequences, e.g. the human 15-lipoxygenase has only 39 % of its 660 amino acids in common with the 5-lipoxygenase [241]. Leucotriene B_4 (LTB_4) has proinflammatory and immune-mediator activities in mammals, and has similar functions in the frog *Rana temporaria* and the trout *Salmo gairdneri* [72, 205]. HETEs and related hydroxy fatty acids with certain localizations and stereo-configurations of the hydroxy group act in lower vertebrates and invertebrates as triggers for oocyte maturation, e.g. 15(S)- and 12(S)-HETE in the frog *Xenopus laevis*, 8(R)-HETE in several starfish, and 11(R)- and 12(R)-HETE in the sea urchin *Strongylocentrotus purpuratus* [88]. The trihydroxy-eicosatrienic, -eicosatetraenic and -docosapentaenic acids, synthesized by the barnacle *Balanus balanoides* and excreted into the water, function as hatching factors [248]. In the sea snail *Aplysia californica*, the 12-ketoeicosatetraenic acid produced from 12-HPETE appears to be an intracellular signal mediating the effects of the peptide hormone FMRF amide [206]. The formation of 8-HETE and other HETEs from arachidonic acid has been demonstrated in the thysanuran *Thermobia domestica;* 5-HETE and 12-HETE have been found in the freshwater mussel *Ligumia subrostrata;* and 15-HPETE has been found in the coral *Pseudoplexaura porosa*, which produces no prostaglandins and accordingly has no cyclooxygenase [79, 212].

15.5.2 Insect Juvenile Hormones

Based upon investigations of the bug *Rhodnius prolixus*, Wigglesworth in the 1930s postulated the existence of a hormonal factor that originated from the corpora allata and, whilst allowing continued growth, prevented metamorphosis. In 1967, Röller defined the chemical structure of such a juvenile hormone (now known as JH I) from *Hyalophora cecropia*. Since then, four further **juvenile hormones** have been detected in the corpora allata or the embryos of this lepidopteran species (Fig. 15.31). Complicated JH spectra appear to be restricted to the Lepidoptera; insects of other orders have only JH III. Early results showing the presence of JH I and JH II in the cockroach *Nauphoeta cinerea*, the housefly *Musca domestica* and the migratory locust *Locusta migratoria* have not survived the test of newer detection methods [31, 62, 150]. The JH spectrum of the lepidoptera changes during development; in *Manduca sexta*, the egg contains mainly JH 0 and JH iso-0, the larvae contain JH II and JH I, and the adults possess JH III and JH II. The basic regulatory mechanisms, at the

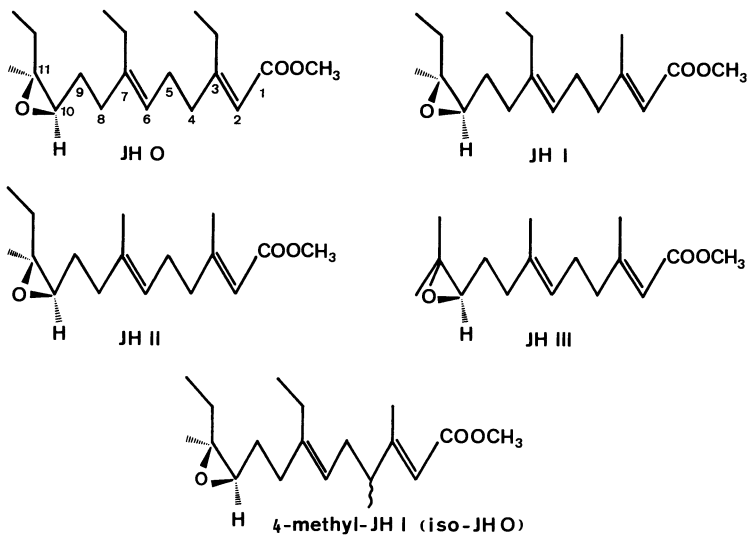


Fig. 15.31. Juvenile hormones [62]

level of either precursor concentrations or the enzymes for formation and degradation, have not yet been described in detail. In the emperor moths, there is a marked difference between the sexes in the JH titre: freshly hatched males contain 16.9 ng JH II/g and 0.4 ng JH I/g compared with 0.06 and 0.01 ng/g, respectively, in the females. The biological significance of this sexual dimorphism awaits explanation. Juvenile hormone is passed from male to female during mating; however, allatectomied males and females produce viable progeny [8]. Whereas mosquito larvae have only JH III, *Drosophila melanogaster* also produces the unusual JHB_3 (JH bisepoxide) [222]. Traces of JH III have been detected in the crustacean *Libinia emarginata* but this is probably an oxidation artefact of the 1000-fold higher concentration of methylfarnesinic acid [147]. JH III has even been isolated from plants [263].

The process of **juvenile hormone synthesis** has been studied with tracer experiments and in vitro investigations of many insect species [234]. The biosynthesis of JH III occurs via the classical isoprenoid pathway (see Fig. 16.2, p. 625) via farnesyl pyrophosphate, which by the cleavage of the pyrophosphate residue, oxidation to the acid, epoxidation and methyl transfer is converted to JH III. In the lepidopterans, the epoxidation occurs first, followed by the methyl transfer; the reverse is found in the Blattodea, Mantodea and Orthoptera [171]. The origin of the ethyl branching in the other juvenile hormones is something of a mystery. They occur from otherwise unknown homoisoprene subunits (Fig. 15.32), which may be produced from one molecule of propionyl-CoA and two molecules of acetyl-CoA with homoevalonic acid as the intermediate product. Tracer experiments with propionic acid and acetic acid have confirmed this hypothesis and shown that the propionyl-CoA in *Manduca* originates from valine or isoleucine metabolism [27]. Analogous to the known isoprene subunits (Fig. 16.2), the homoisoprenes are homoisopentenyl pyrophosphate and the isomeric ethylmethylallyl pyrophosphate; these can be interconverted by the action of a specific isomerase [62, 234]. The type of juvenile hormone produced is dependent upon the ratio of homoisoprene to isoprene subunits: JH 0 (3:0), JH I (2:1), JH II (1:2) and JH III (0:3).

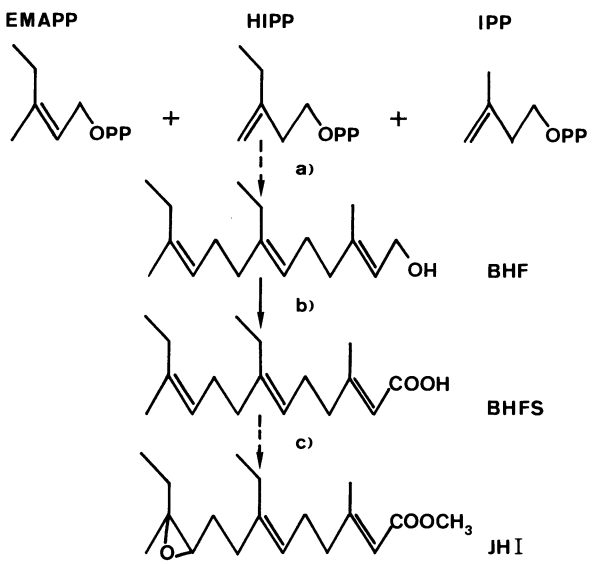


Fig. 15.32. Biosynthesis of the juvenile hormone JH I [62]. A multiple-step reaction from one molecule each of ethylmethylallyl pyrophosphate *(EMAPP)*, homoisopentenyl pyrophosphate *(HIPP)* and isopentenyl pyrophosphate *(IPP)* results in 7,11-bishomofarnesol *(BHF)*, which is oxidized to 7,11-bishomofarnesinic acid *(BHFS)*, this in turn is converted to JH I by epoxidation and esterification

Not much is known about **enzymes of JH biosynthesis**. 3-Hydroxy-3-methylglutaryl-CoA reductase (HMG reductase), which is responsible for the production of mevalonic acid as the precursor of the reactive isoprene subunits (Fig. 16.2), appears to be just as rate limiting for JH biosynthesis as for the synthesis of steroids. It is subject to feedback inhibition by isoprenoid, and in both the corpora allata and mammalian liver this involves inactivation by phosphorylation [234]. Isopentenyl pyrophosphate isomerase, which interconverts the two isomeric homoisoprene subunits, and farnesyl pyrophosphate synthase have been partially purified from *Bombyx mori* [139, 140]. Relatively more is known about the enzymes of the last two steps of JH biosynthesis, the epoxidation and the esterification. The epoxidase is an NADPH-dependent microsomal monooxygenase of the cytochrome P450 type, and the soluble O-methyl transferase utilizes S-adenosylmethionine as the methyl group donor. The corpora allata of males of the large silkmoth *Attacus atlas*, in which almost all the steps of JH synthesis take place, have no O-methyl transferase activity; most of the esterification of the epoxy acid occurs in the accessory glands of the male genital tract [8].

The inactivation and degradation of juvenile hormones are initiated by the hydration of the epoxy group to 1,2-diol, or by hydrolytic cleavage of the methyl ester. The relative importance of the two degradation pathways varies with the species, the tissue and the developmental stage. In

the wax moth *Galleria melonella*, 50% of JH III degradation in the fat body at the onset of the last larval stage occurs via ester hydrolysis; in the integument the value is 80% and in the silk glands it is 90%. The involvement of epoxide hydration increases during this larval stage. **Epoxide ether hydrolases** (formerly epoxide hydratases) have been found in numerous insects. They are present in various tissues, especially in the fat body and the midgut, but are not found in the haemolymph [83, 282]. **Esterases** with the capacity to cleave juvenile hormones occur in the haemolymph of probably all insects. However, a distinction between specific JH esterases and specific "general" esterases is not easy to make so long as none of them have been purified to homogeneity. The characteristics of such enzymes appear to be that they show little activity with the typical substrates of unspecific esterases, such as α-naphthylacetate; that they cleave not only free juvenile hormone but also that bound to specific carrier proteins; and that they show little inhibition by diisopropylfluorophosphate (DP). Attempts to purify JH esterases have so far been mainly undertaken for the Lepidoptera. The enzymes vary greatly in molecular mass between 34 and more than 100 kDa [53, 225, 269, 288]. The sequencing of a multiple JH esterase from the moth *Heliothis virescens* showed the enzyme to be a serine carboxyl esterase [86].

The haemolymph of many insects contains **JH-binding proteins** (JHBPs), which can be classified as low affinity ($K_D \geq 10^{-5}$ mol/l) or high affinity ($K_D \leq 10^{-7}$mol/l). Amongst the high-affinity JHBPs are some with low (20–40 kDa) and some with high (100–450 kDa) molecular masses. Low molecular weight JHBPs are typical of the Lepidoptera, but also occur together with high molecular weight JHBPs in *Chironomus thummi* [283]. Blattodea, Orthoptera, Coleoptera and Diptera all have exclusively high molecular weight JHBPs [218, 237]. In the cockroaches *Periplaneta americana* and *Leucophaea madeira*, the beetle *Leptinotarsa decemlineata* and the non-biting midge *Chironomus thummi*, but not in the locust *Locusta migratoria*, the high molecular weight JHBP is identical to the lipid shuttle protein lipophorin [136, 218, 283]. The JHBPs immediately take up the juvenile hormone released from the corpora allata, prevent its transfer to lipid stores, and thereby maintain its availability in the haemolymph. The high-affinity JHBPs protect the juvenile hormone from degradation by unspecific esterases, but because of their even higher affinities the specific JH esterases are able to remove juvenile hormone from the carrier proteins and hydrolyse it. Protection of the hormone from esterolytic cleavage by JHBPs is thus only effective so long as the titre of the specific-JH esterases remains low, and the juvenile hormone concentration is regulated not only by biosynthesis but mainly by the specific activity of JH esterases [83].

The only JHBP to have been sequenced so far is a monomer of 25 kDa (226 amino acids); this was from the tobacco hornworm *Manduca sexta*. Its sequence shows no similarity to any other protein and, consequently, constitutes a new protein super-family [149]. The JHBPs are specific for the juvenile hormone type. For example, the low molecular weight JHBPs of *Manduca sexta* and *Galleria melonella* bind the JH I and JH II, which are typical of the Lepidoptera, much better than the ubiquitous JH III; the reverse order is true for the high molecular weight JHBPs from *Leptinotarsus decemlineata, Chironomus thummi* and *Drosophila melanogaster* [200, 237, 283]. The JHBPs are also stereo-specific, e.g. the affinity of the high molecular weight JHBP of the cockroach *Diploptera punctata* is higher for the natural 10(R)-isomer of JH III than for the 10(S)-isomer [131]. **Intracellular JHBPs**, which without doubt also have receptor functions, have been reported from various insects. In the epidermis of *Manduca sexta*, each cell nucleus contains about 10^4 molecules of a 29-kDa protein of high JH affinity, and the cytoplasm has a 38-kDa protein with lower affinity; this may be an intracellular JH carrier protein [202]. A JH III-specific, 85-kDa protein, isolated from the fat body of *Drosophila melanogaster*, stimulates protein synthesis by male accessory glands cultured in vitro, and is apparently a JH receptor [238].

References

1. Abdel-Halim M. and Farah S. I.: Short time regulation of acetyl CoA carboxylase. Comp. Biochem. Physiol. Pt. B 81: 9–19 (1985)
2. Aloia R. C. and Raison J. K.: Membrane function in mammalian hibernation. Biochim. biophys. Acta 988: 123–146 (1989)
3. Araki S. et al.: Structure of a triphosphopentasoylceramide containing 4-O-methyl-N-acetylglucosamine from the skin of the sea hare, Aplysia kurodai. J. biol. Chem. 262: 14141–45 (1987)
4. Araki S. et al.: Novel phosphonoglycosphingolipids containing pyruvylated galactose from the nervous system of Aplasia kurodai. Eur. J. Biochem. 198: 689–695 (1991)

5. Argiles J. M.: Has acetone a role in the conversion of fat to carbohydrate in mammals? Trends biochem. Sci. 11: 61–63 (1986)
6. Avrova N. F.: The effect of natural adaptations of fishes to environmental temperature on brain ganglioside fatty acid and long chain base composition. Comp. Biochem. Physiol. Pt. B 78: 903–909 (1984)
7. Avrova N. F. et al.: Systematic position of fish species and ganglioside composition and content. Comp. Biochem. Physiol. Pt. B 83: 669–676 (1986)
8. Baker F. C. et al.: Identification of the juvenile hormones from adult Attacus atlas. Insect Biochem. 15: 321–324 (1985)
9. Barry R. J. et al.: Elucidation of the 2-aminoethylphosphonate biosynthetic pathway in Tetrahymena pyriformis. Biochem. biophys. Res. Commun. 153: 177–182 (1988)
10. Bartelt R. J., Krick T. P. and Jones R. L.: Cuticular hydrocarbones of the yellowhead spruce sawfly, Pikonema alaskensis. Insect Biochem. 14: 209–213 (1984)
11. Bartelt R. J. et al.: Comparative analysis of cuticular hydrocarbons in the Drosophila virilis species group. Comp. Biochem. Physiol. Pt. B 83: 731–742 (1986)
12. Beenakkers A. M. T., van der Horst D. J. and van Marrevijk W. J. A.: Biochemical processes directed to flight muscle metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 451–486. Pergamon Press, Oxford 1985
13. Beis, A., Zammit V. A. and Newsholme E. A.: Activities of 3-hydroxybutyrate dehydrogenase, 3-oxoacid-CoA-transferase and acetoacetyl-CoA thiolase in relation to ketone-body utilization in muscles from vertebrates and invertebrates. Eur. J. Biochem. 104: 209–215 (1980)
14. Bell M. V., Henderson R. J. and Sargent J. R.: The role of polyunsaturated fatty acids in fish (Minireview). Comp. Biochem. Physiol. Pt. B 83: 711–719 (1986)
15. Bell M. V. and Tocher D. R.: Molecular species composition of the major phospholipids in brain and retina from rainbow trout (Salmo gairdneri). Occurrence of high levels of di-(n-3)-polyunsaturated fatty acid species. Biochem. J. 264: 909–915 (1989)
16. Beninger P. G. and Stephan G.: Seasonal variations in the fatty acids of the triacylglycerols and phospholipids of two populations of adult clam (Tapes decussatus L. and T. philippinarum) reared in a common habitat. Comp. Biochem. Physiol. Pt. B 81: 591–601 (1985)
17. Bieber L. L.: Carnitine. Annual Rev. Biochem. 57: 261–283 (1988)
18. Blomquist G. J., Chu A. J. and Remaley S.: Biosynthesis of wax in the honeybee, Apis mellifera L. Insect Biochem. 10: 313–321 (1980)
19. Blomquist G. J., Toolson E. C. and Nelson D. R.: Epicuticular hydrocarbons of Drosophila pseudoobscura (Diptera, Drosophilidae). Identification of unusual alkadiene and alkatriene positional isomers. Insect Biochem. 15: 25–34 (1985)
20. Blomquist G. J., Roubik D. W. and Buchmann S. L.: Wax chemistry of two stingless bees of the Trigonisca group (Apidae: Meliponinae). Comp. Biochem. Physiol. Pt. B 82: 137–142 (1985)
21. Blomquist G. J. and Dillwith J. W.: Cuticular lipids. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 3, pp. 117–154. Pergamon Press, Oxford 1985
22. Blomquist G. J., Nelson D. R. and de Renobales M.: Chemistry, biochemistry, and physiology of insect cuticular lipids (Review). Arch. Insect Biochem. Physiol. 6: 227–265 (1987)
23. Blomquist G. J., Borgeson C. E. and Vundla M.: Polyunsaturated fatty acids and eicosanoids in insects. Insect Biochem. 21: 99–106 (1991)
24. Bowman E. D. et al.: Purification and characterization of the Tetrahymena pyriformis P-C bond forming enzyme phosphoenolpyruvate phosphomutase. Biochemistry 29: 7059–63 (1990)
25. Brady U. E.: Prostaglandins in insects. Insect Biochem. 13: 443–451 (1983)
26. Brenner R. R. and Bernasconi A.: Prostaglandin biosynthesis in the gonads of the hematophagous insect Triatoma infestans. Comp. Biochem. Physiol. Pt. B 93: 1–4 (1989)
27. Brindle P. A. et al.: Comparative metabolism of branched-chain amino acids to precursors of juvenile hormone biogenesis in corpora allata of lepidopterous versus nonlepidopterous insects. J. Biol. Chem. 263: 10653–57 (1988)
28. Buckner J. S. et al.: Novel surface lipids of diapausing Manduca sexta pupae. Long chain oxoalcohol esters of acetoacetic, hydroxybutyric, and acetic acids. J. biol. Chem. 259: 8461–70 (1984)
29. Bunbdy G. L. et al.: Discovery of an arachidonic acid C-8 lipoxygenase in the gorgonian coral Pseudoplexaura porosa. J. biol. Chem. 261: 747–751 (1986)
30. Burken R. R., Wertz P. W. and Downing D. T.: A survey of polar and nonpolar lipids extracted from snake skin. Comp. Biochem. Physiol. Pt. B 81: 315–318 (1985)
31. Camps F. et al.: Identification of juvenile hormone III in the hemolymph of Blattella germanica adult females by gas chromatography-mass spectrometry. Arch. Insect Biochem. Physiol. 6: 181–189 (1987)
32. Cane J. H. and Brooks R. W.: Dufour's gland lipid chemistry of three species of centris bees (Hymenoptera: Apoidea, Anthoporidae). Comp. Biochem. Physiol. Pt. B 76: 895–897 (1983)
33. Carballeira N. et al.: Biosynthetic studies of marine lipids: 5. The biosynthesis of long-chain branched fatty acids in marine sponges. J. org. Chem. 51: 2751–56 (1986)
34. Carey M. C., Small D. M. and Bliss C. M.: Lipid digestion and absorption. Annual Rev. Physiol. 45: 651–677 (1983)
35. Chamberlin M. E. and Phillips J. E.: Oxidative metabolism in the locust rectum. J. comp. Physiol. B 151: 191–198 (1983)
36. Chang S. I. and Hammes G. G.: Homology analysis of the protein sequences of fatty acid synthetases from chicken liver, rat mammary gland, and yeast. Proc. Nat. Acad. Sci. USA 86: 8373–76 (1989)
37. Chapelle S.: Aspects of phospholipid metabolism in crustaceans as related to changes in environmental temperatures and salinities (Review). Comp. Biochem. Physiol. Pt. B 84: 423–439 (1986)
38. Chapelle S.: Plasmalogens and O-alkylglycerophospholipids in aquatic animals (Review). Comp. Biochem. Physiol. Pt. B 88: 1–6 (1987)
39. Chase J. et al.: Biosynthesis of methyl branched hydrocarbons of the German cockroach Blatella germanica (L.) (Orthoptera, Blatellidae). Insect Biochem. 20: 149–156 (1990)

40. Cheesbrough T. M. and Kolattukudy P. E.: Purification and characterization of a fatty acyl-CoA hydrolase from the uropygial glands of peking ducks (Anas domesticus). Arch. Biochem. Biophys. 237: 208–216 (1985)
41. Chelomin V. P. and Zhukova N. V.: Lipid composition and some aspects of aminophospholipid organization in erythrocyte membrane of the marine bivalve mollusc Scapharca broughtoni (Schrenck). Comp. Biochem. Physiol. Pt. B 69: 599–604 (1981)
42. Chirala S. S. et al.: Complementation of mutations and nucleotide sequence of FAS1 gene encoding β-subunit of yeast fatty acid synthase. J. biol. Chem. 262: 4231–40 (1987)
43. Christiansen J. A.: Changes in phospholipid classes and fatty acids and fatty acid desaturation and incorporation into phospholipids during temperature acclimation of green sunfish Lepomis cyanellus R. Physiol. Zool. 57: 481–492 (1984)
44. Chu A. J. and Blomquist G. J.: Decarboxylation of tetracosanoic acid to n-tricosane in the termite Zootermopsis angusticollis. Comp. Biochem. Physiol. Pt. B 66: 313–317 (1980)
45. Ciereszko L. S. et al.: Prostaglandin (15S)-PGA, derivatives in the gorgonian Plexaura homomalla (Esper.), forma kükenthali Moser. Experientia 41: 37–38 (1985)
46. Ciminiello P. et al.: Furan fatty acid steryl esters from the marine sponge Dictyonella incisa which show inflammatory activity. Experientia 47: 739–743 (1991)
47. Collins D. M. and Hargis P. S.: Distribution of fatty acid binding proteins in tissues and plasma of Gallus domesticus. Comp. Biochem. Physiol. Pt. B 92: 283–289 (1989)
48. Colton VI S. W. and Downing D. T.: Variation in skin surface lipid composition among the Equidae. Comp. Biochem. Physiol. Pt. B 75: 429–433 (1983)
49. Colton VI S. W. et al.: Skin surface lipids of the mink. Comp. Biochem. Physiol. Pt. B 84: 369–371 (1986)
50. Cosper C. I. and Ackman R. G.: Occurrence of cis-9,10-methylenehexadecanoic and cis-9,10-methyleneoctadecanoic acids in the lipids of immature and mature Fundulus heteroclitus (L.) and in roe. Comp. Biochem. Physiol. Pt. B 75: 649–654 (1983)
51. Cossins A. R. and Macdonald A. G.: Homoviscous adaptation under pressure: III. The fatty acid composition of liver mitochondrial phospholipids of deep-sea fish. Biochim. biophys. Acta 860: 325–335 (1986)
52. Cripps C. et al.: The delta-12 desaturases from the house cricket, Acheta domesticus (Orthoptera: Gryllidae): Characterization and forms of the substrate. Arch. Biochem. Biophys. 278: 46–51 (1990)
53. Dabrowski U. et al.: Novel phosphorus-containing glycosphingolipids from the blowfly Calliphora vicina Meigen. Strutural analysis by H1- and H1[P32]-edited NMR spectroscopy at 500 and 600 megahertz. J. Biol. Chem. 265: 9737–43 (1990)
54. Dadd R. H. and Kleinjan J. E.: Prostaglandin synthetase inhibitors modulate the effect of essential dietary arachidonic acid in the mosquito Culex pipiens. J. Insect Physiol. 30: 721–728 (1984)
55. Daum G.: Lipids of mitochondria (Review). Biochim. biophys. Acta 822: 1–42 (1985)
56. Dennis R. D. et al.: Glycosphingolipids in insects. Chemical structures of ceramide monosaccharide, disaccharide, and trisaccharide from pupae of Calliphora vicina (Insecta: Diptera). Eur. J. Biochem. 146: 51–58 (1985)
57. Devaux P. F.: Static and dynamic lipid asymmetry in cell membranes. Biochemistry 30: 1163–73 (1991)
58. Diagne A. et al.: Studies on ether phospholipids. II. Comparative composition of various tissues from human, rat and guinea pig. Biochim. biophys. Acta 793: 221–231 (1984)
59. Dikarev V. P., Svetashev V. I. and Vaskovsky V. E.: Noctiluca miliaris – one more protozoan with unusual lipid composition. Comp. Biochem. Physiol. Pt. B 72: 137–140 (1982)
60. Dimarzo V. et al.: A novel multifunctional metabolic pathway in a marine mollusc leads to unprecedented prostaglandin derivatives (prostaglandin-1,15-lactones). Biochem. J. 273: 593–600 (1991)
61. Dorset D. L. and Ghiradella H.: Insect wax secretion. The growth of tubular crystals. Biochim. biophys. Acta 760: 136–142 (1983)
62. Downer R. G. H. and Laufer H. (eds.): Endocrinology of insects. Alan R. Liss, New York 1983
63. Downer R. G. H.: Lipid metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 77–113. Pergamon Press, Oxford 1985
64. Downing D. T. and Lindholm J. S.: Skin surface lipids of the cow. Comp. Biochem. Physiol. Pt. B 73: 327–330 (1982)
65. Fahimi H. D. and Sies H. (eds.): International symposium on peroxisomes in biology and medicine. Springer, Berlin 1987
66. Farkas T.: Adaptation of fatty acid composition to temperature – A study on carp (Cyprinus carpio L.) liver slices. Comp. Biochem. Physiol. Pt. B 79. 531–535 (1984)
67. Florin-Christensen J. et al.: Phospholipases and phosphonolipids in a ciliate: an attack and defense system? Trends biochem. Sci. 11: 354–355 (1986)
68. Fricke H. and Oehlenschläger J.: Fatty acid and sterol composition of the Antarctic amphipod Themisto gaudichaudii Guerin 1828. Comp. Biochem. Physiol. Pt. B 89: 39–42 (1988)
69. German J. B., Bruckner G. G. and Kinsella J. E.: Lipoxygenase in trout gill tissues acting on arachidonic, eicosapentaenoic and docosahexaenoic acids. Biochim. biophys. Acta 875: 12–20 (1986)
70. van Golde L. M. G.: Synthesis of surfactant lipids in the adult lung. Annual Rev. Physiol. 47: 765–774 (1985)
71. Goubern M and Cadot M.: Ketone bodies in cold-acclimated rats. Comp. Biochem. Physiol. Pt. B 76: 741–744 (1983)
72. Green F. A. et al.: Leukotrienes and related eicosanoids are produced by frog leukocytes. Biochem. biophys. Res. Commun. 142: 309–314 (1987)
73. Griffiths M.: The biology of monotremes. Acad. Press, New York 1978
74. Guo L. and Blomquist G. J.: Identification, accumulation, and biosynthesis of the cuticular hydrocarbons of the southern armyworm, Spodoptera eridania (Cramer) (Lepidoptera, Noctuidae). Arch. Insect Biochem. Physiol. 16: 19–30 (1991)
75. Guppy M., Hulbert W. C. and Hochachka P. W.: Metabolic sources of heat and power in tuna muscles: II. Enzyme and metabolite profiles. J. exp. Biol. 82: 303–320 (1978)
76. Gurr M. I. and Harwood J. L.: Lipid biochemistry. An introduction. 4th ed.. Capman & Hall, New York 1990

77. Hadley N. F., Jackson L.L. and Leader J.: Cuticular hydrocarbons of the New Sealand alpine weta Hemideina maori (Orthoptera, Stenopelmatidae). Comp. Biochem. Physiol. Pt. B 91: 685–689 (1988)
78. Hagar A. F. and Dietz T. H.: Seasonal changes in the lipid composition of gill tissue from the freshwater mussel Carunculina texasensis. Physiol. Zool. 59: 419–428 (1986)
79. Hagar A. F., Hwang D. H. and Dietz T. H.: Lipoxygenase activity in the gills of the freshwater mussel, Ligumia subrostrata. Biochim. biophys. Acta 1005: 162–169 (1989)
80. Hahn S. et al.: Unusual pattern of fatty acid biosynthesis. Evidence for C-19 desaturase activity in freshwater sponges. J. Biol. Chem. 264: 21043–46 (1989)
81. Halarnkar P. P. and Blomquist G. J.: Comparative aspects of propionate metabolism. Comp. Biochem. Physiol. Pt. B 92: 227–231 (1989)
82. Halver J. E. (ed.): Fish nutrition. Second edition. Acad. Press, San Diego 1989
83. Hammock B.: Regulation of juvenile hormone titer: Degradation. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 7, pp. 431–472. Pergamon Press, Oxford 1985
84. Hanahan D. J.: Platelet activation factor: A biologically active phosphoglyceride. Annual Rev. Biochem. 55: 483–509 (1986)
85. Hansen H. O., Jensen S. S. and Knudsen J.: Absence of monoacylglycerol pathway for triacylglycerol synthesis in goat mammary gland. Biochem. J. 238: 173–176 (1986)
86. Hanzlik T. N. et al.: Isolation and sequencing of cDNA clones coding for juvenile hormone esterase from Heliothis virescens. J. Biol. Chem. 264: 12419–25 (1989)
87. Haunerland N. H. and Chisholm J. M.: Fatty acid binding protein in flight muscle of the locust, Schistocerca gregaria. Biochim. biophys. Acta 1047: 233–238 (1990)
88. Hawkins D. J. and Brash A. R.: Lipoxygenase metabolism of polyunsaturated fatty acids in oocytes of the frog Xenopus laevis. Arch. Biochem. Biophys. 268: 447–455 (1989)
89. Hawthorne J. N. and Ansell G. B. (eds.): Phospholipids. New comprehensive biochemistry Vol 4. Elsevier, Amsterdam 1982
90. Hayashi A. and Matsubara T.: A new homologue of phosphonoglycosphingolipid, N-methylaminoethylphosphonyltrigalactosylceramide. Biochim. biophys. Acta 1006: 89–96 (1989)
91. Hayashi A., Nishimura Y. and Matsubara T.: Occurrence of ceramide digalactoside as the main glycosphingolipid in the marine sponge Halichondria japonica. Biochim. biophys. Acta 1083: 179–186 (1991)
92. Hazel J. R. and Zerba E.: Adaptation of biological membranes to temperature: Molecular species compositions of phosphatidylcholine and phosphatidylethanolamine in mitochondrial and microsomal membranes from thermally acclimated rainbow trout. J. comp. Physiol. B 156: 665–674 (1986)
93. Hazel J. R. et al.: Thermal adaptation in biological membranes: interacting effects of temperature and pH. J. comp. Physiol. B 162: 593–601 (1992)
94. Henderson R. J. and Sargent J. R.: Chain-length specificities of mitochondrial and peroxisomal β-oxidation of fatty acids in livers of rainbow trout (Salmo gairdneri). Comp. Biochem. Physiol. Pt. B 82: 79–85 (1985)
95. Hilbig R.: Structure related phylogenetic variations in brain gangliosides of vertebrates. Comp. Biochem. Physiol. Pt. B 77: 151–160 (1984)
96. Hilderson H. J. (ed.): Intracellular transfer of lipid molecules. Subcellular Biochemistry, Vol. 16. Plenum, New York 1990
97. Holz G. G. jr. et al.: Biosynthesis of the novel fatty acid, 17-methyl-cis-9,10-methyleneoctadecanoic acid, by the parasitic protozoan, Herpetomonas megaseliae. Lipids 18: 607–610 (1983)
98. Holzer K. P., Liu W. and Hammes G. G.: Molecular cloning and sequencing of chicken liver fatty acid synthase cDNA. Proc. Nat. Acad. Sci. USA 86: 4387–91 (1989)
99. Horst M. N.: Dolichol phosphorylation occurs via a CTP-dependent reaction in Artemia larvae. J. exp. Zool. 252: 16–24 (1989)
100. Hovius R. et al.: Improved methods to isolate and subfractionate rat liver mitochondria. Lipid composition of the inner and outer membrane. Biochim. biophys. Acta 1021: 217–226 (1990)
101. Hutzell P. A. and Krushberg L. R.: Fatty acid compositions of Caenorhabditis elegans and C. briggsae. Comp. Biochem. Physiol. Pt. B 73: 517–520 (1982)
102. Irazu C. E., Pollero R. J. and Brenner R. R.: Occurrence of 22:2 nonmethylene interrupted dienoic fatty acid and its seasonal distribution among lipids and tissues of the freshwater bivalve Diplodon delodontus from an isolated environment. Lipids 19: 649–655 (1984)
103. Isay S. V. and Bustarova N. G.: Study on fatty acid composition of marine organisms – I. Unsaturated fatty acids of Japan sea invertebrates. Comp. Biochem. Physiol. Pt. B 77: 803–810 (1984)
104. Isay S. V., Osheva O. N. and Makarchenko M. A.: Study on glyceryl ethers – II. α-Glyceryl ether content in tissues of Octopus dofleini. Comp. Biochem. Physiol. Pt. B 77: 799–801 (1984)
105. Itasaka O. et al.: Di- and trimannosylceramides in hepatopancreas of a fresh-water bivalve, Hyriopsis schlegelii. Biochim. biophys. Acta 751: 8–13 (1983)
106. Itonori S. et al.: Characterization of a new phosphono-cerebroside, N-methyl-2-aminoethylphosphonylglucosylceramide, from the Antarctic crill, Euphausia superba. Biochim. biophys. Acta 1081: 321–327 (1991)
107. Jackson L. L.: Epicuticular lipid composition of the sand cockroach, Arenivaga investigata. Comp. Biochem. Physiol. Pt. B 74: 255–257 (1983)
108. Jacob J. and Ziswiler V.: The uropygial gland. In: Farner D. S., King J. R. and Parkes K. C. (eds.): Avian biology, Vol. 6, pp. 199–324. Acad. Press, New York 1982
109. Jacob J.: Stomach oils. In: Farner D. S., King J. R. and Parkes K. C. (eds.): Avian biology, Vol. 6, pp. 325–340. Acad. Press, New York 1982
110. Jacob J.: Chemosystematik in der Zoologie. Verh. Dtsch. Zool. Ges. 77: 69–85 (1984)
111. Jacob J. and Hoerschelmann H.: Similarities within the pattern of uropygial gland secretion from tinamous, kiwis and galliform species. Comp. Biochem. Physiol. Pt. B 81: 945–947 (1985)
112. Jacobs G., Herrmann H. and Gercken G.: Incorporation of (1-C14)acetate into fatty acids and aliphatic moieties of glycerolipids in Leishmania donovani promastigotes. Comp. Biochem. Physiol. Pt. B 73: 367–373 (1982)

113. Jang S. H., Cheesbrough T. M. and Kolattukudy P. E.: Molecular cloning, nucleotide sequence, and tissue distribution of malonyl-CoA decarboxylase. J. Biol. Chem. 264: 3500–05 (1989)
114. Juarez M. P. and Brenner R. R.: The epicuticular lipids of Triatoma infestans: II. Hydrocarbon dynamics. Comp. Biochem. Physiol. Pt. B 82: 793–803 (1986)
115. Juarez P. and Brenner R. R.: Fatty acid biosynthesis in the integument tissue of Triatoma infestans. Comp. Biochem. Physiol. Pt. B 93: 763–772 (1989)
116. Kaib M., Brandl R. and Bagine R. K. N.: Cuticular hydrocarbon profiles: A valuable tool in termite taxonomy. Naturwissenschaften 78: 176–179 (1991)
117. Kaikaus R. M., Bass N. M. and Ockner R. K.: Function of fatty acid binding proteins (Review). Experientia 46: 617–630 (1990)
118. Kameda K. and Goodridge A. G.: Isolation and partial characterization of the gene for goose fatty acid synthase. J. Biol. Chem. 266: 419–426 (1991)
119. Kaneniwa M. et al.: Fatty acids in Holothurioidea: Occurrence of cis-14-tricosenoic acid. Comp. Biochem. Physiol. Pt. B 84: 451–455 (1986)
120. Kaneniwa M., Okamoto J. and Tagaki T.: Fatty acids in Asteroidea. Occurrence of unusual 5-olefinic acids. Comp. Biochem. Physiol. Pt. B 87: 1037–40 (1987)
121. Karlsson K. A., Leffler H. and Samuelsson B. E.: Characterization of cerebroside (monoglycosylceramide) from the sea anemone, Metridium senile. Identification of the major long-chain base as an unusual dienic base with a methyl branch at a double bond. Biochim. biophys. Acta 574: 79–93 (1979)
122. Kasai R. et al.: Regulatory mechanism of desaturation activity in cold acclimation of Tetrahymena pyriformis, with special reference to quick cryoadaptation. Biochim. biophys. Acta 836: 397–401 (1985)
123. Kasama K., Blank M. L. and Snyder F.: Identification of 1-alkyl-2-acyl-3-(2',3'-diacylglycerol) glycerols, a new type of lipid class, in harderian gland tumors of mice. J. Biol. Chem. 264: 9453–61 (1989)
124. Kates M. and Moanson L. A. (eds.): Membrane fluidity. Plenum, New York 1984
125. Kaya K., Uchida K. and Kusumi T.: Tetrahydroxystearic acid-containing taurolipid isolated from Tetrahymena thermophila. Biochim. biophys. Acta 875: 97–102 (1986)
126. Kaya K., Itoh H. and Nohara K.: 2-(7,13-dihydroxy-2-trans-octadecenoylamino)-ethanesulfonic acid (lipotaurine) as an intermediate of taurolipids biosynthesis. Biochim. biophys. Acta 1042: 338–343 (1990)
127. Kaya K. and Kusumi T.: Identification of pentahydroxystearic acid-containing taurolipid (taurolipid C) isolated from Tetrahymena thermophila. Biochim. biophys. Acta 1042: 198–203 (1990)
128. Keller A. et al.: The pulmonary surfactant protein-C (SP-C) precursor is a type-II transmembrane protein. Biochem. J. 277: 493–499 (1991)
129. Kilby P. M., Radda G. K. and Wegener G.: 2-Aminoethylphosphonic acid is the main phosphorus compound in locust hemolymph. Naturwissenschaften 78: 514–517 (1991)
130. Kilponen J. M., Palosaari P. M. and Hiltunen J. K.: Occurrence of a long chain delta-3,delta-2-enoyl-CoA isomerase in rat liver. Biochem. J. 269: 223–226 (1990)
131. King L. E. and Tobe S. S.: The identification of an enantioselective JH III binding protein from the haemolymph of the cockroach, Diploptera punctata. Insect Biochem. 18: 793–805 (1988)
132. Kitamoto T. et al.: The quaternary structure and activity of newly purified fatty acid synthetase from the Harderian gland of Guinea-pig. Biochim. biophys. Acta 827: 164–173 (1985)
133. Kolattukudy P. E. (ed.): Chemistry and biochemistry of natural waxes. North-Holland, London 1976
134. Kolattukudy P. E. and Rogers L.: Biosynthesis of 3-hydroxy fatty acids, the pheromone components of female mallard ducks, by cell-free preparations from the uropygial gland. Arch. Biochem. Biophys. 252: 121–129 (1987)
135. Komnick H.: Fat transport in the insect intestine (In German). Verh. Dtsch. Zool. Ges. 77: 123–126 (1984)
136. Koopmanschap A. B. and de Kort C. A. D.: Isolation and characterization of a high molecular weight JH-III transport protein in the hemolymph of Locusta migratoria. Arch. Insect Biochem. Physiol. 7: 105–118 (1988)
137. Koudelka A. P. et al.: Oleic acid desaturation in Tetrahymena pyriformis. Biochim. biophys. Acta 751: 129–137 (1983)
138. Koval M. and Pagano R.: Intracellular transport and metabolism of sphingomyelin (Review). Biochim. biophys. Acta 1082: 113–125 (1991)
139. Koyama T., Matsubara M. and Ogura K.: Isoprenoid enzyme systems of silkworm. I. Partial purification of isopentenyl pyrophosphate isomerase, farnesyl pyrophosphate synthetase, and geranylpyrophosphate synthetase. J. Biochem Tokyo 98: 449–456 (1985)
140. 5Koyama T., Matsubara M. and Ogura K.: Isoprenoid enzyme systems of silkworm. II. Formation of the juvenile hormone skeletons by farnesyl pyrophosphate synthetase II. J. Biochem. Tokyo 98: 457–464 (1985)
141. Krook M., Marekov L. and Jornvall H.: Purification and structural characterization of placental NAD-linked 15-hydroxyprostaglandin dehydrogenase. The primary structure reveals the enzyme to belong to the short-chain alcohol dehydrogenase family. Biochemistry 29: 738–743 (1990)
142. Kuksis A. (ed.): Fat absorption, 2 vol. set. CRC Press, Boca Raton USA 1986
143. Lam W. et al.: Biosynthetic studies of marine lipids. 30. Metabolism of unusual membrane phospholipids in the marine sponge Microciona prolifera. Biochemistry 30: 372–377 (1991)
144. Lambremont E. N. and Dial P. F.: Fatty acid composition of major phospholipids from the fat body, flight muscle and testis of the house cricket, and positional distribution of fatty acids in testicular phospholipids. Comp. Biochem. Physiol. Pt. B 66: 327–330 (1980)
145. Lange A. B.: The transfer of prostaglandin-synthesizing activity during mating in Locusta migratoria. Insect Biochem. 14: 551–556 (1984)
146. Larsen-Rapport E.: Fatty acid composition of wild type and mutant Drosophila. Insect Biochem. 16: 653–657 (1986)
147. Laufer H. et al.: Identification of a juvenile hormone-like compound in a crustacean. Science 235: 202–205 (1987)

148. Lee J. A. C. and Cossins A. R.: Temperature adaptation of biological membranes: differential homoeoviscous responses in brush-border and basolateral membranes of carp intestinal mucosa. Biochim. biophys. Acta 1026: 195–203 (1990)
149. Lerro K. A. and Prestwich G. D.: Cloning and sequencing of a cDNA for the hemolymph juvenile hormone binding protein of larval Maduca sexta. J. Biol. Chem. 265: 19800–06 (1990)
150. Lessman C. A. et al.: Detection of juvenile hormone I, II and III in adult monarch butterflies (Danaus plexippus). Insect Biochem. 19: 431–433 (1989)
151. Li Y. T., Ishikawa Y. and Li S. C.: Occurrence of ceramide-glycanase in the earthworm, Lumbricus terrestris. Biochem. biophys. Res. Commun. 149: 167–172 (1987)
152. Linko R. R., Kaitaranta J. K. and Vuorela R.: Comparison of the fatty acids in baltic herring and available plancton feed. Comp. Biochem. Physiol. Pt. B 82: 699–705 (1985)
153. Litchfield C. and Greenberg A. J.: Compositional topography of melon lipids in the Amazon river dolphin, Inia geoffrensis: implications for echolocation. Comp. Biochem. Physiol. Pt. A 63: 183–187 (1979)
154. Lockey K. H.: Lipids of the insect cuticle: Origin, composition and function (Review). Comp. Biochem. Physiol. Pt. B 89: 595–645 (1988)
155. Lockey K. H. and Oraha V. S.: Cuticular lipids of adult Locusta migratoria migratorioides (R and F), Schistocerca gregaria (Forskål) (Acrididae) and other orthopteran species. II. Hydrocarbons. Comp. Biochem. Physiol. Pt. B 95: 721–725 (1990)
156. Lockey K. H.: Insect hydrocarbon classes: Implications for chemotaxonomy. Insect Biochem. 21: 91–97 (1991)
157. Lok C. M. and Folkersma B.: Composition of wax esters and triacylglycerols in the melon and blubber fats of a young Sowerby's whale Mesoplodon bidens. Lipids 14: 872–875 (1979)
158. Lok C. M. and van der Horst D. J.: Chiral 1,2-diacylglycerols in the haemolypmph of the locust, Locusta migratoria. Biochim. biophys. Acta 618: 80–87 (1980)
159. Lovern J. A.: Neutral fats and oils. In: Florkin M. and Stotz E. H. (eds.): Comprehensive biochemistry, Vol. 6, pp. 52–82. 1965
160. Malinski E. et al.: The composition of hydrocarbons of the larvae of the khapra beetles Trogoderma granarium. Comp. Biochem. Physiol. Pt. B 84: 211–213 (1986)
161. Malinski E. et al.: Cuticular hydrocarbons of the colorado beetle Leptinotarsa decemlineata Say. Z. Naturforsch. Sect. B 41: 567–574 (1986)
162. Martin M. M. and Lieb T. J.: Patterns of fuel utilization by the thoracic muscles of adult worker ants. The use of lipid by a hymenopteran. Comp. Biochem. Physiol. Pt. B 64: 387–390 (1979)
163. Mason R. T., Chinn J. W. and Crews D.: Sex and seasonal differences in the skin lipids of garter snakes. Comp. Biochem. Physiol. Pt. B 87: 999–1003 (1987)
164. Mason R. T. et al.: Characterization of fulgorid waxes (Homoptera: Fulgoridae: Insecta). Insect Biochem. 19: 737–740 (1989)
165. Matsubara Y. et al.: Molecular cloning and nucleotide sequence of cDNAs encoding the precursors of rat long chain acyl-coenzyme A, short chain acyl-coenzyme A, and isovaleryl-coenzyme A dehydrogenase. Sequence homology of four enzymes of the acyl-CoA dehydrogenase family. J. Biol. Chem. 264: 16321–31 (1989)
166. McConville M. J. et al.: Structures of the glycoinositolphospholipids from Leishmania major. A family of novel galactofuranose-containing glycolipids. J. Biol. Chem. 265: 7385–94 (1990)
167. McConville M. J. et al.: Structure of the lipophosphoglycan from Leishmania major. J. Biol. Chem. 265: 19611–23 (1990)
168. Megias A., Municio A. M. and Olleros T.: Phospholipid requirement for fatty acid desaturase activity in microsomes of Ceratitis capitata. Comp. Biochem. Physiol. Pt. B 70: 53–56 (1981)
169. Merlie J. P. et al.: Isolation and characterization of the complementary DNA for sheep seminal vesicle prostaglandin endoperoxide synthase (cyclooxygenase). J. Biol. Chem. 263: 3550–53 (1988)
170. Meyer R., Becker W. and Klimkewitz M.: Investigations on the ketone body metabolism in Biomphalaria glabrata: Influence of starvation and of infection with Schistosoma mansoni. J. comp. Physiol. B 156: 563–571 (1986)
171. Meyer W. R. and Lanzrein B.: Degradation of juvenile hormone and methylation of juvenile hormone acid by corpora cardiaca-corpora allata of the cockroach, Nauphoeta cinerea. I. Biochemical aspects. Arch. Insect Biochem. Physiol. 10: 303–316 (1989)
172. Miceli M. V., Henderson T. O. and Myers T. C.: Alkylphosphonic acid distribution in the planorbid snail Helisoma sp. Comp. Biochem. Physiol. Pt. B 88: 603–611 (1987)
173. Milborrow B. V., Kennedy J. M. and Dollin A.: Composition of wax made by the Australian stingless bee Trigona australis. Aust. J. biol. Sci. 40: 15–25 (1987)
174. Mills G. L., Taylor D. C. and Williams J. F.: Lipid composition of the helminth parasite Taenia crassiceps. Comp. Biochem. Physiol. Pt. B 69: 553–557 (1981)
175. Minami-Ishii N. et al.: Molecular cloning and sequence analysis of the cDNA for rat mitochondrial enoyl-CoA hydratase. Structural and evolutionary relationships linked to the bifunctional enzyme of the peroxisomal β-oxidation system. Eur. J. Biochem. 185: 73–78 (1989)
176. Mita M. and Ueta N.: Phosphatidylcholine metabolism for energy production in sea urchin spermatozoa. Biochim. biophys. Acta 1047: 175–179 (1990)
177. Mohamed A. H. et al.: Primary structure of the multifunctional αsubunit protein of yeast fatty acid synthase derived from FAS2 gene sequence. J. Biol. Chem. 263: 12315–25 (1988)
178. Mommsen T. P., French C. J. and Hochachka P. W.: Sites and patterns of protein and amino acid utilization during the spawning migration of salmon. Can. J. Zool. 58: 1785–99 (1980)
179. Morell P. (ed.): Myelin, 2. ed. Plenum, New York 1984
180. Morii H. and Kaneda T.: Biosynthesis of branched-chain amino acids in subcutaneous tissues of the marine little-toothed whale, Stenella caeruloalba. Comp. Biochem. Physiol. Pt. B 71: 357–365 (1982)
181. Mortensen P. B.: C6-C10 dicarboxylic acids in liver and kidney tissue in normal, diabetic ketotic and clofibrate-treated rats. Biochim. biophys. Acta 878: 14–19 (1986)

182. Moschidis M. C.: Evidence for the presence of glycerophosphonolipids in the land snail Eobania vermiculata. Z. Naturforsch. Sect. C 41: 1124–26 (1986)
183. Muallem S. et al.: Classification of prostaglandin receptors based on coupling to signal transduction systems. Biochem. J. 263: 769–774 (1989)
184. Nakamura K., Inagaki F. and Tamai Y.: A novel ganglioside in dogfish brain. Occurrence of a trisialoganglioside with a sialic acid linked to N-acetylgalactosamin. J. Biol. Chem. 263: 9896–9900 (1988)
185. Nakhel I. C. et al.: Phospho- and phosphonolipids in the Aegean pelagic scyphomedusa Pelagia noctiluca. Biochim. biophys. Acta 958: 300–307 (1988)
186. Natarajan V. et al.: Occurrence of N-acylethanolamine phospholipids in fish brain and spinal chord. Biochim. biophys. Acta 835: 426–433 (1985)
187. Naughton J. M. and Trewhella M. A.: Presence of 5,11,14-eicosatrienoic acid in liver phosphoglycerides of the echidna (Tachypleus aculeatus). Comp. Biochem. Physiol. Pt. B 85: 139–141 (1986)
188. Needleman P. et al.: Arachidonic acid metabolism. Annual Rev. Biochem. 55: 69–102 (1986)
189. Nelson D. R., Fatland C. L. and Baker J. E.: Mass spectral analysis of epicuticular n-alkadienes in three Sitophilus weevils. Insect Biochem. 14: 435–444 (1984)
190. Nelson D. R., Nunn N. J. and Jackson L.: Reanalysis of the methylalkanes of the grasshoppers Melanoplus differentialis, M. packardi and M. sanguinipes. Insect Biochem. 14: 677–683 (1984)
191. Nelson D. R. and Carlson D. A.: Cuticular hydrocarbons of the tsetse flies Glossina morsitans morsitans, G. austeni and G. pallipes. Insect Biochem. 16: 403–416 (1986)
192. Nelson D. R. et al.: Biosynthesis of the very long-chainded methylbranched alcohols, their acetate esters and hydrocarbons in pupae of the tobacco hormworm, Manduca sexta. Insect Biochem. 20: 821–827 (1990)
193. Nevenzel J. C. and Menon N. K.: Lipids of midwater fish: family Gonostomatidae. Comp. Biochem. Physiol. Pt. B 65: 351–355 (1980)
194. Nevenzel J. C., Gibbs A. and Benson A. A.: Plasmalogens in the gill lipids of aquatic animals. Comp. Biochem. Physiol. Pt. B 82: 293–297 (1985)
195. Numa S. (ed.): Fatty acid metabolism and its regulation. New Comprehensive Biochemistry Vol 7. Elsevier, Amsterdam 1984
196. O'Brien R. W. and Breznak J. A.: Enzymes of acetate and glucose metabolism in termites. Insect Biochem. 14: 639–643 (1984)
197. Okamura N. et al.: Phylogenetic distribution of nerve glycosphingolipids. Proc. Nat. Acad. Sci. USA 82: 6779–82 (1985)
198. Oku H. et al.: Distribution of branched-chain fatty acid in the skin surface lipid of laboratory animals. Comp. Biochem. Physiol. Pt. B 96: 475–482 (1990)
199. Ostrander G. K. et al.: Isolation and characterization of four major neutral glycosphingolipids from the liver of the English sole (Parophrys vetulus). Presence of a novel branched lacto-ganglio-iso-globo hybrid structure. J. Biol. Chem. 263: 18716–25 (1988)
200. Ozyhar A. and Kochman M.: Juvenile hormone-binding protein from the hemolymph of Galleria melonella (L.) – Isolation and characterization. Eur. J. Biochem. 162: 675–682 (1987)
201. Pace-Asciak C. and Granström E. (eds.): Prostaglandins and related substances. New Comprehensive Biochemistry Vol. 5. Elsevier, Amsterdam 1983
202. Palli S. R. et al.: Juvenile hormone receptors in insect larval epidermis: Identification by photoaffinity labeling. Proc. Nat. Acad. Sci. USA 87: 796–800 (1990)
203. Palauskis J. D. and Sul H. S.: Structure of mouse fatty acid synthase mRNA. Identification of the two NADPH binding sites. Biochem. biophys. Res. Commun. 158: 690–695 (1989)
204. Peschke K. and Metzler M.: Cuticular hydrocarbons and female sex pheromones of the rove beetle, Aleochara curtula (Goeze) (Coleoptera: Staphylinidae). Insect Biochem. 17: 167–178 (1987)
205. Pettitt T. R., Rowley A. F. and Barrow S. E.: Synthesis of leukotriene B and other conjugated triene lioxygenase products by blood cells of the rainbow trout, Salmo Gairneri. Biochim. biophys. Acta 1003: 1–8 (1989)
206. Piomelli D. et al.: Formation and biological activity of 12-ketoeicosatetranoic acid in the nervous system of Aplysia. J. Biol. Chem. 263: 16591–96 (1988)
207. Poulos A. et al.: Occurrence of unusual molecular species of sphingomyelin containing 28–34-carbon polyenoic fatty acids in ram spermatozoa. Biochem. J. 248: 961–964 (1987)
208. Poulose A. J. et al.: Cloning and sequence of the cDNA for S-acyl fatty acid synthase thioesterase from the uropygial gland of mallard duck. J. biol. Chem. 260: 15953–58 (1985)
209. Prasad M. R. et al.: Evidence for multiple condensing enzymes in rat hepatic microsomes. J. biol. Chem. 261: 8213–17 (1986)
210. Previato J. O. et al.: Primary structure of the oligosaccharide chain of lipopeptidophosphoglycan of epimastigote forms of Trypanosoma cruzi. J. Biol. Chem. 265: 2518–26 (1990)
211. Putney J. W. (ed.): Phosphoinositides and receptor mechanisms. Alan R. Liss, New York 1986
212. Ragab A. et al.: The lipoxygenase pathway of arachidonic acid metabolism in reproductive tissues of the firebrat, Thermobia domestica (Thysanura). Insect Biochem. 21: 321–326 (1991)
213. Rahmann H. und Probst W.: Modulator-Funktion von Hirngangliosiden bei der Temperaturadaptation und -regulation. BIONA-Report 4: 77–83 (1986)
214. Rahmann H., Hilbig R. and Geiser F.: Brain gangliosides in monotremes, marsupials and placentals: Phylogenetic and thermoregulatory aspects. Comp. Biochem. Physiol. Pt. B 83: 151–157 (1986)
215. Rahmann H.: Gangliosides and modulation of neuronal functions. Springer, Berlin 1987
216. Ramesha C. S. and Thompson G. A.: Phospholipid molecular species alterations in Tetrahymena ciliary membranes following low-temperature acclimation. Biochemistry 21: 3618–22 (1983)
217. Rauch J. C. and Behrisch H. W.: Ketone bodies: A source of energy during hibernation. Can J. Zool. 59: 754–760 (1981)
218. Rayne R. C. and Koeppe J. K.: Relationship of hemolymph hormone-binding protein to lipophorin in Leucophaea maderae. Insect Biochem. 18: 667–673 (1988)
219. de Renobales M., Woodin T. S. and Blomquist G. J.: Drosophila melanogaster fatty acid synthetase. Characteristics and effect of protease inhibitors. Insect Biochem. 16: 887–894 (1986)

220. Rezanka T.: Very-long-chain fatty acids from the animal and plant kingdoms (Review). Progr. Lipid Res. 28: 147–187 (1989)
221. Rhoads D. E. and Kaneshiro E. S.: Fatty acids metabolism in Paramecium. Oleic acid metabolism and inhibition of polyunsaturated fatty acid synthesis by tripanol. Biochim. biophys. Acta 795: 20–29 (1984)
222. Richard D. S. et al.: Juvenile hormone bisepoxide biosynthesis in vitro by the ring gland of Drosophila melanogaster: A putative juvenile hormone in the higher diptera. Proc. Nat. Acad. Sci. USA 86: 1421–25 (1989)
223. Richard M. J., Holck J. T. and Beitz D. C.: Lipogenesis in liver and adipose tissue of the domestic cat (Felis domestica). Comp. Biochem. Physiol. Pt. B 93: 561–564 (1989)
224. Roberts D. D. and Ginsburg V.: Sulfated glycolipids and cell adhesion. Arch. Biochem. Biophys. 267: 405–415 (1988)
225. Robertson B., van Golde L. M. G. and Batenburg J. J. (eds.): Pulmonary surfactant. North-Holland, London 1985
226. Robinson B. S., Johnson D. W. and Poulos A.: Unique molecular species of phosphatitylcholine containing very-long-chain (C24-C38) polyenoic fatty acids in rat brain. Biochem. J. 265: 763–767 (1990)
227. Rotstein N. P. and Aveldaño M. I.: Synthesis of very long chain (up to 36 carbon) tetra, penta and hexaenoic fatty acids in retina. Biochem. J. 249: 191–200 (1988)
228. Rudnicka M and Jones D.: Characterization of homogenous juvenile hormone esterase from the larvae of Trichoplusia ni. Insect Biochem. 17: 373–382 (1987)
229. Ryan R. O. et al.: Biosynthesis of myristrate in an aphid: Involvement of a specific acylthioesterase. Arch. Biochem. Biophys. 213: 26–36 (1982)
230. Saavedra R. A. et al.: The myelin proteins in the shark brain are similar to the myelin proteins of the mammalian peripheral nervous system. J. mol. Evol. 29: 149–156 (1989)
231. Sams G. H., Hargis B. M. and Hargis P. S.: Isolation and characterization of a fatty acid binding protein in adipose tissue of Gallus domesticus. Comp. Biochem. Physiol. Pt. B 96: 585–590 (1990)
232. Schaub J. (ed.): Composition and physiological properties of human milk. Elsevier, Amsterdam 1985
233. Schneider P. et al.: Structure of the glycosylphosphatidylinositol membrane anchor of the Leishmania major promastigote surface protein. J. Biol. Chem. 265: 16955–64 (1990)
234. Schooley D. A. and Baker F. C.: Juvenile hormone biosynthesis. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 7, pp. 363–389. Pergamon Press, Oxford 1985
235. Schulz H.: Beta-oxidation of fatty acids (Review). Biochim. biophys. Acta 1081: 109–120 (1991)
236. Shan J. and Bailey E.: Enzymes of ketogenesis in the fat body and the thoracic muscle of the adult cockroach. Insect Biochem. 6: 251–254 (1976)
237. Shemshedini L. and Wilson T. G.: A high affinity, high molecular weight juvenile hormone binding protein in the hemolymph of Drosophila melanogaster. Insect Biochem. 18: 681–689 (1988)
238. Shemshedini L., Lanoue M. and Wilson T. G.: Evidence for a juvenile homeone receptor involved in protein synthesis in Drosophila melanogaster. J. Biol. Chem. 265: 1913–18 (1990)
239. Shetty S. et al.: Biochemical properties of pigeon milk and its effect on growth. J. comp. Physiol. B 162: 632–636 (1992)
240. Shimomura K. et al.: An unusual glucocerebroside in the crustacean nervous system. Science 220: 1392–93 (1983)
241. Sigal E. et al.: Molecular cloning and primary structure of human 15-lipoxygenase. Biochem. biophys. Res. Commun. 157: 457–464 (1988)
242. Singh B. N., Costello C. E. and Beach D. H.: Structure of glycophosphosphingolipids of Tritrichomonas foetus. A novel glycophosphosphingolipid. Arch. Biochem. Biophys. 286: 409–418 (1991)
243. Singh H. and Poulos A.: Distinct long chain and very long chain fatty acyl CoA synthetases in rat liver peroxisomes and microsomes. Arch. Biochem. Biophys. 266: 486–495 (1988)
244. Smirnova G. P., Kochetkov N. K. and Sadovskaya V. L.: Gangliosides of the starfish Aphelasterias japonica, evidence for a new linkage between two N-glycolyl-neuraminic acid residues through the hydroxy group of the glycolic acid residue. Biochim. biophys. Acta 920: 47–55 (1987)
245. Smith J. D. and Lepak N. M.: Purification and characterization of a phosphonic acid-containing glycoprotein from the cell membranes of Tetrahymena. Arch. Biochem. Biophys. 213: 565–572 (1982)
246. Smith S. et al.: Mammalian fatty acid synthetase is a structurally and functionally symmetrical dimer. Eur. J. Biochem. 152: 547–555 (1985)
247. Smith W. L.: The eicosanoids and their biochemical mechanisms of action. Biochem. J. 259: 315–324 (1989)
248. Song W. C. et al.: Identification of novel hydroxy fatty acids in the barnacle Balanus balanoides. Biochim. biophys. Acta 1047: 239–246 (1990)
249. de Sousa S. M., Krishnan K. S. and Kenkare U. W.: Phospholipid methyltransferase from Drosophila melanogaster: purification and properties. Insect Biochem. 18: 377–388 (1988)
250. Stangassinger M., Kaspar W. and Giesecke D.: The role of adipose and hepatic tissues in the lipogenesis of the dog. Comp. Biochem. Physiol. Pt. B 85: 67–69 (1986)
251. Stanley-Samuelson D. W.: 9-Eicosanoic acid: A predominantly male triacylglycerol fatty acid in the waxmoth, Galleria melonella. Comp. Biochem. Physiol. Pt. B 77: 443–445 (1983)
252. Stanley-Samuelson D. W., Loher W. and Blomquist G. J.: Biosynthesis of polyunsaturated fatty acids by the Australian field cricket, Teleogryllus commodus. Insect Biochem. 16: 387–393 (1986)
253. Stanley-Samuelson D. W.: Physiological role of prostaglandins and other eicosanoids in invertebrates. Biol. Bull. 173: 92–109 (1987)
254. Stanley-Samuelson D. W. et al.: Insect immune response to bacterial infection is mediated by eicosanoids. Proc. Nat. Acad. Sci. USA 88: 1064–68 (1990)
255. Stewart J. M. and Driedzic W. R.: Fatty acid binding proteins in teleost fish. Can. J. Zool. 66: 2671–75 (1988)
256. Stoskopf M. K. et al.: Characterization of novel glycolipids from the giant cockroach (Blaberus colosseus). J. Biol. Chem. 264: 4964–71 (1989)
257. Suzuki H. et al.: Structure and regulation of rat long-chain acyl-CoA synthetase. J. Biol. Chem. 265: 8681–85 (1990)

258. Takai T. et al.: Primary structure of chicken liver acetyl-CoA carboxylase deduced from cDNA sequence. J. Biol. Chem. 263: 2651–57 (1988)
259. Takamizawa K. et al.: Gangliosides of bovine buttermilk. J. biol. Chem. 261: 5625–30 (1986)
260. Thiede M. A., Ozols J. and Strittmatter P.: Construction and sequence of cDNA for rat liver stearyl coenzyme A desaturase. J. biol. Chem. 261: 13230–35 (1986)
261. Thiry E. and Hoffmann K. H.: Chemical composition and fatty acids of lipids in an Antarctic beetle (Hydromedion sparsutum, Perimylopidae) and an African beetle (Pachnoda marginata, Scarabaeidae). Comp. Biochem. Physiol. Pt. B 84: 387–392 (1986)
262. Timm S. L., Pereira-Netto A. D. and Oliveira M. M.: Fatty acids of Trypanosoma cruzi. Comp. Biochem. Physiol. Pt. B 71: 397–402 (1982)
263. Toong Y. C., Schooley D. A. and Baker F. C.: Isolation of insect juvenile hormone III from a plant. Nature 333: 170–171 (1988)
264. Tsukamoto Y. et al.: The architecture of the animal fatty acid synthetase complex. J. biol. Chem. 258: 15312–22 (1983)
265. Turner M. R., Leggett S. L. and Lumb R. H.: Distribution of omega-3 and omega-6 fatty acids in the ether- and ester-linked phosphoglycerides from tissues of the rainbow trout, Salmo gairdneri. Comp. Biochem. Physiol. Pt. B 94: 575–579 (1989)
266. Turunen S.: Uptake of dietary lipids; a novel pathway in Pieris brassicae. Insect Biochem. 18: 499–505 (1988)
267. Vaz A. H., Blomquist G. J. and Reitz R. C.: Characterization of the fatty acid elongation reactions involved in hydrocarbon biosynthesis in the housefly, Musca domestica L. Insect Biochem. 18: 177–184 (1988)
268. Veerkamp J. H., Peeters R. A. and Maatman R. G. H. J.: Structural and functional features of different types of cytoplasmic fatty acid-binding proteins (Review). Biochim. biophys. Acta 1081: 1–24 (1991)
269. Venkatesh K. et al.: Characterization of affinity-purified juvenile hormone esterase from the plasma of the tobacco hornworm, Manduca sexta. J. Biol. Chem. 265: 21727–32 (1990)
270. de Vleet E. S. et al.: Neutral lipid components of eleven species od Carribean sharks. Comp. Biochem. Physiol. Pt. B 79: 549–554 (1984)
271. Wakayama E. J., Dillwith J. W. and Blomquist G. J.: Characterization of prostaglandin biosynthesis in the housefly, Musca domestica (L.). Insect Biochem. 16: 903–909 (1986)
272. Wakil S. J.: Fatty acid synthase, a proficient multifunctional enzyme. Biochemistry 28: 4523–30 (1989)
273. Ward J. P., Candy D. J. and Smith S. N.: Lipid storage and changes during flight by triatomine bugs (Rhodnius prolixus and Triatoma infestans). J. Insect Physiol. 28: 527–534 (1982)
274. Weaver T. E. and Whitsett J. A.: Function and regulation of pulmonary surfactant-associated proteins (Review Article). Biochem. J. 273: 249–264 (1991)
275. Wegener G.: Brains burning fat: Different forms of energy metabolism in the CNS of insects. Naturwissenschaften 70: 43–45 (1983)
276. Wertz P. W., Colton VI S. W. and Downing D. T.: Comparison of the hydroxyacids from the epidermis and from the sebaceous glands of the horse. Comp. Biochem. Physiol. Pt. B 75: 217–220 (1983)
277. Wertz P. W., Stover P. M. and Downing D. T.: A survey of polar and nonpolar lipids from epidermis and epidermal appendages of the chicken (Gallus domesticus). Comp. Biochem. Physiol. Pt. B 84: 203–206 (1986)
278. Wertz P. W. and Downing D. T.: omega-Hydroxyacid derivatives in the epidermis of several mammalian species. Comp. Biochem. Physiol. Pt. B 93: 265–269 (1989)
279. Weske B. et al.: Glycosphingolipids in insects: Chemical structures of two variants of a glucuronic-acid-containing ceramide hexasaccharide from pupae of Calliphora vicina (Insecta: Diptera) distinguished by a N-acetylglucosamine-bound phosphoethanolamine sidechain. Eur. J. Biochem. 191: 379–388 (1990)
280. West G. C., Burns J. J. and Modaferri M.: Fatty acid composition of blubber from the four species of Bering Sea phocid seals. Can. J. Zool. 57: 189–195 (1979)
281. Wiegandt H. (ed.): Glycolipids. New Comprehensive Biochemistry Vol 10. Elsevier, Amsterdam 1985
282. Wisniewski J. R., Rudnicka M. and Kochman M.: Tissue specific juvenile hormone degradation in Galleria melonella. Insect Biochem. 16: 843–849 (1986)
283. Wisniewski J. R. and Steuernagel A.: The hemolymph of the midge Chironomus thummi contains both a high and a low molecular weight juvenile hormone-binding protein. Comp. Biochem. Physiol. Pt. B 95: 545–550 (1990)
284. Witkowki A. et al.: Structural organization of the multifunctional animal fatty-acid synthase. Eur. J. Biochem. 198: 571–579 (1991)
285. de Witt D. L. and Smith W. L.: Primary structure of prostaglandin G/H synthase from sheep vesicular gland determined from the complementary DNA sequence. Proc. Nat. Acad. Sci. USA 85: 1412–16, Correction l.c. 85: 5056 (1988)
286. Wix M. A., Wertz P. W. and Downing D. T.: Polar lipid composition of mammlian hair. Comp. Biochem. Physiol. Pt. B 86: 671–673 (1987)
287. Wodtke E.: Adaptation of biological membranes to temperature: Modifications and their mechanisms in the eurythermic carp. BIONA-Report 4: 129–138 (1986)
288. Wozniak M. et al.: Biochemical and immunochemical properties of different electrophoretic forms of juvenile hormone esterase from Trichoplusia ni (Hübner). Biochim. biophys. Acta 926: 26–39 (1987)
289. Yeung D., Nacht S. and Cover R. E.: The composition of skin surface lipids of the gerbil. Biochim. biophys. Acta 663: 524–535 (1981)
290. Zammit V. A. and Newsholme E. A.: Activities of enzymes of fat and ketone-body metabolism and effects of starvation on blood concentrations of glucose and fat fuels in teleost and elasmobranch fishes. Biochem. J. 184: 313–322 (1979)
291. Zhou B. et al.: Isolation and characterization of ceramide glycanase from the leech, Macrobdella decora. J. Biol. Chem. 264: 12272–77 (1989)
292. Zhukova N. V. and Svetashev V. I.: Non-methylene-interrupted dienoic fatty acids in molluscs from the sea of Japan. Comp. Biochem. Physiol. Pt. B 83: 643–646 (1986)
293. Zhukova N. V.: Biosynthesis of non-methylene-interrupted dienoic fatty acids from (C14-)acetate in molluscs. Biochim. biophys. Acta 878: 131–133 (1986)

16 Sterols and Steroids

Sterols and steroids, like the terpenes and carotinoids, are isoprene derivatives. The sterols are components of the membranes of all eukaryotic cells and they bind and condense the phospholipid bilayer. The outer cell membrane is particularly rich in sterols, with a molar ratio of sterols to phospholipids of 0.8–0.1, compared with the usual value of 0.1–0.3 for intracellular membranes. The parent substance of all sterols is cholesterol (Fig. 16.1). The sterols are classified with the steroid hormones and bile salts as steroids because they also have the gonan (formerly steran) four-ring system and have their biosynthetic origin in cholesterol. Steroids with hormonal functions are known from the vertebrates (sex hormones and corticosteroids) and arthropods (ecdysteroids). However, the same molecular species are also found in other metazoans, protozoans, plants, and even in primitive eukaryotes such as yeast and lower fungi; they are clearly phylogenetically very old. The critical event for the evolution of hormone systems based on steroids in the vertebrates and arthropods was not the occurrence of new steroids but the "invention" of steroid receptors [23]. The calciferols or D vitamins (Fig. 16.18) are usually discussed in the context of the steroids, although, strictly speaking, they do not belong together because the B ring of the gonan is in this case open; they arise by UV effects on sterols. Finally, several different animal groups possess pharmacologically active glycosides, saponins and alkaloids derived from cholesterol.

The common **structural element** of all steroids, the gonan, consists of four rings, A-D, with a total of 17 C atoms, of which the six linking the rings are asymmetrical (Fig. 16.1). The ring system is more-or-less flat: the H atoms or substituents of the asymmetrical C atoms which lie beneath the ring are indicated with an interrupted line and the Greek letter "α", whilst those above the plane of the ring are indicated with a continuous line and "β". The position of the H atom on C-5 also indicates the configuration of the A/B ring linkage at C-5: 5α steroids have trans-linked A and B rings, and in 5β steroids they are cis linked. The nature of the link between rings A and B determines the shape of the molecule; the ring system of the 5α steroids is flatter than that of the 5β steroids (Fig. 16.16a). Of the 64 theoretically possible gonan variants, only two are found in nature; the ring connections BC and CD always have the trans configuration. Methyl groups at C-10 and C-13 (both always in the β position) and different side-chains on C-17 (also always β) lead to the formation of other structural elements, from which an almost unlimited **variety of molecular species** is derived by the introduction of double bonds into the ring system or the side-chains and substitutions at different C atoms. Several hundred natural steroids are known to date. As can be easily imagined, there are consi-

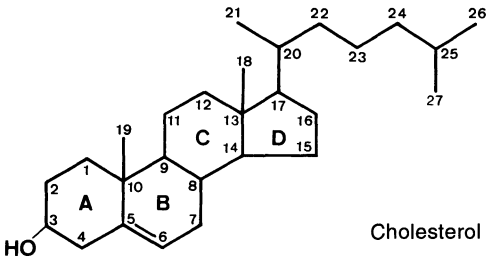


Fig. 16.1. Cholesterol (5-cholesten-3β-ol)

derable problems with nomenclature, and trivial names are often used instead of the systematic names derived, according to certain rules, from the structural elements [25, 133].

16.1 Biosynthesis of Cholesterol

The complicated **reaction series of cholesterol biosynthesis** is more easily understood when it is considered in separate parts [25]. The biosynthetic pathway begins with the production of mevalonate from acetyl-CoA and its conversion to the "active isoprenoids" isopentenyl and dimethylallyl pyrophosphate (Fig. 16.2). Condensation of these C_5 components leads to the C_{15} compound farnesyl pyrophosphate, from which arise various higher polyprenols such as dolichol, terpenes, the ubiquinone side-chain, the juvenile hormones of the insects and, finally, squalene and cholesterol. Two molecules of farnesyl pyrophosphate produce the C_{30} polyprenol squalene by head-to-head condensation. The formation of cholesterol by the cyclization of squalene requires oxygen. Thus, the sterols can have arisen on the earth only after the appearance of oxygen in the atmosphere. In the prokaryotes and several primitive eukaryotes, squalene is cyclicized, without the participation of oxygen, to the so-called hopanoids, which here assume the role of the sterols as membrane components [92]. The ciliate *Tetrahymena* also produces a hopanoid-like triterpene, tetrahymanol (Fig. 16.3). The nature of the

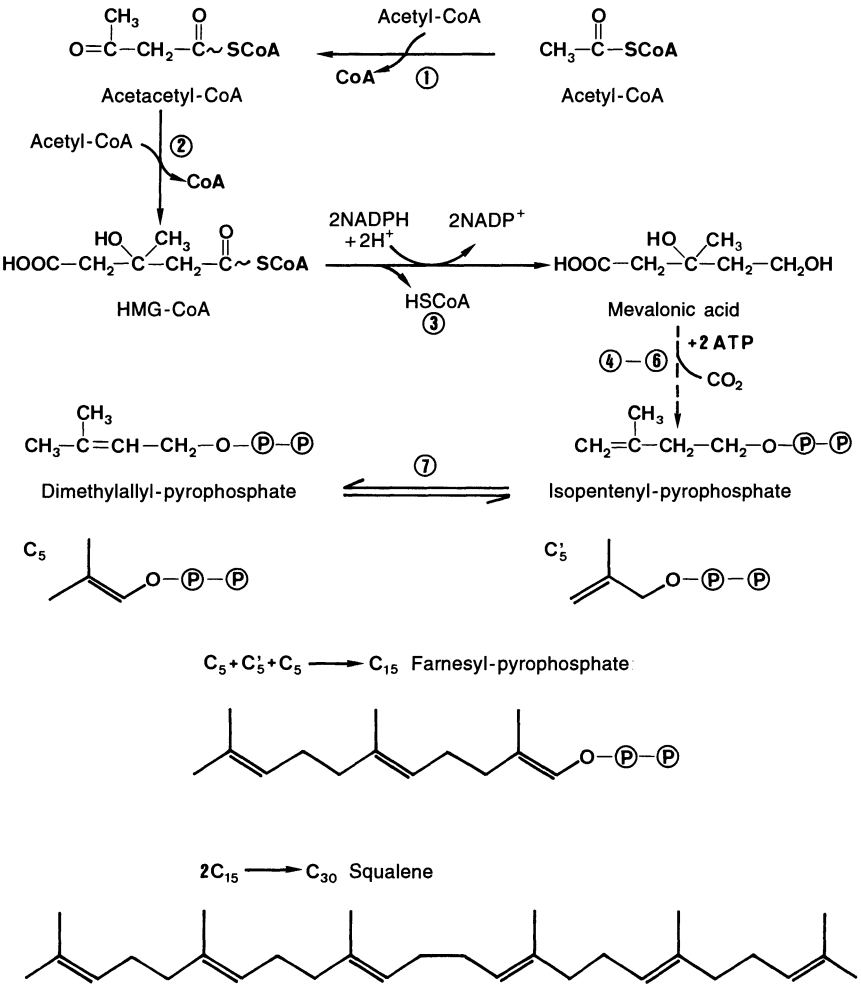


Fig. 16.2. The biosynthesis of active isoprenoids and squalenes. *1*, Acetoacetyl-CoA thiolase; *2*, 3-hydroxy-3-methylglutaryl-CoA synthetase (HMG-CoA synthetase); *3*, HMG-CoA reductase; *4*, mevalonate-5-phosphotransferase; *5*, phosphomevalonate kinase; *6*, pyrophosphomevalonate decarboxylase; *7*, isopentenylpyrophosphate isomerase

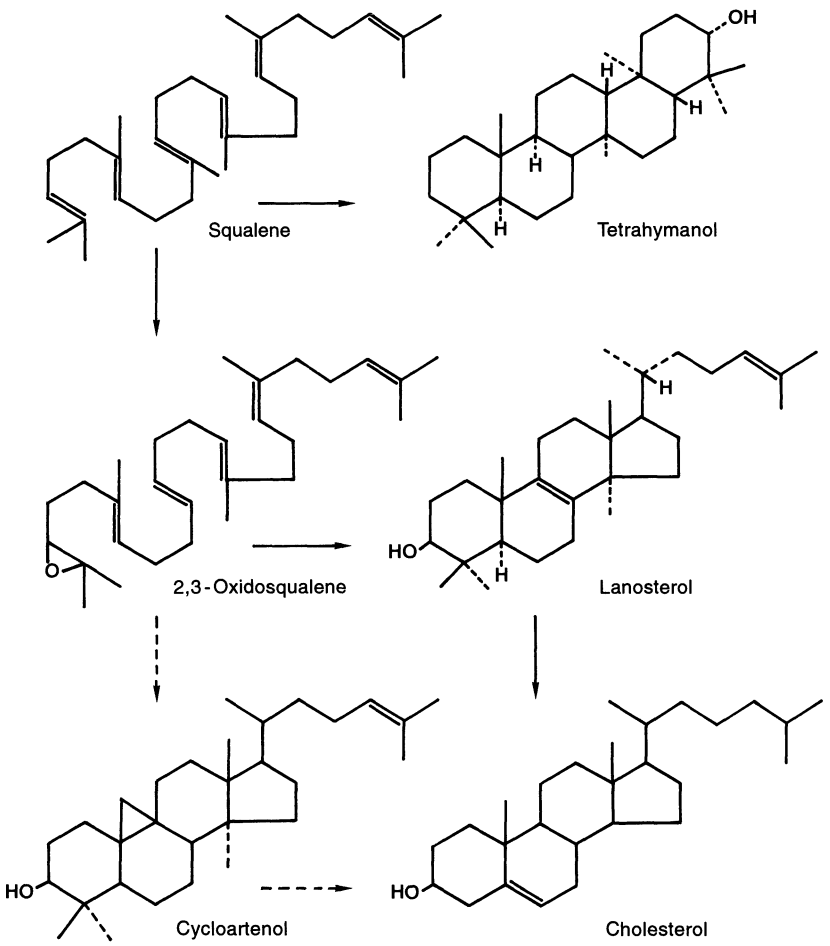


Fig. 16.3. Cyclization of squalene to tetrahymanol (anaerobic) and via 2,3-oxidosqualene to lanosterol or cycloratenol (aerobic)

primary cyclicized intermediate on the pathway from squalene to cholesterol separates organisms into two groups: the "photosynthetica", which with the exception of the photosynthetic bacteria can all synthesize cholesterol, produce primarily cycloartenol; and the "non-photosynthetica", which in so far as they are able to synthesize cholesterol, produce primarily lanosterol (Fig. 16.3). Cycloartenol instead of lanosterol is also the primary cyclization product of amoebae in the genera *Acanthamoeba* and *Naegleria* [91]. The final part of the biosynthetic pathway to cholesterol includes 19 reactions steps in which three methyl groups of lanosterol are removed and double bonds are shifted or reduced [25].

In the animal kingdom most steps of cholesterol biosynthesis and the corresponding enzymes have only been examined in detail in the higher vertebrates, in particular in the rat and the chicken. Acetyl-CoA arises primarily in the mitochondria and for cholesterol and fatty acid biosynthesis must be transported into the cytoplasm. This occurs mostly in the form of citrate or acetoacetate, and less so as acetylcarnitine. In the rat liver, the use of acetyl-CoA for cholesterol or fatty acid synthesis occurs in the ratio of approximately 1:10 [25]. There are both mitochondrial and cytoplasmic isoenzymes of acetoacetate thiolase and 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase; in the mitochondria they take part in ketone body metabolism (see Fig. 15.11, p. 579), and in the cytoplasm they participate in isoprene synthesis. The activity of the cytoplasmic HMG-CoA synthase is drastically reduced by a cholesterol-rich diet, probably as part of an important regulation mechanism of cholesterol biosynthesis.

HMG-CoA reductase (Fig. 16.2) is the best-investigated enzyme of isoprenoid synthesis in terms of comparative biochemistry. It is present not only in vertebrates of all classes but also in insects, higher plants, lower fungi and bacteria. The HMG-CoA reductase of the hamster agrees in 80 % of its amino acids with that of the frog *Xenopus laevis*, and even shows 61–76 % sequence agreement with the enzyme of the sea

urchin *Strongylocentrotus purpuratus*, although only in the terminal regions (amino acids 1–340 and 537–932); the linker region shows hardly any similarity to the mammalian enzyme. The human and yeast enzymes have about 50 % amino acids in common [4, 20, 129]. In the mammals, the highest activities are found in the liver and gut, with significantly less in all other tissues. The enzyme activity shows a marked diurnal rhythm with a maximum during the night, and the enzyme can be inactivated by phosphorylation. The proportion of the active non-phosphorylated form, which in rat liver is normally about 10–20 %, is increased by insulin or glucagon and reduced by the supply of cholesterol. Synthesis of the enzyme also responds to feeding of cholesterol [25]. The condensation of the isoprenoid components is catalysed by several different enzymes, at least in the mammals, and different end products are produced, e.g. dolichol or the side-chain of ubiquinone. The only enzyme to have been isolated, the 80-kDa enzyme which produces farnesyl pyrophosphate, was obtained from the liver of the chicken and several mammals. Condensation of two molecules of farnesyl pyrophosphate gives rise initially to pre-squalene pyrophosphate, which by cleavage of the pyrophosphate and reduction by NADPH is converted to squalene. These two reaction steps may be catalysed by the same microsomal enzyme. Squalene epoxidase, squalene cyclase and the enzymes of the further reaction steps from lanosterol to cholesterol are bound to the membrane of the endoplasmic reticulum [25].

16.2 Sterols of Different Animal Groups

Cholesterol is the predominant sterol of the vertebrates, tunicates, annelids, arthropods and molluscs, but in most cases other sterols are also present. Marine animals, such as the Porifera, Cnidaria and echinoderms, always contain an extraordinarily large variety of sterols. Of the numerous sterols that can be identified in individuals of these animal groups by use of present-day refined analytical techniques, most occur only in amounts of one or a few percent. Marine invertebrates typically possess 10–30 types of sterols; no fewer than 74 different sterols were identified in the sponge *Axinella cannabina* [51]. The marine sterols are indeed a particularly stimulating research field for biochemists! Amongst the marine sterols are many which are not found in terrestrial organisms. For example, many plant sterols have alkyl substituents on C-24, whereas C-22 and C-23 substitutions are only found in marine organisms. Sterols with cyclopropane-substituted side-chains, with reduced A rings (A-norsterols) or without the 19-methyl group (19-norsterols), and polyhydroxy sterols or epidioxy sterols occur only amongst the marine sterols (Fig. 16.9). Given that many other animals, including the mammals, survive with only cholesterol as the predominant sterol component of their cell membranes, the variety of sterols in the marine invertebrates is probably not required to satisfy specific needs as the result of adaptive evolutionary processes. It appears much more likely, as in the case of the fatty acids, that nature is "experimenting" with different structural variants in these primitive organisms and will "chose" the most suitable. The predominance in most animals of cholesterol, which is the starting point for the biosynthesis of the other sterols, is quite compatible with this concept. The presence of a particular sterol in an organism may be the result of de novo synthesis via cholesterol or of uptake in the diet and eventual metabolic modification; it may also be a product of symbiontic algae, fungi or bacteria [37]. Numerous tracer experiments with acetate or mevalonate have been carried out to determine which animals are capable of de novo cholesterol synthesis; unfortunately, the results are in some cases ambiguous. Whilst the sterol components in the diet of terrestrial animals are in general easy to define, the availability of sterols in the diet of marine animals is almost impossible to estimate [37].

Apparently all, or almost all, animals can synthesize isoprenoid components and polyisoprenoids such as dolichol or the side-chains of ubiquinone or menaquinone (vitamin K group); however, many appear unable to produce cholesterol and very often also squalene. The **ability to synthesize cholesterol** has been clearly demonstrated in vertebrates, tunicates, annelids, nemertines and echinoderms. It is also quite certain that the arthropods lack this capacity. Among the molluscs, the gastropods and amphineurans have been shown to carry out limited cholesterol biosynthesis, whereas for the bivalves and cephalopods there are more negative than positive results. The Porifera and Cnidaria appear unable to synthesize any sterols, although the situation is not yet completely clear; in this case, it is difficult to exclude fully the participation of the ever-present symbionts [37, 56].

Probably by means of absorption, catabolic, and excretory processes, all animals are in a posi-

tion to select particular sterols from those available in the diet and introduce them into the sterol spectrum of their tissues. It also appears that all animals, including those unable to synthesize sterols de novo, can **metabolically modify those sterols taken up in the diet**. Very common, though not universal, is the removal of the alkyl substituents on C-24 (24-dealkylation) and the reduction or creation of double bonds. Many sterols can be autoxidized, and the accumulation of autoxidation products is common in marine invertebrates, where storage of excretory products is more usual than in terrestrial animals. Sterol esters can be produced by acyl transfer to the 3β-OH group present in almost all sterols, and cleaved again by specific hydrolases. These are storage and transport forms of sterols and are more usual in intracellular fat droplets than in membranes. Sulphate esters are also found at low levels in probably all animals; only in the case of the echinoderms are most of the sterols sulphated. Sterol glycosides and sulphated steroid saponins are also more-or-less restricted to this phylum. However, various sterol conjugates may be formed by detoxification processes prior to excretion.

16.2.1 Sterols of the Vertebrates

The terrestrial vertebrates have almost exclusively cholesterol; up to one-third of the total sterols in marine fish may be made up of other sterols that are quite clearly derived from sterols in the diet. Marine teleosts also have **bioactive steroid glycosides**, e.g. the six different "pavoninins" of *Pardachirus marmoratus*; these are steroids with O-β-glycosidic-bound N-acetylglucosamine that have a deterrent effect on sharks [115].

The human **cholesterol requirement** is usually covered in equal proportions by the diet and by de novo synthesis, which mainly occurs in the liver. The cholesterol absorbed from the gut lumen with the lipid micelles is incorporated by the gut cells into chylomicrons which enter the lymph and, therefore, eventually the blood circulation. As a result of the exchange and conversion processes between the different types of plasma lipoproteins, the then esterified cholesterol is mostly found in the blood bound to LDL (low-density lipoproteins). The esterification of cholesterol in the circulatory system takes place on HDL (high-density lipoproteins) by the action of the plasma enzyme lecithin:cholesterol acyltransferase. Similar enzymes have been found in the blood plasma of the trout *Salmo alpinus* and in frog blood. The cholesterol transported in the blood provides the starting material for the biosynthesis of steroid hormones in the gonads and the adrenal cortex. The most important degradation products of cholesterol are the bile salts; about 1 g cholesterol is used daily in the synthesis of bile salts in the human liver. Because of the limited water solubility of cholesterol and its precursors, they must be bound to **carrier proteins** for metabolism and transport processes in the liver cells. The squalene-binding protein of rat liver (SCP-1) is essential for the biosynthetic pathway from squalene to lanosterol, and the sterol-binding protein (SCP-2) is required for the pathway from lanosterol to cholesterol and for intracellular cholesterol transport. Together with for example, the "Z protein" and the fatty acid-binding proteins, these two proteins belong to a family of closely related 12- to 16-kDa proteins, representatives of which are found not only in other rat tissues but also in shark and trout liver [5]. A subsection of the 122-amino-acid sequence of SCP-2 from rat liver shows significant homology to the variable domain of the heavy immunoglobulin chain [90].

16.2.2 Sterols of the Insects

Insects, like all other arthropods, are unable to synthesize cholesterol or squalene de novo, and require **sterols as essential nutrients**. The symbiontic microorganisms present in almost all insects participate to a high degree in the satisfaction of the host's sterol requirements, and some insect species require no further supply of exogenous sterols. Cholesterol is, in most cases, the optimal source, although *Drosophila pachea* and *Xyleborus ferrugineus* are reported to require Δ^7-sterols. According to the results of many feeding experiments, sterol requirements can almost always be met by many different sterols, although the demands of individual species can differ markedly [112]. Herbivorous insects in particular have a versatile capacity for converting plant C_{28} and C_{29} sterols into cholesterol, i.e. for removing the alkyl substituent on C-24. Other insects, e.g. the hide beetle *Dermestes maculatus*, the housefly *Musca domestica* and certain predatory Hemiptera, lack the capacity for 24-dealkylation. The silkworm *Bombyx mori* requires sterols with the cholestan side-chain; however, these can be variously ring desaturated. Sterols in the insects function as membrane components and as precursors of the moulting hormones, and these different

functions place different requirements on the species. Thus, a housefly, despite its restricted capacity for sterol conversion, must take up less than 0.5 % of the total sterol in the diet as cholesterol, i.e. less than 0.1 μg; the remaining sterol requirement can be statisfied by other sterols [112].

The **metabolic pathways leading from the phytosterols to cholesterol** have been defined in various insects by tracer experiments and the identification of intermediates. In contrast to previous assumptions, there are marked differences between individual insect species and groups [25, 112]. The removal of the 24-alkyl group always requires several oxidative and reductive reaction steps. The metabolic pathway of sitosterol has been examined in more than a dozen insect species (Fig. 16.4a). Oxidation of the 24-ethyl group to a 24-ethylene group produces fucosterol, which is converted to the 24,28-epoxide by the addition of oxygen. The next identifiable intermediate is desmosterol, which can be reduced to cholesterol. The metabolism of campesterol proceeds in a similar fashion via 24-methylene cholesterol and its 24,28-epoxide to desmosterol and cholesterol (Fig. 16.4a). Stigmasterol is converted via cholesta-5,22,24-trien-3β-ol to desmosterol. In the mealworm *Tribolium confusum*, the pathway proceeds via cholesta-5,7,24-trien-3β-ol to 7-dehydrocholesterol, which comprises up to 70 % of the total sterol of this species; the desaturation of C-7 can occur before or after reduction of the double bond on C-22 (Fig. 16.4b). 7-Dehydrocholesterol, as well as 7-dehydrocampesterol and 7-dehydrositosterol predominate in the prothorax gland of *Bombyx mori*. All haemolymph sterols are apparently oxidized in this organ to the 7-dehydro derivatives [101]. Ergosterols can also be converted to cholesterol by many insects; however, the common cockroach *Blatella germanica* appears incapable of reducing the C-22 double bond and produces 22-dehydrocholesterol [112].

The Mexican bean beetle *Epilachna varivestis* mainly has the saturated sterols cholestanol, campestanol and stigmastanol, in addition to more than 10 % lathosterol (cholest-7-en-3β-ol). Tracer experiments show that the beetle is unable to dealkylate phytosterols to cholesterol but produces instead stanols, which are dealkylated to cholestanol and in part oxidized to lathosterol (Fig. 16.4c) [112, 113]. The ladybird *Coccinella septempunctata*, which feeds off aphids but nevertheless belongs to the same family as the bean beetle (Coccinellidae), appears unable to modify the sterols, which mainly consist of cholesterol, in its diet [112]. There are also phytophagous insect species that lack the ability to convert plant sterols to cholesterol. This is true, for example, for *Oncopeltus fasciatus* and other Hemiptera, which have adapted to the absence of cholesterol to such an extent that the typical moulting hormone of these animals, makisterone A, carries a methyl group on C-24 (Fig. 16.13). The khapra beetle *Trogoderma granarium* is also unable to metabolize phytosterols. However, both of these phytophagous species can selectively absorb the cholesterol and campesterol in their diet. The honeybee *Apis mellifera* is apparently just as inefficient in converting phytosterols to cholesterol, but can accumulate the 24-methyl cholesterol present in, for example, pollen and pass it on to the young [112]. The leaf-cutting ants *Atta cephalotes* cannot dealkylate the ergosterol predominating in their fungus diet. Hence, their tissues contain in particular the related compound 24-methylene-cholesta-5,7,24(28)-trien-3β-ol as well as 22-dihydroergosterol and ergosterol (Fig. 16.4d) [99].

