

The background of the cover is a deep navy blue. It features several translucent jellyfish with long, thin tentacles, appearing to float. Interspersed among the jellyfish are several green, ribbon-like structures that represent the 3D folds of proteins, specifically highlighting the structure of green fluorescent protein (GFP).

CAMPBELL | FARRELL

# BIOCHEMISTRY

**INSIDE**

**BIOCHEMISTRY**

Hot Topics Magazine!

**GREEN** Fluorescent Protein,  
Jellyfish and Green Monkeys.

SEVENTH  
EDITION

## LIST OF ABBREVIATIONS

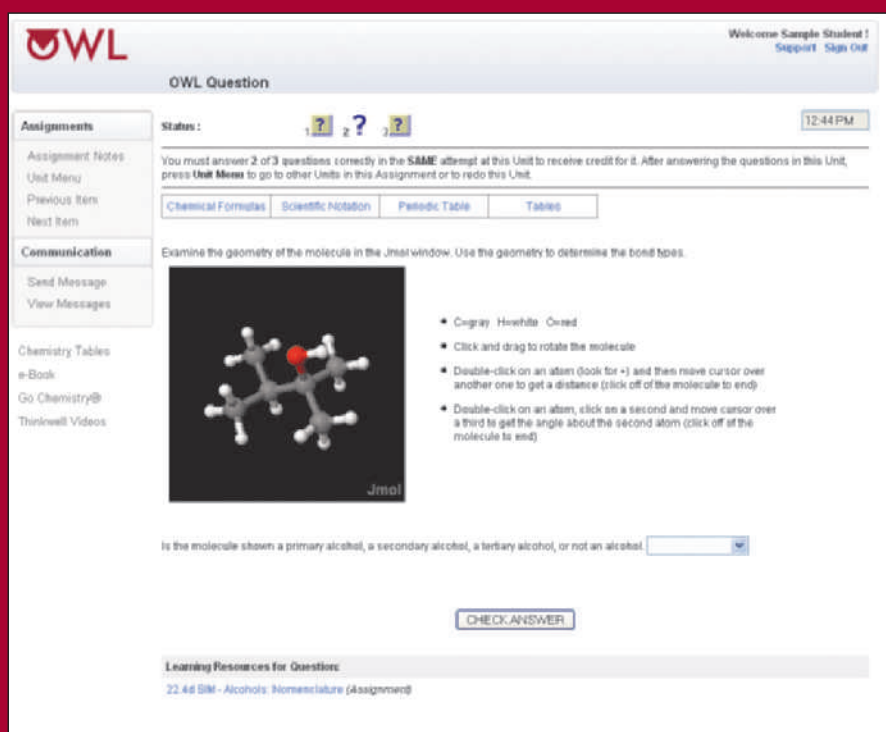
A	Adenine
ACAT	Acyl-CoA cholesterol acyl transferase
ACP	Acyl carrier protein
ADP	Adenosine diphosphate
AIDS	Acquired immunodeficiency syndrome
AMP	Adenosine monophosphate
ATCase	Aspartate transcarbamoylase
ATP	Adenosine triphosphate
bp	Base pairs
C	Cytosine
cAMP	Cyclic adenosine monophosphate
CAP	Catabolite activator protein
CDP	Cytidine diphosphate
Chl	Chlorophyll
CMP	Cytidine monophosphate
CoA (CoA-SH)	Coenzyme A
CoQ	Coenzyme Q
CTP	Cytidine triphosphate
d	Deoxy
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DV	Daily value
EF	Elongation factor
ER	Endoplasmic reticulum
FAD	Flavin adenine dinucleotide (oxidized form)
FADH <sub>2</sub>	Flavin adenine dinucleotide (reduced form)
fMet	N-Formylmethionine
FMN	Flavin mononucleotide
G	Guanine
GDP	Guanosine diphosphate
GMP	Guanosine monophosphate
GSH	Glutathione (reduced form)
GSSG	Glutathione (oxidized form)
GTP	Guanosine triphosphate
Hb	Hemoglobin
HDL	High-density lipoprotein
HIV	Human immunodeficiency virus
HMG-CoA	$\beta$ -Hydroxy- $\beta$ -methylglutaryl-CoA
HPLC	High-performance liquid chromatography