As is illustrated by the processes described for converting phytosterols into cholesterol, insects can carry out numerous reactions of sterol metabolism but these are not realized by all species [21, 25, 112]. This includes the reduction of double bonds, in particular at C-5, C-7, C-22 and C-24, and the reduction of keto groups. Amongst the oxidative processes are the introduction of double bonds and substitution with hydroxy or oxo groups, which are important for the formation and activation of the ecdysteroids, the biosynthesis of defence substances, and preparation for the production of glycosides and other conjugates. Cleavage of akyl substituents on C-24, which arise by oxidation and reduction, is widely found in the insects. Shortening of the cholesterol side-chain to C_{21}, C_{19} and C_{18} steroids occurs during the production of defence substances by the beetle families Dytiscidae and Silphidae. The formation of the C_{21} steroid pregn-5-en-3β,20β-diol (Fig. 16.5a) from cholesterol has also been observed in *Manduca sexta* [118]. The production and hydrolysis of sterol esters has been detected by tracer experiments in many insect species. The microsomal acyl-Coa:cholesterol acyl transferase from the gut and fat body of the pre-pupa of the moth *Heliotis zea* has been examined in some detail [7]. Adults of the common cockroach *Periplaneta americana* have 29 % esterified cholesterol, and the proportion is even higher in insect eggs. Up to 41 % of the esters in the eggs of the housefly *Musca domestica* are completely hydro-

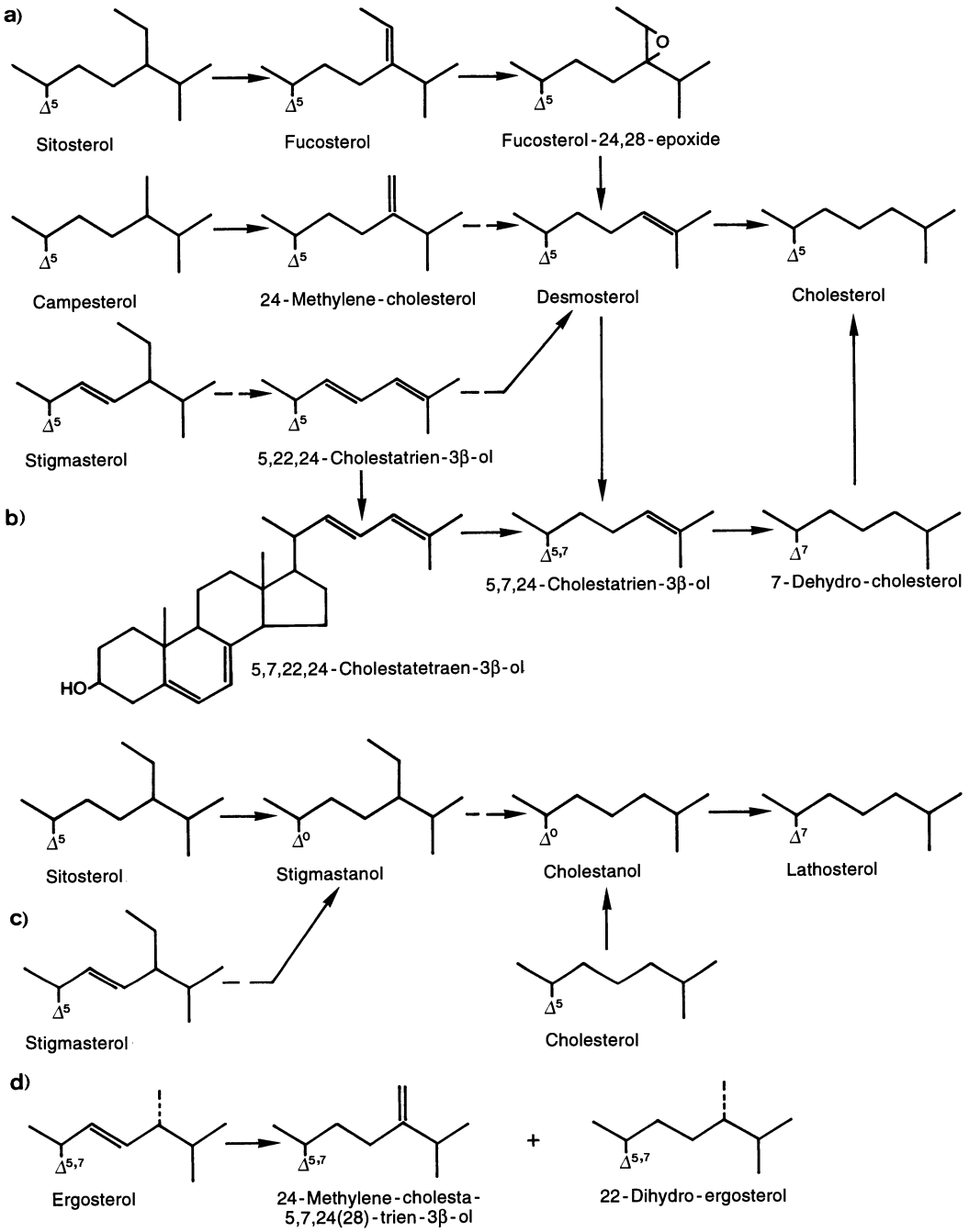


Fig. 16.4a–d. The metabolism of phytosterols in insects [99, 112]. **a** The pathways found in *Bombyx mori, Manduca sexta* and many other insects. **b** The corn beetle *Tribolium confusum* has a different pathway of stigmasterol metabolism that leads via 5,7,24-cholestatrien-3β-ol to 7-dehydrocholesterol. **c** The Mexican bean beetle *Epilachna varivestis* produces mainly saturated sterols such as stigmastanol and cholestanol. **d** The leaf-cutting ant *Atta cephalotes* is unable to 24-dealkylate the ergosterol in its diet

lyzed during embryo development and are thus probably a storage form [112].

For a long time **C_{18}, C_{19} and C_{21} steroid hormones** were considered to be exclusive to the vertebrates. It was therefore a surprise when, in 1966, Schildknecht found large amounts of the corticosteroid cortexon in the secretion from the prothorax defence glands of the great water beetle *Dytiscus marginalis*. Since then the defence secretions of other water beetles (Dytiscidae) have

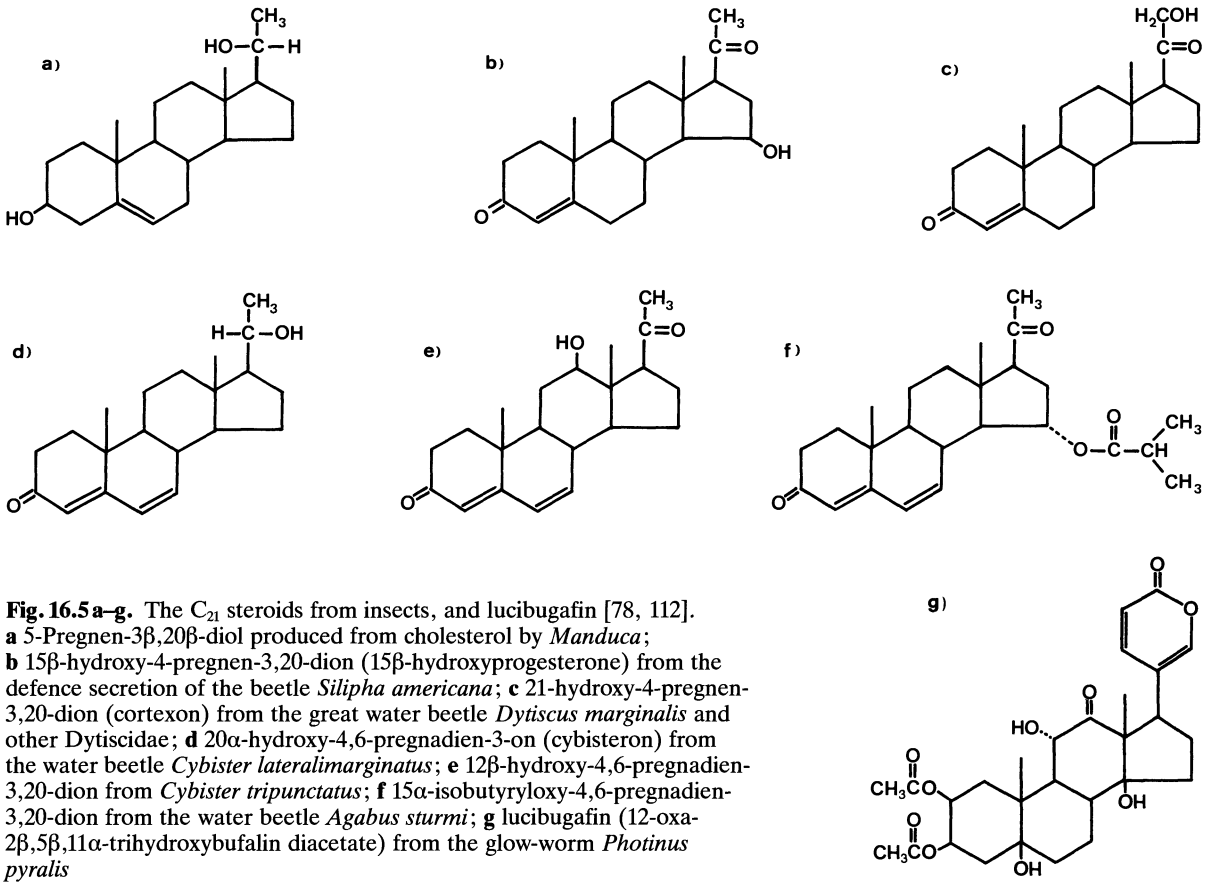


Fig. 16.5 a–g. The C_{21} steroids from insects, and lucibugafin [78, 112]. **a** 5-Pregnen-3β,20β-diol produced from cholesterol by *Manduca*; **b** 15β-hydroxy-4-pregnen-3,20-dion (15β-hydroxyprogesterone) from the defence secretion of the beetle *Silipha americana*; **c** 21-hydroxy-4-pregnen-3,20-dion (cortexon) from the great water beetle *Dytiscus marginalis* and other Dytiscidae; **d** 20α-hydroxy-4,6-pregnadien-3-on (cybisteron) from the water beetle *Cybister lateralimarginatus*; **e** 12β-hydroxy-4,6-pregnadien-3,20-dion from *Cybister tripunctatus*; **f** 15α-isobutyryloxy-4,6-pregnadien-3,20-dion from the water beetle *Agabus sturmi*; **g** lucibugafin (12-oxa-2β,5β,11α-trihydroxybufalin diacetate) from the glow-worm *Photinus pyralis*

been found to contain different C_{21} steroids that are identical or closely related to the vertebrate corticosteroids (Fig. 16.5 c–f). In addition, there are lower amounts of C_{18} and C_{19} steroids that correspond to the sex hormones of the vertebrates, e.g. testosterone, estradiol-17β and estrone. The amount of corticosteroids present in a single beetle may be as much as 1 mg, which exceeds manyfold the hormone content of the mammalian adrenal gland. Tracer experiments have demonstrated the synthesis of these steroids from cholesterol [21, 112]. Several of the C_{21} steroids have numbing effects on fish and deterrent effects on amphibians. Whether they have in addition some regulatory functions in the insects themselves is not known [112]. Testosterone, progesterone and other C_{19} and C_{21} steroids have been identified in numerous insect species. The synthesis of such steroids from radiolabelled cholesterol has so far been shown only in the coleopteran *Acilius sulcatus* (Dytiscidae) and the lepidopteran *Manduca sexta* [10, 114].

15β-Hydroxyprogesterone (Fig. 16.5b) and six other hydroxylated 20-ketopregnans have been found in beetles of the genus *Silpha* (Silphidae); they form part of an odorous defence secretion that is produced in an associated gland of the hindgut and is released from the anus in times of danger [78]. Seven pyrone-substituted steroids known in the glow-worm *Photinus pyralis* (Fig. 16.5 g) are referred to as **lucibufagins** and apparently have deterrent effects on birds. They are probably produced by the animal itself from as yet unknown precursors [112]. In contrast, the **cardiac glycosides** of the cardenolid type, which are stored as defence substances in many insects, are probably mostly obtained from plants in the diet. These are C_{23} steroids with a characteristic 14β-hydroxyl group and a-α,β unsaturated γ-lactone ring (see Fig. 19.1a, p. 717). However, some beetles of the family Chrysomelidae contain cardenolids, although their host plants, mint and rosemary, do not produce cardiac glycosides. In one of these species, *Chrysolina coerulans*, it has been clearly shown by tracer experiments that they are able to convert cholesterol to cardenolids, in the course of which a C_{21} intermediate appears, as in plants. This is apparently a case of convergent development of the same metabolic pathway [87, 89]. The species *Chrysolina hyperici*, which lives on the cardenolid-free *Hypericum perforatum*, contains stigmastan glycosides instead of cardenolids [24].

16.2.3 Sterols of Tunicates, Annelids, Crustaceans and Molluscs

Ascidians, polychaetes, crustaceans, cephalopods and gastropods contain mainly cholesterol [8, 64, 100, 111]; other sterols, usually present at low concentrations, probably originate in the diet. The 24-alkyl sterols in the diet of marine crustaceans are dealkylated to cholesterol; the presumed intermediates, 24-ethylidencholest-5-en-3β-ol and desmosterol (cholest-5,24-dien-3β-ol; Fig. 16.4a), are found in trace amounts [52]. Steroids of the vertebrate hormone type have also been identified in crustaceans [29, 84].

Cholesterol constitutes only about 50 % of the sterols in marine **mussels**; the remainder is a complex spectrum of C_{26} to C_{29} sterols, including 24-methyl and 24-ethyl sterols and, as a special feature, different $\Delta^{5,7}$-sterols and 4α-methyl sterols (Fig. 16.6a), which are a specific product of the bivalves [53, 111, 116]. Desaturation of Δ^5- to $\Delta^{5,7}$-sterols has also been detected in certain protozoans and insects; conversely, the snail *Charonia tritonis*, which feeds on starfish of the species *Acanthaster plancii*, can convert Δ^7-sterols via $\Delta^{5,7}$- to Δ^5-sterols such as cholesterol [37]. In contrast to all other molluscs, the **chitons** (Amphineura) contain predominantly Δ^7-sterols. The chi-

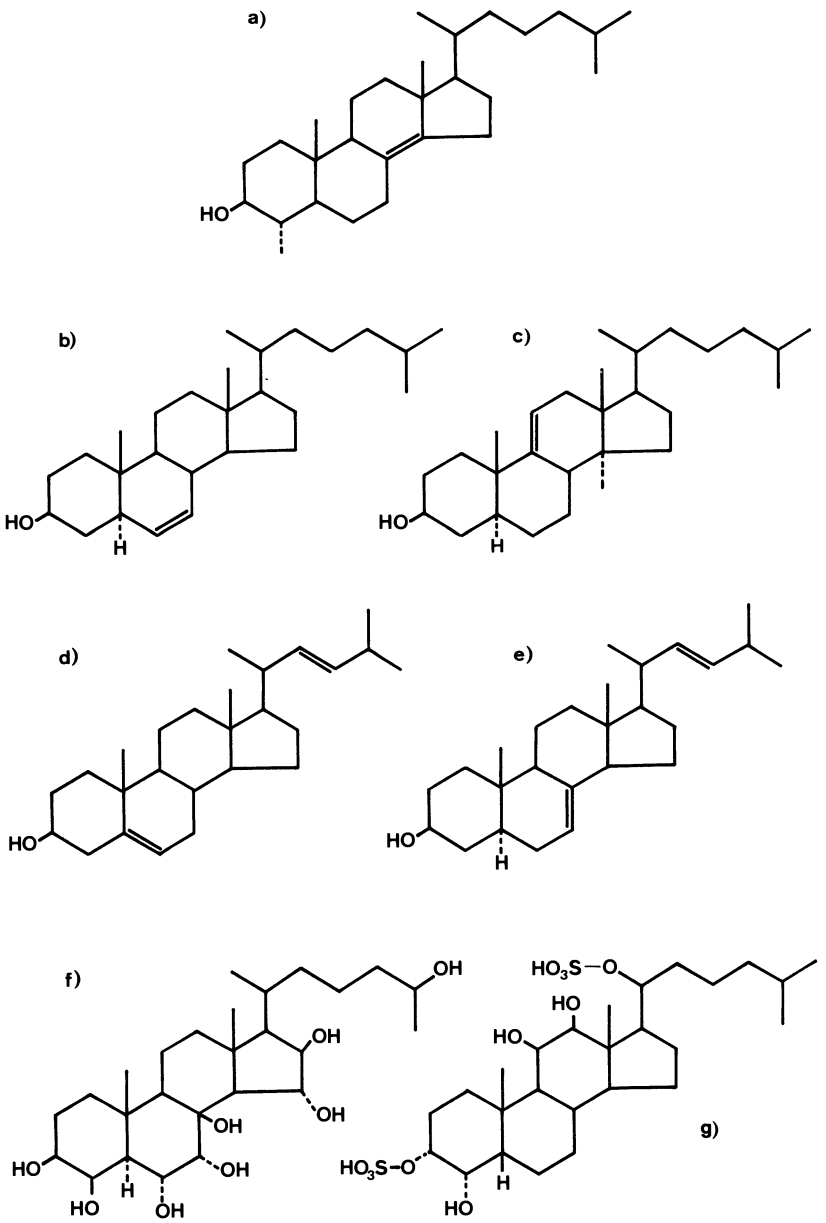


Fig. 16.6a–g. Unusual sterols from **a** molluscs and **b–g** echinoderms. **a** 4α-Methyl-cholest-8(14)-en-3β-ol from the oyster *Crassostrea virginica* [116]; **b** 5α-cholest-7-en-3β-ol is widespread in the asteroids and holothurians [25]; **c** 14α-methyl-5α-cholest-9(11)-en-3β-ol from the holothurian *Psolus fabricii* [25]; **d** 24-norcholesta-5,22-dien-3β-ol is common in the ophiuridans and crinoids [25]; **e** asterosterol is found in many asteroids and holothurians [25]; **f** 5α-cholestan-3β,4β,6α,7α,8,15α,16β,26-octol from the starfish *Protoreaster nodosus* [97]; **g** 5β-cholestan-3α,4α,11β,12β,21-pentol-3,21-bussulphate from the ophiuridan *Ophioderma longicaudum* [98]

ton species *Lioluphura japonica* incorporates ^{14}C-acetate not into cholesterol but into cholest-7-en-3β-ol; the Δ^7-sterol is also produced from labelled cholesterol. In the same way, the 24-alkylated Δ^7-sterols, present as subsidiary components, arise from Δ^5-sterols in the diet [117]. As in other invertebrates, the molluscs also have steroids of the vertebrate hormone type. The production of testosterone from cholesterol has been demonstrated in tracer experiments on the gonads of the **cephalopod** *Sepia officinalis* [37, 95]. Progesterone, testosterone and oestrone have been identified in the free-living nematode *Turbatrix aceti* [69].

16.2.4 Sterols of the Echinoderms

The principal sterols of the echinoderms are C_{27} sterols; Δ^5-sterols, such as cholesterol, occur in the echinoids, ophiurids and crinoids, and Δ^7-sterols, such as 5α-cholest-7-en-3β-ol (Fig. 16.6b), are present in the asteroids and holothurians. The holothurians also have 5α-cholestan derivatives. In particular, the fraction of saturated and Δ^5-unsaturated C_{27} sterols is often more than 50 % sulphate esterified, and there are large differences between the spectra of the free and the sulphated sterols. For example, in the sea cucumber *Psolus fabricii* 14α-methyl-5α-cholest-9(11)-en-3β-ol (Fig. 16.6c) makes up 60–70 % of the free sterols but only 10 % of the sulphate esters; the latter are mainly 5α-cholestan-3β-ol [38]. Sterol sulphate esters have also been found in trace amounts in other invertebrates and the mammals; only in the echinoderms are they found in large amounts. Their biological significance is, however, still something of a puzzle. In addition to the C_{27} sterols, the echinoderms also have C_{26} sterols, such as 24-norcholesta-5,22-dien-3β-ol and asterosterol (Fig. 16.6d–e), as well as more conventional C_{28} and C_{29} sterols. As these animals can synthesize C_{27} but not C_{28} or C_{29} sterols, the latter undoubtedly originate from the diet. The sterol spectrum of the echinoderms, like that of the other marine invertebrates, can be extremely complex. For example, 29 sterols have been identified in the holothurian *Stichopus japonicus*; these include 68 % sulphate esters and also β-xylosides and sterol ethers [60].

The **ability to synthesize sterols** has been shown by tracer experiments in representatives of all five echinoderm classes. *Echinus esculentus* incorporates mevalonic acid into squalene, lanosterol and desmosterol; in experiments with *Asterias rubens* and radioactive mevalonate or cholesterol, 5α-cholest-7-en-3β-ol was most heavily labelled. The conversion of Δ^5- to Δ^7-sterols can occur by two different pathways; the pathway via the 4-en-3-one compound and 5α-stanol has been demonstrated in tracer experiments; the pathway via a $\Delta^{5,7}$-sterol is, as yet, just a probability [37]. Steroids corresponding to the vertebrate steroid hormones have also been detected in the echinoderms. These substances are apparently endogenous in origin; for example, extracts from the starfish *Asterias rubens* converted cholesterol and other sterols in vitro to progesterone, testosterone and 11-deoxycorticosterone. Both of the pathways for steroid biosynthesis recognized in the vertebrates, the Δ^4 and Δ^5 pathways (p. 639), are also active in *Asterias* [124]. *Asterias* and other echinoderms contain estrogen but, as for other invertebrates, they are not able to synthesize this steroid [37].

Numerous pharmacologically active **steroid glycosides and steroid saponins** are known from the echinoderms. The steroid glycosides of the starfish are termed (astero-)saponins because they produce soap-like solutions at very low concentrations; they have haemolytic and membrane-toxic detergent activity. Polyhydroxy sterols in the starfish are mainly found in the saponins but have also been reported as free sterols, e.g. 5α-cholestan-hexol, -heptol and -octol neurotoxins (Fig. 16.6f) in the Pacific asteroid *Protoreaster nodosus*, and 5β-cholestan-pentol bissulphate (Fig. 16.6g) in the ophiurid *Ophioderma longicaudum* [97, 98]. The asterosaponins are always mixtures which vary in composition between the individual organs of an animal [36]. Usually they are $\Delta^{9(11)}$-unsaturated, 3β,6α-hydroxylated 23-oxosteroids with a carbohydrate chain on the 6-hydroxyl and a sulphate residue on the 3-hydroxyl group. 24-O-glycosylated asterosaponins are also found [121]. Hydrolysis of the asterosaponins almost always yields asterone (Fig. 16.7a), which is now accepted as an isolation artefact. The neutral aglycone is thornasterol A (Fig. 16.7b), and there are apparently others [14]. The carbohydrate chain on the 6-hydroxyl group consists of five to six sugars, in particular D-quinovose as well as D-fucose, D-xylose and D-galactose [81]. Species-specific mixtures of highly toxic steroid saponins are found in almost all holothurians; they are stored in the skin glands and the so-called Cuvier organs, sticky tubes that at the approach of danger can be ejected from the body cavity. To date, at least 19 different aglyco-

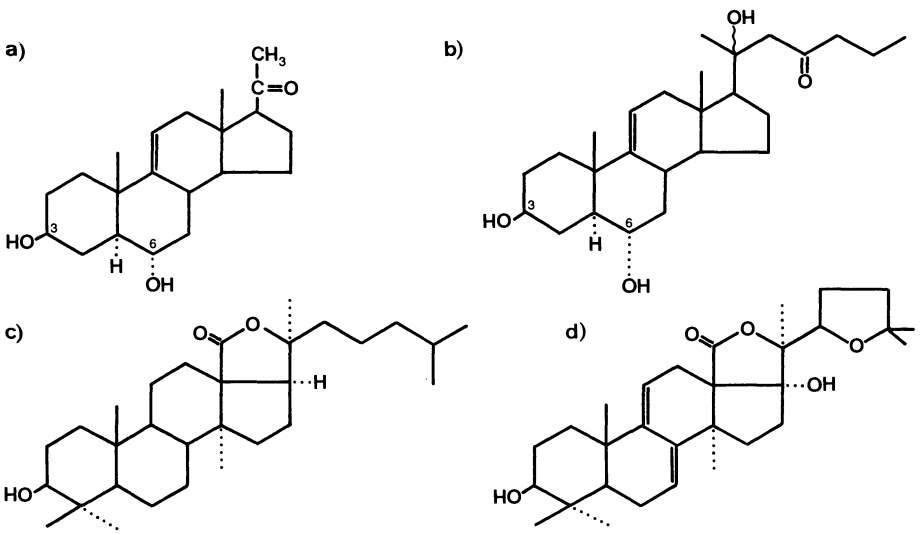


Fig. 16.7a–d. Asterosaponins **(a, b)** and holothurinogenins **(c, d)** [14, 43]. **a** Asterone; **b** thornasterol A; **c** holostanol (basic element of the holothurinogenins); **d** 22,25-epoxy-holothurinogenin

nes have been identified, all of which can be derived from the same basic component, holostanol (Fig. 16.7 c–d) [43]. The carbohydrate chain is bound to the 3-hydroxyl group and, for example, in the holothurin A of *Actinopyga agassizi* consists of one molecule each of D-glucose, D-xylose, D-quinovose, 3-O-methylglucose and sulphate.

16.2.5 Sterols of the Cnidaria

The cnidarians possess complex mixtures of Δ^5-sterols, usually including large quantities of cholesterol. As they are apparently incapable of the de novo biosynthesis of sterols, the cholesterol must arise either by specific accumulation from the diet or by the dealkylation of 24-alkyl sterols [25, 37, 80]. The Alcyonaria have large amounts of 24-methyl and 24-methylene sterols, as well as some with 28-CH_2OH and 28-CHO groups that may be intermediates of 24-dealkylation. The gorgosterol discovered in 1943 in *Plexaura flexuosa* carries a cyclopropane-substituted side chain (Fig. 16.8a). It is the predominant sterol in, for example, many soft corals (Gorgonaria), but only occurs in those cnidarians that host symbiontic dinoflagellates (Zooxanthella). However, isolated dinoflagellates do not produce gorgosterol but do produce 4-methyl sterols; it is possible that the 4α-methyl-5α-gorgosterol is a precursor of gorgosterol [2, 25). Various polyhydroxy sterols have been isolated from Octocorallia, Alcyonaria, Gorgonaria and other cnidarians; examples include hippuring I (Fig. 16.8b) from *Isis hippuris*, and 24-methylcholestan-pentol with the rather unusual 22-hydroxyl group (Fig. 16.8 c) from *Asterospicularia randalli* [25, 73]. 5α,8α-Epidioxy sterols (Fig. 16.8 d), known from other invertebrates, are found in the sea anemone *Metridium senile* [33]. Steroids with less than 27 C atoms are occasionally found in the cnidarians, e.g. derivatives of 20-epicholanic acid (Fig. 16.8 e) occur in the sea-pen *Ptilosarcus gurneyi*, 3-ketopregnans (Fig. 16.8 f) are found in various soft corals [25], and a polyhydroxy androstan is present in *Sarcophyton glaucum*. Five pregnene derivatives with arabinose or xylose have been described in the soft coral *Alcyonium* sp. [18].

16.2.6 Sterols of the Porifera

The sponges, the most primitive of the multicellular animals, have the most complicated sterol spectra. There are many sterols that have been described only in the Porifera. Some sponges contain mainly sterols with unusual ring systems, whereas others have conventional ring structures but unusual side-chains. The variety of sponge sterols most probably arises by metabolic modification of sterols in the diet, as the porifera themselves apparently cannot synthesize sterols de novo; however, only a very few tracer experiments appear to have been carried out with this phylum [25, 37].

Many sponges contain predominantly Δ^5-sterols with 26–30 C atoms [26]. Less frequently

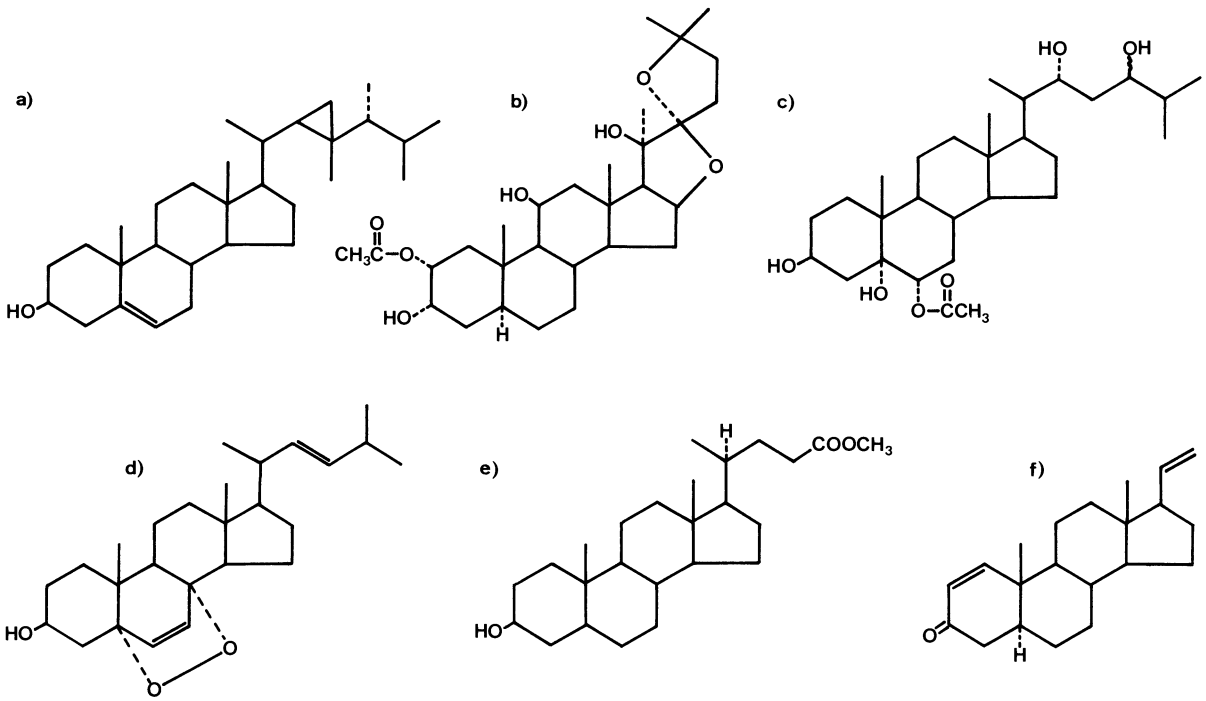


Fig. 16.8a–f. Sterols of the Cnidaria. **a** Gorgosterol predominates in many soft corals (Gorgonaria) [25]; **b** hippurin I from the gorgonian *Isis hipuris* [25]; **c** 24-methylcholestan-3β,5,6β,22,24,pentol-6-acetate from the gorgonian *Asterospicularia randalli* [67]; **d** the most common of the nine 5α,8α-epidioxy sterols from the sea anemone *Metridium senile* [33]; **e** a 20-epicholanic acid derivative from the sea-pen *Ptilosarcus* sp. [25]; **f** a 3-ketopregnan from soft corals [25]

found are saturated sterols, e.g. in *Homaxinella balfourensis* and *Axinella damicornis* [105]; Δ^7-sterols, e.g. in *Agelas oroides*; or $\Delta^{5,7}$-sterols, e.g. in *Spongionella gracilis*, *Ircinia pipetta* and *Dysidea avara* [106]. Sterols with saturated ring systems or with Δ^5, $\Delta^{5,7}$ or Δ^7 double bonds are in fact found in many marine organisms, but *Axinella cannabina* and several other sponge species also have $\Delta^{5,7,9(11)}$-, $\Delta^{7,9,(11)}$-, Δ^8-, $\Delta^{8(14)}$- and $\Delta^{6,8(14)}$-sterols [19, 25]. Even more unusual **ring systems** have been found in the sterols of other *Axinella* species. *A. polyploides* predominantly contains sterols without the 19-methyl group (19-norsterol; Fig. 16.9a), and *A. verrucosa* has sterols with a reduced A ring (A-norsterols; Fig. 16.9b); in contrast, *A. damicornis* has the usual 5α-cholestan derivatives. The species *Acanthella aurantiaca* from the Red Sea contains 16 different 3-hydroxymethyl A-norsterols but has no sterols with the normal ring system. The 19-norsterols and A-norsterols arise from cholesterol, whereby cholest-4-en-3-one has been identified as an intermediate of A-norsterols synthesis. Various 5α,8α-epidioxy sterols (cf. Fig. 16.8d) are also found in the sponges [25, 75].

The **side-chains** of sponge sterols are even more variable than the ring systems. 24-Methyl and 24-ethyl sterols with the usual structure are widely found, and these are derived from the sterols of the plankton diet; however, the sponges also contain 24-isopropyl sterols (Fig. 16.9c). Especially characteristic of porifera sterols are the side-chains on C-26 or C-27 which are extended by additional methyl groups (Fig. 16.9d–f). Sterols with a cyclopropane in the side-chain are known only in marine organisms; in the poriferans, the cyclopropane ring can occupy various positions: with a CH_2 bridge between C atoms 22 and 23, as in the gorgosterols of the corals (Fig. 16.8a); with a bridge between C atoms 23 and 24 as in 23,24-dihydrocalysterol from the sponge *Calyx niceaensis* (Fig. 16.9g); or with bonds between C atoms 20 and 23, 24 and 26 or 26 and 27 (Fig. 16.9h) [18]. Even triple bonds between C atoms, which are otherwise seldom found in biological substances, have been reported in sponge sterols (Fig. 16.9i) [25, 75].

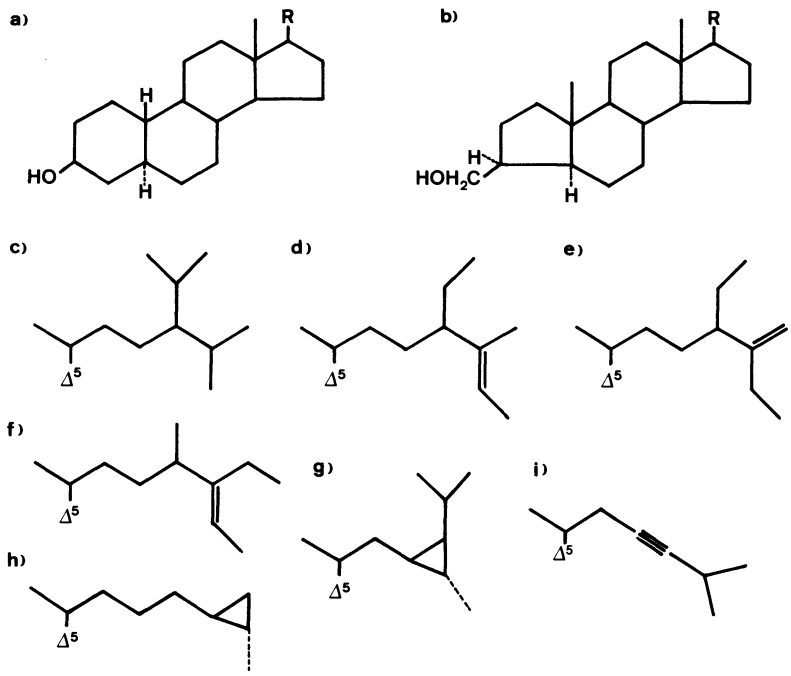


Fig. 16.9 a–i. Unusual ring system **(a, b)** and side-chains **(c–i)** in the sterols of poriferans [95]. **a** Ring system of the 19-norstanols from *Axinella polypoides*; **b** 5β-hydroxymethyl-A-nor-5α-cholestan forms the ring system of the sterols from *Axinella verrucosa*; **c** 24-isopropylcholesterol from *Pseudoaxinyssa* sp.; **d** stelliferasterol from *Jaspis stellifera*; **e** strongylosterol from *Strongylophora purissima*; **f** verongulasterol from *Verongula cauliformis*; **g** 23,24-dihydrocalysterol from *Calyx nicaensis*; **h** petrosterol from *Petrosia ficiformis*; **i** cholest-5-en-23-yn-3β-ol from *Calyx nicaensis*

16.2.7 Sterols of the Protozoa

There is a marked absence of systematic studies of sterol content and metabolism in the Protozoa, and given the heterogeneity of this group, no generalizations are possible. Ciliates of the genus *Tetrahymena* cyclicize squalene in anaerobic reactions to tetrahymanol (Fig. 16.3), which then assumes sterol functions as a membrane component. However, whenever sterols are present in the diet, these are incorporated into the membrane and tetrahymanol synthesis is suppressed [25]. Cholesterol is an essential nutrient for *Paramecium tetraurelia* and rumen Ciliata [96]. The synthesis of cholesterol by the amoebae *Acanthamoeba* and *Naegleria* via cycloartenol (Fig. 16.3), instead of lanosterol as in other animals, has already been mentioned [91].

16.3 Steroid Hormones of the Vertebrates

Certain steroids with 21, 19 or 18 C atoms function as sex hormones or as regulators of carbohydrate or mineral metabolism in all vertebrates. In accordance with the aims of this book, the structure and metabolism of the steroid hormones will be the main concern (as was the case in Chapter 8 for the proteo- and peptide hormones); a discussion of the functional aspects should be sought in comparative accounts of hormone physiology. All steroid hormones arise by shortening of the side-chain of cholesterol (Fig. 16.10). Cleavage of side chain residues with C atoms 22–27 leads to the C_{21} steroids, the basic component of which is referred to as 5α- or 5β-pregnan, depending on the configuration of the A/B ring association. The corticosteroids and progesterone are C_{21} steroids. The former derive their name from their site of synthesis in the adrenal cortex, and they are involved in the regulation of carbohydrate and mineral balance. Progesterone is a female sex hormone produced by the ovary, but is also an important intermediate of steroid hormone biosynthesis in all steroidogenic tissues. The complete cleavage of the side-chain on C-17 leads to the male sex hormones, or androgens, with the C_{19} component 5α- or 5β- androstan. As a consequence of the aromatization of ring A there is no valency free at C-10 for binding of the C-19 methyl group; this leads to the formation of the female sex hormone, oestrogen [39, 74].

For many years, the C_{18}, C_{19} and C_{21} steroids were considered a special feature of the vertebrates. However, it has since been realized that they are widely distributed, if not ubiquitous, in the eukaryotes. Oestrogens are found even in yeast, and their biosynthesis has been demonstrated in plants.

Many invertebrates have C_{18}, C_{19} and C_{21} steroids. Thus, the steroids with hormone action in

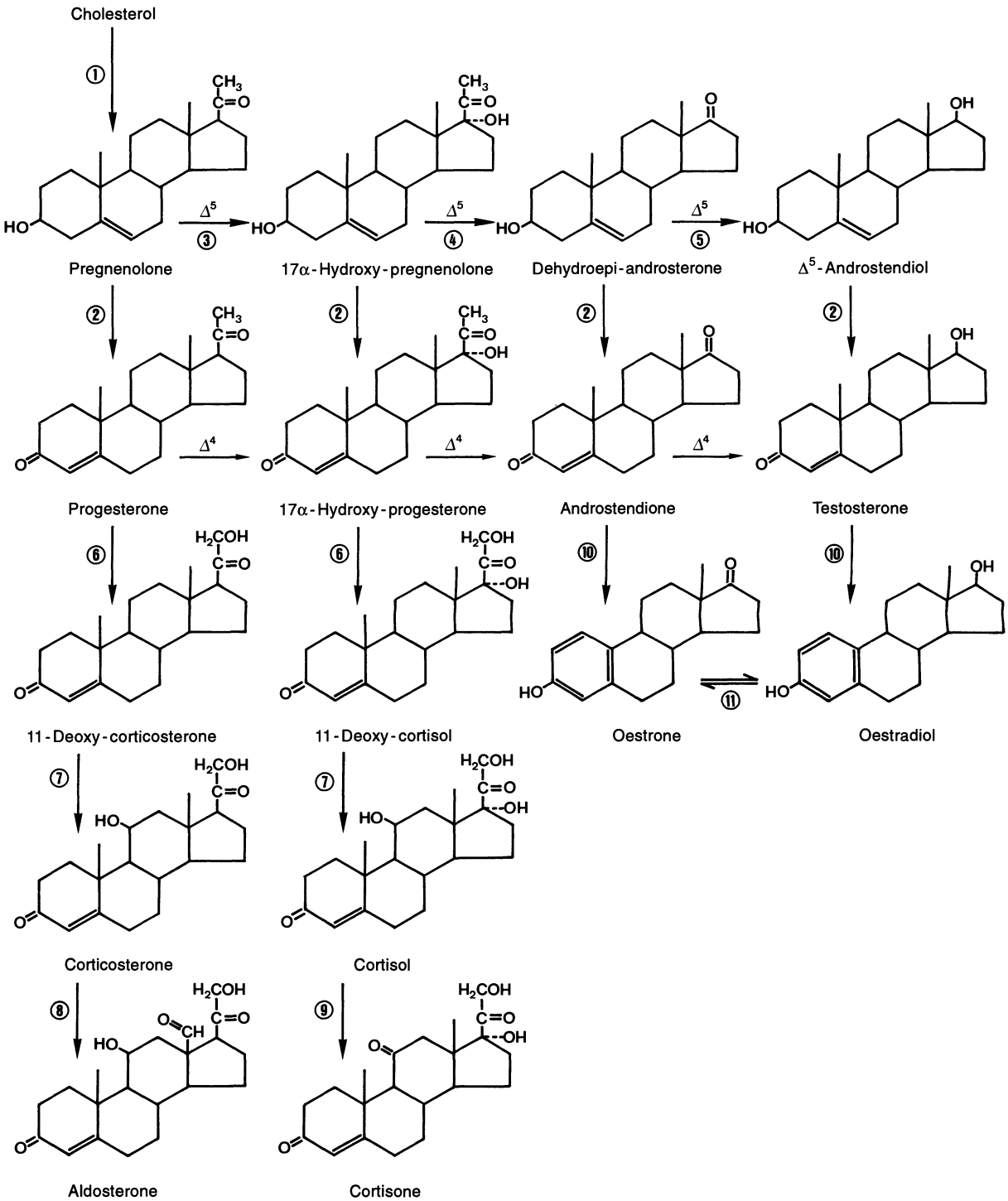


Fig. 16.10. The most important steroid hormones of the vertebrates and their biosynthesis from cholesterol [39]. The reactions of the Δ^4 and Δ^5 pathways (p. 639) are indicated by the corresponding symbols. *1*, Oxidative cleavage of the side-chain on C-20; *2*, oxidation at C-3 and Δ_5/Δ_4 isomerization; *3*, hydroxylation at C-17; *4*, oxidative cleavage of the side-chain at C-17; *5*, reduction at C-17; *6*, hydroxylation at C-21; *7*, hydroxylation at C-11; *8*, oxidation at C-18; *9*, oxidation at C-11; *10*, aromatization of ring A and cleavage of C-19; *11*, oxidation/reduction at C-17

the vertebrates are apparently phylogenetically rather ancient, and the evolution of the vertebrate steroid hormone system began not with the appearance of the steroid structure but with the "invention" of specific steroid receptors. It therefore becomes clear why there is no characteristic structural feature of the vertebrate steroid hormones; even the frequently occurring 3-oxo-4-en structure (Fig. 16.10) is not universal [39, 74].

The steroid hormones are found in much lower **amounts** than the sterols. The human adrenal cortex secretes about 30 mg corticosteroids per day. However, because of their extremely rapid turnover, the blood plasma contains only about 100 μg/l; the concentration of the sex hormones is one or two orders of magnitude lower. The extraction of large quantities of starting material, however, led to the identification of more than 50 corticosteroids in the blood and over 200 steroids and steroid metabolites in the urine. Most of these substances are in fact only intermediates, by-products or end products of metabolism; the **number of natural hormones** is relatively small. Amongst the corticosteroids of the mammals, cortisol, cortisone and corticosterone mainly affect carbohydrate metabolism and are therefore known as glucocorticoids; aldsterone and 11-deoxycorticosterones are the most important mineral corticoids. Dihydrotestosterone and testosterone are the most typical male sex hormones; the female reproductive process is regulated mainly by oestradiol, oestrone and progesterone (for the formulae see Fig. 16.10) [39, 74].

The **spectrum of biologically active steroid hormones** appears to be similar in all vertebrates, although the concentrations of individual components may vary. As yet, very few steroid hormones have been found to be restricted to certain vertebrate groups or species, e.g. equilenin and equilenin of the horse, or 1α-hydroxycorticosterone of the elasmobranch. However, it should be borne in mind that the consistently low concentrations of the steroid hormones have resulted in comparative investigations making use primarily of biological activity or immunological reactivity, and not physicochemical methods, for purposes of identification. The variety of steroid effects in the vertebrates apparently relies more upon differences in relative concentrations and reaction chains than on hormone structure. The **primary target of steroid hormone action** is gene transcription in the nucleus. In contrast to the peptide hormones, the lipophilic nature of the steroids gives them easy access to the inside of the cell across the cell membrane. There they are bound to specific receptors that then interact with certain regulatory DNA sequences (hormone regulatory elements) [6]. The same steroid can have very different effects on different cells or organisms, depending upon the gene transcribed and the secondary effect of its product [39, 74].

The **steroid receptors** belong to the protein super-family of the nuclear hormone receptors. They recognize transcriptional enhancer sequences and control networks of genes that have important effects on growth, development and homeostasis. They bind specific lipophilic hormones, including not only the steroids but also the thyroid hormone and retinoic acid. The chicken ovalbumin upstream promoter transcription factor (COUP-TF) and various receptors of the fly *Drosophila* also belong to this super-family [125, 127]. Despite their functional diversity, the receptors show extensive agreement in domain structure and amino acid sequences. Six domains (A-F) can be distinguished. The DNA-binding domain C is the most highly conserved in the super family, and consists of 66–68 amino acids folded into two zinc-finger motifs. Each finger binds a zinc ion that is coordinately bound to four cysteine residues. The first finger determines DNA-sequence specificity, and the second is involved in protein-protein interactions, such as the dimerization of the receptor. The sequence of the first finger characterizes two subfamilies with the glucocorticoid and oestrogen receptors as nominal representatives. The domain E, with about 250 amino acids, is also well conserved between members of the super-family; this carries the hormone-binding site and is responsible for receptor dimerization, nuclear translocation and hormone-dependent transcriptional activation. An independent transcriptional activation function is located in domains A and B, the lengths and sequences of which vary greatly between the different receptor types. Domain D, carrying the nuclear localization signal, is also relatively poorly conserved, as is the C-terminal domain F. The total length of the nuclear receptors varies between 433 and 934 amino acids. Additional variation is generated by the existence of several genes for the same receptor type, alternative promoters for the same gene, and alternative splicing [54, 125]. Some steroid receptors appear exclusively in the nuclear fraction on cell extraction and others appear also in the cytosol. This is apparently due to the low affinity for chromatin of some receptors. A vertebrate species such as the spiny dogfish, whose tissues have an extremely high osmolality of 1010 mOsmol/kg

water, is reported to have steroid receptors that bind more firmly to chromatin than do those of the mammals, with a tissue osmolality of about 300 mOsmol/kg water. The oestrogen receptor of the dogfish is eluted from DNA-cellulose with 0.55 mol/l NaCl, compared with 0.22 mol/l required for the freshwater ray *Potamotrygon* sp. and the mammals [16]. The sequence similarity between members of the family is ≥42 % in the DNA-binding domain C, usually ≥15 % in the hormone-binding domain E, and negligible in the other domains [28, 127]. Sequence comparisons between receptors of the same type show, for example, 81–82 % agreement between, the mouse progesterone receptor and those of humans and rabbits, and 50 % agreement with that of the chicken [104]; the thyroid receptor of the frog *Xenopus laevis* agrees 92 % with that of the chicken [11]. The human and chicken oestrogen receptors have 80 % of their 589 amino acids in common, and the corresponding receptor from the rainbow trout shows 92–95 % similarity to other species in domain E, but lacks domain A [66, 88].

16.3.1 Biosynthesis of Steroid Hormones

All vertebrate steroid hormones share a common **synthesis pathway** (Fig. 16.10). This begins with the oxidative cleavage of the side-chain on C-20 of cholesterol by an NADPH-dependent mitochondrial hydroxylase with cytochrome P-450. This yields isohexanal and pregnenolone. The ensuing metabolism takes place in the endoplasmic reticulum. Progesterone arises by conversion of the 5-en-3β-hydroxy structure of pregnenolone to the typical 4-en-3-oxo structure of the biologically active steroids by 3β-hydroxy steroid dehydrogenase and Δ^5/Δ^4-isomerase. Metabolic pathways lead from progesterone to corticosterone and aldosterone, and from 17α-hydroxy progesterone to cortisol and cortisone. The biosynthesis of testosterone from pregnenolone requires, on the one hand, the sequence of 17α-hydroxylation, oxidative cleavage of the side-chain on C-17 and reduction of the 17-oxo group to 17α-hydroxyl, and, on the other hand, the conversion of the 5-en-4β-hydroxy to the 4-en-3-oxo structure. Depending on the order of these reactions, one can distinguish the Δ^5 pathway, from pregnenolone to testosterone via 5-androstendiol, and the Δ^4 pathway, via 4-androstendione (Fig. 16.10). The rat and other rodents mainly use the Δ^4 pathway, humans and rabbits use the Δ^5 route, and goats and many other mammals have both pathways. Aromatization of the A ring and the consequent cleavage of the C-19 methyl group leads from testosterone to oestradiol-17β, and from androstendione to oestrone [39, 74].

All steroid-forming tissues have the same **enzyme spectrum** of dehydrogenases, reductases, hydroxylases and isomerases, although perhaps with different relative activities. The species- and organ-specific spectrum of products arises from the preference for certain synthetic pathways and from cessation of synthesis at different steps. In addition to the steroids that are typical of the organ, lower amounts of other steroids may also be produced, e.g. sex hormones in the adrenal cortex, oestrogen in the testis, and androgen in the ovary. The synthesis of atypical steroids may increase markedly in tumours. Certain steroid-metabolizing enzymes are also found in other tissues; noteworthy in this context is the reduction of testosterone to dihydrotestosterone in associated glands of the male genital tract, and the formation of oestrogens from testosterone or androsterone in the liver, brain and blood of mammals. The **inactivation of steroid hormones** occurs mainly in the liver, but in mammals and reptiles it also occurs in the kidney. The degradation pathways are not yet known in detail, even for the mammals. Important steps in inactivation, however, include the reduction of the 4-en-3-oxo configuration and the 20-oxo group. The newly formed 3-hydroxy group is esterified with glucuronate or sulphate and the conjugates are excreted in the urine [9, 39, 74].

16.3.2 Steroid-Binding Proteins

Because of their relatively low solubility in water, the steroid hormones can only be transported in the blood plasma when they are bound to proteins. Serum albumin binds steroids unspecifically with relatively low affinity, and most vertebrates have specific **steroid-binding plasma proteins** [34]. The difference between the affinity constants of the binding proteins (10^{-8} mol/l) and the intracellular receptors (10^{-9} mol/l) thermodynamically favours the uptake of steroid molecules into the cell. Two types of steroid-binding globulins are usually found in the vertebrates from the teleosts to the mammals; of these, the corticosteroid-binding globulin (CBG) or transcortin binds all C_{21} steroids, including progesterone, and the hormone-binding globulin (SHBG) binds C_{18} and

C_{19} steroids. Plasma proteins from sharks, rays and lampreys bind C_{21}, C_{18} and C_{19} steroids with the same low affinity, and the blood of the common hagfish *Myxine* contains no steroid-binding proteins of any sort. **Transcortin-like proteins** are apparently widely distributed from the bony fish onwards. There are, however, large differences in amino acid composition and molecular mass between, for example, human and chicken transcortins. CBG, like other plasma proteins, is secreted by the liver but the corresponding human mRNA is also found in the lung, testis and kidney. Human CBG consists of a polypeptide chain of 383 amino acids and five oligosaccharides, which make up about 23 % of the molecular mass of 58 kDa. The amino acid sequence shows no similarity to other steroid-binding proteins but does exhibit significant homology to α_1-antitrypsin and other serpins [44].

Sex hormone-binding globulin has been found in many mammals, but is absent from rodents, elephants and the birds. Reptiles possess SHBG and, in fact, have separate androgen- and oestrogen-specific proteins. SHBG-like proteins have also been detected in various amphibians and teleosts [102]. The SHBG of the mammals is a dimer of two identical polypeptides of 367–373 amino acids with three carbohydrate chains, although it binds only one steroid molecule [45, 126]. At least in the N-terminal region, it shows sequence similarity to the similar-sized **androgen-binding protein** (ABP) from mammalian testis; the function of ABP appears to be to concentrate androgens and thereby amplify their effects on spermatogenesis and sperm maturation [41, 57]. A similar protein is found in the testis of the spiny dogfish *Squalus acanthias* [73]. The **steroid-binding protein from rat prostate** apparently functions in a feedback mechanism to regulate androgen activity; it is a tetramer of four different polypeptides that both bind to, and are induced by, androgen. This protein shows significant sequence homology to the **uteroglobin** from the uterus and lung of the rabbit; this is a homodimer of about 16 kDa that is both induced by and binds to progesterone. The 91 codons of the gene sequences of the pre-uteroglobins of the rabbit and of the hare *Lepus capensis* have five synonymous substitutions, suggesting that the two species separated about 4 million years ago; there are also six amino acid-exchanging substitutions, which suggests an unusually high rate of evolution [72].

16.3.3 Corticosteroids

Only in the mammals are steroid-producing and catecholamine-producing (chromaffine) cells of the adrenal gland organized into the cortex and medulla of a compact organ; in the birds, reptiles and amphibians they are mixed together in various ways, and in fish they form two separate organs. The corticosteroids carry a side-chain of two C atoms on C-17, and these can be substituted with hydroxy or oxo groups. Additional oxygens on C-3, C-11, C-17 or C-18 result in a large number of related steroids, most of which are inactive metabolic products. Due to the large number of polar groups, the corticosteroids are quite soluble in water. All biologically active corticosteroids influence both the carbohydrate and the mineral balance, but as one function usually predominates a distinction may be drawn between **glucocorticoids and mineral corticoids**. The glucocorticoids all have a hydroxy or oxo group at C-11 (for the formula see Fig. 16.10), whereas the 11-deoxysteroids act as mineral corticoids. One exception to this rule is aldosterone, in which the 11-hydroxy group forms a semi-acetal, together with the aldehyde group on C-18; this is unique to this molecule (Fig. 16.11a). The glucocorticoids stimulate the synthesis of various enzymes of gluconeogenesis, resulting in glycogen storage. The mineral corticoids act primarily on ion transport via the ATPases of the cell membrane, and regulate the excretion of sodium and calcium in the distal tubules of the kidney [39, 74]. The **spectrum of corticosteroids** varies between the different vertebrate groups. Cortisol predominates in most mammals. Only in the case of the rat, mouse and rabbit is there almost no cortisol but only corticosterone. However, exogenous cortisol is still highly active in these species. Birds appear to completely lack cortisol, and corticosterone predominates in the reptiles and the amphibians. Cortisol is, however, the main corticosteroid of the teleosts. Aldosterone appears to be absent from the agnathans, elasmobranch and teleosts, but is present in the dipnoans and all tetrapods. In the teleosts cortisol also regulates sodium uptake by the gills. In the birds, reptiles and amphibians, aldosterone and also corticosterone appear to be involved in regulation of the mineral balance; both act, to some extent, antagonistically. From this it is clear that the classification into glucocorticoids and mineral corticoids is only strictly observed in the mammals [39, 74].

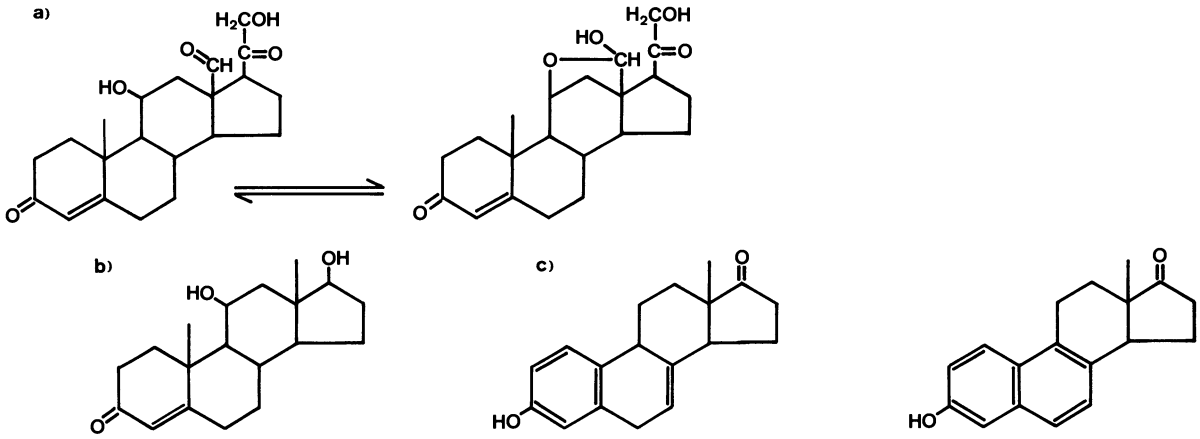


Fig. 16.11a–c. Tautomeric forms of aldosterone and several unusual steroid hormones of vertebrates [39]. **a** Aldehyde and semi-acetal forms of aldosterone; **b** 11β-hydroxytestosterone of teleosts; **c** equilin *(left)* and equilenin *(right)* of the horse

16.3.4 Sex Hormones

Although neither the biosynthesis nor the action of the C_{19} and C_{18} steroids are strictly sex specific, they are usually referred to as male and female hormones. The **androgens** regulate the appearance of the male primary and secondary sexual characteristics. They are produced in the testis but also arise in the adrenal cortex and the ovary. The main product of the mammalian testis is testosterone. In some species, intermediates of testosterone synthesis are also found in the bloodstream, e.g. pregnenolone and progesterone in the rat, and dehydroepiandrosterone and androstenediol in the rabbit. The androgens isolated from the mammalian testis also show species-specific differences in composition; for example, a particularly high proportion of sulphated intermediates occurs in humans, the dog and the pig. Testosterone is reduced to dihydrotestosterone (5α-androstan-17β-ol-3-one) in associated glands of the genital tract or other peripheral organs; this then binds to specific receptors in the target organ and is apparently the biologically active androgen. The ratio of testosterone to dihydrotestosterone in the blood plasma is about 10 : 1.

Even in the anurans, the main product of the testis is dihydrotestosterone, and the other 5α-metabolites arise only in the peripheral organs. Large amounts of 5α-metabolites are otherwise found only in some birds, e.g. the chicken and the Japanese quail *Coturnix coturnix;* the urodelans, reptiles, other birds and the mammals have only trace amounts of these androstan derivatives. The fish are unique amongst the vertebrates in that their gonads have 11-hydroxylase activity, which is otherwise found only in steroid-producing adrenal tissues. In this case, the testis produces mainly 11β-hydroxytestosterone (Fig. 16.11b) and 11-oxotestosterone. 11β-Hydroxytestosterone is more active than testosterone itself and is apparently the natural androgen. The teleost *Coris julis* undergoes a natural sex change in that after laying the eggs the females undergo a change in colour and turn into males; in addition to these secondary types, there are also primary males that look like females. Only the secondary males produce large amounts of 11β-hydroxytestosterone. Their different colouring, is apparently associated with the higher androgen activity of this steroid; primary males and females can be induced to change colour by injection of high doses of testosterone [94].

In addition to progesterone, the mammalian ovary produces mainly oestradiol-17β and oestrone; the horse additionally produces low amounts of the C_{18} steroids equilin and equilenin (Fig. 16.11c), which are more unsaturated and weaker in activity. The **oestrogens** are formed mainly in ripe follicles and **progesterone** is formed in the corpus luteum. In some species, the latter also has the capacity to synthesize oestrogen, e.g. in humans and pigs but not in cows or sheep. Androgens are also formed in the ovary, apparently in cells of the stroma. Thus, the ovary contains different steroid-synthesizing cell types whose individual activities have not yet been completely defined. In humans and all other mammals where pregnancy lasts more than 70 days, the placenta secretes not only progesterone but also large amounts of oestrogens, leading to a marked increase in oestrogen excretion in

the urine. In all mammals except the primates, oestrone is the main oestrogen in the urine, whereas in humans and the anthropoid apes it is oestriol. In addition, the primates produce oestradiol-17β and the other mammals produce oestradiol-17α.

Enzymes that can aromatize the A ring and produce oestrogens have been found in the testis of mammals, reptiles and teleosts. The production of oestrone and oestradiol-17β is especially high in the testes of the horse and pig; these are amongst the most oestrogen-rich tissues and stallion urine was formerly used as a copious source of oestrogen. The possible biological importance of this female hormone in the testis is not known. In males of the zebra finch *Poephila guttata* there is intensive oestrogen synthesis in the hypothalamus-hypophysis region of the brain; the hormone is involved in the regulation of song activity [103]. The biosynthesis of oestrogen in mammals occurs by either the Δ^5 (cattle) or the Δ^4 pathway (horse, pig). Inactivation of oestrogen in this case is not by reduction of the A ring, as for other steroid hormones, but by substitution of various ring positions with hydroxy or oxo groups; this increases the water solubility, and with it the excretability. Thus, oestradiol-17β is converted via oestrone to oestriol, during which the biological activity declines in the ratio 100 : 20 : 1. The female hormones control the reproductive processes in mammals, with oestrogens regulating the growth and progesterone the differentiation of the uterus lining and the maintenance of pregnancy. The oestrogens are also involved in regulating the reproductive rhythms of all other vertebrate classes, but comparative studies are hindered by their relatively low concentrations compared with those of the corticosteroids and androgens. As in the mammals, oestradiol-17β appears to be the most important female sex hormone in the other vertebrates, but the biological role of progesterone is not so clear [39].

Certain steroids in vertebrates function as **sex pheromones**. For example, in various fish species oestrogens excreted into the water by females attract and activate the males. Females of the goldfish *Carassius auratus* produce a steroid (Fig. 16.12a) that as a hormone promotes maturation of oocytes, and as a pheromone stimulates sperm release in the males [110]. Steroidal sex pheromones are also found in mammals, probably including humans. Males of the domestic pig produce in the maxillary gland a fragrant steroid (Fig. 16.12b) that suppresses the movements of the female and thereby facilitates copulation; the steroid has also been detected in the sweat of human males [3].

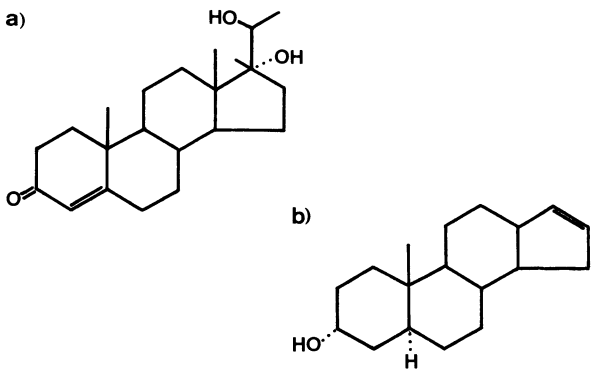


Fig. 16.12a, b. Steroidal sex pheromones of the vertebrates [3]. **a** 17α,20β-Dihydroxy-4-pregnen-3-on is a sexual pheromone of female goldfish; **b** 5α,16-androsten-3α-ol is a sexual pheromone of male domestic pigs (boars)

16.4 Ecdysteroids

In 1954, Butenandt and Karlson isolated 25 mg ecdysone from 500 kg silkworm pupae, but the correct structural formula of this hormone (Fig. 16.13) was first published by Karlson in 1965. In the intervening period, a further moulting hormone was found in various insects and crustaceans; it was known at first as "β-ecdysone", "crustecdysone" or "ecdysterone" but is now designated 20-hydroxyecdysone (Fig. 16.13). Ecdysone and 20-hydroxyecdysone were for a long time considered to be the only moulting hormones of the arthropods, but substances in addition to, or instead of, 20-hydroxyecdysone are now known from several species: makisterone A in insects and a brachyuran, ponasterone A in various brachyurans (Fig. 16.13), and several unidentified molecules in the Pantopoda [12]. All such moulting hormones, together with their metabolites and other similar steroids in animals and plants, are classified as ecdysteroids. They differ from the steroid hormones of the vertebrates in that the cholesterol side-chain is maintained in an extended form, and they are more water soluble due to a higher number of OH groups [65, 132].

The concepts of the cellular **modes of action of the steroid hormones** were first established in the 1970s, mainly from studies of the insect moulting hormones. Of particular help was the fact that

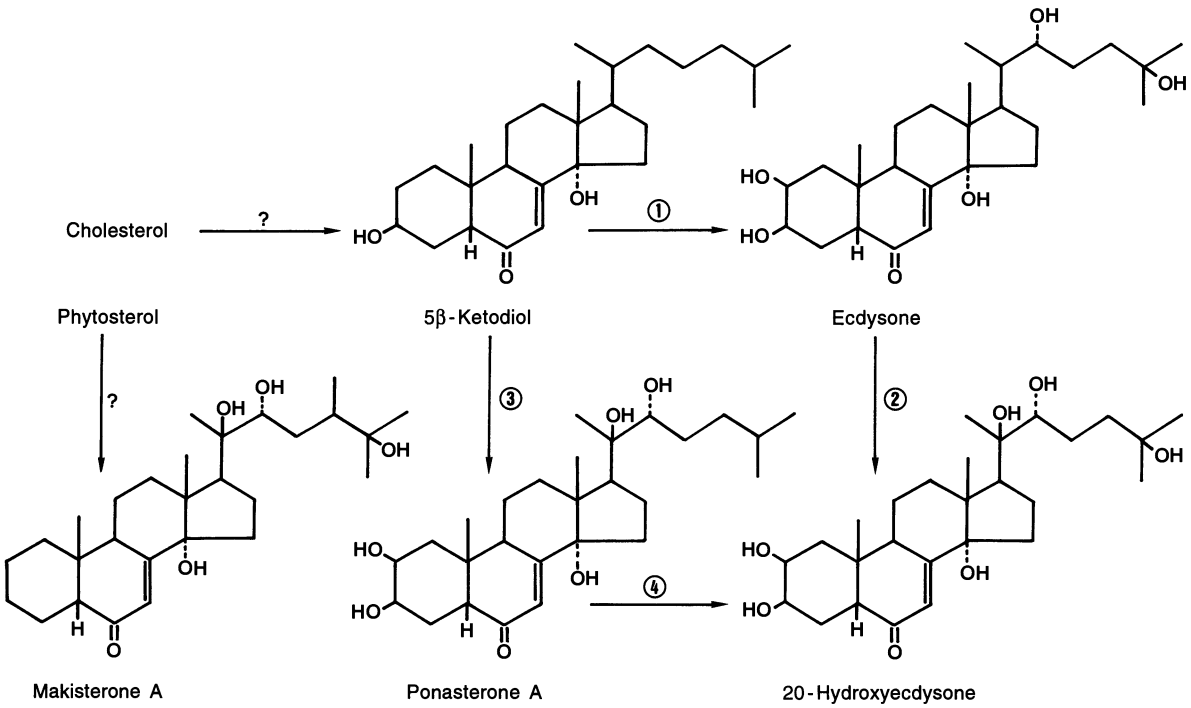


Fig. 16.13. Biosynthesis of ecdysteroids [49]. *1*, Hydroxylation at C-25, C-22 and C2; *2*, hydroxylation at C-20; *3*, hydroxylation at C-22, C-2, C-20 (order?); *4*, hydroxylation at C-25. See text for further explanation

gene activation of ecdysteroids could be directly observed in the formation of so-called puffs on the giant chromosomes from the salivary glands of the Diptera. The ecdysone-dependent heat-shock gene hsp27 of *Drosophila melanogaster* includes sequences with significant similarity to the steroid-response elements of the vertebrates [76]. The ecdysteroids apparently act not only on the transcription of certain genes but also post-translationally, e.g. on the phosphorylation of some proteins [65, 132]. At least seven members of the steroid/thyroid hormone receptor superfamily have been identified so far in *Drosophila melanogaster*. They are involved in the regulation of development and are known after characteristic mutations: knirps (kni), knirps-related (knrl), embryonic gonad (egon), 75B, ultraspiracle (usp or 2C), seven-up (svp) and tail-less (tll). All these receptors have the typical DNA-binding domain with two zinc fingers, and svp, 75B and 2C are also very similar to vertebrate receptors in their hormone-binding domains [48, 125].

Ecdysteroids, like all steroids, are produced from cholesterol. The arthropods themselves are not able to synthesize cholesterol de novo, but many tracer experiments with insects and crustaceans have shown that they can metabolize steroids in their diet to ecdysteroids. The site of **biosynthesis** in insect larvae is the prothorax gland, in crustaceans it is the Y organ at the base of the antennae, and in ticks it is the epidermis [131]. The prothorax gland degenerates towards the end of the pupal resting stage but many adult insects produce ecdysteroids, either in the follicle cells of the ovary or, in some cases, in the oenocytes. At least in the lepidopterans the biosynthesis is regulated by the prothoracicotropic neurohormone (PTTH), the structure and mode of action of which is known for several species (p. 307). It is certain, however, that factors or mechanisms other than PTTH are involved in the regulation of ecdysteroid synthesis and secretion. Only the final steps of the biosynthetic pathway are known in any detail (Fig. 16.13): the first confirmed intermediate is a 5β-ketodiol, which is converted to ecdysone by hydroxylation on C-25, C-22 and C-2. The main product formed from radiolabelled cholesterol in the prothorax glands of *Manduca sexta* and other lepidopterans is not ecdysone but 3-dehydroecdysone, a compound that does not feature in the classical scheme of biosynthesis. Ecdysone can be produced from 3-dehydroecdysone by a 3-dehydroecdysone reductase in the haemolymph and other tissues. This enzyme is found not only in the Lepidoptera but also in Blattaria, Orthoptera, Coleoptera and

Diptera [40, 63]. Little is known about the other enzymes of ecdysteroid synthesis; the 25- and 22-hydroxylases are apparently microsomal, and the 2-hydroxylase is a mitochondrial cytochrome P-450 monooxygenase, but without the characteristic inhibition by CO [49, 58].

The best-known enzyme of ecdysteroid biosynthesis is the **20-monooxygenase** that hydroxylates ecdysone to 20-hydroxyecdysone. This is not present in the prothorax glands but is found in the fat body, the Malpighian tubules, the follicle epithelium and other tissues. It is a typical cytochrome P-450 monooxygenase with microsomal and mitochondrial forms in species-specific proportions [109]. It is still a matter of debate whether **ecdysone or 20-hydroxyecdysone** is the actual moulting hormone. It was shown in 1973 that the two hormones bring about different puff patterns in the giant chromosomes of *Chironomus*. In 1978, experiments with *Drosophila* showed that the two hormones have the same effects but that 20-hydroxyecdysone is 50- to 100-fold more active than ecdysone. Three cell lines of *Drosophila melanogaster* with very different capacities for the hydroxylation of ecdysone showed the same response to the two moulting hormones. Depending upon the test model, ecdysone and 20-hydroxyecdysone can have very different, even complementary, effects. It may be that a hormonal effect of ecdysone that still functions in primitive arthropods has been lost during evolution, so that in some groups it is now essentially only a pro-hormone for 20-hydroxyecdysone [65, 132].

Makisterone A differs from 20-hydroxyecdysone only by having a methyl group on C-24, i.e. it is a C_{28} ecdysteroid (Fig. 16.13). In the insects, this ecdysteroid has so far been detected only in several species of Hemiptera, Diptera and Hymenoptera, and in the crustaceans in the brachyuran *Callinectes sapidus;* it is, however, probably widely distributed in the arthropods. The insects referred to include *Drosophila melanogaster, Apis mellifera* and various phytophagous Hemiptera from the families Pentatomidae (e.g. *Oncopeltus fasciatus* and *Dysdercus cingulatus)* and Alydidae *(Megalotomus quinquespinosus)*. The diet of all these species contains almost no cholesterol and, in addition, they lack the ability to produce cholesterol from ingested C_{28} and C_{29} steroids by cleavage of the 24-alkyl residue. Hence, they synthesize their moulting hormones from 24-methylsterols. However, in the course of evolution they have turned the problem to their advantage and adapted to this rather unusual moulting hormone; in some Hemiptera makisterone A is tenfold more active than 20-hydroxyecdysone. Even the predatory pentatomid *Podisus maculiventris*, which undoubtedly obtains sufficient cholesterol in its diet, still relies upon makisterone A [30, 31, 65, 132]. **Ponasterone A** (Fig. 16.13) was discovered in 1979 in the brachyurans *Callinectes sapidus, Carcinus maenas* and *Gecarcinus lateralis*. This is converted to 20-hydroxyecdysone by 25-hydroxylation and to **inokosterone** by 26-hydroxylation. Both end products are found as natural hormonal components in crustaceans but in some insects, e.g. *Bombyx mori*, they are derived from exogenous ponasterone A [68].

In some adult insects, ecdysteroids that are synthesized in the follicle cells of the ovary are accumulated in species-specific proportions in the haemolymph or eggs; e.g. in the migratory locust *Locusta migratoria* ecdysteroids occur in the ovary and haemolymph or eggs in the ratio 50 : 1, and in the wax moth *Galleria melonella* in the ratio 1 : 1. Compared with the relatively simple ecdysteroid spectrum of the larval stages, the composition of **ecdysteroids in the females and in the eggs** is very complex. Here, there are not only large quantities of synthesis intermediates, such as 2,22-dideoxyecdysone and 2-deoxyecdysone in *Bombyx mori*, but also inactivation products, such as 20,26-dihydroxyecdysone in *Carausius morosus*, ecdysone-22-phosphate in *Bombyx mori*, 26-hydroxyecdysone-26-phosphate in *Manduca sexta* and ecdysonic acid in *Schistocerca gregaria* [32, 35, 86, 119]. The majority of ecdysteroids in the ovary and eggs are always conjugated. The insects mainly contain ecdysteroid-22-phosphate and also 26-phosphate [119], but, in contrast, the ticks have 22-fatty acid esters [22, 23]. Twenty-two long-chain fatty acyl esters of ecdysone have recently been discovered in freshly laid eggs of the house cricket *Acheta domestica* [128]. The most unusual conjugate has been found in freshly laid eggs of *Locusta migratoria* (Fig. 16.14); here, the 22-OH of ecdysone is bound via a phosphate residue to N^6-isopentenyl adenosine (cytokinin), which functions in plants as a growth regulator but is also found as the phosphate in mammalian cells [120]. During embryogenesis in *Schistocerca gregaria* and other orthopterans, the conjugates are gradually hydrolyzed, and the released ecdysteroids are further metabolized to inactive end products. Thus, in these primitive insects, the ecdysteroids of the ovary are involved in the development of the embryo, whereas in the more highly developed insects they regulate the production of yolk pro-

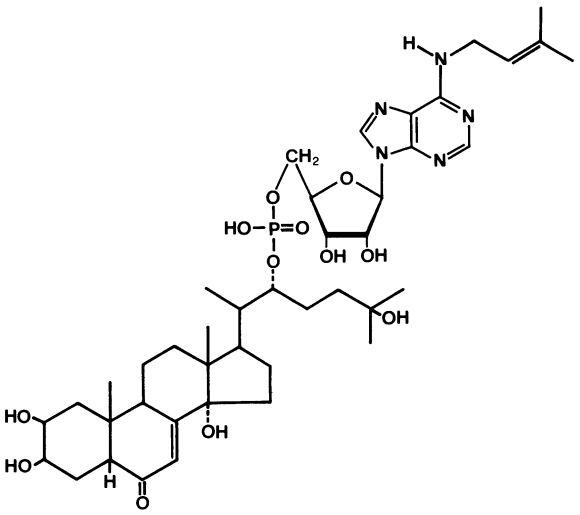


Fig. 16.14. Ecdysone conjugated with cytokinin [N^6(isopentenyl) AMP] from the eggs of the migratory locust *Locusta migratoria* [120]

tein in the fat body. Ecdysone even stimulates yolk protein synthesis in male mosquitoes. The observation that vitellogenin synthesis in some dipterans declines only a little, or not at all, after ovariectomy suggests the existence of further sites of ecdysteroid production; the oenocytes are a prime candidate [49, 71, 119].

The moulting hormones appear to be the subject of very rapid **turnover;** 83 % of 480 ng of 20-hydroxyecdysone (an extremely large amount) is converted to more strongly polar metabolites within 10 min of injection into adult *Drosophila* males [107]. The **metabolism of ecdysone** can occur by various pathways, and more than 60 metabolites have been identified to date. The most important reactions are: C-20 hydroxylation; oxidation at the C-3 position with the formation of 3-dehydro ecdysteroids; formation of the 3α-epimer by reduction of the 3-dehydro bond; hydroxylation at C-26 and further oxidation to ecdysonic acid; and phosphorylation or acetylation of the 3α-, 3β-, 22α- or 26β-hydroxyl groups [65, 108, 132]. Little is known about the enzymes of ecdysone metabolism in insects. The 22-phosphotransferase of the locust *Schistocerca gregaria* has been purified [59]. 14-Deoxyecdysone is a major product of ecdysone metabolism in the mouse, whereas endogenous dehydroxylation of ecdysteroids has never been observed in the insects. The 14-deoxy ecdysteroids detected in *Gryllus bimaculatus* apparently arise from the metabolism of gut symbionts [50]. Ecdysteroids that are phosphorylated or acylated on the 22α-hydroxyl group may be storage and/or inactivation products. The corresponding esters, especially with the fatty acids 16 : 0, 18 : 0, 18 : 1 or 18 : 2, have been observed in insects and also in ticks, spiders, scorpions and millipedes [22, 23, 65, 132]. Ecdysteroids in the haemolymph are probably always transported bound to proteins. In contrast to the **transport proteins** of the juvenile hormones, those for the moulting hormones have been relatively little investigated. In the fly *Calliphora vicina*, the ecdysteroids are bound to the larval haemolymph protein calliphorin. A haemolymph protein that binds 20-hydroxyecdysone in the migratory locust *Locusta migratoria* has been purified to a high degree. It has a mass of 270–280 kDa and is probably a dimer; it binds 20-hydroxyecdysone with a binding constant of about 10^{-7} mol/l but has a 100-fold lower affinity for ecdysone [17].

Surprisingly large amounts of ecdysone, 20-hydroxy-ecdysone and ponasterone A were found in plants as early as 1966. The subsequent systematic investigations have led so far to the identification of about 70 biologically active plant ecdysteroids. Such **"phytoecdysteroids"** are extraordinarily widespread in the plant kingdom; they are found in half of all fern families, almost all families of the gymnosperms, and approximately every fifth family of the angiosperms. The biochemical prerequisites for biological activity are the 6-oxo-7-en structure, the cis association of the A and B rings and a complete side-chain. The hydroxyl groups on C-14 are more important than those on C-3 and C-22, whilst those on C-2 and C-25 appear to have no importance for moulting hormone activity. Many of the phytoecdysteroids are C_{28} or C_{29} steroids with an alkyl group on C-24. The concentrations of ecdysteroids in plants are usually much higher than in arthropods, being as a rule between 10 mg/kg and 1 g/kg dry weight and sometimes exceeding 20 g/kg. The biological significance of the phytoecdysteroids is probably mainly defence against phytophagous insect larvae. There appears to have been coevolution of the regulation of moulting in insects and defence mechanism of plants [27].

In contrast, the importance of the **ecdysteroids of the non-arthropod invertebrates**, i.e. cnidarians, platyhelminths, nematodes, annelids and molluscs, is a complete mystery. Representatives of these groups have been found to contain, for example, ecdysone, 2-deoxyecdysone, 20-hydroxyecdysone, 20,26-dihydroxyecdysone and ecdysone-24-0-β-glucoside. The concentrations

are usually lower than those in the arthropods [79, 82, 85]. A notable exception is the Mediterranean anemone *Gerardia savaglia*, which has about 3 g 20-hydroxyecdysone per kg dry weight [42].

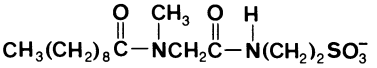


Fig. 16.15. Decanoyl-sarcosyl-taurine, an emulsifier from the crab *Cancer pagurus* [46]

16.5 Bile Acids and Bile Alcohols

The liver of all vertebrates produces a secretion, bile, which in many species is temporarily stored in a gall bladder before it is released into the first section of the small intestine, the duodenum. Bile contains steroids which as bile alcohols carry several hydroxyl groups and as bile acids carry an additional carboxyl group. In native bile, the bile alcohols are always esterified with sulphate and the bile acids are conjugated with taurine or glycine. Both steroid types are therefore present as anions and are referred to collectively as **bile salts**. Because the polar groups of the bile salts lie exclusively, or predominantly, on one side of the flat molecule (Fig. 16.16a), they have marked amphiphilic properties. Hence, bile salts can carry out several **important** functions in **fat digestion and absorption:**

1. They participate in the emulsification of fats in the small intestine and stabilize the fat emulsion by orientating themselves on the fat droplets with their hydrophilic side facing outwards.
2. They stimulate lipases and cholesterol esterases.
3. They allow the formation of the mixed micelles required for the effective transport of lipolysis products through the hydration layer of the microvillous border of the gut epithelium.

In fact, triacylglycerols and phospholipids can be hydrolysed and resorbed in the gut lumen without bile, albeit at about one-third of the efficiency as with bile, but cholesterol and most of the fat-soluble vitamins can only be resorbed with the participation of bile [25].

Bile acids and bile alcohols are found only in vertebrates. However, invertebrates do make use of **biodetergents** that are involved in a similar way in fat digestion and resorption, but which have completely different chemical structures. In the crab *Cancer pagurus* compounds made up of fatty acids, sarcosine and taurine exercise such a function (Fig. 16.15). Emulsifying compounds with fatty acids and amino acids have also been detected in other arthropods, e.g. sarcosine-free acyltaurine in the crayfish *(Astacus, Orconectes)*, and acyl derivatives of various amino acids in spiders *(Tegenaria, Cupiennius)* and insects *(Gryllus)*. A completely different type of emulsifier has been found in the edible snail *Helix pomatia;* hydrolysis of this emulsifier gives sulphate, saturated fatty alcohols (C_{11} to C_{16}), glycerol and fatty acids. The emulsifying activity is apparently based upon the sulphate residues bound to the fatty alcohols and glycerol (123).

The **basic structural elements** of the bile alcohols are 5α- or 5β-cholestan (C_{27}), 27-norcholestan (C_{26}), 26,27-dinorcholestan (C_{25}) and cholan (C_{24}) (Fig. 16.16b). The evolutionarily more ancient C_{27} bile acids are derived from 5α- and 5β-cholestan, and the C_{24} acids are derived from 5α- and 5β-cholan. Unconjugated C_{28} bile acids have only been found in some anurans; they carry a methyl or carboxyl group on C-24 (Fig. 16.16). 5α(Allo)-bile acids make up only a few percent of the total in mammals but are the main bile components in some lower vertebrates. Four basic types can be distinguished amongst the conjugated bile alcohols and acids: (1) sulphate esters of the C_{27} bile alcohols are found in the agnathans to the amphibians; (2) the more primitive C_{27} bile acids are always conjugated with taurine and are widespread in reptiles and amphibians, and are present in some fish; and (3–4) the C_{24} bile acids represent the latest stage of bile acid evolution. They are missing from the agnathans, cartilaginous fish and amphibians at the root of the evolutionary progression, but are typical of the derived, more highly developed groups of teleosts, lizards, snakes, birds and mammals. In most groups they are conjugated with taurine; only the higher mammals (Eutheria) also have glycine-conjugated C_{24} bile acids [46].

16.5.1 Biosynthesis of Bile Acids and Bile Alcohols

The **precursor** of the bile salts is cholesterol; at least in the mammals, the major pathway of cholesterol degradation in the liver is conversion to bile acids. About 1 g bile salts is produced daily in humans, which have a pool of 3–5 g. However, as the majority of bile acids in the gut are reab-

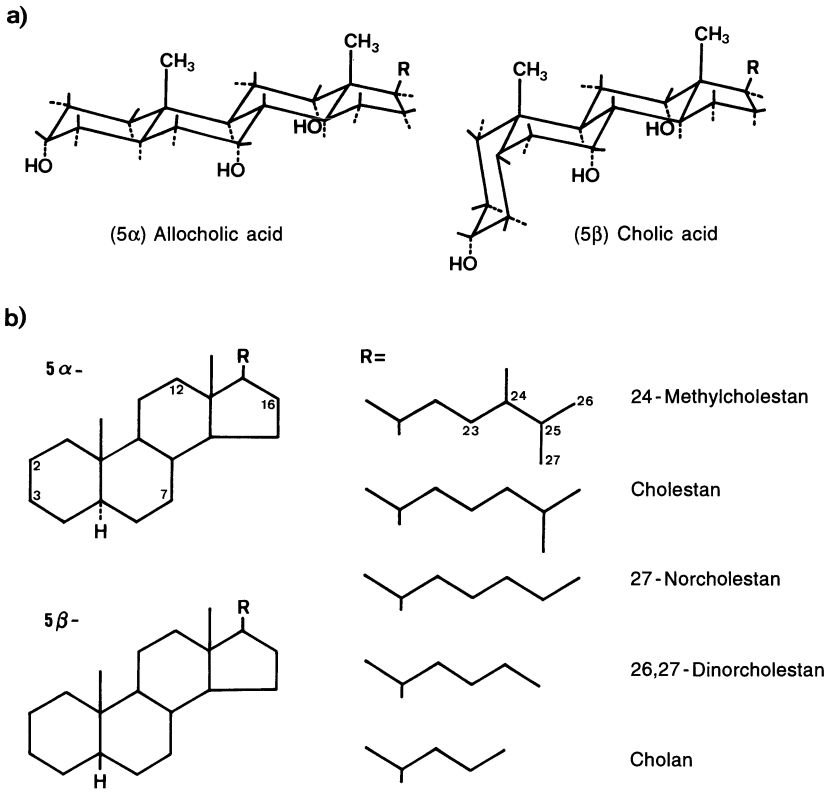


Fig. 16.16a, b. Bile alcohols and bile acids [25, 46]. **a** Spatial structure of the 5α and 5β bile acids, illustrated with (5α) allocholic acid (3α,7α,12α-trihydroxy-5α-cholanic acid) and (5β) cholic acid (3α,7α,12α-trihydroxy-5β-cholanic acid (R = $CH(CH_3)CH_2CH_2COOH$). **b** The ring systems and side-chains of bile acids and alcohols: the possible substitution sites are numbered. For clarity, the conjugation residues have been omitted. However, in the gall bladder all alcohols carry a sulphate residue on one of the primary OH groups of the side-chain; myxinol bissulphate has sulphate residues on C-3 and on C-26 (or C-27). All bile acids are conjugated with taurine (or glycine).

The trivial and systematic names of all bile acids and alcohols mentioned in the text:

Bile alcohols

Arapaimol B	=	5β-Cholestan-2β,3α,7α,12α, 26,27-hexol
Scymnol	=	5β-Cholestan-3α,7α,12α,24,26,27-hexol
5α-Dermophol	=	5α-Cholestan-3α,7α,12α,25,26,27-hexol
5β-Dermophol	=	5β-Cholestan-3α,7α,12α,25,26,27-hexol
Arapaimol A	=	5β-Cholestan-2β-3α,7α,12α,26-pentol
5α-Chimaerol	=	5α-Cholestan-3α,7α,12α,24,26-pentol
5β-Chimaerol	=	5β-Cholestan-3α,7α,12α,24,26-pentol
5α-Bufol	=	5α-Cholestan-3α,7α,12α,25,26-pentol
5β-Bufol	=	5β-Cholestan-3α,7α,12α,25,26-pentol
5α-Cyprinol	=	5α-Cholestan-3α,7α,12α,26,27-pentol (Fig. 16.17)
Latimerol	=	5α-Cholestan-3β,7α,12α,26,27-pentol
26-Deoxy-5α-chimaerol	=	5α-Cholestan-3α,7α,12α,24-tetrol
27-Deoxy-5α-cyprinol	=	5α-Cholestan-3α,7α,12α,26-tetrol
Myxinol	=	5α-Cholestan-3β,7α,16α,26-tetrol
5α-Ranol	=	(24R)-27-Nor-5α-cholestan-3α,7α,12α,24,26-pentol
5β-Ranol	=	(24R)-27-Nor-5β-cholestan-3α,7α,12α,24,26-pentol
5α-Petromyzonol	=	5α-Cholan-3α,7α,12α,24-tetrol

Bile acids

Chenodeoxycholic acid	=	3α,7α-Dihydroxy-5β-cholanic acid (Fig. 16.17)
Deoxycholic acid	=	3α,12α-Dihydroxy-5β-cholanic acid (Fig. 16.17)
Hyodoexycholic acid	=	3α,6α-Dihydroxy-5β-cholanic acid
Lithocholic acid	=	3α,7α-Hyxdroxy-5β-cholanic acid (Fig. 16.17)
α-Muricholic acid	=	3α,6β,7α-Trihydroxy-5β-cholanic acid
β-Muricholic acid	=	3α,6β,7β-Trihydroxy-5β-cholanic acid
ω-Muricholic acid	=	3α,6α,7β-Trihydroxy-5β-cholanic acid
Phocaecholic acid	=	3α,7α,23-Trihydroxy-5β-cholanic acid
Pythocholic acid	=	3α,12α,16α-Trihydroxy-5β-cholanic acid
Ursodeyxycholic acid	=	3α,7β-Dihydroxy-5β-cholanic acid
Varanic acid	=	3α,7α,12α,24-Tetrahydroxy-5β-cholestanic acid

sorbed and recycled into bile **(entero-hepatic circulation)**, about 20–30 g bile acids are released into the gut each day. The bile acids may be chemically modified by gut bacteria during their residence in the gut lumen, and may be further metabolized in the liver after resorption. The metabolites produced by this combination of bacterial and animal metabolism are referred to as **secondary bile acids**. For example, the deoxycholic acid which is commonly found in higher vertebrates arises from the cleavage by gut bacteria of the 7α-hydroxyl group of cholic acid [46].

The **biosynthesis** of various bile acids and bile alcohols from cholesterol has been demonstrated by tracer experiments for all classes of vertebrates. The synthesis of cholic acid from cholesterol initially requires the introduction of hydroxyl groups into the 7α and 12α positions, the epimerization of the 3β-OH to 3α-OH, and the reduction of the Δ^5 double bond. The intermediates formed are 7α-hydroxy-4-cholesten-3-one and 5β-cholestan-3α,7α,12α-triol. Finally, a part of the side-chain on C-24 is cleaved off as propionyl-CoA (Fig. 16.17). The biosynthesis of the bile alcohol 5α-cyprinol proceeds analogously via 5α-cholestan-3α,7α,12α-triol. The numerous tracer experiments on vertebrates of all classes and involving different precursors or intermediates have resulted in a confusing picture of the metabolism of bile salts; the same product can apparently be produced via different pathways from the same precursor, but the same precursor also yields different products depending upon the species and the physiological condition of the experimental animal. A multitude of species-specific differences is even found in the metabolism of the bile alcohols and bile acids, and this is probably based upon differences in enzyme complement. For example, some amphibians and all crocodiles lack the ability to shorten the cholesterol side-chain, and certain hydroxylations are restricted to individual groups or species [25, 46]. The **enzymes** involved in the biosynthesis of bile salts have also been only partly described in the mammals, and there are almost no comparative data. The 7α- and 12α-hydroxylases are microsomal cytochrome P-450 enzymes, but the 27-hydroxylase is located in the mitochondria [15]. Prior to conjugation, bile acids must be activated by the bile acid : CoA ligase. This is the rate-limiting step of bile acid conjugation. The ligase isolated from the liver of the codfish *Ophiodon elongatus* has a much lower affinity than the mammalian enzyme and is present in lower amounts [122]. The bile acid-CoA : taurine/glycine acyltransferases from the chicken and bovine liver have been isolated. The taurine-specific enzyme of the chicken has a mass of 47–51 kDa and is significantly shorter than the glycine/taurine-specific bovine enzyme of 63–65 kDa. It may be that the ability of the bovine enzyme to bind glycine as well as taurine is related to the extension of the polypeptide chain by about 125 amino acids. The transferases from non-mammals are highly specific for taurine. The liver enzyme of the teleost *Ophiodon* metabolizes glycine with a 100-fold lower efficiency than taurine. This teleost liver enzyme is a dimer of 100 kDa, whereas the enzymes from mammals and birds are monomers of 50 and 60 kDa, respectively [122].