IF	Initiation factor
$K_M$	Michaelis constant
LDL	Low-density lipoprotein
Mb	Myoglobin
NAD <sup>+</sup>	Nicotinamide adenine dinucleotide (oxidized form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
NADP <sup>+</sup>	Nicotinamide adenine dinucleotide phosphate (oxidized form)
NADPH	Nicotinamide adenine dinucleotide phosphate (reduced form)
$P_i$	Phosphate ion
PAGE	Polyacrylamide gel electrophoresis
PCR	Polymerase chain reaction
PEP	Phosphoenolpyruvate
PIP <sub>2</sub>	Phosphatidylinositol <i>bis</i> phosphate
PKU	Phenylketonuria
Pol	DNA polymerase
PP <sub>i</sub>	Pyrophosphate ion
PRPP	Phosphoribosylpyrophosphate
PS	Photosystem
RF	Release factor
RFLPs	Restriction-fragment-length polymorphisms
RNA	Ribonucleic acid
RNase	Ribonuclease
mRNA	Messenger RNA
rRNA	Ribosomal RNA
tRNA	Transfer RNA
snRNP	Small nuclear ribonuclear protein
S	Svedberg unit
SCID	Severe combined immune deficiency
SSB	Single-strand binding protein
SV40	Simian virus 40
T	Thymine
TDP	Thymidine diphosphate
TMP	Thymidine monophosphate
TTP	Thymidine triphosphate
U	Uracil
UDP	Uridine diphosphate
UMP	Uridine monophosphate
UTP	Uridine triphosphate
$V_{\max}$	Maximal velocity

# Get a Better Grade in Chemistry!

Log in now to the #1 online homework and tutorial system for chemistry.  
Score better on exams, get homework help, and more!



**OWL Question**

Welcome Sample Student!  
Support Sign Out

Status: 1 ? 2 ? 3 ? 12:44 PM

You must answer 2 of 3 questions correctly in the **SAME** attempt at this Unit to receive credit for it. After answering the questions in this Unit, press **Unit Menu** to go to other Units in this Assignment or to redo this Unit.

Chemical Formulas Scientific Notation Periodic Table Tables

Examine the geometry of the molecule in the Jmol window. Use the geometry to determine the bond types.

Is the molecule shown a primary alcohol, a secondary alcohol, a tertiary alcohol, or not an alcohol.

CHECK ANSWER

Learning Resources for Questions  
22.46 SSM - Alcohols: Nomenclature (Assignment)

- **Master chemistry and improve your grade** using OWL's step-by-step tutorials, interactive simulations, and homework questions that provide instant answer-specific feedback. Available 24/7.

- **Learn at your own pace with OWL**, a study smart system that ensures you've mastered each concept before you move on.

- **Access an e-version of your textbook** enhanced with videos and animations, highlighting, the ability to add notes, and more.

To get started, use the access code that may have been packaged with your text or purchase access online.  
Check with your instructor to verify that OWL is required for your course before purchasing.

# www.cengage.com/OWL



# Biochemistry





# Biochemistry

7th EDITION

**Mary K. Campbell**

*Mount Holyoke College*

**Shawn O. Farrell**



Australia • Brazil • Japan • Korea • Mexico • Singapore • Spain • United Kingdom • United States

This is an electronic version of the print textbook. Due to electronic rights restrictions, some third party content may be suppressed. Editorial review has deemed that any suppressed content does not materially affect the overall learning experience. The publisher reserves the right to remove content from this title at any time if subsequent rights restrictions require it. For valuable information on pricing, previous editions, changes to current editions, and alternate formats, please visit [www.cengage.com/highered](http://www.cengage.com/highered) to search by ISBN#, author, title, or keyword for materials in your areas of interest.



**Biochemistry, 7th Edition****Mary K. Campbell, Shawn O. Farrell**

Publisher: Mary Finch

Developmental Editor: Alyssa White

Assistant Editor: Jon Olafsson

Editorial Assistant: Laura Bowen

Senior Media Editor: Lisa Weber

Media Editor: Stephanie Van Camp

Marketing Manager: Barb Bartoszek

Marketing Assistant: Kevin Carroll

Marketing Communications Manager:  
Linda Yip

Content Project Manager: Teresa L. Trego

Art Director: John Walker

Print Buyer: Judy Inouye

Rights Acquisitions Specialist:  
Dean Dauphinais

Production Service: PreMediaGlobal

Text Designer: RHDG Riezebos Holzbaur

Photo Researcher: Bill Smith Group

Text Researcher: Sarah D'Stair

Copy Editor: PreMediaGlobal

Illustrator: PreMediaGlobal/2064 Design

OWL producers: Stephen Battisti, Cindy  
Stein, David Hart (Center for Educational  
Software Development), University of  
Massachusetts, Amherst

Cover Designer: RHDG Riezebos Holzbaur

Cover Image: Lacey Ann Johnson/  
Getty Images

Compositor: PreMediaGlobal

© 2012, 2009 Brooks/Cole, Cengage Learning

ALL RIGHTS RESERVED. No part of this work covered by the copyright herein may be reproduced, transmitted, stored or used in any form or by any means graphic, electronic, or mechanical, including but not limited to photocopying, recording, scanning, digitizing, taping, Web distribution, information networks, or information storage and retrieval systems, except as permitted under Section 107 or 108 of the 1976 United States Copyright Act, without the prior written permission of the publisher.

For product information and technology assistance, contact us at

**Cengage Learning Customer & Sales Support, 1-800-354-9706**For permission to use material from this text or product,  
submit all requests online at **[cengage.com/permissions](http://cengage.com/permissions)**Further permissions questions can be emailed to  
**[permissionrequest@cengage.com](mailto:permissionrequest@cengage.com)**

Library of Congress Control Number: 2010935869

Student Edition:

ISBN-13: 978-0-8400-6858-3

ISBN-10: 0-8400-6858-1

**Brooks/Cole**

20 Davis Drive

Belmont, CA 94002-3098

USA

Cengage Learning is a leading provider of customized learning solutions with office locations around the globe, including Singapore, the United Kingdom, Australia, Mexico, Brazil, and Japan. Locate your local office at **[www.cengage.com/global](http://www.cengage.com/global)**

Cengage Learning products are represented in Canada by  
Nelson Education, Ltd.

To learn more about Brooks/Cole, visit **[www.cengage.com/brookscole](http://www.cengage.com/brookscole)**

Purchase any of our products at your local college store or at our  
preferred online store **[www.cengagebrain.com](http://www.cengagebrain.com)**

Printed in Canada

1 2 3 4 5 6 7 14 13 12 11 10





---

*To all of those who made this text possible and especially to all of the students who will use it.*

—Mary K. Campbell

*To the returning adult students in my class, especially those with children and a full-time job . . . my applause.*

—Shawn O. Farrell

# About the Authors



## Mary K. Campbell

Mary K. Campbell is professor emeritus of chemistry at Mount Holyoke College, where she taught a one-semester biochemistry course and advised undergraduates working on biochemical research projects. She frequently taught general chemistry and physical chemistry as well. At some point in her 36 years at Mount Holyoke, she taught every subfield of chemistry, except the lecture portion of organic chemistry. Her avid interest in writing led to the publication of the first six highly successful editions of this textbook. Originally from Philadelphia, Mary received her Ph.D. from Indiana University and did post-doctoral work in biophysical chemistry at Johns Hopkins University. Her area of interest includes researching the physical chemistry of biomolecules, specifically, spectroscopic studies of protein–nucleic acid interactions.

Mary enjoys traveling and recently visited parts of Mexico near her current home in Tucson, Arizona. She participates in events at the University of Arizona and enjoys hiking in the desert and the mountains.



## Shawn O. Farrell

Shawn O. Farrell grew up in northern California and received a B.S. degree in biochemistry from the University of California, Davis, where he studied carbohydrate metabolism. He completed his Ph.D. in biochemistry at Michigan State University, where he studied fatty acid metabolism. For 18 years, Shawn worked at Colorado State University teaching undergraduate biochemistry lecture and laboratory courses. Because of his interest in biochemical education, Shawn has written a number of scientific journal articles about teaching biochemistry. He is the coauthor (with Lynn E. Taylor) of *Experiments in Biochemistry: A Hands-On Approach*. Shawn became interested in biochemistry while in college because it coincided with his passion for bicycle racing. An active outdoorsman, Shawn raced competitively for 17 years and now officiates at bicycle races around the world. He is currently the technical director of USA Cycling, the national governing body of bicycle racing in the United States. He is also an avid fly fisherman, a third-degree black belt in Tae Kwon Do, and a first-degree black belt in combat hapkido. Shawn has also written articles on fly fishing for *Salmon Trout Steelheader* magazine. His other passions are soccer, chess, and foreign languages. He is fluent in Spanish and French and is currently learning German and Italian.