16.5.2 Bile Salts of Individual Animal Groups[1]

The gall bladders of **agnathans** in the genera *Myxine* and *Eptatretus*, which are considered to be the most primitive vertebrates, mainly contain myxinol bissulphate. This is also the most primitive bile alcohol that still includes the complete cholesterol basic element (5α-cholestan) and a 3β-hydroxyl; it is the only double sulphated bile alcohol and is much less effective as a detergent than any other bile alcohol. The predominant bile salt of the higher agnathan *Petromyzon marinus* is 5α-petromyzonol sulphate; this is much more developed and has a shortened side-chain and a 3α-hydroxyl group. The **cartilaginous fish** also contain unique bile salts; the 5β-chimaerol[2] from *Chimaera monstruosa* differs from the scymnol, which is common in sharks and rays, by the absence of the 27-OH. Cholic acid is also produced from radiolabelled cholesterol in the dogfish *Scyliorhinus canicula* but is normally present only in trace amounts [46, 61]. The latimerol found in the **crossopterygian** *Latimeria chalumnae*, one of the oldest bony fish, has a primitive 3β-hydroxyl group and appears to be the evolutionary precursor of 5α-cyprinol. In fact, in addition to latimerol, the crossopterygians have the related 5α-cholestan-3β,7α,12α,25,26-pentol and 5α-cyprinol, as well as traces of different C_{27}, C_{26} and even C_{24} bile acids [62].

---

1 The structures of all the bile alcohols and bile acids mentioned in the text are given in Fig. 16.16.

2 As for all bile alcohols, these are also found in the gall bladder as sulphate esters; for the sake of brevity, "sulphate" is omitted from the names of the bile alcohols in the following text.

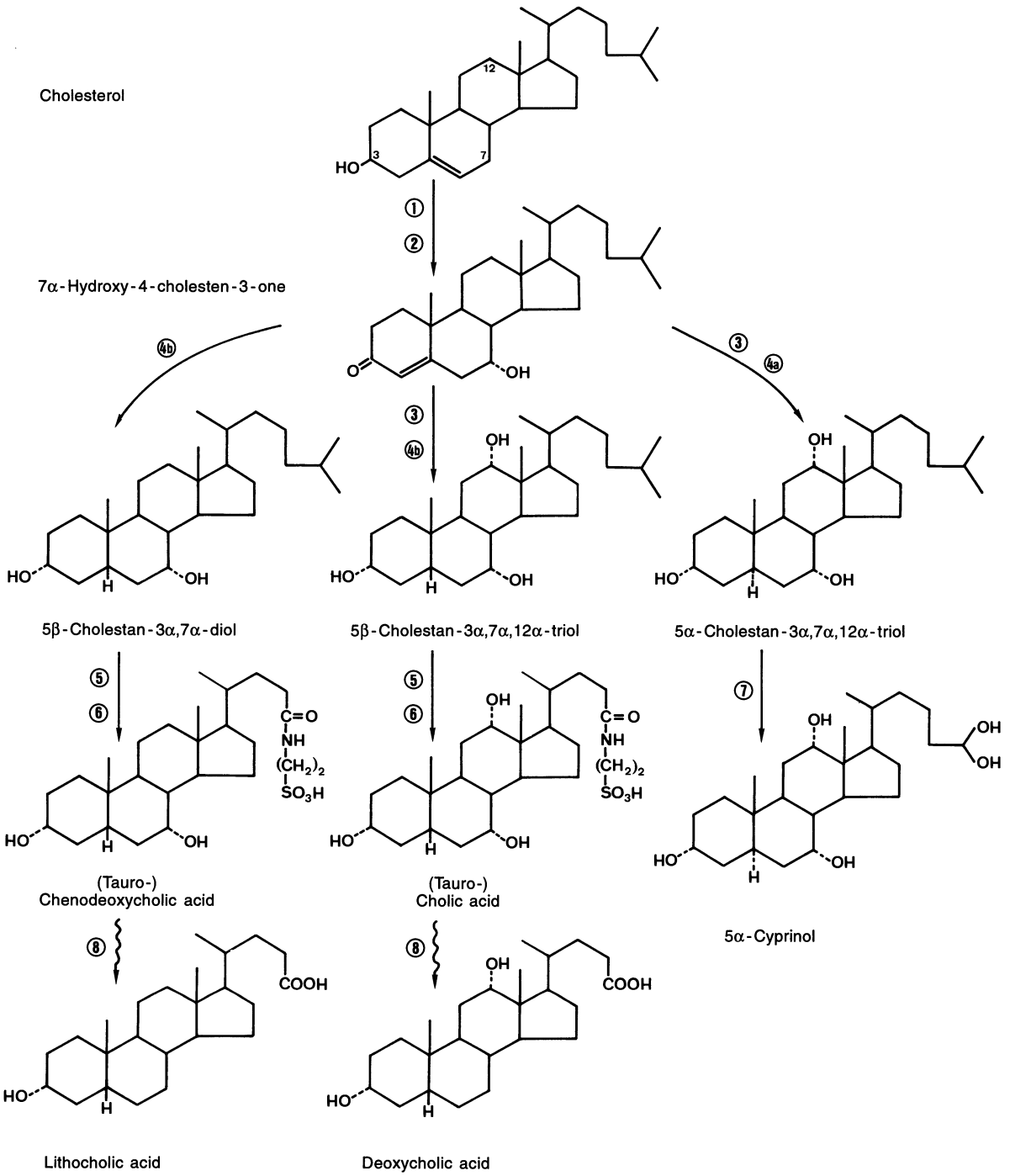


Fig. 16.17. The biosynthesis of bile acids and alcohols [25]. The taurine conjugation is shown only for chenodeoxycholic acid and cholic acid. *1*, 7α-Hydroxylation; *2*, oxidation at C-3 and Δ^5/Δ^4 isomerization; *3*, 12α-hydroxylation; *4a*, Δ^4-3-oxosteroid-5α reduction; *4b*, Δ^4-3-oxosteroid-5β reduction; *5*, oxidative cleavage of the side-chain at C-24 as propionyl-CoA; *6*, conjugation with taurine (or glycine); *7*, 26- and 27-hydroxylation; *8*, formation of secondary bile acids by gut bacteria

The **lung fish (Dipnoi)** are distributed over three continents; *Lepidosiren paradoxa* is found in South America, *Neoceratodus fosteri* exists in Australia, and the four species of *Protopterus* are found in Africa. Nevertheless, the bile salts of the Dipnoi are all very similar. 27-Deoxy-5α-cyprinol is found in all three, 5α-cyprinol in *Protopterus* and *Lepidosiren*, and 5α-bufol in *Lepidosiren* and

Neoceratodus; 5α-chimaerol and 26-deoxy-5α-chimaerol are restricted to *Neoceratodus*. The large proportion of 5α-bufol in *Lepidosiren* is an indication of the relationship between the Dipnoi and the amphibians. The primitive **bony fish**, such as *Acipenser, Polyodon, Amia, Lepisosteus, Polypterus* and *Calamoichthys*, contain predominantly taurocholic acids and only traces of bile alcohols. In contrast, their derived evolutionary descendants, the teleosts, exhibit all possible variations of the bile salt spectrum from the most primitive, which consists mainly of 5α-cyprinol (Cyprinidae, several Catastomidae) or 5α-chimaerol (several Catastomidae) and only traces of taurine-conjugated C_{24} bile acids, to those which consist almost exclusively of such bile acid conjugates, e.g. the catastomid *Carpoides carpio* and the gymnotid *Electrophorus electricus*. The giant Amazonian fish *Arapaima gigas* (Osteoglossidae) is unique in possessing arapaimol A and B in addition to taurocholate [46].

The **amphibians** also possess characteristic bile salts. For example, the gymnophion *Dermophis mexicanus* has 5β-dermophol, the urodelans have 5α-dermophol, 5α-bufol or 5α-cyprinol and, in the genus *Salamandra*, 5α-ranol is also present. To date, a total of 14 different bile alcohols has been identified in the anurans; these include the particularly common 5α- and 5β-ranol and 5β-bufol. Non-conjugated C_{28} bile acids have been identified in the American bullfrog *Rana catesbeiana* and the toad *Bufo bufo formosa*. No bile alcohols are found from the **reptiles** onwards, only C_{27} or C_{24} bile acids; however, the alcohols may appear under pathological conditions, even in man. The bile acids of the reptiles are all taurine conjugated. The gall bladder of the crocodile has mainly 3α,7α,12α-trihydroxycoprostanic acid, which is an intermediate in the synthesis of cholic acid. Thus, the crocodile lacks the ability to shorten the side-chain. The turtles have a C_{27} bile acid, tetrahydroxy-sterocholanic acid, which carries a unique 22-hydroxyl group, but they apparently also have the capacity to produce C_{24} bile acids. The lizards normally synthesize cholic acid or allocholic acid; only the Varanidae are unusual in having a C_{27} bile acid, varanic acid [1]. The snakes, of which more than 80 species have been investigated, also contain mainly 5β-C_{24} bile acids. The pythocholic acid of the boas and pythons is a secondary bile acid. This arises by conversion of the cholic acid, secreted from the liver, to deoxycholic acid by gut bacteria; this is further 16α-hydroxylated to pythocholic acid by the liver following reabsorption. The capacity for 16α-hydroxylation is found only in this group of primitive snakes, and the 23-hydroxylation of bile acids, which occurs in some species of the Viperidae and Colubridae, is also highly characteristic [46].

Chenodeoxycholic acid was for a long time regarded as the typical bile acid of the **birds** but, in fact, cholic acid and ursodeoxycholic acid are also found. β-Phocaecholic acid, previously known from only the seal, has additionally been found in large quantities in the duck *Anas platyrhynchos* [55]. The **mammals** have almost exclusively 5β-C_{24} bile acids which are conjugated with taurine or glycine. Taurine-conjugated trihydroxy acids, in particular cholic acid, normally predominate in carnivorous mammals, and glycine-conjugated dihydroxy acids, especially chenodeoxycholic acid and deoxycholic acid, predominate in the herbivorous mammals, although some trihydroxy acids are also found in the herbivores. Glycine-conjugated bile acids are found only in the higher mammals (Eutheria); in the cloacal (monotremes) and pouched (marsupial) mammals, the usual three bile acids are conjugated with taurine. Whereas in the primates the three bile acids are bound partly to glycine and partly to taurine, the rabbit has almost exclusively glycine-conjugated deoxycholic acid. Amongst the rodents, the beaver *Castor canadensis* has the typical herbivorous spectrum with glycine-conjugated chenodeoxycholic acid as the main component. However, the rat and the mouse have cholic acid and also the typical α- and β-muricholic acid, all mainly conjugated with taurine. In addition, the mouse has ω-muricholic acid, a secondary bile acid of unknown origin. The three typical bile acids of the Carnivora are found predominantly in the taurine-conjugated form. In contrast to the meat-eating bears, the vegetarian bears have glycine-conjugated dihydroxy acids such as chenodeoxycholic acid or ursodeoxycholic acid. In addition to the usual taurine-conjugated bile acids, the marine mammals (Pinnipedia) have a characteristic acid, phocaecholic acid [47]. The Artiodactyla have both glycine-conjugated chenodeoxycholic acid and hyodeoxycholic acid, which may be either primary or secondary bile acids. Thus, chenodeoxycholic acid is converted to hyocholic acid by 6α-hydroxylation in the pig liver, and hyodeoxycholic acid is then produced by cleavage of the 7α-hydroxyl group by gut bacteria; 6α-hydroxylation is a speciality of the pig (Suinae). The usual three bile acids predominate in the ruminants but, in contrast to those found in other herbivorous mammals, they are partly taurine conjugated [46].

16.6 Calciferols

The calciferols are referred to as D vitamins because feeding of these compounds or their sterol precursors alleviates certain deficiency symptoms (rachitis). They are, however, perhaps better categorized as hormones that regulate the calcium balance. This activity, as for that of the classical steroid hormones, involves binding to intracellular receptors and effects on the expression of certain genes. The **biosynthesis of cholecalciferol** (vitamin D_3) occurs as follows: cholesterol is oxidized in the skin to Δ^7-dehydrocholesterol; the bond between C-9 and C-10 is broken by UV irradation; the resulting 9,10-secosteroid (pre-vitamin D_3) converts spontaneously into cholecalciferol (Fig. 16.18). Because the vitamin-D-binding protein in the blood plasma has a greater affinity for vitamin D_3 than for the pre-vitamin or for 7-dehydrocholesterol, the cholecalciferol is selectively transported from the skin by the blood circulation. The analogous ergocalciferol (vitamin D_2) is produced from exogenous ergosterol. The photochemical formation of secosterols, which additionally occurs to a considerable extent in phytoplankton, also functions as a protection against injurious UV irradiation [83].

Cholecalciferol and ergocalciferol are hydroxylated to 25-hydroxycalciferol by a microsomal, NADPH-dependent cytochrome P-450 system, mainly in the liver, although calciferol-25-hydroxylase is also found in the microsomes of the lung and kidney. The further **metabolism of the D vitamins** takes place mainly in the kidney cortex, where the 25-hydroxy-calciferol is converted by hodroxylation at C-1, C-23, C-24 or C-26 into various dihydroxy or trihydroxy metabolites. Two of the metabolites produced in the kidney are the **biologically active hormones** 1α,25-dihydroxycalciferol and 24,25-dihydroxycalciferol. These active substances can also be produced extrarenally; 25-hydroxy-calciferol-1-hydroxylase and 24-hydroxylase have been detected in the bone cells of mammals and birds and in the placenta. The situation is especially complex in pregnant animals, where not only the maternal kidney and placenta but also the kidney and other organs of the embryo are involved in the formation of the two active calciferol metabolites; the placenta allows transport of 25-hydroxycalciferol from the mother to the foetus but prevents the return transport of the dihydroxycalciferols. 25-Hydroxycalciferol-1-hydroxylase is a mitochondrial cytochrome P-450 oxygenase that requires ferredoxin and NADPH. It is regulated by the hormones of the parathyroid gland and other hormone glands, as well as by the concentrations of calcium and phosphate ions. 1α,25-Dihydroxycalciferol inhibits the 25-hydroxycalciferol-1-hydroxylase (product inhibition) but stimulates the 25-hydroxycalciferol-24-hydroxylase.

The conversion of 25-calciferol to 1α,25-dihydroxycalciferol and 24,25-dihydroxycalciferol is apparently possible in all vertebrates. These processes have even been detected in the common hagfish *Eptatretus stoutii* and in various sharks and rays, which have no bones [83]. It may

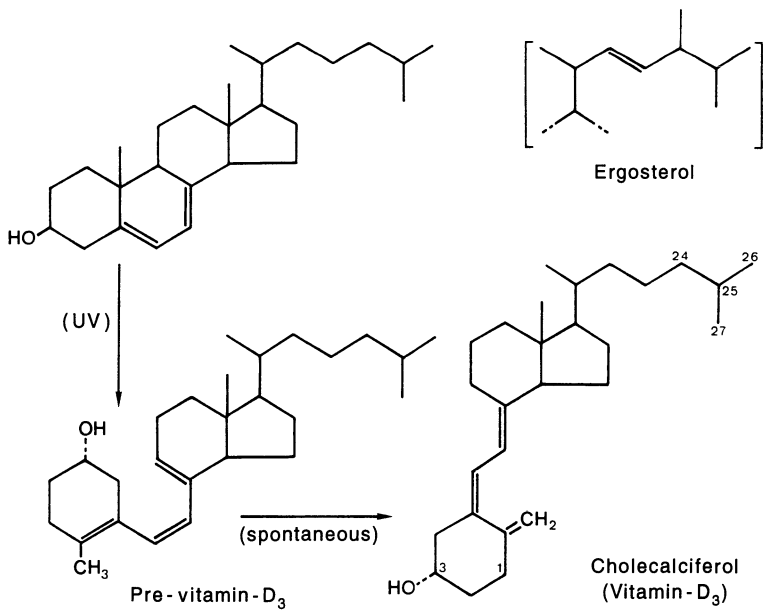


Fig. 16.18. The synthesis of calciferol [83]

be that the calciferol metabolites are involved in calcium homeostasis, maintaining, for example, the calcium concentration in the blood plasma of the shark at a level one-third that of the surrounding seawater. At a concentration of 20–80 μg/l, 25-hydroxycalciferol is the major component in the blood plasma of all birds and mammals examined; this is followed by calciferol, 24,25- and 25,26-dihydroxycalciferol at 1–10 μg/l. The concentration of 1α,25-dihydroxycalciferol is one order of magnitude lower at 0.02–0.06 μg/l. Cholecalciferol (vitamin D_3) and its metabolites are almost always found at concentrations 20- to 100-fold higher than those of ergocalciferol (vitamin D_2); the sheep appears to be an exception in which the concentrations are more-or-less equal [83]. Cholecalciferol and ergocalciferol are equally effective in humans, Old World apes, cattle and rats. In the New World apes, cholecalciferol is clearly superior and must be considered to be the active substance in all other vertebrates; for example, it is 100-fold more effective than ergocalciferol in the chicken. In agreement with this is the observation that 25-hydroxyergocalciferol is less tightly bound than 25-cholecalciferol in the blood plasma of fish, reptiles, birds and monotremes, but in the Placentalia binding is more-or-less equal. The calciferol metabolites in the blood plasma are bound to a specific **vitamin-D-binding protein** (DBP). The DBP of the mammals belong to the α-globulins; these are monomeric glycoproteins of 51–55 kDa with 1 % carbohydrate which, according to the N- and C-terminal sequences, constitute their own protein family [93].

The two dihydroxycalciferols apparently have different **functions** and are not completely interchangeable. Chickens or Japanese quails containing only 1α,25-dihydroxycalciferol have eggs that develop normally but do not hatch; however, if 24,25-dihydroxycalciferol is added to the diet, the hatching rate is the same as with calciferol itself. Similarly, skeleton development in young chickens is normal only when both forms are available. The two forms apparently affect different cell populations of the bone. The osteoclasts, responsible for the absorption of bone and calcium mobilization, are stimulated by 1α,25-dihydroxycalciferol, and the cartilage cells and osteoblasts, responsible for bone formation, are stimulated by 24,25-dihydroxycalciferol. 1α,25-Dihydroxycalciferol additionally stimulates gut resorption of calcium and phosphate ions, as well as calcium reabsorption in the kidney tubules. In summary, 1α,25-dihydroxycalciferol appears to be responsible in particular for calcium homeostasis, and 24,25-dihydroxycalciferol is responsible for bone formation and mineralization [83].

Calciferol receptors are found at all active sites, i.e. in the calcium-transporting cells of the gut, kidney, bone and milk glands, and also in hormone glands whose hormones influence calciferol hydroxlases, e.g. the parathyroid, hypophysis and pancreas. This distribution has been shown in different mammals and birds, and also in lower vertebrates such as the eel *Anquilla anquilla*, where 1α,25-calciferol receptors have been found in the gut, gills, hypophysis, Stannius bodies, liver and brain, but not in the kidneys, muscle or skin. Different receptors that are specific for 1α,25- or 24,25-dihydroxycalciferol have been isolated from the parathyroid of the chicken [83]. 1α,25-Dihydroxycalciferol receptors from various mammals and the chicken have been sequenced via the cDNA. They are polypeptides of about 60 kDa that bind 1α,25-dihydroxycalciferol selectively with high affinity and then interact with the DNA. The cysteine-rich DNA-binding domains in particular are very similar to the corresponding regions of other steroid receptors and the erb-A oncogene [13, 77].

The **calciferol balance** in the fish appears to differ in several important respects from that in the birds and mammals. Although the photochemical formation of cholecalciferol is hardly significant, because of the UV-absorbing nature of water, many fish nevertheless contain large amounts of this substance, and are important sources of calciferol. This is also true of the cartilaginous fish, in which the skeleton is not calcified. Fish blood plasma also contains a vitamin D-binding α-globulin. Although fish liver, gut and kidney have been shown to contain cytochrome P-450, the liver microsomes do not produce polar metabolites other than 25-hydroxycalciferol from calciferol; the kidney metabolizes neither 25-hydroxycalciferol nor calciferol [130]. It is not clear whether 25- and 1α-hydroxylation are possible in invertebrates. There are earlier results showing the formation of 25-hydroxycalciferol in terrestrial snails, but the present conclusion is that neither 25-hydroxycalciferol nor 1α,25-dihydroxycalciferol have been unambiguously demonstrated in either the gastropods, bivalves or echinoids [70].

References

1. Ali S. S., Stephenson E. and Elliott W. H.: Bile acids LXVII. The major bile acids of Varanus monitor. J. Lipid Res. 23: 947–954 (1982)
2. Anderson D. G.: Gorgosterol biosynthesis: Localization of squalen formation in the zooxanthellar component of various gorgonians. Comp. Biochem. Physiol. Pt. B 81: 423–428 (1985)
3. Barbier M.: Introduction to chemical ecology. Longman, Harlow UK 1979
4. Basson M. E. et al.: Structural and functional conservation between yeast and human 3-hydroxy-3-methylglutaryl coenzyme A reductase, the rate limiting enzyme of sterol biosynthesis. Mol. cell. Biol. 8: 3797–3808 (1988)
5. Bauermeister A. et al.: Distribution and some properties of bile salt-binding proteins in rainbow trout Salmo gairdneri. Comp. Biochem. Physiol. Pt. B 78: 195–198 (1984)
6. Beato M.: Induction of transcription by steroid hormones (Review). Biochim. biophys. Acta 910: 95–102 (1987)
7. Billheimer J. T., Tavani D. M. and Ritter K. S.: Acyl coenzyme A:cholesteryl acyltransferase activity in fat body and intestinal microsomes of Heliothis zea. Comp. Biochem. Physiol. Pt. B 76: 127–132 (1983)
8. Blanchier B., Boucaud-Camou E. and Silberzahn P.: Comparative study of the sterol composition of the digestive gland and the gonad of Sepia officinalis L. (Mollusca, Cephalopoda). Comp. Biochem. Physiol. Pt. B 83: 599–602 (1986)
9. Bourne A. R.: Occurrence of 20ß-hydroxysteroid oxidoreductase activity in the kidney and liver of the lizard Tiliqua rugosa. Comp. Biochem. Physiol. Pt. B 77: 221–222 (1984)
10. Bradbrook D. A. et al.: The occurrence of vertebrate type steroids in insects and comparison with ecdysteroid levels. Comp. Biochem. Physiol. Pt. B 95: 365–374 (1990)
11. Brooks A. R., Sweeney G. and Old R. W.: Structural and functional expression of a cloned Xenopus thyroid hormone receptor. Nucleic Acids Res. 17: 9395–9405 (1989)
12. Bückmann D. et al.: Isolation and identification of major ecdysteroids from the pycnogonid Pycnogonum litorale (Ström) (Arthropoda, Pantopoda). J. comp. Physiol. B 156: 759–765 (1986)
13. Burmester J. K. et al.: Structure and regulation of the rat 1,25-dihydroxyvitamin D3 receptor. Proc. Nat. Acad. Sci. USA 85: 9499–9502 (1988)
14. Burnell D. J., Simon J. W. A. and Gilgan M. W.: Occurrence of saponins giving rise to asterone and asterogenol in various species of starfish. Comp. Biochem. Physiol. Pt. B 85: 389–391 (1986)
15. Cali J. J. and Russel D. W.: Characterization of human sterol 27-hydroxylase, a mitochondrial cytochrome P-450 that catalyzes multiple oxidation reactions in bile acid biosynthesis. J. Biol. Chem. 266: 7774–78 (1991)
16. Callard G. V. and Mak P.: Exclusive nuclear location of estrogen receptors in Squalus testis. Proc. Nat. Acad. Sci. USA 82: 1336–40 (1985)
17. Cao M. et al.: Purification of a 20-hydroxyecdysone binding protein from haemolymph of the locust Locusta migratoria. Insect Biochem. 13: 567–575 (1983)
18. Catalan C. A. N. et al.: Minor and tryce sterols in marine invertebrates (39.). Steroids 40: 455–463 (1982)
19. Catalan C. A. N. et al.: Biosythetic studies of marine lipids – 3. Experimental demonstration of the course of side chain extension in marine sterols. Tetrahedron 41: 1073–84 (1985)
20. Chen H. and Shapiro D. J.: Nucleotide sequence and estrogen induction of Xenopus laevis 3-hydroxy-3-methylglutaryl-coenzyme A reductase. J. Biol. Chem. 265: 4622–29 (1990)
21. de Clerck D. et al.: Study of the metabolism of steroids in the larvae of the fleshfly Sarcophaga bullata. Comp. Biochem. Physiol. Pt. B 87: 821–826 (1987)
22. Connat J. L., Diehl P. A. and Thompson M. J.: Possible inactivation of ingested ecdysteroids by conjugation with long-chain fatty acids in the female tick Ornithodorus moubata (Acarina: Argasidae). Arch. Insect Biochem. Physiol. 3: 235–252 (1986)
23. Crosby T. et al.: Identification of ecdyson 22-long-chain fatty acyl esters in newly laid eggs of the cattle tick Boophilus microplus. Biochem. J. 240: 131–138 (1986)
24. Daloze D., Braekman J. C. and Pasteels J. M.: New polyoxygenated steroidal gluyosides from Chrysolina hyperici (Coleoptera: Chrysomelidae). Tetrahedron Letters 26: 2311–14 (1985)
25. Danielsson H. and Sjovall J. (eds.): Sterols and bile acids. New comprehensive biochemistry Vol. 12. Elsevier, Amsterdam 1985
26. Dini A. et al.: Sterol composition of marine sponges Stryphnus ucronatus and Reniera sarai. Comp. Biochem. Physiol. Pt. B 81: 111–114 (1985)
27. Downer R. G. H. and Laufer H. (eds.): Endocrinology of insects. Alan R. Liss, New York 1983
28. Evans R. M.: The steroid and thyroid hormone receptor superfamily. Science 240: 889–895 (1988)
29. Fairs N. J. et al.: Detection of unconjugated and conjugated steroids in the ovary, eggs, and haemolymph of the decapod crustacean Nephrops norvegicus. Gen. comp. Endocrinol. 74: 199–208 (1989)
30. Feldlaufer M. F. et al.: Biosynthesis of makisterone A and 20-hydroxyecdysone from labeled sterols by the honey bee, Apis mellifera. Arch. Insect Biochem. Physiol. 3: 415–421 (1986)
31. Feldlaufer M. F. et al.: The neutral sterols of Megalotomus quinquespinosus (Say) (Hemiptera: Alydidae) and identification of makisterone A as the major free ecdysteroid. Arch. Insect Biochem. Physiol. 3: 423–430 (1986)
32. Feldlaufer M. F. et al.: Fate of maternally-acquired ecdysteroids in unfertilized eggs of Manduca sexta. Insect Biochem. 18: 219–221 (1988)
33. Findlay J. A. and Patil A. D.: A novel sterol peroxide from the sea anemone Metridium senile. Steroids 44: 261–265 (1985)
34. Forest M. G. and Pugeat M. (eds.): Binding proteins of steroid hormones. Jon Libbey Eurotext, Montrouge 1986
35. Fournier B. and Radallah H.: Ecdysteroids in Carausius eggs during embryonic development. Arch. Insect Biochem. Physiol. 7: 211–224 (1988)
36. Garneau F. X. et al.: The distribution of asterosaponins in various body components of the starfish Leptasterias polaris. Comp. Biochem. Physiol. Pt. B 92: 411–416 (1989)
37. Goad L. J.: Sterol biosynthesis and metabolism in marine vertebrates. Pure appl. Chem. 53: 837–852 (1981)

38. Goad L. J. et al.: Composition of free, esterified and sulphated sterols from the sea cucumber Psolus fabricii. Comp. Biochem. Physiol. Pt. B 84: 189–196 (1986)
39. Gorbman A. (ed.): Comparative endocrinology. Wiley & Sons, New York 1983
40. Grieneisen M. L. et al.: A putative role to ecdysteroids. Metabolism of cholesterol invitro by mildly disrupted prothoracic glands of Manduca sexta. Insect Biochem. 21: 41–51 (1991)
41. Griffin P. R. et al.: The amino acid sequence of the sex steroid-binding protein of rabbit serum. J. Biol. Chem. 264: 19066–75 (1989)
42. Guerriero A. and Pietra F.: Isolation in large amounts of the rare plant ecdysteroid ajugasterone C from the Mediterranean zoanthid Gerardia savaglia. Comp. Biochem. Physiol. Pt. B 80: 277–278 (1985)
43. Habermehl G.: Toxic animals and their weapons (In German). 4th ed. Springer, Berlin 1987
44. Hammond G. L. et al.: Primary structure of human corticosteroid binding globulin, deduced from hepatic and pulmonary cDNAs, exhibits homology with serine protease inhibitors. Proc. Nat. Acad. Sci. USA 84: 5153–57 (1987)
45. Hammond G. L. et al.: The cDNA-deduced primary structure of human sex hormone-binding globulin and location of its steroid-binding domain. FEBS Letters 215: 100–104 (1987)
46. Haslewood G. A. D.: The biological importance of bile salts. North-Holland, Amsterdam 1978
47. Hellou J., King A. and Ni I. H.: Bile acids from the harp seals, Phoca groenlandica. Comp. Biochem. Physiol. Pt. A 89: 211–214 (1988)
48. Henrich V. C. et al.: A steroid/thyroid hormone receptor superfamily member in Drosophila melanogaster that shares extensive sequence similarity with a mammalian homologue. Nucleic Acids Res. 18: 4143–48 (1990)
49. Hoffmann J. A.: Ten years of ecdysone workshop. Insect Biochem. 16: 1–9 (1986)
50. Hoffmann K. H., Thiry E. and Lafont R.: 14-Deoxyecdysteroids in an insect (Gryllus bimaculatus). Z. Naturforsch. Sect. C 45: 703–708 (1990)
51. Itoh T., Sica D. and Djerassi C.: Minor and trace sterols in marine invertebrates, part 35. Isolation and structure elucidation of 74 sterols from the sponge Axinella cannabina. J. chem. Soc. Perkin Trans. II. 1983: 147–153
52. Jarzebski A. et al.: Sterol composition of Mesidothea entomon L. (Isopoda, Crustacea). Comp. Biochem. Physiol. Pt. B 81: 733–735 (1985)
53. Jarzebski A., Polak L. and Habermehl G.: Free and esterified sterols of Macoma balthica (L.). Comp. Biochem. Physiol. Pt. B 86: 561–563 (1987)
54. Jeltsch J. M. et al.: Characterization of multiple mRNAs originating from the chicken progesterone receptor gene. Evidence for a specific transcript encoding form A. J. Biol. Chem. 265: 3967–74 (1990)
55. Jirsa M. et al.: Classical bile acids in animals, β-phocaecholic acid in ducks. Comp. Biochem. Physiol. Pt. B 92: 357–360 (1989)
56. Johnson W. J. and Cain G. D.: Biosynthesis of polyisoprenoid lipids in the rat tapeworm Hymenolepis diminuta. Comp. Biochem. Physiol. Pt. B 82: 487–495 (1985)
57. Joseph D. R., Hall S. H. and French F. S.: Rats androgen-binding protein: evidence for identical subunits and amino acid sequence homology with human sex hormone-binding globulin. Proc. Nat. Acad. Sci. USA 84: 339–343 (1987)
58. Kabbouh M. et al.: Further characterization of the 2-deoxyecdysone C-2 hydroxylase from Locusta migratoria. Insect Biochem. 17: 1155–61 (1987)
59. Kabbouh M. and Rees H. H.: Characterization of the ATP-2-deoxyecdysone 22-phosphotransferase (2-deoxyecdysone 22-kinase) in the follicle cells of Schistocerca gregaria. Insect Biochem. 21: 57–64 (1991)
60. Kalinovskaya N. I. et al.: Steroid metabolites of the far eastern holothurian Stichopus japonicus Selenka. Comp. Biochem. Physiol. Pt. B 76: 167–171 (1983)
61. Karlaganis G. et al.: A bile alcohol sulfate as a major component in the bile of the small skate (Raja erinacea). J. Lipid Res. 30: 317–322 (1989)
62. Kihira K. et al.: Bile salts of the coelacanth, Latimeria chalumnae. J. Lipid Res. 25: 1330–36 (1985)
63. Kiriishi S. et al.: Prothoracic gland synthesis of 3-dehydroecdysone and its hemolymph 3-beta-reductase mediated conversion to ecdysone in representative insects. Experientia 46: 716–721 (1990)
64. Kljajic Z., Dogovic N. and Gasic M. J.: Sterols in Adriatic sea ascidians. Comp. Biochem. Physiol. Pt. B 75: 519–521 (1983)
65. Koolman J.: Ecdysteroids (Review). Zool. Sci. 7: 563–580 (1990)
66. Krust A. et al.: The chicken oestrogen receptor sequence: homology with v-erbA and the human oestrogen and glucocorticoid receptors. Embo J. 5: 891–897 (1986)
67. Ksebati M. B. and Schmitz F. J.: 24-Methyl-5α-cholestane-3ß,5,6ß,22R,24-pentol 6-acetate: new polyhydroxylated sterol from the soft coral Asterospicularia randalli. Steroids 43: 639–649 (1984)
68. Lachaise F. and Lafont R.: Ecdysteroid metabolism in a crab: Carcinus maenas L. Steroids 43: 243–259 (1984)
69. Lee H. M., Parish E. J. and Bone L. W.: Occurrence of mammalian sex steroids in the free-living nematode, Turbatrix aceti. Comp. Biochem. Physiol. Pt. A 97: 115–117 (1990)
70. Lehtovaara I and Koskinen T.: Demonstration of vitamin D3 metabolism in Mytilus edulis. Experientia 42: 147–148 (1986)
71. de Loof A., Briers T. and Huybrechts R.: Presence and function of ecdysteroids in adult insects. Comp. Biochem. Physiol. Pt. B 79: 505–509 (1984)
72. Lopez de Haro M. S. and Nieto A.: Nucleotide and derived amino acid sequences of a cDNA coding for the pre-uteroglobin from the lung of the hare (Lepus capensis). Biochem. J. 235: 895–898 (1986)
73. Mak P and Callard G. V.: A novel steroid-binding protein in the testis of the dogfish Squalus acanthias. Gen comp. Endocrinol. 68: 104–112 (1987)
74. Makin H. L. J. (ed.): Biochemistry of steroid hormones. Blackwell, Oxford 1984
75. Manconi R. et al.: Steroids in porifera. Sterols from freshwater sponges Ephydatia fluviatilis (L.) and Spongilla lacustris (L.). Comp. Biochem. Physiol. Pt. B 91: 237–245 (1988)
76. Martinez E., Givel F. and Wahli W.: A common ancestor DNA motif for invertebrate and vertebrate hormone response elements. Embo J. 10: 263–268 (1991)
77. McDonnell D. P. et al.: Molecular cloning of complementary DNA encoding the avian receptor for vitamin D. Science 235: 1214–17 (1987)

78. Meinwald J. et al.: Defensive steroids from a carrion beetle (Silpha americana). Experientia 41: 516–519 (1985)
79. Mercer J. G., Munn A. E. and Rees H. H.: Caenorhabditis elegans: occurrence and metabolism of ecdysteroids in adults and dauer larvae. Comp. Biochem. Physiol. Pt. B 90: 261–267 (1988)
80. Milkova T. S., Popov S. S. and Andreev S. N.: Sterols from black sea Actinaria species. Comp. Biochem. Physiol. Pt. B 87: 267–269 (1987)
81. Minale L. et al.: Starfish saponins – XVI. Composition of the steroidal glycoside sulphates from the starfish Luidia maculata. Comp. Biochem. Physiol. Pt. B 80: 113–118 (1985)
82. Nolte A. et al.: Ecdysteroids in the dorsal bodies of pulmonates (Gastropoda): synthesis and release of ecdysone. Comp. Biochem. Physiol. Pt. A 84: 777–782 (1986)
83. Norman A. W. (ed.): Vitamin D. de Gruyter, Berlin 1985
84. Novak F. et al.: Pregnenolone and estradiol identification in the brine shrimp, Artemia sp., by means of gas chromatographical-mass spectrometrical analysis. Comp. Biochem. Physiol. Pt. B 95: 565–569 (1990)
85. O'Hanlon G. M., Howarth O. W. and Rees H. H.: Identification of ecdysone 24-O-ß-D-glucopyranoside as a new metabolite of ecdysone in the nematode Parascaris equorum. Biochem. J. 248: 305–307 (1987)
86. Ohnishi E. et al.: Isolation and identification of major ecdysteroid conjugates from the ovaries of Bombyx mori. Insect Biochem. 19: 95–101 (1989)
87. van Oycke S., van Braekman J. C. and Daloze D.: Cardenolide biosynthesis in chrysomelid beetles. Experientia 43: 460–462 (1986)
88. Pakdel F. et al.: Full-length sequence and in vitro expression of rainbow trout estrogen receptor cDNA. Mol. cell. Endocrinol. 71: 195–204 (1990)
89. Pasteels J. M. et al.: Evolution of exocrine chemical defence in leaf beetles (Coleoptera, Chrysomelidae) (Review). Experientia 45: 295–300 (1989)
90. Pastuszyn A. et al.: Primary squence and structural analysis of sterol carrier protein 2 from rat liver: homology with immunoglobulins. J. biol. Chem. 262: 13219–27 (1987)
91. Raederstorff D. and Rohmer M.: Sterol biosynthesis via cyclartenol and other biochemical features related to photosynthetic phyla in the amoebae Naegleria lovanensis and Naegleria gruberi. Eur. J. Biochem. 164: 427–434 (1987)
92. Raederstorff D. and Rohmer M.: Polyterpenoids as cholesterol and tetrahymanol surrogates in the ciliate Tetrahymena pyriformis. Biochim. biophys. Acta 960: 190–199 (1988)
93. Ray K. et al.: The rat vitamin-D binding protein (Gc-globulin) gene: Structural analysis, functional and evolutionary considerations. J. Biol. Chem. 266: 6221–29 (1991)
94. Reinboth R. and Becker B.: In vitro studies on steroid metabolism by gonadal tissues from ambisexual teleosts. I. Conversion of C14-testosterone by males and females of the protogynous wrasse Coris julis. L. Gen. comp. Endocrinol. 55: 245–250 (1984)
95. Reis-Henriques et al.: Studies of endogenous steroids from the marine mollusc Mytilus edulis L. by gas chromatography and mass spectrometry. Comp. Biochem. Physiol. Pt. B 95: 303–309 (1990)
96. Rhoads D. E. and Kaneshiro E. S.: Fatty acids metabolism in Paramecium. Oleic acid metabolism and inhibition of polyunsaturated fatty acid synthesis by triparanol. Biochim. biophys. Acta 795: 20–29 (1984)
97. Riccio R. et al.: A novel group of highly hydroxylated steroids from the starfish Protoreaster nodosus. Tetrahedron 38: 3615–22 (1982)
98. Riccio R. et al.: Unusual sulfated marine steroids from the ophiuroid Ophioderma longicaudum. Tetrahedron 41: 6041–46 (1985)
99. Ritter K. S. et al.: Identification of delta-5,7–24-methylene- and methylsterols in the brain and whole body of Atta cephalotes isthmicola. Comp. Biochem. Physiol. Pt. B 71: 345–349 (1982)
100. Romero M. S. and Seldes A. M. Steroids from aquatic organisms – XI. Steroid composition of the prosobranch mollusc Patinigera magellanica. Comp. Biochem. Physiol. Pt. B 84: 125–129 (1986)
101. Sakurai S. et al.: 7-Dehydrosterols in prothoracic glands of the silkworm, Bombyx mori. Experientia 42: 1034–36 (1986)
102. Santa-Coloma T. A., Muschietti J. P. and Carreau E. H.: Sex steroid binding protein from Bufo arenarum: Further characterization. Comp. Biochem. Physiol. Pt. A 85: 401–405 (1986)
103. Schlinger B. A. and Arnold A. P.: Brain is the major site of estrogen synthesis in a songbird. Proc. Nat. Acad. Sci. USA 88: 4191–94 (1991)
104. Schott D. R. et al.: Molecular cloning, sequence analyses, and expression of complementary DNA encoding murine progesterone receptor. Biochemistry 30: 7014–20 (1991)
105. Seldes A. M. et al.: Steroids from aquatic organisms – XII. Sterols from the Antarctic sponge Homaxinella balfouriensis (Ridley and Dendy). Comp. Biochem. Physiol. Pt. B 83: 841–843 (1986)
106. Sica D., Piccialli V. and Pronzato R.: delta-5,7-sterols from the sponges Ircinia pipetta and Dysidea avara. Identification of cholesta-5,7,24-trien-3ß-ol. Comp. Biochem. Physiol. Pt. B 88: 293–296 (1987)
107. Smith T. and Bownes M.: Metabolism of 20-hydroxyecdysone in adult Drosophila melanogaster. Insect Biochem. 15: 749–754 (1985)
108. Sommé-Martin G., Colardeau J. and Lafont R.: Conversion of ecdysone and 20-hydroxyecdysone into 3-dehydro-ecdysteroids is a major pathway in third instar Drosophila melanogaster larvae. Insect Biochem. 18: 729–734 (1988)
109. Srivatsan J., Weirich M. and Agosin M.: Cytochrome P450-catalyzed formation of 20-hydroxyecdysone in larval housefly mitochondria. Biochem. biophys. Res. Commun. 166: 1372–77 (1990)
110. Stacey N. E. et al.: Direct evidence that 17α,20ß-dihydroxy-4-pregnen-3-one functions as a goldfish primer pheromone. Preovulatory release is closely associated with male endocrine responses. Gen. comp. Endocrinol. 75: 62–70 (1989)
111. Stoilov I., Popov S. and Andreev S.: Sterols from the main black sea molluscs. Comp. Biochem. Physiol. Pt. B 79: 493–497 (1984)
112. Svoboda J. A. and Thompson M. J.: Steroids. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 137–175. Pergamon Press, Oxford 1985
113. Svoboda J. A. et al.: New intermediates in the conversion of stigmasterol to cholestanol in the Mexican bean beetle. Lipids 21: 639–642 (1986)

114. Swevers L., Lambert J. G. D. and Deloof A.: Synthesis and metabolism of vertebrate-type steroids by tissues of insects: A critical evaluation. Experientia 47: 687–698 (1991)
115. Tachibana K., Sakaitanai M. and Nakanishi K.: Pavoninins: shark-repelling ichthyotoxins from the defensive secretion of the Pacific sole. Science 226: 703–705 (1984)
116. Teshima S. and Patterson G. W.: Identification of 4α-methylsterols in the oyster, Crassostrea virginica. Comp. Biochem. Physiol. Pt. B 69: 175–181 (1981)
117. Teshima S. et al.: Sterols of the chiton (Liolophura japonica): a new C30-sterol, (24Z)-24-propylidencholest-7-enol, and other minor sterols. Comp. Biochem. Physiol. Pt. B 71: 373–378 (1982)
118. Thompson M. J. et al.: Biosynthesis of a C21-steroid conjugate in an insect. The conversion of (C14)cholesterol to 5-(C14)pregnen-3ß,20ß-diol glucoside in the tobacco hornworm, Manduca sexta. J. biol. Chem. 260: 15410–12 (1985)
119. Thompson M. J. et al.: Profile of free and conjugated ecdysteroids and ecdysteroid acids during embryonic development of Manduca sexta (L.) following maternal incorporation of C14-cholesterol. Arch. Insect Biochem. Physiol. 7: 157–172 (1988)
120. Tsoupras G., Luu B. and Hoffmann J. A.: A cytokinin (isopentenyl-adenosyl-mononucleotide) linked to ecdysone in newly laid eggs of Locusta migratoria. Science 220: 507–509 (1983)
121. Veares M. P., Goad L. J. and Apsimon J. W.: Thornasterol A sulphate, a constituent of the starfish Asterias rubens. Comp. Biochem. Physiol. Pt. B 90: 25–28 (1988)
122. Vessey D. A. et al.: Purification and characterization of the enzymes of the bile acid conjugation from fish liver. Comp. Biochem. Physiol. Pt. B 95: 647–652 (1990)
123. Vonk H. J. and Western J. R. H.: Comparative biochemistry and physiology of digestion. Acad. Press, New York 1984
124. Voogt P. A., Denbesten P. J. and Jansen M.: The delta-5-pathway in steroid metabolism in the sea star Asterias rubens L. Comp. Biochem. Physiol. Pt. B 97: 555–562 (1990)
125. Wahli W. and Martinez E.: Superfamily of steroid nuclear receptors. Faseb J. 5: 2243–49 (1991)
126. Walsh K. A. et al.: Amino acid sequence of the sex steroid binding protein of human blood plasma. Biochemistry 25: 7584–90 (1986)
127. Wang L. H. et al.: COUP transcription factor is a member of the steroid receptor superfamily. Nature 340: 163–166 (1989)
128. Whiting P. and Dinan L.: Identification of the endogenous apolar ecdysteroid conjugates present in newly-laid eggs of the house cricket (Acheta domesticus) as 22-long-chain fatty acyl esters of ecdysone. Insect Biochem. 19: 759–765 (1989)
129. Woodward H. D., Allen J. M. C. and Lennarz W. D.: 3-hydroxy-3-methylglutaryl-coenzyme A reductase of the sea urchin embryo. Deduced structure and regulatory properties. J. Biol. Chem. 263: 18411–18 (1988)
130. Yanda D. M. and Ghazarian J. G.: Vitamin D and 25-hydroxyvitamin D in rainbow trout (Salmo gairdneri): Cytochrome P-450 and biotransformations of the vitamins. Comp. Biochem. Physiol. Pt. B 69: 183–188 (1981)
131. Zhu X. X., Oliver J. H. and Dotson E. M.: Epidermis as a source of ecdysone in an argasid tick. Proc. Nat. Acad. Sci. USA 88: 3744–47 (1991)
132. Special Issue: Ecdysone: From biosynthesis to mode of action. J. Insect Physiol. 34: 549–745 (1988)
133. IUPAC-IUB Joint Commission on Biochemical Nomenclature: The nomenclature of steroids. Recommendations 1989. Eur. J. Biochem. 186: 429–458 (1989)

17 Ester Hydrolases, ATPases and Carboanhydrases

Many biomolecules are esterified with carbonic acid, phosphoric acid or sulphuric acid. The enzymes responsible for the hydrolytic cleavage of the esters are the carboxylester hydrolases, phosphoesterases and sulphatases. In addition to the phosphomonoesterases, we will also consider the phosphodiesterases and nucleases that are specific for cAMP and cGMP, as well as the ATPases that cleave phosphoric acid anhydride bonds. The ion-transporting ATPases will be discussed together with comparative biochemical data on ion channels. The chapter ends with a discussion of carboanhydrases.

17.1 Carboxylester Hydrolases

The carboxylester hydrolases include the (carboxyl) esterases, lipases and cholinesterases. The carboxylester hydrolases sequenced so far show 27–53 % sequence agreement on pairwise comparison and thus belong to the same protein super-family. Unfortunately, the sequence data available are insufficient to draw conclusions about the evolution of this enzyme family and to construct a genealogical tree. All known carboxylester hydrolases have a serine in the active centre, but the serine esterases are not related to the serine proteases [29, 43]. Surprisingly, the amino acids 30–575 of acetylcholinesterase from the electric organ of *Torpedo* agree about 28 % with amino acids 2210–2750 of bovine thyroglobulin; in addition, six of the eight cysteine residues are conserved, resulting in similar folding [146]. It is not easy to differentiate between (carboxyl) esterases and lipases. According to the usual definition, lipases should only be active on substrate aggregates at the ester–water interface; in fact, lipases also cleave substrates in molecular or micellar solution (e.g. paraoxon), whereas microsomal carboxyl esterases also attack emulsified tributyrylglycerol. A practical distinction is that triacylglycerols with long-chain fatty acids are only cleaved by lipases. Amongst the enzymes termed phospholipases some should be classified as carboxyl esterases and others as phosphodiesterases.

17.1.1 Carboxyl, Aryl and Acetyl Esterases

The esterases cleave aliphatic esters of short-chain carboxyl acids (preferentially C_3 to C_6) and also aromatic esters, aromatic amides and thioesters. Some also catalyse acyl transfer to alcohols or amino acids. The **classification** of these enzymes is especially difficult for the following reasons [123]:

1. With few exceptions, neither the natural substrate nor the biological function is known for any esterase; artificial substrates are used for identification and activity determinations, e.g.

the chromogens 4-nitrophenylester and naphthylester.
2. The substrate preferences of the different classes show considerable overlap.
3. Most esterases exist in multiple forms, on the one hand because of marked genetic polymorphism, and on the other hand as the result of variable glycosylation and other post-translational modifications.
4. Other enzymes have esterase activity; examples include serine proteases, and even proteins which are not usually considered to be enzymes, like serum albumins.
5. Only a few esterases have been sequenced so far.

Thus, classification of the esterases relies not only on substrate specificities (which overlap considerably) but also on responses to inhibitors. The Enzyme Commission of the International Union of Biochemistry (IUB) differentiates between four **main classes**. The carboxyl esterases in a narrow sense, named also ali- or B-esterases (EC 3.1.1.1), are inhibited by diisopropylphosphofluoridate (DIPF) and other organophosphates. They therefore belong to the serine hydrolases, which are irreversibly damaged by phosphorylation of the active serine. In contrast, the aryl- or A-esterases (EC 3.1.1.2) not only are insensitive to organophosphates but can even cleave compounds such as paraoxon (diethyl-4-nitrophenylphosphate); however, they are inhibited by *p*-hydroxy- or *p*-chloromercuribenzoate (PHMB, PCMB). The acetylcholinesterases (EC 3.1.1.7) and the so-called non-specific cholinesterases (EC 3.1.1.8, also pseudo- or butyrylcholinesterases) are inhibited not only by DIPF but also by eserine (physostigmine). The last group, the acetyl- or C-esterases (EC 3.1.1.6) are insensitive to all known inhibitors and are characterized by their specificity for acetic esters.

To describe the esterase spectrum of fish muscle, Hart and Cook in 1976 introduced three further classes, of which type ER is just as insensitive as the acetyl esterases but shows broader specificity, ESe are inhibited only by eserine and not by DIPF, and finally Esdp are inhibited by DIPF as well as by PHMB or PCMB [55]. A strong distinction between "A-esterases", which are active against paraoxon, and "aryl esterases", which are active against phenylacetate, has been suggested by investigations of the blood serum of 11 mammalian species, where both activities were always found, and 14 bird species, where higher aryl esterase activity was always present but little or no A-esterase activity occurred [88]. Other authors, however, recommend the rejection of all such classifications and recognition only of those esterases for which the natural substrate is known, such as the juvenile hormone esterases of the insects, or from *Drosophila melanogaster* esterase 6 which cleaves the pheromone cis-vaccenylacetate [123]. At least a dozen esterases can be separated by, for example, electrophoresis in practically all cells of animals down to the Protozoa; they include representatives of several of the classes defined by the Enzyme Commission. However, the same tissue, even in closely related species, may have a completely different pattern. Thus, no conclusions about the biological roles of esterases can be arrived at from this classification.

The **carboxyl esterases in the narrowest sense** (EC 3.1.1.1) have been investigated most thoroughly in mammalian liver. There they are bound to the cytoplasmic side of the endoplasmic reticulum, probably via phospholipids. Their function is not clear but they are at least involved in detoxification processes. Several such enzymes from mouse and rat liver can be separated by electrophoresis; some are monomers of 57–60 kDa and others have the otherwise rather rare trimeric structure. In the trimeric esterase 6A of the mouse, titration of the active centre with paraoxon gives an equivalent mass of 178 kDa; therefore, the active centres of the three subunits are not functionally equivalent. The 14 isoenzymes known so far from the mouse are encoded by two gene clusters on chromosome 8. The trimeric esterase 6A, widely found in many organs, is immunologically very similar to esterase 1F from the blood and esterase 9A from the liver and kidney; the corresponding genes lie adjacent in cluster 1. Apart from this, there are considerable differences in immunological reactivity, substrate specificity, inhibitor responses and quaternary structure between the products of the two gene families [135, 139].

The best-known **aryl esterases** (EC 3.1.1.2) are those from mammalian blood plasma. The bovine and human enzymes have molecular masses of 440 and 240 kDa, respectively; they are associated with high-density lipoproteins (HDL). These enzymes are defined not so much by their substrate specificity, as arylesters can be cleaved by other esterases, carboanhydrases and proteases, but much more by their responses to inhibitors. They are not inhibited by DIPF or eserine but by PHMB, PCMB and other SH reagents as well as by heavy metals. They cleave aromatic esters, in particular acetates, but not aliphatic esters.

Because they also attack phosphoester bonds in organophosphates such as paraoxon, they may also be classified as phosphatases. The blood plasma of many vertebrates also contains an **acetyl esterase** that is specific for the acetate residue in the platelet-activating factor 1-alkyl-2-acetyl-sn-glycero-3-phosphocholine. This enzyme is apparently present in all mammals but is missing from birds (chicken, pigeon, turkey) and anurans *(Rana pipiens);* it has been reported in the snake *Thamnophis sirtalis*, the turtle *Terrapene carolina* and the teleosts *Salmo gairdneri* and *Lepomis macrochirus*. The action of the platelet-activating factor has, as yet, been investigated only in mammals, and thus an explanation for the particularly sporadic distribution of this enzyme is not available [21]. The **insects** have just as complicated a spectrum of esterases as do the vertebrates. For example, 18 esterases have been separated electrophoretically in the mosquito *Culex tarsalis*, of which 14 can be classified as carboxyl esterases, two as acetylcholinesterases and one each as an aryl esterase and an acetyl esterase. Here again, it is impossible to derive the biological importance of these different enzymes. The aryl esterases take part in the detoxification of the organophosphate insecticides and are apparently widespread in insects. Such an enzyme from the bug *Triatoma infestans* is one of the few insect esterases so far subjected to purification attempts [92]. Quite understandable interest has been aroused by the esterases that inactivate juveline hormones by cleavage of the methyl group. The specific juvenile hormone esterases are distinguishable from non-specific esterases in that they are not inhibited by DIPF but are inhibited by certain inhibitors which are only effective in this case, e.g. EPPAT (O-ethyl-S-phenyl-phosphoamidothiolate). These enzymes have been discussed already in connection with the juvenile hormones (p. 615). The insect esterases are model examples of enzyme polymorphism. For this reason, two highly polymorphic esterases from *Drosophila mojavensis* have been isolated and studied in detail. Esterase 4 is a dimer of 62- to 64-kDa subunits with α-naphthylacetate as the optimal substrate. The usual inhibitors are only effective at very high concentrations (DIPF, eserine) or are only very weak (PCMB); thus, the enzyme is classifiable as an acetyl esterase (EC 3.1.1.6). Esterases 4 and 5 differ in only 6 of their 34 N-terminal amino acids, and are apparently products of duplicated genes, such as are found in other *Drosophila* species, but which are expressed in a species-, stage- or organ-specific manner [122]. A carboxyl esterase isolated from *Caenorhabditis elegans* shows significant similarity, at least in the 13 N-terminal amino acids, to the latter two enzymes from *Drosophila mojavensis* [98].

17.1.2 Cholinesterases

Essential components of the cholinergic synapses of probably all metazoans are enzymes that can hydrolase acetylcholine. It is not clear, however, whether these enzymes are comparable with the cholinesterases of the vertebrates, and they are characterized in particular by specific inhibition by eserine. Two types of cholinesterase can be distinguished in the vertebrates. The acetylcholinesterase (EC 3.1.1.7, AChE) of cholinergic synapses is well known for its extremely high turnover rate of 15 000 s^{-1}. In addition, many cells contain an enzyme that cleaves many other choline esters in addition to acetylcholine; this is known as butyrylcholinesterase (BChE) or pseudocholinesterase (EC 3.1.1.8). A clear differentiation between AChE and BChE is only really possible in the higher vertebrates; in lower vertebrates and invertebrates the association of acetylcholine-cleaving activity with one or other of the two types is not always possible. For example, the BChE from the heart of the electric ray *Torpedo* sp. is responsible for the physiological inactivation of the transmitter in the heart, and this enzyme hydrolyses acetylcholine better than butyrylcholine. The cholinesterases from the muscles of the flounder *Pleuronectes flesus* and the enzyme designated as an acetylcholinesterase in *Drosophila* lie between AChE and BChE in their specificities [26, 158]. AChE and BChE are glycoproteins with a high degree of heterogeneity between tissues in their quaternary structure, charge and carbohydrate composition. In the vertebrates, they are found particularly in the nervous system and muscle, but also occur in erythrocytes and other cells, in blood plasma and egg yolk, where their function is completely unknown. The muscle-derived electric organs of the electric eel *Electrophorus electricus* and the electric rays *Torpedo californica, T. marmorata* and *Narce japonica* are especially rich in AChEs [166].

AChEs and BChEs exist in multiple molecular forms. The asymmetrical forms A4, A8 and A12 are made up of one, two or three catalytic tetramers bound to a triple-helical, collagen-like tail unit (Fig. 17.1). This tail unit contains hydroxy-

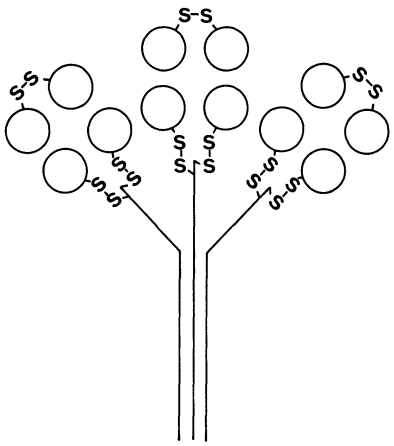


Fig. 17.1. The structure of the asymmetrical acetylcholinesterase A12 from the electric organ of the electric eel *Electrophorus electricus* [95]

proline and hydroxylysine. The A forms are produced intracellularly but are then transported to the basal lamina, where they are bound by ionic interaction with the tail unit. The globular forms G1, G2 and G4 contain one, two or four catalytic subunits. In G2 and G4, the subunits exist as dimers linked by disulphide bridges; however, the dimers do not dissociate under reducing conditions and, therefore, other sorts of bonds between the subunits must be present. The G forms are either water soluble and easily extractable, or membrane bound and only released by detergent treatment ("amphiphilic forms"). The G2 AChE of erythrocytes and many other cells is anchored to the plasma membrane by glycosyl phosphatidylinositol. In contrast, the G4 AChE of mammalian brain has an as yet incompletely characterized hydrophobic anchor of 20 kDa that is bound asymmetrically to only two of the subunits and contains fatty acids but no inositol [19, 26, 166].

The A forms of the AChEs are at their highest densities in the neuromuscular junctions and electric organs, where, in fact, the G forms predominate. Apart from in muscles and electric organs, the A forms are found only in peripheral ganglia; all other cell types contain only G forms. A forms are found in the muscles of all vertebrates from the agnathans onwards, but are not present in the invertebrates. A12 forms of AChE and BChE are known from mammals and birds. The predominant form in the breast muscle of the chicken is a hybrid molecule with an equal number of AChE and BChE subunits [26, 127]. Membrane-bound G2 AChEs are apparently ubiquitous. They have been identified, for example, in electric organs, in flounder muscle, in the erythrocytes of various mammals, in blood platelets, and in the peripheral ganglia and brain of *Drosophila*. In contrast to the peripheral ganglia, the central nervous system contains only G forms of AChE. A membrane-bound G4 AChE from mammalian brain has been especially well investigated. AChEs and BChEs also exist in soluble form in body fluids. About 95% of the cholinesterase activity of human blood plasma is due to G4 BChE [26].

The A forms of AChE and the membrane-bound G2 AChE from the electric organs have catalytic subunits of different sizes (72 and 69 kDa). The electric organs of *Torpedo californica, T. marmorata* and the related species *Narce japonica* contain a third subunit of 76 kDa. The subunits differ in sequence only in a few amino acids at the C-terminus, and arise by alternative splicing of the same gene. The AChE gene of *Torpedo* consists of three exons for amino acids 1–1511, 1512–1678 and 1679–1718 of the A subunit. The newly synthesized subunit of the amphiphilic G2 contains a C-terminal, non-polar extension which is replaced by glycolipid in the mature protein [147, 153, 166]. Complete, or almost complete, amino acid sequences are so far available for only the *Torpedo* AChE, *Drosophila* AChE, foetal bovine AChE, human BChE and rabbit BChE [26, 70, 146]. As yet, it is not possible to compare the sequences of AChE and BChE of the same species. However, bovine AChE shows greater similarity to the *Torpedo* AChE sequence (60%) than to the human BChE sequence (50%). The catalytic subunits of all these enzymes are similar in length: 537 amino acids for human hydrophilic BChE, 574 amino acids for the hydrophilic *Torpedo* AChE, 537 amino acids for mature amphiphilic *Torpedo* AChE, and 577 amino acids for amphiphilic *Drosophila* AChE. The sequence around the active serine is always Gly-Glu-Ser-Ala-Gly [26].

Of the invertebrate cholinesterases, only those of some insects have been examined in any detail. The main form of AChE from *Drosophila melanogaster* is an amphiphilic dimer, but amphiphilic monomers are also present. The hydrophilic dimers and monomers reported are possibly proteolytic artefacts. The polypeptide chains of the amphiphilic dimers, which are anchored in the membrane at the C-terminus via a glycosyl phosphatidylinositol, are post-translationally cleaved into two fragments of 55 and 16 kDa. The AChE gene of *Drosophila* has ten exons and is thus much more fragmented than the vertebrate gene,

but the last two introns have the same position [2, 44, 51]. G1 AChEs and G2 AChEs are also detectable in other insect species; the report in 1978 of an A form in *Musca domestica* probably requires reexamination [12]. In contrast to other esterases, there is no known polymorphism of the *Drosophila* AChE; no electrophoretic variants were found in studies of 25 populations of *Drosophila melanogaster* and 30 populations of *D. pseudoobscura*. The results of density gradient centrifugation suggest the existence of G1, G2 and G4 forms of AChE in the ascidian *Ciona intestinalis;* asymmetric forms, on the other hand, appear to be absent. AChE aggregates 8S (450 kDa), 10S, 18S and 32S have been isolated from the trematode *Schistosoma mansoni* [165].

17.1.3 Lipases and Cholesterol Ester Hydrolases

There are four lipolytic activities in **mammalian gut:** (1) the preduodenal lipases or gastric lipases, which are produced in glands beneath the tongue, in the pharynx or in the stomach, depending upon the species; (2) the colipase-dependent pancreas lipase; (3) the bile salt-dependent lipases of the pancreas and the milk; and (4) the phospholipase A_2, also from the pancreas.

The preduodenal lipases of humans, rats, mice, cattle and sheep are mainly produced in the Ebner glands beneath the tongue, and those of the guinea-pig, rabbit, dog, horse, and pig are produced in the major cells of the stomach. They are especially active in the stomach, where in humans 30% of triacylglycerols (TAGs) are already cleaved. Their activity varies greatly with the species [48, 104]. The rat and human lingual lipases are glycoproteins with 377–379 amino acids and show 78% sequence agreement. The enzyme acts as a typical lipase only on aggregated substrates, but in contrast to pancreas lipase preferentially hydrolyses the sn-3 ester bond and shows marked product inhibition; thus, TAG cleavage only results in diacylglycerol (DAG). The pH dependence is broad, with an optimum of 4.5, and it is quite possible that the enzyme functions at neutral pH in the small intestine. Whatever the case, the lingual lipases promote the emulsification of fats in the small intestine; DAG is produced and free fatty acids accumulate at the fat-water interface [13, 48].

The classical **pancreas lipase** (EC 3.1.1.3) is classified as bile salt-independent because it is fully active in the absence of bile salts and is, in fact, inhibited by bile acids above their critical micellar concentration of 0.03–0.07%; it is stimulated by calcium ions. Human pancreas lipase consists of 465 amino acids and shows 85% sequence agreement with the pig enzyme, and 70% agreement with the dog enzyme [85]. Micellar TAGs are attacked by pancreas lipase only after partial lipolysis by gastric lipase [13]. The pH optimum is 9. The pancreas lipase shows maximum activity when it is anchored by colipase to the surface of lipid micelles. **Colipase** is present in the pancreatic secretion as an inactive proenzyme that is activated by the tryptic hydrolysis of an Arg-Gly bond; in the process, an oligopeptide of four or five amino acids is cleaved off. The heat- and acid-stable colipase has been reported from vertebrates of all classes down to the common hagfish *Myxine glutinosa;* the complete sequence is known for pig, horse and human colipase and the N-terminal sequence is known for the enzyme from the chicken and the shark *Acanthias vulgaris*. The polypeptide is very conservative: the N-terminal sequences of shark and man agree 57%, and those of the shark and chicken agree 64%. Nevertheless, the heterogeneous system of human lipase and shark colipase is fivefold less active than the human homologous system. This is probably due to differences in the charge of the colipase, which has a pI of 10.2 in the shark compared with 5.8 in humans [85, 157].

The **bile salt-dependent lipase of the pancreas** or **cholesterol esterase** is secreted as an inactive monomer of 65 kDa; in the rat it is polymerized to active hexamers by trihydroxy bile acids and to higher aggregates by dihydroxy bile acids. The porcine and human pro-enzymes are converted to active dimers by the action of cholic acid. The enzyme acts upon soluble esters, such as triacetyl or tripropionyl glycerol, methylbutyrate or lysolecithin, but also upon insoluble esters of cholesterol and vitamins A, D and E in bile acid-stabilized micelles; it has little activity against emulsified fats [113]. Human and gorilla milk contain a lipase that is activated by bile acids and is of critical importance for milk digestion in infants. It is synthesized in the milk glands and makes up about 1% of the total protein of human milk. It is apparently encoded by the same gene as the bile salt-dependent lipase of the pancreas. The sequence consists of 722 amino acids and is homologous to other serine esterases [4, 113]. The milk enzyme is not present in other mammals such as the cow, goat, pig, horse, rat or guinea-pig; nor is it present in rhesus monkeys, and it was therefore considered to be a relatively recent

acquisition in the evolution of the higher primates. More recently, it has been found in the milk of dogs at about the same level as in humans, and in the cat at somewhat lower levels. The activity in these two animals is, however, more dependent upon the lactation phase than in humans, which explains previous negative results with the dog. The human, dog and cat enzymes have corresponding pH optima of 7.3–8.5, and are also all strongly inhibited by eserine; in addition, they all have an absolute requirement for glycocholic or taurocholic acid and are inhibited by secondary bile acids such as taurodeoxycholic and glycodeoxycholic acids [46].

Lipases are found in the spectrum of **digestive enzymes** of all animals examined but have rarely been examined in detail. The gut lipases of the insects are similar to the pancreas lipases in that they are stimulated by calcium ions, preferentially cleave primary ester bonds of TAG, and produce sn-2-monoacylglycerol (MAG) [36]. In contrast, the TAG/DAG lipase secreted together with other hydrolases into the medium by the ciliate *Tetrahymena* is not stimulated by Ca^{2+}, and also has a much smaller size of 30.5 kDa [42]. **Intracellular lipolytic activity** is found in almost all animal cells. However, the assay methods used did not distinguish clearly between lipases and esterases in many cases, and only a few such enzymes have been examined further. The hormonally regulated lipase from mammalian and avian fat tissue is especially interesting; adrenalin, noradrenalin, glucagon and ACTH bring about the phosphorylation of the enzyme via a cAMP-dependent protein kinase and, thereby, its activation. The fat-tissue cells (adipocytes) of the rat and chicken contain a large lipid-rich complex with the ability to cleave acyl residues from TAG, DAG, MAG and cholesterol esters. A 42-kDa TAG/MAG lipase can be separated from a 45-kDa MAG lipase. The former is also phosphorylated by a cAMP-dependent protein kinase but is not activated; in this case activation requires interaction with other polypeptides [16].

Lipoprotein lipase and **hepatic lipase** are involved in the metabolism of plasma lipoproteins. The lipoprotein lipase supplies the ca. 100 g plasma fatty acids that are metabolized daily in humans [183]. This enzyme is located on the lumen side of the endothelial cells and produces TAG chylomicron residues and IDL (intermediate lipoproteins) by hydrolysis of chylomicrons and VLDL (very low density lipoproteins). Its main function is apparently to supply fatty acids to various tissues. Lipoprotein lipase is found in several different tissues, e.g. fat tissue, heart, skeletal muscle, lung, milk glands, brain and kidney; it is also detectable in milk itself. The enzyme is a glycoprotein whose sequence of 448–450 amino acids is very similar in humans, mice and cattle [183]. It requires apolipoprotein C-II as the cofactor for maximal activity and is inhibited by 1 mol/l NaCl. The hepatic lipase requries no cofactor and is active in 1 mol/l NaCl; however, the two enzymes are very similar and in vitro can cleave phospholipids, TAG, DAG and MAG. The hepatic lipase is located mainly on the sinusoidal surface of the liver, and its apparent function is to hydrolyse TAG in IDL and phospholipids in HDL. The enzyme in the rat is a glycoprotein with 8% carbohydrate and a polypeptide chain of 472 amino acids that carries two N-glycosylation sites [79]. Lipoprotein lipase and hepatic lipase belong to the super-family of the serine esterases. A hypothetical genealogical tree can be constructed to show how gene duplication, exon shuffling and loss of introns combined to convert an ancestral gene with 14 introns into the gene of the colipase-dependent pancreas lipase with 12 introns, and the genes of the lipoprotein lipase and hepatic lipase with 9 introns. The tree also includes vitellogenins YP1 to YP3 of *Drosophila;* these show 35–40% sequence agreement with lipoprotein lipase in part of their sequences [77, 124].

Like the adipocyte lipase of the vertebrates, the **lipolytic activity in insect fat bodies** is also hormonally regulated; the adipokinetic hormone (AKH) from the corpora cardiaca phosphorylates, and thereby activates, the lipase via a cAMP-dependent protein kinase; this stimulates the release of sn-1,2-DAG into the haemolymph where it is bound to the transport protein lipophorin. It is not clear whether formation of the DAG involves the cleavage only of the sn-3 bond or whether sn-2-MAG is produced initially and then converted to sn-1,2-DAG by a specific acyl transfer. The **flight muscles** of the Lepidoptera contain high lipolytic activity that is mainly involved in the energy-yielding metabolism of lipids; MAG is cleaved with high efficiency. The TAG- and DAG/MAG-specific activities can be separated electrophoretically, but it is not known whether an enzyme complex, analogous to that in the fat tissue of the chicken, is present [36]. The enzyme in the flight muscle of the migratory locust *Locusta migratoria* is inhibited by the C_L protein of the haemolymph, Conversion of lipoprotein A^{yellow} to A^+ in the haemolymph as a result of AKH activity occupies the majority of the C_L

protein and thus brings about a four- to fivefold increase in the hydrolysis of DAG in the lipoprotein [59, 179]. Whilst membrane binding of mammalian lipoprotein lipase occurs via glycosaminoglycan, the locust enzyme does not bind to heparin and therefore cannot be purified by bioaffinity chromatography [59].

Various types of **cholesterol ester hydrolases** are known in mammals. The lysosomes have an enzyme with a pH optimum of 4.5, and the cytoplasm of most cells contains a neutral enzyme with a pH optimum of 7.5. Little is known in detail about a third type from the mitochondria. The cytoplasmic cholesterol ester hydrolase of the vertebrates is very similar to the hormone-regulated lipase of the adipocytes: both are activated by ACTH via a cAMP-dependent protein kinase and are inhibited by DIPF; they are both 84–90 kDa in size and have a very similar cleavage-peptide pattern. The cholesterol esterase of the pancreas is identical to the bile salt-dependent lipase [113]. Rat testis contains two chromatographically distinct cholesterol ester hydrolases which agree in many of their properties, although they show rather different temperature stabilities; whereas one of them, like the corresponding enzyme from the liver and kidney, is stable up to 37 °C, the other is rapidly inactivated at temperatures above 32 °C. It may be remembered that spermatogenesis in mammalian testis ceases completely at temperatures above this approximate limit. The temperature-labile enzyme in rats who have had the hypophysis removed is induced by FSH but not by LH and is, therefore, probably located in the Sertoli cells. The temperature-stable enzyme, on the other hand, is FSH- and LH-dependent and is apparently present in both the Sertoli and the Leydig cells. The unusual properties and specific location of the temperature-labile enzyme suggests that it is specifically adapted to a particular biological function in the testis [38].

17.1.4 Phospholipases

There exist various **enzyme types** that hydrolyse phospholipids [175]. The phospholipases A_1, A_2 and B are carboxylester hydrolases of different specificity, whereas phospholipases C and D may be classified as phosphodiesterases (Fig. 17.2). **Phospholipases of type A_1** cleave the ester bonds at sn-1 of various gylcerophospholipids, thereby producing sn-2-lysophospolipids. They are found in two quite different forms in almost all cells.

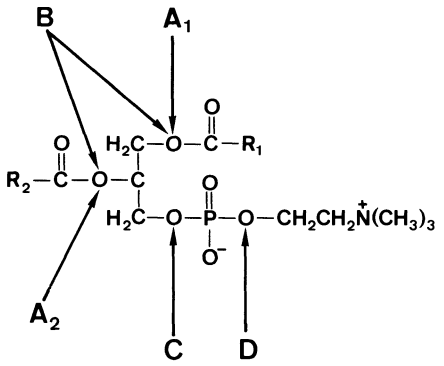


Fig. 17.2. The sites at which phospholipases A_1, A_2, B, C and D attack the lecithin molecule

The membrane-bound phospholipase A_1 is localized in cell membranes, microsomes and the Golgi apparatus; it has an alkaline pH optimum and is stimulated by Ca^{2+} ions. The soluble phospholipase A_1 is found in the cytoplasm and the lysosomes; it has an acidic pH optimum and is not stimulated by calcium ions but may be inhibited by them. Whereas high activities of phospholipase A_2 are found in pancreas secretions and animal toxins, vertebrate phospholipase A_1 is never extracellular. The identification of **phospholipases in mammalian blood** is not without its problems; this is because lecithin:cholesterol acyltransferase also shows A_2 activity and TAG lipase has A_1 activity. Thus, it is particularly interesting that a phospholipase detected in the **haemolymph** of the moth *Manduca sexta* is clearly identifiable as A_1 on the basis of experiments with specifically labelled phosphatidylethanolamine. The enzyme is unusually large with a molecular mass of 155 kDa, is stimulated by Ca^{2+} and is active against phosphatidylethanolamine and phosphatidylglycerol but not the predominant phosphatidylcholine of the haemolymph or TAG [112].

The **phospholipases A_2** cleave the sn-2 ester bonds in glycerophospholipids and produce sn-1 lysophospholipids. Specific lysophospholipases are responsible for cleavage of the second acyl residue from the lysophospholipids that result from phospholipases A_1 and A_2 activity. Phospholipases A_2 have their highest activities in the toxins of snakes, bees and scorpions, and in the pancreatic secretion in mammals. Lower intracellular activities are found in all cell compartments, where, with the exception of the mitochondria, they are always accompanied by phospholipase A_1 activity. The intracellular phospholipases A_2 are much less well known than the extracellular forms. The lysosomal enzymes have acid pH

optima and are inhibited by Ca^{2+}; the other cell organelles contain enzymes with pH optima of 8–9.5 that are stimulated by calcium ions. The enzymes from the blood platelets of the rat or the human spleen are atypical in that they are synthesized with a signal sequence but the membrane-bound, mature polypeptide of 124–125 amino acids has no transmembrane segment. The enzyme from human spleen shows only 39 % agreement with that of the pancreatic secretion [64, 73]. All phospholipases A_2 are monomers or dimers of small rigid polypeptides with about 125 amino acids and six or seven disulphide bridges. More than 30 complete and several partial sequences are known. The enzymes from the pancreas and snake venom are sequence homologues and have very similar tertiary structures [37], but, nevertheless, can be subdivided into several structural classes. The first class includes the enzymes from the pancreas and the venoms of the more primitive cobras (Elapidae) and the sea snakes (Hydrophiidae). Characteristic of these enzymes is a disulphide bridge between 11-Cys and 69-Cys. The enzymes of the second class, from the venom of the pit vipers (Crotalidae), have a C-terminal extension of six amino acids, with the terminal cysteine linked by a disulphide bridge to the 50-Cys of the active centre. Invariant in both classes are the 48-His, the 49-Asp of the calcium-binding site, and the loop with 28-Tyr, 30-Gly and 32-Gly; the characteristic residues 48-His and 49-Asp are also found in the phospholipase A_2 of bee venom. Thus, it is rather strange to find that the venoms of the water moccasin *Agkistroson piscivorus* and the fer-de-lance *Bothrops atrox* contain not only typical A_2 enzymes of the second class but also A_2 with a lysine residue in position 49. In contrast to all the others, these enzymes do not bind calcium in the absence of the substrate; the order in which the catalytically active ternary complex is formed is such that calcium is bound to the enzyme after, rather than before, the phospholipid [94]. Without taking into account the 49-Lys, a **dendogram** has been constructed in which three groups can be distinguished. The first includes the A_2 from the pancreas and the venoms of the Asiatic Elapidae, the second the A_2 from the Hydrophiidae venoms and the Australo-Asiatic Elapidae, and the third the venoms of the Crotalidae and Viperidae. Surprisingly, all neurotoxic venoms belong in the second group. The dendogram probably contains paralogous sequences (p. 155) and therefore cannot be accepted as a reliable description of A_2 evolution [37]. The phospholipase A_2 from bee venom agrees with the enzymes from mammalian pancreas and snake venom in the position of the cysteine residues, but otherwise shows little similarity [80].

The existence of **phospholipases B**, which can cleave fatty acids from both phospholipids and lysophospholipids, was for some time controversial but is now accepted. Phosphatidylinositol-specific **phospholipases C** (PI-PLC) are components of a widely found signal transduction system in which they generate the secondary messengers inositol trisphosphate and diacylglycerol by hydrolysis of phosphatidylinositol-4,5-bisphosphate. This signal transduction pathway links very different stimuli with effector processes such as cell growth and differentiation, contraction, engram formation and phototransduction in invertebrates. As in the case of the control of adenylate cyclases and cGMP phosphodiesterases, activation of PI-PLC probably involves as yet undefined G proteins. There are at least five isoenzymes of PI-PLC in mammals; these are subdivided into the four classes α–δ and differ widely in their molecular size and amino acid sequence. PLC-α has a mass of 56.6 kDa (cDNA from rat basophilic leukaemic cells), PLC-β 138.2 kDa, PLC-γ 148.4 kDa and PLC-δ 85.8 kDa (all cDNAs from rat or bovine brain). The sequences of PLC-β, -γ and -δ are only similar in certain regions, and PLC-α is completely different from the other three classes. Thus, it is surprising that these enzymes all have very similar catalytic properties. They hydrolyse phosphatidylinositol (PI), phosphatidylinositol-4-phosphate (PIP) and phosphatidylinositol-4,5-bisphosphate (PIP_2), but not phosphatidylcholine or phosphatidylethanolamine, and they are Ca^{2+}-dependent [133]. A cDNA and corresponding PLC protein isolated from *Drosophila melanogaster* shows extensive similarity in its sequence of 1095 amino acids to PLC-δ from bovine brain. This is a product of the gene norpA (no receptor potential A); mutants with this gene have no PLC activity in their eyes, and consequently no receptor potential and are blind. Thus, the PLC encoded by norpA apparently belongs to the phototransduction system of the *Drosophila* eye [170]. A further PLC, expressed in *Drosophila* brain, has been sequenced via the cDNA and designated plc-21. Its sequence of 1305 or 1312 amino acids is clearly different from the norpA sequence [149]. A PLC with a unique specificity is found in *Trypanosoma brucei*. This cleaves the glycosyl phosphatidylinositol, by means of which the variable surface glycoprotein (VSG) is anchored in the membrane, but has little activity

against free phosphatidylinositol. The enzyme is activated under certain conditions and then releases the VSG. The sequence of this PLC has no significant similarity to the PI-PLCs of mammals or to other proteins [58]. **Phospholipases D** are widespread in plants but are probably of no importance in animals [175].

17.2 Phosphatases and Phosphodiesterases

Amongst the phosphatases (phosphomonoesterases), membrane-bound alkaline phosphatase and the mainly lysosomal acid phosphatase are ubiquitous in their distribution. Both enzymes act upon a large variety of substrates. In addition, there is a series of phosphomonoesterases of narrower specificity, e.g. for sugar phosphates, acyl phosphates, dolichyl phosphates or phosphoproteins. Phospholipases C and D and the nucleotidases are also specific phosphomonoesterases; although these enzymes are found in vertebrates of all classes, there have been few comparative biochemical studies [41]. The enzymes specific for cyclic nucleotides are amongst the most interesting of the phosphodiesterases because of their importance in cellular regulation. The nucleases are also phosphodiesterases, and in this case there is a wealth of comparative biochemical data. A pyrophosphohydrolase degrades diguanosine tetraphosphate Gp_4G, a purine reserve of the brine shrimp *(Artemia salina)*, asymmetrically into GTP and GMP, and also cleaves the adenosine compounds Ap_4A and Ap_4G. Asymmetric Np_4N-pyrophosphohydrolases have also been isolated from various mammalian tissues [128].

17.2.1 Alkaline Phosphatases

The membrane-bound alkaline phosphatases in mammals have their highest activities in the gut and placenta, followed by bone and kidney; lower activities are found in the liver, lung and spleen. These are dimeric glycoproteins of about 160 kDa with two zinc ions per subunit. An enzyme that has been partially purified from the lizard *Agama agama* has similar characteristics. Under physiological conditions, the mammalian enzyme shows such strong negative cooperativity that only one of the two substrate-binding centres is actually active; however, at the optimal pH of 9 the enzyme shows Michaelis-Menten kinetics. At least two **isoenzymes** of alkaline phosphatase are present in mammals; these are distinguished by their inhibition by the amino acids L-phenylalanine and L-homoarginine. The organ-specific isoenzyme of the liver, bone, kidney and placenta has a higher affinity for homoarginine and the intestinal enzyme has a greater affinity for phenylalanine. A further organ-specific isoenzyme is found in the human placenta and in that of other higher primates; in contrast to the isoenzyme of the liver, bone and kidney (L/B/K isoenzyme) this has a higher affinity for L-phenylalanine than for homoarginine. The human placental enzyme is first expressed in the 4th month; up to this time, human placenta, like that of other mammals, contains the L/B/K isoenzyme. The human enzyme is unique in showing high polymorphism, with 3 common and at least 15 rarer variants. The genes of the intestinal and placental isoenzymes, together with related genes such as placental-like and foetal-intestinal, form a cluster on chromosome 2; the gene for the L/B/K isoenzyme is found on chromosome 1, and because of its large intron of more than 50 kb is at least fivefold larger than the other genes. The human intestinal pre-phosphatase agrees in 86.5 % of its 528 amino acids with the placental isoenzyme but only 56.6 % with the L/B/K isoenzyme. The L/B/K isoenzyme was apparently the first to separate in the evolution of the alkaline phosphatases, and the placental isoenzyme first appeared in the primates. Human and rat L/B/K pre-phosphatases have 88 % identical amino acids, and the human placental alkaline phosphatase still agrees 33 % with the alkaline phosphatase of *E. coli*. Thus, the rate of evolution of the alkaline phosphatases is relatively low [57, 78, 167].

The variety of the alkaline phosphatases is increased by the presence of differentially glycosylated forms in different organs. In a large comparative study, an organ-specific intestinal isoenzyme was found in all mammals and birds examined as well as in the tortoise *Geoclemmys reevesi*. In contrast, the gut, liver and kidney of the lizard *Anolis carolinensis*, the snake *Elaphe climacophora*, and various amphibians and fish all contained an isoenzyme that is resistant to both L-phenylalanine and homoarginine, and consequently represents a new type of alkaline phosphatase [185]. An alkaline protein is also a normal component of milk. The enzyme finds its way into the milk-gland secretion during the merocrine secretion of milk fat together with the membranes of the ejected cells. The milk enzyme corresponds immunologically to the organ-

specific isoenzyme but differs from the liver enzyme in its sialic acid content [54].

Alkaline phosphatases have been partially purified from and characterized for a series of **invertebrate animals**, e.g. annelids [130], crustaceans [27, 131], insects [63, 116] and molluscs [129]. Although many of these investigations stem from the same laboratory, they present a quite heterogeneous picture. Electrophoretically or chromatographically distinct isoenzymes were only observed in a minority of species, e.g. four in *Spirographis* and no fewer than nine in *Culex*. The molecular masses, as determined by gel filtration, range from 285 to 295 kDa in *Squilla* and *Helix* (Cepaea) to 800 kDa or more in *Asteria* and *Octopus*. Values between 207 and 609 kDa were recorded for the four isoenzymes from *Spirographis*. The pH optima range from 8 *(Culex)* to 11 *(Asterias, Octopus)*. The substrate specificity of the invertebrate phosphatases appears to be even lower than that of the vertebrate phosphatases. Determinations for *Squilla* by using 4-nitrophenylphosphate showed that ATP, ADP, AMP, glucose-1-phosphate, glucose-6-phosphate and even inorganic pyrophosphate were competitive inhibitors but cAMP was not; the affinities for ATP and ADP were even higher than for 4-nitrophenylphosphate. Nucleotides and sugar phosphates were also competitive in the annelids, but ATP and ADP were either non- or uncompetitive inhibitors. The annelid enzymes are inhibited by phenylalanine but those of the crustaceans and molluscs are not; cysteine is much more effective than phenylalanine as an inhibitor in *Culex*.

17.2.2 Acid Phosphatases

Most vertebrate tissues contain several **cell compartment-specific isoenzymes** of acid phosphatase which are clearly encoded by different genes. The heterogeneity of the enzymes is further increased by post-translational modifications, in particular by variable glycosylation. Mammals have a microsomal type I of 200 kDa, a lysosomal type II of 90–120 kDa and two cytosomal types of lower molecular mass: type III of 40 kDa and type IV of 8–18 kDa. The isoenzymes also differ in their substrate specificity and response to inhibitors: I and II have a broad specificity and, in contrast to III and IV, are inhibited by fluoride and tartrate; III appears to act preferentially on nucleotide triphosphates, whereas IV is specific for flavin monophosphate and 17β-oestradiol-3-phosphate. However, the biological roles of the different isoenzymes are far from clear. Unfortunately, very few sequences have been determined so far. The acid phosphatase from lysosomes of the rat liver has a transmembrane segment and nine N-glycosylation sites in a polypeptide chain of 393 amino acids. It is synthesized as a membrane protein with a signal sequence of 30 amino acids, transported to the lysosome via the Golgi apparatus without a mannose-6-phosphate marker, and is then partly released into the lysosomal matrix after cleavage of a C-terminal fragment [60]. The human and rat lysosomal enzymes agree 89 % in their sequences and the soluble prostate enzymes agree 75 %; in either species these two enzymes agree by only about 50 % [125, 138].

Lower vertebrates also have acid phosphatases but these are not directly comparable with the isoenzymes of the mammals. The liver of the frog *Rana esculenta* contains four isoenzymes with different substrate specificities and molecular masses of 240, 110, 38 and 17 kDa [68, 120]. The acid phosphatase from the tadpole tail tissue of *Xenopus laevis* has received particular attention; four isoenzymes are recognised and the activity increases markedly during metamorphosis. Three isoenzymes, I–III, can be distinguished chromatographically in the carp liver. These are all dimers with subunits of 59 kDa (I and II) or 46.5 kDa (III). Isoenzyme I is apparently a sialic acid-rich derivative of isoenzyme II [119]. The liver phosphatase of the catfish *Ictalurus punctatus* is a homodimer of 35-kDa subunits, the two forms of which differ only in their carbohydrate fractions [67].

The acid phosphatases of many *Drosophila* species have been thoroughly investigated genetically because of their marked polymorphism. Null variants of acid phosphatase exist and show no reduction in viability or fertility under laboratory conditions. The acid phosphatases of *Drosophila*, like the lysosomal isoenzymes of the vertebrates, are homodimeric glycoproteins with 50 to 60-kDa subunits. They are also inhibited by tartrate and fluoride and cleave various phosphomonoesters, including AMP and glucose-6-phosphate. Choline phosphate is metabolized by the *Drosophila* phosphatase to the same limited extent as it is metabolized by the liver enzyme, but it is attacked by the prostate enzyme [109]. Five acid phosphatase isoenzymes have been isolated from the silkworm *Bombyx mori* [62]. Several acid phosphatase (AcP) isoenzymes from the eggs and embryos of the sea urchins have also

been described. Of the two soluble enzymes, AcP-1 resembles the cytoplasmic isoenzyme IV of the mammals in that it has a molecular mass of 14–16 kDa, metabolizes mainly FMN as well as 4-nitrophenylphosphate, and is not inhibited by fluoride. AcP-2 resembles the cytoplasmic isoenzyme III of the mammals in its molecular mass of 42–48 kDa and its relative specificity for ATP and ADP; however, in contrast to the mammalian enzyme, it is inhibited by fluoride. In addition to the soluble enzymes, sea-urchin eggs also contain a phosphatase of 600–700 kDa that is bound to the yolk platelets and, like the typical lysosomal enzyme, is inhibited by fluoride and tartrate. The yolk platelets contain further typical lysosomal enzymes and therefore probably originate from lysosomes; however, the mechanism that prevents the release of stored nutrients before fertilization but activates release thereafter is not known [184]. The ciliate *Tetrahymena* possesses a phosphatase that corresponds to the lysosomal prostate enzyme in its pH optimum of 4.5–5.0, its dimeric structure with subunits of about 55–60 kDa and its broad specificity; unlike the mammalian enzyme, it attacks the terminal phosphoric acid bonds of ATP and ADP and is secreted into the environment [5].

17.2.3 Substrate-Specific Phosphomonoesterases

Fructose-1,6-bisphosphatase (FBPase) is a key enzyme of gluconeogenesis. However, it can also form a substrate cycle (futile cycle) together with phosphofructokinase (PFK); this may enhance the allosteric effects at this important glycolysis regulation point or simply serve to produce heat (p. 527). As described in detail in Chapter 14, FBPase and PFK are counter-regulated; AMP and fructose-2,6-bisphosphate inhibit the phosphatase and simultaneously stimulate the kinase. The liver and kidney have by far the highest FBPase activities, but it is found in most other tissues. Two or three organ-specific isoenzymes of FBPase can be distinguished in the mammals; there is consistently a liver type and a muscle type that is also present in, for example, the kidney. Immunological, electrophoretic and kinetic investigations have further shown the existence of a third small intestine-specific isoenzyme in the mouse, rat and golden hamster but not in the guinea-pig, rabbit or dog [101].

All FBPases are heterotetramers with subunits of about 36.5 kDa. The 335- to 336-amino-acid sequences of the enzymes from the pig kidney and sheep liver are known; although they belong to different isoforms, they differ in only 34 amino acids. In contrast, rat liver contains an FBPase which has a 27-amino-acid C-terminal extension and with a mass of 40 kDa is therefore somewhat larger; however, it shows no differences in functional properties. Comparative investigations of the immunological cross-reactivity and molecular masses of enzymes from rodent species from five families showed that the 40-kDa FBPase only occurs in wild and laboratory brown rats *Rattus norvegicus* and the domestic rat *Rattus rattus* [39, 134]. The FBPase from rat brain is also atypical. Whilst the FBPase from bovine brain, like other mammalian FBPases, has a K_m of 2 μmol/l, an absolute requirement for divalent metal ions, and is non-competitively inhibited by AMP, the enzyme from rat brain has a K_m of 130 μmol/l, no requirement for exogenous metal ions and no AMP inhibition [83].

In the vertebrates the highest activities of **glucose-6-phosphatase** are found in the liver and kidney, where it is involved in blood sugar production. The enzyme is intrinsic to the membrane of the endoplasmic reticulum and consists of two components: a highly specific translocase that transports glucose-6-phosphate from the cytoplasm into the lumen of the endoplasmic reticulum (or into isolated microsomes), and a non-specific hydrolase that only metabolizes substrates presented from the lumen side of the membrane. These vectorial properties of the native enzyme explain the different substrate specificities and affinities of microsomes after treatment with detergents. The glucose-6-phosphatase from the liver of the carp *Cyprinus carpio* appears to be very similar to the enzyme from mammalian liver [154]. The activity in vertebrate skeletal muscle is always very low, but the activity is somewhat higher in heart muscle. Much higher activities are found in the flight muscles of Hymenoptera, Diptera and some Lepidoptera; the flight muscles of the wasp *Vespa vulgaris* and some species of the bumble bee genus *Bombus* have activities as high as those in mammalian liver. It is possible in this case that the enzyme is part of a substrate cycle allowing more effective regulation of glucose phosphorylation.

The biological function of **trehalose-6-phosphatase** is the cleavage of trehalose-6-phosphate, the initial product of trehalose biosynthesis. However, the enzyme from the fat body of the fly *Phormia regina* can attack glucose-6-phosphate, and this activity is stimulated by free

trehalose. In contrast, the corresponding enzyme from the cockroach *Periplaneta americana* cannot cleave glucose-6-phosphate. To determine the distribution of these two different enzyme types, the enhancement of glucose-6-phosphate cleavage by trehalose has been investigated in crude extracts for many insect species. The stimulation factors are 4.0 and about 2.0 for the purified enzyme and crude extract, respectively, of Phormia, and 1.0 for the two types of preparation from *Periplaneta*. Values above 1.0 are found in Diptera from the families Sirphidae, Sarcophagidae, Calliphoridae, Tachinidae and Tephritidae, but not from the Culicidae, Tipulidae and Drosophilidae or from non-dipterans. The reason for the occurrence of a special type of trehalose-6-phosphatase in only one evolutionary line of the Diptera is a matter for debate [47].

The **substrate-specific phosphatases** also include the phosphoprotein phosphatases, described in Chapter 3. Enzymes that specifically cleave acylphosphate bonds are widely found in the muscles of vertebrates and invertebrates. Three types are found in vertebrate muscle. In Type I, the polypeptide chain of the enzyme is linked to glutathione by a disulphide bridge, and type II is a disulphide dimer of the chain itself. The structure of the third type, which makes up only a few percent, has not yet been determined. The sequences are known for several of these enzymes. Acylphosphatase 1 from turkey muscle consists of 120 amino acids; 25-Cys is bound to glutathione by a disulphide bridge. The corresponding enzyme from the horse is 4 amino acids shorter at the N-terminus but otherwise shows only 17 amino acid differences to the turkey enzyme [22]. A phosphatase that is specific for dolichol phosphate, but was previously known only from bacteria and higher plants, has also been found in the ciliate *Tetrahymena pyriformis;* the enzyme has a pH optimum of about 5–6 and is only slightly inhibited by tartrate. An equally unusual enzyme from the sea anemone *Metridium senile* is specific for phosphonomonoester bonds, e.g. between aminoethyl phosphonate (AEP) and glycoproteins. This is a zinc protein that can be assayed with 4-nitrophenylphosphonate and shows no activity with phosphoric acid esters.

17.2.4 cAMP and cGMP Phosphodiesterases

The cAMP and cGMP phosphodiesterases convert the cyclic nucleoside monophosphates cAMP and cGMP to the corresponding 5′-nucleoside monophosphates, and thereby regulate the cAMP and cGMP levels. Several of these enzymes are allosterically regulated by cyclic nucleotides and others are regulated by interaction with proteins, e.g. Ca^{2+}-calmodulin or transducin. The vertebrates have various **isoenzymes** of these phosphodiesterases with different properties and organ distribution. This multiplicity of forms has an important, if not completely understood, role in the specific cell reactions to various hormones. The number of isoenzymes varies from one (uterus) to six (cerebellum), but only three types have been investigated in detail. The first type includes cytoplasmic enzymes with low affinity which are activated 5- to 60-fold by Ca^{2+}-calmodulin. They have a significantly higher affinity for cGMP than for cAMP, although cAMP is turned over more rapidly. These enzymes consist of subunits of about 60 kDa which are present in various degrees of aggregation. The second type is insensitive to Ca^{2+} and calmodulin, also consists of subunits of 50–60 kDa, but is probably membrane bound. For the third type, cAMP hydrolysis is stimulated by cGMP. The activites with cAMP and cGMP are about equal but the affinity for cGMP is much higher [111].

The proportion of the different phosphodiesterases differs greatly between organs. The first type is found in only trace amounts in the liver, and predominates in the heart and brain. The second type is present in many cells, and the third type is reported to be present in the adrenal gland and bovine heart. Despite considerable functional differences, these enzymes are apparently closely related; specific cGMPase and specific cAMPase from porcine kidney show immunological cross-reactivity [111]. The calmodulin-regulated phosphodiesterase from bovine brain, the cGMP-regulated enzyme from bovine heart, the cAMP-regulated enzyme encoded by the $dunce^{+}$ gene of *Drosophila melanogaster*, and the cAMP phosphodiesterase encoded by the PDEE21 gene of bakers' yeast differ in size between 40 and 105 kDa, but all four enzymes have a homologous segment of about 275 amino acids in the region of the catalytic domain [25, 33]. The cGMP phosphodiesterase in the outer segment of the rods in the vertebrate eye and its role in the process of phototransduction will be dealt with in Chapter 19.

Whereas Ca^{2+}-calmodulin-activated phosphodiesterases are present in almost all vertebrate tissues, they have seldom been found in the insects, despite the fact that insects have both calmodulin

and various phosphodiesterases. About 40 % of the cAMP-specific phosphodiesterases in the head of *Drosophila* are calmodulin-dependent. A calmodulin-activated phosphodiesterase has been purified from the fat body of the silkworm *Bombyx mori* and accounts for 3–5 % of the total activity. The enzyme cleaves cAMP and cGMP, is activated two- to fourfold by Ca^{2+}-calmodulin and has a mass of 115 kDa. Kinetic analysis suggests that the enzyme has two substrate-binding sites; the G site is catalytic only for cGMP, but the binding of cAMP as well as cGMP to this site has allosteric effects that reduce the affinity of the A site for its substrate, cAMP [106]. The dunce mutants of *Drosophila melanogaster* have a defective memory and reduced activity of the cAMP-dependent phosphodiesterase but have normal activity of the calmodulin-regulated enzyme. The dunce gene has a length of more than 100 kb and contains several other genes within its enormous intron of 80 kb. The dunce$^+$ protein has 362 amino acids and shows significant similarity to the vertebrate phosphodiesterases. Several homologues of the dunce gene have been identified in the rat genome. One of these encodes a polypeptide of 610 amino acids; in the domain of 275 amino acids conserved in all phosphodiesterases this shows no less than 75 % sequence agreement with the *Drosophila* enzyme [33, 164].

17.2.5 Nucleases

Amongst the polynucleotide-cleaving enzymes, a distinction based on substrate specificity can be drawn between DNases, RNases and non-specific nucleases; and based on the mode of action a distinction can be drawn between endo- and exonucleases. The endonucleases produce either 5′- or 3′-phosphate, and the exonucleases attack either from the 5′ or the 3′ end. Endonucleases are usually base specific and preferentially or exclusively cleave phosphodiester bonds of certain nucleotide residues. Restriction endonucleases with specificity for certain recognition sequences of four or six nucleotides are found in large numbers in microorganisms but are apparently not present in animals. Individual examples of all the other nucleases mentioned above are known in animals [93]. An interesting example is a uracil-specific deoxyribonuclease from *Drosophila melanogaster;* it is present almost exclusively in the third larval stage. Whereas this enzyme can cleave already present uracil-containing DNA, the deoxyuridine triphosphatase (dUTPase) in the embryo can prevent the incorporation of uracil into DNA by removal of the precursor dUTP [50].

Most vertebrate ribonucleases are specific for the 3′-phosphate groups of pyrimidine nucleotides and reduce RNAs to a mixture of mono- and oligonucleotides with terminal pyrimidine-3′-phosphate. They make up a protein super-family which also includes the angiogenic peptide angiogenin. A distinction can be made between secretory and non-secretory pyrimidine-specific RNases. The secretory enzymes are found in almost all secretory tissues and have pH optima of 8.0; typical examples are the pancreas RNases. The bovine RNase produced in the seminal vesicles and excreted in large quantities in the seminal plasma agrees up to 81 % with the pancreas RNase [163]. The non-secretory RNases have optimal activity at pH 6.5–7.0 and are present, in particular, in liver, lung, spleen and leukocytes. In most tissues these enzymes are found complexed with a ribonuclease inhibitor. The bovine liver RNase is only 47 % similar to that from the pancreas [61]. Both RNases are found together in body fluids such as urine, blood plasma and semen. The non-secretory RNase from human urine has about 30 % sequence similarity to the known secretory RNases of mammals [10].

The **pancreas RNases** have especially high activities in ruminants and in mammals with ruminant-like or caecal digestion. They apparently serve to degrade the large amounts of ribonucleic acids that enter the gut with the symbiontic microorganisms from the stomach and the caecum. About 40 pancreas RNases from mammals have been sequenced so far; most have 124 amino acids and some are extended C-terminally to 126 (horse) or 128 amino acids (two-toed tree sloth *Choloepus hoffmanni* and several RNases from the Hystricomorpha). The RNase of the turtle *Chelydra serpentina*, the only one known outside of the mammals, shows 70 differences in its 119 amino acids to the bovine enzyme. Amongst the mammals, closely related species such as the domestic cow *(Bos taurus)* and the bison *(Bison bison)*, or the sheep *(Ovis aries)* and the goat *(Capra hircus)*, have identical sequences. In contrast, the bovine and porcine RNases already differ at 26 positions. The guinea-pig *(Cavia porcellus)* possesses two non-allelic isoenzymes which differ in 31 amino acids. Allelic polymorphism of individual positions has been observed in various species [11, 145].

The asparagine residues at positions 21, 34, 62 and 76 can be glycosylated; however, the **carbohydrate content of pancreas RNases** varies widely

with the species. The RNases of herbivores in which the caecum functions as a fermentation chamber, e.g. the pig, horse and rodents, are particularly highly glycosylated. In contrast, the RNases of herbivores in which a part of the stomach serves as a fermentation chamber inhabited by symbionts, are only weakly glycosylated (ruminants, hippopotamus *Hippopotamus amphibius*, tree sloth *Choloepus hoffmanni)* or may even be completely lacking in carbohydrate (kangaroo *Macropus giganteus)*. The individual RNase molecules of a single animal may be variably glycosylated. The level of glycosylation clearly depends upon the activity of the corresponding enzymes (glycosyl-transferases), but also to some extent upon the nature of the amino acids adjacent to the asparagines. Thus, for example, there are two variants of the sequence 34–37 (Asn-Met-Thr-Gln/Lys) in the hippopotamus, and only that with 37-Gln is glycosylated [8, 9, 11].

17.3 ATPases (With an Excursus on Ion Channels)

Instead of ester bonds, some enzymes hydrolyse acid anhydride bonds of phosphoric acid. Apart from ATPase, this class of enzymes also includes, for example, **apyrase**, which cleaves terminal phosphates from ATP and ADP, thereby degrading ATP to AMP and two P_i residues. This enzyme was discovered in yeast in 1945 but its existence in animals was for a long time controversial as the same reaction can also occur by the interaction of (myosin) ATPase and adenylate kinase. However, genuine apyrase has since been demonstrated in the pancreas of the rat and pig, albeit without any indication of its function. The enzyme is found in the salivary glands of a wide range of blood-sucking arthropods, e.g. the bug *Rhodnius prolixus*, the mosquito *Aedes aegypti*, the tsetse fly *Glossina austeni* and the tick *Ixodes dammini*, where it probably has an antihaemostatic effect [143]. A further unusual phosphoric acid anhydride hydrolase is the **diguanosine tetraphosphatase** which cleaves the diguanosine tetraphosphate of the brine shrimp *Artemia salina* into GTP and GMP. A protein with the same specificity but somewhat larger (20 instead of 17.5 kDa) is also found in the rat liver; both enzymes cleave various dinucleoside tetraphosphates and would be better referred to as dinucleoside tetraphosphatases. They are inhibited by nucleoside tetraphosphate [171].

In the dynamic equilibrium of living organisms, just as much ATP is cleaved as is produced. The release of the two energy-rich bound phosphates of ATP can also occur via indirect pathways, e.g. transfer reactions produce phosphoric acid esters which are later hydrolysed by phosphatases, or there may be cleavage of the inorganic pyrophosphate ions arising from ligase reactions. Quantitatively more important is the cleavage of ATP by myosin ATPase during contractile processes. However, by far the greatest proportion of ATP consumption lies with the ion-transporting membrane ATPases, the properties and evolution of which will be discussed in this section. The membrane-bound ATPases also include the terminal transphosphorylase of oxidative phosphorylation; this normally makes use of an electrochemical gradient for the synthesis of ATP from ADP and inorganic phosphate, but can also cleave ATP. This will be discussed in connection with the electron transport system of the mitochondria. The membrane ATPases responsible for active ion transport will be discussed together with ion channels and the Na^+-dependent membrane transporter [156].

17.3.1 Na^+,K^+-ATPase

The best-known ion-stimulated ATPase is the Na^+,K^+-ATPase of the cell membrane; as the "sodium–potassium pump" this removes excess sodium from the cell and admits potassium. This enzyme is found in all cells, and apart from in various mammalian organs has been studied especially in the electric organs of fish [75, 115], in the insects *Drosophila melanogaster* and *Locusta migratoria* [1, 81] and in the nauplius larvae of the brine shrimp *Artemia salina* [7, 15]. The Na^+,K^+-ATPase is an integral membrane protein that directly couples the hydrolysis of ATP with the counter-current transport of sodium and potassium ions. The enzyme consists of a catalytic α subunit of about 112 kDa and a β subunit of 35 kDa (without the carbohydrate portion) which is a sialoglycoprotein of unknown function [3, 17]. It is not clear how far a small membrane protein of 7.7 kDa (68 amino acids) that is also present during the isolation of the α and β subunits should be considered as the "γ subunit" of Na^+,K^+-ATPase [30]. The **functional cycle** of the Na^+,K^+-pump begins with the binding of an ATP and three sodium ions to the cytoplasmic side of the α subunit. As with all ATPases, the real substrate is an Mg^{2+}–ATP com-

plex. Phosphorylation of a certain aspartate residue brings about a change in conformation and ion affinity; as a result, the three sodium ions are released to the outside of the cell membrane and two potassium ions are bound. The subsequent hydrolytic cleavage of the phosphate residue reverses the conformational change, the potassium ions are released into the cytoplasm, and the cycle begins again. Thus, approximately three sodium ions are transported inward and two potassium ions are transported outward per ATP. The majority of the Na^+,K^+-ATPases are inhibited by cardiac glycosides such as digoxin and ouabain. The described conformational change involves a region of more than 80 amino acids; this changes between an α-helix and a β-fold according to the state of phosphorylation/dephosphorylation. Because the binding of Na^+ or ATP to the cytoplasmic side and the binding of K^+ or ouabain to the outside influences the conformation, an energy transfer must occur between the cytoplasmic, transmembrane and extracellular domains.

The number of ATPase molecules can be determined from the **ouabain binding;** the combined activities of cellular ATPase activity and the Na^+,K^+-ATPase give a turnover rate, e.g. in the Malpighian tubules of the migratory locust *Locusta migratoria*, of 2645 ATP per ATPase molecule per minute [1]. However, ouabain-insensitive Na^+,K^+-ATPases also exist. In fibroblast cell cultures mutants can be induced in which ouabain sensitivity is eliminated but ion transport is unaffected. The planarian *Dugesia gonocephala* is insensitive to as much as 8 mol/l ouabain. Ouabain-insensitive Na^+,K^+-ATPases have also been found in the gills of various molluscs. The reduced ouabain sensitivity of the Na^+,K^+-ATPase from the caterpillars of the monarch butterfly *Danaus plexippus* makes good biological sense: this species feeds on milkweed and consequently receives considerable quantities of cardiac glycosides in its diet. The differences in ouabain sensitivity are determined not so much by the binding of ouabain as by its dissociation. Sensitive enzymes form a stable, enzymatically inactive complex with the glycoside, whereas insensitive enzymes lack the lipid boundary that retards dissociation of the complex [151]. All Na^+,K^+-ATPases are inhibited by orthovanadate, and herein lies the biological role of this trace element.

The mammals possess tissue-specific isoenzymes of Na^+,K^+-ATPase which have different α chains (α1, α2, α3) but always the same β chain; these differ in Na^+ affinity, regulation by insulin and thyroid hormone, and sensitivity to cardiac glycosides. In the rat, α1 (previously α) is expressed mainly in the kidney and lungs, α2 (previously α^+) is expressed in muscle, and α3 (previously αIII) is expressed in the brain. Foetal and neonatal heart tissue contains predominantly α3, whereas juveniles and adults possess α2. Pairwise comparison of the α subunits shows 85–86% sequence agreement. Three homologous α subunits are also recognized in humans; however, the human genome contains at least five α genes or pseudogenes [117]. Isoforms of Na^+,K^+-ATPase have also been observed in invertebrates. The nauplius larvae of the brine shrimp *Artemia salina* show polymorphism of the catalytic subunit α-Gly/Lys; the forms probably also differ in carbohydrate content [107].

The α chains have lengths of 1013–1018 amino acids. In most cases, the primary translation product has an additional five N-terminal amino acids that are removed post-translationally; however, a signal peptide as such is not found. The domain structure in mammals and electric rays is identical. The N-terminal-third of the molecule has four hydrophobic transmembrane domains (H1–H4); this is followed by the cytoplasmic domain of more than 400 amino acids, and this carries, e.g. in the enzyme of sheep kidney, the phosphorylation site at 369-Asp and the ATP-binding site at 501-Lys. Four further transmembrane domains (H5–H8) are found toward the C-terminal end of the molecule. Ouabain binding to the Na^+,K^+-ATPase of sheep kidney is probably at 310-Trp [151]. The Ca^{2+}-ATPase from the sarcoplasmic reticulum of rabbit muscle has a similar molecular organization. A model derived for this latter molecule, based upon electron microscope observations, shows a stemmed sphere and this can probably also be assumed for the Na^+,K^+-ATPase (Fig. 17.3).

The sequences of the human, rat, sheep and pig α subunits agree in 96–97% of their amino acids; the α1 chains of the chicken and ray *Torpedo* are also very similar to each other [74]. The genome of *Drosophila melanogaster* contains only one gene for the α chain of Na^+,K^+-ATPase; the encoded sequence of 1038 amino acids agrees by about 80% with that of the mammalian α subunits [81]. In contrast, the α subunit of the brine shrimp *Artemia salina* is shorter at the N-terminus compared with those of both mammals and *Drosophila;* it has a length of only 996 amino acids. The sequence agreement with other α subunits amounts to only 69–72% [7]. The β subunits of

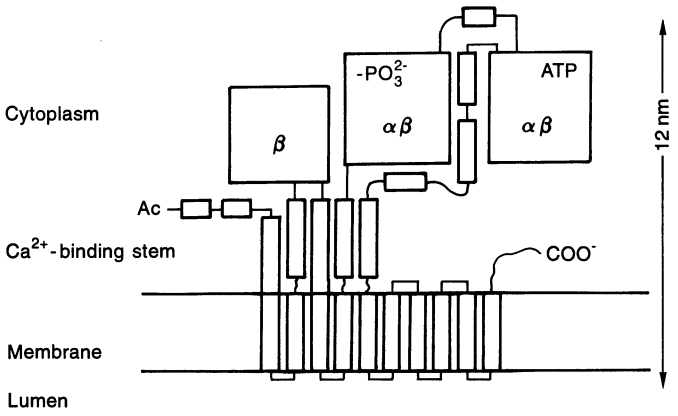


Fig. 17.3. A model of the Ca^{2+}-ATPase from the sarcoplasmic reticulum of rabbit muscle [89]. For clarity, the arrangement of the α-helices *(small rectangles)* and the main domains with αβ-sheet or αβ structure *(squares)* are drawn in one plane, in fact, the ATP-binding domain lies close to, and interacts with, the two other main domains

the Na^+,K^+-ATPases from mammalian kidney consist of 302 amino acids. The N-terminal amino acids 1–33 probably lie in the cytoplasm. There follows a transmembrane domain that spans positions 34–60, and the rest of the chain is extracellular. The three oligosaccharides on the β subunit are apparently bound to the asparagine residues 157, 192 and 264, and these lie in typical -Asn-X-Ser/Thr- sequences. The α and β subunits are bound by non-polar interactions [150]. The β subunit of the Na^+,K^+-ATPase from the brine shrimp *Artemia* is significantly homologous to those of various mammals, the chicken, *Xenopus* and *Torpedo*, although there are considerable differences, particularly in the sequence of the cytoplasmic, N-terminal domain [15].

17.3.2 Ca^{2+}-ATPase

The cytosolic concentration of free calcium plays an important role in all cells in the control of various physiological processes. The Ca^{2+} concentration in resting cells is about 0.1–0.2 μmol/l and can increase by two orders of magnitude after stimulation. The calcium concentration of the extracellular fluids is about four orders of magnitude higher than the cytosolic resting level. All cells possess several mechanisms for regulating the calcium level: voltage- or ligand-dependent ion channels allow the entry of calcium into cells along the electrochemical gradient; cell organelles with calcium-binding proteins and ATP-driven calcium pumps serve in the intracellular storage of Ca^{2+} ions; and Ca^{2+}-ATPases and Na^+,Ca^{2+}-exchange proteins in the plasma membrane remove excess calcium from the cell. The calcium-storing cell organelle of the skeletal muscle is the sarcoplasmic reticulum, and in non-muscle cells it is the endoplasmic reticulum or the recently described "calcisome". The Ca^{2+}-ATPases of the sarco(endo)plasmic reticulum are known as SERCA and those from the plasma membrane as PMCA [18]. The Ca^{2+}-ATPase of the plasma membrane is activated by calmodulin. Phosphorylation is involved in the regulation of the muscle enzyme; this is achieved by a cAMP-dependent protein kinase in heart and skeletal muscle and by a cGMP-dependent enzyme in smooth vascular muscle [53]. The best-known Ca^{2+}-ATPase is that from the **membrane of the sarcoplasmic reticulum** (SR). The enzyme makes up about 90 % of the total membrane protein and can therefore be easily isolated. The functional cycle corresponds to that described above for the Na^+,K^+-ATPase, in this case with two Ca^{2+} transported per ATP. Two calcium ions and Mg^{2+}-ATP are bound to the cytosolic side with high affinity; a conformational change, triggered by a phosphorylation, brings the calcium into the lumen of the SR vesicle, where its reduced binding affinity results in its release. Dephosphorylation restores the original high-affinity state of the enzyme; the inorganic phosphate ion is released into the cytosol. Regulation of the Ca^{2+}-ATPase involves a 22-kDa protein of the SR membrane, **phospholamban**, which is phosphorylated at one site by a cAMP-dependent protein kinase and at another site by a calmodulin-dependent kinase; phosphorylation of the two sites brings about an independent but additive stimulation of Ca^{2+}-ATPase [132].

All Ca^{2+}-ATPases belong to the same protein super-family as the other ATP-driven ion pumps, produce a phosphorylated intermediate during the functional cycle, and have a similar molecular structure (Fig. 17.3). Despite their undoubted homology, the sequence agreement between Ca^{2+}-ATPases and Na^+,K^+-ATPases only exceeds 30 % in particular regions. In addition to the

highly conserved domains that are responsible for ATP binding and formation of the phosphorylated intermediate, the Ca^{2+}-ATPases of the plasma membrane contain a further regulatory domain with the calmodulin-binding site and phosphorylation site. The regulatory domain consists of 120 amino acids in a C-terminal extension and is responsible for the larger molecular mass of PMCA (about 1200 amino acids = 135 kDa) compared with that of SERCA, Na^+,K^+-ATPases and K^+,H^+-ATPases (1000–1043 amino acids = 100–110 kDa) [160]. The mammals possess tissue-specific isoenzymes of SERCA and PMCA and these are encoded by different genes or arise by alternative splicing. PMCA1 is the "housekeeping form" of the enzyme present in the plasma membrane of all cells; PMCA2 is mainly expressed in the brain and heart, and PMCA3 in the brain and skeletal muscle. The three isoenzymes show 81–85 % sequence agreement [53, 160]. The sarcoplasmic reticulum of adult fast-twitch skeletal muscle contains the isoenzyme SERCA1a, whereas neonatal muscle contains the alternatively spliced SERCA1b. Slow-twitch skeletal muscle and heart muscle are characterized by the presence of SERCA2a, and smooth muscle and non-muscle cells have the alternatively spliced SERCA2b. SERCA3, which agrees with the other isoenzymes in 75–77 % of positions, is found in various muscle and non-muscle tissues [18, 187]. Ca^{2+}-ATPases are found in all animals down to the protozoans [49]. The enzyme of the brine shrimp *Artemia salina* agrees in about 70 % of its 1003 amino acids with the mammalian SERCAs [118].

17.3.3 K^+,H^+-ATPase and Anion-Sensitive ATPase

K^+,H^+-ATPases are enzymes that generate proton gradients by the hydrolysis of ATP or, conversely, exploit the energy potential of H^+ gradients for ATP synthesis. They can be subdivided into three families. The K^+,H^+-ATPases of mammalian stomach and the H^+-ATPase of the plasma membrane belong, together with the Na^+,K^+-ATPase and Ca^{2+}-ATPase, to the P-ATPases, which produce a phosphorylated intermediate during the functional cycle. These have a catalytic subunit of about 100 kDa and an H^+/ATP stoichiometry of 1.0. F-ATPases (or F_0F_1 ATPases) are found in the plasma membranes of the Eubacteria and in the inner membrane of the mitochondria and chloroplasts; they function without a phosphorylated intermediate mainly in the synthesis of ATP, but they can also cleave ATP and pump H^+ ions with an H^+/ATP stoichiometry of 3.0. They are made up of multi-subunit complexes of about 500 kDa. V-ATPases are components of the inner membranes of eukaryote cells, e.g. those of vacuoles, lysosomes, Golgi vesicles and coated vesicles. They transport H^+ ions with an H^+/ATP stoichiometry of 2.0. Like the F-ATPases, they produce no phosphorylated intermediate and are multi-subunit complexes of about 500 kDa. However, they are sufficiently different to the F-ATPases to constitute their own family [110].

The cells of the stomach mucosa are able to generate an H^+ gradient of $1:10^6$ by the activity of an ATPase. This enzyme has been examined in detail in the frog and various mammals; it catalyses the electroneutral exchange of H^+ and K^+ and is highly specific for Mg^+-ATP. It is a glycoprotein with a carbohydrate fraction of 20 %, which is rich in glucose but free of sialic acid. The human, pig and rat enzymes are composed of a catalytic α subunit of 114 kDa and a β subunit of 34 kDa. The α sequences agree by up to 60 % with the corresponding subunits of Na^+,K^+-ATPase, and the β subunits agree by 29–37 % [90, 91, 152].

The stomach mucosa also contains an **anion-sensitive ATPase** which was earlier thought to be responsible for chloride transport in exchange for hydrogen carbonate. It appears, however, that in the secretion of hydrochloric acid in the stomach only the H^+ ions and not the Cl^- ions are actively transported. Anion-sensitive ATPases are found in many cells, not only in the stomach mucous membranes of mammals, reptiles and amphibians, but also in the intestinal cells of mammals, fish and insects, in mammalian erythrocytes and many mammalian organs, in the gills of fish and crustaceans, in the mantle tissue of mussels, and in the Malpighian tubules of insects [180]. These enzymes require magnesium or manganese ions and are sensitive to ouabain. They are stimulated by various quite different anions, such as hydrogen carbonate, chloride, selenite, sulphite, arsenite and borate, to a degree depending upon the species and cell; they are inhibited by thiocyanate and nitrate. The specificity for the nucleoside triphosphate is relatively low. GTP and ITP have the same activity as ATP with the enzyme from the rabbit kidney, but less than 50–60 % of the ATP activity with the stomach enzyme. Part of the anion-sensitive ATPase activity of the cell is located in the mitochondria, but there is also activity in the cell membrane and in the microsomes.

17.3.4 Ion Channels

In contrast to the ion pumps, ion channels do not allow transport against a concentration gradient, but in their activated "open" state they can achieve transport rates that come close to those of free diffusion. Channel opening can be triggered by a chemical signal (ligand-gated channels) or by a change in the membrane potential (voltage-gated channels). The channel proteins include the connexins from the gap junctions (p. 238) and the synaptophysins from the synaptic vesicles. The two proteins have no sequence similarity but show a very similar topography with four transmembrane segments. The activating signals for these proteins are unfortunately not known. Knowledge of the molecular structure of the ion channels has increased rapidly in recent years with the help of cDNA sequencing. However, it is not yet clear how many protein super-families are involved, nor is it known to what extent diversification and convergence have played a role in the evolution of ion channels [14, 66, 102].

Sodium channels have been characterized in muscles, electric organs and nerve tissue of vertebrates, and from the brain of *Drosophila melanogaster* [14]. They are composed of a glycosylated α subunit of 260 kDa (protein component 208 kDa) and in mammalian muscle a β_1 subunit of 38 kDa and in the brain a β_1 subunit of 36 kDa and a β_2 subunit of 33 kDa (Fig. 17.4a). The amino acid sequences of the α subunits from the electric eel and rat brain and from a sodium channel from *Drosophila melanogaster* have been determined from the cDNA or gene sequences, and have been shown to be significantly homologous. Of the 1829–2008 amino acids, there is 60 % agreement between the electric eel and rat brain and approximately 50 % between *Drosophila* and vertebrates [23, 114, 121, 142].

Each chain is made up of four homologous domains of 215–272 amino acids and the domains are joined by linkers of varying length; it would appear that the evolution of the sodium channels involved two gene duplications. The four homologous domains each consist of six α-helical transmembrane segments (S1–S6) and together form a channel (Fig. 17.4a). The N- and C-termini of the α chain are localized on the cytoplasmic side of the membrane. The S4 helix is unique amongst the transport segments in that it includes regularly arranged arginine or lysine residues (Fig. 17.4b). S4 is apparently important as a voltage sensor; engineered changes in the S4 sequence affect voltage sensitivity [162]. These regions are extremely well conserved in evolution; *Drosophila* and the rat have 19 out of 24 amino acids in common in the S4 regions of the fourth domain. The positively charged basic amino acids of S4 helices are each neutralized by negative charges in the lipid bilayer. According to the **"sliding helix model"**, the resting potential stabilizes the S4 helix by forces exerted on these charges, which in the case of the positive charges are directed towards the channel lumen and for the negative charges are directed away from the lumen. Membrane depolarization diminishes these stabilizing forces and facilitates changes in conformation, in which the S4 helices undergo rotational movement with new charge pairing, and the channel opens. The sodium channel and other voltage-gated channels appear to have several conductance states [99]. The expression of α mRNAs from different sources in oocytes of *Xenopus laevis* have shown that the α subunits are sufficient for the formation of functional channels in the plasma membrane, but the β subunits greatly influence the transport kinetics of the channels [14]. Each vertebrate species possesses several tissue-specific types of sodium channels

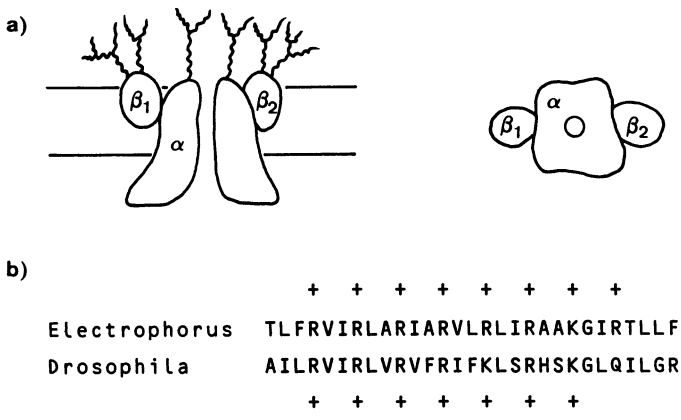


Fig. 17.4a, b. Voltage-dependent ion channels. **a** The sodium channel of the electric eel *Electrophorus electricus* consists of an α subunit, whose four transmembrane domains form the ion channel, and two different β subunits [23]. **b** The sequence of the fourth S4 region (amino acids 1415–1443) in the α subunit of the sodium channel from *E. electricus* is significantly homologous to a region (amino acids 359–387) in the product of the *shaker* locus of *Drosophila melanogaster* [121]. The basic amino acids marked with + are responsible for the voltage-dependent opening of the sodium channel

that have differing physiological and pharmacological properties and vary in their α subunits. For example, the sodium channels in nerves and muscles are blocked by even nanomolar concentrations of tetrodotoxin, whereas those in the heart require micromolar concentrations [136]. The para locus of *Drosophila melanogaster* produces a polypeptide of 1820 amino acids; this shows significant homology to the α subunit of rat, and generates several variants by alternative splicing [84].

The **calcium channels** of skeletal and heart muscle have the same molecular architecture as the sodium channels [14, 103, 181]. The α_1 subunit of the calcium channel shows significant sequence similarity to the α subunit of the sodium channel, particularly in the S4 region. In addition, there are subunits α_2 (143 kDa), β (54 kDa), γ (30 kDa) and δ (24–27 kDa). The α_2 and δ subunits are linked by a disulphide bridge and arise from the same primary gene product. All subunits except β are glycosylated [69, 72]. The calcium channels of humans and other mammals exist in several different tissue-specific forms, and show large differences in α_1 chain sequence. The α_1 chain from heart muscle (2171 amino acids) shows 66 % agreement with that from skeletal muscle (1873 amino acids); two α_1 chains from rabbit brain (2273 and 2424 amino acids) show only 40 % agreement with those from the heart and skeletal muscle, and show only 33 % agreement with one of the four forms from rat brain (rbA = 2212 amino acids) [100, 105, 155]. The calcium channel from the muscle of the carp *Cyprinus carpio* has an α_1 subunit (1852 amino acids) that is highly homologous to that of mammalian muscle; α_2 subunits are detectable but there is no indication of either β or γ subunits [52].

Potassium channels make up the most diverse family of voltage-sensitive channel proteins. A breakthrough in the definition of their structure occurred with the analysis of the *shaker* mutants of *Drosophila*. The shaking of the extremities of ether-anaesthetized specimens, giving rise to the name of the mutant, is due to a disturbance of the K^+ flux [82]. The *shaker* locus consists of >100 kb and produces various polypeptides of 500-600 amino acids by alternative splicing; these have six transmembrane segments and typical S4 sequences, and are very similar in individual domains to the α and α_1 subunits of the sodium and calcium channels. Four such subunits make up a potassium channel. The splice variants of the *shaker* locus differ in their N- and C-terminal sequences. Expression in *Xenopus* oocytes results in different combinations of these variant channels with different physiological properties [96]. Other genes encode *shaker*-related proteins *(shab, shaw, shal)* which show more than 50 % sequence agreement with the *shaker* protein; these also probably form K^+-selective channels [20, 177]. The *eag* locus of *Drosophila* codes for a K^+-channel protein with a length of 1174 amino acids and seven transmembrane segments; this has no sequence similarity to *shaker* and has a very different structure [176]. Using *shaker* as a probe, sequences with similarities to *shaker* and *shaker*-related proteins have been detected in mammalian brain. On the other hand, a gene isolated from mammalian kidney defines a second gene family of K^+-channel proteins that are not related to *shaker* [97, 177]. The diversity of K^+-channel proteins in the mammals, as in *Drosophila*, is based on the existence of multiple genes and splice variants, although many of the mammalian genes are intron-less and therefore show no alternative splicing [24, 87].

The **ligand-gated ion channels** with reaction times of less than 1 ms are much faster than the G-protein-coupled receptors. These include as cation-selective ion channels the nicotinic acetylcholine receptor, described in Chapter 8, and the neuronal glutamate receptor, and as chloride-specific ion channels mainly the glycine- and the γ-aminobutyrate (GABA) receptors [161]. Because excitation involves cation influx and inhibition involves an increase in Cl^- permeability, glycine and GABA are inhibitory transmitters. The glycine receptor from mammalian brain is a pentameric channel of five homologous subunits: three α (48 kDa) and two β (58 kDa). The α subunit carries the ligand-binding site. Multiple forms of both subunits are known, and the C-termini each have four transmembrane segments (M1–M4). The nicotinic acetylcholine receptor and the GABA receptor have a similar pentameric architecture and homologous transmembrane segments. The M2 helix of all ligand-gated receptors contains mainly polar, uncharged amino acids and makes up the hydrophilic inner lining of the channel lumen. At both ends of M2 are ion-selective filter structures, consisting of positively charged amino acids in the case of the anion-selective glycine and GABA receptors, and negatively charged amino acids in the case of the cation-selective acetylcholine receptor. Just one type of subunit expressed in *Xenopus* oocytes produces functional ion channels, and the combination of different subunits produces functionally varied receptors [14]. A cDNA isolated from

the ray *Torpedo marmorata* and expressed in *Xenopus* oocytes results in a homodimeric Cl^--specific channel. This is a protein of 805 amino acids with 12–13 transmembrane segments and shows no sequence similarity to the glycine and GABA receptors or to the HCO_3^-/Cl^- exchanger of erythrocytes (band-3 protein). Similar chloride channels probably occur in mammalian kidney [71].

17.4 Sulphatases

The sulphatases are subdivided according to their substrate specificity into aryl, steroid, glyco- and chondroitin sulphatases, although in many cases the distinction is problematic. Because of the frequent overlapping of the substrate specificity of individual enzymes, this classification must be considered provisional; the real relationships between the sulphatase types cannot be determined in the absence of sequence data. The enzymes designated as aryl sulphatases cleave various aryl and alkyl sulphate esters; 4-nitrophenylsulphate or catechol sulphate are used for enzyme assays and in most cases the physiological substrates are not known. Aryl sulphatases have been detected in bacteria, lower fungi, vertebrates and in many phyla of the invertebrates; they have especially high activities in molluscs. Enzyme preparations from the edible snail *Helix pomatia* and from other snails are commercially available for analytical purposes.

The **aryl sulphatases of the mammals** are divided into two subclasses according to their subcellular localization. Type I (sulphatase C) includes membrane-bound, microsomal enzymes with neutral or weakly alkaline pH optima; these are only weakly inhibited by sulphate and phosphate. Many C sulphatases function as steroid sulphatases [141, 148]. The sulphatases of type II (sulphatases A and B) are mainly found in lysosomes, have acid pH optima and are strongly inhibited by sulphate and phosphate. Two groups (A and B) of type-II aryl sulphatases can be distinguished [169]. The sulphatase A enzymes are anionic glycoproteins which are strongly inhibited by silver ions and exist as monomers of 80–182 kDa at pH 7.5. The enzymes of the higher mammals aggregate to form dimers at pH 4.5, but those of the marsupials, such as the opossum and the red kangaroo, do not [168, 174]. Frequently occurring sulphate esters, such as cerebroside sulphate, ascorbic acid sulphate and tyrosine sulphate, may be the natural substrates of sulphatase A. The B sulphatases are cationic glycoproteins which are not inhibited by silver ions and are always monomeric. Sulphatase B is presumably involved in the metabolism of glycosaminoglycans by cleaving N-acetylgalactosamine-4-sulphate compounds. Both A and B sulphatases exist in multiple forms. The human aryl sulphatases A and B, together with steroid sulphatase and glucosamine-6-sulphatase make up a family of homologous proteins with sequence agreements of 18–27 % on pairwise comparison [126]. A- and B-like aryl sulphatases can also be distinguished in the other vertebrate classes down to the amphibians and fish [31]. A-like sulphatases in the salmonids and other fish possibly have the special function of cleaving ascorbic acid sulphate. Ascorbic acid is essential for salmon, trout and other fish, and the much more stable sulphate ester probably plays an important role as a source of ascorbate.

Multiple aryl sulphatases have been found in gastropods but these cannot be directly compared with the vertebrate enzymes. The gastropod enzymes are also glycoproteins but, unlike the vertebrate enzymes, they contain no sialic acid [28, 140]. Like the mammalian type A, the majority of **invertebrate enzymes** are active against cerebroside sulphate. This is not true, however, of the enzyme in the seminal plasma of the sea urchin *Strongylocentrotus intermedius;* as in other sea urchins, the enzyme facilitates penetration of the sperm through the gelatinous egg membrane. 4-Nitrophenylsulphate in the medium competitively inhibits the enzyme and retards fertilization. This enzyme was not found in the starfish. The aryl sulphatase in the larvae (plutei) of the sea urchin *Hemicentrotus pulcherrimus* makes up about 0.5 % of the total protein. The native enzyme is about 670 kDa and is composed of 63-kDa subunits with 551 amino acids, whose sequence shows no similarity to any other protein [144, 182].

Chondroitin sulphatases and glycosulphatases, which cleave ester sulphates from glycan sulphates and sugar sulphates, have been detected in bacteria, mammals, arthropods, echinoderms and molluscs; they are often difficult to distinguish from one another. They have been examined in detail in the gastropods, where they show particularly high activity and apparently have digestive functions. The hepatopancreas of the Japanese snail *Charonia lampas*, which produces glycan sulphate (charonin sulphate), contains a cellulose sulphatase which is inactive

against chondroitin sulphate, and a genuine chondroitin sulphatase and two glycosulphatases which can attack, for example, glucose-6-sulphate. A sulphatase isolated from the mussel *Anomalocardia brasiliensis* can completely desulphatize chondroitin-4-sulphate and chondroitin-6-sulphate without concomitant depolymerization. This degradation pathway for chondroitin sulphate differs fundamentally from that in the vertebrates, in which the polysaccharide is first reduced to monosaccharides by the sequential action of endoglycosidase, β-glucuronidases and β-N-acetyl-galactosaminidases; only then is it desulphatized by 4- and 6-N-acetylgalactosamine sulphatases [34].

17.5 Carboanhydrases

The carboanhydrases (CA) or carbonate dehydratases catalyse the reversible reaction,

$$CO_2 + H_2O \leftrightarrows H^+ + HCO_3^-. \qquad (17.1)$$

They are present in vertebrates, invertebrates, higher plants and bacteria, and are involved in various processes, e.g. gas exchange, regulation of the acid/base balance, ion transport, the deposition and mobilization of calcium carbonate in skeletal structures, and photosynthesis. All known CAs are zinc proteins and make up a protein family. All animal CAs also have esterase activity, e.g. against 4-nitrophenylacetate, and are strongly inhibited by acetazolamide ($K_i = 10^{-6}$ to 10^{-8} mol/l) and less strongly inhibited by azide, cyanide and halides. The plant CAs are much less sensitive to acetazolamide ($K_I > 10^{-5}$ mol/l) and have no esterase activity; the substrate-binding groove in this case is apparently much flatter than in the animal enzyme [35].

At least 6, and possibly 10–12, **CA isoenzymes** are found in mammals; these are encoded by different genes, show large structural and functional differences, and to some extent are tissue specific [40, 108]. The highest CA activities in all vertebrates are found in the erythrocytes, which are therefore commonly the starting point for investigations of CA. Human erythrocytes, and those of most other mammals, have two isoenzymes: the highly active CA II, and CA I, which at physiological pH values has about 10- to 20-fold lower specific activity. Apart from in erythrocytes, CA I is found in stomach mucosa and kidney; CA II is found in many tissues. A third type, CA III, with an even lower catalytic efficiency was first discovered in 1978 in skeletal muscle, where it is restricted to slow, oxidative, red-muscle fibres; it apparently facilitates the diffusion of CO_2. CA III also occurs in the liver and lung and even in human erythrocytes [173, 178]. All three isoenzymes are monomeric proteins of 29 kDa with 259–260 amino acids and one zinc atom. The strength of zinc binding varies; for example, it is weak in the common hagfish *Myxine glutinosa* and particularly strong in the kangaroo *Macropus eugenii*. The sequence differences between the isoenzymes are reflected in the amino acid composition; CA I of mammals always contains 28–33 serine residues, whereas CA II has only 16–22 serine residues. CA I and CA II usually contain only one cysteine; sequences with no or two cysteines are rare, e.g. in the main and subsidiary CA II of equine erythrocytes. In contrast, equine and bovine CA III have five cysteines and rabbit CA III has six. The functional properties of the three isoenzymes vary somewhat with the species, but CA activity, esterase activity and sensitivity to acetazolamide declines consistently in the order II > I > III [178].

The relative proportion of the two **CA types in erythrocytes** varies greatly with the species. The ratio of I to II is 1:2 in mouse and 5:1 in humans; human CA I is the second most abundant erythrocyte protein at 4 g/l cells. Whereas the content and properties of isoenzyme II in human red blood cells is very constant, CA I activity is very variable and may even be completely absent without causing deficiency symptoms. Some mammalian species have no CA I in their erythrocytes, e.g. the ruminants such as cattle, sheep, goat and elk, the cat, the dolphin and the kangaroo *Macropus eugenii*. However, that the CA I gene is present may be concluded from the presence of CA I in the rumen of cattle and sheep. Thus, it is questionable whether CA I has any sort of important function in red blood cells. CA I in particular shows considerable polymorphism in humans and many other mammals. Allelic variants are found at 11 positions of the CA I chain of the horse; the N-terminus of CA II may be either acetylserine or acetylthreonine [65]. The CA II genes of the laboratory mouse breeding lines YBR and BALB/c differ by one amino acid-exchanging substitution and two synonymous substitutions [172]. Apart from differences in the coding sequences, multiple CAs may also arise by post-translational modification, e.g. by step-wise deamidation of asparagine or glutamine, as occurs in human CA I, or by formation of a mixed disulphide with glutathione, as occurs in horse and sheep CA II.

In addition to the well-known types CA I–III, **other CA types** are known in mammals; these have different structural and functional characteristics and are probably encoded by their own genes. Membrane-bound carboanhydrases (CA IV), which are only released by detergent treatment, are found in the lung, kidney, skeletal muscle and hypophysis. The subunits of CA IV from human lung have a mass of 35 kDa and are therefore significantly larger than the CAs I–III from the erythrocytes [40, 188]. The mitochondrial CA V of hepatocytes is presumably involved in urea synthesis, gluconeogenesis and fatty acid synthesis. CA VI is produced in the parotid gland and excreted in the saliva. It was discovered in sheep but is also found in humans, cattle and mice. Sheep CA VI shows only 33 % agreement in its sequence of 307 amino acids with the small CA II [40].

The mammalian CAs are amongst the most frequently investigated proteins; numerous CA I, II and III enzymes have been sequenced. These proteins have changed only slowly during evolution (see Table 4.12, p. 161). Thus, all known sequences of CA I from mammals agree at 76 % of positions, and those of CA II agree 71 %; CA III of the horse and cattle are 91 % identical, and CA II of humans and the chicken are 70 % identical. A partial sequence from the turtle *Macaclemys terrapin* shares a surprising 60–66 % identity with human CAs I and II. The differences between the isoenzymes are much greater: human CA I and CA II have only 60 % of their amino acids in common, and all known CAs I, II and III have only 40 % in common. Thus, the gene duplications that separated the isoenzymes occurred very early in the evolution of the vertebrates. Despite the differences in sequence, the spatial structures of human CA I and CA II agree almost completely [6, 137, 172, 178]. Each of the isoenzymes I–III contains 20–32 amino acids which are characteristic of that particular isoenzyme and are to a large extent invariable in the mammals [45].

The **CA II gene** of the mouse has a length of 16 kb and consists of seven exons. No coincidence is found between the exons and the protein domains; on the contrary, the amino acids forming the substrate groove of the active centre, for example, are encoded by four different exons (2, 3, 5 and 6) [172]. The chicken CA II gene is 17 kb long and agrees with the mouse gene in the positions of five out of six introns; however, the fourth intron has shifted during evolution, lying at codon 143 in the mouse and between codons 147 and 148 in the chicken [186]. The genes for CA I and CA II in the mouse lie tightly linked on chromosome 3, and differ by 39 % amino acid-exchanging substitutions. The CA I genes of the mouse and rabbit, the evolutionary lines of which separated about 80–90 million years ago, differ by 11 % such substitutions. According to these data, the duplication creating isoenzymes I and II occurred about 300–320 million years ago [45].

CA is found in the erythrocytes and various cells of **animals other than the mammals**. However, the activities vary greatly, even between closely related species. For example, systematic investigations of amphibians have shown relative activities of 82 *(Xenopus laevis)* and 4964 *(Bufo marinus)* in the blood, and between 19 *(Rana catesbeiana)* and 2035 *(Rana pipiens)* in the kidney. Only one CA was detected initially in the erythrocytes of lower vertebrates, and this corresponded in many ways to mammalian CA II. It was concluded that CA I arose as an evolutionary novelty at the same time as the mammals about 100–120 million years ago. This concept was abandoned with the discovery of a typical CA I in the gut tissue of the chicken, in addition to CA II in the erythrocytes; furthermore, the turtle *Malaclemys terrapin* has two CAs in its erythrocytes, one of which is present at a 10-fold lower concentration but has a 36-fold higher specific activity. However, these two isoenzymes cannot be compared directly with mammalian Cas I and II.

Almost all erythrocyte CAs of the vertebrates have a molecular mass of 29–30 kDa, corresponding to 260 amino acids. The only exceptions are those of the sharks, e.g. *Squalus acanthias* with 34 kDa, *Carcharias leucas* with 35.9 kDa and *Galeocerda cuvieri* with 38.6 kDa. These shark enzymes contain no fewer than 18–25 cysteine residues per subunit, and most of them are oxidized to form disulphide bridges. The erythrocyte CAs of most non-mammals contain 4–7 cysteine residues; for example, there are 4 in the frog *Rana catesbeiana*, 6 in the trout *Salmo gairdneri*, and 7 in the chicken. In contrast to the situation in the elasmobranchs, most of these cysteine residues are not oxidized to disulphides; thus, the enzymes have their maximal activities in the presence of protective SH compounds and are inhibited by *p*-chloromercuribenzoate (PCMB). This is not true, however, for all the CAs from lower vertebrates; for example, the teleosts *Oncorhynchus gorbusha* and *Archosargus probatocephalus* are more like the mammals and have only one cysteine per CA chain [76].

CAs have been reported in many **invertebrates**, such as the poriferans, cnidarians, annelids,

molluscs and arthropods, but have not been characterized in any detail. For a long time it was assumed that the red blood cells of invertebrate animals, in contrast to those of vertebrates, have no CA. However, CA activity has since been recorded in the haemoglobin-containing cells of the polychaete *Glycera americana* and in the haemerythrin-containing cells of the sipunculid *Phascolopsis gouldi* [56]. The red blood cells found in some mussels are free of CA, although the haemolymph of the oyster *Crassostrea virginica* has a soluble CA. In its native form, this is a high molecular weight aggregate of 26.6S, and even after SDS treatment remains as aggregates of 500 kDa with 13 zinc atoms. The estimated equivalent mass of 38 kDa lies in the same range as that of the elasmobranch CA. The haemolymph of the spider *Eurypelma californica* contains a soluble CA of about 40 kDa. The enzyme is associated with a 16S lipoprotein that occurs in the haemolymph as a non-respiratory protein and constitutes about 20 % of the total protein [159].

CA was detected in the **insects**, in the non-biting midge *Chironomus*, as early as 1932, shortly after the discovery of the enzyme in mammalian blood. However, it was not until 1979 that the first insect CA was characterized, namely that from the caterpillars of *Manduca sexta*. CA is mainly found in the fat bodies, midgut and integument of *Manduca*. It is detectable in the haemolymph, salivary glands, midgut and brain of pupation-mature larvae of the fly *Musca autumnalis*, but not in the Malpighian tubules, fat body, trachea or the remainder of the animal [32]. A fundamental difference exists between the *Manduca* enzyme and mammalian CA: whereas the latter is significantly inhibited by KCl and choline chloride, the insect enzyme is strongly activated by KCl and only weakly inhibited by choline chloride. This effect is apparently related not to differences in the active centres but to allosteric effects of the potassium ion. The CA in the gills of crustaceans and molluscs, like the enzymes of fish gills, are considered to have a role in maintaining the ion balance. Because of its possible involvement in calcification of the shell, the mantle tissue CA of the molluscs has received particular attention. The CA of the predatory snails *Purpura lapillus* (Nuricidae) and *Pollinices lewisii* has the interesting special function of boring through mussel shells. The CA activity in the boring organ of *Pollinices* is threefold higher than in the mantle and sixfold higher than in the gills; the enzyme has a molecular weight of over 100 kDa. The boring process can be inhibited with acetazolamide. The boring sponges (Clionidae) probably exploit CA for the same purpose.

References

1. Anstee J. H., Baldrick P and Bowler K.: Studies on ouabain-binding to Na,K-ATPase from malpighian tubulus of the locust, Locusta migratoria L. Biochim. biophys. Acta 860: 15–24 (1986)
2. Arpagaus M., Fournier D. and Toutant J. P.: Analysis of acetylcholinesterase molecular forms during the development of Drosophila melanogaster. Evidence for the existence of an amphiphilic monomer. Insect Biochem. 18: 539–549 (1988)
3. Arzamazova N. M. et al.: Detailed structural analysis of exposed domains of membrane-bound Na,K-ATPase: a model of transmembrane arrangement. FEBS Letters 217: 269–274 (1987)
4. Baba T. et al.: Structure of human milk bile salt activated lipase. Biochemistry 30: 500–510 (1991)
5. Banno Y. and Nozawa Y.: Purification and characterization of extracellular acid phosphatase of Tetrahymena pyriformis. Biochim. biophys. Acta 799: 20–28 (1984)
6. Barlow J. H. et al.: Human carbonic anhydrase I cDNA. Nucleic Acids Res. 15: 2386 (1987)
7. Baxterlowe L. A. et al.: Molecular cloning of the Na,K-ATPase alpha-subunit in developing brine shrimp and sequence comparison with higher organisms. FEBS Letters 257: 181–187 (1989)
8. Beintema J. J.: Evolutionary role of posttranslational modifications of proteins, as illustrated by the glycosylation characteristics of the digestive enzyme pancreatic ribonuclease. J. mol. Evol. 24: 118–120 (1986)
9. Beintema J. J., Fitch W. M. and Carsana A.: Molecular evolution of pancreatic type ribonucleases. Mol. Biol. Evol. 3: 262–275 (1986)
10. Beintema J. J. et al.: Amino acid sequence of the non-secretory ribonuclease of human urine. Biochemistry 27: 4530–38 (1988)
11. Beintema J. J.: The primary structure of langur (Presbytis entellus) pancreatic ribonuclease: Adaptive features in digestive enzymes in mammals. Mol. Biol. Evol. 7: 470–477 (1990)
12. Belzunces L. P., Toutant J. P. and Bounias M.: Acetylcholinesterase from Apis mellifera head. Evidence for amphiphilic and hydrophilic forms characterized by triton X-114 phase separation. Biochem. J. 255: 463–470 (1988)
13. Bernbaeck S., Blaeckberg L. and Hernell O.: Fatty acids generated by gastric lipase promote human milk triacylglycerol digestion by pancreatic colipase-dependent lipase. Biochim. biophys. Acta 1001: 286–293 (1989)
14. Betz H.: Homology and analogy in transmembrane channel design: Lessons from synaptic membrane proteins. Biochemistry 29: 3591–99 (1990)
15. Bhattacharyya K. K., Bergstrom E. E. and Hokin L. E.: Molecular cloning of the beta-subunit of the Na,K-ATPase in the brine shrimp, Artemia. FEBS Letters 269: 233–238 (1990)
16. Borgström B. and Brockman H. L. (eds.): Lipases. American Elsevier, New York 1984

17. Brown T. A. et al.: Molecular cloning and sequence analysis of the Na,K-ATPase β subunit from dog kidney. Biochim. biophys. Acta 912: 244–253 (1987)
18. Burk S. E. et al.: cDNA cloning, functional expression, and mRNA tissue distribution of a third organellar Ca2+ pump. J. Biol. Chem. 264: 18561–68 (1989)
19. Butikofer P. et al.: Molecular species analysis of the glycosyl-phosphatidylinositol anchor of Torpedo marmorata acetylcholinesterase. J. Biol. Chem. 265: 18983–87 (1990)
20. Butler A., Wei A. and Salkoff L.: Shal, Shab and Shaw: three genes encoding potassium channels in Drosophila. Nucleic Acids Res. 18: 2173–74 (1990)
21. Cabot M. C. et al.: Vertebrate class distribution of 1-alkyl-2-acetyl-sn-glycero-3-phosphocholine acetylhydrolase in serum. Comp. Biochem. Physiol. Pt. B 78: 37–40 (1984)
22. Camici G. et al.: The primary structure of turkey muscle acylphosphatase. Eur. J. Biochem. 137: 269–277 (1983)
23. Catterall W. A.: Genetic analysis of ion channels in vertebrates. Annual Rev. Physiol. 50: 395–406 (1988)
24. Chandy K. G. et al.: A family of three mouse potassium channel genes with intronless coding regions. Science 247: 973–975 (1990)
25. Charbonneau H. et al.: Identification of a conserved domain among cyclic nucleotide phosphodiesterases from diverse species. Proc. Nat. Acad. Sci. USA 83: 9308–12 (1986)
26. Chatonnet A. and Lockridge O.: Comparison of butyrylcholinesterase and acetylcholinesterase (Review). Biochem. J. 260: 625–634 (1989)
27. Chuang N. N.: A heat-stable alkaline phosphatase from Penaeus japonica Bate (Crustacea: Decapoda): a phosphatidylinositol-glycan anchored membrane protein. Comp. Biochem. Physiol. Pt. B 95: 165–169 (1990)
28. Clark A. G. and Jowett D. A.: Purification and properties of an arylsulphatase from the digestive gland of Haliotis iris. Comp. Biochem. Physiol. Pt. B 89: 409–417 (1988)
29. Collet C. et al.: Molecular analysis of duplicated esterase genes in Drosophila melanogaster. Mol. Biol. Evol. 7: 9–28 (1990)
30. Collins J. H. and Leszyk J.: The „γ-subunit" of Na,K-ATPase: a small amphiphilic protein with an unique amino acid sequence. Biochemistry 26: 8665–68 (1987), Corr. l.c. 27: 8508 (1987)
31. Daniel W. L., Leonard L. and Carmody T.: Comparative biochemistry of fish and amphibian arylsulfatases. Comp. Biochem. Physiol. Pt. B 87: 385–390 (1987)
32. Darlington M. V., Meyer H. J. and Graf G.: Carbonic anhydrases in the face fly, Musca autumnalis (deGeer) (Diptera: Muscidae). Insect Biochem. 15: 411–418 (1985)
33. Davis R. L. et al.: Cloning and characterization of mammalian homologs of the Drosophila dunce+ gene. Proc. Nat. Acad. Sci. USA 86: 3604–08 (1989)
34. Desousa J. F., Nader H. B. and Dietrich C. P.: Sequential degradation of chondroitin sulfate in molluscs: Desulfation of chondroitin sulfate without prior depolymerization by a novel sulfatase from Anomalocardia brasiliana. J. Biol. Chem. 265: 20150–55 (1990)
35. Dodgson S. J. et al. (eds.): Carbonic anhydrases. Cellular physiology and molecular genetics. Plenum, New York 1991
36. Downer R. G. H.: Lipid metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 77–113. Pergamon Press, Oxford 1985
37. Dufton M. J. and Hider R. C.: Classification of phospholipases A2 according to sequence. Evolutionary and pharmacological implications. Eur. J. Biochem. 137: 545–551 (1983)
38. Durham L. A. and McLean Grogan W.: Characterization of multiple forms of cholesteryl ester hydrolase in rat testis. J. biol. Chem. 259: 7433–38 (1984)
39. El-Maghrabi M. R. et al.: cDNA sequence of rat liver fructose-1,6-bisphosphatase and evidence for down-regulation of its mRNA by insulin. Proc. Nat. Acad. Sci. USA 85: 8430–34 (1988)
40. Fernley R. T. et al.: Tissue and species distribution of the secreted carbonic anhydrase isoenzyme. Biochem. J. 259: 91–96 (1989)
41. Fini C. et al.: 5'-Nucleotidase from bull seminal plasma, chicken gizzard and snake venom is a zinc metalloprotein. Biochim. biophys. Acta 1038: 18–22 (1990)
42. Florin-Christensen J. et al.: Phospholipase A1 and triacylglycerol lipase: two novel enzymes from Tetrahymena extracellular medium. Comp. Biochem. Physiol. Pt. B 85: 149–155 (1986)
43. Fontaine R. N., Carter C. P. and Hui D. Y.: Structure of the rat pancreatic cholesterol esterase gene. Biochemistry 30: 7008–14 (1991)
44. Fournier D. et al.: Drosophila melanogaster acetylcholinesterase gene: Structure, evolution and mutations. J. mol. Biol. 210: 15–22 (1989)
45. Fraser P. J. and Curtis P. J.: Molecular evolution of the carbonic anhydrase genes: calculation of divergence time for mouse carbonic anhydrase I and II. J. mol. Evol. 23: 294–299 (1986)
46. Freed L. M. et al.: Bile salt-stimulated lipase in non-primate milk: longitudinal variation and lipase characteristics in cat and dog milk. Biochim. biophys. Acta 878: 209–215 (1986)
47. Friedman S. and Hsueh T. F.: Insect trehalose-6-phosphatase: the unactivated type, as illustrated in Periplaneta americana, and a survey of the ordinal distribution of the two presently known types. Comp. Biochem. Physiol. Pt. B 64: 339–344 (1979)
48. Gargouri Y., Moreau H. and Verger R.: Gastric lipases: biochemical and physiological studies (Review). Biochim. biophys. Acta 1006: 255–271 (1989)
49. Ghosh J. et al.: A high affinity Ca-ATPase on the surface membrane of Leishmania donovani promastigote. J. Biol. Chem. 265: 11345–51 (1990)
50. Giroir L. E. and Deutsch W. A.: Drosophila deoxyuridine triphosphatase. Purification and characterization. J. biol. Chem. 262: 130–134 (1987)
51. Gnagey A. L., Forte M. and Rosenberry T. L.: Isolation and characterization of acetylcholinesterase from Drosophila. J. biol. Chem. 262: 13290–98 (1987)
52. Grabner M. et al.: Calcium channels from Cyprinus carpio skeletal muscle. Proc. Nat. Acad. Sci. USA 88: 727–731 (1991)
53. Greeb J. and Shull G. E.: Molecular cloning of a third isoform of the calmodulin-sensitive plasma membrane Ca-transporting ATPase that is expressed predominantly in brain and skeletal muscle. J. Biol. Chem. 264: 18569–76 (1989)
54. Hamilton T. A., Gornicki S. Z. and Sussman, H. H.: Alkaline phosphatase from human milk. Comparison with isoenzymes from placenta and liver. Biochem J. 177: 197–201 (1979)

55. Hart N. H. and Cook M.: Comparative analysis of tissue esterases of the zebra danio (Brachydanio rerio) and the pearl danio (B. albolineatus) by disc gel electrophoresis. Comp. Biochem. Physiol. Pt. B 54: 357–364 (1976)
56. Henry R.: Invertebrate red blood cell carbonic anhydrase. J. exp. Zool. 242: 113–116 (1987)
57. Henthorn P. S. et al.: Sequence and characterization of the human intestinal alkaline phosphatase gene. J. Biol. Chem. 263: 12011–19 (1988)
58. Hereld D., Hart G. W. and Englund P. T.: cDNA encoding the glycosyl-phosphatidylinositol-specific phospholipase C of Trypanosoma brucei. Proc. Nat. Acad. Sci. USA 85: 8914–18 (1988)
59. van Heusden M. C. et al.: Partial purification of locust flight muscle lipoprotein lipase (LpL): apparent differences from mammalian LpL. Comp. Biochem. Physiol. Pt. B 88: 523–527 (1987)
60. Himeno M. et al.: Isolation and sequencing of a cDNA clone encoding acid phosphatase in rat liver lysosomes. Biochem. biophys. Res. Commun. 162: 1044–53 (1989)
61. Hofsteenge H., Matthies R. and Stone S. R.: Primary structure of a ribonuclease from porcine liver, a new member of the ribonuclease superfamily. Biochemistry 28: 9806–13 (1989)
62. Horii S. and Eguchi M.: Comparison of the property of a novel isozyme E (formerly null mutant, 0) with other isozymes of hemolymph acid phosphatase of the silkworm. Comp. Biochem. Physiol. Pt. B 93: 443–449 (1989)
63. Houk E. J. and Hardy J. L.: Alkaline phosphatases of the mosquito, Culex tarsalis Coquillet. Comp. Biochem. Physiol. Pt. B 78: 303–310 (1984)
64. Ishizaki J. et al.: cDNA cloning and sequence determination of rat membrane-associated phospholipase A2. Biochem. biophys. Res. Commun. 162: 1030–36 (1989)
65. Jabusch J. R. and Deutsch H. F.: Sequence of a high-activity equine erythrocyte carbonic anhydrase: N-terminal polymorphism (acetyl-ser/acetyl-thr) and homologies to similar mammalian isozymes. Biochem. Genetics 22: 357–367 (1984)
66. Jan L. Y and Jan Y. N.: Voltage-sensitive ion channels (Review). Cell 56: 13–25 (1989)
67. Janska H. et al.: Catfish liver acid phosphatases – differently glycosylated enzyme molecules with altered kinetic properties. Comp. Biochem. Physiol. Pt. B 85: 753–758 (1986)
68. Janska H., Kubicz A. and Szalewicz A.: The lower molecular weight acid phosphatase from the frog liver: Isolation of homogenous AcPase III and IV representing glycoforms with different bioactivity. Comp. Biochem. Physiol. Pt. B 92: 341–346 (1989)
69. Jay S. D. et al.: Primary structure of the gamma-subunit of the DHP-sensitive calcium channel from skeletal muscle. Science 248: 490–492 (1990)
70. Jbilo O. and Chatonnet A.: Complete sequence of rabbit butyrylcholinesterase. Nucleic Acids Res. 18: 3990 (1990)
71. Jentsch T. J., Steinmeyer K. and Schwarz G.: Primary structure of Torpedo marmorata chloride channel isolated by expression cloning in Xenopus oocytes. Nature 348: 510–514 (1990)
72. de Jongh K. S., Warner C. and Catterall W. S.: Subunits of purified calcium channels: alpha-2 and delta are encoded by the same gene. J. Biol. Chem. 265: 14738–41 (1990)
73. Kanda A. et al.: The primary structure of a membrane-associated phospholipase A2 from human spleen. Biochem. biophys. Res. Commun. 163: 42–48 (1989)
74. Kano I. et al.: Structure of the alpha-1 subunit of horse Na,K-ATPase gene. FEBS Letters 250: 91–98 (1989)
75. Kawakami K. et al.: Primary structure of the α-subunit of Torpedo californica Na,K-ATPase deduced from cDNA sequence. Nature 316: 733–736 (1985)
76. Kim J. S., Gay C. V. and Schraer R.: Purification and properties of carbonic anhydrase from salmon erythrocytes. Comp. Biochem. Physiol. Pt. B 76: 523–527 (1983)
77. Kirchgessner T. G. et al.: Organization of the human lipoprotein lipase gene and evolution of the lipase gene family. Proc. Nat. Acad. Sci. USA 86: 9647–51 (1989)
78. Knoll B. J., Rothblum K. M. and Longley M.: Nucleotide sequence of the human placental alkaline phosphatase gene. Evolution of the 5'-flanking region by deletion/substitution. J. Biol. Chem. 263: 12020–27 (1988)
79. Komaromy M. C. and Schotz M. C.: Cloning rat hepatic lipase cDNA: evidence for a lipase gene family. Proc. Nat. Acad. Sci. USA 84: 1526–30 (1987)
80. Kuchler K. et al.: Analysis of the cDNA for phospholipase A2 from honeybee venom glands. The deduced amino acid sequence reveals homology to corresponding vertebrate enzymes. Eur. J. Biochem. 184: 249–254 (1989)
81. Lebovitz R. M., Takeyasu K. and Fambrough D. M.: Molecular characterization and expression of the (Na,K)-ATPase alpha-subunit in Drosophila melanogaster. Embo J. 8: 193–202 (1989)
82. Lichtinghagen R. et al.: Molecular basis of altered excitability in shaker mutants of Drosophila melanogaster. Embo J. 9: 4399–4407 (1990)
83. Liu F. and Fromm H. J.: Purification and characterization of fructose-1,6-biphosphatase from bovine brain. Arch. Biochem. Biophys. 260: 609–615 (1988)
84. Loughney K., Kreber R. and Ganetzky B.: Molecular analysis of the para locus, a sodium channel gene in Drosophila. Cell 58: 1143–54 (1989)
85. Lowe M. E., Rosenblum J. L. and Strauss A. W.: Cloning and characterization of human pancreatic lipase cDNA. J. Biol. Chem. 264: 20042–48 (1989)
86. Lowe M. E. et al.: Cloning and characterization of the human colipase cDNA. Biochemistry 29: 823–828 (1990)
87. Luneau C. J. et al.: Alternative splicing contributes to K+ channel diversity in the mammalian central nervous system. Proc. Nat. Acad. Sci. USA 88: 3932–36 (1991)
88. Mackness M. I. et al.: Distinction between „A"-esterases and arylesterases. Implications for esterase classification. Biochem. J. 245: 293–296 (1987)
89. MacLennan D. H. et al.: Amino-acid sequence of a Ca,Mg-dependent ATPase from rabbit muscle sarcoplasmic reticulum, deduced from its complementary DNA sequence. Nature 316: 696–700 (1985)
90. Maeda M., Ishizaki J. and Futai M.: cDNA cloning and sequence determination of pig gastric (H+,K+)-ATPase. Biochem. biophys. Res. Commun. 157: 203–209 (1988)
91. Maeda M. et al.: Human gastric (H+,K+)-ATPase gene. J. Biol. Chem. 265: 9027–32 (1990)
92. de Malkenson N. C., Wood E. J. and Zerba E. N.: Isolation and characterization of an esterase of Triatoma infestans with a critical role in the degradation of organophosphorus esters. Insect Biochem. 14: 481–486 (1984)

93. Maouri A. and Georgatsos J. G.: Specificity and other properties of three ribonucleases of Tetrahymena pyriformis. Eur. J. Biochem. 168: 523–528 (1987)
94. Maraganore J. M. et al.: A new class of phospholipase A2 with lysine in place of aspartate 49. Functional consequences for calcium and substrate binding. J. biol. Chem. 259: 13839–43 (1984)
95. Massoulié J.: The polymorphism of cholinesterase and its physiological significance. Trends biochem. Sci. 5: 160–164 (1980)
96. McCormack K. et al.: Shaker K+ channel subunits form heteromultimeric channels with novel functional properties. Biochem. biophys. Res. Commun. 171: 1361–71 (1990)
97. McCormack T., Vega-Saez de Miera E. C. and Rudy B.: Molecular cloning of a member of a third class of shaker-family K+ channel genes in mammals. Proc. Nat. Acad. Sci. USA 87: 5227–31 (1990)
98. McGhee J. D.: Purification and characterization of a carboxylesterase from the intestine of the nematode Caenorhabditis elegans. Biochemistry 26: 4101–07 (1987)
99. Meves H. and Nagy K.: Multiple conductance states of the sodium channel and of other ion channels. Biochim. biophys. Acta 988: 99–105 (1989)
100. Mikami A. et al.: Primary structure and functional expression of the cardiac dihydropyridine-sensitive calcium channel. Nature 340: 230–233 (1989)
101. Mizunuma H. and Tashima Y.: Evidence for the intestinal type of fructose-1,6-biphosphatase in mouse, rat, and golden hamster. Arch. Biochem. Biophys. 217: 512–516 (1982)
102. Montal M.: Molecular anatomy and molecular design of channel proteins. Faseb J. 4: 2623–35 (1990)
103. Morad M. et al. (eds.): Calcium channel. Structure, function and implications. Springer, Berlin 1988
104. Moreau H. et al.: Screening of preduodenal lipases in several mammals. Biochim. biophys. Acta 959: 247–252 (1988)
105. Mori Y. et al.: Primary structure and functional expression from complement DNA of a brain calcium channel. Nature 350: 398–402 (1991)
106. Morishima I., Inoue Y. and Ueno T.: Purification and properties of calmodulin-dependent cyclic nucleotide phosphodiesterase in silkworm fat body. Insect Biochem. 15: 325–331 (1985)
107. Morohashi M. and Kawamura M.: Solubilization and purification of Artemia salina (Na,K)-activated ATPase and N-terminal amino acid sequence of its larger subunit. J. biol. Chem. 259: 14928–34 (1984)
108. Murakami H. and Sly W. S.: Purification and characterization of human salivary carbonic anhydrase. J. biol. Chem. 262: 1382–88 (1987)
109. Narise S.: Purification and properties of acid phosphatase from Drosophila virilis. Insect Biochem. 14: 473–480 (1984)
110. Nelson N. and Taiz L.: The evolution of H+-ATPases. Trends biochem. Sci. 14: 113–116 (1989)
111. Nemoz G. and Pringent A. F.: Cyclic nucleotide phosphodiesterase in the vertebrates. Biochimie 66: 139–150 (1984)
112. Nieder M. and Law J. H.: A phospholipase A1 from the hemolymph of the tobacco hornworm, Manduca sexta. J. biol. Chem. 258: 4304–10 (1983)
113. Nilsson J. et al.: cDNA cloning of human-milk bile-salt-stimulated lipase and evidence for its identity to pancreatic carboxylic ester hydrolase. Eur. J. Biochem. 192: 543–550 (1990)
114. Noda M. et al.: Primary structure of Electrophorus electricus sodium channel deduced from cDNA sequence. Nature 312: 121–127 (1984)
115. Noguchi S. et al.: Primary structure of the β-subunit of Torpedo californica Na,K-ATPase deduced from the cDNA sequence. FEBS Letters 196: 315–320 (1986)
116. Okada N., Azuma M. and Eguchi M.: Alkaline phosphatase isozymes in the midgut of silkworm: purification of high pH-stable microvillus and labile cytosolic enzymes. J. comp. Physiol. B 159: 123–130 (1989)
117. Orlowski J. and Lingrel J. B.: Tissue-specific and developmental regulation of rat Na,K-ATPase catalytic α isoform and β subunit mRNAs. J. Biol. Chem. 263: 10436–42 (1988)
118. Palmero I. and Sastre L.: Complementary DNA cloning of a protein highly homologous to mammalian sarcoplasmic reticulum Ca-ATPase from the crustacean Artemia. J. mol. Biol. 210: 737–748 (1989)
119. Panara F. and Pascolini R.: Acid phosphatases from liver of Cyprinus carpio. Comp. Biochem. Physiol. Pt. B 92: 751–754 (1989)
120. Panara F., Angiolillo A. and Pascolini R.: Acid phosphatases from liver of Rana esculenta. Subcellular localization and partial characterization of multiple forms. Comp. Biochem. Physiol. Pt. B 93: 877–882 (1989)
121. Papazian D. M. et al.: Ion channels in Drosophila. Annual Rev. Physiol. 50: 379–394 (1988)
122. Pen J., van Beeumen J. and Beintema J. J.: Structural comparison of two esterases from Drosophila mojavensis isolated by immunoaffinity chromatography. Biochem. J. 238: 691–699 (1986)
123. Pen J. and Beintema J. J.: Nomenclature of esterases. Biochem. J. 240: 933 (1986)
124. Persson B. et al.: Structural feature of lipoprotein lipases. Lipase family relationships, binding interactions, non-equivalence of lipase cofactors, vitellogenin similarities and functional subdivision of lipoprotein lipase. Eur. J. Biochem. 179: 39–45 (1989)
125. Peters C. et al.: High degree of homology between primary structure of human lysosomal acid phosphatase and human prostatic acid phosphatase. Biol. Chem. Hoppe-Seyler 370: 177–181 (1989)
126. Peters C. et al.: Phylogenetic conservation of arylsulfatases. cDNA cloning and expression of human arylsulfatases B. J. Biol. Chem. 265: 3374–81 (1990)
127. Pezzementi L. et al.: Molecular forms of acetylcholinesterase from the skeletal muscle of the ammocoete of the lamprey Petromyzon marinus. Comp. Biochem. Physiol. Pt. B 92: 385–387 (1989)
128. Prescott M., Milne A. D. and McLennan A. G.: Characterization of the bis(5'-nucleosidyl)tetraphosphate pyrophosphohydrolase from encysted embryos of the brine shrimp Artemia. Biochem. J. 259: 831–838 (1989)
129. Principato G. B. et al.: Partial purification and characterization of an alkaline phosphatase in Helix nemoralis and in Octopus vulgaris. Comp. Biochem. Physiol. Pt. B 72: 325–328 (1982)
130. Principato G. B. et al.: Purification and characterization of four different alkaline phosphatases from Spirographis spallanzanii. Comp. Biochem. Physiol. Pt. B 78: 485–491 (1984)
131. Principato G. B. et al.: Characterization of the soluble alkaline phosphatase from hepatopancreas of Squilla mantis L. Comp. Biochem. Physiol. Pt. B 80: 801–804 (1985)

132. Rega A. F. and Garrahan T. J.: The Ca-pump of plasma membranes. CRC Press, Boca Raton USA 1986
133. Rhee S. G. et al.: Studies of inositol phospholipid-specific phospholipase C. Science 244: 546–550 (1989)
134. Rittenhouse J., Nevo E. and Marcus F.: Distribution of a COOH-terminal amino acid extension of liver fructose-1,6-bisphosphatase among rodent species. Comp. Biochem. Physiol. Pt. B 82: 507–509 (1985)
135. Robbi M. and Beaufay H.: Purification and characterization of various esterases from rat liver. Eur. J. Biochem. 137: 293–301 (1983)
136. Rogart R. B. et al.: Molecular cloning of a putative tetrodotoxin-resistant rat heart Na+ channel isoform. Proc. Nat. Acad. Sci. USA 86: 8170–74 (1989)
137. Rogers J. H.: Sequence of carbonic anhydrase II cDNA from chick retina. Eur. J. Biochem. 162: 119–122 (1987)
138. Roiko K., Jaenne O. A. and Vihko P.: Primary structure of rat secretory acid phosphatase and comparison to other acid phosphatases. Gene 89: 223–229 (1990)
139. Ronai A. et al.: Immunochemical interrelationships between carboxylesterase isozymes (EC 3.1.1.1) of the house mouse, Mus musculus. Comp. Biochem. Physiol. Pt. B 82: 347–355 (1985)
140. Roy A. B. and Williams E. A.: The sulphatase of Helix pomatia. Purification and kinetic properties. Comp. Biochem. Physiol. Pt. B 93: 229–237 (1989)
141. Ruoff B. M. and Daniel W. L.: Comparative biochemistry of mammalian arylsulfatase-C and steroid sulfatase. Comp. Biochem. Physiol. Pt. B 98: 313–322 (1991)
142. Salkoff L. et al.: Genomic organization and deduced amino acid sequence of a putative sodium channel gene in Drosophila. Science 237: 744–749 (1987)
143. Sarkis J. J. F., Guimaraes J. A. and Ribeiro J. M. C.: Salivary apyrase of Rhodnius prolixus. Kinetics and purification. Biochem. J. 233: 885–891 (1986)
144. Sasaki H. et al.: cDNA cloning, nucleotide sequence and expression of the gene for arylsulfatase in the sea urchin (Hemicentrotus pulcherrimus) embryo. Eur. J. Biochem. 177: 9–13 (1988)
145. Schueller C. et al.: Evolution of nucleic acids coding for ribonucleases. The mRNA sequence of mouse pancreatic ribonuclease. Mol. Biol. Evol. 7: 29–44 (1990)
146. Schumacher M. et al.: Primary structure of Torpedo californica acetylcholinesterase deduced from its cDNA sequence. Nature 319: 407–409 (1986)
147. Schumacher M. et al.: Multiple messenger RNA species give rise to the structural diversity in acetylcholinesterase. J. Biol. Chem. 263: 18979–87 (1988)
148. Shankaran R. et al.: Characterization of arylsulfatase-C isozymes from human liver and placenta. Biochim. biophys. Acta 1078: 251–257 (1991)
149. Shortridge R. D. et al.: A Drosophila phospholipase-C gene that is expressed in the central nervous system. J. Biol. Chem. 266: 12474–80 (1991)
150. Shull G. E., Lane L. K. and Lingrel J. B.: Amino-acid sequence of the β-subunit of the Na,K-ATPase deduced from a cDNA. Nature 321: 429–431 (1986)
151. Shull G. E., Greeb J. and Lingrel J. B.: Molecular cloning of three distinct forms of the Na,K-ATPase from rat brain. Biochemistry 25: 8125–32 (1986)
152. Shull G. E.: cDNA cloning of the β-subunit of rat gastric H,K-ATPase. J. Biol. Chem. 265: 12123–26 (1990)
153. Sikorav J. L. et al.: Complex alternative splicing of acetylcholinesterase transcripts in Torpedo electric organ: Primary structure of the precursor of the glycolipid-anchored form. Embo J. 7: 2983–93 (1988)
154. Simon L. M., Földenyi R. and Boross L.: Studies on the properties of glucose-6-phosphatase from carp liver microsomes (Cyprinus carpio L.). Comp. Biochem. Physiol. Pt. B 78: 227–229 (1984)
155. Starr T. V. B., Prystay W. and Snutch T. P.: Primary structure of a calcium channel that is highly expressed in the rat cerebellum. Proc. Nat. Acad. Sci. USA 88: 5621–25 (1991)
156. Stein W. D.: Channels, carriers and pumps. An introduction to membrane transport. Acad. Press, New York 1990
157. Sternby B., Engström Å. and Hellman U.: Purification and characterization of pancreatic colipase from the dogfish (Squalus acanthias). Biochim. biophys. Acta 789: 159–163 (1984)
158. Stieger S., Gentinetta R. and Brodbeck U.: Cholinesterases from flounder muscle. Purification and characterization of glycosyl-phosphatidylinositol-anchored and collagen-tailed forms differing in substrate specificity. Eur. J. Biochem. 181: 633–642 (1989)
159. Stratakis E. and Linzen B.: Carbonate dehydratase (carbonic anhydrase) in a spider: association with the hemolymph lipoprotein. Hoppe Seyler's Z. physiol. Chem. 365: 1187–97 (1984)
160. Strehler E. E. et al.: mRNAs for plasma membrane calcium pump isoforms differing in their regulatory domain are generated by alternative splicing that involves two internal donor sites in a single exon. Proc. Nat. Acad. Sci. USA 86: 6908–12 (1989)
161. Stroud R. M., McCarthy M. P. and Shuster M.: Nicotinic acetylcholine receptor superfamily of ligand-gated ion channels. Biochemistry 29: 11009–23 (1990)
162. Stuehmer W. et al.: Structural parts involved in activation and inactivation of the sodium channel. Nature 339: 597–603 (1989)
163. Suzuki H. et al.: Complete amino acid sequence of bovine seminal ribonuclease, a dimeric protein from seminal plasma. Biol. Chem. Hoppe-Seyler 368: 1305–12 (1987)
164. Swinnen J. V., Joseph D. R. and Conti M.: Molecular cloning of rat homologues of the Drosophila melanogaster dunce gene cAMP phosphodiesterase: Evidence for a family of genes. Proc. Nat. Acad. Sci. USA 86: 5325–29 (1989)
165. Tarrab-Hazdai R. et al.: Acetylcholinesterase of Schistosoma mansoni. Molecular forms of the solubilized enzyme. Biochim. biophys. Acta 790: 61–69 (1984)
166. Taylor P.: The cholinesterases (Minireview). J. Biol. Chem. 266: 4025–28 (1991)
167. Terao M. et al.: Isolation and characterization of the mouse liver/bone/kidney-type alkaline phosphatase gene. Biochem. J. 268: 641–648 (1990)
168. Thompson B. D., Daniel W. L. and Glaser J. H.: Comparative studies of rodent anionic arylsulfatases. Comp. Biochem. Physiol. Pt. B 82: 55–61 (1985)
169. Thompson D. B. and Daniel W. L.: Comparative biochemistry of mammalian arylsulfatases A and B. Comp. Biochem. Physiol. Pt. B 90: 823–831 (1988)
170. Toyoshima S. et al.: Purification and partial amino acid sequences of phosphoinositide-specific phospholipase-C of Drosophila eye. J. Biol. Chem. 265: 14942–48 (1990)

171. Vallejo C. G. et al.: Dinucleosidtetraphosphatase in rat liver and Artemia salina. Biochim. biophys. Acta 438: 304–309 (1976)
172. Venta P. J. et al.: Structure and exon to protein domain relationships of the mouse carbonic anhydrase II gene. J. biol. Chem. 260: 12130–35 (1985)
173. Wade R. et al.: Nucleotide sequence, tissue-specific expression, and chromosome location of human carbonic anhydrase III: The human CAIII gene is located on the same chromosome as the closely liked CAI and CAII genes. Proc. Nat. Acad. Sci. USA 83: 9571–75 (1986)
174. Waheed A., Risley J. M. and van Etten R. L.: Structural and immunochemical relationships among mammalian arylsulfatase A enzymes. Comp. Biochem. Physiol. Pt. B 82: 855–862 (1985)
175. Waite M.: Phospholipases. Plenum, New York 1987
176. Warmke J., Drysdale R. and Ganetzky B.: A distinct potassium channel polypeptide encoded by the Drosophila-EAG locus. Science 252: 1560–62 (1991)
177. Wei A.: K+ current diversity is produced by an extended gene family conserved in Drosophila and mouse. Science 248: 599–603 (1990)
178. Wendorff K. M. et al.: The sequence of equine muscle carbonic anhydrase. J. biol. Chem. 260: 6129–32 (1985)
179. Wheeler C. H., Boothby K. M. and Goldsworthy G. J.: C1-proteins and the regulation of lipoprotein lipase activity in locust flight muscle. Biol. Chem. Hoppe-Seyler 367: 1127–33 (1986)
180. Willidsen P., Riding G. A. and McKenna R. V.: Adenosine triphosphatases in the cattle tick Boophilus microplus. Insect Biochem. 17: 127–131 (1987)
181. Wray D. W., Norman R. I. and Hess P. (eds.): Calcium channels. Structure and function. New York Acad. Sciences, New York 1989
182. Yamada K., Akasaka K. and Shimada H.: Structure of sea-urchin arylsulfatase gene. Eur. J. Biochem. 186: 405–410 (1989)
183. Yang C. Y. et al.: Structure of bovine milk lipoprotein lipase. J. Biol. Chem. 264: 16822–27 (1989)
184. Yokota Y.: Purification and characterization of particulate acid phosphatase from eggs of Mediterranean sea urchins. Comp. Biochem. Physiol. Pt. B 84: 255–260 (1986)
185. Yora T. and Sakagishi Y.: Comparative biochemical studies of alkaline phosphatase isozymes in fish, amphibians, reptiles, birds and mammals. Comp. Biochem. Physiol. Pt. B 85: 649–658 (1986)
186. Yoshihara C. M., Lee J. D. and Dodgson J. B.: The chicken carbonic anhydrase II gene: evidence for a recent shift in intron position. Nucleic Acids Res. 15: 753–770 (1987)
187. Zarain-Herzberg A., MacLennan D. H. and Periasamy M.: Characterization of rabbit cardiac sarco(endo)-plasmic reticulum Ca-ATPase gene. J. Biol. Chem. 265: 4670–77 (1990)
188. Zhu X. L. and Sly W. S.: Carbonic anhydrase IV from human lung. Purification, characterization, and comparison with membrane carbonic anhydrase from human kidney. J. Biol. Chem. 265: 8795–8801 (1990)

18 Oxidative Metabolism

The extraction of energy from substrates always involves the complete or partial oxidation of the substrate, in the course of which electrons or hydrogen atoms are transferred to one or more consecutively activated acceptors. The primary acceptor must always be reoxidized, whilst the final acceptor accumulates in the reduced form. Oxygen is the ideal terminal acceptor because its reduction product, water, can be accumulated without harm. Under aerobic conditions, nutrients (with the exception of nitrogenous excretions) are oxidized completely to carbon dioxide and water. The majority of biologically useful energy, in the form of ATP, is obtained in this respiratory chain by electron transfer to oxygen (oxidative phosphorylation).

The **processes of aerobic energy-yielding metabolism** can be divided into four phases (Fig. 18.1). In phase I, which may already occur in the gut lumen, nutrient compounds are degraded to their basic components without the release of energy. In phases II and III, hydrogen atoms or their electrons are transferred pairwise from the nutrient molecules to acceptors, and individual carbon atoms are released as CO_2. Whereas phase III (the citric acid cycle) is the final phase of oxidative degradation that is common to all nutrients, phase II consists of various reactions that introduce different nutrients individually into the cycle. Such reactions of phases II and III that do not include dehydrogenation or decarboxylation may be considered as preparatory to such steps. The stepwise transfer of electrons from the primary acceptor to oxygen in the respiratory chain (phase IV) results in the formation of up to three energy-rich ATP residues per electron pair. Without the involvement of the respiratory chain, ATP can also be formed in one reaction of the citric acid cycle, or in two reactions of phase II of carbohydrate catabolism (substrate chain phosphorylation).

The components of the respiratory chain and oxidative phosphorylation are located in the inner membrane of the mitochondria, and the enzymes of the citric acid cycle, fatty acid β-oxidation and oxidative decarboxylation are located in the mitochondrial matrix. Because of their high solubility, the enzymes citrate synthase, aconitase, NAD-specific isocitrate dehydrogenase, succinate-CoA synthase, fumarase and malate dehydrogenase were thought to exist freely in the matrix. More careful extraction, however, has shown that these enzymes from rat liver mitochondria can be isolated as a high molecular weight complex which is apparently associated in vivo with the matrix surface of the inner membrane. The 2-oxoglutarate dehydrogenase complex and succinate dehydrogenase are tightly bound to the membrane [140].

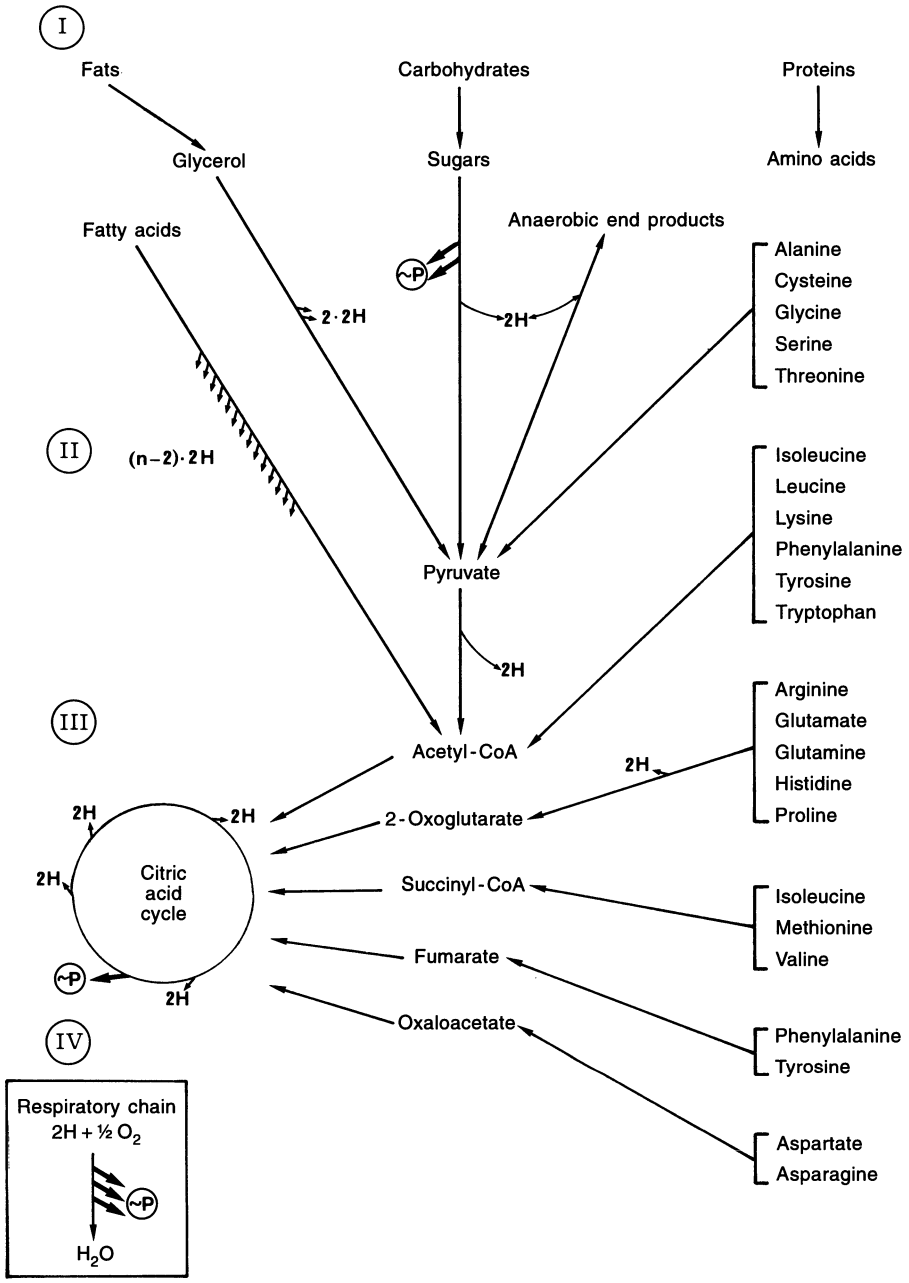


Fig. 18.1. A schematic representation of energy-yielding metabolism. In phase *I*, complex compounds in the diet are degraded to their basic components; in phases *II* and *III*, pairs of hydrogen atoms or their electrons are transferred to acceptors; and in phase *IV*, these are oxidized by oxygen to water. Dehydrogenations are shown by $\rightarrow 2H$, and phosphorylations from ADp to ATP by $\rightarrow(\sim P)$

Energy-yielding metabolism requires **transport processes through the mitochondrial membranes;** oxygen, ADP, inorganic phosphate and substrates enter the mitochondria and water, carbon dioxide and ATP are removed. Furthermore, some biosynthetic reactions involve both mitochondrial and cytosolic enzymes; examples include urea biosynthesis, gluconeogenesis and ammoniogenesis in the kidney, and here too the transport of substances through both mitochondrial membranes is required. The outer membrane contains voltage-dependent, anion-selective channels with an effective diameter of 1.7–2.0 nm; in their open state these allow the free passage of molecules up to 3 kDa. They consist of basic proteins of 30–35 kDa, the porins, which are homologous in all eukaryotes and have no sequence similarity to the bacterial porins [34, 184]. Transport of substances through the inner membrane requires specific transport proteins (TP) (transporter proteins, carrier proteins or translocases). As many as 12 specific transporters have been identified so far in mammalian mitochondria on the basis of their interaction with different substrates and

inhibitors (Table 18.1), but they are not all present in all mitochondria. The transporters for adenine nucleotides, phosphate and pyruvate are present in all mitochondria, and the carnitine transporter is present in mitochondria of all fatty acid-oxidizing tissues. Other transporters have a more restricted distribution, e.g. the proton-dependent glutamate transporter (glutamate hydroxyl carrier) which is involved in the synthesis of urea and glucose from glutamate in the liver, the glutamine transporter required for ammoniogenesis in the kidney, the dicarboxyl transporter in all tissues carrying out gluconeogenesis, and the tricarboxyl transporter in all fatty acid-synthesizing tissues [38, 87].