On his fiftieth birthday, he had his first downhill skiing lesson and now cannot get enough of it. Never tired of education, he visited CSU again, this time from the other side of the podium, and earned his Master of Business Administration in 2008.

# Brief Contents

## Magazine: Hot Topics in Biochemistry

Green Fluorescent Protein: Jellyfish and Green Monkeys HT3

DNA and Family Trees: Who Is a Relative? HT6

Diabetes: An Epidemic for Modern Times HT10

Just One Word: Nanotechnology HT13

Small, Smaller, Smallest: Beyond the Electron Microscope to Single Molecules HT17

HPV Vaccines: Waging the War on Cervical Cancer HT20

Stem Cells: Hope or Hype? HT23

Doping in Sports: Good Science Gone Bad HT28

- 1 Biochemistry and the Organization of Cells 1
- 2 Water: The Solvent for Biochemical Reactions 35
- 3 Amino Acids and Peptides 61
- 4 The Three-Dimensional Structure of Proteins 83
- 5 Protein Purification and Characterization Techniques 117
- 6 The Behavior of Proteins: Enzymes 139
- 7 The Behavior of Proteins: Enzymes, Mechanisms, and Control 165
- 8 Lipids and Proteins Are Associated in Biological Membranes 193
- 9 Nucleic Acids: How Structure Conveys Information 227
- 10 Biosynthesis of Nucleic Acids: Replication 253
- 11 Transcription of the Genetic Code: The Biosynthesis of RNA 281
- 12 Protein Synthesis: Translation of the Genetic Message 323

- 13** Nucleic Acid Biotechnology Techniques 355
- 14** Viruses, Cancer, and Immunology 395
- 15** The Importance of Energy Changes and Electron Transfer in Metabolism 429
- 16** Carbohydrates 451
- 17** Glycolysis 481
- 18** Storage Mechanisms and Control in Carbohydrate Metabolism 507
- 19** The Citric Acid Cycle 533
- 20** Electron Transport and Oxidative Phosphorylation 563
- 21** Lipid Metabolism 591
- 22** Photosynthesis 627
- 23** The Metabolism of Nitrogen 653
- 24** Integration of Metabolism: Cellular Signaling 687

# Contents

## Magazine: Hot Topics in Biochemistry

**Green Fluorescent Protein: Jellyfish and Green Monkeys** HT3

**DNA and Family Trees: Who Is a Relative?** HT6

**Diabetes: An Epidemic for Modern Times** HT10

**Just One Word: Nanotechnology** HT13

**Small, Smaller, Smallest: Beyond the Electron Microscope to Single Molecules** HT17

**HPV Vaccines: Waging the War on Cervical Cancer** HT20

**Stem Cells: Hope or Hype?** HT23

**Doping in Sports: Good Science Gone Bad** HT28

## 1 Biochemistry and the Organization of Cells 1

**1.1 Basic Themes** 1

**1.2 Chemical Foundations of Biochemistry** 3

**1.3 The Beginnings of Biology: Origin of Life** 4

The Earth and Its Age 4

Biomolecules 6

Molecules to Cells 10

**1.4 The Biggest Biological Distinction—Prokaryotes and Eukaryotes** 13

**1.5 Prokaryotic Cells** 16

**1.6 Eukaryotic Cells** 16

**1.7 Five Kingdoms, Three Domains** 21

*Biochemical Connections:* Extremophiles: The Toast of the Industry 23

**1.8 Common Ground for All Cells** 24

**1.9 Biochemical Energetics** 26

**1.10 Energy and Change** 27

**1.11 Spontaneity in Biochemical Reactions** 28

**1.12 Life and Thermodynamics** 28

*Biochemical Connections:* Predicting Reactions 29

**Summary** 30

**Review Exercises** 31

**Annotated Bibliography** 33

## 2 Water: The Solvent for Biochemical Reactions 35

**2.1 Water and Polarity** 35

Solvent Properties of Water 36

**2.2 Hydrogen Bonds** 39

Other Biologically Important Hydrogen Bonds 42

*Biochemical Connections:* How Basic Chemistry Affects Life: The Importance of the Hydrogen Bond 43