The ATP/ADP TP, phosphate TP, 2-oxoglutarate TP and the uncoupling protein belong to the same protein super-family [144]. Their sequences consist of three homologous repeats of about 100 amino acids; these repeats form six transmembrane segments and were apparently formed by gene triplication. The human and rat phosphate TPs show about 90 % sequence agreement and are cross-reactive with those of cattle and the blowfly [135]. Tissue-specific isoforms of the ATP/ADP and phosphate TPs are known from mammals. For example, there are two bovine ATP/ADP TPs, encoded by different genes, which differ in 33 of their 297 amino acids and predominate in either cardiac muscle or the gut. The human and bovine genomes contain a whole family of ATP/ADP TP homologous gene sequences [62, 132].

Table 18.1. Transport proteins (TP) of the inner mitchondrial membrane [38]

(Type) Name	Physiological substrate
(Electroneutral; proton compensated)	
Glutamate hydroxyl TP	Glutamate
Pyruvate TP	Monocarboxy acids, ketone bodies, branched keto acids
Phosphate TP	Phosphate, arsenate
Ornithine TP	Ornithine, citrulline, lysine
(Electroneutral anion exchangers)	
Dicarboxylate TP	Phosphate, malate, succinate, oxaloacetate
2-Oxoglutarate TP	Malate, 2-oxoglutarate, succinate, oxaloacetate
Tricarboxylate TP	Citrate, isocitrate, phosphoenolpyruvate, malate, succinate
(Neutral)	
Carnitine TP	Carnitine, acylcarnitine
Neutral amino acid TP	Neutral amino acids
Glutamine TP	Glutamine
(Electrogens)	
Adenine nucleotide TP	ADP, ATP
Glutamate/aspartate TP	Glutamate/aspartate

The degradation of carbohydrates by glyceraldehyde phosphate dehydrogenase in the cytosol results in the formation of NADH which is reoxidized under aerobic conditions in the mitochondria. However, the inner mitochondrial membrane is impermeable to NADH and NAD^+, and the transport of reducing equivalents through the mitochondrial membranes is carried out by so-called **shuttles**, the most important of which are the malate/aspartate shuttle and the glycerol-3-phosphate shuttle (α-glycerophosphate shuttle). The 2-oxoglutarate/malate and the glutamate/aspartate transporters take part in the malate/aspartate shuttle. Because the ratio NADH: NAD^+ is higher in the mitochondria than in the cytosol, the inward transport of reducing equivalents takes place against a gradient with the consumption of energy. The energy-consuming step is actually the outward transport of the aspartate ion, together with a proton, in exchange for non-dissociated glutamic acid. No transporter is involved in the glycerol-3-phosphate shuttle because the mitochondrial glycerol-3-phosphate dehydrogenase binds its substrate to the outer side of the membrane and only transports reducing equivalents through the membrane to the electron transport system. This transport requires no energy but yields only two, instead of three, ATPs per electron pair [38]. The malate/aspartate shuttle is the more important in vertebrate mitochondria, whereas the glycerol-3-phosphate shuttle is at its highest activity in insect flight muscles. The malate/aspartate shuttle is very important to insects in certain circumstances, e.g. in the eggs of *Bombyx mori* at the end of the diapause, when the large accumulation of sorbitol is converted back to glycogen and the resulting NADH must be reoxidized by the mitochondria [179].

The translocases, which make available substrates and cofactors but with the consumption of energy, the citric acid cycle, which produces reducing equivalents, and the respiratory chain, which reoxidizes NADH and yields ATP, show a strict **functional and regulatory interdependency**. A central role in this regulation is played by adenylate, i.e. the relative concentrations of ATP, ADP and AMP or quotients calculated from them, for example the "energy charge". The ratio [ADP]:[ATP] regulates the citric acid cycle at the NAD-specific isocitrate dehydrogenase [20], and the available ADP limits mitochondrial respiration (respiratory control) [38]. These regulatory

mechanisms are important points of interaction for the adaptation of mitochondrial metabolism to different environmental conditions. Investigations of the respiration-stimulating effects of exogenous substrates on isolated mitochondria are classic experiments of cell biology, with the mitochondria of rat liver and insect flight muscles acting as model systems. Species- and tissue-specific substrate preferences are dependent in a complicated manner upon the available transporters and enzymes.

The integrated multi-enzyme system of the citric acid cycle, the respiratory chain and oxidative phosphorylation appears in principle to be the same in the mitochondria of all animals that have predominantly aerobic metabolism. Adaptive differences are mainly quantitative and involve the overall intensity of the oxidative process as well as the proportion of different substrates consumed in energy-yielding metabolism. Marked deviations in **mitochondrial function** are found in animals that are especially adapted to anaerobic energy metabolism, e.g. the parasitic worms and certain species from tidal zones (p. 549). The complete enzyme system of the citric acid cycle is found even in the parasitic worms, although several enzymes, such as aconitase and isocitrate dehydrogenase, have very low activities (Table 18.2). The existence of the citric acid cycle in animals that have continuously anaerobic metabolism is explained by the reverse use of the last part of the cycle for catalysing the reductive formation of succinate from malate, and by the necessity for the cycle, also under anaerobic conditions, as the starting point for the synthesis of essential cellular components.

The cytochrome system is also present in the parasitic worms but often differs from that in the mammals by the existence of a branch or parallel pathway leading to a different terminal oxidase, the so-called cytochrome o. Part of the electron transport system is necessary for the reduction of fumarate to succinate. Oxidative phosphorylation and respiratory control by the ADP concentration have been demonstrated, for example, in the mitochondria of the roundworm *Ascaris lumbricoides* [9]. Fumarate reduction is also the preferred direction of mitochondrial metabolism in various Kinetoplastida *(Leishmania* spp., *Trypanosoma cruzi)* and the ciliate *Tetrahymena* [23, 101, 157]. The facultatively anaerobic invertebrates of tidal zones show a relatively rapid switch between the different functional conditions of the mitochondria. Some protozoans, e.g. the trichomonads *Entamoeba hystolytica* and *Giardia lamblia*, lack mitochondria and thus lack both the enzymes of the citric acid cycle and the cytochrome system; however, a special sort of electron transport system is present instead [177].

Apart from the terminal oxidase of the respiratory chain, the cytochrome oxidase, there are several hundred other enzymes that can react directly with molecular oxygen. Several of them, e.g. the D- and L-amino acid oxidases, the oxygenases involved in tyrosine and tryptophan degradation and various hydroxylases, have been referred to already. The microsomal hydroxylases or mixed-function oxygenases, which contain cytochrome P-450 as the prosthetic group, are especially interesting for comparative biochemistry and will be discussed in detail (Sect. 18.5.2). Electron transfer to oxygen can give rise to particularly reactive and toxic reaction products, in particular the superoxide ion, O_2^-, and hydrogen peroxide, H_2O_2. These are detoxified by the superoxide dismutases, catalases and peroxidases, about which much is known that is of interest to comparative biochemistry.

Table 18.2. The activities (mU/mg protein) at 30 °C of the enzymes of the citric acid cycle in mitochondria of the endoparasitic worms *Ascaris lumbricoides* and *Fasciola hepatica* compared with those of the rat liver [9]

	Ascaris	*Fasciola*	Rat
Pyruvate dehydrogenase	32	18	23
Citrate synthase	10	15	25
Aconitase	0.8	5	80
Isocitrate dehydrogenase			
NAD specific	0.4	0	1
NADP specific	5	17	263
2-Oxoglutarate decarboxylase	4	22	18
Succinate dehydrogenase	215	28	21
Fumarase	225	100	432
Malate dehydrogenase	7650	2180	3100

18.1 The Citric Acid Cycle

The **central position of the citric acid cycle in metabolism** allows the introduction of carbon chains from very different biological substances (Fig. 18.1). In addition, intermediates of the cycle can be diverted and used for biosynthesis. The reducing equivalents, in the form of NADH or NADPH, required for various biosynthetic processes can be made available by the mitochondrial dehydrogenases or their cytosolic isoenzymes. Few animals are able to convert fat to carbohydrate, as is possible, for example, in germinating

plant seeds. For this purpose, the citric acid cycle must be modified to the glyoxylate cycle (Fig. 18.4). The citric acid cycle involves two decarboxylations and, on balance, completely degrades the introduced activated acetic acid to CO_2; the acetyl-CoA released by β-oxidation from fatty acids cannot be used for gluconeogenesis via oxaloacetate. The net synthesis of carbohydrates from fatty acids is possible only if the decarboxylation can be circumvented; in the glyoxylate cycle this occurs by the action of the enzymes isocitrate lyase and malate synthase. If, as a result of the utilization of intermediates for biosynthesis, the oxaloacetate that is essential for the continuous functioning of the cycle is insufficient, this must be supplied by so-called **anaplerotic reactions**, of which the carboxylation of pyruvate or phosphoenolpyruvate is especially important (Fig. 18.2).

The flux through the citric acid cycle is regulated on the one hand by the availability of substrates and cosubstrates, and on the other hand by the allosteric properties of various cycle enzymes, in particular NAD-dependent isocitrate dehydrogenase, but also citrate synthase, 2-oxoglutarate decarboxylase and succinate dehydrogenase. The pyruvate dehydrogenase complex, which forms the connection between glycolysis and the citric acid cycle, is also allosterically regulated in a complex manner. In such a central and ubiquitous enzyme system as the citric acid cycle, changes in the kinetic-regulatory properties of enzymes are especially important mechanisms of adaptive evolution.

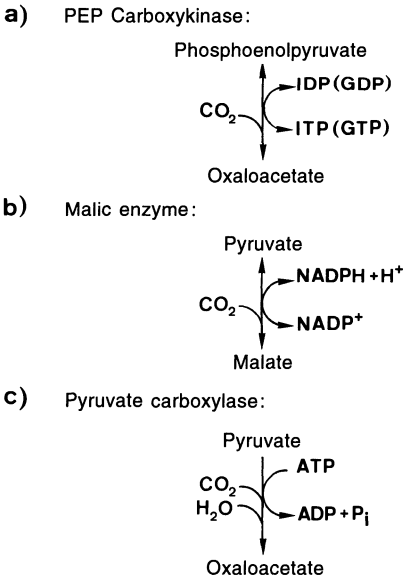


Fig. 18.2a–c. Enzymes of the carboxylation of phosphoenolpyruvate and pyruvate

18.1.1 Pyruvate Dehydrogenase Complex and Citrate Synthase

The **pyruvate dehydrogenase complex** (PDC) catalyses the irreversible reaction

$$\text{Pyruvate} + \text{CoA} + \text{NAD}^+ \rightarrow \text{acetyl-CoA} + \text{CO}_2 + \text{NADH} + \text{H}^+ \quad (18.1)$$

and thereby introduces pyruvate from glycolysis or from the degradation of amino acids into the citric acid cycle. The PDC in the roundworm *Ascaris lumbricoides* also catalyses the condensation of pyruvate with acetaldehyde to acetoin (Fig. 18.3).

Three α-ketoacid dehydrogenase complexes with similar structure are known in mammals: PDC, the 2-oxoglutarate dehydrogenase complex and the branched chain α-ketoacid dehydrogenase complex. These are all multi-enzyme complexes that each have three enzyme activities: decarboxylase, dihydrolipoamide acetyltransferase and dihydrolipoamide dehydrogenase. The first two enzymes are complex specific but the last enzyme is common to all three complexes [142]. The mammalian PDC is composed of six components: several $\alpha_2\beta_2$ tetramers of pyruvate decarboxylase (E_1), the multimeric dihydrolipoamide acetyltransferase (E_2), several homodimers of dihydrolipoamide dehydrogenase (E_3) and a few molecules each of the E_1-specific protein kinase, the phospho-E_1-phosphatase, and one 50-kDa polypeptide (protein X) which apparently has a regulatory function. Pyruvate decarboxylase (E_1) consists of two α subunits of 41 kDa, two β subunits of 36 kDa and two molecules of thiamine pyrophosphate. E_1 is inactivated by phosphorylation of the α subunit and activated by dephosphorylation. NADH and acetyl-CoA have inactivating effects on the PDC by influencing the protein kinase, and NAD^+, pyruvate, ADP and Ca^{2+} have activating effects. Three phosphoserine resi-

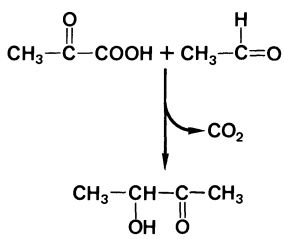


Fig. 18.3. In the roundworm *Ascaris*, pyruvate dehydrogenase also catalyses the condensation of pyruvate with acetaldehyde to acetoin [9]

dues have been identified in the inactivated α sequence [127, 142]. The multimeric aggregate of dihydrolipoamide acetyltransferase molecules (E_2) appears to form the core of the PDC. The subunits of the human and porcine homodimeric dihydrolipoamide dehydrogenases agree in >98 % of their 509 amino acids; the enzyme thus shows very conservative evolution [130]. High molecular weight PDCs with similar kinetic-regulatory properties have been isolated not only from the tissues of mammals and birds and from various microorganisms but also from the fat bodies of the silkworm *Bombyx mori*, the gills of the mussel *Modiolus demissus*, and the roundworm *Ascaris lumbricoides* [104, 128]. The enzyme complexes of the different species, however, differ significantly in the critical value of the NAD^+:NADH ratio above which PDC is inhibited. This has a value of 10 for the rat liver, 5 for mussel gills, and 0.07–0.7 for the roundworm; in the latter case, the PDC is adapted to the higher level of reduction in *Ascaris* mitochondria [128]. Various termite species contain no pyruvate dehydrogenase activity whatsoever, and in this case the energy requirements are covered mainly by acetate that is produced by gut symbionts [120].

Citrate synthase (CS) catalyses the aldol condensation of oxaloacetate and acetyl-CoA to citrate and CoA. The mammalian enzyme is a dimer of about 100 kDa. It is subjected to feedback inhibition by citrate and is also strongly inhibited by ATP and mono- and divalent metal ions. The less well-known CSs of the invertebrates agree to a large extent with the mammalian enzyme in their molecular mass and response to inhibitors. CSs have been examined from the adductor muscle of the scallop *Pecten alba*, from sea-urchin eggs, from the gut sacs of the starfish *Luidia clathrata*, from the agent of Chagas disease *Trypanosoma cruzi*, and from the parasitic flagellate of insects *Crithidia fasciculata* [23, 102, 147]. The CS of *Tetrahymena pyriformis* is unusual and, like other enzymes of this ciliate, has some prokaryotic features; it agrees with the CSs of other eukaryotes, but also with those of Gram-positive bacteria such as *Bacillus subtilis*, in having a molecular mass of 120 kDa, compared with 250 kDa for *Escherichia coli* and other Gram-negative bacteria; it is also inhibited by ATP but not by NADH. However, the *Tetrahymena* CS is very different from the CSs of other animals and higher plants and resembles bacterial enzymes in that it is stimulated by KCl at higher concentrations and is specifically inhibited by DTNB (dithiobis-2-nitrobenzoate) [41].

18.1.2 Aconitase and Isocitrate Dehydrogenases

Aconitase catalyses the stereo-specific isomerization of citrate to threo-D_s-isocitrate such that the secondary hydroxyl group is bound to that part of the citrate molecule that arises from oxaloacetate. The enzyme from invertebrates has been little investigated, but the aconitase from the hepatopancreas of the oyster *Crassostrea virginica* appears to be very similar to the rat liver enzyme [156].

The **isocitrate dehydrogenases** (IDH) catalyse the oxidative decarboxylation of isocitrate to 2-oxoglutarate. Two IDH types can be distinguished in vertebrates, insects and molluscs: the NAD-specific IDH from the mitochondria, and the NADP-specific IDH, which in animals is present as different cytosolic and mitochondrial isoenzymes. The intracellular distribution of the two enzyme types led to the idea that the NADP-IDH is not really an enzyme of the citric acid cycle but delivers NADPH for biosynthetic purposes. And, in fact, flight-muscle mitochondria, for example, contain much more NAD-IDH than NADP-IDH; this corresponds to the lower concentration of NADPH in these mitochondria. Furthermore, NAD-IDH in insects and mammals is part of a constant proportion group together with malate dehydrogenase and cytochrome oxidase, whereas NADP-IDH shows very variable activities relative to these enzymes of energy-yielding metabolism [22]. However, in many cells the NAD-specific IDH has a much lower activity than that of the NADP-IDH, for example in vertebrate cardiac muscle, avian breast muscle, the red muscle and liver cells of fish, and the muscles of annelids, molluscs, crustaceans, xiphosurans and parasitic intestinal worms (Table 18.2) [9, 115]. The mantle muscle and hepatopancreas of the edible mussel *Mytilus edulis* contains almost exclusively NADP-IDH [143], whilst both types could be isolated from the shell-adductor muscle [79]. The NAD-specific enzyme is completely absent from the turbellarian *Polycelis nigra* and the protozoans *Trypanosoma cruzi, Leishmania* sp. and *Tetrahymena pyriformis* [23, 41, 101, 157]. As there is still a functioning citric acid cycle in these animals, it can be assumed that the mitochondrial NADP-IDH has the appropriate activity. The problem that the respiratory chain can only reoxidize NADH is solved by an NADPH:NAD^+ transhydrogenase, which is known from the mitochondria of mammals, fish, insects and parasitic worms [22]. It has been demonstrated that the

cytoplasmic NADP-IDH, for example, in the mantle tissue and hepatopancreas of the edible mussel, in the fly *Drosophila* and in vertebrates, together with glucose-6-phosphate dehydrogenase, 6-phosphogluconate dehydrogenase and malic enzyme, produces NADPH for the biosynthesis of fatty acids and sterols.

In contrast to the NADP-IDH, the NAD-IDH has marked **regulatory properties** and is probably the most important regulation enzyme of the citric acid cycle. ADP strongly stimulates the NAD-IDH of, for example, vertebrates, insects and mussels, and this effect may be such that no enzyme activity is measurable in the absence of ADP; ATP is inhibitory [20, 22, 79]. Ca^{2+} activates the NAD-IDH from the cardiac muscle of mammals, birds, amphibians and fish, but not from insect flight muscle, higher plants or bacteria [105]; in fact, there is evidence that Ca^{2+} inhibits the insect enzyme [22]. Ca^{2+} regulation of NAD-IDH, like that of 2-oxoglutarate dehydrogenase, is apparently a novelty of vertebrate evolution. It allows the hormonal regulation of the citric acid cycle with Ca^{2+} as the secondary messenger without the far-reaching effects of a change in the ATP:ADP ratio or NADH:NAD^+ ratio [105]. NADP-IDH usually shows no allosteric properties, although it has been reported that the enzyme from the hepatopancreas of the edible mussel is inhibited, in order of effectiveness, by ATP > ADP > AMP [143].

18.1.3 The Citric Acid Cycle Between 2-Oxoglutarate and Oxaloacetate

The **2-oxoglutarate dehydrogenase** of the mammals is a multi-enzyme complex, similar to that of pyruvate dehydrogenase, which requires CoA, NAD^+, Mg^{2+} and thiamine pyrophosphate, is stimulated by ADP and is inhibited by ATP and arsenite. The rather scarce comparative data show similar characteristics for the cardiac enzyme complex of the mammals, birds, amphibians and fish as well as of various insects and the gills of the mussel *Modiolus demissus*. In contrast, the activation of 2-oxoglutarate dehydrogenase by Ca^{2+}, like that of NAD-IDH, is restricted to the vertebrates [22, 105, 128]. There are also few comparative biochemical data for the subsequent enzymes of the citric acid cycle. **Succinate thiokinase** catalyses a substrate phosphorylation. The mammalian enzyme is specific for GDP or IDP and shows little activity with ADP. On the other hand, the enzyme isolated from the flight muscle of the fly *Calliphora erythrocephala* prefers ADP [22]. Of the two succinate thiokinases isolated from *Trypanosoma brucei*, one is ADP specific and the other GDP specific [70]. **Succinate dehydrogenase** is a component of the respiratory chain and will be discussed below together with fumarate reductase, which in anaerobic metabolism catalyses the opposite reaction. **Fumarase** is a highly conserved enzyme in evolution; the human and porcine liver enzymes agree in 96 % of their 466 amino acids [145]. The cytoplasmic and mitochondrial fumarases of the rat liver are encoded by the same gene; however, the mRNA of the mitochondrial enzyme codes for an additional N-terminal sequence of 41 amino acids [165].

There is relatively more information about **malate dehydrogenase** (MDH). This catalyses the reversible oxidation of malate to oxaloacetate and is strictly NAD specific in all animals. Most eukaryotes possess two types of MDH, mitochondrial (mMDH) and cytoplasmic (cMDH); MDH is also found in the glycosomes of *Trypanosoma cruzi* and *Leishmania mexicana* [23, 114]. Whereas mMDH is an enzyme of the citric acid cycle, the function of cMDH lies in the supply of NADH for cytoplasmic biosynthetic processes, such as gluconeogenesis and lipid synthesis, and in preparations for the transport of reduction equivalents across the mitochondrial membrane (malate/aspartate shuttle). In anaerobic metabolism, the NADH produced by glyceraldehyde phosphate dehydrogenase is reoxidized by cMDH (p. 550). The cMDH and mMDH of all eukaryotes are dimers with subunits of 33–35 kDa; the MDHs of 33–39 kDa isolated from *Leishmania* probably arise by dissociation of dimers [114]. The ciliate *Tetrahymena pyriformis* is again an exception in that both its MDH isoenzymes are tetramers of 122 kDa and are of a type that is otherwise found only in Gram-positive bacteria; the kinetic properties of the *Tetrahymena* enzyme also correspond more to those of the bacterial MDH [42].

All these enzymes are homologues; even the bacterial MDH and the NADP-specific MDH from the chloroplasts of higher plants belong to the same protein family [48]. The genes of murine mMDH and cMDH have a similar organization with nine exons [155]. The pre-mMDH of the mouse consists of the 314 amino acids of the mature enzyme and a signal sequence (leader) of 24 amino acids; this is typical of a mitochondrial enzyme and includes three basic but no acidic or hydrophobic residues. The sequence agreement

between the enzymes from the mouse and the pig is 95% for mMDH and 93% for cMDH (334 amino acids), but mMDH and cMDH of the mouse have only 23% common amino acids. Rather surprisingly, the agreement between the mouse cMDH and the MDH from *Thermus flavus*, and the agreement between the mouse mMDH and the MDH from *Escherichia coli*, are greater than that between the two mouse isoenzymes [73, 74].

Reptiles, birds and mammals have only one **cMDH locus;** in contrast, most fish and amphibians have two cMDH loci, the products of which form a three-banded pattern of two homodimers and one heterodimer. Two cMDH loci are already found in the cartilaginous fish and apparently even in the acranian *Branchiostoma lanceolatum* and the ascidian *Ciona intestinalis*. Several polyploid species or species hybrids of the teleosts have as many as three or four cMDH loci. In the teleost *Leiostomus xanthurus* the homodimer AA is more temperature stable and corresponds to the cMDH of the higher vertebrates; BB, on the other hand, is less temperature stable but has a higher activity at lower temperatures [152]. Several electrophoretically distinct MDHs are also usually found in parasitic trematodes, cestodes and nematodes; for example, the canine roundworm *Toxocara canis* and the porcine roundworm *Ascaris suum* have three cMDHs and one mMDH, although these differ little in their functional properties [100].

18.2 Carboxylases

18.2.1 Pyruvate Carboxylase

Pyruvate carboxylase (PC) catalyses the formation of oxaloacetate from pyruvate (Fig. 18.2c). It is always located in the mitochondria, and in mammals has the structure of a tetramer of about 500 kDa with four biotinyl groups [81]. In vertebrates, PC is found especially in the liver and kidney, but also occurs in the fat tissue, brain and adrenal gland; in insects, it is present in the fat body [22]. Its main function in these organs is probably the provision of oxaloacetate for gluconeogenesis and other biosynthetic processes. PC is absent from vertebrate skeletal muscle, but is found in the cardiac muscle of the frog and rat and has a high activity in the flight muscles of insects, where oxaloacetate is required for the activation of the citric acid cycle [22]. In the flagellate *Leishmania donovani* from the Kinetoplastida, PC has an activity that is 30- to 50-fold higher than that of citrate synthase, and it serves to metabolize part of the pyruvate formed in glycolysis to succinate via oxaloacetate, malate and fumarate; phosphoenolpyruvate carboxykinase (PEPCK) is not present [101]. *Crithidia fasciculata* from the same group of flagellates has some PC activity, but the carboxylation step in succinate production is probably carried out by PEPCK; finally, *Trypanosoma cruzi* has no PC activity at all but does have an active PEPCK [23]. All PCs examined are only active in the presence of acetyl-CoA, which increases the affinity for HCO_3^- and ATP [57, 81]. The ciliate *Tetrahymena pyriformis* also has no PC and, as in plants, the anaplerotic function is taken over by a **phosphoenolpyruvate carboxylase**, a nucleotide-dependent enzyme that carboxylates phosphoenolpyruvate (PEP) to oxaloacetate. PEP carboxylase, like PC, is activated by acetyl-CoA. Flagellates of the genus *Leishmania* also possess a PEP carboxylase [101, 157].

18.2.2 Phosphoenolpyruvate Carboxykinase

The direction of the reversible reaction catalysed by PEPCK varies according to the function (Fig. 18.2a). In gluconeogenesis, the enzyme has a decarboxylating action and in cooperation with PC brings about the production of phosphoenolpyruvate from pyruvate. In specialized anaerobes and in many protozoans, PEPCK catalyses the carboxylation step in the metabolic pathway from the C_3 compound phosphoenolpyruvate to the C_4 compound succinate. The reasons for the different directions of the PEPCK reaction, for example in vertebrates and parasitic worms, were discussed on p. 550. PEPCK in the vertebrate liver serves in gluconeogenesis, and has higher activity with GDP/GTP than with IDP/ITP. Depending upon the species, the enzyme is found exclusively in the cytosol, e.g. in the rat, mouse and golden hamster, or almost entirely in the mitochondria, as in the rabbit and chicken, or equally distributed between the two compartments, as in humans, cattle, sheep, pigs, guinea-pigs, and the frog *Rana catesbeiana* [53, 57, 121]. The PEPCKs from mammalian liver cytoplasm and mitochondria differ in their structural, immunological and kinetic properties [51, 57]. The only sequence so far recorded is that of the 622 amino acids of the chicken PEPCK [31].

The two isoenzymes in the frog and tadpole liver are identical immunologically and

increase in parallel during spontaneous or triiodothyronine-induced metamorphosis. In vitro tracer experiments show that both are taken up into intact mitochondria and are therefore the same enzyme, corresponding apparently to the mitochondrial mammalian type [53]. The enzyme from the shark *Squalus acanthias* is 90 % mitochondrial and differs from all other PEPCKs in its low substrate affinity (K_m = 1 μ mol/l) and its activation by urea [75]. The tapeworm *Moniezia expansa* also possesses cytoplasmic and mitochondrial PEPCKs which are very similar, if not identical [9]. In contrast, the PEPCK in the nematodes *Ascaris lumbricoides* and *Turbatrix aceti* is predominantly cytoplasmic. The *Ascaris* enzyme, like that of the mammals, has a mass of about 65 kDa but differs from the latter in having a higher activity with IDP than with GDP [9]. PEPCK is localized in the glycosomes of the flagellates *Trypanosoma cruzi, T. brucei* and *Leishmania mexicana* [114]. The enzyme from *T. brucei* shows significant homology to the enzymes from bakers' yeast and *E. coli*, but no sequence similarity to the PEPCK from the rat, chicken or *Drosophila* [126].

18.2.3 Malic Enzyme

The reaction of the malic enzyme (ME) is, in principle, bidirectional (Fig. 18.2b). However, most MEs are much more active in decarboxylation than in the opposite direction. The only exception is that the activities with pyruvate and malate are equally high, for example for the cytoplasmic ME of the abdominal muscle of the prawn *Crangon crangon* [13]. All MEs require divalent metal ions such as Mn^{2+} or Mg^{2+}. Most prefer NADP as the **coenzyme**, but many mitochondria also have NAD-specific enzymes. NADP-specific MEs are also found in the cytoplasm of many vertebrates and invertebrates; the NAD-specific enzyme is always mitochondrial. NADP-specific MEs show only little activity with NAD^+ but the coenzyme specificity of the "NAD-specific" type is not so strict and these are better referred to as NAD(P) specific [168]. For example, the NAD-specific MEs from various rat organs show with $NADP^+$ 62–71 % of the activity achieved with NAD^+, and the ME from the flight muscle mitochondria of the beetles *Catharius* sp. and *Leptinotarsa decemlineata* shows 30–40 % [67, 176]. In addition, the coenzyme specificity, like the kinetic properties of the ME, is dependent upon the divalent metal ion involved; for example, the NAD(P)-specific MEs of various mammals and insects are active with NADP only in the presence of Mn^{2+} but not of Mg^{2+} [176].

NADP- and NAD-specific MEs vary widely in their species- and organ-specific distribution. In the rat, they have approximately equal activity in the testis and lung, the NADP type predominates in the brain, heart, kidney and muscle, and the NAD type predominates in the small intestine, spleen and thymus; the liver lacks both activities. In contrast, cardiac muscle mitochondria of the rabbit, guinea-pig, pigeon and rainbow trout and the breast muscle of the pigeon have both types together [158]. The NADP-specific MEs of vertebrate red muscle are mainly located in the mitochondria and those of white muscle in the cytoplasm. The main function of the cytosolic ME is probably the production of NADPH for biosynthetic purposes. The function of the mitochondrial NADP-specific ME is less clear but may include the production of pyruvate. This enzyme is commonly found in tissues that show a high rate of cell renewal [98].

In insects, the NADP-specific enzyme is found mainly in fat bodies and the NAD-specific enzyme occurs in flight muscles [23]. In insect flight muscles that are specialized in the utilization of proline, the NAD-specific ME makes available pyruvate as the acceptor for the transamination reaction (see Fig. 12.6, p. 421). The enzyme from the beetles *Catharius* sp. and *Leptinotarsa decemlineata* has been isolated and characterized [67, 176]. The NADP-specific ME from the abdominal muscle of the prawn *Crangon crangon* is exclusively cytoplasmic, and that of the crayfish *Orconectes limosus* is predominantly mitochondrial [13, 159]. In the mitochondria of animals which are specialized for anaerobic energy extraction, ME catalyses the oxidative partial reaction of the dismutation of malate to pyruvate and succinate (p. 551). These MEs are NAD specific in the parasitic nematodes and some tapeworms, but NADP specific in other cestode species and various annelids. The NAD-specific MEs of the roundworms *Ascaris lumbricoides* and *Toxocara canis* have been isolated and characterized [9, 45].

The MEs are tetramers with a **molecular mass** usually of the order of 260 kDa [13, 67, 158]. There are only a few exceptions: 130 kDa for the two MEs from the flagellate *Trypanosoma cruzi* [23], 190 kDa for the NAD-specific ME from trout heart [158], 220 kDa for the NAD-specific ME from the flight muscles of the potato beetle [176], and 320 kDa for the liver ME of the teleost

Dicentrarchus labrax [98]. The ME gene of the rat is unusually large at 95 kb and contains introns of >15 kb [112]. In contrast to that of the cytoplasm, the mitochondrial ME from the muscles of vertebrates and arthropods shows **cooperativity** that is dependent upon the pH and divalent metal ions. The Hill constant of the ME from the abdominal muscle of the crayfish *Orconectes limosus* increases from 1.1 to 1.8 between pH 7 and 8. The NAD-specific ME from the flight muscle of the beetle *Catharsius* sp. shows a lower cooperativity, but a higher substrate affinity with Mn^{2+} than with Mg^{2+} [67]. All MEs are inhibited by oxaloacetate and ATP and many are also inhibited by succinate, ADP and AMP [67, 176]. Depending upon their functional roles, the mitochondrial and cytoplasmic MEs may respond rather differently. In the flagellate *Trypanosoma cruzi*, the cytoplasmic ME-II is only slightly inhibited by oxaloacetate (K_i = 90 μmol/l), and is even stimulated by succinate and aspartate; it apparently serves in the degradation of excessive C_4 acids such as aspartate. The mitochondrial ME-I has a much higher affinity for oxaloacetate (K_i = 9 μ mol/l) and is probably involved in the citric acid cycle and succinate formation [23].

18.3 The Glyoxylate Cycle

In addition to the typical citric acid cycle enzymes citrate synthase, aconitase and malate dehydrogenase, the glyoxylate cycle (Fig. 18.4) includes two further enzymes: **isocitrate lyase**, which reversibly cleaves isocitrate to succinate and glyoxylate, and **malate synthase**, which produces malate from glyoxylate and acetyl-CoA. This pathway circumvents the decarboxylation steps of the citric acid cycle, on balance produces one succinate from two acetyl-CoAs, and allows the conversion of fatty acids to carbohydrates. The glyoxylate cycle is widely found in bacteria, yeast, fern spores and plant seeds. In the animal kingdom, the enzymatic prerequisites for a functional glyoxylate cycle are found in the protozoans *Tetrahymena pyriformis* and *Leishmania* sp., several parasitic and free-living nematodes, the liver fluke *Fasciola hepatica*, the marine mussel *Petricola pholadiformis* and the tick Hyalomma dromedarii [12, 27, 122, 173]. In the vertebrates, following some rather doubtful reports, e.g. for the bladder of the toad *Bufo marinus*, a potentially functional glyoxylate cycle has been found in the rat liver and in the brown fat of the black bear [32, 33]. All attempts to demonstrate the glyoxylate cycle in insects and other arthropods have so far failed. In all cases, only one or the other of the

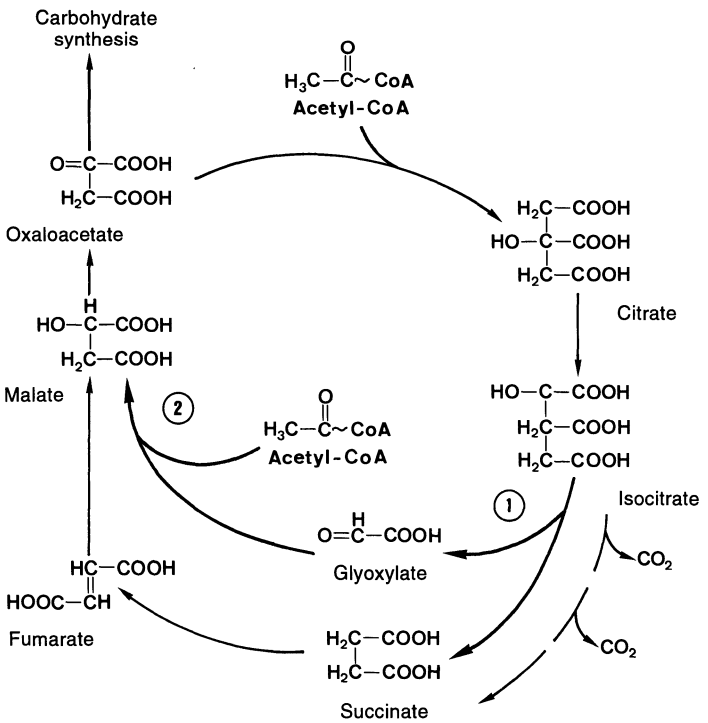


Fig. 18.4. The glyoxylate cycle produces, on balance, one oxaloacetate from two acetyl-CoA residues, and thereby allows the net synthesis of carbohydrates from fatty acids. In addition to enzymes of the citric acid cycle, it requires two further enzymes: *1*, isocitrate lyase; and *2*, malate synthase

two key enzymes was found to be present at low levels, e.g. isocitrate lyase in *Prodenia eridania* and *Tribolium confusum*, and malate synthase in *Aedes solicitans* and *Periplaneta americana;* the complete cycle was never present. Isocitrate lyase has even been isolated from the eggs of the tick *Hyalomma dromedarii* [22, 27, 77]. Attempts to demonstrate a functional glyoxylate cycle by tracer experiments suffer from the problem that, because of the stereo-specificity of the aconitase, the carbon from acetyl-CoA is incorporated into oxaloacetate and glucose without any increase in their net synthesis.

In the roundworm *Ascaris lumbricoides*, intensive gluconeogenesis from stored lipids occurs between days 10 and 25 of development; both the enzymes of the glyoxylate cycle and PEPCK operate at their maximum activities during this time. At present, it is only possible to speculate about the **biological importance of the glyoxylate cycle** in this case; it is possible that it functions in the preparation of carbohydrate reserves required during migration of the larvae in the body of the host. However, in contrast to *Ascaris*, the adults of free-living nematodes also have a functioning glyoxylate cycle. In the vinegar threadworm *Turbatrix aceti*, the cycle may be involved in the utilization of the acetic acid in the natural environment; in starving animals of this species and in *Panagrellus redivivus* gluconeogenesis takes place at the expense of stored lipid. In response to induced anhydrobiosis, *Aphelenchus avena* produces large amounts of trehalose and glycerol from lipids via the glyoxylate cycle [99]. The glyoxylate cycle in the liver fluke *Fasciola hepatica* may utilize ketogenic amino acids from the host's blood to replenish the large carbohydrate reserves which make up 15–20 % and 32 % of the live weight of adults and eggs, respectively [27].

The glyoxylate cycle in *Ascaris* and *Turbatrix aceti* is not localized, as in plants, in special cell organelles (glyoxysomes) but is found together with the citric acid cycle in the mitochondria. Isocitrate dehydrogenase and isocitrate lyase thus compete for the same substrate. The NADP-specific isocitrate dehydrogenase has a higher activity and substrate affinity than the lyase and must, therefore, be down-regulated so long as the glyoxylate cycle is to function. Several inhibitory mechanisms have been proposed: the concerted effects of oxaloacetate and glyoxylate, phosphorylation of the enzyme, or an increase in the ratio $NADPH:NADP^+$ [137]. The nematode *Caenorhabditis elegans* is reported to contain special organelles which are comparable with the glyoxysomes of higher plants [27, 173]. Acetate can be exploited in gluconeogenesis by the ciliate *Tetrahymena*, and here the isocitrate lyase and malate synthase are located in the peroxisomes. Whereas the plant glyoxysomes contain the full complement of enzymes for the glyoxylate cycle, the *Caenorhabditis* glyoxysomes and the *Tetrahymena* peroxisomes lack citrate synthase, aconitase and malate dehydrogenase, and they must therefore function in cooperation with the mitochondria [27, 173]. Flagellates of the genus *Leishmania* possess isocitrate lyase and malate synthase but citrate production is so low in this case that the glyoxylate cycle is probably of little importance [101]. Of the two key enzymes of the glyoxylate cycle, only the **isocitrate lyases** of some nematodes have been characterized in any detail. Five forms can be separated in *Turbatrix aceti;* these range from tetramers of 480 kDa to monomers of 123 kDa. The enzyme from *Caenorhabditis elegans* has a mass of 256 kDa and that of *Ascaris lumbricoides* a mass of 214 kDa. All isocitrate lyases require Mg^{2+} and are specific for threo-D_s-isocitrate; they are not active with the threo-L_s [27].

18.4 Electron Transport and Oxidative Phosphorylation

The inner mitochondrial membrane contains a series of electron-transporting proteins which, according to the chemiosmotic theory, are energetically coupled to ATP synthesis by an electrochemical proton gradient. The energy conversion system of the mitochondria and the corresponding systems in the thylakoid membranes of chloroplasts and the cell membranes of bacteria are the most complicated multi-enzyme systems known. Amongst the animals, the mitochondria of bovine hearts have been most intensively examined. The inner membranes of these mitochondria consist of 70 % protein and 30 % lipid, of which 40 % is phosphatidylcholine, 35 % phosphatidylethanolamine and 15 % cardiolipin. The enzymes of electron transport and oxidative phosphorylation comprise approximately half of the membrane proteins, and of the remaining proteins many are involved, for example, as transport proteins and dehydrogenases in energy-yielding metabolism [54]. The energy-conversion system of the mitochondrial membrane can be divided into five functionally independent **multi-enzyme complexes** (Table 18.3). Complexes I, III and IV

Table 18.3. The functional complexes of the respiratory chain and oxidative phosphorylation [28, 38, 118]

		Subunits	Prosthetic groups
I	NADH:coenzyme Q oxidoreductase	16–25	1 FMN 16–24 non-haem irons (5–6 centres)
II	Succinate:coenzyme Q oxidoreductase	4	1 FAD 8 non-haem irons (3 centres)
	Ubiquinone		
III	Coenzyme Q: cytochrome c oxidoreductase	6–8	2 b-type haems 1 c-type haem 2 non-haem irons (1 centre)
	Cytochrome c		
IV	Ferrocytochrome c:O_2 oxidoreductase	6–7	2 a-type haems 2 Cu
V	ATP synthase	12	

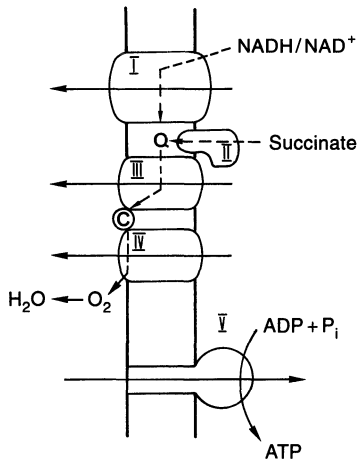


Fig. 18.5. The organization of complexes *I–V* of electron transport and oxidative phosphorylation in the inner mitochondrial membrane can only be shown schematically [118]. The *dashed line* shows electron transport and the *continuous lines* show proton transport. *Q*, Coenzyme; *C*, cytochrome c

make up the electron transport system (ETS) together with ubiquinone (coenzyme Q) and cytochrome c; this is also known as the respiratory chain. Strictly speaking, the succinate:coenzyme Q oxidoreductase (complex II) does not belong to the ETS but, like other enzymes of the inner mitochondrial membrane, plays a supportive role in that it transfers electrons from the substrate to coenzyme Q, but without taking part in the formation of the proton gradient [38, 124]. The ATP synthase constitutes complex V [54]. The individual complexes are typical integral membrane proteins which undergo diffusion motion at the membrane level; thus, their organization can only be represented schematically (Fig. 18.5). Of a total of about 60 different polypeptides making up the five complexes, 13 are encoded by mtDNA (p. 52).

In the first part of the respiratory chain, the redox centres are quinoid structures (FMN, FAD, ubiquinone) and Fe/S clusters, and in the latter part of the chain they are the haem iron of cytochrome and protein-bound copper. Four main groups of cytochromes (a–d) can be distinguished on the basis of the structure of the prosthetic haem group and their protein bonds (Fig. 18.6). Cytochromes of the b group contain protohaem IX, as is found in haemoglobin; in the haem group of the a cytochromes, the protohaem methyl group on C-18 is replaced by a formyl residue and the vinyl group on C-3 is replaced by a methyl-branched side-chain. All the cytochromes with covalent bonds between the haem side-chains and the protein are classified as c cytochromes. The cytochromes of the d group contain a dihydroporphyrin (chlorin) as the prosthetic group; the presence of d cytochromes, which are widespread in bacteria, has often been claimed in various protozoans but has never been unambiguously demonstrated. In general, the different cytochromes are identified by their spectral properties; particularly sharp spectra are obtained by cooling the cytochromes to 77 K in liquid nitrogen, although the spectra thereby obtained are shifted about 3 nm towards shorter wavelengths. The cytochromes, like the haemoglobins, show three absorption bands (α, β and Soret bands). For the exact designation of the individual cytochromes, either a number is added to the group letter (e.g. cytochrome b_5 and c_1) or the position of the α band at room temperature is given (e.g. cytochrome b-566). The general agreement of the cytochrome spectra from bovine heart, the oligochaete *Tubifex* sp. and yeast clearly shows the conservative evolution of the respiratory chain [72]. The cytochrome variants found in some protozoans and specialized anaerobes will be discussed below.

18.4.1 Complexes I–III

Complex I (NADH:coenzyme Q oxidoreductase) is probably the most complicated and, up to now, the least well-known part of the respiratory chain. The isolation of an intact complex I has so

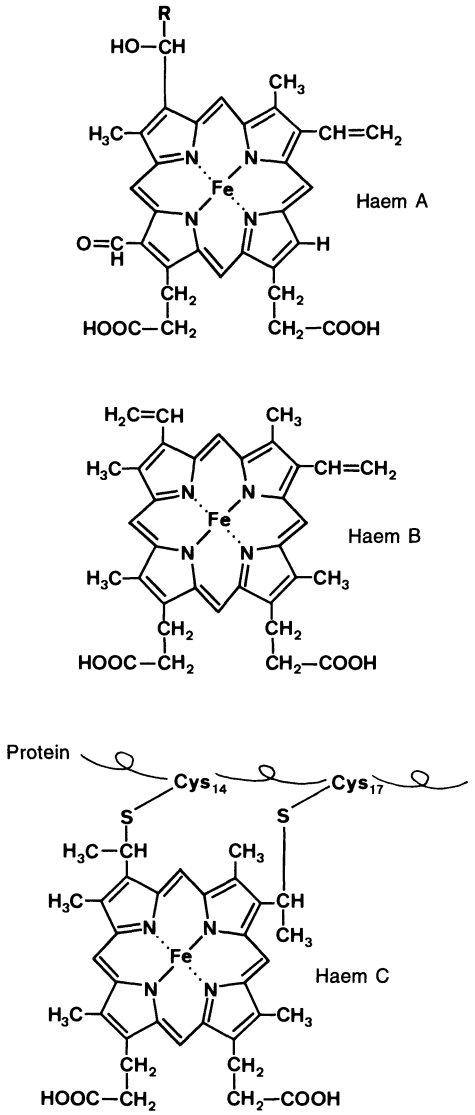


Fig. 18.6. The prosthetic groups of the three cytochrome types of animals. *R* is a multiply methyl-branched hydrocarbon chain, the structure of which is shown in different ways in current textbooks

far only been achieved from bovine heart; inserted into an artificial membrane, this proved capable of proton translocation. Immunoprecipitation with antibodies against the whole of complex I has led to the isolation of about 25 different polypeptides. Using the same method, complexes from human, rat and rabbit liver have been isolated; these have a similar polypeptide spectrum to that of the bovine heart complex [28, 178].

Further pathways of electron transfer lead to coenzyme Q with the participation of flavoproteins but without the translocation of protons; these include the succinate:coenzyme Q oxidoreductase (complex II) and the electron-transferring flavoproteins of fatty acid β-oxidation, which are associated with the matrix side of the inner mitochondrial membrane, and glycerol-3-phosphate:coenzyme Q oxidoreductase, which lies at the outer side of the membrane and thus has access to substrates in the cytoplasm. **Succinate dehydrogenase**, which is usually assigned to the citric acid cycle, should not be considered as an independent enzyme as it is only active when isolated with artificial acceptors, such as ferricyanide or phenazine methosulphate (PMS). The natural functional unit that is competent for electron transfer to coenzyme Q is **complex II**, which consists of four different polypeptides (Table 18.3). Two of these make up succinate dehydrogenase, which is structurally similar in all prokaryotes and eukaryotes; these are a 70-kDa protein with an FAD and four Fe/S clusters and a 27-kDa protein with a further three Fe/S clusters. The two smaller proteins of complex II, with masses of 15.5 and 13.5 kDa, are encoded in the chromosome and form cytochrome b-560. This is not the same as the cytochrome b from complex III, which is encoded by mtDNA.

The NADH-specific **fumarate reductase** that is active in the succinate pathway of anaerobic metabolism includes not only a succinate:coenzyme Q oxidoreductase but also an NADH:coenzyme Q oxidoreductase and is, therefore, competent for proton translocation (see Fig. 14.11, p. 551). The complex II isolated from *Ascaris suum* contains cytochrome b-558; in the presence of the NADH:cytochrome c oxidoreductase (complexes I and III), cytochrome b-558 is reduced with NADH and reoxidized with fumarate. The relative activities of complexes II and III in the mitochondria of *Ascaris* are 20:1 and in rat liver mitochondria are 1:5. In *Ascaris*, the coenzyme Q involved in electron transfer between NADH and cytochrome b-558 is rhodoquinone (Fig. 18.7b), the more negative redox potential of which is responsible for the fact that fumarate is reduced and succinate is not oxidized. The fumarate reductases from other parasites, molluscs and oligochaetes and from mammalian heart also contain cytochrome b [82, 166].

Coenzyme Q of the mitochondrial respiratory chain introduces electrons into complex II, not only from the NADH of complex I but also from various other substrates. In vertebrates, the substrate is ubiquinone-10 (with ten isoprene units in the side-chain), although ubiquinones with shorter side-chains function equally well and are preferred in experimental studies because of their higher water solubility [118]. Ubiquinones with

shorter or longer side-chains are common in protozoans and invertebrate metazoans, e.g. ubiquinone-8 in *Plasmodium lophurae*, and ubiquinone-9 in various *Crithidia* and *Trypanosoma* species and the nematode *Metastrongylus elongatus*. Instead of ubiquinone, the tapeworm *Moniezia expansa*, the roundworm *Ascaris lumbricoides* and several other cestodes and nematodes have rhodoquinone (Fig. 18.7b), which is unknown in the mammals; both ubiquinone and rhodoquinone are found in the larvae of the trichina *Trichinella spiralis* and in various other larval and adult nematodes [9]. The specificity of complex I for NADH results in the indirect coupling of NADPH to the respiratory chain. The responsible enzyme is an **$NADP^+$:NADH transhydrogenase** which catalyses the following reaction:

$$NADP^+ + NADH \leftrightarrows NADPH + NAD^+. \quad (18.2)$$

This enzyme is located in the inner mitochondrial membrane and can translocate protons. The formation of NADPH requires energy which originates in the electrochemical proton gradient. The transhydrogenase isolated from bovine heart is a homodimer with subunits of 109 kDa. The subunits carry 14 transmembrane segments in their sequence of 1043 amino acids and are homologous to the enzyme from *E. coli*. The NAD-binding site at the N-terminus and the NADP-binding site at the C-terminus lie outside of the inner mitochondrial membrane in the mitochondrial matrix [180]. In some animals that utilize the succinate pathway of anaerobic metabolism, e.g. the rat tapeworm *Hymenolepis diminuta* and the annelids *Arenicola marina* and *Tubifex* sp., the mitochondrial malic enzyme (ME) produced mainly NADPH; in this case the transhydrogenase connects the ME with the NADH-specific fumarate reductase. However, ME is NAD specific in the roundworm *Ascaris lumbricoides* and the liver fluke *Fasciola hepatica*. The transhydrogenase present in the liver fluke probably functions in the production of NADPH as in the mammals; *Ascaris* has no transhydrogenase activity.

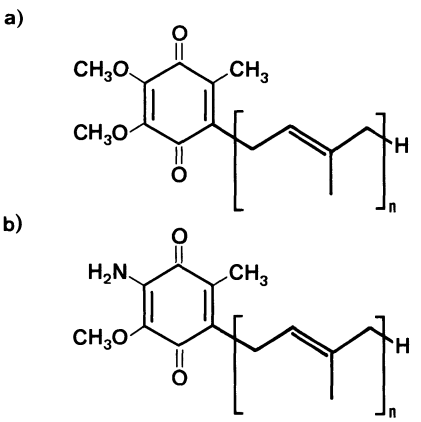


Fig. 18.7a, b. Coenzyme Q [9]. **a** Ubiquinone; **b** rhodoquinone; *n* gives the number of isoprene subunits in the side-chain

Complex III (coenzyme Q:cytochrome c oxidoreductase) from the bovine heart consists of ten subunits: two core proteins, cytochromes b and c_1, an iron-sulphur protein, a ubiquinone-binding protein (QP-C), a hinge protein, and three subunits with lower molecular masses. Similar patterns are found for the mitochondria of other mammals, birds and the tuna fish *Thunnus thynnus* [39, 164]. Various models that have been developed for electron transfer through complex III and the associated proton translocation assume electron transfer from reduced ubiquinone, on the one hand via the Fe/S protein and cytochrome c_1 to cytochrome c, and on the other hand via cytochrome b-566 to b-562. The **Rieske Fe/S protein** (named after its discoverer) carries a 2 Fe/2 S cluster near the C-terminus of its chain of 196 amino acids; it is responsible for electron transport between ubiquinone and cytochrome c_1 [148]. **Cytochrome c_1** consists of two polypeptides, the sequences of which are known in the case of bovine heart. A 27.9-kDa protein of 241 amino acids carries the haem group bound covalently to 37-Cys and 40-Cys; the smaller 9.2-kDa protein with 78 amino acids is remarkable in that positions 5–12 are occupied by eight consecutive glutamate residues. The latter protein is responsible for the interaction between cytochromes c and c_1. The two haem groups of cytochrome b have different spectral and functional properties, but sit on the same mtDNA-encoded polypeptide chain of 43 kDa with eight transmembrane domains. The cytochrome b forms from various vertebrates, *Drosophila, Trypanosoma*, higher plants, yeast and bacteria are remarkably homologous [63].

18.4.2 Cytochrome c

Cytochrome c is a small haem protein of only 13 kDa which functions in electron transport from complex III to cytochrome oxidase (complex IV). Because of its ubiquitous distribution, small size and simplicity of isolation this was the first protein for which systematic **sequence comparisons** were carried out and the construction of a molecular genealogical tree attempted (Table 18.4). The

Table 18.4. Species for which the complete sequence of cytochrome c is known [4, 68, 111]

18	Mammals
7	Birds
2	Reptiles *(Chelydra serpentina, Cortalus adamanteus)*
1	Amphibian *(Rana catesbeiana)*
2	Teleosts *(Thunnus thynnus, Katsuwonus wagrans)*
1	Elasmobranch *(Squalus sucklii)*
1	Agnathan *(Entosphenus tridentatus)*
1	Echinoderm *(Asterias rubens)*
1	Mollusc *(Helix aspersa)*
1	Crustacean *(Macrobrachium malcomsonii)*
8	Insects
1	Annelid *(Eisenia foetida)*
1	Ciliate *(Tetrahymena pyriformis)*
2	Flagellates *(Euglena gracilis, Crithidia oncopelti)*
26	Plants
8	Fungi
6	Prokaryotes

polypeptide chain of cytochrome c consists of 103–104 amino acids in the vertebrates, 104–107 amino acids in insects, 107–109 amino acids in yeast and lower fungi, and 111–112 amino acids in higher plants. Among all the sequences known 36 positions are invariant and the majority of amino acid substitutions are conservative (Fig. 18.8). Amino acids 14-Cys and 17-Cys are bound to side-chains 2 and 4 of the haem via covalent thioether bonds; interaction of the haem iron with 80-Met and 18-His prevents the molecule reacting with O_2, CO or other ligands. The tertiary structures of all cytochrome-c species known are the same, and show the same high degree of conformational change in the transition from the reduced to the oxidized state. Cytochrome c is a very conserved protein in evolutionary terms; the cytochromes from yeast and horse differ by only 48 amino acids, and complete sequence agreement is found between the chicken and turkey, between the pig, cow and sheep, and between humans and the chimpanzee. The low **rate of evolution** of cytochrome c (see Table 4.12, p. 161) is due to the fact that it must interact functionally with two different multienzyme complexes of the respiratory chain. Closer analysis of the functional restrictions has been carried out with natural and synthetic cytochrome-c fragments, but the details of this cannot be presented here [129].

In the reduced state, cytochrome c of the trypanosomes has an absorption maximum of 556–558 nm compared with the normal 550–552 nm. This is apparently due to loss of one of the two thioether bonds between the haem and the apoprotein [44]. In contrast to the cytochromes from the Metazoa, that of the insect-parasitic flagellate *Crithidia oncopelti* is highly methylated; this begins at the N-terminus with a dimethylproline and includes trimethyllysine at positions 8 and 72 (numbered as for vertebrate cytochrome). A cytochrome-c methyltransferase has been isolated from this unicellular species [172]. The cytochrome c of the American bullfrog *Rana*

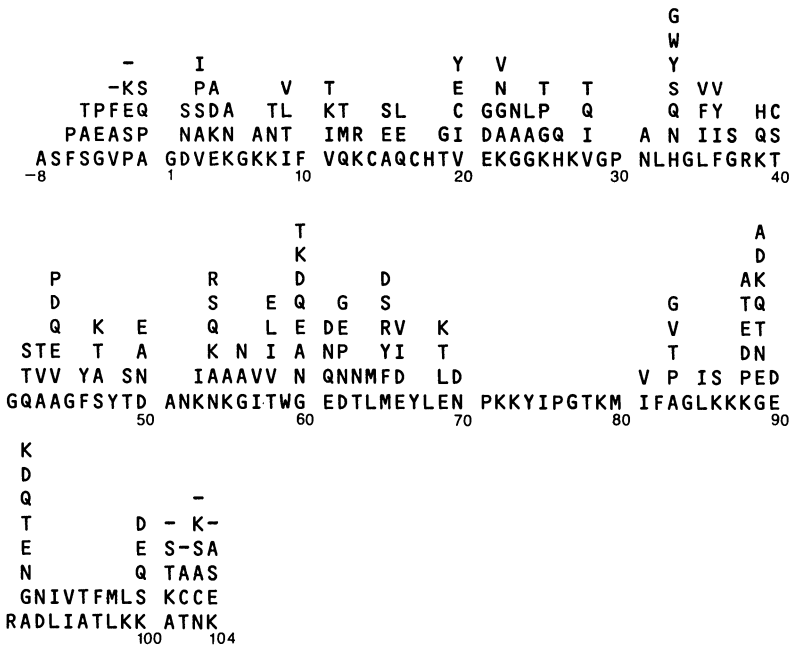


Fig. 18.8. Cytochrome c [96]. A summary of the amino acids occupying the individual positions of 32 sequences. The numbering is related to the cytochrome c of mammals

catesbeiana is unique in that it contains two further cysteines (20-Cys and 102-Cys), in addition to the two that are bound to the haem; these form an intramolecular disulphide bridge but do not significantly alter the tertiary structure [17].

Despite the conservative nature of this protein, the evolution of cytochrome c has included several gene duplications which led to remarkably different **isoforms**. For example, in the housefly *Musca domestica* the larvae contain 90 % of a larval cytochrome c and the flies contain more than 70 % of an adult cytochrome c which differs in 6 out of the 107 amino acids. Compared with other insect cytochromes, the larval isocytochrome has some original properties; 58-Ser is found only in locusts, and 96-Ala occurs in locusts and lepidopterans but not in the adult isoprotein of *Musca*. Furthermore, the larval isocytochrome of the housefly predominates in all tissues of the adult except flight and leg muscles; thus, the adult isoform is a specific muscle cytochrome c [68]. Two genes (DC3 and DC4) coding for different cytochrome-c forms have been isolated from *Drosophila melanogaster*. The sequence of 107 amino acids derived from DC4 corresponds, with a few corrections, to the original directly obtained sequence and is very similar to other cytochromes. DC3 codes for a cytochrome of only 104 amino acids which differs in no fewer than 32 positions from the DC4 protein. The DC3-encoded protein also differs at 13 positions from the consensus sequence of the eukaryotes, including the otherwise invariant position 70. The two genes are expressed in a tissue- and stage-specific manner but are otherwise not comparable with the *Musca* isocytochromes. They are separated by only 3.5 kb and are thus apparently the result of a gene duplication. If one assumes that 20 million years are required for the evolution of a 1 % difference (UEP, unit evolutionary period), this duplication occurred 600 million years ago [96]. The third known case of isocytochromes is that of the mouse, in which there is a sperm cell-specific form which differs in 13 of the 104 amino acids from the common cytochrome c found in all cells. A testis-specific cytochrome c is also present in the rat and shows four differences from the mouse protein; in contrast, the somatic proteins of these two species are identical [175].

The conservatism of cytochrome c allows the gene from one organism to be used as a probe in another species, and in this way the **cytochrome-c genes** have been isolated from, for example, two yeasts, *Drosophila melanogaster*, the chicken, the rat and the mouse. All non-mammals have only 1 or 2 genes, whereas mammals have 20–30, most of which are pseudogenes. There are two introns in the functional genes of the rat and mouse, one in codon 56 and the other in the 5′ non-coding region. The chicken gene has only the 5′ intron and the *Drosophila* and yeast genes have no introns. In contrast to the active genes, three pseudogenes isolated from the mouse also contain no introns and thus arose via the mRNA [95, 96]. All the mouse pseudogenes are duplicates of the present active genes. The human pseudogenes, however, fall into two groups, one of which is derived from an ancestral form of the cytochrome-c gene [40].

18.4.3 Cytochrome Oxidase (Complex IV)

The ferrocytochrome c:O_2 oxidoreductase (cytochrome oxidase) of the mitochondria has been isolated from and characterized for various vertebrates, e.g. mammals and birds [15, 18, 19, 38, 54, 110, 118, 167], amphibians, and fish [110]. There appear to have been no recent investigations of the structure of complex IV in invertebrates. In the cytochrome oxidases isolated from mammals, 13 polypeptides can be distinguished, of which the 3 largest (I–III) are encoded by mtDNA. The complete sequences of all 13 polypeptides from bovine heart are now known (Table 18.5). In most cases, sequencing is

Table 18.5. The polypeptides of the cytochrome oxidases from bovine heart (VIIb is from porcine heart) [24, 38, 94, 162]

Nomenclature according to Kadenbach (after Buse)		Size (Da)	Site of synthesis	N-terminal sequence
I	(I)	56 993	mito	f-Met-Phe-Ile-Asn-
II	(II)	26 049	mito	f-Met-Ala-Tyr-Pro-
III	(III)	29 918	mito	(Met)-Thr-His-Gln-
IV	(IV)	17 153	cyto	Ala-His-Gly-Ser-Val-
Va	(V)	12 436	cyto	Ser-His-Gly-Ser-His-
Vb	(VIa)	10 670	cyto	Ala-Ser-Gly-Gly-Gly-
VIa	(VIb)	9 419	cyto	Ala-Ser-Ala-Ala-Lys-
VIb	(VII)	10 068	cyto	Ac-Ala-Glu-Asp-Ile-
VIc	(VIc)	8 490	cyto	Ser-Thr-Ala-Leu-
VIIa	(VIIIc)	6 244	cyto	Phe-Glu-Asn-Arg-Val-
VIIb	(–)	6 350	cyto	Ile-His-Gln-Lys-Arg-
VIIc	(VIIIa)	5 541	cyto	Ser-His-Tyr-Glu-Glu-
VIII	(VIIIb)	4 962	cyto	Ile-Thr-Ala-Lys-Pro-

f-Met, Formylmethionine; Ac-Ala, N-acetylalanine

complicated by the existence of multiple genes and pseudogenes. The **redox centres** consist of two haem groups (a, a_3) and two copper atoms, which are bound in an as yet unknown manner to the subunits I and II and have different molecular environments. Cytochrome oxidase catalyses electron transfer from the one-electron donor ferrocytochrome c to the four-electron acceptor O_2. In the process, electrons are probably taken over from cytochrome c by haem a and the associated Cu-A, and are transferred to O_2 stepwise via haem a_3/Cu-B, the incorporation of four protons giving rise to two molecules of water [19, 24].

All cytochrome oxidases contain the **three subunits I–III encoded by mtDNA**, of which I and II carry the redox centres. Subunit III has no bound metal ion but is nevertheless as indispensable to the proton-translocating complex as are subunits I and II [18, 19]. Homologues of subunits I–III are already found in prokaryotes. In addition, the cytochrome oxidases of the eukaryotes all contain several **chromosome-encoded subunits** of unknown function; the slime mould *Dictyostelium discoideum* has five such subunits, yeast has six, fish and birds have seven to eight, and mammals have ten [146]. Nine polypeptides of the yeast complex are all homologous to the corresponding proteins of bovine heart; only polypeptides VIa, VIb and VIc of mammals (nomenclature of Kadenbach) have no homologous partners in yeast. The 38–56 % sequence agreement between yeast and bovine mtDNA-encoded polypeptides is significantly higher than that between their chromosome-encoded polypeptides (13–30 %). The mtDNA-encoded subunit III of the carp (261 amino acids) agrees by 80–81 % with that of various mammals, 90 % with that of *Xenopus laevis*, and 66 % with that of *Drosophila* [64]. The chromosome-encoded subunits of the mouse and rat consistently show more than 90 % sequence agreement, and the agreement between representatives of various mammal orders is about 80 %. Tissue-specific isoforms are known for most of the chromosome-encoded subunits, e.g. in cattle, pigs or rats for IV, VIa, VIb, VIc, VIIa, VIIb, VIIc and VIII; in contrast, no isoforms have been described for the mtDNA-encoded subunits I–III [146, 183].

Depending upon their origin, the cytochrome oxidases that have been isolated from mammals are **monomers** of 5–7S, **dimers** of 10–11S or **higher aggregates** of more than 17S. The form of the enzyme that lies in the mitochondrial membrane is not yet clear. Preparations from bovine heart contain about 90 % dimers and 10 % higher aggregates, but in alkaline conditions these dissociate to monomers without loss of activity. Preparations of cardiac mitochondria from the camel and the hammerhead shark *Sphyrna lewini* consist almost exclusively of monomers, from the spiny dogfish *Squalus acanthias* and the chicken they contain about equal parts of monomers and dimers, and from the rat they have 85 % monomers and 15 % higher aggregates [15].

18.4.4 ATP Synthase (Complex V)

The ATP synthase of the inner mitochondrial membrane is a reversible, proton-translocating ATPase. In contrast to other ion-transporting ATPases, such as the Na^+, K^+-ATPase of the cell membrane or the Ca^{2+}-ATPase from the sarcoplasmic reticulum, ATP synthase has a completely different and much more complicated structure. Electron microscope pictures of the energy-converting membranes of mitochondria, chloroplasts or bacterial cells show the ATP synthase to be a sphere with a diameter of 9 nm and bound to the membrane by a 3-nm-thick and 4-nm-long stem (Fig. 18.5). The catalytic subunit projecting above the membrane surface is known as F_1 and the membrane portion is known as F_0. The F_1 of all ATP synthases consists of five types of subunit in the stoichiometric ratio $\alpha_3\beta_3\gamma\delta\varepsilon$. The F_0 of *E. coli* is made up of three types of subunit according to the formula ab_2c_{10}. F_0 preparations from bovine heart contain eight to ten polypeptides of 6–24 kDa, of which subunits 6 and 9 (DCCD-binding proteolipid) are homologous to subunits a and c from *E. coli*, and the others are unique to eukaryote F_0: these are oligomycin sensitivity-conferring protein (OSCP), coupling factor 6, coupling factor B, subunit 8, and two proteins of 20 and 24 kDa that are not found in all preparations. The eukaryote polypeptide corresponding to subunit b of *E. coli* has not yet been defined. Subunits 6 and 8 of the mitochondrial ATPase are mtDNA encoded (p. 53). Although many of the eukaryotic ATPase subunits have been sequenced, their stoichiometry, function and, in particular, their organization in the membrane remain undefined [55, 76]. The F_1 subunits α and β not only are clearly homologous in cattle and *E. coli* but also show certain, albeit lower, agreement with each other and with other ATP-binding proteins such as myosin, the Ca^{2+}-ATPase and the adenine nucleotide transporter.

The **efficiency of energy coupling** of electron transport to the ATP synthase derives from the

ratio $H^+/2e^-$ for the electron transport system and H^+/ATP for ATP synthase,

$$ATP/2e^- = (H^+/2e^- / (H^+/ATP). \quad (18.3)$$

The determination of $H^+/2e^-$ takes into account the loss of proteins through "leaks" in the membrane, and the value of H^+/ATP is influenced by the energy required for the transport of ADP and ATP through the mitochondrial membrane. The theoretical ratios $ATP/2e^-$ correspond to the experimental P/O ratios, which lie between 2 and 3 for NADH and between 1.5 and 2 for reduced flavin.

18.4.5 Heat Production in Brown Fat Tissue

The brown fat tissue found in some mammals is much richer in mitochondria than is the more common white fat tissue. The single function of brown fat tissue is the production of heat by fatty acid β-oxidation. The maximal heat production of brown fat cells is about 300 W/kg and is thus manyfold higher than the average heat production of a few W/kg by the whole animal. Brown fat tissue is found especially in newborn animals and infant mammals, which lack heat insulation and cannot otherwise maintain their body temperature, in hibernating species, where it is required during reawakening, and in cold-adapted animals. Activation of the brown fat occurs via the sympathetic nervous system, and experimentally also by noradrenalin; it is accompanied by a 20- to 40-fold increase in oxygen consumption. However, ATP production is negligible, and the efficiency of the mitochondrial ATP synthase is essentially zero. This is due to the activity of a specific protein, the uncoupling protein (UCP), which is found only in brown fat tissue and makes up about 10–15 % of the total protein of the inner mitochondrial membrane. The active UCP functions as an H^+ channel, reduces the electrochemical proton gradient of the inner mitochondrial membrane, and thus uncouples electron transport and ATP production.

The UCP is inhibited by GTP and GDP, and also by ATP and ADP. Free fatty acids reduce its affinity for ATP and thereby allow UCP activity even at physiological ATP concentrations. The regulation of heat production in brown fat tissue is not yet totally understood. Adrenalin brings about a reduction in the electrochemical proton potential of isolated mitochondria, and the UCP is directly involved in the effect of adrenalin [58]. The UCPs belong to the same protein superfamily as the ATP/ADP transport protein. Their sequences of about 300 amino acids are divided into three homologous domains of 100 amino acids and each has two transmembrane segments. In the UCP gene of the mouse, each of the six exons corresponds to a transmembrane segment. This fits the hypothesis that they originated by a gene triplication. In contrast to most other mitochondrial proteins, the mature protein, except for the N-terminal methionine, corresponds to the primary translation product, and the usual N-terminal translocation signal is absent [83, 86, 138]. The UCPs are evolving relatively slowly; the rat protein (306 amino acids) agrees 91 % with that of the golden hamster and 86 % with that of the rabbit [5].

18.4.6 Alternative Pathways of Mitochondrial Electron Transport

The terminal oxidase of some protozoans, nematodes, trematodes and cestodes is a CO-binding, auto-oxidizing cytochrome of the b type; it is cyanide resistant, preferentially produces H_2O_2 and is usually referred to as cytochrome o. In the mitochondria of the roundworm *Ascaris lumbricoides*, the electron transport system (ETS) branches at coenzyme Q, in this case rhodoquinone. About 20–30 % of the electrons travel the pathway leading to cytochrome aa_3, corresponding to the usual respiratory chain, whereas the remaining 70 % pass to the alternative cytochrome. Oxidative phosphorylation takes place at the usual steps in the pathway to cytochrome aa_3, but there is no indication of ATP formation by the alternative pathway. This type of branched respiratory chain is also found in other parasitic nematodes, such as *Nippostrongylus brasiliensis* and *Ascaridia galli* [123], free-living nematodes, such as *Aphelenchus avenae* [108], trematodes, such as *Fasciola hepatica*, and cestodes, such as *Moniezia expansa* and *Hymenolepis diminuta* [106]. One can imagine that branched ETSs are an adaptation to the marked fluctuations in the oxygen partial pressure experienced by these species in their habitats. There are, however, parasitic worms whose only terminal oxidase is cytochrome aa_3, for example adults of *Schistosoma mansoni* and *Metastrongylus elongatus* [9].

Many animals from sulphide-rich marine habitats contain haematins which catalyse the oxidation of the sulphides and protect the animal from sulphide poisoning. The haematin is located in the coelom fluid of *Urechis caupo*, and together

with haemoglobin in the gills of the clams *Solemya reidi, Calyptogena magnifica* and *Lucinoma annulata. S. reidi* has no gut and utilizes sulphide as an energy source; thiosulphate formed by oxidation of sulphide in the tissues of the mussel is converted by symbiontic bacteria in the gills to compounds that can serve as substrates. In addition, sulphide can be used in the mitochondria of the gills and the symbiont-free foot muscle as a substrate for ATP production by oxidative phosphorylation. This is the first evidence for the direct exploitation of an inorganic energy source by animal cells. Oxidative phosphorylation in mammalian mitochondria would be completely inhibited by the sulphide concentrations of up to 20 µmol/l that are commonly found in mussel tissues [131].

Cytochrome aa_3 is the predominant terminal oxidase in the soil amoeba *Acanthamoeba castellanii* [153], whereas the so-called cytochrome o is found in several parasitic protozoans. In the culture form (epimastigote) of *Trypanosoma rhodesiense* it is the only terminal oxidase; others, such as *Crithidia fasciculatea* and the epimastigotes of *Trypanosoma brucei*, also possess cytochrome aa_3; in other words, they have a branched respiratory chain [171]. In fact, three pathways are found in *T. cruzi;* one is inhibited by CN^- and N_3^-, the second only by CN^-, and the third by neither inhibitor [3]. Cyanide-resistant respiration can therefore apparently stem from terminal oxidases other than cytochrome o, as has already been proposed in the case of the ciliate *Paramecium aurelia* [36]. In addition to cytochrome o, *Tetrahymena* has a cytochrome a-620 which clearly differs both spectroscopically and functionally from cytochrome aa_3; it cannot oxidize mammalian cytochrome c but can oxidize the cytochrome c-553 found in this unicellular animal; this latter cytochrome has unusual properties as an electron donor [97]. The cytochrome a-620 of *Tetrahymena* has been referred to as cytochrome d but should not be confused with bacterial cytochrome d. Some protozoans, e.g. the human parasites *Entamoeba histolytica, Giardia lamblia* and *Trichomonas vaginalis*, and the symbiotic *Dasytricha ruminantium* of the ruminant stomach, have neither cytochrome nor mitochondria. Trichomonads and rumen ciliates, however, have H_2-producing hydrogenosomes, and *Dasytricha* has a cyanide- and azide-resistant NADH oxidase activity, with an unknown terminal oxidase, which is possibly involved in oxygen detoxification [181]. In *Entamoeba*, and probably also *Giardia*, there is an ETS leading from various substrates via flavin, non-haem iron (ferredoxin) and coenzyme Q to an as yet unidentified terminal oxidase. The biological significance of this ETS is presently a mystery; it is not necessary for protection from oxygen, which for these protozoans is not toxic [177].

18.5 Microsomal Electron Transport Systems

Electron transport systems are also present in the membrane of the endoplasmic reticulum and thus are found in the microsome fraction after homogenization and fractionating centrifugation. The microsomal flavin-containing monooxygenases (FMOs), which in humans and various mammals are found in at least three tissue-specific isoforms, catalyse the NADPH-dependent oxidation of numerous foreign substances [35]. A further type of microsomal oxidoreductase contains the characteristic components cytochrome b_5 and NADH cytochrome b_5 reductase. It is involved in various reactions of lipid metabolism, for example the desaturation and elongation of fatty acids and the biosynthesis of steroids and plasmalogens. A third type is represented by the NADPH-dependent monooxygenases, which contain cytochrome P-450 and NADPH-cytochrome P-450 reductase and can introduce oxygen into organic substrates, in particular catalysing the hydroxylation of steroids and foreign substances. Several of the steroid-specific enzymes are localized in the mitochondria.

18.5.1 Cytochrome b_5

Probably the best-investigated cytochrome-b_5 enzyme system in the microsomes of mammalian liver is the stearoyl-CoA Δ^9-desaturase, a complex of NADH:cytochrome b_5 reductase, cytochrome b_5 and a cyanide-sensitive terminal oxidase. Cytochrome b_5 from the liver of different mammals is a monomer of about 16 kDa. The polypeptide of 133 amino acids consists of an N-terminal hydrophilic domain, carrying a protohaem as the prosthetic group, and a C-terminal hydrophobic domain, which anchors the protein in the membrane. A cDNA from chicken liver codes for a protein of 138 amino acids; this is 6 amino acids longer at the N-terminus and 1 amino acid shorter at the C-terminus [182]. The NADH:cytochrome b_5 reductase with FAD as

the prosthetic group has a similar structure. Because of their amphiphilic character, these proteins tend to aggregate in solution.

Enzyme systems with a similar structure are probably quite widespread, but in the invertebrates have been examined in detail in only a few species. All known cytochromes of the b_5 type carry protohaem as the prosthetic group but apparently differ greatly in chain length. The microsomal cytochrome b_5 from the larvae of the Mediterranean fruit fly *Ceratitis capitata* has a native molecular mass of about 125 kDa and appears to be a hexamer of 21-kDa subunits; with 194–196 amino acids it is clearly longer than the mammalian enzyme. A reductase isolated from the same species has subunits of 31 kDa and reduces ferricyanide or dichlorinphenol with NADH but not with NADPH; cytochrome b_5 with NADH also reduces cytochrome c [107]. The cytochrome b_5 from the microsomes of the moth *Spodoptera eridania* and the (probably) microsomal cytochrome b-555 from the larvae of the housefly *Musca domestica* are less well characterized [107]. An NADH-specific reductase of 55 kDa and a cytochrome b_5 of about 100 kDa have been isolated from the muscle microsomes of the roundworm *Ascaris lumbricoides* where they probably form an aggregate. The ciliate *Tetrahymena pyriformis* contains a system comprising a cytochrome b-560, similar to mammalian cytochrome b_5, and an NADH:cytochrome b_5 reductase which, like the corresponding system in rat or chicken liver, is involved in fatty acid desaturation; mixed systems with *Tetrahymena* cytochrome and rat reductase are functional [85]. Cytochrome b-560 differs from mammalian cytochrome b_5 not only in its absorption spectrum but also in its larger molecular mass of 22 kDa. The 70-kDa NADPH:cytochrome c reductase present in *Tetrahymena* microsomes, together with the NADH : ferricyanide reductase, is able to reduce both mammalian cytochrome b_5 and *Tetrahymena* cytochrome b-560 with NADPH [49].

18.5.2 Cytochrome P-450

The cytochrome P-450 discovered in 1958 in the endoplasmic reticulum of cells of the mammalian liver is an atypical b type with protohaem as the prosthetic group. The name indicates that the absorption band (Soret band) of the CO complex in the short-wave length range is shifted by 30 nm from the usual 420 nm of the haem proteins. This is due to the binding of cysteine to the haem iron as the axial ligand. Cytochromes of this type catalyse numerous monooxygenation reactions, in which one oxygen atom from O_2 is incorporated in the substrate and the other is incorporated into H_2O. These reactions require NADPH, although NADH has a sparing effect or is synergistic. Electron transport from NADPH to cytochrome P-450 involves the action of an NADPH:cytochrome P-450 reductase; cytochrome b_5 can also link cytochrome P-450 with NADH via NADH:-cytochrome b_5 reductase. Various cytochrome P-450 monooxygenases are constitutive enzymes which serve especially in steroid metabolism; however, cytochrome P-450 enzymes can also be induced to a unique extent by foreign substances, and they have been intensively researched because of the importance of cytochrome-P-450-catalysed reactions in the metabolism of pharmaceuticals and other foreign substances.

Several hundred substances are known to induce one or more different types of cytochrome P-450. The majority of mammalian cytochrome P-450 types fit into one of three **classes:** those of the first group are induced by phenobarbital (PB); the second group are induced by 3-methylcholanthrene (3MC) or, more effectively, by 2,3,7,8-tetrachlorodibenzo-*p*-dioxin (TCDD); and the third group are induced by steroids such as pregnenolone-16α-carbonitrile (PCN) [66]. There is also an ethanol-inducible cytochrome P-450, the sequences of which in the rat and humans show only 48 % agreement with the PB-induced type [160]. Members of these classes also differ in the position of the Soret band; hence, those induced by 3MC and TCDD are known more specifically as **cytochrome P-448** to distinguish them from the pB-induced cytochrome P-450. The best-known example of the 3MC-induced enzymes is aryl hydrocarbon hydroxylase, which metabolizes, for example, benzo(a)pyrene. 3MC induction involved a **cytosolic receptor** that binds the inducer and transports it to the nucleus. The intracellular receptor of the PB-induced P-450 cytochrome has not been identified. Regulation of cytochrome P-450 activity occurs not only by the activation of transcription but also by various combinations of post-transcriptional and post-translational mechanisms. In addition to the inducible enzymes that are mainly involved in the metabolism of foreign substances, there are also constitutive enzymes, which take part in the metabolism of endogenous compounds, such as steroids, bile acids, fatty acids, prostaglandins, leucotrienes and biogenic amines, and may be sex specific [117].

All types of cytochrome P-450 are polypeptides with about 500 amino acids, and the haem group is bound to a cysteine-rich sequence of 22 amino acids near the C-terminus. Up to the end of October 1990, the **sequences** of 154 genes and 6 pseudogenes from the protein super-family of cytochrome P-450 have been determined; these came from 11 vertebrate species (9 mammals, the chicken and the trout), 3 invertebrate species (2 insects and the pond snail), 1 higher plant (avocado), 8 lower fungi and 6 prokaryotes. Genes with more than 40 % sequence agreement have been placed in one family; up to now, 27 such families have been described, 10 of which are found amongst the vertebrates. Some of the vertebrate families can be divided into subfamilies with at least 55 % sequence agreement. This prolificacy of cytochrome P-450 types brings with it certain problems of nomenclature. In 1987, Nebert proposed that the gene families be given Roman numerals, the subfamilies capital letters and the individual genes Arabic numerals: a certain gene or the encoded protein would thus be known as "P450IIIE1" or simply as "IIE1". The unified nomenclature proposed in 1991 makes use of the abbreviation "CYP" for the super-family and suggests the more readable Arabic numerals for designating families; hence, the gene referred to above and its cDNA in the mouse become *"CYP2E1"* and *"CYP2e1"*, and the mRNA or the protein in all species becomes "CYP2E1" or "P4502E1" or simply "2E1". Numerous allelic variants of human cytochrome P-450 and those of laboratory rodents are known [117].

A genealogical tree, generated by the unweighted pair group method (UPGM) (p. 157), shows more than 30 gene duplications in the evolution of the CYP2 family alone. In some cases, the time point of the gene duplications can be estimated. Thus, in the mammals two genes, 1A1 and 1A2, are recognized but in the trout there is only one 1A1 gene; 1A2 apparently arose first with the evolution of the terrestrial vertebrates [117]. Duplications of CYP genes have also occurred more recently; for example, rabbits in the subfamily CYP2E have two genes, but humans and the rat have only one. Species-specific gene duplications or gene conversions in the subfamilies 2A, 2B, 2C, 2D and 4A make it difficult to determine orthologous relationships between genes of different species [103, 117]. Examination of the genealogical tree reveals that a period of 3.3 to 5.0 million years (UEP) is required for the occurrence of a 1 % sequence difference, and thus cytochrome P-450 belongs to the proteins with a relatively high rate of evolution (see Table 4.12, p. 161). The functions and inducibilities of the different types of cytochrome P-450 are not a reliable guide to their relationships. The substitution by genetic engineering of just one of the 500 amino acids can drastically alter the substrate specificity. Not all genes of a family or subfamily are inducible in the same way and, conversely, enzymes coinduced by the same substance in the same animal may differ considerably in sequence.

Monooxygenase systems of NADPH:cytochrome P-450 reductase and cytochrome P-450 have been found in the liver microsomes of vertebrates from all classes, but have been characterized in only a few cases [119, 139]. Among the lower vertebrates, the **fish** have been examined the most comprehensively, with the major interest being ecotoxicology. In 1967, the first evidence was obtained for the existence of cytochrome P-450 in the liver of the rainbow trout *Salmo gairdneri*, and in 1976 a benzo(a)pyrene hydroxylase was reconstructed from cytochrome P-450, NADPH:cytochrome P-450 reductase and phospholipid from the ray *Raja erinacea*. Several different types of cytochrome P-450 have been isolated from these two species and from the marine teleost *Stenotomus chrysops;* attempts to derive homologies between these and certain mammalian enzymes remain questionable in the absence of sequence analysis [84, 109]. Monooxygenase systems with cytochrome P-450 have been detected in many **insects**, but only a few sequences are known as yet. The cDNA of a cytochrome P-450 was isolated from PB-treated houseflies *(Musca domestica)* and shows 27 % agreement in its sequence of 509 amino acids with mammalian proteins of the family CYP3, but is considered to represent a new family, CYP6. The sequence is known of a cytochrome P-450 from the cockroach *Blaberus discoidalis;* this cytochrome is induced by the hypertrehalosaemic hormone. It shows 32–36 % sequence agreement with the family CUP4, which is involved in the metabolism of fatty acids and prostaglandins, and is known as CYP4C1 [16, 47]. Most of the insect cytochrome P-450 types are comparable with the mammalian enzymes only in their inducibility or inhibition characteristics. For example, in mammals the P-448 enzyme aryl hydrocarbon hydroxylase is induced by β-naphthoflavone (βNF) and TCDD, but in the moth *Spodoptera eridania* it is induced by βNF but not by TCDD or 3MC. In contrast to the rat enzyme, that of *Spodoptera* is induced by pentamethylbenzol and naphthaline. The binding

and effect of inhibitors such as α-naphthoflavone and phenylimidazole also do not allow assignment of the insect enzymes to one of the mammalian classes P-448 or P-450 [26].

Multiply inducible P-450 monooxygenases in insects are mostly found in the midgut but also occur in the fat body, integument and other organs [30, 46, 61, 139]. They take part in many reactions involving endogenous and foreign substances, e.g. juvenile hormone epoxidation in the corpora allata, 20-hydroxylation of ecdysone in the fat body mitochondria of *Manduca sexta* and in the microsomes of the fat body and Malpighian tubules of *Locusta migratoria*, the inactivation of ecdysteroids, the biosynthesis of some pheromones such as disparlure, and the metabolism of insecticides. They are apparently also important for resistance to plant secondary metabolites [29, 163]. Amongst 35 lepidopteran species, the highest cytochrome P-450 activities (measured as aldrin epoxidation) were found in the midgut of the 12 polyphagous species (mean 294 pmol/mg protein min^{-1}); lower activities (mean 91 pmol/mg protein min^{-1}) were recorded in 15 oligophagous species, and the lowest activities (mean 20 pmol/mg protein min^{-1}) were in monophagous species. Plant extracts can induce up to 45-fold increases in the total cytochrome P-450 activity or individual activities in insects [61]. Enzyme systems with cytochrome P-450 are apparently ubiquitous from the bacteria and lower eukaryotes to the higher plants and animals [139]. Apart from the data for insects, there are unfortunately only sporadic data for other invertebrates, e.g. for the protozoans *Tetrahymena pyriformis* and *Trypanosoma cruzi* [88], molluscs [151] and crustaceans [69, 71]. In view of the otherwise ubiquitous distribution of this enzyme, it is surprising that two nematodes, the parasitic *Heligmosomoides polygyrus* and the free-living *Panagrellus redivivus*, appear to possess neither cytochrome P-450 nor cytochrome b_5 [133].

The **NADPH:cytochrome P-450 reductases** also reduce cytochrome c; partial proteolysis of the human enzyme produces a shortened form that can metabolize cytochrome c but can no longer metabolize cytochrome P-450. Comparative investigations are often carried out on preparations that still have activity against dichlorophenolindophenol or cytochrome c but not against cytochrome P-450. However, intact enzymes have been isolated from the flies *Musca domestica* and *Phormia regina* and these allow the reconstitution of a monooxygenase system. Gel electrophoresis indicates that they have a lower molecular mass (74 kDa) than the mammalian enzyme (82–83 kDa). Comparison of the amino acid composition gives a difference (ΔSn) of 50 between these fly species, i.e. they are closely related (p. 150), but a difference of 500 with the mammalian enzyme; this suggests only limited sequence similarity [174]. Immunological comparisons also present evidence for a relatively rapid evolution of NADPH:cytochrome P-450 reductase. Antibodies against the enzyme from the moth *Spodoptera eridania* inhibit the enzymes of other Lepidoptera, Diptera and Orthoptera but not those of other insects, the earthworm or various vertebrates.

18.6 Oxygen-Detoxifying Enzymes

Complete reduction of a molecule of oxygen requires four electrons. During stepwise electron transfer, intermediates may arise that, because of their extreme reactivity, are extremely toxic:

$$O_2 \xrightarrow{e^-} O_2^- \xrightarrow{e^- + 2H^+} H_2O_2 \xrightarrow{e^- + H^+} \begin{matrix} OH^\cdot \xrightarrow{e^- + H^+} H_2O. \\ \searrow H_2O \end{matrix} \qquad (18.4)$$

Cytochrome oxidase, which is responsible for most of the oxygen reduction that takes place in the cell, does not release such intermediates. However, they arise by the reactions of many other oxidases and must be detoxified by the enzymes superoxide dismutase (for the superoxide ion), catalase and peroxidase (for the hydrogen peroxide). There is no specific enzyme for the hydroxyl radical; the formation of this extremely reactive radical is apparently effectively prevented by the high catalytic efficiency of the enzymes for the preceding intermediates.

18.6.1 Superoxide Dismutases

The superoxide dismutases (SOD) catalyse the reaction

$$2O_2^- + 2H^+ \rightarrow H_2O_2 + O_2. \qquad (18.5)$$

These are metalloenzymes which can be divided into three groups according to their structure and distribution. The Cu,Zn-SODs are homodimers of 16-kDa subunits with one Cu and one Zn; they are found in the cytoplasm of almost all eukaryotic cells, and are also present in bacteria. The Mn-SODs are homotetramers of 80 kDa and are typical of many prokaryotes and the mitochond-

ria of eukaryotes. Fe-SODs are present in many bacteria, and have also been detected in the eukaryotic alga *Euglena gracilis*, the higher plants *Brassica campestris, Gingko biloba* and *Nuphar luteum*, several Kinetoplastida and the ciliate *Tetrahymena pyriformis*. They are mostly homodimers of 40 kDa, but homotetramers of 80 kDa are found in some bacteria and *Tetrahymena* [8, 37, 169]. Fe-SOD and Mn-SOD are closely related; the sequence of human mitochondrial Mn-SOD agrees 43–50 % with bacterial Mn-SODs and up to 37–42 % with bacterial Fe-SODs [125]. The three SOD types are easily distinguished by their response to inhibitors. Cu,Zn-SOD is inhibited by cyanide and peroxide, Fe-SOD is inhibited by azide and peroxide, and the Mn-SOD is not inhibited by any of these three inhibitors [169].

The **amino acid sequences** are known for the human Cu,Zn-SOD and for those of various other mammals [136, 150, 161], the frog *Xenopus laevis* [25], the swordfish *Xiphias gladius* [141], various *Drosophila* species [89, 90, 91, 154], the yeast *Saccharomyces cerevisiae*, and the bacteria *Photobacter leiognathi* and *Brucella abortus* [11]. All Cu,Zn-SODs have lengths in the range 150–153 amino acids (Fig. 18.9). The eukaryotic enzymes are evolutionally very conservative. The sequence agreement of Cu,Zn-SOD from humans, cattle, horses, *Drosophila* and yeast is 39 %, between the four animals it is 50 % and between the three mammals 74 % [92]. All of them contain a disulphide bridge at approximately the same position (in Fig. 18.9 between positions 58 and 160), six metal-binding histidines (positions 47, 49, 76, 85, 94 and 134), the aspartate that is also involved in metal binding (position 90), and the arginine of the active centre (position 157). Of the 25 glycine residues in the SOD of cattle, 17 are invariant [92]. The SOD of the frog *Rana catesbeiana* also agrees in 13 out of 24 N-terminal amino acids with the cattle enzyme; there is no difference between the enzymes from the liver and erythrocytes, or between those from the tadpole and the adult frog [1]. Acetylation of the N-terminus is found only in the mammals [1, 52]. The catalytic activity of Cu,Zn-SOD approaches the limit set by substrate diffusion [149]. As the reaction of SOD involves interaction of the superoxide anion with positive charges at the active centre, in particular with the arginine residue mentioned above, the higher number of basic amino acids surrounding the active centre of the sheep enzyme, compared with the number in the enzyme of cattle, results in a reduction in specific activity [150]. Electrophoresis of Cu,Zn-SOD normally produces three bands which arise by partial deamidation of asparagine or glutamine residues [52]. Only in the case of the frog *Xenopus laevis* are two isoenzymes of Cu,Zn-SOD found to be encoded by different genes; they differ in 19 out of 150 amino acids [25].

The discovery in the bacterium *Photobacter leiognathi* of a Cu,Zn-SOD, which was otherwise known only in eukaryotes, triggered a lively debate about its evolutionary origin. *P. leiognathi* is

```
Human          Ac-ATKAVCVLKG DGPVQGIINF EQKESNGPVK VWGSIKG-LT
Drosophila        VVKAVCVING DAKGTVFFEQ ESS--GTPVK VSGEVCG-LA
Yeast              VQAVAVLKG DAGVSGVVKF EQASESEPTT VSYEIAGNSP
Xiphias           VLKAVCVLRG AGETTGTVYF EQEGNANAVG KGIILKG-LT
Photobacter        QDLTVKMTD LQTGKPVGTI ELSQNKYGVV FTPELAD-LT

Hu     EGLHGFHVHE FGDNTAGCTS ---------- --AGPHFNP- LSRKHGGPKD
Dr     KGLHGFHVHE FGDNTNGCMS ---------- --SGPHFNP- YGKEHGAPVD
Ye     NAERGFHIHE FGDATDGCVS ---------- --AGPHFNP- FKKTHGAPTD
Xi     PGEHGFHVHG FGDNTNGCIS ---------- --AGPHFNP- ASKKHAGPKD
Ph     PGMHGFHIHQ NG----SCAS SEKDGKVVLG GAAGGHYDPE HTNKHGFPWT

Hu     EERHVGDLGN VTADKDGVAD VSIEDSVISL SGDHCIIGRT LVVHEKADDL
Dr     ENRHLGDLGN IEATGDCPTK VNITDSKITL FGADSIIGRT VVVHADADDL
Ye     EVRHVGDMGN VKTDEDGVAK GSFKDSLIKL IGPTSVVGRS VVIHAGQDDL
Xi     EDRHVGDLGN VTADANGVAK IDITD-KISL TGPYSIIGRT MVIHEKADDL
Ph     DDNHKGDLPA LFVSANGLAT NPVLAPRLTL K---ELKGHA IMIHAGGDNH

Hu     GKGGNEESTK TGNAGSRLAC GVIGIAQ
Dr     GQGGHELSKS TGNAGARIGC GVIGIAK
Ye     GKGDTEESLK TGNAGPRPAC GVIGLTN
Xi     GRGGNEESLK TGNAGSRLAC GVIGTE-
Ph     ----SDMPKA LGGGGARVAC GVIQ---
```

Fig. 18.9. The amino acid sequences of superoxide dismutases from humans, the fly *Drosophila melanogaster*, yeast, the swordfish *Xiphias gladius* and the luminescent bacterium *Photobacter leiognathi* [56, 93]

found not only free-living but also in symbiontic association with the luminescent ponyfish *Leiognathus splendens*. Although the sequence of the Cu,Zn-SOD from *L. splendens* is not known, the sequence agreement of about 33 % between the Cu,Zn-SOD from *P. leiognathi* and that of the fish *Xiphias gladus* and other eukaryotes (Fig. 18.9) suggested the occurrence of a gene transfer between the fish and its symbiont [6]. Cu,Zn-SODs have since been found in several other bacterial species. The Cu,Zn-SODs of the bacterial *P. leiognathi* and *Brucella abortus* show far greater sequence agreement with each other than with the enzyme from any vertebrate. This speaks for the existence of an independent bacterial type of Cu,Zn-SOD [11].

As befits its **protective function**, Cu,Zn-SOD is generally distributed in eukaryotic cells. It may be assumed that its presence is required wherever abnormally high oxygen partial pressures occur, e.g. in the gas glands of the teleosts. SOD activity is in fact correlated with the proportion of oxygen in the swim-bladder gas in surface fish [113]; investigations of deep-sea fish would be particularly interesting. Various toxic oxygen compounds arise from H_2S and O_2. Deep-sea animals such as the pogonophoran *Riftia pachyptila* and the mussel *Calyptogena magnifica*, which live in the neighbourhood of hydrothermal vents (p. 269) in hot, sulphide-rich waters, are exposed to such a danger. The muscles of *Riftia* contain several electrophoretically separable Cu,Zn-SODs and an Mn-SOD, both of which are probably of animal origin. The sulphide-oxidizing symbiontic bacteria that inhabit the gland-like tissue of the trophosome have an Fe-SOD and several Mn-SODs, and satisfy the greater part of the host's energy requirement by their chemosynthetic metabolism. The mussel also possesses an Mn-SOD and Cu,Zn-SOD in the bacteria-inhabited gills and bacteria-free muscles [14].

Human blood plasma, lymph and synovial fluid, and those of various other mammals, contain an **extracellular SOD** (EC-SOD), which also occurs intracellularly. This is a tetrameric glycoprotein of about 120 kDa containing one Cu and one Zn per subunit; it produces H_2O_2 like a Cu, Zn-SOD and is inhibited by cyanide, azide and H_2O_2. The activities of the EC-SODs in the blood plasma vary greatly with the species; they are 10- to 100-fold higher in the rabbit, rat and mouse than in humans, the dog and cat, although the intracellular activities of these species are very similar [80]. The EC-SODs show no similarity in amino acid composition or immunological properties to other Cu,Zn-SODs; the amino acid sequence derived from the cDNA shows that they contain sequence elements of different origin. As expected for a secretory protein, they have a signal sequence of 18 amino acids. The first 95 of the 222 amino acids of the mature protein show no similarity to other known protein sequences; however, positions 96–193 of EC-SOD agree up to 50 % with the C-terminal two-thirds of all eukaryote Cu,Zn-SODs. The C-terminal end of the EC-SOD (positions 194–222) is strongly hydrophilic and contains nine positively charged amino acids; this sequence is apparently responsible for the strong binding of EC-SOD to heparin and heparan sulphate. The large difference in sequence between EC-SOD and other Cu, Zn-SODs suggests that EC-SOD appeared before the separation of the fungi, plants and animals [59].

There exists little comparative biochemical data for the **Mn-SODs**. These enzymes are homotetramers of about 84 kDa; the subunits of human, rat and mouse Mn-SOD have been sequenced. They consist of 222 amino acids, including an N-terminal signal sequence of 24 amino acids for the intracellular target (the leader peptide). The sequences of the mature proteins agree by about 94 % but the leader peptides agree by only about 54 % [60]. The N-terminal sequence of the Mn-SOD from the liver of the American bullfrog *Rana catesbeiana* is significantly homologous to the enzymes of mammals and birds [2]. Two allelic variants of the Mn-SOD subunit (S and F) are found in the pike, giving a five-banded electrophoretic pattern corresponding to the tetrameric structure of the enzyme. At least three isoforms of **Fe-SOD** are detectable in the cytoplasm of the flagellate *Crithidia fasciculata*. The main form is a dimer of 43 kDa with three Fe atoms and it resembles the bacterial enzyme of this type in its amino acid composition. Like other Fe-SODs, it is cyanide resistant but is inhibited by azide and peroxide. SODs with similar responses to inhibitors are found in other Kinetoplastida, e.g. *Leishmania tropica, Trypanosoma brucei* and *T. cruzi* [169].

18.6.2 Catalases

The **catalase reaction** begins with the formation of a primary complex known as "compound I":

$$\text{Catalase} + H_2O_2 \rightarrow \text{compound I.} \qquad (18.6)$$

At higher H_2O_2 concentrations, a second H_2O_2 molecule serves as a further electron donor (cata-

lase activity: 18.7a), and at lower H_2O_2 concentrations an alcohol or other oxidizable compound serves as electron donor (peroxidase-like activity: 18.7b):

$$\text{Compound I} + H_2O_2 \rightarrow \text{catalase} + 2H_2O + O_2, \quad (18.7a)$$

$$\text{Compound I} + AH_2 \rightarrow \text{catalase} + 2H_2O + A. \quad (18.7b)$$

The turnover rate of $10^5\ s^{-1}$ measured for the catalase activity of the mammalian liver enzyme is the highest value recorded for any enzyme. The peroxidase-like activity is probably important mainly for the detoxification of H_2O_2 in living cells. The degradation of alcohol in the mammalian liver also occurs mainly as the result of peroxidase-like catalase activity. Catalases are ubiquitous in aerobic organisms but are lacking in strict anaerobes, such as *Entamoeba histolytica*, the trichomonads and certain stages of the Kinetoplastida *Leishmania mexicana, L. donovani* and *Trypanosoma brucei*. Amongst the platyhelminths, they are present in free-living turbellarians but not in endoparasitic trematodes and cestodes [10].

3-Amino-1,2,4-triazole (AT) is an irreversible **inhibitor** of catalase which binds to the 74-His of the active centre, but only in compound I. In living animals, AT completely inactivates the available catalase within several hours but has no influence on de novo synthesis. Thus, the subsequent increase in activity is a measure of the rate of synthesis [116]. The catalases are haem proteins with subunits of 53–65 kDa, each with a protohaem IX as the prosthetic group. In most cases, they are homotetramers of 210–280 kDa, e.g. in the vertebrates, the earthworm *Lumbricus terrestris* and *Drosophila melanogaster* [7, 116, 134]. However, catalases with deviant **molecular masses and quaternary structures** have also been reported, e.g. the 345-kDa enzyme from the mouse liver, dimers of 130 kDa in the tick *Hyalomma dromedarii*, and dimers of 117 kDa in the explosion chamber of the bombardier beetles *Brachinus crepitans* and *B. explodens* [78]. Almost complete **amino acid sequences** are so far available only for the catalases from bovine and rat liver and from human and bovine erythrocytes. The sequences consist of 506 to >520 amino acids, but it is unclear whether the differences in length are species or cell specific or are due to partial degradation. The sequences of the catalases from bovine liver and from bovine erythrocytes agree completely, but differ in at least 41 positions from the enzyme from human erythrocytes [50]. Immunological comparisons between the bovine enzymes and those of *Drosophila* and *Lumbricus* show no cross-reactivity [116, 134]. Electrophoresis of the liver catalases from the rat and mouse always gives two to five bands; however, there is only one gene and these bands are probably the result of post-translational modifications. The liver enzyme of the frog *Discoglossus pictus* also gives two bands but that of *Rana ridibunda* gives only one [7].

Catalase is the main enzyme found in a certain type of intracellular organelle, the **peroxisome**, which otherwise contains different H_2O_2-producing oxidases [43]. In contrast to the mitochondria, the peroxisomes are not active in energy extraction and their oxygen consumption is strongly dependent on the oxygen partial pressure; about one-third of the oxygen consumption of liver cells occurs in the peroxisomes. Almost all mammalian tissues additionally contain catalase particles that are smaller than peroxisomes and which also differ greatly in their fine structure (microperoxisomes). Only in the case of the mouse is the majority of liver catalase located in the peroxisomes; in the rat, the soluble and peroxisomal activities are about equal, and in the remaining mammals to have been examined the cytosolic enzyme activity is by far the highest. As was described earlier (p. 576), the peroxisomes play an important role in lipid metabolism; in plants and some protozoans, they are also involved in gluconeogenesis, the glyoxylate cycle and amino acid synthesis.

18.6.3 Peroxidases

Peroxidases oxidize various substrates and use H_2O_2 as the oxidation agent:

$$AH_2 + H_2O_2 \rightarrow A + 2H_2O. \quad (18.8)$$

They are also active in the formation of organic halogen compounds with halide ions and an appropriate acceptor, such as tyrosine, other substituted phenols, β-keto acids or cyclic β-diketones. They are almost ubiquitous in plants and animals, but even in the vertebrates they have not been sufficiently well examined in their structure and function. Several closely related haem peroxidases are present in mammalian secretions, such as milk, saliva and tears, and also in the cells of the thyroid gland and in leukocytes. The lactoperoxidase in cow's milk is a monomeric glycoprotein of 612 amino acids, including 15 hemicystines. One of the cysteine residues forms a

disulphide bridge with a protohaem IX in which a methyl is replaced by a mercaptomethyl group. There are four or five potential N-glycosylation sites, and the carbohydrate portion comprises up to 10 % of the molecular mass. Different carbohydrate fractions, deamidation and partial proteolysis are some of the possible factors that give rise to the heterogeneity of lactoperoxidases, heterogeneity that is observed even between the milk of individual animals. Bovine lactoperoxidase shows 54–55 % sequence agreement with human myeloperoxidase and eosinophil peroxidase, and 45 % agreement with human thyroperoxidase. Fourteen of the 15 hemicystines are located at identical positions in each of these four enzymes [21].

The homology of the few characterized peroxidases of lower vertebrates and invertebrates with those of the mammals cannot at present be estimated. The sponge *Iotrochota birotulata*, which like all sponges with a horn-like skeleton can halogenate proteins, and the crayfish *Orconectes* sp. possess at least two different peroxidases [65]. In the fertilization of sea-urchin eggs, the protein of the fertilization membrane is cross-linked and rigidified by an ovoperoxidase through the linking of tyrosine side-chains to bistyrosyl residues. The H_2O_2 required for this reaction is produced by the ovoperoxidase itself. This attains its NAD(P)H:O_2 oxidoreductase activity only after being activated in an undefined way by **ovothiol**, a multiply methylated 4-thiohistidine (Fig. 18.10) [170].

In view of the dangers associated with H_2O_2, it is rather strange to find present in some animals an enzyme, **glucose oxidase**, whose biological function appears in fact to be the production of H_2O_2. This typical enzyme of plants and bacteria is also present in the pharyngeal gland of the honeybee and in honey, where the H_2O_2 produced may have bactericidal activity. However, glucose oxidase is also found in the cuticula of the migratory locust *Schistocerca gregaria*, where it probably produces the H_2O_2 needed for the formation of the bis- and tertyrosine residues of resilin. Its presence in the hard cuticula of the lepidopteran *Calpodes ethlius* may be related to the formation of sclerotizing substances; why glucose oxidase is also present in the haemolymph is at present a mystery. The rat tapeworm *Hymenolepis diminuta*, like all parasitic worms, has no catalase but possesses a **cytochrome-c peroxidase**, the function of which may be the removal of the H_2O_2 that arises under aerobic conditions in the mitochondria [106].

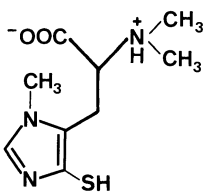


Fig. 18.10. Ovothiol (1-methyl-αN,αN-dimethyl-4-thiohistidine) confers on the ovoperoxidase of sea-urchin eggs the activity of an H_2O_2-producing NAD(P)H:O_2 oxidoreductase [170]

References

1. Abe Y. et al.: Copper and zinc superoxide dismutase from the liver of bullfrog Rana catesbeiana. Comp. Biochem. Physiol. Pt. B 77: 125–130 (1984)
2. Abe Y. and Okazaki T.: Purification and properties of the manganese superoxide dismutase from the liver of bullfrog, Rana catesbeiana. Arch. Biochem. Biophys. 253: 241–248 (1987)
3. Affranchino J. L. et al.: Terminal oxidases in the trypanosomatid Trypanosoma cruzi. Comp. Biochem. Physiol. Pt. B 85: 381–388 (1986)
4. Ambler R. P. and Daniel M.: Rattlesnake cytochrome-C. A re-appraisal of the reported amino acid sequence. Biochem. J. 274: 825–831 (1991)
5. Balogh A. G. et al.: Rabbit brown adipose tissue uncoupling protein mRNA: Use of only one of the two polyadenylation signals in its processing. Biochem. biophys. Res. Commun. 161: 156–161 (1989)
6. Bannister J. V. and Parker M. W.: The presence of a copper/zinc superoxide dismutase in the bacterium Photobacter leiognathi: a likely case of gene transfer from eukaryotes to prokaryotes. Proc. Nat. Acad. Sci. USA 82: 149–152 (1985)
7. Barja de Quiroga G., Gil P. and Alonso-Bedate M.: Catalase enzymatic activity and electrophoretic patterns in adult amphibians – a comparative study. Comp. Biochem. Physiol. Pt. B 80: 853–858 (1985)
8. Barra D. et al.: A tetrameric iron superoxide dismutase from the eucaryote Tetrahymena pyriformis. J. Biol. Chem. 265: 17680–87 (1990)
9. Barrett J.: Biochemistry of parasitic helminths. Macmillan, London 1981
10. Barrett J. and Beis I.: Catalase in free-living and parasitic platyhelminths. Experientia 38: 536 (1982)
11. Beck B. L., Tabatabai L. B. and Mayfield J. E.: A protein isolated from Brucella abortus is a Cu-Zn superoxide dismutase. Biochemistry 29: 372–376 (1990)
12. Benevides J. M. et al.: Determination of isocitrate lyase and malate synthase activities in a marine bivalve mollusc by a new method of assay. Comp. Biochem. Physiol. Pt. B 94: 779–782 (1989)
13. Biegniewska A. and Skorkowski E. F.: Isolation and some properties of malic enzyme from the shrimp abdominal muscle. Comp. Biochem. Physiol. Pt. B 74: 627–632 (1983)
14. Blum J. and Fridovich I.: Enzymatic defenses against oxygen toxicity in the hypodermal vent animals Riftia pachyptila and Calyptogena magnifica. Arch. Biochem. Biophys. 228: 617–620 (1984)
15. Bolli R., Nalecz K. A. and Azzi A.: The aggregation state of bovine heart cytochrome c oxidase and its kinetics in monomeric and dimeric form. Arch. Biochem. Biophys. 240: 102–116 (1985)

16. Bradfield J. Y., Lee Y. H. and Keeley L. L.: Cytochrome-P450 family-4 in a cockroach. Molecular cloning and regulation by hypertrehalosemic hormone. Proc. Nat. Acad. Sci. USA 88: 4558–62 (1991)
17. Brems D. N., Cass R. and Stellwagen E.: Conformational transitions of frog heart cytochrome c. Biochemistry 21: 1488–93 (1982)
18. Brunori M. et al.: Cytochrome-c oxidase – Subunit structure and proton pumping. Eur. J. Biochem. 169: 1–8 (1987)
19. Brunori M. and Chance B. (eds.): Cytochrome oxidase. Structure, function and physiopathology. New York Acad. Sci., New York 1989
20. Bulos B. A. et al.: Regulation of pyruvate oxidation in blowfly flight muscle mitochondria: requirement for ADP. Arch. Biochem. Biophys. 234: 382–393 (1984)
21. Cals M. M. et al.: Primary structure of bovine lactoperoxidase, a forth member of a mammalian heme peroxidase family. Eur. J. Biochem. 198: 733–739 (1991)
22. Candy D. J.: Intermediary metabolism. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 1–41. Pergamon Press, Oxford 1985
23. Cannata J. J. B. and Cazzulo J. J.: The aerobic fermentation of glucose by Trypanosoma cruzi. Comp. Biochem. Physiol. Pt. B 79: 297–308 (1984)
24. Capaldi R. A.: Structure and function of cytochrome c oxidase. Annual Rev. Biochem. 59: 569–596 (1990)
25. Carri M. T. et al.: Xenopus laevis Cu,Zn superoxide dismutase-B cDNA. Nucleic Acids Res. 18: 1641 (1990)
26. Chang K. M. et al.: Aryl hydrocarbon hydroxylase in larvae of the southern armyworm (Spodoptera eridania). Insect Biochem. 13: 87–94 (1983)
27. Cioni M, Pinzauti G. and Vanni P.: Comparative biochemistry of the glyoxylate cycle (Review). Comp. Biochem. Physiol. Pt. B 70: 1–26 (1981)
28. Cleeter M. W. and Ragan C. I.: The polypeptide composition of the mitochondrial NADH:ubiquinone reductase complex from several mammalian species. Biochem. J. 230: 739–746 (1985)
29. Cohen M. B., Berenbaum M. R. and Schuler M. A.: Immunochemical analysis of cytochrome-P450 monoxygenase diversity in the black swallowtail caterpillar, Papilio polyxenes. Insect Biochem. 20: 777–783 (1990)
30. Collins P. J.: Induction of the polysubstrate monooxygenase system of the native budworm Heliothis punctiger (Wallengren) (Lepidoptera: Noctuidae). Insect Biochem. 15: 551–555 (1985)
31. Cook J. S. et al.: Nucleotide sequence of the mRNA encoding the cytosolic form of phosphoenolpyruvate carboxykinase (GTP) from the chicken. Proc. Nat. Acad. Sci. USA 83: 7583–87 (1986)
32. Davis W. L., Matthews J. L. and Goodman D. B. P.: Glyoxylate cycle in the rat liver: effect of vitamin D3 treatment. Faseb J. 3: 1651–55 (1989)
33. Davis W. L. et al.: Hibernation activates glyoxylate cycle and gluconeogenesis in black bear brown adipose tissue. Biochim. biophys. Acta 1051: 276–278 (1990)
34. Depinto V. et al.: Characterization of pore-forming activity in liver mitochondria from Anguilla anguilla. Two porins in mitochondria. Biochim. biophys. Acta 1061: 279–286 (1991)
35. Dolphin C. et al.: Cloning, primary sequence and chromosomal mapping of human flavin-containing monooxygenase (FMO1). J. Biol. Chem. 266: 12379–85 (1991)
36. Doussière J. and Vignais P. V.: AMP-dependence of the cyanide insensitive pathway in the respiratory chain of Paramecium tetraurelia. Biochem. J. 220: 787–794 (1984)
37. Duke M. V. and Salin M. L.: Purification and characterization of an iron-containing superoxide dismutase from a eukaryote, Ginkgo biloba. Arch. Biochem. Biophys. 243: 305–314 (1985)
38. Ernster L.: Bioenegetics. New Comprehensive Biochemistry Vol. 9. Elsevier, Amsterdam 1984
39. Esposti M. D. et al.: Comparative biochemistry of the ubiquinol-cytochrom c oxidoreductase (EC 1.10.2.2) isolated from different heart mitochondria. Comp. Biochem. Physiol. Pt. B 85: 543–552 (1986)
40. Evans M. J. and Scarpulla R. C.: The human somatic cytochrome c gene. Two classes of processed pseudogenes demarcate a period of rapid molecular evolution. Proc. Nat. Acad. Sci. USA 85: 9625–29 (1988)
41. Fabregat I., Satrustegui J. and Machado A.: Citrate synthase of Tetrahymena pyriformis: evolutionary and regulatory aspects. Arch. Biochem. Biophys. 220: 354–360 (1983)
42. Fabregat I., Satrustegui J. and Machado A.: Malate dehydrogenase of Tetrahymena pyriformis: evolutionary and regulatory aspects. Comp. Biochem. Physiol. Pt. B 77: 319–323 (1984)
43. Fahimi H. D. and Sies H. (eds.): Peroxisomes in biology and medicine. Springer, Berlin 1987
44. Fairlamb A.: Biochemistry of trypanosomiasis. Trends biochem. Sci. 7: 249–253 (1982)
45. Ferreira M. F. A., Andrade C. M. and Ribeiro L. P.: Mitochondrial malic enzyme of Toxocara canis muscle. Comp. Biochem. Physiol. Pt. B 81: 939–944 (1985)
46. Feyereisen R. and Farnsworth E. E.: Developmental changes of microsomal cytochrome P-450 monooxygenases in larval and adult Diploptera punctata. Insect Biochem. 15: 755–761 (1985)
47. Feyereisen R. et al.: Isolation and sequence of cDNA encoding a cytochrome P-450 from an insecticide-resistant strain of the house fly, Musca domestica. Proc. Nat. Acad. Sci. USA 86: 1465–69 (1989)
48. Fickenscher K., Scheibe R. and Marcus F.. Amino acid sequence similarity between malate dehydrogenase (NAD) and pea chloroplast malate dehydrogenase (NADP). Eur. J. Biochem. 168: 653–658 (1987)
49. Fukushima H. et al.: Topographical orientation of electron transport enzymes and glucose-6-phosphatase in Tetrahymena microsomes. Comp. Biochem. Physiol. Pt. B 81: 443–445 (1985)
50. Furuta S. et al.: Complete nucleotide sequence of cDNA and deducted amino acid sequence of rat liver catalase. Proc. Nat. Acad. Sci. USA 83: 313–317 (1986)
51. Gallwitz W. E. et al.: Purification and characterization of the isozymes of phosphoenolpyruvate carboxykinase from rabbit liver. Biochim. biophys. Acta 964: 36–45 (1988)
52. Galtieri A. et al.: Isolation and characterization of Cu,Zn-superoxide dismutase of the shark Prionace glauca. Comp. Biochem. Physiol. Pt. B 83: 555–559 (1986)
53. Goto Y. et al.: Cytosolic and mitochondrial phosphoenolpyruvate carboxykinase of the bullfrog, Rana catesbeiana, liver. Biochim. biophys. Acta 657: 383–389 (1981)
54. Hatefi Y.: The mitochondrial electron transport and oxidative phosphorylation system. Annual Rev. Biochem. 54: 1015–69 (1985)

55. Hekman C., Tomich J. M. and Hatefi Y.: Mitochondrial ATP synthase complex: Membrane topography and stoichiometry of the F0 subunits. J. Biol. Chem. 266: 13564–71 (1991)
56. Hering K. et al.: The primary structure of porcine Cu-Zn superoxide dismutase: evidence for alloenzymes of superoxide dismutase in pigs. Biol. Chem. Hoppe-Seyler 366: 435–445 (1985)
57. Hers H. G. and Hue L.: Gluconeogenesis and related aspects of glycolysis. Annual Rev. Biochem. 52: 617–653 (1983)
58. Himms-Hagen J.: Brown adipose tissue thermogenesis: interdisciplinary studies. Faseb J. 4: 2890–98 (1990)
59. Hjalmarsson K. et al.: Isolation and sequence of complementary DNA encoding human extracellular superoxide dismutase. Proc. Nat. Acad. Sci. USA 84: 6340–44 (1987)
60. Ho Y. S. and Crapo J. D.: Isolation and characterization of complementary DNAs encoding human manganese-containing superoxide dismutase. FEBS Letters 229: 256–260 (1988)
61. Hodgson E.: The significance of cytochrome P-450 in insects (Short review). Insect Biochem. 13: 237–246 (1983)
62. Houldsworth J. and Attardi G.: Two distinct genes for ADP/ATP translocase are expressed at the mRNA level in adult human liver. Proc. Nat. Acad. Sci. USA 85: 377–381 (1988)
63. Howell N.: Evolutionary conservation of protein regions in the protonmotive cytochrome b and their possible roles in redox catalysis. J. mol. Evol. 29: 157–169 (1989)
64. Huang C. J. et al.: Nucleotide sequence of carp mitochondrial cytochrome c oxidase III. Nucleic Acids Res. 18: 1056 (1990)
65. Ilgner R. H. and Woods A. E.: Purification, physical properties and kinetics of peroxidases from freshwater crayfish (genus Orconectes). Comp. Biochem. Physiol. Pt. B 82: 433–440 (1985)
66. Imai Y., Komori M. and Sato R.. Comparison of primary structures deduced from cDNA nucleotide sequences for various forms of liver microsomal cytochrome P-450 from phenobarbital-treated rats. Biochemistry 27: 80–88 (1988)
67. Imbuga M. O. and Pearson D. J.: The kinetic properties of NAD-linked malic enzyme from muscle of the dung beetle Catharsius. Insect Biochem. 12: 55–61 (1982)
68. Inoue S. et al.: Complete amino-acid sequences of two isocytochromes c of the housefly, Musca domestica L., and their developmental variation in different tissues. Biochim. biophys. Acta 790: 188–195 (1984)
69. James M. O. and Shiverick K. T.: Cytochrome P-450-dependent oxidation of progesterone, testosterone, and ecdysone in the spiny lobster, Panulirus argus. Arch. Biochem. Biophys. 233: 1–9 (1984)
70. Jenkins T. M., Eisenthal R. and Weitzman P. D. J.: Two distinct succinate thiokinases in both bloodstream and procyclic forms of Trypanosoma brucei. Biochem. biophys. Res. Commun. 151: 257–261 (1988)
71. Jewell C. S. E. and Winston G. W.: Characterization of the microsomal mixed-function oxygenase system of the hepatopancreas and green gland of the red swamp crayfish, Procambarus clarkii. Comp. Biochem. Physiol. Pt. B 92: 329–339 (1989)
72. Joergensen L.: Identification of mitochondrial cytochromes of Tubifex sp. using fourth-order finite difference analysis of low temperature spectra. Comp. Biochem. Physiol. Pt. B 69: 769–774 (1981)
73. Joh T. et al.: Cloning and sequence analysis of cDNAs encoding mammalian cytosolic malate dehydrogenase. Comparison of the amino acid sequences of mammalian and bacterial malate dehydrogenase. J. biol. Chem. 262: 15127–31 (1987)
74. Joh T. et al.: Cloning and sequence analysis of cDNAs encoding mammalian mitochondrial malate dehydrogenase. Biochemistry 26: 2515–20 (1987)
75. Johnson W. V., Kemp J. R. and Anderson P. M.: Purification and properties of mitochondrial phosphoenolpyruvate carboxykinase from liver of Squalus acanthias. Arch. Biochem. Biophys. 280: 376–382 (1990)
76. Joshi S. and Burrows R.: ATP synthase complex from bovine heart mitochondria, Subunit arrangement as revealed by nearest neighbor analysis and suceptibility to trypsin. J. Biol. Chem. 265: 14518–25 (1990)
77. Kamel M. Y. and Fahmy A. S.: Biochemical studies on tick embryogenesis. V. Purification and partial characterization of isocitrate lyase from eggs of the tick Hyalomma dromedarii. Comp. Biochem. Physiol. Pt. B 72: 107–115 (1982)
78. Kamel M. Y. and Hamed R. R.: Purification and properties of catalase from developing embryos of Hyalomma dromedarii (Acarina: Ixodidae). Insect Biochem. 12: 481–491 (1982)
79. Kargbo A. O. and Swift M. L.: NAD-dependent isocitrate dehydrogenase from the oyster, Crassostrea virginica Gmelin. Comp. Biochem. Physiol. Pt. B 76: 123–126 (1983)
80. Karlsson K. and Marklund S. L.: Extracellular superoxide dismutase in the vascular system of mammals. Biochem. J. 255: 223–228 (1988)
81. Keech D. B. and Wallace J. C. (eds.): Pyruvate carboxylase. CRC Press, Boca Raton USA 1985
82. Kita K. et al.: Complex II is a major component of the respiratory chain in muscle mitochondria of Ascaris suum with high fumarate reductase activity. Comp. Biochem. Physiol. Pt. B 89: 31–34 (1988)
83. Klingenberg M.: Mechanism and evolution of the uncoupling protein of brown adipose tissue. Trends biochem. Sci. 15: 108–112 (1990)
84. Klotz A. V. et al.: Cytochrome P-450 isozymes from the marine teleost Stenotomus chrysops: their roles in steroid hydroxylation and the influence of cytochrome b5. Arch. Biochem. Biophys. 249: 326–338 (1986)
85. Koudelka A. P.: A cytochrome b5 electron transport chain in Tetrahymena. Biochim. biophys. Acta 751: 121–126 (1983)
86. Kozak L. P. et al.: The mitochondrial uncoupling protein gene. Correlation of exon structure to transmembrane domains. J. Biol. Chem. 263: 12274–77 (1988)
87. Kraemer R. and Palmieri F.: Molecular aspects of isolated and reconstituted carrier proteins from animal mitochondria. Biochim. biophys. Acta 974: 1–23 (1989)
88. Kuwahara T. et al.: A cytosolic FAD-containing enzyme catalyzing cytochrome c reduction in Trypanosoma cruzi. I. Purification and some properties. Arch. Biochem. Biophys. 239: 18–28 (1985)
89. Kwiatowski J., Patel M. and Ayala F. J.: Drosophila melanogaster Cu,Zn superoxide dismutase gene sequence. Nucleic Acids Res. 17: 1264 (1989)
90. Kwiatowski J. and Ayala F. J.: Drosophila virilis Cu,Zn superoxide dismutase gene sequence. Nucleic Acids Res. 17: 2133 (1989)

91. Kwiatowski J., Gonzalez F. and Ayala F. J.: Drosophila simulans Cu,Zn superoxide dismutase gene sequence. Nucleic Acids Res. 17: 6735 (1989)
92. Lee Y. M., Friedman D. J. and Ayala F.: Superoxide dismutase: an evolutionary puzzle. Proc. Nat. Acad. Sci. USA 82: 824–828 (1985)
93. Lee Y. M., Friedman D. J. and Ayala F.: Complete amino acid sequence of copper-zinc superoxide dismutase from Drosophila melanogaster. Arch. Biochem. Biophys. 241: 577–589 (1985)
94. Lightowlers R. et al.: Cloning and sequencing of the cDNA for a 13th different subunit (IHQ) of beef heart cytochrome c oxidase. J. Biol. Chem. 264: 16858–61 (1989)
95. Limbach K. J. and Wu R.: Characterization of a mouse somatic cytochrome c gene and three cytochrome c pseudogenes. Nucleic Acids Res. 13: 617–630 (1985)
96. Limbach K. J. and Wu R.: Characterization of two Drosophila melanogaster cytochrome c genes and their transcripts. Nucleic Acids Res. 13: 631–644 (1985)
97. LLoyd D. et al.: Cytochrome a-620 in Tetrahymena pyriformis. Reactions with carbon monoxide and oxygen at subzero temperatures and photochemical action spectra. Biochem. J. 206: 367–372 (1982)
98. Madero E., Gallego-Iniesta M. and Garrido-Pertierra A.: Properties and function of a mitochondrial malate enzyme from bass liver. Comp. Biochem. Physiol. Pt. B 83: 861–866 (1986)
99. Madin K. A. C., Loomis S. H. and Crowe J. H.: Anhydrobiosis in nematodes – Control of carbon flow through the glyoxylate cycle. J. exp. Zool. 234: 341–350 (1985)
100. Mansini E., Oestreicher E. G. and Ribeiro L. P.: Purification and properties of mitochondrial malate dehydrogenase of Toxocara canis muscle. Comp. Biochem. Physiol. Pt. B 85: 223–228 (1986)
101. Marr J. J.: Carbohydrate metabolism in Leishmania. In: Levandowsky M. and Hutner S. H. (eds.): Biochemistry and physiology of Protozoa, 2. ed., Vol. 3, pp. 313–340. Acad. Press, New York 1980
102. Marsh A. and Lawrence J.: The effects of cations on the activity of citrate synthase (EC 4.1.3.7) in Lucidia clathrata (Say) (Echinodermata: Asteroidea). Comp. Biochem. Physiol. Pt. B 81: 767–770 (1985)
103. Matsunaga E., Umeno M. and Gonzalez F. J.: The rat P450 IID subfamily: Complete sequences of four closely linked genes and evidence that gene conversions maintained sequence homogeneity at the heme-binding region of the cytochome P450 active site. J. mol. Evol. 30: 155–169 (1990)
104. Matuda S. et al.: Immunological identification of a new component of Ascaris pyruvate dehydrogenase complex. Biochim. biophys. Acta 926: 54–60 (1987)
105. McCormack J. G. and Denton R. M.: A comparative study of the regulation by calcium of the activities of the 2-oxoglutarate dehydrogenase complex and NAD-isocitrate dehydrogenase from a variety of sources. Biochem J. 196: 619–624 (1981)
106. McKelvey J. R. and Fioravanti C. F.: Localization of cytochrome c oxidase and cytochrome c peroxidase in mitochondria of Hymenolepis diminuta (Cestoda). Comp. Biochem. Physiol. Pt. B 85: 333–335 (1986)
107. Megias A., Saborido A. and Municio A. M.: Cytochrome b5 from the insect Ceratitis capitata. Biochim. biophys. Acta 872: 116–124 (1986)
108. Mendis A. H. W. and Evans A. A. F.: First evidence for the occurrence of cytochrome-o in a freeliving nematode. Comp. Biochem. Physiol. Pt. B 78: 729–735 (1984)
109. Miranda C. L. et al.: Comparison of rainbow trout and mammalian cytochrome-P450 enzymes. Evidence for structural similarity between trout P450 LMC5 and human P450IIIA4. Biochem. biophys. Res. Commun. 176: 558–563 (1991)
110. Montecucco C. et al.: Isolation and characterization of cytochrom c oxidase from bird and fish heart mitochondria. Comp. Biochem. Physiol. Pt. B 87: 851–856 (1987)
111. Moore G. R. and Pettigrew G. W.: Cytochromes c. Evolutionary, structural and physicochemical aspects. Springer, New York 1990
112. Morioka H. et al.: Structural characterization of the rat malic enzyme gene. Proc. Nat. Acad. Sci. USA 86: 4912–16 (1989)
113. Morris S. M. and Albright J. T.: Glutathione peroxidase, and superoxide dismutase in the rete mirabile and gas gland epithelium of six species of marine fishes. J. exp. Zool. 232: 29–39 (1984)
114. Mottram J. C. and Coombs G. H.: Purification of particulate malate dehydrogenase and phosphoenolpyruvate carboxykinase from Leishmania mexicana mexicana. Biochim. biophys. Acta 827: 310–319 (1985)
115. Mourik J.: Oxidations in the tricarboxylic acid cycle by intact mitochondria isolated from the lateral red muscle of goldfish (Carassius auratus L.). Effects of anoxia on the oxidation of pyruvate and glutamate. Comp. Biochem. Physiol. Pt. B 76: 851–859 (1983)
116. Nahmias J. A. and Bewley G. C.: Characterization of a catalase purified from Drosophila melanogaster by hydrophobic interaction chromatography. Comp. Biochem. Physiol. Pt. B 77: 355–364 (1984)
117. Nebert D. W. et al.: The P450 superfamily: Update on new sequences, gene mapping, and recommended nomenclaturs. DNA and Cell Biol. 10: 1–14 (1991)
118. Nichols D. G.: Bioenergetics. Acad. Press, New York 1982
119. Noshiro M. and Omura T.: Microsomal monooxygenase system in frog livers. Comp. Biochem. Physiol. Pt. B 77: 761–767 (1984)
120. O'Brien R. W. and Breznak J. A.: Enzymes of acetate and glucose metabolism in termites. Insect Biochem. 14: 639–643 (1984)
121. Ogata K. et al.: Mitochondrial phosphoenolpyruvate carboxykinase (GTP) and the regulation of gluconeogenesis and ketogenesis in avian liver. J. biol. Chem. 257: 5385–91 (1982)
122. O'Riordan V. B. and Burnell A. M.: Intermediary metabolism in the dauer larvae of the nematode Caenorhabditis elegans. II. The glyoxylate cycle and fatty-acid oxidation. Comp. Biochem. Physiol. Pt. B 95: 125–130 (1990)
123. Paget T. A., Fry M. and Lloyd D.: Haemoprotein terminal oxidases in the nematodes Nippostrongylus brasiliensis and Ascaridia galli. Biochem. J. 256: 295–298 (1988)
124. Papa S., Chance B. and Ernster L. (eds.): Cytochrome systems. Molecular biology and bioenergetics. Plenum, New York 1988
125. Parker M. W. and Blake C. C. F.: Iron- and manganese-containing superoxide dismutases can be distinguished by analysis of their primary structures. FEBS Letters 229: 377–382 (1988)

126. Parsons M. and Smith J. M.: Trypanosomal glycosomal protein P60 is homologous to phosphoenolpyruvate carboxykinase (ATP). Nucleic Acids Res. 17: 6411 (1989)
127. Patel M. S. and Roche T. E.: Molecular biology and biochemistry of pyruvate dehydrogenase complexes. Faseb J. 4: 3224–33 (1990)
128. Paynter K. T. et al.: Pyruvate dehydrogenase complex from ribbed mussel gill mitochondria. J. exp. Zool. 236: 251–257 (1985)
129. Poerio E., Parr G. R. and Taniuchi H.: A study of roles of evolutionary invariant proline 30 and glycine 34 of cytochrome c. J. biol. Chem. 261: 10976–89 (1986)
130. Pons G. et al.: Cloning and cDNA sequence of the dihydrolipoamide dehydrogenase component of human α-ketoacid dehydrogenase complex. Proc. Nat. Acad. Sci. USA 85: 1422–26 (1988)
131. Powell M. A. and Arp A. J.: Hydrogen sulfide oxidation by abundant nonhemoglobin heme compounds in marine invertebrates from sulfide-rich habitats. J. exp. Zool. 249: 121–132 (1989)
132. Powell S. J. et al.: Two bovine genes for mitochondrial ADP/ATP translocase expressed differentially in various tissues. Biochemistry 28: 866–873 (1989)
133. Precious W. Y. and Barrett J.: The posssible absence of cytochrome P-450 linked xenobiotic metabolism in helminths. Biochim. biophys. Acta 992: 215–222 (1989)
134. Prento P. and Prento A.: Crystalline catalase from the earthworm Lumbricus terrestris (Oligochaeta: Annelida): purification and properties. Comp. Biochem. Physiol. Pt. B 77: 325–328 (1984)
135. Rasmussen U. B. and Wohlrab H.: Conserved structural domains among species and tissue-specific differences in the mitochondrial phosphate-transport protein and the ADP/ATP carrier. Biochim. biophys. Acta 852: 306–314 (1986)
136. Reinecke K. et al.: The amino acid sequence of rabbit Cu-Zn superoxide dismutase. Biol. Chem. Hoppe-Seyler 369: 715–725 (1988)
137. Revilla E., Fabregat I. and Machado A.: NADPH/NADP ratio could regulate the glyoxylate cycle in Tetrahymena pyriformis. Comp. Biochem. Physiol. Pt. B 88: 851–854 (1987)
138. Ricquier D., Casteilla L. and Bouillaud F.: Molecular studies on the uncoupling protein. Faseb J. 5: 2237–42 (1991)
139. Riviere J. L. and Cabanne F.: Animal and plant cytochrome P-450 systems. Biochimie 69: 743–752 (1987)
140. Robinson J. B. jr. and Srere P. A.: Organization of Krebs tricarboxylic acid cycle enzymes in mitochondria. J. biol. Chem. 260: 10800–05 (1985)
141. Rocha H. A., Bannister W. H. and Bannister J. V.: The amino-acid sequence of copper/zinc superoxide dismutase from swordfish liver. Comparison of copper/zinc superoxide dismutase sequences. Eur. J. Biochem. 145: 477–484 (1984)
142. Roche T. E. and Patel M. (eds.): Alpha-keto acid dehydrogenase complexes. Organization, regulation, and biomedical ramifications. New York Acad. Sci., New York 1989
143. Ruano A. R., Santos J. H. and Ruiz-Amil M.: NADP-dependent isocitrate dehydrogenase from hepatopancreas of mussel (Mytilus edulis L.). Comp. Biochem. Physiol. Pt. B 81: 953–957 (1985)
144. Runswick M. J. et al.: Sequence of the bovine 2-oxoglutarate/malate carrier protein. Structural relationship to other mitochondrial transport proteins. Biochemistry 29: 11033–40 (1990)
145. Sacchettini J. C. et al.: Amino acid sequence of porcine heart fumarase. Biochem. biophys. Res. Commun. 153: 435–440 (1988)
146. Saccone C., Pesole G. and Kadenbach B.: Evolutionary analysis of the nucleus-encoded subunits of mammalian cytochrome-c oxidase. Eur. J. Biochem. 195: 151–156 (1991)
147. Salvarrey M. S. and Cazzulo J. J.: Citrate synthase from Crithidia fasciculata: inhibition by adenine nucleotides and Suramin. Comp. Biochem. Physiol. Pt. B 72: 165–168 (1982)
148. Schägger H. et al.: Isolation and amino acid sequence of the „Rieske" iron sulfur protein of beef heart ubiquinol:cytochrome c reductase. FEBS Letters 219: 161–168 (1987)
149. Schinina M. E. et al.: Primary structure of porcine Cu,Zn superoxide dismutase. FEBS Letters 186: 267–270 (1985)
150. Schinina M. E. et al.: Primary structure of a cationic Cu-Zn superoxide dismutase: the sheep enzyme. FEBS Letters 207: 7–10 (1986)
151. Schlenk D. and Buehler D. R.: Determination of multiple forms of cytochrome P450 in microsomes from the digestive gland of Cryptochiton stelleri. Biochem. biophys. Res. Commun. 163: 476–480 (1989)
152. Schwantes M. L. B. and Schwantes A. R.: Adaptive features of ectothermic enzymes – I. Temperature effects on the malate dehydrogenase from a temperate fish Leiostomus xanthurus. Comp. Biochem. Physiol. Pt. B 72: 49–58 (1982)
153. Scott R. I. and Lloyd D.: Photochemical action spectra indicate that cytochrome a/a3 is the predominant haemoprotein terminal oxidase in Acanthamoeba castellanii. Biochem. J. 210: 721–725 (1983)
154. Seto N. O. L., Hayashi S. and Tener G. M.: Cloning, sequence analysis and chromosomal localization of the Cu-Zn superoxide dismutase gene of Drosophila melanogaster. Gene 75: 85–92 (1989)
155. Setoyama C. et al.: Structural organization of the mouse cytosolic malate dehydrogenase gene: Comparison with that of the mouse mitochondrial malate dehydrogenase gene. J. mol. Biol. 202: 355–364 (1988)
156. Shoukry M. I.. Aconitase from the oyster Crassostrea virginica. Comp. Biochem. Physiol. Pt. B 72: 321–324 (1982)
157. Shrago E. and Elson C.. Intermediary metabolism in Tetrahymena. In: Levandowsky M. and Hutner S. H. (eds.): Biochemistry and physiology of Protozoa, 2. ed., Vol. 3, pp. 287–312. Acad. Press, New York 1980
158. Skorkowski E. F. et al.: Malic enzymes of salmon trout heart mitochondria: separation and some physicochemical properties of NAD-preferring and NADP-specific enzymes. Comp. Biochem. Physiol. Pt. B 80: 901–907 (1985)
159. Skorkowski E. F.: Mitochondrial malic enzyme from crustacean and fish muscle. Comp. Biochem. Physiol. Pt. B 90: 19–24 (1988)
160. Song B. J. et al.: Complementary DNA and protein sequences of ethanol-inducible rat and human cytochrome P-450s. Transcriptional and post-transcriptional regulating of the rat enzyme. J. biol. Chem. 261: 16689–97 (1986)

161. Steffens G. J. et al.: The amino-acid sequence of rat Cu,Zn superoxide dismutase. Biol. Chem. Hoppe-Seyler 367: 1017–24 (1986)
162. Stroh A. and Kadenbach B.: Tissue-specific and species-specific distribution of -SH groups in cytochrome c oxidase subunits. Eur. J. Biochem. 156: 199–204 (1986)
163. Sundseth S. S., Nix C. E. and Waters L. C.: Isolation of insecticide resistance-related forms of cytochrome P-450 from Drosophila melanogaster. Biochem. J. 265: 213–217 (1990)
164. Suzuki H. et al.: Cloning and sequencing of a cDNA for human mitochondrial ubiquinone-binding protein of complex III. Biochem. biophys. Res. Commun. 156: 987–994 (1988)
165. Suzuki T. et al.: Rat liver mitochondrial and cytosolic fumarases with identical amino acid sequences are encoded from a single gene. J. Biol. Chem. 264: 2581–86 (1989)
166. Takamiya S., Furushima R. and Oya H.. Electron-transfer complexes of Ascaris suum muscle mitochondria. II. Succinate-coenzyme Q reductase (complex II) associated with substrate-reducible cytochrome. Biochim. biophys. Acta 848: 99–107 (1986)
167. Takamiya S., Lindorfer M. A. and Capaldi R. A.: Purification of all thirteen polypeptides of bovine heart cytochrome c oxidase from one aliquot of enzyme: Characterization of bovine fetal heart cytochrome c oxidase. FEBS Letters 218: 277–282 (1987)
168. Taroni F., Gellera C. and di Donato S.: Evidence for two distinct mitochondrial malic enzymes in human skeletal muscle: purification and properties of the NAD(P)-dependent enzyme. Biochim. biophys. Acta 916: 446–454 (1988)
169. Le Trant N. et al.: Iron-containing superoxide dismutase from Crithidia fasciculata. Purification, characterization, and similarity to leishmanial and trypanosomal enzymes. J. biol. Chem. 258: 125–130 (1983)
170. Turner E. et al.: Ovothiol: a novel thiohistidine compound from sea urchin eggs that confers NAD(P)H-oxygen oxidoreductase activity to ovoperoxidase. J. biol. Chem. 261: 13056–63 (1986)
171. Turrens J. F.: The role of succinate in the respiratory chain of Trypanosoma brucei procyclic trypomastigotes. Biochem. J. 259: 363–368 (1989)
172. Valentine J. and Pettigrew G. W.: A cytochrome c methyltransferase from Crithidia oncopelti. Biochem. J. 201: 329–338 (1982)
173. Vanni P. et al.: Comparative structure, function and regulation of isocitrate lyase, an important assimilatory enzyme (Review). Comp. Biochem. Physiol. Pt. B 95: 431–458 (1990)
174. Vincent D. R. and Terriere L. C.: Characterization of NADPH-cytochrom P-450 reductase from house flies (Musca domestica L.) susceptible and resistant to insecticides, and the blowfly (Phormia regina, Meigen). Insect Biochem. 15: 299–307 (1985)
175. Virbasius J. V. and Scarpulla R. C.: Structure and expression of rodent genes encoding the testis-specific cytochrome c. Differences in gene structure and evolution between somatic and testicular variants. J. Biol. Chem. 263: 6791–96 (1988)
176. Weeda E.: Some properties of mitochondrial NAD linked malic enzyme and malate dehydrogenase from the flight muscles of Leptinotarsa decemlineata. Insect Biochem. 11: 679–684 (1981)
177. Weinbach E. C.: Biochemistry of enteric parasitic Protozoa. Trends biochem. Sci. 6: 254–257 (1981)
178. Weiss H. et al.: The respiratory-chain NADH dehydrogenase (complex-I) of mitochondria (Review). Eur. J. Biochem. 197: 563–576 (1991)
179. Yaginuma T. and Yamashita O.: Malate-aspartate cycle as an effective hydrogen shuttle at the termination of diapause in the eggs of Bombyx mori. Insect Biochem. 16: 677–685 (1986)
180. Yamaguchi M. et al.: The primary structure of the mitochondrial energy-linked nicotinamide nucleotide transhydrogenase deduced from the sequence of cDNA clones. J. Biol. Chem. 263: 2761–67 (1988)
181. Yarlett N., Lloyd D. and Williams A. G.: Respiration of rumen ciliate Dasytricha ruminantium. Biochem J. 206: 259–266 (1982)
182. Yoo M. and Steggles A. W.: The complete nucleotide sequence of human liver cytochrome b5 mRNA. Biochem. biophys. Res. Commun. 156: 576–580 (1988)
183. Zeviani M. et al.: Sequence of cDNAs encoding subunit Vb of human and bovine cytochrome c oxidase. Gene 65: 1–11 (1988)
184. Multi-author-review: Pore-forming proteins of biological membranes. Experientia 46: 131–200 (1990)

19 Secondary Metabolites

The components of macromolecules and supramolecular cell structures as well as the substrates of energy-yielding processes of cell metabolism are generally the same in all organisms. In addition, higher plants, fungi and bacteria in particular contain a large number of low molecular weight substances which are not part of this central anabolic and catabolic metabolism, and are therefore referred to as secondary metabolites [118]. The secondary metabolites of plants arise from the action of specific enzymes on products of primary metabolism, e.g. acetate, mevalonate, amino acids and carbohydrates. The complement of such enzymes, and thus the spectrum of secondary metabolites, varies with the species and the group, and is exploited by taxonomists. Several secondary metabolites are of great practical importance for the production of pharmaceuticals.

The **number of secondary metabolites** in animals is nothing like as large as that in plants. Nevertheless, more than 1000 such substances have already been isolated from marine invertebrates, especially sponges and corals. These are particularly interesting as many of them have pharmacological effects, and thus the potential for medicinal application [7, 33, 161]. The source for the unique variety of sterols and carotenoids in marine animals probably lies in the prolific species of marine algae. Seawater is rich in halogens; this explains not only the wealth of organohalogen compounds in marine organisms but perhaps also the numerous marine terpenes, which can arise by bromine-induced cyclization. On the other hand, compared with the terrestrial plants and animals, marine organisms are poorer in alkaloids [62]. Numerous secondary metabolites in particular signalling substances are known from the insects; these have received systematic attention because of their possible application in biological pest control, especially for certain groups of Lepidoptera [182]. Finally, there is a wealth of comparative biochemical data on the pigmented substances or biochromes [31] and the toxins [75, 155, 184].

Only a few secondary metabolites are synthesized de novo by animals, sometimes probably with the help of symbiontic microorganisms. Others find their way into animals via the **food-chain** and are stored temporarily, either unmodified or after metabolic conversion. Many insects are poisonous, and so protected from predators, as the result of the ingestion and storage of plant substances in their diet: examples include cardiac glycosides of the cardenolid type (Fig. 19.1a) in the locust *Poekilocerus bufonius*, the aphid *Aphis nerii*, the bug *Oncopeltus fasciatus*, the green lacewing *Chrysopa* sp. (Neuroptera) and the monarch butterfly *Danaus plexippus;* aristolochic acids (Fig. 19.1b) in the lepidopteran family Papilionidae; pyrrolizidine alkaloids (Fig. 19.1 c) in the Arctiidae and Noctuidae; and mustard oils (Fig. 19.1 d) in the Pieridae [24, 56, 75]. In contrast to most products of primary metabolism, secondary metabolites are often found only in

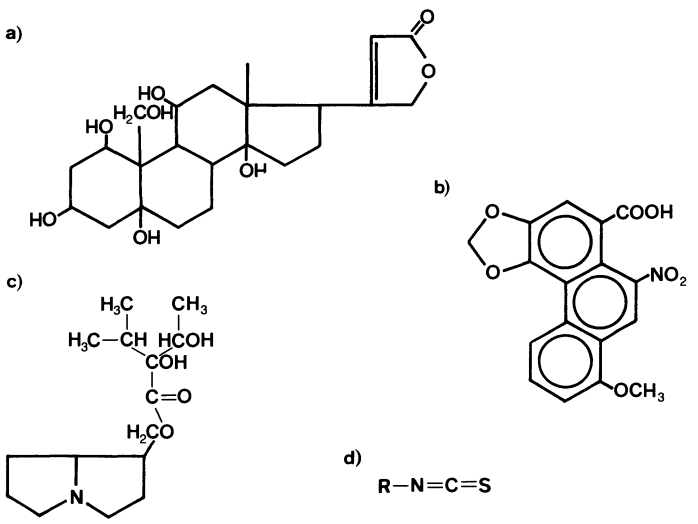


Fig. 19.1a–d. Plant toxins accumulated by insects [24, 75]. **a** Ouabagenin (cardenolide), the aglycone of the heart glycoside ouabain; **b** aristolochic acid; **c** lindelophin (pyrrolizidine alkaloid); **d** mustard oil (general formula)

certain cells or organs, e.g. in glands or in the integument. They often take the form of highly complicated multi-component systems. For example, more than 68 volatile substances have been isolated and identified from the red forest ant *Formica rufa*, more than 50 from the Dufour glands of the ant *Lasius niger*, and about 45 from the preorbital gland of the antelope *Cephalophus monticola* [8, 34, 67].

Although they are produced in side reactions of cell metabolism and often only in small amounts, some secondary metabolites fulfil important **functions** in the life of the animal, e.g. as end products or excretions, as biochromes or luminous substances, as toxins and defence substances, or as chemical signals. The signal substances can be classified according to the species and their effects: interspecific effects (allelochemical effects) to the advantage of the producer are associated with the allomones (e.g. defence substances, toxins, prey attractants), and to the advantage of the receiver in the case of the kairomones (warning and deterrent substances). The pheromones have intraspecific effects (sexual attractants, social pheromones, alarm substances, marking substances). Many substances have no known function, whereas others are clearly vital. In this situation, the term "secondary" metabolite is nothing more than a convenient way of solving the awkward problem of the classification of the numerous different partly species-specific low molecular weight substances according to their biological functions. The presence of a certain substance in an animal may have **a selective value** in a specific function, but it may also simply indicate that this animal species can tolerate a particular allelochemical component in its vegetable diet, thereby achieving an advantage over competitors for a source of food. Or it may be that the species is resistant to a self-produced defence substance. Thus, for example, the locust *Poekilocerus bufonius*, which feeds on spurge, is 300-fold less sensitive to the cardiac glycoside ouabain (Fig. 19.1a), a highly effective inhibitor of Na^+, K^+ ATPase, than are related species with another food plant. The beetle *Caryedes brasiliensis* can even use L-canavanine (see Fig. 12.2, p. 410) as a nitrogen source, whereas this structural analogue of arginine is toxic for most other insects. The cyanogenic butterflies of the genus *Zygaena* are resistant to HCN [15].

The uptake or production, storage, release and effect of secondary metabolites involve many different themes of functional morphology, cell biology, physiology, toxicology and ecology, e.g. the construction and function of glands, the effects of toxins and defence substances on the target organism, detoxification of allelochemicals, the dispersal of pheromones into the environment and their perception by the receiver of the signal. All these aspects are dealt with in the many relevant textbooks and monographs available. In accordance with the aims of this book, the secondary metabolites of animals will be dealt with primarily according to their chemical structure and not to their function or biological roles. In the process, it is possible to choose only individual examples from the several thousand known compounds which have some special biochemical or biological features. The subdivisions of this chapter are based upon the simple classification into N-free and N-containing substances.

Classes of compounds with particularly interesting structural or functional features are dealt with in separate sections.

19.1 Nitrogen-Free Substances

19.1.1 Aliphatic Compounds

Aliphatic hydrocarbons are present in much greater variety in insects than in any other animal group. The cuticular lipids of the insects contain a species-specific spectrum of unbranched and branched alkanes and alkenes, mostly with 20–40 C atoms (p. 592). In addition to their protective function against moistening, the cuticular hydrocarbons also have communication functions. Due to the extremely low volatility of these hydrocarbons, direct contact between the individuals is probably necessary. In female houseflies *Musca domestica*, the main component of the contact pheromone which induces the males to copulate is (Z)-9-tricosene (Fig. 19.2a). This effect is amplified by C_{30}-methylalkane, (Z)-9,10-epoxytricosane and (Z)-14-tricosen-10-one (Fig. 19.2b, c). (Z)-9-Tricosene and the methylalkanes arise by the elongation and decarboxylation of unsaturated and methyl-branched fatty acids, and the epoxide and ketone arise from (Z)-9-tricosene by the action of a cytochrome P-450 monooxygenase [1]. Cuticular hydrocarbons with a similar structure and function are known from many Muscidae, and also from dipterans of other families and from other insects [27].

Research into **sex pheromones** began with the isolation by Butenandt and coworkers in 1959 of 6 mg of the scent (E, Z) 10,12-hexadecadienol (bombycol, Fig. 19.2g) from 500000 silkworm females. Since then, the female sex pheromones of numerous other Lepidoptera have been identified, in particular from species of economical importance as pests of various crops. In general, they are usually derivatives of unbranched C_{10} to C_{18} alcohols, either monounsaturated, rarely saturated, or multiply unsaturated. The most frequent derivative is the acetate, but free alcohols and aldehydes are also found. The number of possible structural variants is large, but is still not sufficient to denote specifically every species. Hence, species specificity of the signals is, as a rule, achieved by **multi-component systems**. Usually, the individual components of such systems differ from each other in only one structural feature: an unsaturated alcohol is esterified with acetic acid or oxidized to the aldehyde; the double bond is shifted by two positions; a C_2 subunit is introduced into the chain; a double bond is saturated such that the saturated compound is formed from the monoene and the monoene from the diene; the geometry of the double bond changes from Z (= cis) to E (= trans) or vice versa [27, 182]. Relationships based upon such a one-step model can already be seen in the pheromone mixture of the silkworm moth, which in addition to (E, Z) 10,12-hexadecadienol also contains (E, E) 10,12-hexadecadienol and (Z, E) 10,12-hexadecadienol [21]. Even larger multicomponent systems are known from other Lepi-

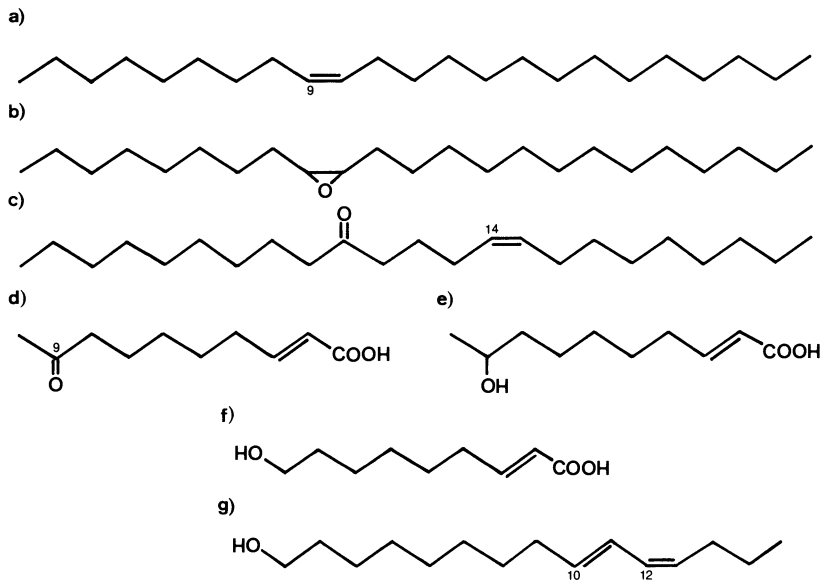


Fig. 19.2a–g. Pheromones of the housefly *Musca domestica* **(a–c)**, the honeybee *Apis melifera* **(d–f)** and the silkworm *Bombyx mori* **(g)** [27, 182]. **a** (Z)-9-tricosene (muscalure); **b** (Z)-9,10-epoxytricosane; **c** (Z)-14-tricosen-10-one; **d** 9-oxo-(E)-2-decenoic acid; **e** 9-hydroxy-(E)-2-decenoic acid; **f** 10-hydroxy-(E)-2-decenoic acid; **g** (E, Z)-10,12-hexadecadienol (bombycol)

doptera, and the individual components are all related by such one-step changes (Fig. 19.3). As experiments with *Manduca sexta* have shown, not all components of the pheromone blend are essential for triggering the complete behavioural sequence (anemotaxis, approach and contact with the pheromone source, attempts at copulation) [189]. The relative proportions of the individual components are very important for the species specificity of the signal effect, although there may be differences between widely separated populations of the same species [15]. The amount of sex pheromone released by a female is only about 1–100 ng/h. The sensitivity of the receptors on the antennae of the males is correspondingly very high; arousal may result from but a single molecule [15]. The individual pheromone components may affect different sense cells, e.g. silkworm males possess different receptors for bombycol and bombycal. Olfaction requires specific proteins for transport, reception and degradation of the scent molecule. Pheromone-binding proteins (PBPs) of 142 amino acids have been identified in the antennae of male *Antherea pernyi, A. polyphemus* and *Manduca sexta;* these are homologous between these species but have no similarity to the olfactory proteins of the rat and frog [150]. A protein with limited homology to the male PBP is expressed in the antennae of the females of *A. pernyi* [28]. The degradation of the species-specific sexual attractant in male antennae occurs by esterification, oxidation and ester formation to aldehydes, acids and esters with long-chain fatty acids. Several antennae-specific enzymes, such as aldehyde oxidases or dehydrogenases, have been described in different lepidopteran species [95, 147, 183].

The pheromone components apparently arise from the common fatty acids, which in a species-specific manner, are desaturated, shortened, reduced to alcohols and esterified with acetate. The **biosynthesis of pheromones** has been studied

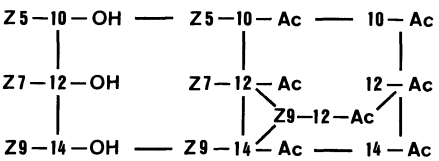


Fig. 19.3. The components of the sex pheromones of *Agrotis segetum*. The *number* after the *Z* indicates the position of the double bond, which is always in the Z configuration; the *other numbers* give the number of C atoms. *-OH* designates alcohol and *-Ac* its acetic acid ester. Z5-10-OH is therefore (Z)-5-decenol. All the components are derived from each other by one-step reactions [27]

by tracer experiments with various lepidopteran species. From this it is clear that most of the known pheromone components can be produced by a few desaturation reactions in combination with chain shortening (limited β-oxidation) or chain elongation. Some of these desaturation reactions have not been observed in any other eukaryote. Δ^{11}-Desaturation has been shown to be a key step in the production of Δ^{11}, Δ^{9}, Δ^{7} and Δ^{5} components in several species. The Δ^{11}-desaturase of most lepidopteran species produces Z and E isomers in certain proportions, whereas that of *Trichoplusia ni* produces only the Z isomer. A Δ^{19}-desaturase is involved in the synthesis of (Z)-8-tetradecenyl acetate and (Z)-10-tetradecenyl acetate in *Planototrix excessana*. In *Thaumatopoea pityocampa*, the Δ^{11}-desaturase produces a triple bond and is responsible, together with a Δ^{13}-desaturase, for the synthesis of cis-13-hexadecen-11-ynyl acetate. The synthesis of (E)- and (Z)-12-tetradecenyl acetate in *Ostrinia furnacalis* occurs by Δ^{14}-desaturation of palmitic acid followed by a chain shortening [5, 27, 66, 123, 212]. Whereas most Δ^{11}-desaturases are not strongly stereo-specific, the acetyltransferases of many lepidopterans are highly specific for the Z isomer of their monounsaturated fatty alcohol substrates [93].

Little is known about the female **sex pheromones in other insect groups**. The pheromones of the honeybee have been investigated in some detail, especially the "queen substance" 9-oxo-(E)-2-decenoic acid (Fig. 19.2 d) which is produced in the mandibular glands of the queen bee. Together with the reduced form (9-hydroxy-(E)-2-decenoic acid; Fig. 19.2 e), this constitutes the sexual attractant for the drones during the marriage flight, but also functions as an attractant for the workers in the hive; it inhibits the development of ovaries and prevents the construction of queen cells. The related compound 10-hydroxy-(E)-2-decenoic acid (Fig. 19.2 f) is produced in large amounts in the mandibular glands of workers and is the main component of larval food [13]. 1,2-Dioleyl-3-palmitoyl glycerol brings about the attraction and clustering of workers around the queen [112]. Beeswax owes its characteristic odour to the presence of oxygenated volatiles which originate partly from pollen and nectar and, like the main component decenal, are produced in part by the workers [23]. Long-chain hydrocarbons, ketones, acids and esters, which in most species are methyl-branched, serve as sex pheromones in the females of the Coleoptera and Homoptera [109,

182]. An interesting case is the production of (Z)-9-tetradecenyl acetate, (Z)-9-tetradecenal and (Z)-11-hexadecenal by the spider *Mastophora cornigera* as components of a mixture of scents that attract prey [175]. The males of some lepidopteran species also produce pheromones but their signal functions have been defined in only a few cases. They are mostly completely different classes of compounds from those produced by the females, e.g. terpenes, and aromatic or heterocyclic compounds [182]. No fewer than 63 different carboxylic acids have been identified together with N-heterocyclic (pyrrolizine) compounds in the male abdominal hair pencils of different species from the genus *Amauris* (Danainae) [163]. Polychaetes of the family Nereidae exhibit a special reproductive behaviour (swarming and nuptial dance) which is triggered by signal substances in the coelomic fluid. 5-Methyl-3-heptanone has been identified as a sex pheromone in *Platynereis dumerilii* [211]. Fatty acid derivatives can also serve as **alarm substances** [22]. This is the function of (E)-2-hexenal and (E)-2-octenal in the bedbug and other Hemiptera, whereas in the honeybee it is usually the acetates of short-chain fatty acids that have this function. Short-chain esters and ketones are also found as alarm signals in other bees and some ants. Otherwise, terpenes often have this function, e.g. in mites (Acari), homopterans, soldier termites, many ants, and bees of the genus *Trigona* [22]. Methyl-branched short-chain alcohols and ketones are typical defence substances of the suborder Palpatores of the craneflies (Opiliones), whereas the Laniatores produce alkylbenzoquinones and phenols [57].

Free formic acid is the most important toxin and defence substance of ants from the subfamily Formicinae, some carabids (Carabidae) and moth caterpillars from the family Notodontidae [27]. Formicin toxins contain up to 60% formic acid [13]. In the defence secretions, the acid is usually mixed with long-chain hydrocarbons, ketones, alcohols or esters which facilitate penetration of the arthropod cuticula. Acetic acid also occurs, together with formic acid, in defence secretions, e.g. in the carabid *Thermophilium homoplatum;* it is especially characteristic, together with acetic acid esters, as the defence substance of many bugs from the family Coridae [27]. Short-chain fatty acids and their derivatives are usually not produced from longer-chain fatty acids but from amino acids. Thus, in the biosynthesis of formic acid a C_1 fragment of serine or glycine is transferred to tetrahydrofolic acid, and formiate is released by the action of 10-formyltetrahydrofolate synthetase (Fig. 19.4). Methacrylic acid and isobutyric acid are produced from valine by the carabid *Carabus taedatus* and caterpillars of the butterfly *Papilio aegus*, respectively; ethacrylic acid and α-methylbutyric acid are produced from isoleucine [27]. The biosynthesis of the 2-ketones (methylketones) and 3-ketones (ethylketones), e.g. 2-heptanone or 3-methyl-3-heptanone, in the secretions of ants has not yet been explained [20].

Amongst the common long-chain **hydroxy acids**, myrmicacine (3-hydroxydecanoic acid) is of special interest; this is secreted by the fungus-rearing, leaf-cutting ant *Atta sexdens* and is distributed in the nest together with phenyl and indole acetic acid, where it inhibits germination of the spores of all fungi except the desired species [13]. Hydroxy acids can produce ring-formed **lactones** by internal ester bonding, and these are found in the secretions of ants, bees of the genus *Trigona* and beetles of the genus *Bledius* (Staphylinidae) (Fig. 19.5a). An unsaturated lactone has the function of a sex pheromone in the beetle *Popillia japonica* (Fig. 19.5b); only the (R) isomer illustrated is found in the animal and the synthetic (S) isomer is inhibitory [27]. The 4-hydroxy acid lactones in the orbital gland secretions of various deer also function as sex pheromones [13, 34]. ω-Hydroxy acids can give rise to lactones with many-membered rings (macrocyclic lactones) (Fig. 19.5c). Such compounds are known from the Porifera, from the defence secretions of some termites and beetles, and from the secretions of the Dufour glands of the bee families Halictidae and Colletidae, where they occur together with the corresponding ω-hydroxy acids [27]. The aplysia toxin from the sea snail *Styocheilus longicauda* and other "sea hares" is also a macrocyclic lactone, coupled to a bromophenol (Fig. 19.5d).

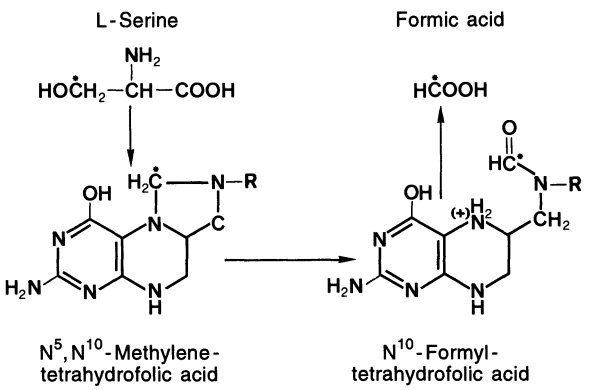


Fig. 19.4. The biosynthesis of formic acid from L-serine [27]

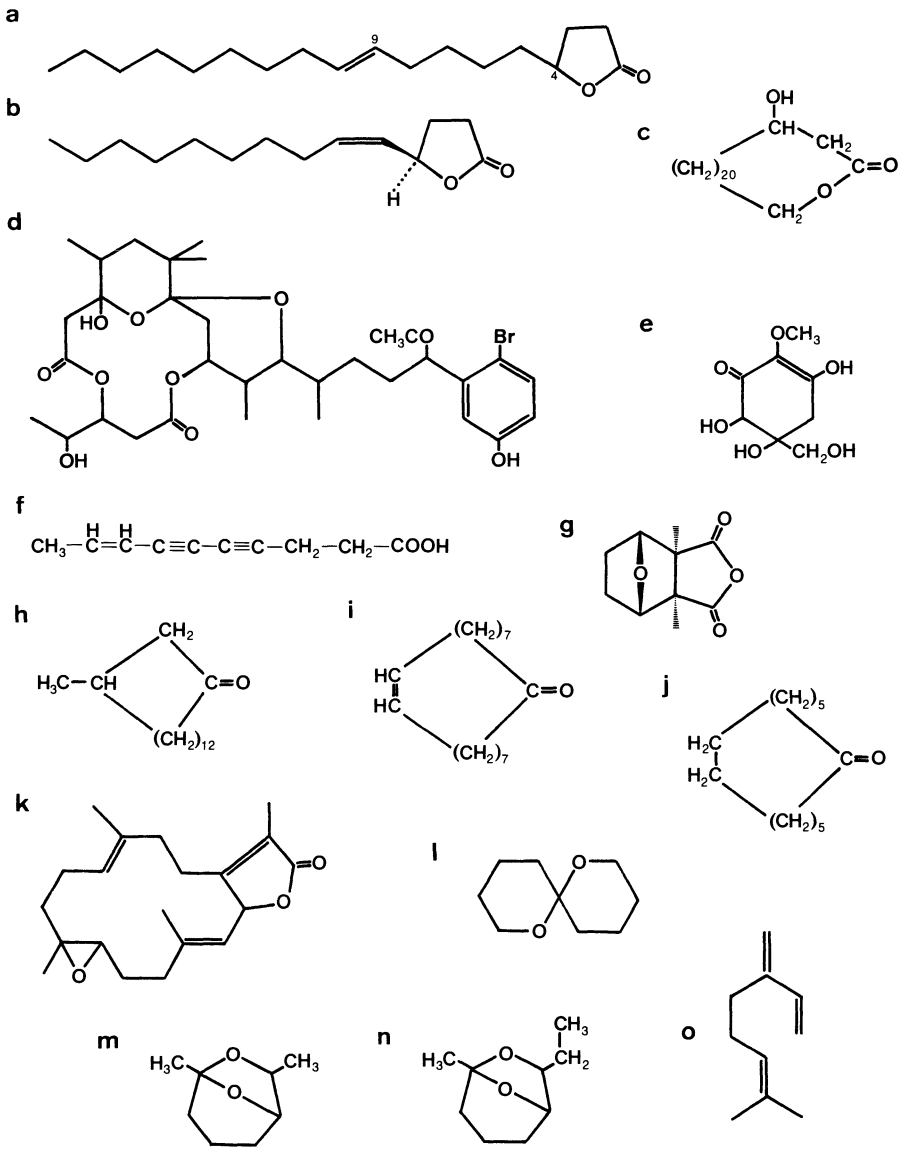


Fig. 19.5 a–o. Nitrogen-free substances. **a** 4-Hydroxy-(E)-9-octadecenoic acid lactone (4-octadec-9-enolide) from *Lasius flavus* [27]; **b** 4-hydroxy-(Z)-5-tetradecenoic acid lactone from *Popillia japonica* [27]; **c** 3-hydroxy-24-tetracosanolide from *Armitermes* sp. [27]; **d** aplysia toxin from *Stylocheilus longicauda* [155]; **e** gadusol from marine teleosts [72]; **f** dihydromatricaria acid from the beetle *Chauliognathus lecontei* [27]; **g** cantharidine from the oil beetle (Meloidae) [75]; **h** muscone from the musk ox [13]; **i** civettone from the civet [13]; **j** exaltone from the muskrat [13]; **k** sarcophine from the gorgonian *Sarcophyton glaucum* [155]; **l** 1,7-dioxaspiro-5,5-undecane from the olive fly *Dacus oleae* [27]; **m–o** frontaline **(m)**, exo-brevicomine **(n)** and myrcene **(o)** from *Dendroctonus brevicomis* [13]

Certain mussel and coral toxins are based upon complicated **polyether structures**. The occasional poisoning epidemics that result from the consumption of mussels are caused by toxins from the algal diet of the mussel. The N-containing saxitoxin from the dinoflagellates of the genus *Gonioaulax* (p. 743) has been known for some time. However, N-free compounds with a very different structure are produced by dinoflagellates and accumulated in mussels and these can also lead to poisoning symptoms. Amongst these are polyether lactones, such as dinophysistoxins from the dinoflagellate *Dinophysis acuminata* (Fig. 19.6a), the similarly constructed brevetoxins from *Gymnodinium breve*, and the pectenotoxins (Fig. 19.6b) from an as yet unknown dinoflagellate species [11, 61, 104]. The most complicated and the most poisonous of all marine toxins is the palytoxin of the crust coral (Zoantharia) *Palythoa toxica*, which inhabits the coastal waters of

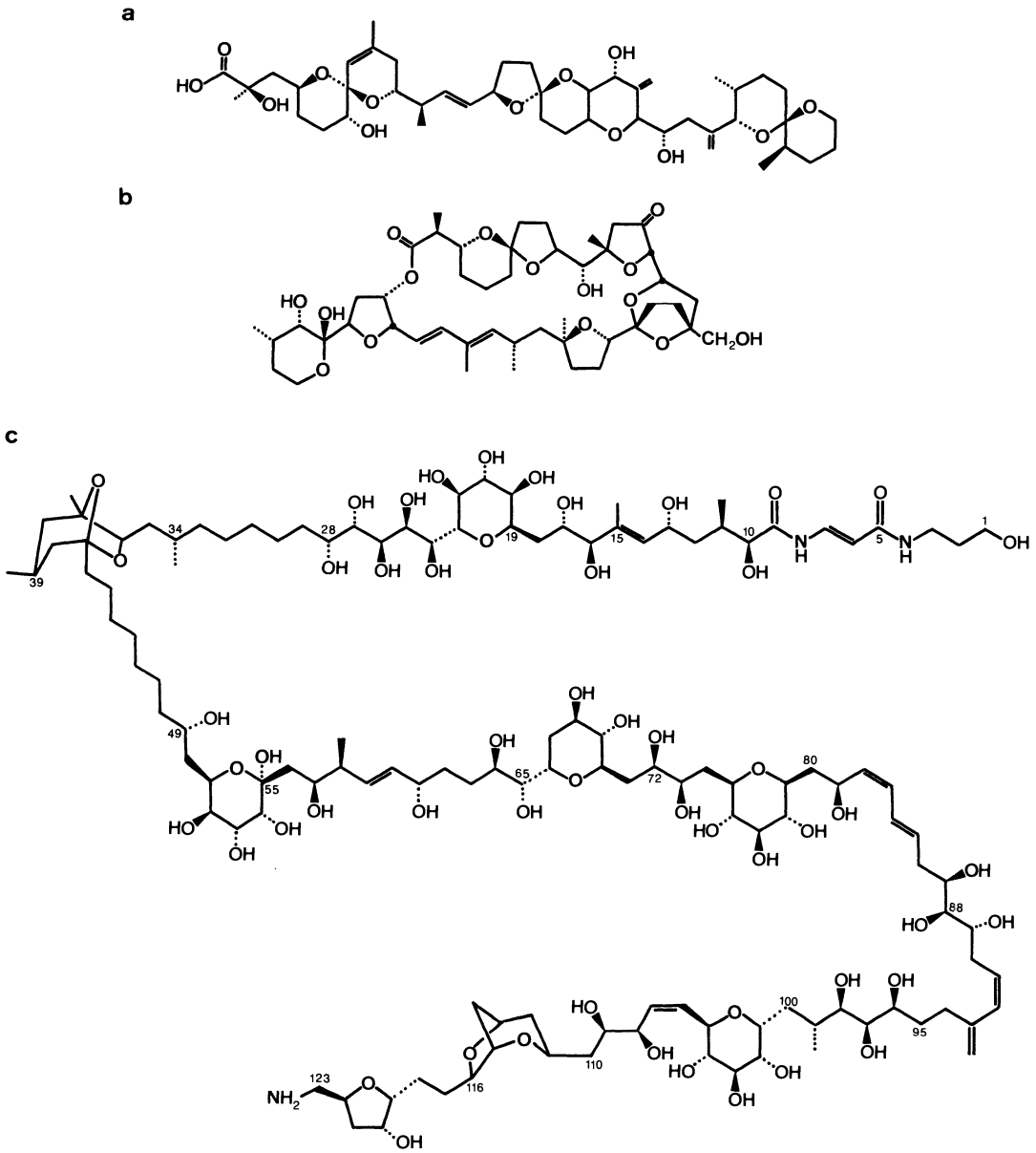


Fig. 19.6 a–c. Mussel and coral toxins. **a** Dinophysistoxin 1 [104]; **b** pectenotoxin [104]; **c** palytoxin [135]

Hawaii, and other Palythoa species (Fig. 19.6 c). Definition of the structure of the 115 C-atom molecule proved to be extraordinarily difficult and was first achieved in 1981. Palytoxin has a mammalian LD_{50} of 10–100 ng/kg and is thus the most potent of all animal poisons. It binds to Na^+,K^+-ATPase, converting it into a permanently open channel which leads to lethal ion loss through the plasma membrane [74, 135].

Some aliphatic substances have unusual structural properties. For example, **triple bonds** are seldom found between carbon atoms in animal compounds. However, such a bond is created by the processionary moth *Thaumatopoea pityocampa* during the biosynthesis of its pheromone (Z)-13-hexadecen-11-ynyl acetate [5]. The oxytoxins produced by the sea snail *Oxynoe olivacea* (Opisthobranchia) from precursors present in its food alga *Caulerpa prolifera* also contain triple bonds [41], as does the dihydromatricaria acid (Fig. 19.5 f) obtained by the soldier beetle *Chauliognathus lecontei* from its plant diet (Compositae) and used unmodified as a defence substance [27]. The anal gland secretions of the North American skunk *Mephitis mephitis*, its South American counterpart *Conepatus suffocans* and

the Philippine *Mydaus marchei*, all of which belong to the marten family (Mustelidae), contain stinking mixtures of **mercaptans and sulphides** [13]. Dimethylsulphide and dimethyltrisulphide are also found in the defence glands of the stinging ants (Ponerinae), where they are apparently synthesized from methionine [27, 75]. **Nitro-substituted substances**, such as the 1-nitro-(E)-1-pendadecene in the defence secretion of the termite *Prorhinotermes simplex*, are quite rare [13, 27].

19.1.2 Carbocyclic Compounds

The secondary metabolites include various **cyclohexane derivatives**, e.g. gadusol, which is present in the ovaries of marine teleosts at levels of up to 4 g/kg dry weight (Fig. 19.5 e); similar compounds are present in corals, mussels, and the brine shrimp *Artemia salina*, but their origin is in all cases unknown [72]. The cyclohexane derivative cantharidine (Fig. 19.5 g), found as a defence substance in the "Spanish fly" *Lytta vesicatoria* and other oil beetles (Meloidae), has been known for centuries as an aphrodisiac; the use of this highly toxic compound has unfortunately caused many deaths [75]. The musk substances of mammals, such as the muscone of the musk deer *Moschus moschiferus*, the civettone of the civet *Viverra zibetha*, or exaltone, which is found together with civettone in the muskrat *Ondatra zibethica*, are **macrocyclic ketones** (Fig. 19.5 h–j). Similar substances have been found in the duck *Anas moschata*, the turtle *Kinosternon pennsylvanicum* and several alligators [13]. Sarcophine (Fig. 19.5 k) from the alcyonarian corals is an effective fish poison and is one of the most complicated macrocyclic substances [155]. **Spiranes** are bicyclic hydrocarbons whose two rings have only one C atom in common. Spiroacetals and spiroketals are found together with terpenes as components of the sex pheromones in some species of the Hymenoptera (e.g. *Andrena* sp.), Coleoptera and Diptera [27]. The simplest spiroacetal is 1,7-dioxaspiro-(5,5)-undecane of the olive fly *Dacus oleae* (Fig. 19.5 l). The scolytid beetle (Ipsidae) species *Dendroctonus brevicomis* is attracted by an aggregation pheromone made up of three synergistic components (Fig. 19.5 m–o). This includes two related **bicyclic ketals**, the frontaline produced by the males and exo-brevicomine produced by the females, and the monoterpene myrcene produced by the infested Ponderosa pine [13, 15]. The aggregation pheromone of other scolytids, e.g. *Ips paraconfusus*, is made up of various terpenes (p. 726).

19.1.3 Aromatic Compounds

The majority of aromatic substances in animals originate from aromatic components of the diet, in particular the amino acids phenylalanine and tyrosine. However, certain aromatic compounds can be synthesized by animals from non-aromatic precursors, whereby it is assumed that the intermediates are **polyketides** formed by the aldol condensation of four or more acetate residues (Fig. 19.7). Tracer experiments have been carried out on, for example, the beetle *Eleodes longicollis*, whose defence secretion contains three different 1,4-quinones (Fig. 19.8 a–c). Of these, benzoquinone can only be produced from phenylalanine or tyrosine, whereas the two substituted quinones were synthesized only from acetate and not from the amino acids [27]. Substituted 1,4-quinones are widely found in the defence secretions of beetles in the families Tenebrionidae and Carabidae. The primary product of the polyketide pathway is probably 6-methylsalicylate (Fig. 19.7), which is found in ants and the carabids; the corresponding aldehyde has been reported in a beetle from the family Cerambycidae. Isocoumarins arise by the polyketide pathway in plants and fungi; their origin in the insects is in most cases unclear. Mellein (Fig. 19.8 d), which is a member of this class of substances, is a typical metabolic product of fungi but is also found in ants and lepidopterans [13]. In the wax moth *Aphomia sociella*, mellein is the major component of the male sex pheromone which triggers search behaviour in the females. The gut of larvae and the infested bumble bee nests contain the mellein-producing fungus *Aspergillus ochraceus*, which apparently supplies this pheromone component [106].

The amino acids phenylalanine and tyrosine have a C_6–C_3 backbone; stepwise shortening of the side-chain leads via C_6–C_2 and C_6–C_1 com-

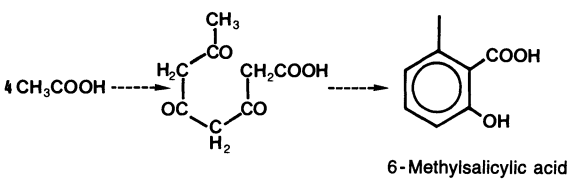


Fig. 19.7. The biosynthesis of aromatic compounds by the polyketide pathway [27]

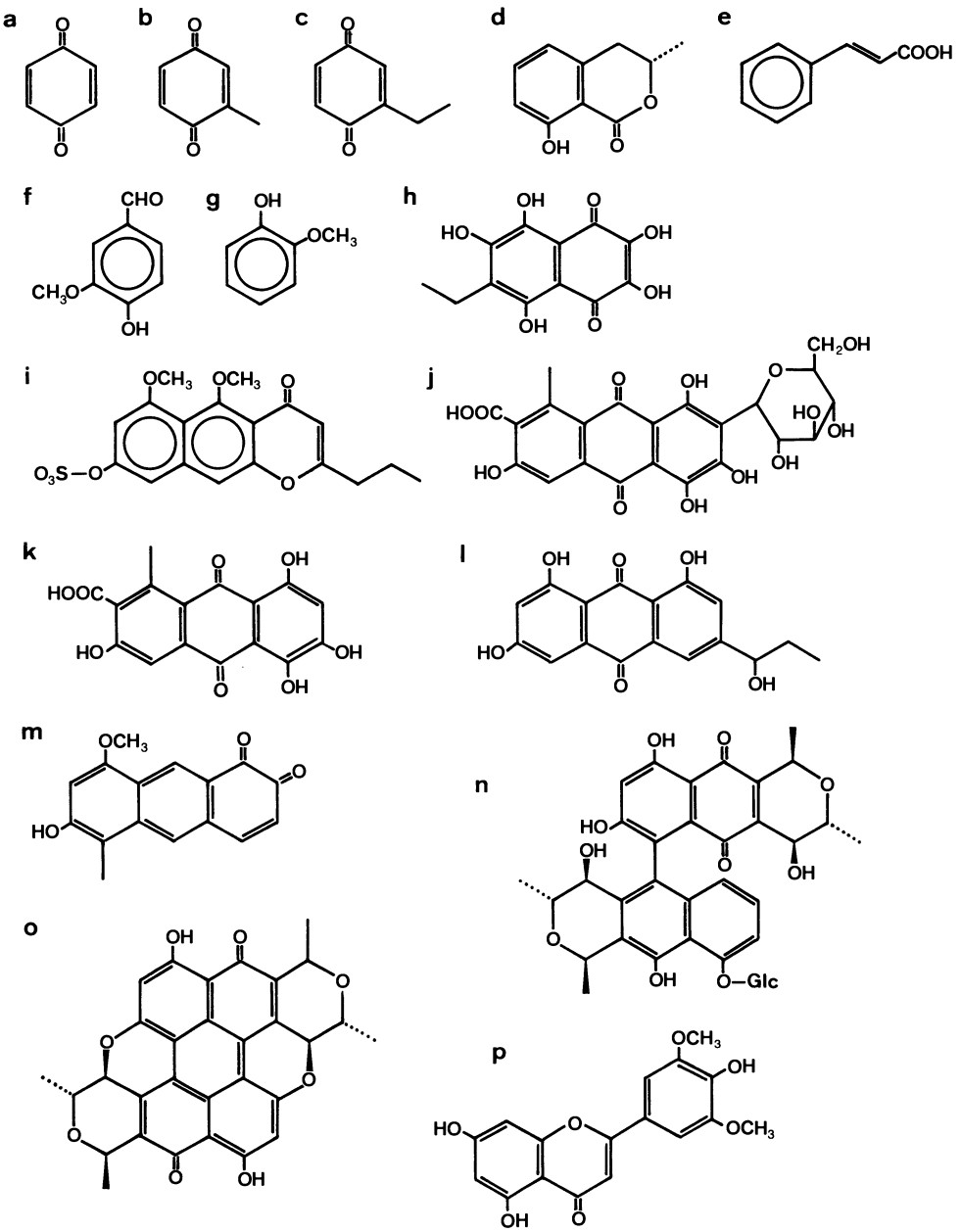


Fig. 19.8 a–p. Aromatic substances [27, 31, 75]. **a** Benzoquinone; **b** toluquinone; **c** ethyl benzoquinone; **d** mellein; **e** trans-cinnamic acid; **f** vanillin; **g** guaiacol; **h** echinochrome A; **i** a polyketide ester from the crinoid *Comatula* sp.; **j** carminic acid; **k** kermesinic acid; **l** rhodoptilometrin; **m** hallachrome; **n** protoaphine from *Aphis fabae*; **o** erythroaphine from *Aphis fabae;* **p** tricine from *Melanargia galathea*

pounds to unsubstituted aromatic substances. Naturally occurring substances with a C_6–C_3 backbone are relatively rare; esters of trans-cinnamic acid (Fig. 19.8 e) or the corresponding alcohol have been identified in the pheromones of male lepidopterans and bugs [27]. C_6–C_2 compounds can arise by decarboxylation from phenylalanine and tyrosine. The most common substance of this type is **phenylethanol**, which is a pheromone component of male Lepidoptera, Hymenoptera and Hemiptera, and is also part of the aggregation pheromone of the scolytid beetle *Dendroctonus breviconis*. Phenylethylester and phenylacetaldehyde have been shown to be pheromone components in these insect groups; phenylacetic acid is dispersed as a bactericide in the nest of the leaf-cutting ant *Atta sexdens*, but is also present in a gland secretion of the gerbil *Meriones unguli-*

culatus where it is used to mark territory [13, 27]. 2′-Hydroxy-4′-methoxyacetophenone and the corresponding propiophenone derivative (C_6–C_3) are part of the defence secretion of the beetle *Tribolium castaneum;* they are very effective inhibitors of prostaglandin synthesis in vertebrates and insects [92].

Benzoic acid arises in plants by β-oxidation of cinnamic acid (Fig. 19.8e) and thiolytic cleavage to acetate and benzoyl-CoA; the benzoic acid released can be reduced to benzaldehyde and benzyl alcohol. It is not clear to what extent these reactions are possible in arthropods or whether the corresponding C_6–C_1 compounds must be taken up in the diet. **Benzyl aldehyde**, however, is widely found, e.g. in millipedes, beetles, ants, bees and lepidopterans [18, 27, 182]. It is produced in certain millipedes and beetles together with HCN by the cleavage of mandelonitrile (Fig. 19.16a). Benzoic acid, *p*-hydroxybenzoic acid, 3,4-dihydroxybenzoic acid (protocatechuic acid) and their esters are present in the defence secretions of water beetles (Dytiscidae) [27]. Protocatechuic acid and other substituted benzoic acids are involved in insects in the tanning of non-cuticular proteins (p. 431). The pleasant-smelling vanillin (Fig. 19.8f) is part of the male pheromones of the moth *Eldana saccharina* and the bug *Leptoglossus corculus* [27, 182].

Oxidative decarboxylation of *p*-hydroxybenzoic acid gives **1,4-hydroquinone**, which is easily converted to **1,4-benzoquinone**. Both compounds are common in defence secretions, especially in beetles, but also in termites, locusts and cockroaches [27]. Diplopods of the orders Julida, Spirobolida and Spirostrepida could just as well be termed "quinone millipedes". Handling these animals often produces red staining of the fingers owing to protein tanning by quinones [18]. That the frequently substituted 1,4-benzoquinones of beetles in the families Tenebrionidae and Carabidae can arise only by the polyketide pathway was mentioned above. In many insects, the very reactive and toxic 1,4-hydroquinones are present as β-glucosides, from which they are released by glycosidase. The bombardier beetle and carabids from the subfamilies Brachininae and Ozaeninae exploit the high reactivity of the hydroquinone in a remarkable defence mechanism. The defence gland secretes a mixture of about 10% hydroquinone and 25% hydrogen peroxide into an internal storage cavity. The brief opening of a separating flap allows a portion of this mixture to enter a chitin-lined reaction chamber, where the action of catalase and peroxidase produces oxygen and 1,4-quinone in an explosive reaction; the hot aqueous solution is ejected forcibly through the opening of the gland to the outside. The enzymes involved are adapted to the high temperatures produced in the reaction chamber [27, 52, 75].

Phenols are relatively rare amongst the animal substances. Unsubstituted phenol is secreted by females of the beetle *Costelytra zealandica* and attracts males. Conversely, it is a major component of the defence secretions of the cockroach *Archiblatta hoeveni* and the locust *Romalea microptera*. It is accompanied by ***p*-cresol** in both species and in the locust also by ***o*-cresol**. This suggests that the biosynthetic pathways of phenol, *p*-cresol and *o*-cresol are linked. The completely different distribution of ***m*-cresol**, which is found mainly in carabids, indicates a different pathway, probably via polyketides. Phenol and *o*-methoxyphenol (guaiacol, Fig. 19.8g) are found in millipedes of the order Polydesmide, where they have been shown to be synthesized from tyrosine; guaiacol is also found in the Hemiptera [18, 27]. The females of various tick species (Ixodidae) attract males by 2,6-dichlorophenol; the cholesteryloleate of the body cuticula then functions as a sex pheromone to induce mounting [77]. Marine algae produce different bromophenols and these are also found in marine animals such as polychaetes, phoronids and hemichordates. The 2,4-dibromophenol from *Saccoglossus kowalewskii* has bactericidal properties [99].

The red, purple and blue pigments of the sea urchins, starfish and crinoids are known as echinochromes or spinochromes; they are **naphthoquinones** and exist as about 20 variants. They were first discovered in spines and calciferous shells, where they contribute greatly to the coloration, but they are also present in somatic cells [31]. Echinochrome A from the coelomocytes of the sea urchin *Echinus esculentus* (Fig. 19.8h) has bactericidal activity [167]. The polyketide sulphates produced as several variants by the crinoids *Comatula pectinata* and *C. perplexa* (Fig. 19.8i) have two condensed aromatic rings; these are not particularly toxic but have deterrent effects against fish [75]. Various important biochromes are found amongst the **anthraquinones**. The C-glucosidic carminic acid (Fig. 19.8j) from the cactus-inhabiting cochineal *Dactylopius coccus (Coccus cacti)* and the closely related kermesinic acid (Fig. 19.8k) from the mealy bug *Kermes quercus* which inhabits oaks, have been used for centuries as dyes. *Ptilometra australis* and other Australian crinoids owe their purple colour to rhodoptilometrin (Fig. 19.8l) and related sub-

stances, which are very similar to some fungal metabolites. Hallochrome (Fig. 19.8 m) from the polychaetes *Halla parthenopeia* and *Lumbriconereis impatiens* (Eunicidae) has the seldom found 1,2-anthraquinone structure [31]. The **aphines** are extremely complicated quinones that are present in a large number of different species-specific structural variants and stereoisomers in the aphids (Aphidae). They arise from protoaphines in the haemolymph (Fig. 19.8 n) and these come from naphthoquinones. After the death of the animal, the red erythroaphines are produced from protoaphine via yellow and orange intermediates (Fig. 19.8 o). All aphines are "stretched quinones" in that the two oxygen atoms are separated by a complicated conjugated multiple-ring system [31]. **Flavonoids**, pigments with the three-ring system of flavone, are common in plants and include important pigments like anthocyans and also the colourless flavones and flavonol. Herbivorous animals consume large quantities of these substances in their diet but usually appear to make no further use of them. Only in the case of some butterflies is it known that flavonoids, taken up at the larval stage, are stored in the wings, e.g. tricine (Fig. 19.8 p) and luxetine by the "chessboard" *Melanargia galathea* [96].

19.2 Terpenes

Terpenes arise by condensation from isopentenyl pyrophosphate, the "activated isoprene" whose synthesis was described earlier (p. 625). The number of coupled isoprene units distinguishes monoterpenes (two units = C_{10}), sesquiterpenes (C_{15}), diterpenes (C_{20}), sesterpenes (C_{25}), triterpenes (C_{30}) and tetraterpenes (C_{40}); these may have a cyclic or an acyclic structure. The terpenes as volatile "ethereal oils" are typical plant substances but numerous terpene derivatives are also found in the animals. Squalene and steroids are triterpenes, and carotenoids are tetraterpenes. In many cases, the terpene derivatives, or their precursors which are found in animals originate in the diet. However, tracer experiments with acetate or mevalonate have clearly shown the capacity for **de novo synthesis of mono-, sesqui- and diterpenes** in ants, termites, lepidopterans and homopterans, although none of these produce squalene or steroids. Citral is apparently an important intermediate of these syntheses [18, 84].

The de novo synthesis of terpenes has also been shown in nudibranch gastropods (Opisthobranchia) but it was not found in the Porifera [9, 40, 84]. The reactions of terpene metabolism in the animal kingdom and the related enzymes are mostly unknown. Terpene metabolism in marine organisms has some interesting features. Whereas land plants produce predominantly volatile monoterpenes, the emphasis in marine algae is on diterpenes. The wealth of halogens, especially bromine, in the marine environment presents many opportunities for terpene cyclization, and this explains the great variety of halogenated terpenes in marine algae; these then find their way into algae-consuming marine animals. For example, 400 different halogen terpenes have been identified in the seaweeds of the genus *Laurencia* [62]. Isocyanoterpenoids are found in the marine poriferans [37]. In many, but certainly not all, cases the animal terpenes can be assigned biological functions: as defence substances or toxins in marine poriferans and slugs [13, 48]; as alarm signals and chemical marking substances in mites, aphids termites, ants and bees; and as sex or aggregation pheromones in scolytid beetles, mealy bugs and butterflies of the family Pieridae [15, 18, 22, 182].

Amongst the **acyclic monoterpenes**, the compounds derived from geranial and neral (the trans- and cis- isomers of citral) and from citronellal are the most widely found (Fig. 19.9a–c). This is not surprising bearing in mind that the biosynthesis of all monoterpenes starts from geranyl pyrophosphate. The three aldehydes, and sometimes the corresponding alcohols and esters, are components of pheromones and defence secretions of mites, bees (e.g. *Trigona subterranea)*, ants (e.g. *Atta sexdens, Acanthomyops claviger)*, butterflies (Pieridae) and beetles (Staphylinidae) [13, 22, 75, 182]. Complex mixtures of all these substances, including joint compounds such as geranylcitronellol or geranylgeranyl acetate, are found, for example, in the labial glands of various species of the bumble bee genus *Bombus* and in the Dufour glands of the bee genus *Andrena* [15]. A new type of acyclic monoterpene is β-acaridial, the sex pheromone of the acarid mite *Caloglophus polyphyllae* [110]. Male scolytids of the species *Ips paraconfusus* secrete in their faeces an aggregation pheromone which contains as synergistic components the acyclic terpenes ipsdienol and ipsenol as well as bicyclic cis-verbenol (Fig. 19.9 d, e, k); the individual components are practically inactive [13, 15]. A total of seven different halogenated acyclic monoterpenes have been identified in the large sea slugs *Aplysia limacina* and *A. californica* (Fig. 19.9 f); three of these

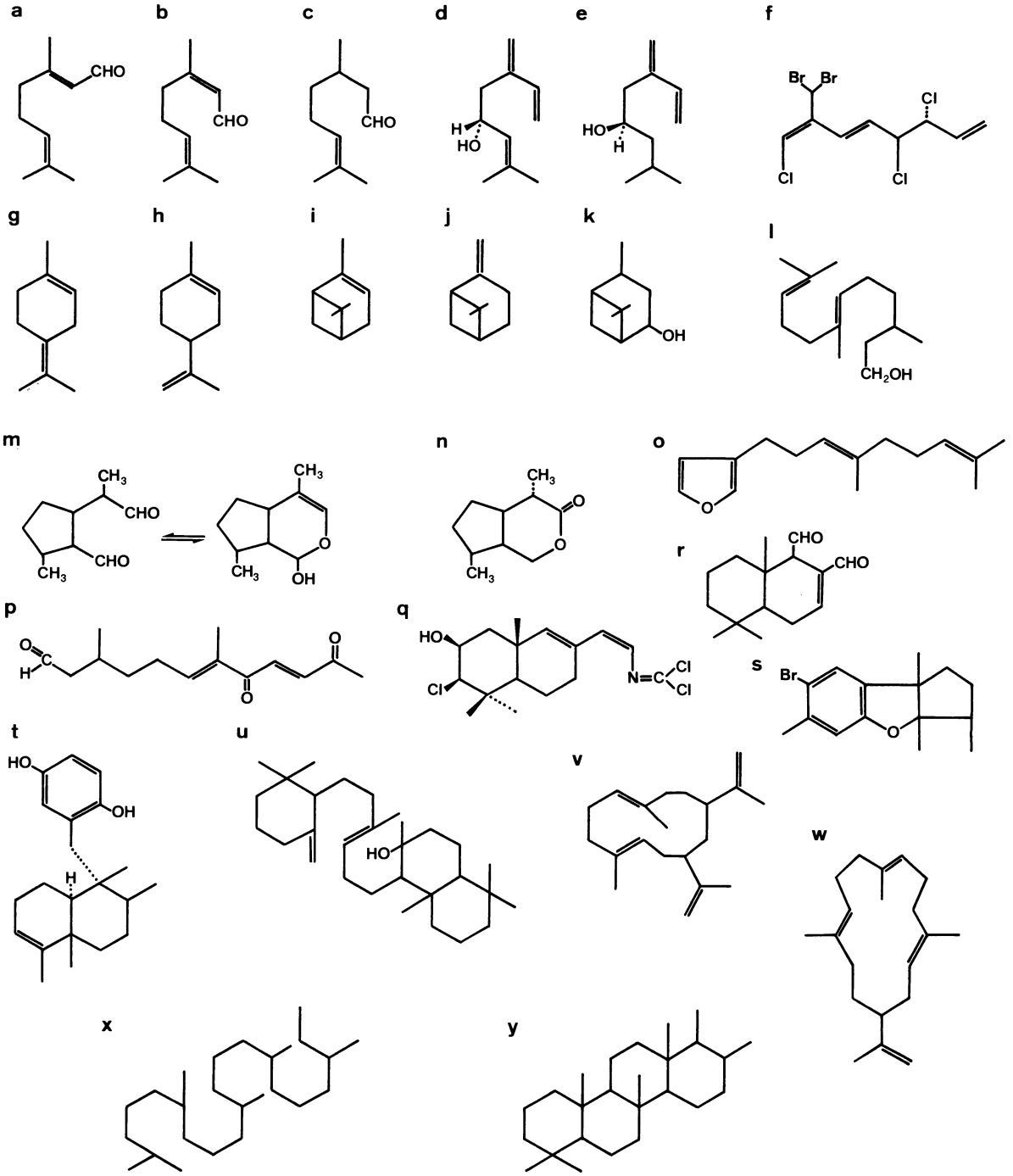


Fig. 19.9 a–y. Terpenes [13, 27]. **a** Geranial; **b** neral; **c** citronellal; **d** ipsdienol from *Ips paraconfusus*; **e** ispenol from *I. paraconfusus*; **f** a halogenated terpene from *Aplysia limacina* [139]; **g** terpinolene; **h** limonene; **i** α-pinene; **j** β-pinene; **k** cis-verbenol from *I. paraconfusus*; **l** farnesol; **m** iridodial from various Dolichoderinae; **n** iridomyrmecine from *Iridomyrmex humilis*; **o** dendrolasine from *Dendrolasius fuliginosus* [75]; **p** gyrinidal from *Gyrinus* sp. [75]; **q** a halogenated terpene from *Pseudoaxinyssa petys* (Porifera) [62]; **r** polygodial [40]; **s** aplysine from *Aplysia kurodai* [155]; **t** avarol from *Dysidea avara* [36, 136]; **u** ambreine from the ambergris of the sperm whale; **v** cubitene from *Cubitermes umbratus* [15]; **w** cembrene A from *C. umbratus* [15]; **x** a marine sesterpene from Porifera and Nudibranchia [48]; **y** a marine sesterpene from Porifera [48]

are also found in algae in the same location [139].

The **monocyclic terpenes** limonene and terpinolene (Fig. 19.9 g, h) serve as components of alarm pheromones and defence substances for the soldiers of various termite species [15, 22]. A dioxo derivative of terpinolene (Fig. 19.9 g), known as robinal, is found in the acarid mite *Rhizoglyphus robini* [111]. The **bicyclic terpenes** α-pinene and β-pinene (Fig. 19.9 i, j) are found in, for example, termites, aphids and the Pieridae, whereas the structurally related cis-verbenol (Fig. 19.9 k) is only found as a component of the aggregation pheromones of *Ips paraconfusus* and related species. Verbenol is apparently produced in these scolytid species by *Bacillus cereus* in the hind gut from α-pinene obtained from the host plant [15, 59]. Many ants exploit terpenes as signal and defence substances. The species *Tapinoma nigerrimum* of the family Dolichoderinae has a rather remarkable defence mechanism. It sprays attackers with a mixture of the ketones methylheptonone and propylisobutylketone and the terpene dialdehyde iridodial (Fig. 19.9 m) which polymerizes and renders them immobile. Many ants of the genera *Iridomyrmex* and *Dolichoderus* have iridodial, whereas others have dolichodial. Iridomyrdmecine (Fig. 19.9 n), which has an insecticide activity superior to that of DDT, is produced by *Iridomyrmex humilis* [13, 15].

The production of **sesquiterpenes** occurs via farnesyl pyrophosphate, which is also an intermediate in the synthesis of the juvenile hormone JH III (p. 613). Hence, the acyclic farnesol (Fig. 19.9 l) is particularly widespread amongst the insects. For example, 2,3-dihydrofarnesol is present in the mandibular gland of the bumble bee *Bombus terrestris;* farnesene and homofarnesene are present in ants of the genus *Formica* [2]; and (E)-β-farnesene occurs as a defence substance in the larvae of the lepidopteran *Papilio helenus* and as a pheromone in aphids [13, 22, 84]. (E, E)-α-Farnesene and (E)-β-farnesene are present at a significantly higher concentration in the urine of dominant male mice [141]. Dendrolasine (Fig. 19.9 o), a modified sesquiterpene, is a defence substance of *Dendrolasius fulginosus;* it particularly affects other ants, and is also found in plants [13, 75]. The (nor)-sesquiterpene gyrinidal (Fig. 19.9 p), which is one C atom shorter, is produced together with a related compound in the pygidial gland of water beetles from the genus *Gyrinus*, and apparently acts as a deterrent against fish [13, 75]. Numerous acyclic and cyclic sesquiterpenes, including many halogenated compounds (Fig. 19.9 q), are known from sea slugs and sponges. The sea slug *Dendrodoris limbata* uses polygodial (Fig. 19.9 r) as a fish deterrent, a substance which it can synthesize de novo [40]. In contrast, the slug *Aplysia kurodai* apparently obtains the poisonous bromoterpene aplysine (Fig. 19.9 s) and the related aplysinol from its algal diet [13]. *Aplysia punctata* contains seven cyclic polyhalogenated monoterpenes which have been shown to originate in the red alga *Plocamium coccineum* [149]. The sesquiterpenoid hydroquinone avarol from the sponge *Dysidea avara* (Fig. 19.9 t) is of medicinal interest because of its pronounced antimicrobial and cytostatic properties [36, 136]. The **triterpene alcohol** ambreine (Fig. 19.9 u) is a major component of ambergris, a wax-like, pathological gut secretion of the sperm whale *Physeter catodon;* this was earlier valued for scent production but has now been replaced by synthetic compounds. Ambergris itself is odourless, but oxidative cleavage produces smaller fragments, several of which have characteristic odours. Terpenes with complicated structures are found in the defence secretions of termite soldiers (Fig. 19.9 v, w) and also in marine sponges and slugs (Fig. 19.9 x, y) [15, 48]. The steroids, steroid glycosides and steroid saponins that serve as toxins and defence substances in fish, insects, starfish, holothurians and corals were described in Chapter 16.

19.3 Carotenoids

Carotenoids are the most common biological pigments; they are found in all groups of organisms from bacteria to the mammals, although not in all species. They are tetraterpenes of eight isoprene subunits whose biosynthesis in plants occurs via the acyclic C_{40} compound lycopene (Fig. 19.10a). The numerous conjugated double bonds which confer the colours to carotenes usually all have the trans configuration. Cyclization produces methyl-substituted cyclohexene rings at both ends. The molecule of β-carotene (Fig. 19.10 b) is built up symmetrically and the C atoms are numbered 1–20 in one half and 1′–20′ in the other half. Whereas the two end groups in β-carotene have the same structure, the isomer α-carotene has a β ring and an ε ring (Fig. 19.10 c). The oxygenated derivatives of the carotenoids are known as xanthopylls. The oxygen is present mainly as hydroxyl groups, as in zeaxynthin and lutein

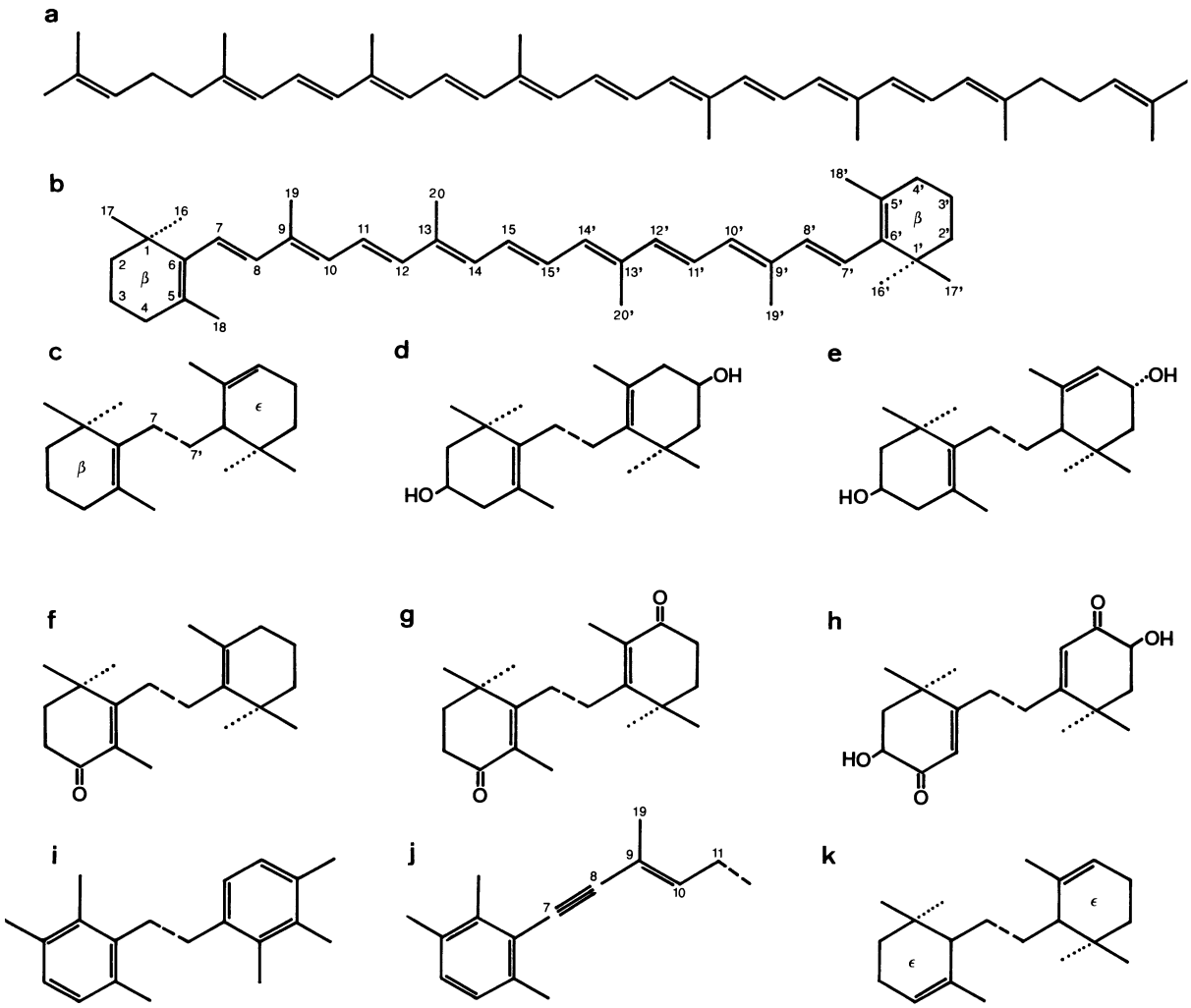


Fig. 19.10 a–k. Carotenoids [71]. **a** Lycopene; **b** β-carotene; **c** α-carotene; **d** zeaxanthin; **e** lutein; **f** echinenone; **g** canthaxanthin; **h** 3S,3′S-astaxanthin; **i** renieratene; **j** 7,8-dihydrorenieratene; **k** ε,ε-carotene

(Fig.19.10 d, e), oxo groups, as in echinenone and canthaxanthin (Fig. 19.10 f, g), or both together, as in astaxanthin (Fig. 19.10 h). The hydroxy carotenes can form esters with long-chain fatty acids. The basic structure of the carotenoids allows numerous variants with different numbers of C atoms, different end groups, different numbers and positions of double bonds and methyl groups, and various oxygen functions. To date, more than 600 naturally occurring carotenoids have been identified, many of which have trivial names. The systematic nomenclature is based on the basic substance "carotene"; the structure of the end groups are denoted by Greek letters. Thus, α-carotene has the systematic name of β-ε-carotene, and zeaxanthin is 3,3′-dihydroxy-β,β′-carotene. The biological roles of the carotenoids depend often, but not always, upon their light-absorbing properties. In plants they are essential for photosynthesis and light protection. In animals they are precursors of the A vitamins (retinols), from which are derived the visual pigments and important cell-signalling molecules such as retinoic acid. However, by no means all the carotenoids detected in animals can be metabolized to vitamin A. Some of the carotenoids that are not suitable as pro-vitamins probably have as yet unknown functions [16], and others originate by chance in the diet and have no function. In fact, many animals survive with no, or almost no, carotenoids [71, 94].

Animals are not able to synthesize carotenoids de novo. In many cases the pigments taken in with the diet are stored without modification. Hence, the same carotenoid may be found in several segments of a food-chain, e.g. in alder lea-

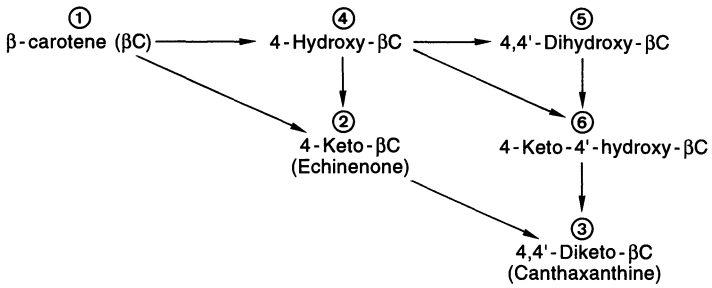


Fig. 19.11. Metabolic pathways from β-carotene to canthaxanthin [71]. See text for further details; the intermediates and end products are referred to in the text by the numbers shown here

ves, in the caterpillars that feed on them and in the great tit which eats the caterpillars. However, **metabolic modifications of ingested carotenoids** can also occur in animal tissues, but there is almost no information about the enzymes responsible. Hydroxylation at C-3 or C-4 is especially common, and in insects hydroxylation at C-2 is also quite frequent; the 4-hydroxy groups in particular can be further oxidized to oxo groups. The metabolization of β-carotene to canthaxanthin or astaxanthin has been demonstrated repeatedly, in particular in birds, fish, molluscs, insects and crustaceans (Fig. 19.11). The order in which the oxygen functions are introduced into the molecule varies from species to species [119]. Reductive metabolic pathways for the carotenoids have been detected only more recently in animals, e.g. the conversion of canthaxanthin and astaxanthin (Fig. 19.10 g, h) to dihydroxycarotene and tetrahydroxycarotene in the snail *Fusinus perplexus* [128]. A particularly interesting reaction of carotene metabolism is the contraction of cyclohexene to a five-ring system by the removal of C-2, as occurs during the formation of actinioerythrin in *Actinia equina* and other sea anemones (Fig. 19.12). Carotenoids with aromatic end groups, such as renieratene (Fig. 19.10 i), are found in the animal kingdom only in the sponges; the aromatization is apparently a function of the animal but the mechanism has not been defined [71]. The origin of the triple bonds (Fig. 19.10 j) present in carotenoids from sponges, starfish and mussels is also unknown [71, 121, 144]. ε,ε-Carotenoids (Fig. 19.10 k) were initially known only in vertebrates, e.g. from the integument and eggs of fish, the skin of the frog *Atelopus chiriquiensis*, the retinal oil drops of the turtle *Chelonia midas*, the egg yolk of the chicken and the feathers of various birds ("canary xanthophyll"). More recently, they have been detected in molluscs. The chicken can form the ε ring from the β end group of zeaxanthins and luteines [71, 187]. The so-called apocarotenoids, which carry only one cyclic end group, occur only occasionally in the animal kingdom, e.g. in the retina of the chicken and in poriferans, echinoderms and prosobranch gastropods [205]. A reaction of carotenoid metabolism which is apparently widespread in animals in the **production of the A vitamins** retinol and 3,4-dihydroretinol by cleavage of a carotene molecule between C atoms 15 and 15′. The pro-vitamin must have a non-substituted β ring and a side-chain of the correct length with conjugated double bonds. Hence, β-carotene gives rise to two molecules of vitamin A and α-carotene yields only one; zeaxanthin and lutein are inactive. A double bond at C-3 has no influence on vitamin synthesis. Vitamin A biosynthesis has been examined in detail only in the mammals, where it occurs in the gut wall and not, as previously thought, in the liver. The enzyme responsible for cleavage of the carotene is β-carotene-15,15′-dioxygenase, and the reaction proceeds via a peroxide intermediate [31]. The aldehyde formed initially is reduced to retinol by retinaldehyde reductase. Retinol is bound to cytoplasmic **retinol-binding proteins** (CRBPs), of which there are two types in mammals and these differ in their tissue distribution. The CRBP II of the rat, which is present only in the gut and plays an important role in carotene absorption and vitamin A synthesis, agrees in 56 % of its amino acids with the other type. Both belong to the same protein family, together with a retinoic acid-binding pro-

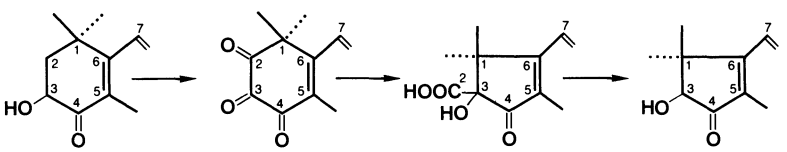


Fig. 19.12. The formation of the norcarotenoid actinioerythrin [31]. Only the left end group is illustrated

tein and the three fatty acid-binding proteins of the gut, liver and heart. In contrast, CRBP II is only partly homologous with the 182-amino-acid retinol-binding protein (RBP) in the blood plasma [53]. The uterus lining (endometrium) of the pig secretes at least four different RBPs with N-terminal sequences which agree 30–70 % with the human plasma RBP [43]. Retinoic acid is a retinol metabolite with considerable effects on cell growth and differentiation [160]. These effects are mediated by specific **retinoic acid receptors** which, together with the receptors of steroid hormones and thyroid hormones, constitute the nuclear receptor super-family. Three types of human and murine retinoic acid receptors (RAR) are known, RAR-α, -β and -γ. Comparison of the different receptors within each species shows much less agreement than that between individual receptor types in the two species. Thus, the three receptor types are evolutionally very old and probably have different functions [103].

The **carotene contents of animals** show large qualitative and quantitative differences, even though the pigments illustrated in Fig. 19.10a–h are found in almost all animal groups. Several carotenoids are always found together, e.g. in the ladybird *Coccinella septempunctata* there are no fewer than 22 types. The similarity of the chemical properties of the components prevents conclusive analysis of these mixtures; the solution of this problem awaits improvements in the analytical methods. The existence of **stereoisomers** must also be taken into account. For example, of the three possible isomers of astaxanthin, only 3S,3′S is present in the starfish *Martasterias glacialis* (Fig. 19.10 h), whereas crustacyanin in the lobster also includes 3R,3′R and the mesomer 3R,3′S [71]. Lutein has three chirality centres and can therefore form eight isomeric forms, A–H; A is generally distributed in plants and animals, and B, D, F and G have so far been identified only in the skin of various fish [126].

The carotenoid spectra of numerous species of the sponges, annelids, mussels, gastropods, crustaceans, insects, fish and birds are available in table form [71], and only a few more unusual details will be mentioned here. The carotenoid spectrum of the **Porifera** is dependent not only upon the plants and animals of the plankton diet but also upon algal and fungal symbionts in the sponges. In addition, the Porifera are capable of rather unusual carotenoid metabolism, e.g. aromatization of the cyclohexene ring, its reduction to a five-ring system, and the creation of triple bonds in the chain (Fig. 19.10 i, j). Among the **echinoderms**, the herbivorous echinoids, crinoids and holothurians contain relatively few carotenoids, mainly carotene hydrocarbons and echinenone, whereas the carnivorous asteroids and ophiurids possess relatively large amounts of predominantly hydroxylated carotenoids [188]. Large variation in the carotene content is also characteristic of the **molluscs:** the gonads of the limpet *Patella vulgata* have the extremely high value of 2.8 g/kg fresh weight; in comparison, most cephalopods are very low in carotene [71, 120].

The **crustaceans** contain free carotenoids, i.e. hydrocarbons and unesterified xanthophylls, together with xynthophyll esterified to long-chain fatty acids and a wealth of carotenoid protein complexes which are discussed below. Many crustacean carotenoids carry keto groups on C-4 and/or C-4′, often together with hydroxyl groups on C-3 and/or C-3′. In the Anostraca and Isopoda non-hydroxylated ketocarotenoids, such as canthaxanthin (Fig. 19.10 g), predominate, and 3-hydroxylated forms, such as astaxanthin (Fig. 19.10 h), are common in the copepods, Euphausiacea and Decapoda (Fig. 19.10 h). The metabolic pathways leading from β-carotene in the diet to canthaxanthin and astaxanthin apparently vary from species to species. Thus, from the schemes of canthaxanthin synthesis given in Fig. 19.11, the brine shrimp *Artemia salina* utilizes the pathway 1-2-3 but not 5-6-3 or 4-2-3; in contrast, reactions 4-2 and 5-6-3 have been demonstrated in the water flea *Daphnia magna* [71]. Amongst the arachnids, carotenoid metabolism has been examined in some detail in the spider mites *Tetranychus pacificus* and *T. urticae*. Mutants with defective carotenoid metabolism have been found, e.g. "albino", which cannot convert β-carotene to ketocarotenoids; "lemon", which has a similar defect but accumulates β-carotene in the haemolymph and thereby assumes its characteristic colour; and "flamingo", which does not esterify hydroxyketocarotenoids but stores them as granulae in its legs, leading to an aberrant pigment distribution [71]. Amongst the **insects**, the lepidopterans have received particular attention. Their carotenoids are derived almost unmodified from the diet, with some species showing a certain selectivity and others showing none at all. Conversion of β-carotene to astaxanthin has been shown, for example, in the migratory locust *Locusta migratoria*. Stick insects and, among the Lepidoptera, the greater puss moth *Cerura vinula* introduce a hydroxyl group at C-2 β-carotene, apparently forming a ketocarotenoid as inter-

mediate [71]. A novel oxidation product of β-carotene, 3,4-didehydro-β,β-carotene, has been identified in the Japanese stick insect *Neophirasea japonica* [129]. In some species, e.g. the aphid *Macrosiphum liliodendri* and the ladybird *Coccinella septempunctata*, part of the carotenoids originate not from the diet but from symbiontic microorganisms. The aphids contain β,γ-carotene and γ,γ-carotene and, in certain colour variants, also lycopene, a typical fungal carotenoid is not present in the diet. The ladybird also has lycopene and 21 further, in part completely unusual, carotenoids; although it feeds on aphids, at least a part of its carotenoids appear to originate from its symbiontic red yeast. In insects, as in vertebrates, a carotene deficiency leads to a loss of visual capacity [71].

Carotenoids have been analysed from more than 200 **fish species**. Yellow and red coloration of the body surface results from the accumulation of carotenoids in various types of chromatophores, of which the xanthophores of the trout have mainly lutein and the erythrophores have astaxanthin [71]. An unusual compound in many fish is tunaxanthin with two ε rings; this is not present in phytoplankton and is thus apparently produced by the fish itself. Conversion of β-carotene or lutein to astaxanthin has been clearly shown, e.g. in cyprinids. In contrast, the salmonids and apparently many other saltwater fish can oxidize neither lutein nor canthaxanthin to astaxanthin, and are therefore dependent upon the supply of ketocarotenoids in the diet. In the **birds**, carotenoid metabolism is often discussed in connection with one much-cited example: flamingos are often used as evidence for the inability of animals to synthesize carotenoids de novo; their pink colour stems from the ingestion of astaxanthin and canthaxanthin from their shrimp diet. In captivity they retain their colour only if their food is supplemented with carotenoids. This is true also for the xanthophylls in the feathers of the roseate spoonbill *Ajaia ajaja* and the scarlet ibis *Guava rubra* [32]. The existence of pigmented droplets in the retina of birds has been known for a long time; at least six different types are found. These droplets contain various carotenes, including ε,ε-carotene, the origin of which has not yet been determined. The oxygen-containing, ε,ε-carotenoids in liver and egg yolk apparently come from lutein and zeaxanthin. Thus birds, like fish, have the ability to convert the β ring into an ε ring, a process which is lacking in plants [71, 127]. **Mammals**, in contrast to most other animals, appear to be unable to oxidize the carotenes taken up in the diet. Their carotenoid metabolism is restricted to the production of vitamin A and its derivatives. The β-carotene-15,15′-oxygenase required is missing from the cat and probably also the mink, and these animals can thus not satisfy their vitamin requirement by ingesting carotenes as a pro-vitamin. A few mammals, such as humans, cattle and the horse, accumulate carotenes and xanthophylls in their bodies, whilst other mammals, like the buffalo, goat, sheep, pig, dog, cat, rat and rabbit, do not [71].

Because of their non-polar character, carotenoids are fat soluble but they may become water soluble by the binding of proteins. **Carotenoid-protein complexes** are widespread in animals; they are well known in several cnidarians, molluscs, echinoderms and insects, and have been most thoroughly examined in the lobster and several other crustaceans [210]. A distinction can be drawn between carotenoproteins, which consists only of polypeptides and carotenoid molecules in stoichiometric proportions, and the carotenolipo-(glyco)-proteins, which in addition to carotenoid include other lipid substances and do not always show stoichiometric relationships. The carotenoid component of both types is usually astaxanthin, but canthaxanthin, lutein and other carotenoids are also found as components of such complexes. The protein components always consist of two to five different polypeptides; the complete amino acid sequences have been determined for the crustacyanin subunits A_2 and C_1 from the lobster *Homarus gammarus*, and sequences have been determined for the N-terminus of the carotenoproteins from the starfish *Asterias rubens* and the siphonophore *Velella velella* [210]. Carotenoproteins usually have a completely different colour to that of the free pigment, ranging from purple through blue, green and yellow to red. The absorption spectra of most carotenoproteins are shifted to higher wavelengths (bathochrome shift). For example, the absorption maximum λ of free astaxanthin is 492 nm, whereas the maxima of astaxanthin proteins range from 536 nm for the red complex of the crustacean *Eriphia spinifrons* to 650 nm for the purple complex from the siphonophore *Porpita* sp. [71]. An explanation for this interesting alteration of the light-absorbing properties by the binding of protein awaits more detailed information about the nature of the carotenoid-protein bond and the resulting interaction of the components. The shift in the absorption maximum is to shorter wavelengths for some carotenoproteins (hypsochrome shift), for example the yellow

carotenoprotein from the lobster armour (carapace) has a maximum of 409 nm. The only explanation so far available here is the interaction between astaxanthin molecules that lie in stacks of 20 molecules [71].

The carotenoproteins of the lobster *Homarus gammarus* have received particular attention. The armour (carapace) contains the blue α-crustacyanin (λ = 628 nm), which consists of 16 subunits of 20 kDa each with an astaxanthin. Alteration of the pH and the ionic strength causes dissociation of the α-crustacyanin to eight dimeric molecules of the purple β-crustacyanin. Heat releases the astaxanthin and results in the characteristic red colour of the cooked lobster. The subunits of crustacyanin belong, together with the insecticyanin and the bilin-binding protein of the mammals, to the super-family of the retinol-binding proteins. There are five isoforms of the subunits and these fall into two groups according to their sequence homology. Group 1 includes the subunits C_1, C_2 and A_1, which agree completely in the 30 N-terminal positions, and group 2 is made up of the subunits A_2 and A_3 with 47 identical N-terminal amino acids. A_2 (172 amino acids) has only 65 amino acids in common with C_1 (181 amino acids). In β-crustacyanin, one A_2 and one C_1 chain are organized head-to-head; α-crustacyanin contains two tetrameric arrangements of such β-crustacyanin dimers [97, 210]. The carapace also contains the yellow carotenoprotein (λ = 409) already mentioned and two further yellow proteins, Y_1 and Y_2, which may, however, be artefacts [71]. The eggs of the lobster contain the green ovoverdin (λ = 460 + 640 nm), which has a molecular mass of about 300 kDa and includes two or three astaxanthin molecules with an apparently loose stoichiometry [71, 210]. In contrast to that of its European counterpart, the blue crustacyanin from the American lobster has only two types of subunit (H_1 and H_2), but is otherwise very similar [210]. Blue carotenoproteins of a similar type have also been reported in *Procambarus clarkii, Astacus leptodactylus* and *Upogebia pusilla* [153]; their close relationship is indicated by the formation of hybrid molecules between crustacyanin subunits from *Homarus americanus* and *P. clarkii*. The carapace of the crayfish *A. leptodactylus* is green-brown due to the presence of three carotenoproteins, one blue, one yellow and one red. The yellow protein (λ = 385 nm) has a native molecular mass of 564 kDa and consists of polypeptides of 195, 151, 126 and 95 kDa together with 12 carotenoid molecules (6 each of astaxanthin and zeaxanthin) [134]. A red astaxanthin lipoprotein (λ = 482 nm) has been described from the carapace of the crayfish *P. clarkii* [133], and one with a purple colour (λ = 597 nm) from the carapace of the anomuran *Galathea strigosa* [70]. The liphorin of insect haemolymph is always bound to carotenoids and thus, according to the above definition, it is a caroteno-lipo-glycoprotein. However, the only genuine carotenoprotein of insects to have been characterized so far is the violet protein found as the main component of the haemolymph in the mosquito *Rhynchosciara americana*. This is a homotetramer of 157 kDa carrying three molecules of echinenone and a canthaxanthin [19]. The skin of the starfish *Linckia laevigata* includes a yellow carotenoprotein (λ = 403 nm) and the blue linckiacyanin, which has two different types of carotene-binding sites and therefore two absorption maxima (λ = 395 and 612 nm). Native linckiacyanin has a molecular mass of more than 1000 kDa and carries at least 200 carotenoid molecules. In addition to astaxanthin and zeaxanthin, these include the previously unknown hydroxyclathriaxanthin with an aromatic end group. The purple asteriarubin from the starfish *Asterias rubens* also contains a mixture of different carotenoids [42, 209].

19.4 Visual Pigments and Phototransduction

In all known cases the light-absorbing **prosthetic group** (chromophore) of the visual pigments is identical or closely related to **retinal**, the aldehyde of vitamin A_1 (retinol). Whereas all the double bonds of the retinol that is newly formed from carotenes have the trans configuration (Fig. 19.13a), the visual pigments always have the 11-cis isomer (Fig. 19.13b), a molecule that is bent into a hook shape. It has even been shown in the unicellular alga *Chlamydomonas* sp. that the disturbed light response of "blind" mutants can be normalized by the supply of 11-cis-retinal [65]. Only a few vertebrates and invertebrates have visual pigments with chromophores other than retinal or its 11-cis isomer. It has been known for more than 100 years that the rods in the retina of many freshwater species do not contain "visual purple" (rhodopsin) with an absorption maximum of 500 nm, but instead contain porphyropsin with a maximum of about 520 nm. Here the chromophore is **3,4-didehydroretinal**, which is derived from vitamin A_2 (3,4,didehydroretinol)

(Fig. 19.13 c). The distribution of these two types of pigment is remarkable: rhodopsin and vitamin A_2 are present in all terrestrial and marine vertebrates, and the visual pigments of most invertebrates are rhodopsin-like. Porphyropsin and vitamin A_2 are found in many (but not all!) freshwater vertebrates. They are widely found in freshwater teleosts, and in fish that migrate between freshwater and seawater the chromophore changes with the surroundings. The tadpoles of the Ranidae and Hylidae have vitamin A_2, but the larvae of the toads (Bufonidae) always have only vitamin A_1, even when they inhabit the same ponds. Adult anurans have predominantly or exclusively A_1. Both visual pigments are found in some species. Thus, in the American bullfrog *Rana catesbeiana* the dorsal half of the retina, which generates the underwater image, contains considerable quantities of porphyropsin, but the ventral half of the retina has only rhodopsin. In contrast, the "four-eyed" fish *Anablebs anablebs*, whose eyes are subdivided into clearly distinct areas for air and underwater vision, has only rhodopsin. The natural diet of the freshwater fish and amphibians contains vitamin A_1 but no A_2. In at least some freshwater fish, vitamin A_2 is apparently obtained from pro-vitamins with 3,4-didehydro- end groups, whereas in the tadpoles of *Rana catesbeiana* retinol can be converted to 3,4-didehydroretinol [185].

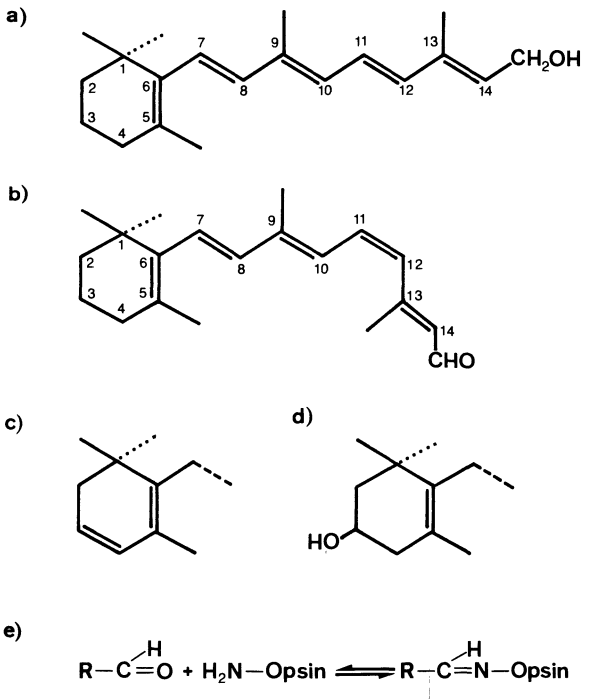


Fig. 19.13 a–e. Vitamin A and the prosthetic groups of visual pigments. **a** all-trans-retinol; **b** 11-cis-retinal (in rhodopsin); **c** 3,4-didehydroretinal (in porphyropsin); **d** 3-hydroxyretinal (in xanthopsin); **e** 11-cis-retinal is bound as a Schiff base to a lysine side-chain (in bovine rhodopsin 296-Lys; see Fig. 8.7b, p. 310)

A new chromophore, **3-hydroxyretinal** (Fig. 19.13 d), was discovered in insects a few years ago; the corresponding visual pigment is known as xanthopsin. 3-Hydroxyretinol was thought at first to be characteristic of the Diptera and Lepidoptera, but was later found also in the Odonata, Neuroptera, Homoptera, several Coleoptera (Cerambycidae), Hemiptera, Plecoptera, Megaloptera and Hymenoptera. Amongst the Diptera, retinal is found in the Simuliidae and 3-hydroxyretinal is found in 13 further families. The Odonata, Hemiptera, Mecoptera and Trichoptera have both retinal and 3-hydroxyretinal [166, 172]. The crayfish *Procambarus clarkii* is the only invertebrate to possess both rhodopsin and porphyropsin. 3-Hydroxyretinal is not found in this species but apparently the alcohol 3-hydroxyretinol is present; this is possibly an intermediate in the synthesis of 3,4-didehydroretinal [179]. The luminous cephalopod *Watasenia scintillans* has three different visual pigments. The main pigment of the whole retina ($\lambda \approx 484$ nm) contains retinal. A second pigment ($\lambda \approx 500$ nm) in the proximal section of the long rhabodome of the ventral retina contains 3,4-dihedydroretinal, and a third pigment ($\lambda \approx 470$ nm) from the distal section has a new chromophore, 4-hydroxyretinal [125]. Lepidopterans are apparently dependent upon 3-hydroxyxanthophyll as the precursor for the synthesis of 3-hydroxyretinal, whereas dipterans can hydroxylate β-carotene and possibly also retinol in the 3- position [193]. In addition to 3-hydroxyretinal, the visual pigment of the higher dipterans appears to carry the corresponding alcohol (3-hydroxyretinol), which has a UV-sensitizing effect, but in contrast to the chromophore does not alter in absorption properties upon illumination. The absorption spectra of the primitive dipterans and lepidopterans show no evidence of a UV-sensitizing pigment [100].

In bovine rhodopsin, 11-cis-retinal is bound as a Schiff base to the side-chain of 296-Lys (Fig. 19.13 e); this shifts the absorption maximum from 380 to about 500 nm (bathochrome effect). The rhodopsins extracted from the retina of various vertebrates have three **absorption maxima**, of which $\alpha = 493–505$ nm and $\beta = 340–350$ nm are due to the chromophore and disappear on light absorption and excitation ("bleaching" of the visual pigment); the absorption band $\gamma = 280$ nm stems from the aromatic amino acids of the protein

components. For the many rhodopsins that have been investigated, in most cases the α bands lie in the given region of 493–505 nm. However, exceptions are found among deep-sea elasmobranchs, teleosts and crustaceans, where the maxima are shifted to shorter wavelengths (blue shift), and among fish that inhabit very cloudy coastal waters; in this case the maxima lie above 505 nm. These deviations are most likely adaptations to the particular light conditions. The capacity for colour resolution among vertebrates and invertebrates is determined by the presence of photoreceptor cells with differing spectral sensitivity maxima, presumably due to the presence of different visual pigments. Much less is known biochemically about the pigments of the colour-sensitive cones of the vertebrate retina than about the rhodopsins and porphyropsins of the rods. Four cone visual pigments have been isolated from the chicken retina and these correspond to the four cone types present: violet, blue, green and red. The sequence of the red pigment ("iodopsin") has been determined via the cDNA [107, 143]. The DNA sequence of other cone visual pigments are known but the pigments themselves have not been isolated, e.g. the opsins of the three human receptor types, postulated by the trichromatic theory of colour vision, and the opsins of the blind cave fish *Astyanax fasciatus*, which are homologous to the human green and red cone pigments [206]. In general, comparative details of the number and biochemical properties of the visual pigments from different animals with colour vision are not yet available.

The **amino acid sequence** of bovine rhodopsin was determined directly and involved years of tedious effort. The new methods of DNA sequencing then rapidly generated further sequence information about, for example, the cone visual pigments of humans, the chicken, and cave fish; the rhodopsins of humans, various other mammals, the chicken and the lamprey *Lampetra japonica;* the four rhodopsins, Rh1 to Rh4, of *Drosophila melanogaster;* the Rh1 homologue of *Calliphora erythrocephala;* the rhodopsin of the squid *Loligo forbesi;* and three bacterial rhodopsins [12, 76, 82, 85, 107, 145, 177, 181, 206]. Although these opsins have very few amino acids in common, they are nevertheless very similar in their basic structure. They are all integral membrane proteins with seven transmembrane helices and carry in their sequences of usually 341–381 amino acids one to two glycosylation sites and C-terminally at least six serines or threonines that can be phosphorylated (see Fig. 8.7b), p. 310). The rhodopsin of *Loligo forbesi* has a C-terminal extension with many proline residues and negatively charged amino acids, and consequently has the unusual total length of 452 amino acids [76]. The protein super-family to which the opsins belong also includes the G-protein-coupled hormone receptors, e.g. the β-adrenalin receptors and muscarinic acetylcholine receptors (p. 309). Although the hormone receptors agree with the opsins in only 12–17 % of their amino acids, they do have the same characteristic architecture with seven transmembrane domains (see Fig. 8.7a). The human, bovine and murine rhodopsins (all with 348 amino acids) show 93–94 % sequence agreement with each other, 87 % with chicken rhodopsin, and 78–82 % with that from the lamprey *Lampetra japonica* (341 amino acids) [12, 82, 181]. The opsin of the human blue receptor has the same length but only 42 % similarity compared with the rod pigment; the red- and green-specific human cone pigments have 96 % of their 363 amino acids in common but show only 43–44 % agreement with the opsin of the blue receptor and 40–41 % agreement with that of the rods [207, 214]. The human X chromosome carries one to three green pigment genes but always has only one red pigment gene. These genes agree in 98 % of their DNA sequences and are apparently the result of relatively recent gene duplications [194]. The cave fish *Astyanax fasciatus* also has one red pigment gene and multiple green pigment genes that are homologous to those of humans. However, sequence comparisons show that the genes of fish and mammals arose by independent gene duplications [206]. Iodopsin (chicken red pigment) has 80 % of its 362 amino acids in common with the opsin of the human red receptor but shows only 43 % agreement with chicken rhodopsin [107]. Using a cDNA of bovine rhodopsin as a probe, homologous genes have been found not only in vertebrates of all classes and in *Drosophila*, but also in the crayfish *Procambarus blandingii*, the squid *Argonauta argo*, the alga *Chlamydomonas reinhardtii* and the similarly light-orientated bacterium *Halobacterium halobium* [122].

The faceted eye of *Drosophila* contains three classes of photoreceptor. Each of the approximately 800 ommatidia has six outer cells, R1–R6, and two inner cells, R7 and R8, which can be divided into several subclasses. Mutants with a reduced rhodopsin content, prolonged after-potential and concomitant desensitization are known as "nina" (neither inactivation nor after-potential). Eight genes (nina A-H) are known so

far, of which ninaE encodes opsin R1–6 [196], and ninaA encodes a cyclophilin-like protein of 237 amino acids which may be responsible for the correct folding and stability of opsin R1–6 [162, 168]. The sequence of opsin R1–6 agrees in 36 % of its 373 amino acids with that of mammalian rhodopsin; an important difference, however, is that the *Drosophila* opsin has 10 additional amino acids in the third cytoplasmic loop (between helices 5 and 6). This probably explains some of the characteristic functional differences from vertebrate rhodopsins: the metarhodopsin which appears as an intermediate following light excitation of rhodopsin is much more stable in the physiological temperature range in *Drosophila* and other invertebrates than in the vertebrates; the chromophore in *Drosophila* remains bound to opsin during the whole photocycle, whereas that of the vertebrates is released; the visual pigment of *Drosophila* and other invertebrates is much less mobile in the membrane than are those of the vertebrates [214]. Only the newly synthesized insect opsin is glycosylated, and the mature protein is not. The opsin R1–6 of the distantly related fly *Calliphora erythrocephala* shows 86 % sequence agreement with that of *Drosophila* [85]. Four opsin genes have been identified so far in *Drosophila*. Gene Rh1 is identical to ninaE and is expressed in the photoreceptor cells R1–6. The opsin genes Rh3 and Rh4 are expressed in two subsets of the photoreceptor cells R7, and gene Rh2 is expressed in R8. Rh1 is also expressed in the paired internal photoreceptor organs of the larvae, and Rh2 is expressed in the three dorsal ocelli of the adult [145]. Opsin R8 has 381 amino acids with a sequence that is 67 % similar to that of opsin R1–6, and also has 10 additional amino acids in the 5/6 loop. Despite the similarity of the coded amino acid sequences, the structure of the two *Drosophila* genes is very different; only one of the four introns in the R8 gene is comparable with those of ninaE [46].

Rhodopsin is excited by **light absorption**, and the process has been studied in detail with bovine rhodopsin. It occurs via several intermediate steps with characteristic absorption maxima, of which the first are extremely short-lived at room temperature and can be defined only at very low temperatures (Fig. 19.14a). The 11-cis-retinal is already isomerized to all-trans-retinal in the very first step, the formation of bathorhodopsin. The active, excited form of rhodopsin in the ensuing process chain is metarhodopsin II, which has a half-life of several minutes before it dissociates into opsin and free retinal [214]. The conformational changes associated with rhodopsin excitation have also been examined in the cephalopod *Euprymna morsei* [81]. **Regeneration of rhodopsin** involves the reisomerization of all-trans-retinal to the 11-cis isomer. It was discovered only relatively recently that retinal is reduced to retinol in the process, and that the retinol isomerase is localized in the pigment epithelium cells of the retina [29].

Retinol is bound to a specific protein, the interstitial retinoid-binding protein (IRBP), for transport between the receptor cells and the pigment epithelium; this protein is present in the extracellular matrix of the vertebrate retina. The human and bovine polypeptide chains of 1230 and 1264 amino acids, respectively, have four homologous repeats of about 300 amino acids, carry four noncovalently bound fatty acids, and can bind two molecules of retinol or another retinoid. IRBP has a mass of about 140 kDa in the terrestrial vertebrates but only 70 kDa in the teleosts [25, 114]. A retinaldehyde-binding protein (RALBP) is responsible for the intracellular transport of 11-cis-retinol and 11-cis-retinal; the bovine and human sequences of 316 amino acids show 92 % agreement [47]. Proteins with immunological similarities to IRBP and RALBP have been demonstrated in various cephalopods [64].

Two segments in the rods of the vertebrate retina can be distinguished; the outer segment consists of a stack of several hundred light-sensitive membranes (discs) whose predominant protein component is rhodopsin. In darkness, cations leave the inner segment and enter the outer segment; light brings about the closure of the ion channels in the outer segment, and thus the hyperpolarization of the receptor cells. The process cascade between the light stimulus and the neural response of the receptor cell is known as **phototransduction** [176, 214, 215]. Here, there is space to give only some basic details of the process (Fig. 19.14b). The cause of the closure of the ion channels in the outer segment of the receptor cells is a reduction in the cGMP concentration, brought about by activation of cGMP phosphodiesterase. In fact, only part of the cGMP in the rods is light dependent; the rest is bound and not influenced by illumination. The functional coupling of light-activated rhodopsin and the phosphodiesterase is established by transducin (Fig. 19.14b). The photosensitive membranes of the rod outer segments contain about 20 different proteins in the frog *Rana catesbeiana;* of these, 70 % are rhodopsin ($3 \cdot 10^9$ molecules per cell), 17 % transducin ($3 \cdot 10^8$ molecules per cell)

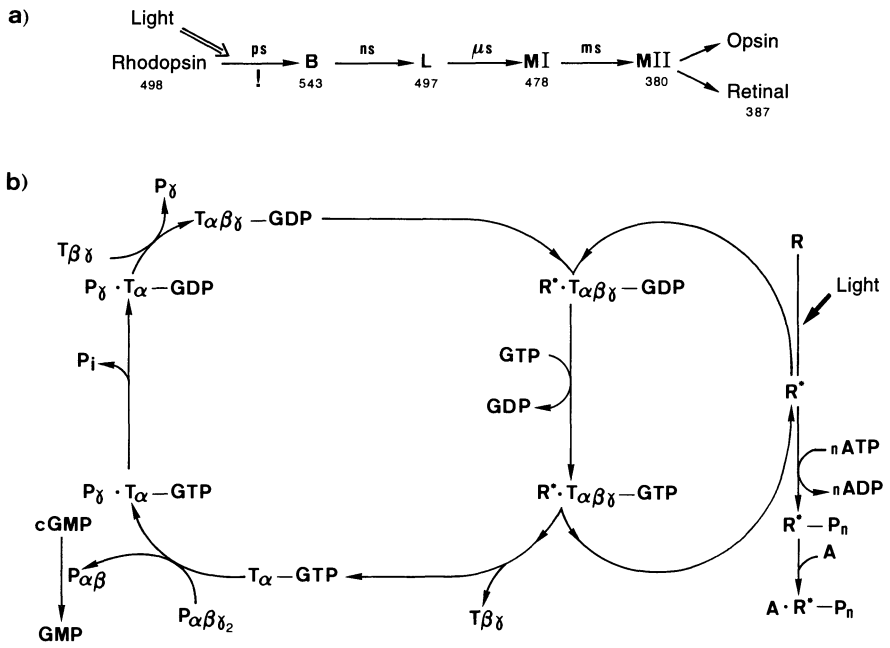


Fig. 19.14 a, b. Phototransduction in the vertebrate retina [176, 214, 215]. **a** Stimulation of the visual pigment: rhodopsin is transformed to the excited state MII by absorption of a photon and changes its absorption maximum (bleaching). The process proceeds via the short-lived intermediates bathorhodopsin *(B)*, lumirhodopsin *(L)* and metarhodopsins I and II (*MI* and *MII*). The isomerization of 11-cis-retinal to all-trans-retinal *(!)* occurs in the first step. Finally, the all-trans-retinal is released from opsin. The order of magnitude of the half-life at body temperature is given above each reaction step; the absorption maxima in nm are shown under each intermediate. **b** Phototransduction in the rods of the vertebrate retina: the rhodopsin stimulated by light (R^+) catalyses the exchange of GDP and GTP on several transducin molecules ($T\alpha\beta\gamma$) and the cleavage into $T\beta\gamma$ and $T\alpha$-GTP. R^+ is gradually inactivated by phsophorylation and binding of the protein arrestin *(A)*. The complex $T\alpha$-GTP activates the cGMP phosphodiesterase ($P\alpha\beta\gamma_2$) by binding its inhibitory γ chains. The complex $P\gamma$-$T\alpha$-GTP dissociates after hydrolytic cleavage of GTP; the released γ chains of the phosphodiesterase can again attach to and inactivate the $P\alpha\beta$ complex; the α chain of transducin *(Tα-GDP)* can bind with $T\beta\gamma$ to give the completely inactivated transducin ($T\alpha\beta\gamma$-GDP). See the text for further details

and 1.5 % phosphodiesterase ($1.5 \cdot 10^7$ molecules per cell) [78]. The interaction between rhodopsin and transducin as well as that between transducin and phosphodiesterase involves lateral diffusion in the membrane. About 15 % of the membrane surface is occupied by rhodopsin molecules; a more dense packing would be possible but this would reduce the mobility of the various protein molecules [215]. The light-induced isomerization of retinal takes place deep inside the membrane (see Fig. 8.7b, p. 310); the resulting conformational change must be communicated to the cytoplasmic surface of the rhodopsin molecule. The interaction with transducin takes place on the second and third cytoplasmic loops and on a fourth loop of amino acids 310–323, which is formed by binding of 322-Cys and 323-Cys with palmitic acid of the bilayer [101]. On the edges of the disc membrane is a specific protein, peripherin, which has 345 amino acids and four transmembrane segments but no known function. In addition to rhodopsin, the plasma membrane of the rod outer segments contains the cGMP-gated cation channel, an oligomer of 80-kDa subunits, and an Na,Ca-exchanger of 220 kDa [45, 176].

Transducin belongs to the same protein family as the G-proteins of the hormone receptors (p. 309). Like the latter, it is also an $\alpha\beta\gamma$ heterotrimer in which the α chain of 39 kDa is loosely bound to a tight complex of one β chain of 36 kDa and one γ chain of 8 kDa. The α chains carry the binding site for guanine nucleotide and this has a similar sequence in all G proteins. The light-activated rhodopsin (metarhodopsin II) causes the exchange of GDP for GTP on transducin. One activated rhodopsin molecule triggers several hundred such exchanges. This is the first of several amplifying effects in the process cascade of phototransduction, and confers the ability of rod receptors to detect single photons [176].

The α chain of bovine rod transducin consists of 350 amino acids, corresponding to 39 kDa; it differs in only one amino acid from the mouse α chain and shows 88 % agreement with the human rod transducin but only 65 % agreement with the 41-kDa bovine cone transducin [54, 151]. There are two structurally and functionally different bovine γ chains, of which γ-2 (6 kDa) allows the interaction of Tαβγ with the excited rhodopsin, and γ-1 (8 kDa) has no known function. A similar heterogeneity of the Tαβγ complex is found in the frog *Rana catesbeiana* [68, 142]. Transducin-like proteins have been detected immunologically not only in the retina of vertebrates of all classes and in the pineal and parapineal organs of fish and amphibians but also in the retina of various cephalopods [174].

Bovine rod **cGMP phosphodiesterase** (PDE) consists of three subunits, α (88 kDa), β (84 kDa) and γ (11 kDa), in the stoichiometry of $\alpha\beta\gamma_2$. The α and β subunits are homologous. The bovine β chain agrees in 72 % of its 853 amino acids with the bovine α chain [115]. The bovine and murine γ chains consist of 87 amino acids and differ at only two positions [191]. Two Tα-GTPs interact one after the other with the inactive $P\alpha\beta\gamma_2$, bind the two inhibitory γ chains and release the active Pαβ [176]. The bovine cone PDE consists of two identical α′ subunits of 93.5 kDa and an unknown number of smaller polypeptides of 11, 13 and 15 kDa. The α and β subunits of the rod enzyme show more agreement with each other than with the α′ subunit of the cone PDE [69, 115]. About 20–30 % of the PDE activity of the rod outer segment is not membrane bound. This soluble isoenzyme consists of four types of subunit, α_{sol} (88 kDa), β_{sol} (84 kDa), γ_{sol} (11 kDa) and δ (15 kDa). The δ subunit has no inhibitory properties and is possibly responsible for the solubility of the isoenzyme [69]. Hydrolysis of the transducin-bound GTP to GDP leads to dissociation of the Pγ-Tα-GDP complex. The released γ chains of PDE can combine with the active αβ complex to give inactive phosphodiesterase, and the remaining GDP-bound transducin α chain combines with the corresponding βγ complex to form inactive GDP-transducin; the latter is converted again to active GTP-transducin by active rhodopsin [176, 215]. Re-establishment of the dark state requires not only the deactivation of PDE and transducin but also the quenching of the excited rhodopsin and the formation of fresh cGMP. The rapid deactivation of the phototransduction cascade is achieved by phosphorylation of rhodopsin on multiple serine and threonine residues in the C-terminal region by a cytosolic 68-kDa rhodopsin kinase, and binding of the cytosolic 48-kDa protein arrestin to the phosphorylated rhodopsin. Several minutes pass before the all-trans-retinal is replaced by 11-cis-retinal to regain the initial state of the rhodopsin. *Drosophila melanogaster* also has a protein of 364 amino acids which shows >40 % sequence agreement with human and bovine arrestin [171, 176]. The activation of guanylate cyclase in bovine rods involves the binding of the 26-kDa protein recoverin, which is highly homologous to visinin from chicken cones. Recoverin belongs to the EF-hand super-family of calcium-binding proteins and has three calcium-binding sites. In the dark state, Ca^{2+} influx through the cGMP-gated channel and the equivalent Ca^{2+} efflux through the Na,Ca-exchanger generates a Ca^{2+} concentration of about 500 nmol/l; this causes recoverin to be loaded with Ca^{2+} and prevents it from binding to guanylate cyclase. Illumination blocks the Ca^{2+} influx, but the Ca^{2+} efflux continues. In this way, the Ca^{2+} concentration drops to about 50 nmol/l and Ca^{2+}-free recoverin is produced; this binds to and activates guanylate cyclase [108, 176]. Phototransduction in the cones corresponds more-or-less to that in the rods, even though many of the components involved show differences [176].

The velocity and sensitivity of the process cascade of phototransduction in vertebrate photoreceptors has been finely tuned during evolution. It is therefore surprising that the equally effective phototransduction in the faceted eyes of the **arthropods** and in the retina of the **cephalopods** display fundamental differences. Two segments of photoreceptor cells can be distinguished in the xiphosuran *Limulus polyphemus*, and only one carries microvilli. The membranes of these microvilli contain rhodopsin molecules which are many times less mobile than those in vertebrate rod membranes. In this case, instead of causing closure light brings about the opening of ion channels and depolarization of the receptor cells. The concentrations of cGMP and Ca^{2+} do not decline after illumination but increase. A transducin-like protein is also present in the cephalopods; however, this does not affect a cGMP phosphodiesterase but activates a phospholipase C which cleaves inositol trisphosphate (IP_3) from phosphatidylinositol bisphosphate (PIP_2). The IP_3 probably regulates the ion permeability of the photoreceptor membrane by the release of Ca^{2+} from intracellular storage pools [203]. The photophobic and phototactic reactions of the ciliate *Stentor coerulus* have a completely different molecular

basis. The pigment granulae serving here as the primary photoreceptors contain two chromoproteins, stentorins I and II, which have extended quinones of complicated structure as the prosthetic groups. The primary photo-process appears to be one of proton dissociation [98, 173].

19.5 Nitrogenous Substances

19.5.1 Amino Compounds

Histamine, serotonin, dopamine and other **biogenic amines** are frequent components of animal toxins (p. 320); thus, in addition to their varied functions as internal signal substances these compounds have pharmacological effects on other individuals. Even substances as poisonous as trimethylammonium are found, for example, in the nematocysts of sea anemones and in the salivary glands of sea snails [13]. The trapping threads in the web of the garden spider *Araneus diadematus* owe their sticky property to the presence of droplets of an aqueous fluid containing γ-aminobutyramide (1.9 mol/l), choline (1.5 mol/l) and isethionic acid (1.1 mol/l) [195]. The haemocytes of the ascidian *Halocynthia roretzi* contain tetrapeptide-like substances with antimicrobial activity, the halocyamines, which are made up of three amino acid residues and 6-bromo-8,9-didehydrotryptamine [10]. In some marine gastropods, the hypobranchial glands produce toxic **cholinesters:** murexine (Fig. 19.15a) in *Murex brandaris*, senecioylcholine in *Thais floridana* and acrylylcholine in *Buccinum undatum* [155]. The polychaete *Lumbriconereis heteropoda* has in its epidermis an amine with an unusual dithiolane ring and which is known as nereistoxin (Fig. 19.15c). Surprisingly, this substance is an effective insecticide and served as a model for the development of the synthetic insecticide padan [62, 155]. The ladybird (Coccinellidae) defends itself very effectively against attackers by exuding droplets of haemolymph at its leg joints (reflex bleeding); this contains various alkaloids and the long-chained diamine whose structure is shown in Fig. 19.15b [26]. The skin glands of the amphibians produce a variety of very different types of highly poisonous substances, including numerous pharmacologically active peptides (p. 293), alkaloids and also several amines. For example, the skin of frogs of the genus *Leptodactylus* contains various serotonin derivatives and also a phenol-betaine known as leptodactyline (Fig. 19.15e). Several rare **amino acids** also have pharmacological effects, e.g. N-methyl-D-aspartate from the sea snail *Scapharca broughtonii* has neuroexcitatory activity [159]. The mercaptohistidine ovothiol (see Fig. 18.10, p. 710) from the eggs of the sea urchin *Strongylocentrotus purpuratus* is apparently involved as a cofactor in oxidative metabolism; similar substances have meanwhile been found in the eggs of the starfish *Evasterias troschelii* and the mussel *Chlamys hastata* [190]. Thiolhistidine is also present in adenochrome, an Fe(III) pigment produced in the octopus in the "white bodies" of the gill hearts and spread in the body by circulating amoebocytes. Adenochrome is a mixture of peptides made up of two molecules of glycine and one molecule of one of the adenochromines A–C (Fig. 19.15d). These unusual amino acids consist of two molecules of thiolhistidine bound to dopa via a thiolester bond [138]. The eyes of many tropical marine fish contain the UV-absorbing protective pigment palythine (λ = 320 nm), asterina-330 (λ = 330 nm), palythinol (λ = 332 nm) and palythene (λ = 360 nm).

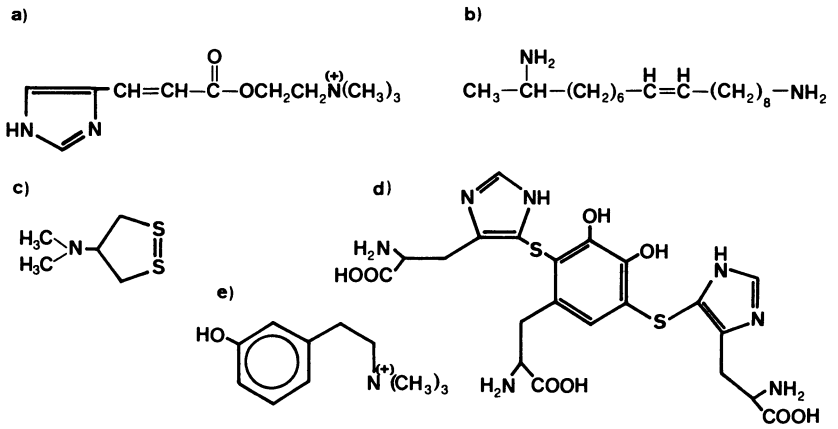


Fig. 19.15a–e. Amino compounds. **a** Murexine [13]; **b** (Z)-1,17-diaminooctadec-9-ene from coccinellids [26]; **c** nereistoxin [155]; **d** adenochromine A [138]; **e** letpodactyline [75]

These are all unusual amino acids with a carbon skeleton that is similar to that of the N-free compound gadusol (Fig. 19.5 e) [55].

19.5.2 Cyanogenic Compounds

The ability to release HCN from cyanogenic glycosides is widely found in plants; about 2000 HCN-producing species and more than 40 different cyanogenic glycosides are known. In contrast, cyanogenesis in the animal kingdom is found only in several arthropods. Most diplopods from the order Polydesmida and many chilopods possess **mandelonitrile**, from which they can release HCN by means of an α-hydroxynitrile lyase in a special reaction chamber (Fig. 19.16a). Benzoic acid esters of mandelonitrile and benzoylcyanide (Fig. 19.16b) are also present as cyanogenic substances. These compounds are synthesized in the animal from phenylalanine as the precursor [18]. Mandelonitrile has also been found as a cyanogenic substrate in the carabid *Megacephala virginica* and other species of the carabid subfamily Cicindelinae. The substrate of cyanide production in several leaf-beetle larvae (Chrysomelidae) has not yet been clearly identified [50]. Cyanogenesis is widely found in moths of the families Zygaenidae and Lymphalidae. All developmental stages contain the two cyanogenic glucosides **linamarin** and **lotaustralin**, which are also found in plants (Fig. 19.16c). HCN release requires cleavage of the sugar by a β-glucosidase. These two glucosides are also found in moths that live on non-cyanogenic host plants [50, 202]. Tracer experiments carried out on several representatives of both families of lepidopterans have shown that linamarin is produced from valine and lotaustralin is produced from isoleucine, with the corresponding nitriles appearing as intermediates (Fig. 19.16c) [51]. Caterpillars of *Zygaena trifolii* contain β-cyanoalanine, which is produced from serine and HCN in an enzymatic reaction. Both cyanogenic and non-cyanogenic lepidopterans are insensitive to HCN, but the mechanism of their resistance is not yet understood [50, 202].

19.5.3 N-Heterocyclic Compounds and Alkaloids

Several **nitrogen-containing rings** (Fig. 19.17) are ubiquitous as structural elements of the amino acids: pyrrolidine in proline, imidazole in histidine, and indole in tryptophan. In addition, N-heterocyclic compounds have attained special functions in individual species. For example, a pyrrole derivative functions as a trail pheromone in ants of the genus *Atta* (Fig. 19.18a). The large group of pyrrole pigments are of particular interest for comparative biochemistry and are considered in Section 19.6. A pyrimidine compound acts as a signal substance in the symbiosis between fish and sea anemones: the amphikuemine produced by the sea anemone *Radianthus keucken-*

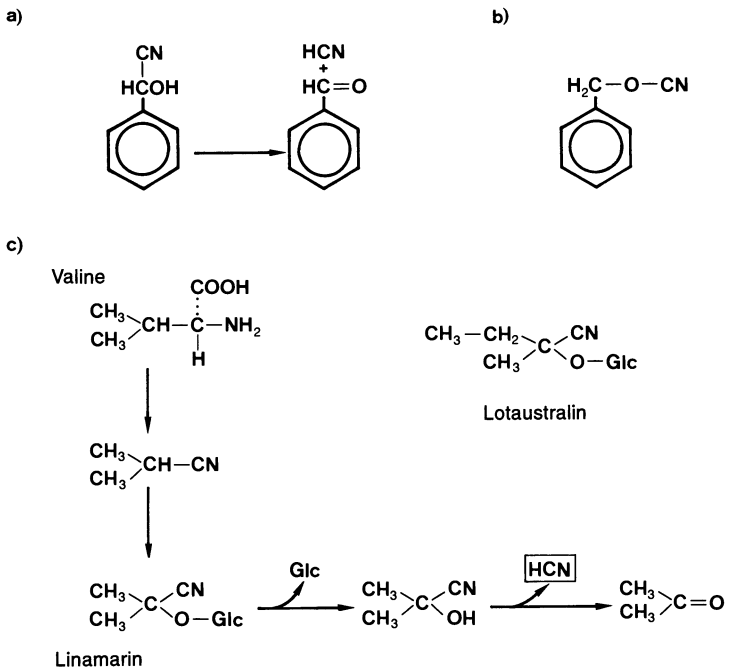


Fig. 19.16a–c. Cyanogenic substances [50]. **a** The release of HCN from mandelonitrile; **b** benzoylcyanide; **c** linamarin: biosynthesis from valine and release of HCN; lotaustralin is formed from isoleucine and is cleaved in an analogous process

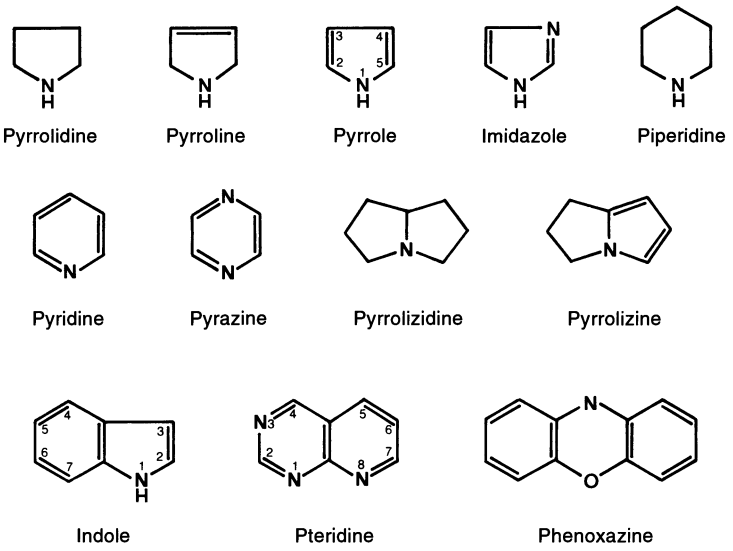


Fig. 19.17. N-Heterocyclic basic structures

thali attracts the fish *Amphiprion perideraion* even at a concentration of 10^{-10} mol/l [137]. Numerous alkyl-substituted pyrazines are found in ants, spider wasps (Sphecidae) and potter wasps (Eumenidae), where they function as defence substances and alarm pheromones. Many ants of the subfamily Myrmicinae utilize 3-ethyl-2,5-dimethylpyrazine as a trail pheromone [89]. The European ants of the genus *Monomorium* synthesize various alkyl pyrrolidines and alkyl pyrrolines which are used to kill prey and to repel other ant species [113]. The venom glands of the "fire ants" of the genus *Solenopsis*, whose toxicity and readiness to attack makes them a danger even to humans, contain species-specific mixtures of alkyl-substituted piperidines, pyrrolines, pyrrolidines and pyrrolizidines [27]. Oxygen-containing derivatives of pyrrolizines are released by special structures (hair brushes) on the males of the monarch butterfly (Nymphalidae), and these compounds act upon the females as sex pheromones. These substances have been shown to originate in pyrrolizidines of the plant diet [27, 164, 200]. Indole (Fig. 19.17) and its 3-methyl derivative skatole are present in the defence secretions of ants and trichopterans [27]; more than 70 different indole derivatives are known from marine animals, including many halogenated forms, e.g. 4,6-dibromindole from the hemichordate *Glossobalanus* sp. [80]. The venom of the spider *Argiope lobata* contains coloured, low molecular weight toxins that block the postsynaptic glutamate receptors of their prey; these are known according to the chromophore groups as argiopin (with 2,4-dihydroxyphenylacetic acid), argiopinin (with 4-hydroxyindole-3-acetic acid) and pseudoargiopinin (with indole-3-acetic acid). Indole derivatives are also constituents of the nephilatoxins of *Nephila clavata* [73]. Riboflavin, which is derived from alloxazine, a pteridine with a condensed benzol ring, is ubiquitously distributed as the prosthetic group of oxidoreductases. Just occasionally, it is involved in determining body colour, as in *Holothuria forskali* [31].

Purine and pyrimidine compounds are present in all animals as components of nucleic acids, coenzymes and cosubstrates. Guanine and other purines are also involved in determining body colour, especially in fish. Micro-crystals, stored in special guanophores or iridophores, reflect light of all wavelengths and confer on the fish a white or silver colour, or gold when the crystals are overlaid with carotenoids, as in the goldfish [31]. Several sponges contain relatively large quantities of nucleosides with unusual structures, e.g. arabinosylthymine and -uracil (spongothymidine and -uridine) and 2-methoxyadenosine (spongosine) in *Thetya (Cryptothetya) crypta;* these are absent from the related species *T. aurantia* and other sponge species. Unusual purine nucleosides are a feature of sea slugs (Opisthobranchia): isoguanosine (Fig. 19.18 c) is present in *Diaulula sandiegensis*, and 1-methylisoguanosine in *Anisodoris nobilis* [197]. 1-Methyladenine induces gonad ripening in the starfish *Asterina pectinifera;* this substance is not derived from methylated nucleic acid but is synthesized de novo. The hermaphrodite gland of the edible snail contains much free inosine. N-Heterocyclic compounds are also the basis of important groups of biochromes: the porphyrins and bilins (with pyrrole), the pterins (with pteridine), the ommochromes (with phenoxazine) and the melanins (with indole-5,6-quinone and pyrrole); these are all dealt with in separate sections.

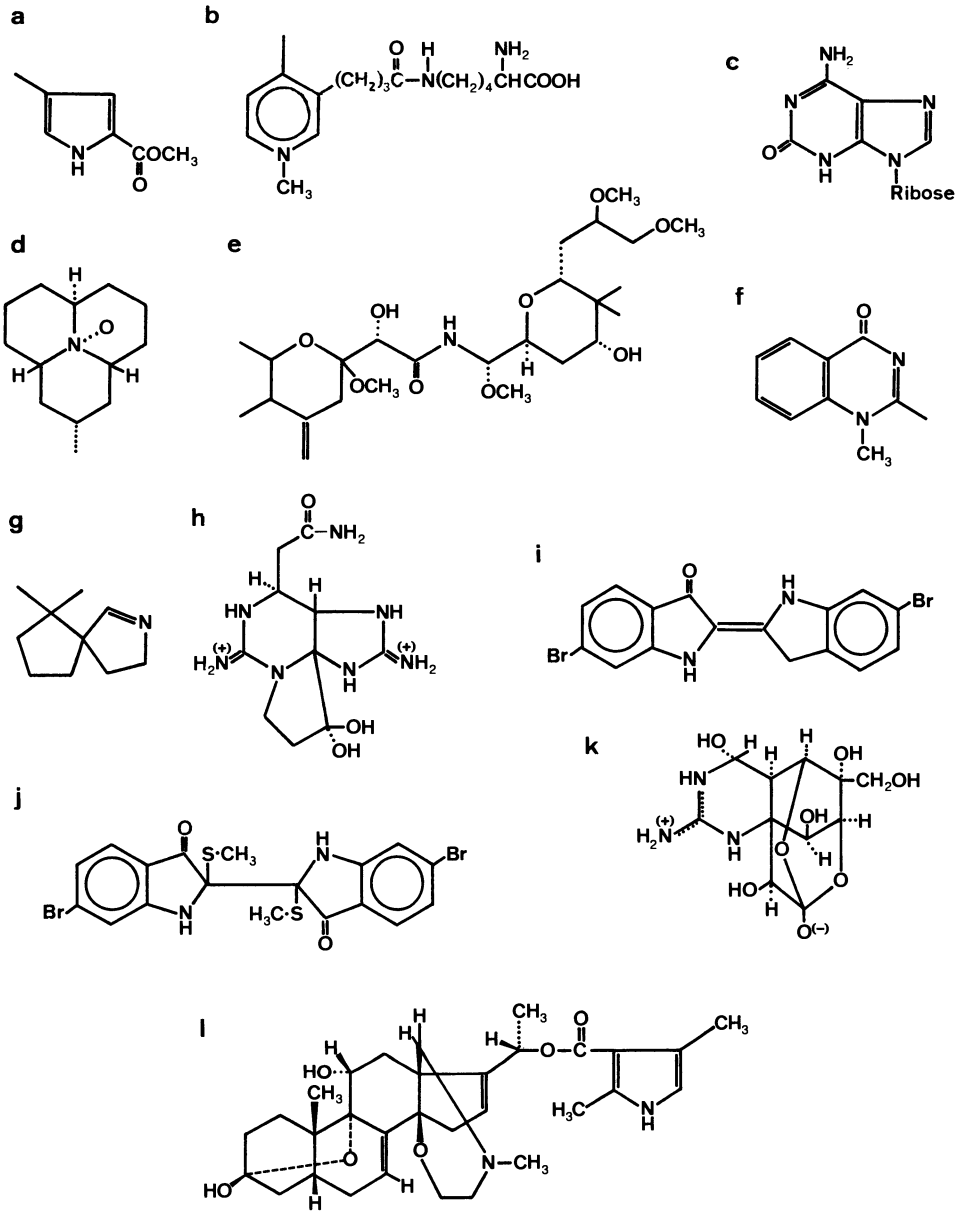


Fig. 19.18 a–l. N-Heterocyclic substances. **a** Methyl-4-methylpyrrole-2-carboxylate [27]; **b** amphikuemine [137]; **c** isoguanosine [197]; **d** coccinellin [75]; **e** pederin [75]; **f** glomerin [13]; **g** polyzonimin [18]; **h** saxitoxin [155]; **i** 6,6′-dibromindigo (Tyrian purple) [31]; **j** tyriverdin [31]; **k** tetrodotoxin [90]; **l** batrachotoxin [90]

N-Heterocyclic substances are frequently referred to as **"alkaloids"**. This term was coined in 1819 by the pharmacist Meissner for alkaline plant substances, and is therefore not a strict chemical definition. Terrestrial plants produce a wealth of alkaloids as deterrents against organisms that eat them; these find their way into the bodies of herbivorous insects, where they can be stored and again serve a deterrent function (p. 716). The N-heterocyclic compounds in the defence secretions of ants (described above) are often included amongst the alkaloids. Alkaloids of rather unusual structure are also found in the fluid exuded by the coccinellids during reflex bleeding (p. 739); they are all based upon the azaphenalene ring system (Fig. 19.18 d). Feeding experiments indicate that the coccinellin of the ladybirds is produced from polyketides. The oxygen-free precoccinellin and other compounds of this group have also been detected in the defence secretions of a beetle from a completely different family, the cantharid *Chauliognathus*

pulchellus [27]. Despite its complicated structure (Fig. 19.18 e), pederin from the staphylinid *Paederus fuscipes* can be synthesized by these animals from simple polyketide precursors [13]. Diplopoda of the family Glomeridae have the characteristic habit of rolling up into a ball when danger threatens and enveloping themselves in a sticky, poisonous fluid that contains the toxic compound glomerin (Fig. 19.18 f) and the N-ethyl analogue homoglomerin; this is the only known quinazolinone of animals [13, 18, 90]. Of all the substances identified in the defence secretions of the diplopods, probably the most unusual are the N-terpene polyzonimin (Fig. 19.18 g) and the related tricyclic nitropolyzonimin from *Polyzonium rosalbum* [18, 75].

The **"paralysing mussel toxins"** are produced by dinoflagellates of the genus *Gonyaulax* and in these organisms find their way into mussels, crabs, echinoderms and other marine animals. The toxin primarily affects the central nervous system, and in about 8 % of cases leads to death by inhibiting respiration [75]. The first of these toxins to be described (in 1975) was saxitoxin (Fig. 19.18 h); further related compounds under the name gonyautoxin have since been discovered [155]. Under certain conditions, the sea snail *Babylonia japonica*, which is a popular Japanese delicacy, contains a toxin that has effects similar to those of saxitoxin, although the complicated structures of these toxins have little in common. The poison is found only in animals that originate from Suruga Bay, or are transplanted there, and is thus referred to as surugatoxin; it is safe to assume also in this case that the source is planktonic organisms [75, 155]. 6,6-Dibromindigo (Fig. 19.18 i), which as "Tyrian purple" is of great historical interest, is also an N-heterocyclic compound and is obtained from the purple snails (Muricidae) of the genera *Murex, Mitra* and *Nucella*. The name comes from the port of Tyre where the Phoenicians first built up a successful industry to produce the dyestuff. Tyrian purple was more valuable than silver in those times, and in ancient Rome only senators and emperors were allowed to wear purple togas. The compound arises initially through the effect of sunlight and atmospheric oxygen on a green precursor, tyriverdin (Fig. 19.18 j), which accumulates in the purple gland (hypobranchial gland) of the snail. When homogenized tissues of the snail are exposed to sunlight they gradually change from green to blue and red, and eventually to purple [31].

One of the most famous poisons is the **tetrodotoxin** (TTX) (Fig. 19.18 k) which is present in more than 50 species of puffer fish (Tetraodontidae), burr fish (Diodontidae) and sunfish (Molidae). The toxin is mainly found in internal organs such as the gonads and liver but does not occur in muscles. Preparation of the fish for consumption, prized in the raw state by Japanese gourmets, obviously requires special care; lethalities are still recorded. The LD_{50} is 80 μg/kg for mice, and TTX is thus one of the most poisonous non-protein substances; it blocks the sodium channels of the cell membrane without affecting the potassium channels [75, 90]. TTX is in no way restricted to the teleosts mentioned, and is also found in the Californian newt *Taricha torosa*, Costa Rican frogs of the genus *Atelopus* and numerous invertebrates, e.g. cephalopods, gastropods, asteroids, crustaceans, xiphosurans, turbellarians and nemertines [4, 86, 116, 132, 148]. This sporadic distribution strongly suggests that the toxin originates in bacteria, and in fact puffer fish that are raised in the aquarium are toxin-free; furthermore, the skin of the puffer fish *Fugu poecilonatus* has been shown to harbour bacteria of the genus *Pseudomonas*, which produce TTX [148]. Various TTX analogues have been identified in newts and in frogs of the genera *Atelopus* and *Brachycephalus* [165, 208].

The skin-gland poisons of the Central and South American frogs from the family Dendrobatidae are used in the production of poisoned arrows by some Indians and these poisons are also alkaloids. The first of these **dendrobatid toxins** to be described in detail was the batrachotoxin from *Phyllobates aurotaenia* (Fig. 19.18 l); with an LD_{50} of 2 μg/kg in mice this compound is extremely toxic. It inhibits neuromuscular transmission by rendering cells permeable to sodium without influencing potassium transport. Thus, it has exactly the opposite effect of TTX. In addition to batrachotoxin, the skin of *P. aurotaenia* contains three further closely related steroid alkaloids. Possession of this toxin is characteristic of all five species of the genus *Phyllobates*. Species of the related genus *Dendrobates* contain a series of toxic alkaloids with a somewhat simpler structure; these are known as pumiliotoxins A and C, histrionicotoxins and gephyrotoxins. In all, more than 200 dendrobatid toxins are now known; they may be subdivided into five classes according to the five basic substances mentioned [49, 75]. More recently, substances that are very similar to the *Dendrobates* alkaloids have been identified in anurans of completely unrelated groups: in the toad *Melanophryscus moreirae* (Bufonidae) from Brasil, and the frogs *Pseudophryne semimarmo-*

rata (Myobatrachidae) from Australia and *Mantella aurantiaca* (Ranidae) from Madagascar. It remains a matter of debate whether these complicated molecular structures evolved convergently more than once in different anuran groups, or whether at one time all anurans had this capacity but it was lost in most evolutionary lines.

19.6 Pyrrole Pigments

All animal pyrrole pigments are tetrapyrroles. In the porphyrins, four pyrroles are combined by four carbon bridges to give a ring system. According to the old numbering system of Fischer, the eight outer, substituted C atoms of the pyrroles are given the numbers 1 to 8, and the C bridges are given the Greek letters α to δ (Fig. 19.19). In the new International Union of Pure and Applied Chemistry (IUPAC) numbering systems, all the C atoms are numbered consecutively 1 to 20, and the N atoms are numbered 21 to 24. The linear bilins arise from the porphyrins by oxidative cleavage of a bridge C, usually of the α atom. These are in principle 1,19-dihydroxy compounds but are usually present in the tautomeric bislactam form (Fig. 19.20).

The **biosynthesis of the porphyrin ring system** always follows the scheme shown in the mammals (Fig. 19.19). Two molecules of 5-aminolaevulinic acid, produced from succinyl-CoA and glycine, combine to give porphobilinogen; four molecules of this pyrrole subunit are linked to give uroporphyrinogen, with the isomeric type III usually being produced instead of the substitution type I. Dehydrogenation at the C bridge, and modification of the substituents lead to uroporphyrin, coproporphyrin and protoporphyrin (Fig. 19.19), and chelation with Fe(II) leads to protohaem IX, the prosthetic group of haemoglobin. The degradation of haem in the vertebrate liver involves cleavage of the porphyrin ring system at the α bridge by the enzyme haemoxygenase, and the production of stoichiometric amounts of biliverdin IXα, CO and Fe. The haemoxygenases from the chicken liver and human liver agree in 62 % of their 296 amino acids [58]. Most bilins originate from protohaem IX by cleavage at the α bridge; however, bilins of types IXβ, IXγ and IXδ are also known and these arise by cleavage of the β, γ or δ bridge, respectively. Cleavage of the porphyrin ring system to give bilin always occurs on a metal complex such as haem, and bilins are therefore typical products of haemoglobin degradation. Where bilins contribute substantially to body coloration in the absence of haemoglobin, e.g. in some insects, it must be assumed that porphyrin-metal complexes exist that have no other function than that of an intermediate in pigment synthesis [96]. Haem proteins such as cytochromes, catalases and peroxidases are ubiquitous in all organisms, and thus the capacity for porphyrin biosynthesis is undoubtedly also widespread. This has been demonstrated in tracer experiments with several insects: the incorporation of ^{14}C-glycine into haem has been shown in *Chironomus* and *Buenoa*, into biliverdin IXα in *Mantis religiosa* and *Locusta migratoria*, and into biliverdin IXγ in *Pieris brassicae* [96]. Kinetoplastida of the genera *Trypanosoma* and *Crithidia* are unable to synthesize porphyrin and require haem derivatives as essential nutrients, so long as they are not host to endosymbionts with such biosynthetic ability. However, *T. cruzi* has several of the enzymes of this metabolic pathway and it would thus appear that these flagellates were originally able to synthesize prophyrin [157]. Apart from this, there is negligible comparative biochemical information about the enzymes of porphyrin and bilin synthesis.

Large amounts of **free porphyrins** are found only in a few animals. The earthworm *Lumbricus terrestris* owes the red colour of its dorsal surface to protoporphyrin laid down in the skin. The shells of some snails contain uroporphyrin, often type I. Protoporphyrin is found in the egg shells of some birds and coproporphyrin III occurs in the feathers; the glowing red feathers of the touraco are due to the copper chelate of uroporphyrin III [31]. **Bilins** are found in the bile of all vertebrates. The initial product of haemoglobin catabolism in the cells of the reticulo-endothelial system is biliverdin IXα; this is rapidly reduced to bilirubin, bound to serum albumin in the blood and transported to the liver, where it is conjugated with glucuronic acid. Mammals secrete predominantly bilirubin in the bile, whereas birds, reptiles and amphibians secrete biliverdin; fish bile contains either bilirubin or biliverdin according to the species [31, 60]. The small amounts of bilins with the IXβ structure, found, for example, in the bile of humans and Rhesus monkeys, possibly arise by the non-enzymatic cleavage of protohaem IX. However, about 17 % of the bilirubin in the bile of the eel *Anguilla japonica* is the IXβ isomer [156]. Bilins play a small role in the coloration of vertebrates; only the shells of some bird eggs are coloured green by biliverdin IXα. In contrast, the green surface coloration of many

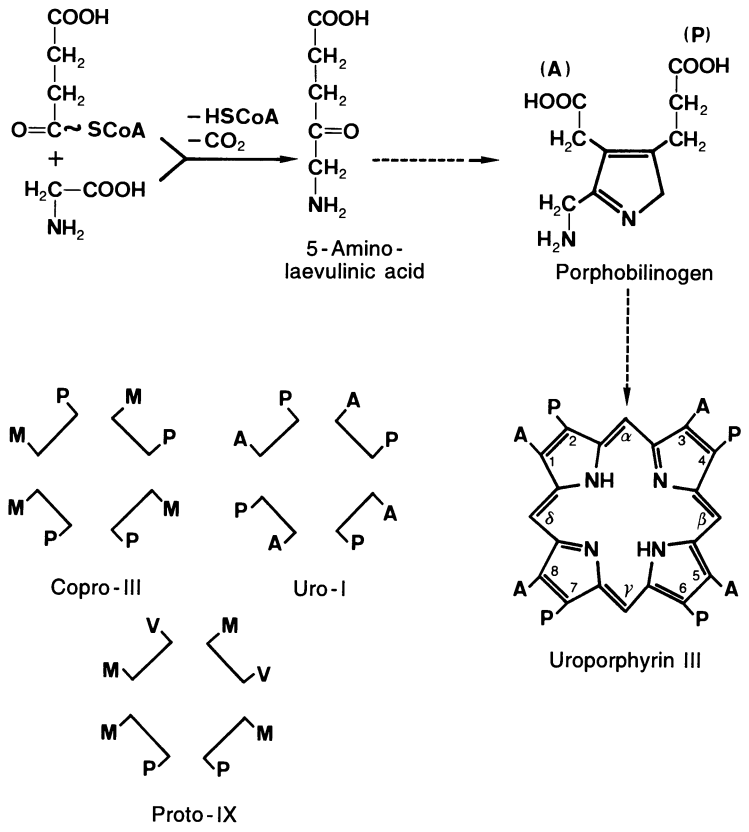


Fig. 19.19. Porphyrins: biosynthesis and structure [31]. The primary porphyrin in all animals is porphobilinogen, which is produced from succinyl-CoA and glycine via 5-aminolaevulinic acid. The most important porphyrins are uroporphyrin III, uroporphyrin I *(uro-I)*, coproporphyrin III *(copro-III)* and protoporphyrin IX *(proto-IX)*. The substituents of the outer C atoms are methyl groups *(M)*, acetyl residues *(A)*, propionyl residues *(P)* and vinyl groups *(V)*

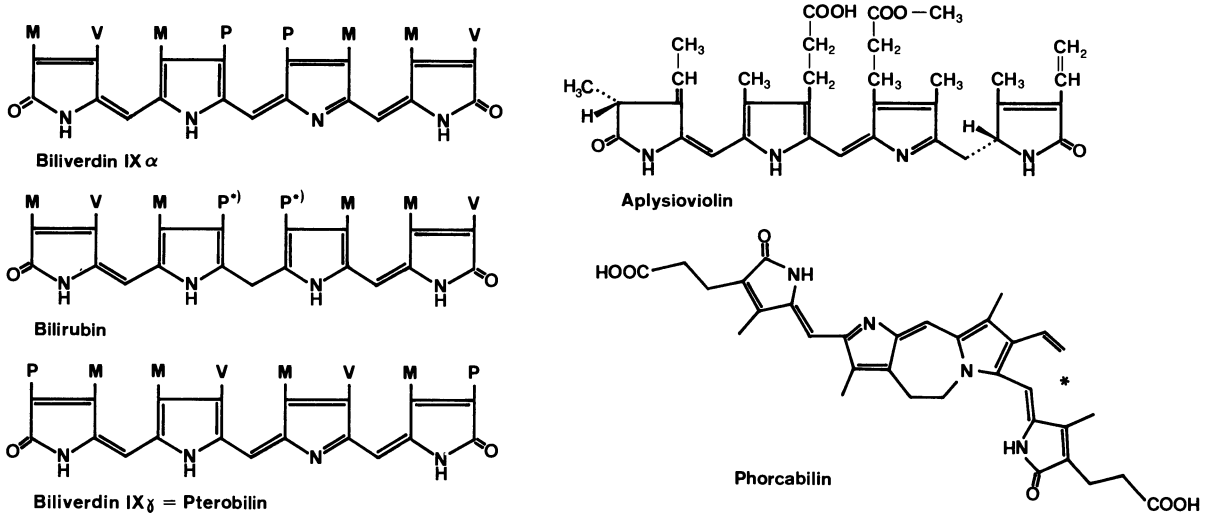


Fig. 19.20. Various animal bilins [31]. Removal of the α bridge in protoporphyrin IX leads to the blue-green biliverdin IXα, and removal of the γ bridge leads to biliverdin IXγ; reduction of biliverdin results in the orange-red bilirubin, which may carry β-glucuronyl residues on the propionic acid residues marked with •). The substituents are methyl groups *(M)*, propionyl residues *(P)* and vinyl groups *(V)*. Aplysioviolin (the methyl ester of phycoviolin) is found as a protein complex in the defence secretion of the sea slug *Aplysia* sp. Several lepidopterans produce the pentacyclic phorcabilin from which the hexycyclic sarpedobilin can arise by ring closure at the position marked with *

invertebrates is due to biliverdin IX or to the related mesobiliverdin (glaucobilin), in which the vinyl substituents are reduced to ethyl groups. Some insects have biliverdin IXγ (Fig. 19.20). Biliverdin IXα is characteristic of hemimetabolic insects such as the Odonata, Phasmida, Orthoptera and Mantodea, but is also found in the green lacewings *Chrysopa* sp. (Neuroptera) and the moth *Thaumatopoea;* otherwise, the Lepidoptera produce the isomer biliverdin IXγ (pterobilin). In addition to pterobilin, the papilionids and saturnids possess the pentacyclic phorcabilin (Fig. 19.20) and the hexacyclic sarpedobilin. Tracer experiments with *Actias selene* have shown that these pigments are produced from pterobilin; the animal clearly possesses a specific enzyme for this purpose, but the two compounds can also be formed non-enzymatically in vitro by the action of light. Sarpedobilin is a very reactive substance and can form stable complexes with many natural polyanions [14]. Biliverdin IXδ and neobiliverdin IXδ have been found in the ovaries of the marine snail *Turbo cornutus* [17].

Complexes between bilins and proteins (phycoerythrins and phycocyanins) are very common in the algae and are also found in some molluscs. One well-characterized example is aplysioviolin (Fig. 19.20), which is bound to protein in the violet defence secretion of the sea slug *Aplysia* sp. [31]. **Biliverdin proteins** are especially widespread in the insects. The blue haemolymph proteins from various lepidopterans, the bean bug *Riptortus clavatus* and the migratory locust *Locusta migratoria* owe their colour to non-covalently bound biliverdin, which has been shown to be the IXγ isomer in the "insectacyanin" of *Manduca sexta* and the bilin protein of *Pieris brassicae*. The bilin proteins of these two lepidopterans are homotetramers, the subunits of which belong to the retinol-binding-protein super-family. The sequences of the subunits in *M. sexta* (189 amino acids) and *P. brassicae* (173 amino acids) agree by 42.8 % [83, 178]. The biliverdin IXα proteins in the haemolymph of *L. migratoria* and *R. clavatus* are homotetramers with a different structure, having subunits of 83 and 76 kDa, respectively. The biliverdin-containing lipoglycoprotein from the haemolymph of *Trichoplusia ni* contains an apoprotein of 150 kDa [39, 91]. However, biliverdin proteins are not restricted to the insects. The haemolymph of the brine shrimp *Artemia salina* contains blue-green biliverdin IXα-containing glycolipoproteins (artemiocyanin) with a complicated quaternary structure, one subunit of which has proteinase activity [102]. In eels and other teleosts, the further metabolism of biliverdins is prevented by binding of the molecules to specific plasma proteins of very varied structure [30, 61]. Biliverdin IXα-proteins have also been detected in the skin of tree frogs.

19.7 Pterins

Compounds with the core structure of pteridine (Fig. 19.17) are ubiquitous as cofactors in cell metabolism: e.g. tetrahydrofolic acid is the coenzyme of C_1 metabolism, and tetrahydrobiopterin is the electron donor in the hydroxylation of aromatic amino acids [213]. Folic acid is an essential nutrient of vertebrates, insects and probably all other animals, whereas biopterin can be produced by animals [140]. The biosynthetic pathway in insects and lower vertebrates leads to a large number of pigments, mostly derived from 2-amino-4-hydroxypteridine (pterin; Fig. 19.21a) and less often from 2,4-dihydroxypteridine (lumazines, e.g. violapterin, shown in Fig. 19.21b). The numerous variants differ in the substituents at C-6 and C-7 as well as the oxidation state of the nitrogen atoms N-5 and N-8. The dimer drosopterin has an especially complicated structure (Fig. 19.21b). **Pterin biosynthesis** begins with guanosine triphosphate (Fig. 19.22); as well as in bacteria and mammals, the pathway has been studied in the fly *Drosophila melanogaster* and the butterflies *Pieris brassicae* and *Colias eurytheme* [31, 96, 140, 213]. In the first step, the 5-ring (imidazole ring) of purine is enlarged to a 6-ring (pyrazine ring) by opening of the imidazole ring, removal of C-8, and introduction of atoms C-1 and C-2 from ribose to give 7,8-dihydroneopterin triphosphate. The enzyme responsible for this rather complicated reaction, GTP cyclohydrolase I, has been characterized in bacteria, the chicken, *Drosophila melanogaster* and various mammals [198]. The second reaction step leads to 6-pyruvoyl-tetrahydropterin by removal of the three phosphates; the responsible enzyme, sepiapterin synthase A, better known as 6-pyruvoyl-H_4-pterin synthase, has also been partially purified from *Drosophila* [180]. In this fly, three biosynthetic pathways lead from 6-pyruvoyl-tetrahydropterin to sepiapterin and biopterin, to pterins like isoxanthopterin, and to drosopterin. Several of the enzymes involved have been partially purified and characterized [63, 96]. Two molecules of dihydroneopterin triphosphate are required for the synthesis of the

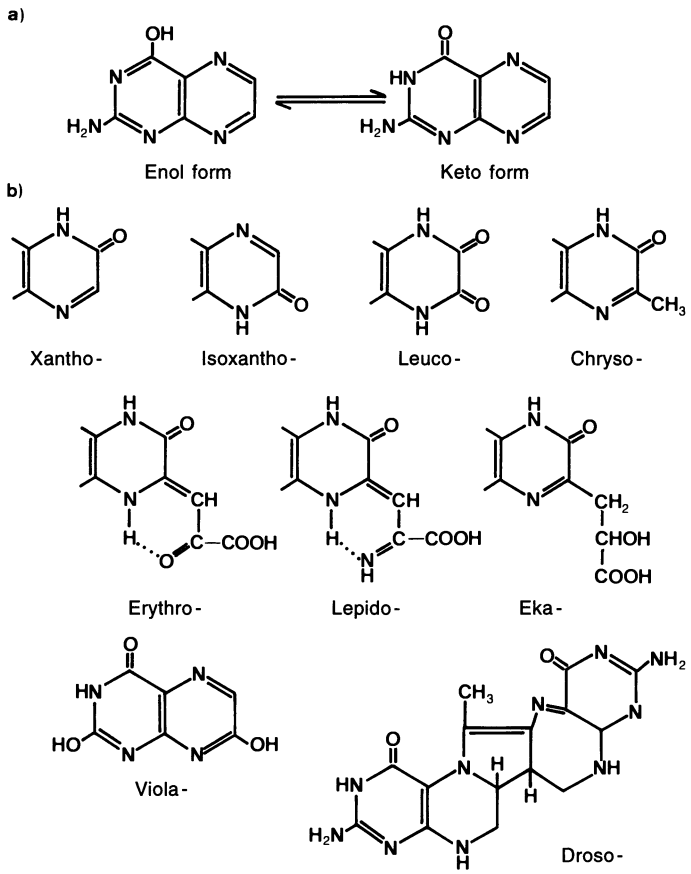


Fig. 19.21 a, b. Pterins [96]. **a** Tautomerism of pterin (2-amino-4-hydroxypteridine). **b** The structure of various pterins; in most cases, only the variable right half of the molecule is illustrated with the different substituents on C-6 and C-7 and the variable oxidation states of N-5 and N-8; the left half of the molecule has the structure shown in **a**

dimeric drosopterins, one of which is reduced to pyrimidodiazepin [96, 199]. In *Drosophila* numerous eye-colour mutants are known in which single enzyme defects change the spectra of pterins and ommochromes in the eye or cause the accumulation of specific intermediates [63]. The deamination of pterins leads to lumazines such as violapterin (7-hydroxylumazine; Fig. 19.21b), which has been identified in the honeybee and other insects [96].

The yellow to red body coloration of many **fish, amphibians and reptiles** stems from pterins stored in special pigment cells (xanthophores or erythrophores); these are sepiapterin or its dimer,

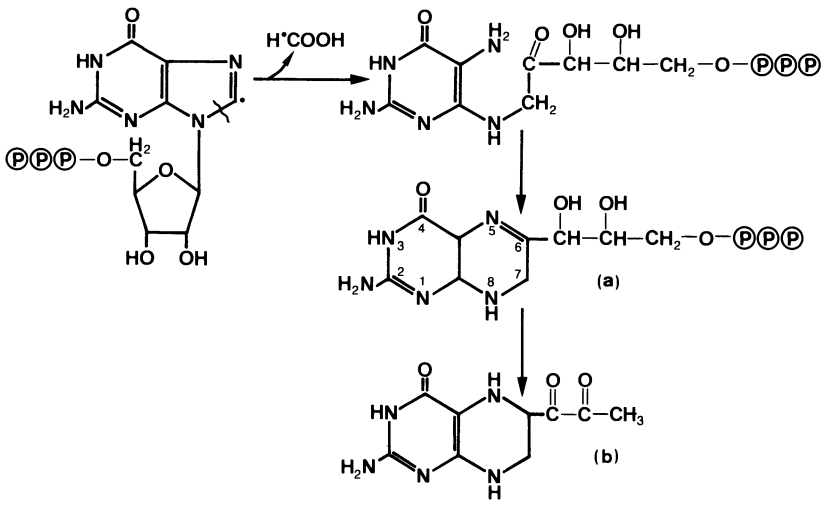


Fig. 19.22 a, b. Pterin biosynthesis [96, 213]. Guanosine trisphosphate gives rise to 7,8-dihydroneopterin triphosphate **(a)** and 6-pyruvoyl-tetrahydropterin **(b)**; see text for further details

pigments which are comparable with drosopterin. In the **insects**, xanthopterin and isoxanthopterin (Fig. 19.21b) in particular are found in almost all orders, with an especially rich spectrum of pterins in bugs and butterflies, notably in the whites (Pieridae). Pterins are mostly localized in the epidermal cells but are also deposited in the form of extracellular crystalline "pterinosomes" in lepidopteran wing scales. White leucopterin is found in large amounts in the cabbage whites *Pieris brassicae* and *P. rapae*, yellow chrysopterin occurs in, for example, the brimstone *Gonepteryx rhamni*, and red erythropterin is found in the aurora *Euchloe (Anthocharis) cardamines*. The yellow of the wasp *Vespula vulgaris* stems from xanthopterin, but the syrphids (hoverflies) have no pterins in their integument, despite their wasp-like appearance. In the dipterans, pterins are found together with ommochromes in the pigment cells of the eye; in particular, sepiapterin and, in *Drosophila*, drosopterin are present. The eyes of many lepidopterans contain ekapterin and lepidopterin, but the dimeric pterorhodin is found only in *Ephestia* and *Ptychopoda;* it is also present in the wing scales of the Pieridae [31, 96]. The colour of pterin-containing tissue is often different from that of the isolated pigment; this may be due to the bathochromic effects of protein binding or to stabilization of the tautomeric keto form [31].

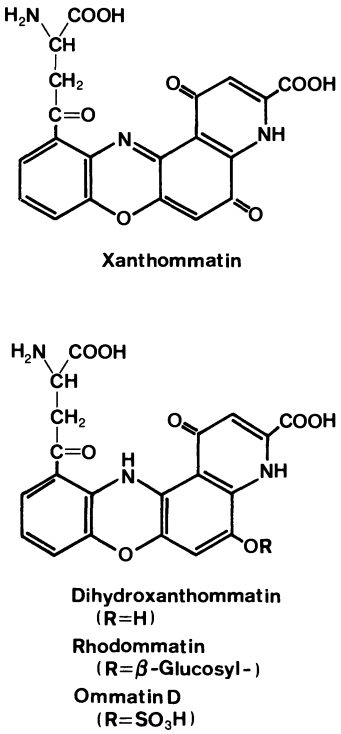


Fig. 19.23. Ommochromes [96]

19.8 Ommochromes and Papiliochromes

The only animal pigments to be based on the N,O-heterocyclic phenoxazine (Fig. 19.17) are the ommochromes, which were first discovered in the complex eyes of insects and other arthropods. They are derived from the tryptophan metabolite 3-hydroxykynurenine (see Fig. 12.9, p. 425). **Xanthommatin**, the simplest of these pigments, is formed non-enzymatically by the oxidation of 3-hydroxykynurenine with hexacyanoferrate (III). Xanthommatin has the unusual property of changing colour from yellow to red upon reduction. For **rhodommatin and ommatin D** the reduced state is fixed by conjugation of dihydroxyxanthommatin with glucose or sulphate (Fig. 19.23). In addition to the above-mentioned ommatins, there are also the sulphur-containing, high molecular weight **ommins** and the **ommidins**, which also contain sulphur but have lower molecular weights. The exact structures of these substances await determination. The first steps of ommochrome synthesis from tryptophan to 3-hydroxykynurenine are well known from investigations of eye-colour mutants of the Mediterranean flour moth *Ephestia kühniella* and *Drosophila melanogaster* (p. 425, see Fig. 12.9). The further steps from 3-hydroxykynurenine to the ommatins, ommatidins and ommins, however, have not yet been completely defined, although a "phenoxazinone synthase" from *Drosophila* has been partially purified [96]. Somewhat more is known about the xanthommatin reductase of *Drosophila*, an NADH-dependent enzyme of 55 kDa which is specifically inhibited by pterins [158]. Ommochromes are found in the eyes of all arthropods and are common in the integument. They have also been reported in the eggs of the echiurid *Urechis caupo*, in the tentacles of sedentary polychaetes (Sabillidae), in the integument of the nemertine *Linaeus* and in the eyes of the jellyfish *Spirocodon*. Insect eyes have predominantly ommins; however, the eyes of flies from the super-family Cyclorrhapha (e.g. *Musca, Calliphora, Drosophila)* contain only xanthommatin; ommatins are mainly found in secondary pigment cells and ommins are present in primary pigment cells and receptor cells. Ommins predominate in the integument of most insects, but they are never found in the wing scales of lepidopterans. The

ommochrome complement of the orthopterans has various special features: the eyes of the Gryllidae and Acrididae contain the unusual ommidins, and acridiommatins, whose structure is not yet known, are found only in the integument of the Acrididae and Tettigoniidae. The conjugated ommatins rhodommatin and ommatin D (Fig. 19.23) are found only in the Lepidoptera. They are the main pigment in the wings of the Nymphalidae, and are also found in the faeces of the larvae and in some internal organs. Apart from the water-soluble conjugated ommatins, ommochromes in cells may be found in the form of specific high-density granules together with calcium phosphate and proteins. The biological importance of the ommochromes is still a matter of debate. Their production possibly serves to eliminate surplus tryptophan at certain stages in development; they may have receptor functions in the eye and confer body colour [96].

The tryptophan metabolite kynurenine also gives rise to a group of pigments that are found only in the wing scales of the family Papilionidae and are known as **papiliochromes**. The light-yellow papiliochromes II and III, the dark-yellow papiliochrome M and the red-brown papiliochrome R have been described. They are all complexes of kynurenine, β-alanine and dopamine, but only in the case of papiliochrome II is the exact structure known. Enzyme preparations from the left colleterial gland of the praying mantis *Tenodera aridifolia* and from the cuticula of the giant silkmoth *Dictyoploca japonica* can produce papiliochrome II from N-β-alanyldopamine and kynurenine [192, 204].

19.9 Melanins

The term melanins was coined for the black pigments that are widely distributed in vertebrates and invertebrates, and are polymeric and insoluble in all solvents. Similar in structure to these black **eumelanins** are the brown, red and yellow phaemolanins that are frequently present in mammal hair and bird feathers. The phaeomelanins, like the eumelanins, contain about 9% nitrogen, but also 9–10% sulphur; furthermore, they are soluble in dilute alkali. Melanins are practically the only pigments that contribute to hair and body coloration in mammals, and are of great importance as a feather pigment in birds. Phaeomelanins are apparently restricted to mammals and birds, whereas the black eumelanins also play an important role in the body coloration of fish and many invertebrates. Light-absorbing melanin layers are the basis of colour generation by physical effects (structure colours), and melanins also serve to protect photoreceptors as well as being present in internal organs. The ink of the cephalopods is pure melanin and is sold as Indian ink under the name "sepia" [31, 96]. Melanins are mostly produced in special melanocytes or melanophores. It should not be assumed that all dark pigments are melanins; ommochromes may also produce dark body coloration, e.g. in the insects. Sclerotization of insect cuticulae can lead to dark structures (p. 430). However, although dopa derivatives play a role in both cases, melanization and sclerotization are two completely different processes. There exist albino forms of the migratory locust without melanin but with a normal sclerotized cuticula [96].

The insolubility of the melanins, their polymeric nature and their binding to proteins make for great difficulties in structure determinations. Clearly homogenous fractions cannot be produced, spectroscopic methods cannot be applied, and defined fragments cannot be obtained. Present-day views of **melanin structure** are based, on the one hand, on analysis of hydrolysis products and, on the other hand, on examination of polymers produced in vitro from various precursors by the action of phenoloxidase. According to the classical Raper-Mason model, eumelanin is a linear polymer of indole-5,6-quinone, which is produced in a series of oxidative reactions from tyrosine. However, both natural and synthetic eumelanins contain 5,6-dihydroxyindole, the 2-carboxylic acids of quinone and hydroquinone, and the corresponding pyrrole derivatives (Fig. 19.24a) [87]. The pyrrole units arise by peroxidative ring cleavage from indole-5,6-quinone. The first two steps of **melanin biosynthesis** from tyrosine to dopa and further to dopa quinone, and possibly also the first polymerization steps, are catalysed by phenoloxidases (tyrosinases). The reactions leading from dopa quinone imine to 5,6-dihydroxyindole require a thermolabile dopa quinone imine conversion factor (QICF) of unknown structure [6]. The sulphur-containing phaeomelanins arise from the reaction of dopa quinone with cysteine to give 5-cysteinyl dopa (Fig. 19.24b). The predominant SH compound of the cells, however, is not free cysteine but reduced glutathione. Thus, it is possible that the primary product is glutathionyl dopa and that this is further metabolized to cysteinyl dopa by the action of γ-glutamyl transpeptidase and peptidase [88].

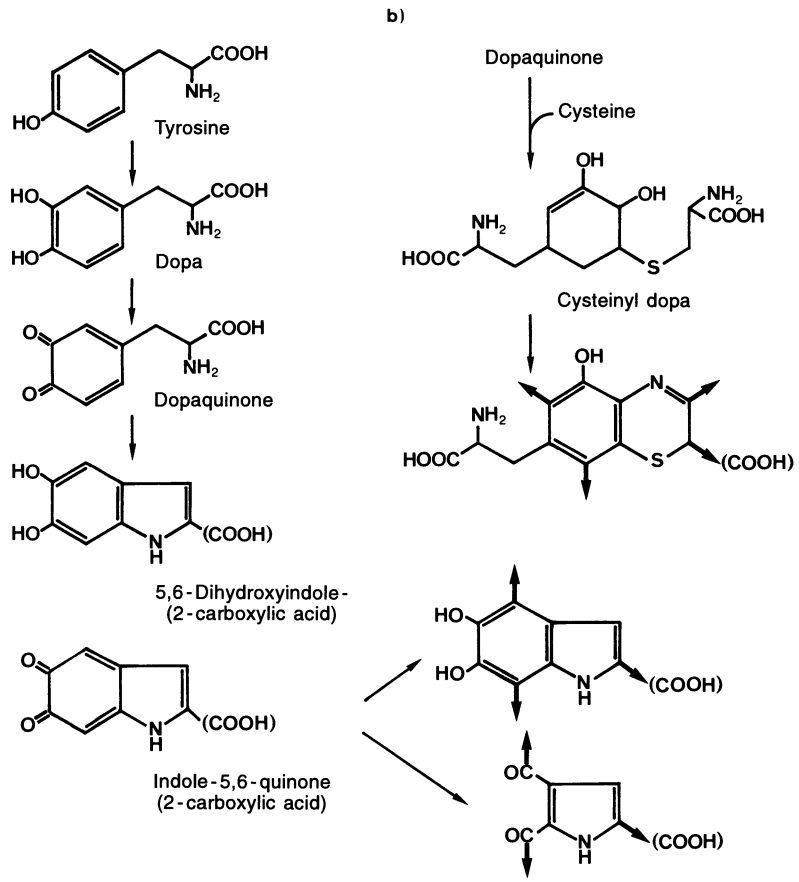


Fig. 19.24 a, b. The biosynthesis of the basic components of **a** eumelanins and **b** phaeomelanins [87, 88]. The carboxyl groups at C-2 may be present or absent; the *thick arrows* on the melanin components indicate possible bonding sites in the polymer

The **phenoloxidases** are copper enzymes which have been studied mainly in fungi and higher plants, but also in the insects and amphibians. The various phenoloxidases from the haemolymph and cuticula of insects and which are involved in sclerotization were discussed on p. 433. Animal phenoloxidases occur mainly as inactive pro-enzymes. The pro-phenoloxidase from the skin of the frog *Rana ridibunda* is activated by partial proteolysis. The pro-enzyme is a dimer of 68 kDa and the active enzyme is a tetramer with subunits shortened to 62 kDa by cleavage of a short peptide. Conflicting information is available for the phenoloxidase from the skin of the clawed frog *Xenopus laevis;* in this case the isolated pro-enzyme is activated not by trypsin but by anionic detergents. The active enzyme here is a glycoprotein of 175 kDa [201]. An interesting source of tyrosinase is cephalopod ink; the isolated enzyme from *Octopus vulgaris* has a mass of 205 kDa and that from *Sepia officinalis* and *Loligo vulgaris* 125–135 kDa. These enzymes, like that from the frog skin, can also produce melanin from D-tyrosine; their affinity for the D-isomer is greater than that for the L-isomer.

19.10 Bioluminescence

There are several thousand species of luminescent animals, most of them living in the sea; almost all animals that live at depths below 700 m have luminescent properties. There is only one known freshwater luminescent species, the New Zealand limpet *Latia neritoides*. Of the many luminescent species of terrestrial animals, probably the best known are the fireflies; however there are other insect groups that have this characteristic, as well as collembolans, diplopods, chilopods and earthworms (oligochaetes). In the marine habitat bioluminescence is found among animals of very different groups: radiolarians, dinoflagellates, hydrozoans, ctenophorans, nemertines, polychaetes, pantopods, molluscs, echinoderms, and tunicates (Thaliacea), but in particular cephalopods and teleosts. Bioluminescence fulfils various **biological functions:** intraspecific communication (for fireflies, several fish, and the polychaete *Odontosyllis*), protection from attack (for the ostracod *Cypridina* and cephalopods), attraction of prey (in the angler fish), and illumination of the

surroundings ("flashlight fish"). The great importance of self-generated light in the life of animals is indicated by the evolution of complicated aids for the amplification, focusing or temporary shading of light. On the other hand, the biological role of the luminescent capacity is in many cases far from clear. In luminescent bacteria and unicellular organisms the phenomenon may simply be part of a metabolic function, e.g. detoxification of oxygen [79, 130].

In some marine animals (fish, cephalopods, tunicates) the light capacity comes from a symbiosis with luminescent bacteria, but in most cases it is generated by the animal itself. The biochemical mechanisms of bioluminescence are very varied, and it would appear that the phenomenon has arisen on several different occasions during evolution. Common to all bioluminescence reactions is that a reactive molecule (luciferin) reacts with oxygen and is thereby brought to an excited state; the excitation energy is released as light energy. It appears that the reactions always proceed via peroxide as the intermediate. So far, seven different **luciferins** are known in detail, five from metazoans and one each from the dinoflagellates and bacteria (Fig. 19.25). The emission maxima of the excited substances usually lie in the range 460–500 nm (blue-green), but for the firefly luciferin it is 560 nm (yellow). However, in many cases fluorescing proteins are present which shift the maximum of the emitted light into the red part of the spectrum [35]. Almost nothing is

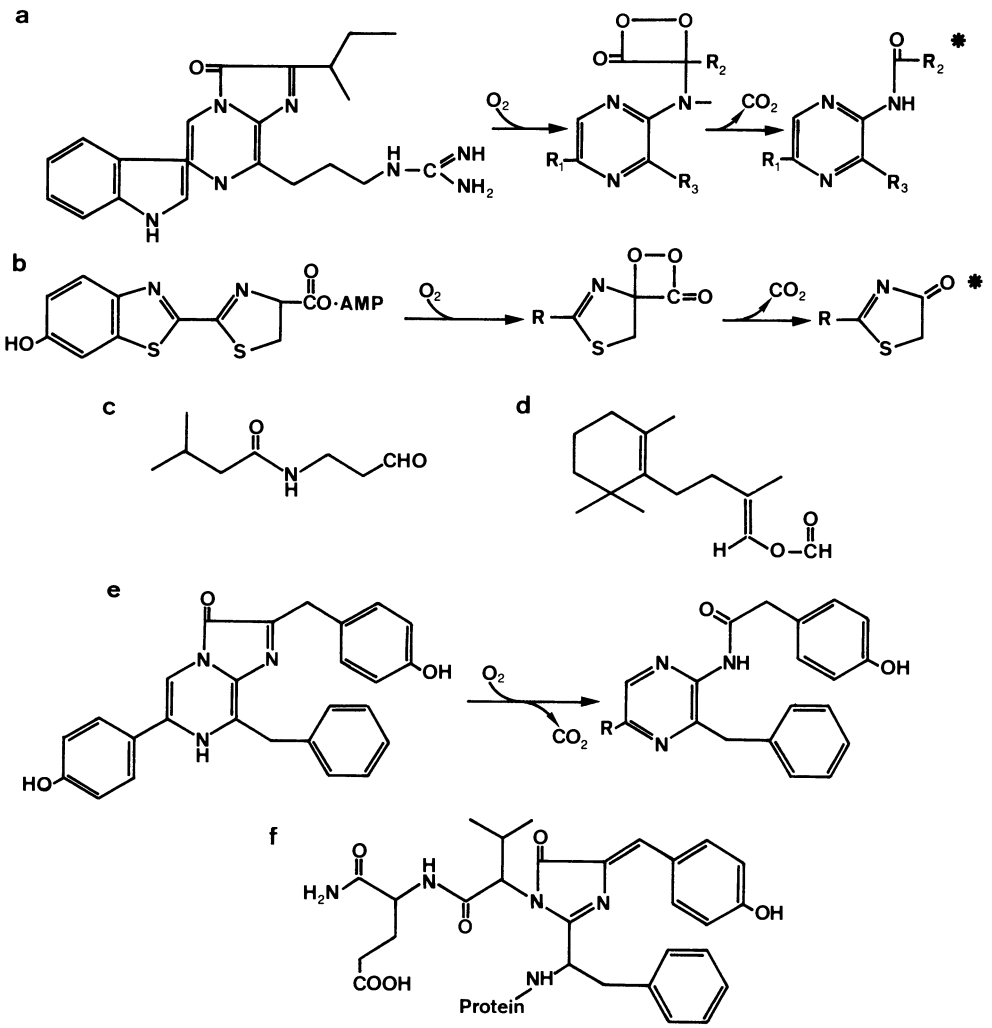


Fig. 19.25 a–f. Luciferins and chromophores of bioluminescence systems [79, 130, 131, 170]. **a** Luciferin from *Cypridina hilgendorfii* (Crustacea: Ostracoda); **b** firefly luciferin; **c** luciferin from *Diplocardia longa* (oligochaete); **d** luciferin from *Latia neritoides* (gastropod); **e** coelenterazine; **f** chromophore of the green-fluorescing protein (GFP) of *Aequorea aequorea*

known about the regeneration or biosynthesis of luciferin or about the enzymes (luciferases) involved in the light reaction. Details of the regulation of the light generation are available for only a few cases [79, 130, 131].

A simple **bioluminescence system** is found in *Cypridina hilgendorfii*. Enormous numbers of this ostracod inhabit the Sea of Japan, and they emit a cloud of a luminescent substance when danger threatens; in this way they obscure their exact whereabouts from the aggressor. The cloud of light arises from the mixing of two secretions, one of which contains luciferin and the other luciferase. With the aid of the luciferase, the luciferin reacts with oxygen; the light-emitting product and free carbon dioxide are formed, probably via the peroxidic intermediate with the 4-ring dioxethanone whose structure is shown in Fig. 19.25a [79, 130]. A simple two-component system of luciferin and luciferase is also responsible for the bioluminescence of the earthworms (oligochaetes), some species of which react to stimuli by the exudation of a luminous fluid from dorsal pores. This light system is apparently located in the free chloragogue cells. The only detailed example is that of *Diplocardia longa*. The luciferin is 3-(isovaleryl-amino)propanol (Fig. 19.25c), and the luciferase is a glycoprotein of 300 kDa which carries up to four copper atoms per molecule [154]. The bioluminescence system of the small limpet *Latia neritoides* from New Zealand, the only freshwater luminescent species known, has unfortunately been only partially characterized. The luciferin contains no nitrogen and has a very unusual structure with an enol formiate group (Fig. 19.25d). The light reaction requires in addition an apparently catalytic "purple protein" and a reducing agent; the reaction produces carbon dioxide and free formic acid [130].

Amongst the **insects**, bioluminescence is found in the larvae of the fungus flies (families Platyuridae, Bolitophilidae) and in a South American cicada (*Fulgora lanternaria*), but occurs mainly in the numerous beetle species of five families (Lampyridae, Elateridae, Phengodidae, Drilidae and Rhagopthalmidae). Biochemically, the best-known system is that of the lampyrid *Photinus pyralis*. However, the information on this species does not necessarily apply only to the Lampyridae; the Elateridae have been shown to possess the same luciferin, and it is probable that the light reaction is similar for all beetles [44]. The reaction scheme for *Photinus* is more complicated than that of *Cypridina* in that the luciferin must be activated before the oxidative reaction. This is accomplished by the transfer of an adenyl residue from Mg-ATP to the carboxyl group of luciferin, and the release of pyrophosphate. This process corresponds to the activation of fatty acids and amino acids in the production of acetyl-CoA and aminoacyl-tRNA. The reaction of the luciferin adenylate with oxygen leads via a peroxidic intermediate with the 4-ring dioxethanone and the dianionic excited molecule to oxyluciferin (Fig. 19.25b). Both reaction steps are catalysed by the luciferase and thus take place on the same enzyme-substrate complex. Firefly luciferin can be synthesized from 2-cyano-6-hydroxybenzthiazol and cysteine, whereby an excitable molecule is produced only with D-cysteine; luciferin produced from L-cysteine is converted to oxyluciferin by luciferase but no light is produced. The luciferases of the Japanese firefly *Luciola cruciata* and the North American *Photinus pyralis* agree by up to 67% in their sequences of 548–549 amino acids [124]. The firefly luciferase shows marked product inhibition; the light reaction ceases already after the conversion of a few luciferin molecules, which gives a light flash of only about 5-5 duration. The luciferases of other Lampyridae are related immunologically to that of *P. pyralis;* the differences in immunological cross-reactivity in the larval and adult luciferins of *Photuris pennsylvanica* suggest the existence of development-dependent isoenzymes. The bioluminescence system of *P. pyralis* has in vivo and in vitro an emission maximum in the yellow-green at 560 nm. Other firefly species produce light of somewhat different wavelengths, and this can only be due to differences in the luciferase because the luciferins are identical. Artificially produced substituted luciferins show alterations in emission maxima due to changes in enzyme binding, e.g. 6-aminoluciferin emits in the red [131]. ATP activation is also found in the light system of the diplopod *Luminodesmus sequoiae*, although the luciferin is not known.

The bioluminescence system of the hydromedusa *Aequorea aequorea* consists of two proteins, the photoprotein **aequorin** and the green-fluorescing protein (GFP). Aequorin has coelenterazine (Fig. 19.25e) as the prosthetic group. This is similar to the *Cypridina* luciferin in that it also contains an imidazopyrazine skeleton; the light-generating reactions are correspondingly similar. The special feature of the *Aequorea* system is that the peroxidic, excited intermediate produced in the reaction with oxygen forms a stable compound with the protein, and this only dissociates in the presence of calcium with the

simultaneous emission of light. This calcium-induced light production also occurs in the complete absence of oxygen. Isolated aequorin has an emission maximum at 469 nm (blue); however, the living animal generates green light at 509 nm because of interference by the GFP. The chromophore of the GFP (Fig. 19.25 f) is closely related to coelenterazine. Analysis of the cDNA shows that the aequorin polypeptide chain consists of 196 amino acids. In the directly obtained sequence, seven amino acids missing from the N-terminus were apparently lost during isolation of the protein. The aequorin sequence has an internal periodicity with three homologous Ca^{2+}-binding domains with the typical EF-hand structure (p. 348). Thus, aequorin belongs to the same protein super-family of Ca^{2+}-binding proteins as do calmodulin and troponin C. The dissociation constant of the Ca^{2+}-binding site is about 10^{-7} mol/l. Due to its high Ca^{2+} sensitivity, aequorin is used as a calcium detection reagent. *Aequorea aequorea* possesses at least three isoforms of aequorin; these differ in 23 amino acids and are encoded by different genes [38, 146].

Ca^{2+}-sensitive proteins that are comparable with aequorin are found in many cnidarians, ctenophorans and in species of the unicellular radiolarians. As for aequorin, they are named after the genera in which they are found; in the absence of sequence data, little can be said about their phylogenetic relationships. Obelin is found in the hydroid polyp *Obelia*, phialidin and halistaurin occur in the hydromedusae *Phialidium* and *Halistaura*, mnemiopsin and berovin are present in the ctenophorans *Mnemiopsis* and *Beroe*, and thalassicolin occurs in the radiolarian *Thassicola* [169]. In contrast to aequorin, mnemiopsin is inactivated by light, probably because coelenterazine and oxygen dissociate from the protein [3]. In the hydrozoans, the photoprotein is always accompanied by a GFP. In the anthozoan *Renilla reniformis*, the luciferin (coelenterazine) is also bound to a Ca^{2+}-triggered luciferin-binding protein (LBP) and can only generate light with luciferase and oxygen in the presence of calcium. The LBP belongs to the same super-family of calcium-binding proteins as do calmodulin and aequorin, and it carries three EF-hand domains in its sequence of 184 amino acids [105]. The *Renilla* luciferase, whose unique sequence of 314 amino acids was determined from the cDNA, catalyses in vitro the oxidative decarboxylation of coelenterazine with the production of blue light (λ = 480 nm). *Renilla* also possesses a GFP which shifts the colour of the bioluminescence in the living animal towards the green at 509 nm [117]. Coelenterazine is in no way restricted to the so-called coelenterates (i.e. cnidarians and ctenophorans), but is also found as the luciferin in fish, cephalopods and crustaceans. In addition, it is present in non-luminescent species, especially of fish and crustaceans, where it is probably taken up in the diet [79, 152, 170]. The light organs of the cephalopod *Watasenia scintillans* contain sulphated coelenterazine, whereas the hepatopancreas has the free substance. The light reaction is ATP dependent but not known in detail [186].

References

1. Ahmad S, Kirkland K. E. and Blomquist G. J.: Evidence for a sex pheromone metabolizing cytochrome P-450 mono-oxygenase in the housefly. Arch. Insect Biochem. Physiol. 6: 121–140 (1987)
2. Ali M. F. et al.: The Dufour substances of the workers of Formica fusca and Formica lemani (Hymenoptra: Formicidae). Comp. Biochem. Physiol. Pt. B 88: 59–63 (1987)
3. Anctil M. and Shimomura O.: Mechanism of photoinactivation and re-activation in the bioluminiscence system of the ctenophore Mnemiopsis. Biochem. J. 221: 269–272 (1984)
4. Aarakawa A. E. et al.: Tetrodotoxin and related substances in a ribbon worm Cephalotrix linearis (nemertean). Toxicon 28: 1083–93 (1990)
5. Arsequell G., Fabrias G. and Camps F.: Sex pheromone biosynthesis in the processionary moth Thaumetopoea pityocampa by delta-13 desaturation Arch. Insect Biochem. Physiol. 14: 47–56 (1990)
6. Aso Y., Imamura Y. and Yamasaki N.: Further studies on DOPA chinone imine concersion factor from cuticles of Manduca sexta (L.). Insect Biochem. 19: 401–407 (1989)
7. Atta-ur-Rahmann and Le Quesne P. W.: Natural products chemistry III. Springer-Verlag, New York 1988
8. Attygalle A. B. et al.: New chemicals from the Dufour gland of the formicine ant Lasius niger (Hymenoptera: Formicidae). Insect Biochem. 17: 219–225 (1987)
9. Avila C. et al.: Biosynthetic origin and anatomical distribution of the main secondary metabolites in the nudibranch mollusc Doris verrucosa. Comp. Biochem. Physiol. Pt. B 97: 363–368 (1990)
10. Azumi K., Yokosawa H. and Ishii S.: Halocyamines: Novel antimicrobial tetrapeptide-like substances isolated from the hemocytes of the solitary ascidian Halocynthia roretzi. Biochemistry 29: 159–165 (1990)
11. Baden D. G.: Brevetoxins: unique polyether dinoflagellate toxins. Faseb J. 3: 1807–17 (1989)
12. Baehr W. et al.: Isolation and analysis of the mouse opsin gene. FEBS Letters 238: 253–256 (1988)
13. Barbier M.: Introduction to chemical ecology. Longman, Harlow UK 1979
14. Barbier M.: Butterfly and moth neopterobilins: sarpedobilin as a natural metachromatic pigment. Comp. Biochem. Physiol. Pt. B 84: 619–621 (1986)

15. Bell W. J. and Cardé R. T.: Chemical ecology of insects. Chapman & Hall, London 1984
16. Bendich A. and Olson J. A.: Biological actions of carotenoids. Faseb J. 3: 1927–32 (1989)
17. Benedikt E. et al.: Biliverdin IX-delta and neobiliverdin IX-delta, isolated from the ovaries of the marine snail, Turbo cornutus. Eur. J. Biochem. 175: 643–648 (1988)
18. Bettini S. (ed.): Arthropod venoms. Hb. exp. Pharmacol. Vol. 48. Springer, Berlin 1978
19. de Bianchi A. G. and Marinotti O.: Physiological pattern and further characterization of the haemolymph violet carotenoprotein of the fly Rhynchosciara americana. Comp. Biochem. Physiol. Pt. B 83: 637–642 (1986)
20. Billen J. P. J., Jackson B. D. and Morgan E. D.: The contents of the pygidial gland of the primitive ant Nothomyrmecia macrops (Hymenoptera, Formicidae). Experientia 44: 794–797 (1988)
21. Bjostad L. B. and Roelofs W. L.: Sex pheromone biosynthetic precursors in Bombyx mori. Insect Biochem. 14: 275–278 (1984)
22. Blum M. S.: Alarm pheromones. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 9, pp. 193–224. Pergamon Press, Oxford 1985
23. Blum M. S. et al.: Oxygenated compounds in beeswax: Identification and possible significance. Comp. Biochem. Physiol. Pt. B 91: 581–583 (1988)
24. Boppré M.: Insects pharmacophagously utilizing defensive plant chemicals (pyrrolizidine alkaloids). Naturwissenschaften 73: 17–26 (1986)
25. Borst D. E. et al.: Interphotoreceptor retinoid-binding protein. Gene characterization, protein repeat structure, and its evolution. J. Biol. Chem. 264: 1115–23 (1989)
26. Braconnier M. F. et al.: (Z)-1,17-diaminooctadec-o-ene, a novel aliphatic diamine from Coccinellidae. Experientia 41: 519–520 (1985)
27. Bradshaw J. W. S.: Insect natural products – Components derived from acetate, shikimate and amino acids. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 11, pp. 655–703. Pergamon Press, Oxford 1985
28. Breer H., Krieger J. and Raming K.: A novel class of binding proteins in the antennae of the silk moth Antheraea pernyi. Insect Biochem. 20: 735–740 (1990)
29. Bridges C. D. and Alvarez R. A.: The visual cycle operates via an isomerase acting on all-trans retinol in the pigment epithelium. Science 236: 1678–80 (1987)
30. Brittain T. and Wells R. M. G.: Characteristics of a biliprotein complex in the serum of the stargazer fish (Genyagus novazelandiae: Uroscopidae) and its possible role in camouflage. Comp. Biochem. Physiol. Pt. B 92: 303–306 (1989)
31. Britton G.: The biochemistry of natural pigments. Cambridge Univ. Press, Cambridge 1983
32. Brush A. H.: Metabolism of carotenoid pigments in birds. Faseb J. 4: 2969–77 (1990)
33. Buckingham J. (ed.): Dictionary of natural products. Chapman & Hall, New York 1993
34. Burger B. V. and Pretorius P. J.: Mammalian pheromone studies. VI. Compounds from the preorbital gland of the blue duiker, Cephalophus monticola. Z. Naturforsch. Sect. C 42: 1355–57 (1987)
35. Campbell A. K. and Herring P. J.: A novel red fluorescent protein from the deep sea luminous fish Malacosteus niger. Comp. Biochem. Physiol. Pt. B 86: 411–417 (1987)
36. Cariello L. et al.: Antimicrobial activity of avarol, a sesquiterpenoid hydroquinone from the marine sponge, Dysidea avara. Comp. Biochem. Physiol. Pt. B 71: 281–283 (1982)
37. Chang C. W. J. and Scheuer P. J.: Biosynthesis of marine isocyanoterpenoids in sponges. Comp. Biochem. Physiol. Pt. B 97: 227–233 (1990)
38. Charbonneau H. et al.: Amino acid sequence of the calcium-dependent photoprotein aequorin. Biochemistry 24: 6762–71 (1985)
39. Chinzei Y. et al.: Purification and characterization of biliverdin-associated cyanoprotein from eggs and hemolymph of the bean bug, Riportus clavatus (Heteroptera: Alydidae). Insect Biochem. 20: 545–555 (1990)
40. Cimino G. et al.: Observations on the toxicity and metabolic relationships of polygodial, the chemical defense of the nudibranch Dendrodoris limbata. Experientia 41: 1335–36 (1985)
41. Cimino G. et al.: Oxytoxins, bioactive molecules produced by the marine opisthobranch mollusc Oxynoe olivacea from a diet-derived preursor. Experientia 46: 767–770 (1990)
42. Clark R. J. H. et al.: The carotenoproteins of the starfish Linckia laevigata (Echinodermata: Asteroidea): A resonance Raman and circular dichroism study. Comp. Biochem. Physiol. Pt. B 95: 847–853 (1990)
43. Clawitter J. et al.: A novel family of progesterone-induced, retinol-binding proteins from uterine secretions of the pig. J. Biol. Chem. 265: 3248–55 (1990)
44. Colepicolo P., Pagni D. and Bechara E. J. H.: Luciferin biosynthesis in larval Pyrearinus termitilluminans (Coleoptera: Elateridae). Comp. Biochem. Physiol. Pt. B 91: 143–147 (1988)
45. Connell G. J. and Molday R. S.: Molecular cloning, primary structure, and orientation of the vertebrate photoreceptor cell protein peripherin in the rod outer segment disk membrane. Biochemistry 29: 4691–98 (1990)
46. Cowman A. F., Zuker C. S. and Rubin G. M.: An opsin gene expressed in only one photoreceptor cell type of the Drosophila eye. Cell 44: 705–710 (1986)
47. Crabb J. W. et al.: Cloning of the cDNAs encoding the cellular retinaldehyde-binding protein from bovine and human retina and comparison of the protein structures. J. Biol. Chem. 263: 18688–92 (1988)
48. Crews P. and Naylor S.: Sesterpenes: An emerging group of metabolites. Progr. Chem. organic natural Products Vol. 48, pp. 203–269. Springer, Wien 1985
49. Daly J. W.: Alkaloids of neotropical poison frogs (Dendrobatidae). Progr. Chem. organic natural Products Vol. 41., pp. 205–340 Springer, Wien 1982
50. Davis R. H. and Nahrstedt A.: Cyanogenesis in insects. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 11, pp. 635–654. Pergamon Press, Oxford 1985
51. Davis R. H. and Nahrstedt A.: Biosynthesis of cyanogenic glucosides in butterflies and moths. Effective incorporation of 2-methylpropanenitril and 2-methylbutanenitril into linamarin and notaustralin by Zygaena and Heliconius species (Lepidoptera). Insect Biochem. 17: 689–693 (1987)
52. Dean J. et al.: Defensive spray of the bombardier beetle: A biological pulse jet. Science 248: 1219–21 (1990)

53. Demmer L. A. et al.: The cellular retinol binding protein II gene. J. biol. Chem. 262: 2458–67 (1987)
54. van Dop C., Medynski D. C. and Apone L. M.: Nucleotide sequence for a cDAN encoding the α-subunit of retinal transducin (GNAT1) isolated from the human eye. Nucleic Acids Res. 17: 4887 (1989)
55. Dunlap W. C. et al.: Biochemical photoadatation in vision: u.v.-absorbing pigments in fish eye tissues. Comp. Biochem. Physiol. Pt. B 93: 601–607 (1989)
56. Dussourd D. E. et al.: Biparental defensive endowment of eggs with acquired plant alkaloid in the moth Utetheisa ornatrix. Proc. Nat. Acad. Sci. USA 85: 5992–96 (1988)
57. Ekpa O. et al.: Ketones and alcohols in the defensive secretion of Leiobunum townsendi Weed and a review of the known exocrine secretions of Palpatores (Arachnida: Opiliones). Comp. Biochem. Physiol. Pt. B 81: 555–557 (1985)
58. Evans C. O. et al.: Cloning, sequencing and expression of cDNA for chick liver haem oxygenase: comparison of avian and mammalian cDNAs and deduced proteins. Biochem. J. 273: 659–666 (1991)
59. Everaerts C. et al.: (+)-alpha-pinene in the defensive secretion of Nasutitermes princeps (Isoptera, Termitidae). Experientia 46: 227–230 (1990)
60. Fang L. S.: Study on the heme catabolism of fish. Comp. Biochem. Physiol. Pt. B 88: 667–673 (1987)
61. Fang L. S. and Bada J. L.: The blue-green blood plasma of marine fish (Review). Comp. Biochem. Physiol. Pt. B 97: 37–45 (1990)
62. Fenical W.: Natural products chemistry in the marine environment. Science 215: 923–928 (1982)
63. Ferré J. et al.: Pigment patterns in mutants affecting the biosynthesis of pteridines and xanthommatin in Drosophila melanogaster. Biochem. Genetics 24: 545–569 (1986)
64. Fong S. L. et al.: IRBP-like proteins in the eyes of six cephalopod species: Immunochemical relationship to vertebrate interstitial retinol-binding protein (IRBP) and cephalopod retinal-binding protein. Vision Res. 28: 563–573 (1988)
65. Foster K. W. et al.: A rhodopsin is the functional photoreceptor for phototaxis in the unicellular eukaryote Chlamydomonas. Nature 311: 756–758 (1984)
66. Foster S. P.: Biosynthesis of a monene and a conjugated diene sex pheromone component of the light-brown apple moth by delta-11 desaturation. Experientia 46: 269–273 (1990)
67. Francke W., Borchert J. and Klimetzek D.: Volatile constituents in the red wood ant Formica rufa L. (Hymenoptera: Formicidae). Z. Naturforsch. Sect. C 40: 661–664 (1985)
68. Fukada Y. et al.: Beta/gamma subunit of bovine transducin composed of two components with distinctive gamma-subunits. J. Biol. Chem. 264: 5937–43 (1989)
69. Gillespie P. G. et al.: A soluble form of bovine rod photoreceptor phosphodiesterase has a novel 15-kDa subunit. J. Biol. Chem. 264: 12187–93 (1989)
70. Gomez R. et al.: A purple carotenoprotein from the carapace of Galathea strigosa. Comp. Biochem. Physiol. Pt. B 90: 53–57 (1988)
71. Goodwin T. W.: The biochemistry of the carotenoids, 2. ed., 2 vol. set. Chapman & Hall, London 1980–84
72. Grant P. T. et al.: The isolation of four aminocyclohexenimines (mycosporines) and a structurally related derivative of cyclohehane-1,3-dione (gadusol) from the brine shrimp, Artemia. Comp. Biochem. Physiol. Pt. B 80: 755–759 (1985)
73. Grishin E. V., Volkova T. M. and Arseniev A. S.: Isolation and structure analysis of components from venom of the spider Argiope lobata. Toxicon 27: 541–549 (1989)
74. Habermann E.: Palytoxin acts through Na,K-ATPase (Review article) Toxicon 27: 1171–87 (1989)
75. Habermehl G.: Toxic animals and their weapons (In German). 4th ed. Springer, Berlin 1987
76. Hall M. D. et al.: Molecular cloning and primary structure of squid (Loligo forbesi) rhodopsin, a phospholipase C-directed G-protein-linked receptor. Biochem. J. 274: 35–40 (1991)
77. Hamilton J. G. C., Sonenshine D. E. and Lusby W. R.: Cholesteryl oleate: mounting sex pheromone of the hard tick Dermacentor variabilis (Say) (Acari: Ixodidae). J. Insect Physiol. 35: 873–879 (1989)
78. Hamm H. E. and Bownds M. D.: Protein complementation of rod segments of frog retina. Biochemistry 25: 4512–23 (1986)
79. Hastings J. W.: Biological diversity, chemical mechanisms, and the evolutionary origins of bioluminescent systems. J. mol. Evol. 19: 309–321 (1983)
80. Higa T., Ichiba T. and Okuda R. K.: Marine indoles of novel substitution pattern from the acorn worm Glossobalanus sp. Experientia 41: 1487–88 (1985)
81. Hiraki K. et al.: Conformation changes of cuttlefish (Euprymna morsei) rhodopsin following photoconversion. Biochim. biophys. Acta 1077: 355–361 (1991)
82. Hisatomi O. et al.: Isolation and characterization of lamprey rhodopsin cDNA. Biochem. biophys. Res. Commun. 174: 1125–32 (1991)
83. Holden H. M. et al.: The molecular structure of insecticyanin from the tobacco hornworm Manduca sexta L. at 2.6 Å. Embo J. 6: 1565–70 (1987)
84. Honda K.: GC-MS and C13-NMR studies on the biosynthesis of terpenoid defensive secretions by the larvae of papilionid butterflies (Luehdorfia and Papilio). Insect Biochem. 20: 245–250 (1990)
85. Huber A. et al.: Opsin of Calliphora peripheral photoreceptors R1–6. Homology with Drosophila Rh1 and posttranslational processing. J. Biol. Chem. 265: 17906–10 (1990)
86. Hwang D. F., Cueh C. H. and Jeng S. S.: Occurrence of tetrodotoxin in the gastropod mollusk Natica lineata (lined moon shell). Toxicon 28: 21–27 (1990)
87. Ito S.: Reexamination of the structure of eumelanin. Biochim. biophys. Acta 883: 155–161 (1986)
88. Ito S. et al.: Incorporation of sulfhydryl compounds into melanins in vitro. Biochim. biophys. Acta 964: 1–7 (1988)
89. Jackson B. D. et al.: Trail pheromone of the ant Tetramorium merdinale. Naturwissenschaften 77: 294–296 (1990)
90. Johne S.: The quinazoline alkaloids. Progr. Chem. organic natural Products Vol. 46, pp. 159–229. Springer, Wien 1984
91. Jones G. et al.: Purification and characterization of a very high density chromolipoprotein from the hemolymph of Trichoplusia ni (Huebner). Arch. Insect Biochem. Physiol. 7: 1–11 (1988)
92. Jurenka R. A., Howard R. W. and Blomquist G. J.: Prostaglandin synthetase inhibitors in insect defensive secretions. Naturwissenschaften 73: 735–737 (1986)
93. Jurenka R. A. and Roelofs W. L.: Characterization of the acetyltransferase used in pheromone biosynthesis

in moths: specificity for the Z isomer in Tortricidae. Insect Biochem. 19: 639–644 (1989)
94. Karnaukhov V. N.: Carotenoids: Recent progress, problems and prospects (Review). Comp. Biochem. Physiol. Pt. B 95: 1–20 (1990)
95. Kasang L. G., Nicholls M. and Vonproff L.: Sex pheromone conversion and degradation in antennae of the silkworm moth Bombyx mori. Experientia 45: 81–87 (1989)
96. Kayser H.. Pigments. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 10, pp. 367–415. Pergamon Press, Oxford 1985
97. Keen J. N. et al.: Complete sequence and model for the A2 subunit of the carotenoid pigment complex, crustacyanin. Eur. J. Biochem. 197: 407–417 (1991)
98. Kim I. H. et al.: Structure and function of the photoreceptor stentorins in Stentor coeruleus (I). Biochim. biophys. Acta 1040: 43–57 (1990)
99. King G. M.: Inhibition of microbial activity in marine sediments by a bromophenol from a hemichordate. Nature 323: 257–259 (1986)
100. Kirschfeld K.: The role of photostabile pigments in eyes and photoreceptor cells (In German). Verh. Dtsch. Zool. Ges. 78: 101–117 (1985)
101. Koenig B. et al.: Three cytoplasmic loops of rhodopsin interact with transducin. Proc. Nat. Acad. Sci. USA 86: 6878–82 (1989)
102. Krissansen G. W., Trotman C. N. A. and Tate W. P.: Identification of the blue-green chromophore of an abundant biliprotein from the haemolymph of Artemia. Comp. Biochem. Physiol. Pt. B 77: 249–252 (1984)
103. Krust A. et al.: A third human retinoic acid receptor, hRAR-gamma. Proc. Nat. Acad. Sci. USA 86: 5310–14 (1989)
104. Kumagai M. et al.: Okadaic acid as the causative toxin of diarrhetic shellfish poisoning in Europe. Agric. biol. Chem. 50: 2853–58 (1986)
105. Kumar S. et al.: Amino acid sequence of the Ca-triggered luciferin binding protein of Renilla reniformis. FEBS Letters 268: 287–290 (1990)
106. Kunesch G. et al.: A fungal metabolite as the male wing gland pheromone of the bumble-bee wax moth, Aphomia sociella L. Z. Naturforsch. Sect. C 42: 657–659 (1987)
107. Kuwata O. et al.: The primary structure of iodopsin, a chicken red-sensitive cone pigment. FEBS Letters 272: 128–132 (1990)
108. Lambrecht H. G. and Koch K. W.: A 26 kd calcium binding protein from the bovine rod outer segments as modulator of photoreceptor guanylate cyclase. Embo J. 10: 793–798 (1991)
109. Lanne B. S. et al.: Biosynthesis of 2-methyl-3-buten-2-ol, a pheromone component of Ips typographus (Coleoptera: Scolytidae). Insect Biochem. 19: 163–167 (1989)
110. Leal W. S. et al.: β-Acaridial, the sex pheromone of the acarid mite Caloglyphus polyphyllae. Naturwissenschaften 76: 332–333 (1989)
111. Leal W. S., Kuwahara Y. and Suzuki T.: Robinal, a highly conjugated monoterpenoid from the mite Rhizoglyphus robini. Naturwissenschaften 77: 387–388 (1990)
112. LeConte Y. et al.: Identification of a brood pheromone in honeybees. Naturwissenschaften 77: 334–336 (1990)
113. Lemaire M. et al.: Alkaloid venom of European ants in the genus Monomorium. Site of synthesis, identification and quantification. Exp. Biol. 48: 27–40 (1988)
114. Liou G. I. et al.: Human interstitial retinoid-binding protein. Gene structure and primary sequence. J. Biol. Chem. 264: 8200–06 (1989)
115. Lipkin V. M. et al.: β-Subunit of bovine photoreceptor cGMP phosphodiesterase. Comparison with the phosphodiesterase family. J. Biol. Chem. 265: 12955–59 (1990)
116. Llewellyn L. E. and Endean R.: Toxins extracted from Australian specimens of the crab, Eriphia sebana (Xanthidae). Toxicon 27: 579–586 (1989)
117. Lorenz W. W. et al.: Isolation and expression of a cDNA encoding Renilla reniformis luciferase. Proc. Nat. Acad. Sci. USA 88: 4438–42 (1991)
118. Mann J.: Secondary metabolism, 2nd ed. Clarendon Press, Oxford 1987
119. Maoka T. and Matsuno T.: Metabolism of carotenoids in terrestrial snail Euhadra callizona amaliae. Comp. Biochem. Physiol. Pt. B 92: 41–43 (1989)
120. Maoka T., Yokoi S. and Matsuno T.: Comparative biochemical studies of carotenoids in nine species of Cephalopoda. Comp. Biochem. Physiol. Pt. B 92: 247–250 (1989)
121. Maoka T., Tsushima M. and Matsuno T.: New acetylenic carotenoids from the starfish Asterina pectinifera and Asterias amurensis. Comp. Biochem. Physiol. Pt. B 93: 829–834 (1989)
122. Martin R. L. et al.: Visual pigment homologies revealed by DNA hybridisation. Science 232: 1266–69 (1986)
123. Martinez T., Fabrias G. and Camps F.: Sex pheromone biosynthetic pathway in Spodoptera littoralis and its activation by a neurohormone. J. Biol. Chem. 265: 1381–87 (1990)
124. Masuda T., Tatsumi H. and Nakano E.: Cloning and sequence analysis of cDNA for luciferase of a Japanese firefly, Luciola crusiata. Gene 77: 265–270 (1989)
125. Matsui S. et al.: 4-Hydroxyretinal, a new visual pigment chromophore found in the bioluminescent squid, Watasenia scintillans. Biochim. biophys. Acta 966: 370–374 (1988)
126. Matsuno T. et al.: Comparative biochemical studies of carotenoids in fishes – XXIX. Isolation of new luteins, lutein F and lutein G from marine fishes. Comp. Biochem. Physiol. Pt. B 85: 77–80 (1986)
127. Matsuno T. et al.: Isolation of three new carotenoids and proposed metabolic pathways of carotenoids in hen's egg yolk. Comp. Biochem. Physiol. Pt. B 84: 477–481 (1986)
128. Matsuno T. and Tsushima M.: Carotenoids of shellfishes. X. Reductive metabolic pathway of echinenone and fritschiellaxanthin in the spindle shell Fusinus perplexus. Comp. Biochem. Physiol. Pt. B 92: 189–193 (1989)
129. Matsuno T., Maoka T. and Toriiminami Y.: Carotenoids in the Japanese stick insect Neophirasea japonica. Comp. Biochem. Physiol. Pt. B 95: 583–587 (1990)
130. McCapra F.: The chemistry of bioluminescence (Review lecture). Proc. Royal Soc. London 215: 247–272 (1982)
131. McElroy W. D. and DeLuca M.: Biochemistry of insect bioluminescence. In: Kerkut G. A. and Gilbert

L.I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 4, pp. 553–563. Pergamon Press, Oxford 1985
132. Mebs D. and Schmidt K.: Occurrence of tetrodotoxin in the frog Atelopus oxyrhynchus. Toxicon 27: 819–822 (1989)
133. Milicua J. C. G. et al.: A red carotenoprotein from the carapace of the crayfish, Procambarus clarkii. Comp. Biochem. Physiol. Pt. B 81: 1023–25 (1985)
134. Milicua J. C. G. et al.: A yellow carotenoprotein from the carapace of the crayfish Astacus leptodactylus. Comp. Biochem. Physiol. Pt. B 85: 615–619 (1986)
135. Moore R. E.: Structure of palytoxin. Progr. Chem. organic natural Products Vol. 48, pp. 81–202. Springer, Wien 1985
136. Müller W. E. G. et al.: Potent antileucemic activity of the novel cytostatic agent avarone and its analogues in vitro and in vivo. Cancer Res. 45: 4822–27 (1985)
137. Murata M. et al.: Charaterization of compounds that induce symbiosis between sea anemone and fish. Science 234: 585–587 (1986)
138. Nardi G., Palumbo A. and Prota G.: The role of the white bodies in the biosynthesis of adenochrome. Comp. Biochem. Physiol. Pt. B 71: 297–300 (1982)
139. Naylor S. et al.: Chemical and biological aspects of marine monoterpenes. Progr. Chem. organic natural Products Vol. 44, pp. 189–241. Springer, Wien 1983
140. Nichol C. A., Smith G. K. and Duch D. S.: Biosynthesis and metabolism of tetrahydrobiopterin and molybdopterin. Annual Rev. Biochem. 54: 729–764 (1985)
141. Novotny M., Harvey S. and Jemiolo B.: Chemistry of male dominance in the house mouse, Mus domesticus. Experientia 46: 109–113 (1990)
142. Ohguro H. et al.: Functional heterogeneity of beta/gamma-subunit of frog transducin. Comp. Biochem. Physiol. Pt. B 95: 763–765 (1990)
143. Okano T. et al.: Purification of cone visual pigments from chicken retina. Biochemistry 28: 8848–56 (1989)
144. Partali V., Tangen K. and Liaaen-Jensen S.: Carotenoids in food chain studies. III. Resorption and metabolic transformation of carotenoids in Mytilus edulis (edible mussel). Comp. Biochem. Physiol. Pt. B 92: 239–246 (1989)
145. Pollock J. A. and Benzer S.: Transcript localization of four opsin genes in three visual organs of Drosophila: Rh2 is ocellus specific. Nature 333: 779–783 (1988)
146. Prasher D. C. et al.: Sequence comparisons of complementary DNAs encoding aequorin isotypes. Biochemistry 26: 1326–32 (1987)
147. Prestwich G. D. et al.: Enzymatic processing of pheromones and pheromone analogs (Review). Experientia 45: 263–270 (1989)
148. Prince R. C.: Tetrodotoxin. Trends biochem. Sci. 13: 76–77 (1988)
149. Quiñoa E., Castedo L. and Riguera R.: The halogenated monoterpenes of Aplysia punctata: A comparative study. Comp. Biochem. Physiol. Pt. B 92: 99–101 (1989)
150. Raming K., Krieger J. and Breer H.: Primary structure of a pheromone-binding protein from Antheraea pernyi: Homologies with other ligand-carrying proteins. J. comp. Physiol. B 160: 503–509 (1990)
151. Raport C. J., Dere B. and Hurley J. B.: Characterization of the mouse transducin α-subunit gene. J. Biol. Chem. 264: 7122–28 (1989)
152. Rees J. F. et al.: Detection of coelenterazine and related luciferase activity in the tissues of the luminous fish, Vinciguerria attenuata. Comp. Biochem. Physiol. Pt. A 96: 425–430 (1990)
153. Rivas J. D. L., Milicua J. C. G. and Gomez R.: Further studies on the blue carotenoprotein from Astacus leptodactylus. Comp. Biochem. Physiol. Pt. B 89: 65–68 (1988)
154. Rudie N. G., Mulkerrin M. G. and Wampler J. E.: Earthworm bioluminescence: Characterization of high specific activity Diplocardia longa luciferase and the reaction it catalyzes. Biochemistry 20: 344–350 (1981)
155. Russell F. E.: Marine toxins and venomous and poisonous marine plants and animals (invertebrates). Adv. marine Biol. 21: 59–217 (1984)
156. Sakai T. et al.: Occurrence of bilirubin-IXß in the gallbladder bile of eel, Anguilla japonica. Biochim. biophys. Acta 993: 128–1330 (1989)
157. Salzmann T. A., del C. Batlle A. M. and de Souza W.: Hemesynthesis in Trypanosoma cruzi: influence of the strain and culture medium. Comp. Biochem. Physiol. Pt. B 83: 57–61 (1986)
158. Santoro P and Parisi G.. Biosynthesis of dihydroxanthommatin. In vitro enzymatic reduction of xanthommatin. Insect Biochem. 17: 635–638 (1987)
159. Sato M. et al.: The occurrence of N-methyl-D-aspartic acid in muscle extracts of blood shell, Scapharca broughtonii. Biochem. J. 241: 309–311 (1987)
160. Saurat J. H. (ed.): Retinoids: 10 years on. Karger, Basel 1991
161. Scheuer P. J. (ed.): Bioorganic marine chemistry, Vol. 1–3. Springer, Berlin 1987
162. Schneuwly S. et al.: Drosophila ninaA gene encodes an eye-specific xyclophilin (cylosporine A binding protein). Proc. Nat. Acad. Sci. USA 86: 5390–94 (1989)
163. Schulz S., Francke W. and Boppré M.: Carboxylic acids from hairpencils of male Amauris butterlies (Lep.:Danainae). Biol. Chem. Hoppe-Seyler 369: 633–638 (1988)
164. Schulz S. et al.: Volatile compounds from androconial organs of danaine and ithomiine butterfles. Z. Naturforsch. Sect. C 43: 99–104 (1988)
165. Sebben A. et al.: A tetrodotoxin-like substance found in the Brazilian frog Brachycephalus ephippium. Toxicon 24: 799–806 (1986)
166. Seki T., Fujishita S. and Obana S.: Composition and distribution of retinal and 3-hydroxyretinal in the compound eye of the dragonfly. Exp. Biol. 48: 65–75 (1989)
167. Service M. and Wardlaw A. C.: Echinochrome-A as a bactericidal substance in coelomic fluid of Echinus esculentus (L.). Comp. Biochem. Physiol. Pt. B 79: 161–165 (1984)
168. Shieh B. H. et al.: The ninaA gene required for visual transduction in Drosophila encodes a homologue of cyclosporin A-binding protein. Nature 338: 67–70 (1989)
169. Shimomura O. and Shimomura A.: Halistaurin, phialidin and modified forms of aequorin as calciumion indicator in biological systems. Biochem. J. 228: 745–749 (1985)
170. Shimomura O.: Presence of coelenterazine in non-bioluminescent marine organisms. Comp. Biochem. Physiol. Pt. B 86: 361–363 (1987)
171. Smith D. P., Shieh B. H. and Zuker C. S.: Isolation and structure of an arrestin gene from Drosophila. Proc. Nat. Acad. Sci. USA 87: 1003–07 (1990)

172. Smith W. C. and Goldsmith T. H.: Phyletic aspects of the distribution of 3-hydroxyretinal in the class insecta. J. mol. Evol. 30: 72–84 (1990)
173. Song P. S. et al.: Structure and function of the photoreceptor stentorins in Stentor coeruleus. II. Primary photoprocess and picosecond time-resolved fluorescence. Biochim. biophys. Acta 1040: 58–65 (1990)
174. Stieve H. and Lumme G.: Similarities between G-proteins in visual cells of sepia and cattle. FEBS Letters 253: 6–10 (1989)
175. Stowe M. K., Tumlinson J. H. and Heath R. R.: Chemical mimicry: Bolas spiders emit components of moth prey species sex pheromones. Science 236: 964–967 (1987)
176. Stryer L.: Visual excitation and recovery (Minireview). J. Biol. Chem. 266: 10711–14 (1991)
177. Sugiyama Y. et al.: Isolation of a gene that encodes a new retinal protein, archaerhodopsin, from Halobacterium sp.aus-1. J. Biol. Chem. 264: 20859–62 (1989)
178. Suter F., Kayser H. and Zuber H.: The complete amino-acid sequence of the bilin-binding protein from Pieris brassicae and its similarity to a family of serum transport proteins like the retinol-binding proteins. Biol. Chem. Hoppe-Seyler 369: 497–505 (1988)
179. Suzuki T. and Miyata S.: 3-Hydroxyretinol in the rhodopsin-porphyropsin retina of an invertebrate (Procambarus clarkii). Comp. Biochem. Physiol. Pt. B 99: 307–312 (1991)
180. Switchenko A. C. and Brown G. M.: The enzymatic conversion of dihydroneopterin triphosphate to tripolyphosphate and 6-pyruvoyl-tetrahydropterin, an intermediate in the biosynthesis of other pterins in Drosophila melanogaster. J. biol. Chem. 260: 2945–51 (1985)
181. Takao M., Yasui A. and Tokunaga F.: Isolation and sequence determination of the chicken rhodopsin gene. Vision Res. 28: 471–480 (1988)
182. Tamaki Y.: Sex pheromones. In: Kerkut G. A. and Gilbert L. I. (eds.): Comprehensive insect physiology, biochemistry and pharmacology, Vol. 9, pp. 145–191. Pergamon Press, Oxford 1985
183. Tasayco M. L. and Prestwich G. D.: Aldehyde oxidases and dehydrogenases in antennae of 5 moth species. Insect Biochem. 20: 691–700 (1990)
184. Teuscher E. und Lindequist D. U.: Animal toxins (In German). G. Fischer, Stuttgart 1987
185. Tsin A. T. C. et al.: Conversion of retinol to 3,4-didehydroretinol in the tadpole. Comp. Biochem. Physiol. Pt. B 81: 415–421 (1985)
186. Tsuji F. I.: ATP-dependent bioluminescence in the firefly squid, Watasenia scintillans. Proc. Nat. Acad. Sci. USA 82: 4629–32 (1985)
187. Tsushima M., Maoka T. and Matsuno T.: Comparative biochemical studies of carotenoids in marine invertebrates: The first positive identification of epsilon,epsilon-carotene derivatives and isolation of two new carotenoids from chitons. Comp. Biochem. Physiol. Pt. B 93: 665–671 (1989)
188. Tsushima M, and Matsuno T.: Comparative biochemical studies of carotenoids in sea urchins (I.). Comp. Biochem. Physiol. Pt. B 96: 801–810 (1990)
189. Tumlinson J. H. et al.: Identification of a pheromone blend attractive to Manduca sexta (L.) males in a wind tunnel. Arch. Insect Biochem. Physiol. 10: 255–271 (1989)
190. Turner E. et al.: Ovothiols, a family of redox-active mercapto-histidine compounds from marine invertebrate eggs. Biochemistry 26: 4028–36 (1987)
191. Tuteja N. and Farber D. B.: gamma-Subunit of the mouse retinal cyclic-GMP phosphodiesterase: cDNA and corresponding amino acid sequence. FEBS Letters 232: 182–186 (1988)
192. Umebachi Y.: Kynurenine, β-alanine, and dopamine in the deep yellow pigment of Papilio machaon (Lepidoptera: Papilionidae). Comp. Biochem. Physiol. Pt. B 94: 207–211 (1989)
193. Vogt K.: The chromophore of the visual pigment in some insect orders. Z. Naturforsch. Sect. C 39: 196–197 (1984)
194. Vollrath D., Nathans J. and Davis R. W.: Tandem array of human visual pigment genes at Xq28. Science 240: 1669–72 (1988)
195. Vollrath F. et al.: Compounds in the droplets of the orb spider's viscid spiral. Nature 345: 526–528 (1990)
196. Washburn T. and O'Tousa J. E.: Molecular defects in Drosophila rhodopsin mutants. J. Biol. Chem. 264: 15464–66 (1989)
197. Weber J. F. et al.: Isolation of allantoin and adenosine from the marine sponge Tethya aurantia. Comp. Biochem. Physiol. Pt. B 70: 799–801 (1981)
198. Weisberg E. P. and O'Donnell J. M.: Purification and characterization of GTP cyclohydrolase I from Drosophila melanogaster. J. biol. Chem. 261: 1453–58 (1986)
199. Wiederrecht G. J. and Brown G. M.: Purification and properties of the enzymes from Drosophila melanogaster that catalyze the conversion of dihydroneopterin triphosphate to the pyrimidodiazepine precursor of the drosopterins. J. biol. Chem. 259: 14121–27 (1984)
200. Wink M., Schneider D. and Witte L.: Biosynthesis of pyrrolizidine alkaloid-derived pheromones in the arctiid moth Creatonotos transiens: Stereochemical conversion of heliotrine. Z. Naturforsch. Sect. C 43: 737–741 (1988)
201. Wittenberg C. and Triplett E. L.: A detergent-activated tyrosinase from Xenopus laevis. I. Purification and partial characterization. J. biol. Chem. 260: 12535–41 (1985)
202. Witthohn K. and Naumann C. M.: Active cyanogenesis – in zygaenids and other Lepidoptera. Z. Naturforsch. Sect. C 42: 1319–22 (1987)
203. Wood S. F., Szuts E. Z. and Fein A.: Inositol trisphosphate production in squid photoreceptors: Activation by light, aluminium fluoride, and guanine nucleotides. J. Biol. Chem. 264: 12970–76 (1989)
204. Yago M.: Enzymic synthesis of papiliochrome II, a yellow pigment in the wings of papilionid butterflies. Insect Biochem. 19: 673–678 (1989)
205. Yamashita E. and Matsuno T.: A new apocarotenoid from the sea hare Aplysia kurodai. Comp. Biochem. Physiol. Pt. B 96: 465–470 (1990)
206. Yokoyama R. and Yokoyama S.: Covergent evolution of the red-like and green-like visual pigment genes in fish, Astyanax fasciatus, and human. Proc. Nat. Acad. Sci. USA 87: 9315–18 (1990)
207. Yokoyama S. and Yokoyama R.: Molecular evolution of human visual pigment genes. Mol. Biol. Evol. 6: 186–197 (1989)
208. Yotsu M., Iorizzi M. and Yasumoto T.: Distribution of tetrodotoxin, 6-epitetrodotoxin, and 11-deoxytetrodotoxin in newts. Toxicon 28: 238–241 (1990)

209. Zagalsky P. F. et al.: Studies on a blue carotenoprotein, linckicyanin, isolated from the starfish Linckia laevigata (Echinodermata: Asteroidea). Comp. Biochem. Physiol. Pt. B 93: 339–353 (1989)
210. Zagalsky P. F., Eliopoulos E. E. and Findlay J. B. C.: The architecture of invertebrate carotenoproteins (Review). Comp. Biochem. Physiol. Pt. B 97: 1–18 (1990)
211. Zeeck E. and Hardege J. D.: Sex pheromones in nereids: Volatile compounds of the coelomic fluid of Nereis virens. Z. Naturforsch. Sect. C 45: 924–925 (1990)
212. Zhao C., Loefstedt C. and Wang X.: Sex pheromone biosynthesis in the Asian corn borer Ostrinia furnacalis (II): Biosynthesis of (E)- and (Z)-12-tetradecenyl acetate involves 14-desaturation. Arch. Insect Biochem. Physiol. 15: 57–65 (1990)
213. Ziegler I.: Pterins: Pigments, cofactors, and signal connection in cellular interaction (In German). Naturwissenschaften 74: 563–572 (1987)
214. Special Issue on Vision: Photobiochem. Photobiophysics 13: 213–414 (1986)
215. Special Topic: Phototransduction in vertebrates. Annual Rev. Physiol. 49: 711–812 (1987)

Appendix

Animal Systematics

The natural system attempts to classify the approximately 1.5 million known animal species according to the degree of their phylogenetic relationships. The categories (taxa), phylum – class – order – family – genus, bring together species with increasingly closer relationships. In the following, the first category are taxa for which there is no completely secure evidence of mutual genealogical relationships. Subgroups of these higher taxa are placed in the subsequent categories, whereby some groups of animals which have only a few species, and which are not mentioned in this book, have been omitted.

I have deliberately resisted referring to the individual categories as phylum, class, order, etc. to avoid giving the impression that groups of animals in the same category are equivalent in the degree of their relationship or their phylogenetic age. The approximate number of species is given for the taxa of the first category.

Protozoa (unicellular animals)

Sarcomastigophora: 25 000 species
  Mastigophora = Flagellata
  Opalinata
  Sarcodina = Rhizopoda
Apicomplexa: 4800 species
Microspora: 800 species
Ciliophora = Ciliata: 7500 species

Metazoa (multicellular animals)

Mesozoa: 50 species
Placozoa: 2 species
Porifera (sponges): 5000 species

Eumetazoa (animals with tissues)

Cnidaria: 9000 species
  Anthozoa (sea anemones and corals)
  Scyphozoa (jellyfish)
  Hydrozoa (hydroids)
Ctenophora (comb jellyfish)
(Coelenterata = Cnidaria and Ctenophora together)
Nemathelminthes: 12 000 species
  Nematoda (round- or threadworms)
  Rotatoria (rotifers or wheel animalcules)
  Acanthocephala (spiny-headed worms)
Priapulida: 4 species
Pogonophora: 50 species
Sipunculida: 250 species
Echiuroida: 150 species
Kamptozoa = Entoprocta: 60 species
Platyhelminthes (flatworms): 12 000 species
  Turbellaria (eddyworms)
  Trematoda (flukes)
  Cestoda (tapeworms)
Nemertini (ribbon worms): 800 species
Mollusca: 130 000 species
  Polyplacophora (chitons)
  Gastropoda (snails and slugs)
    Prosobranchia
    Pulmonata
    Opisthobranchia
  Bivalvia = Lamellibranchia (mussels)
  Cephalopoda (octopuses, cuttlefish and squids)
Annelida (ringed or segmented worms): 9000 species
  Polychaeta (bristle worms)
  Clitellata
    Oligochaeta (earthworms)
    Hirudinea (leeches)
Arthropoda: 800 000 species
  Tardigrada (bear animalcules)
  Onychophora
  Chelicerata
    Xiphosura (horseshoe crabs)
    Arachnida
      Scorpiones (scorpions)
      Uropygi
      Amblypygi
      Araneae (spiders)
      Palpigradi
      Pseudoscorpiones (false scorpions)
      Solifugae

Opiliones (harvestmen)
Acari (mites)
Crustacea (crabs)
Chilopoda (centipedes)
Diplopoda (millipedes)
Insecta = Hexapoda (insects)
Collembola (springtails)
Pterygota (flying insects): subdivisions given below
(Articulata = Annelida and Arthropoda together)
Tentaculata: 5000 species
Phoronida
Bryozoa = Polyzoa
Brachiopoda (lamp shells)
Hemichordata: 80 species
Enteropneusta
Echinodermata: 6000 species
Crinoides (sea lilies, feather stars)
Asteroidea (starfish)
Ophiuroidea (brittle stars)
Echinoidea (sea urchins)
Holothurioidea (sea cucumbers)
Chaetognatha (arrow worms): 50 species
Chordata: 60 000 species
Acrania = Cephalochordata
Tunicata = Urochordata
Vertebrata
Agnatha (jawless fish: hagfish and lampreys)
Chondrichthyes (cartilaginous fish)
Elasmobranchii (sharks and rays)
Holocephali (chimaeras)
Osteichthyes (bony fish)
Actinopterygii
Chondrostei (sturgeons)
Polypteri
Holostei (garpikes and bowfins)
Teleostei ("modern" bony fish)
Dipnoi (lungfish)
Crossopterygii (Latimeria)
Amphibia
Urodela (newts and salamanders)
Anura (frogs and toads)
Reptilia
Chelonia (tortoises and turtles)
Squamata
Sauria (lizards)
Ophidia (snakes)
Crocodilia (crocodiles and alligators)
Aves (birds)
Mammalia (mammals)
Monotremata (egg-laying mammals)
Marsupialia = Metatheria = Didelphia
Placentalia = Eutheria

The most important insect orders

Ephemeroptera (mayflies)
Odonata (dragonflies)
Plecoptera (stoneflies)
Dermaptera (earwigs)
Mantodea (praying mantids)
Blattodea (cockroaches)
Isoptera (termites)
Phasmatodea (stick and leaf insects)
Saltatoria = Orthoptera (locusts and grasshoppers)
Psocoptera (book lice)
Phthiraptera (animal lice)
Thysanoptera (thrips)
Hemiptera = Rhynchota (bugs, aphids and cicadas)
Megaloptera (snake flies)
Planipennia = Neuroptera (alderflies and lacewings)
Coleoptera (beetles)
Strepsiptera (twisted-wing insects)
Hymenoptera (bees, wasps and ants)
Trichoptera (caddis flies)
Lepidoptera (butterflies and moths)
Siphonaptera (fleas)
Diptera (flies)

Biochemical Nomenclature

The nomenclature commissions set up by the IUB (International Union of Biochemistry) and IUPAC (International Union of Pure and Applied Chemistry) have strived for many years to bring uniformity to biochemical nomenclature and have published their recommendations in universally available biochemical journals. Despite these efforts, variation in the names given to individual substances and enzymes is still common. This makes it especially difficult for biologists to read biochemical research papers. I have attempted to present a uniform nomenclature in this book and to deviate as little as possible from the nomenclature of the primary literature.

Information on the positions of substituents in aliphatic carbonic acids is either given by the letters α, β, γ, etc. for the C atoms after the carboxyl C, or the C atoms are numbered consecutively with the carboxyl C as C-1. The positions of substituents on the glycerol molecule are described according to special rules (see Fig. 15.2, p. 568). Oxo-substituted carbonic acids are given as either keto acids or oxo acids, e.g. α-ketoglutarate is identical with 2-oxoglutarate.

The enantiomers of sugars, organic acids and amino acids with asymmetric C atoms are described with the prefix D- or L-. For natural products as in Chapter 19, the arrangement of ligands at the chiral centre is usually given by the "sequence rules" using the symbols (S) and (R). A chiral centre with four tetrahedrically arranged ligands (e.g. an asymmetric C atom) is considered from the side opposite the ligands with the lowest preference. Where the line joining the other three ligands, in order of declining preference according to the sequence rule, runs clockwise, the symbol (R) is used, and where it runs anti-clockwise, (S) is used. The sequence rule gives, for example, the series –I > –Br > –Cl > –OH > $–NH_2$ > –COOH > isopropyl >n-propyl > $–CH_3$ > –H.

2. Position \ 1. Position	A	C	G	U
A	AAA Lys AAG	CAA Gln CAG	GAA Glu GAG	UAA Stop UAG
	AAC Asn AAU	CAC His CAU	GAC Asp GAU	UAC Tyr UAU
C	ACA ACG Thr	CCA CCG Pro	GCA GCG Ala	UCA UCG Ser
	ACC ACU	CCC CCU	GCC GCU	UCC UCU
G	AGA Arg AGG	CGA CGG Arg	GGA GGG Gly	UGA Stop UGG Trp
	AGC Ser AGU	CGC CGU	GGC GGU	UGC Cys UGU
U	AUA Ile AUG Met	CUA CUG Leu	GUA GUG Val	UUA Leu UUG
	AUC Ile AUU	CUC CUU	GUC UUU	UUC Phe UUU

For the steric configuration at double bonds in products of secondary metabolism, most authors use (Z) (from German "zusammen" = together) and (E) (for German "entgegengesetzt" = opposite) instead of cis and trans, respectively.

T Thymidine
U Uridine
ψ Pseudouridine
R Purine nucleoside
Y Pyrimidine nucleoside
N Not specified

Amino acid residues in peptide and protein sequences are given by either the single-letter or the three-letter codes, according to the space available:

A	Ala	Alanine
R	Arg	Arginine
D	Asp	Aspartic acid
N	Asn	Asparagine
C	Cys	Cysteine
E	Glu	Glutamic acid
Q	Gln	Glutamine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
L	Leu	Leucine
K	Lys	Lysine
M	Met	Methionine
F	Phe	Phenylalanine
P	Pro	Proline
S	Ser	Serine
T	Thr	Threonine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
V	Val	Valine
B	Asx	Aspartic acid or asparagine
Z	Glx	Glutamic acid or glutamine
X		Not identified

The following abbreviations are used for sugar residues in compound carbohydrates and glycolipids:

Non-substituted sugars:

Ara	Arabinose
Fru	Fructose
Fuc	Fucose
Gal	Galactose
Glc	Glucose
Man	Mannose
Xyl	Xylose

Sugar derivatives (e.g. of glucose):

GlcN	Glucosamine
GlcNac	N-Acetylglucosamine
GlcUA	Glucuronic acid

Sialic acids:

Neu	Neuraminic acid
NeuAc	N-Acetylneuraminic acid
NeuGc	N-Glycolylneuraminic acid

Glycosidic bonds are described by α or β, for the steric configuration, followed by, in parentheses, the numbers of the bound C atoms. For example, Glcα(1–4)Glc is maltose (4-O-α-D-glucopyranosyl-glucopyranose). In Chapter 14, these symbols are used also for sugar phosphates, e.g. Fru-1,6-P_2 for frutose-1,6-bisphosphate.

Nomenclature problems for other groups of substances are discussed in the relevant sections,

e.g. fatty acids and lipids in Chapter 15, steroids in Chapter 16, and carotenoids in Chapter 19.

Mass units are usually given according to the SI system. Molecular masses are given with the relative atomic mass unit, Da (dalton), or derived units, kDa (kilodalton) or MDa (megadalton). The SI unit of enzyme activity (1 kat = 1 mol/s) is inappropriately large; here I have used the unit, 1 U = 1 μmol/min = 16.67 nkat and the 1000-fold smaller unit mU.

Abbreviations

bp	Base pair
D	Genetic distance (after Nei) (p. 152)
Da	Dalton (unit of relative atomic mass)
ΔT_m	Difference in dissociation temperature between homologous and heterologous DNA double helices
(E)	trans-Configuration at double bonds [the opposite of (Z)]
H	Mean heterozygosity (p. 131)
kb	Kilobases (in DNA or RNA)
kDa	Kilodalton (1000 daltons)
kDNA	Kinetoplast DNA
LD_{50}	Lethal Dose (50 %)
MDa	Megadalton (1 million daltons)
mtDNA	Mitochondrial DNA
mU	Milliunit of enzyme activity (nmol/min)
N_e	Effective population size
5'-NT	5' Non-translated
3'-NT	3' Non-translated
nt	Nucleotide
P	Proportion of polymorphic loci (p. 131)
(R)	Steric configuration at a chiral centre [the opposite of (S)]
rDNA	The gene of ribosomal RNA
rRNA	Ribosomal RNA
(S)	Steric configuration at a chiral centre [the opposite of (R)]
SDS	Sodium dodecyl sulphate
sp.	Non-identified species
T_m	Dissociation (melting) temperature of the DNA double helix
U	Unit of enzyme activity (μmol/min)
v_0	Neutral mutation rate (p. 141)
(Z)	cis-Configuration at double bonds [the opposite of (E)]

Genera Index

D

E

M

R

S

T

U

V

W

X

Subject Index

A. Matsumoto, S. Ishii (Eds.)

Atlas of Endocrine Organs

Vertebrates and Invertebrates

1992. XII, 307 pp. 550 figs., mostly in color.
Including three-dimensional ills.,
photomicrographs, and schematic drawings.
ISBN 3-540-53158-0

This atlas gives a comprehensive insight into the morphology of the endocrine system and its adaptive and evolutionary aspects. With its beautifully prepared, multicolor, three-dimensional drawings and photomicrographs, the macro- and micro-morphology of the endocrine systems in vertebrates as well as invertebrates is explained.

Springer

Tm.BA94..2.1

Klaus Urich

Comparative Animal Biochemistry

1. Position									
		A		C		G			
2. Position	A	AAA	Lys	CAA	Gln	GAA	Glu	UAA	
		AAG		CAG		GAG		UAG	
		AAC	Asn	CAC	His	GAC	Asp	UAC	
		AAU		CAU		GAU		UAU	
	C	ACA	Thr	CCA	Pro	GCA	Ala	UCA	
		ACG		CCG		GCG		UCG	
		ACC		CCC		GCC		UCC	
		ACU		CCU		GCU		UCU	
	G	AGA	Arg	CGA	Arg	GGA	Gly	UGA	
		AGG		CGG		GGG		UGG	
		AGC	Ser	CGC		GGC		UGC	
		AGU		CGU		GGU		UGU	
	U	AUA	Ile	CUA	Leu	GUA	Val	UUA	
		AUG	Met	CUG		GUG		UUG	
		AUC	Ile	CUC		GUC		UUC	
		AUU		CUU		GUU		UUU	

Components of DNA or RNA

A	Adenosine	U	Uridine	Y	Pyrimidine nucleoside
G	Guanosine	ψ	Pseudouridine		
C	Cytidine	R	Purine nucleoside	N	Not specifie
T	Thymidine				

Amino Acid Residues in Peptides
(Single- and three-letter codes)

A	Ala	Alanine	L	Leu	Leucine
R	Arg	Arginine	K	Lys	Lysine
D	Asp	Aspartic acid	M	Met	Methionine
N	Asn	Asparagine	F	Phe	Phenylalanin
C	Cys	Cysteine	P	Pro	Proline
E	Glu	Glutamic acid	S	Ser	Serine
Q	Gln	Glutamine	T	Thr	Threonine
G	Gly	Glycine	W	Trp	Tryptophan
H	His	Histidine	Y	Tyr	Tyrosine
I	Ile	Isoleucine	V	Val	Valine

B	Asx	Aspartic acid or asparagine
Z	Glx	Glutamic acid or glutamine
X		Not identified

Sugar Residues in Compound Carbohydrates

Non-substituted Sugars

Ara	Arabinose	Glc	Glucose
Fru	Fructose	Man	Mannose
Fuc	Fucose	Xyl	Xylose
Gal	Galactose		

Sugar Derivatives (e.g. of glucose):

GlCn	Glucosamine
GlcNac	N-Acetylglucosamine
GlcUA	Glucuronic acid

Sialic Acids:

Neu	Neuraminic acid
NeuAc	N-Acetylneuraminic acid
NeuGc	N-Glycolylneuraminic acid

Klaus Urich

Comparative Animal Biochemistry

Abbreviations in the text

bp	Base pair
D	Genetic distance (after Nei) (p. 152)
Da	Dalton (unit of relative atomic mass)
ΔT_m	Difference in dissociation temperature between homologous and heterologous DNA double helices
(E)	trans-Configuration at double bonds [the opposite of (Z)]
H	Mean heterozygosity (p. 131)
kb	Kilobases (in DNA or RNA)
kDa	Kilodalton (1000 daltons)
kDNA	Kinetoplast DNA
LD_{50}	Lethal dose (50 %)
MDa	Megadalton (1 million daltons)
mtDNA	Mitochondrial DNA
mU	Milliunit of enzyme activity (nmol/min)
N_e	Effective population size
5'-NT	5' Non-translated
3'-NT	3' Non-translated
nt	Nucleotide
P	Proportion of polymorphic loci (p. 131)
(R)	Steric configuration at a chiral centre [the opposite of (S)]
rDNA	The gene of ribosomal RNA
rRNA	Ribosomal RNA
(S)	Steric configuration at a chiral centre [the opposite of (R)]
SDS	Sodium dodecyl sulphate
sp.	Non-identified species
T_m	Dissociation (melting) temperature of the DNA double helix
U	Unit of enzyme activity (μmol/min)
v_0	Neutral mutation rate (p. 141)
(Z)	cis-Configuration at double bonds [the opposite of (E)]