**2.3 Acids, Bases, and pH** 44

**2.4 Titration Curves** 47

**2.5 Buffers** 49

*Biochemical Connections:* Buffer Selection 55

*Biochemical Connections:* Some Physiological Consequences of Blood Buffering 56

*Biochemical Connections:* Lactic Acid—Not Always the Bad Guy 57

**Summary** 57

**Review Exercises** 58

**Annotated Bibliography** 60

## 3 Amino Acids and Peptides 61

**3.1 Amino Acids Exist in a Three-Dimensional World** 61

**3.2 Individual Amino Acids: Their Structures and Properties** 62

Uncommon Amino Acids 66

**3.3 Amino Acids Can Act as Both Acids and Bases** 67

*Biochemical Connections:* Amino Acids to Calm Down and Pep Up 68

### 3.4 The Peptide Bond 72

### 3.5 Small Peptides with Physiological Activity 74

*Biochemical Connections:* Amino Acids Go Many Different Places 75

*Biochemical Connections:* Peptide Hormones—Small Molecules with Big Effects 77

*Biochemical Connections:* Phenylketonuria—Little Molecules Have Big Effects 78

*Biochemical Connections:* Peptide Hormones Revisited 79

**Summary** 79

**Review Exercises** 80

**Annotated Bibliography** 81

## 4 The Three-Dimensional Structure of Proteins 83

### 4.1 Protein Structure and Function 83

### 4.2 Primary Structure of Proteins 84

### 4.3 Secondary Structure of Proteins 84

*Biochemical Connections:* Complete Proteins and Nutrition 85

Periodic Structures in Protein Backbones 85

Irregularities in Regular Structures 87

Supersecondary Structures and Domains 88

The Collagen Triple Helix 90

Two Types of Protein Conformations: Fibrous and Globular 93

### 4.4 Tertiary Structure of Proteins 93

Forces Involved in Tertiary Structures 94

Myoglobin: An Example of Protein Structure 96

Denaturation and Refolding 99

### 4.5 Quaternary Structure of Proteins 100

Hemoglobin 100

Conformational Changes That Accompany Hemoglobin Function 102

### 4.6 Protein Folding Dynamics 106

Hydrophobic Interactions: A Case Study in Thermodynamics 107

The Importance of Correct Folding 109

Protein-Folding Chaperones 110

*Biochemical Connections:* Protein Folding Diseases 111

**Summary** 113

**Review Exercises** 114

**Annotated Bibliography** 115

## 5 Protein Purification and Characterization Techniques 117

### 5.1 Extracting Pure Proteins from Cells 117

### 5.2 Column Chromatography 118

### 5.3 Electrophoresis 126

### 5.4 Determining the Primary Structure of a Protein 127

Cleavage of the Protein into Peptides 129

Sequencing of Peptides: The Edman Method 129

*Biochemical Connections:* The Power of Mass Spectrometry 134

*Biochemical Connections:* Pulling It All Together 135

**Summary** 135

**Review Exercises** 136

**Annotated Bibliography** 138

## 6 The Behavior of Proteins: Enzymes 139

### 6.1 Enzymes Are Effective Biological Catalysts 139

### 6.2 Kinetics versus Thermodynamics 139

*Biochemical Connections:* Enzymes as Markers for Disease 142

### 6.3 Enzyme Kinetic Equations 142

### 6.4 Enzyme–Substrate Binding 144

### 6.5 Examples of Enzyme-Catalyzed Reactions 146

*Biochemical Connections:* Enzymes and Memory 147

### 6.6 The Michaelis–Menten Approach to Enzyme Kinetics 148

*Biochemical Connections:* Enzyme Lets You Enjoy Champagne 154

### 6.7 Enzyme Inhibition 155

*Biochemical Connections:* Practical Information from Kinetic Data 157

*Biochemical Connections:* Enzyme Inhibition in the Treatment of AIDS 160

**Summary** 160

**Review Exercises** 161

**Annotated Bibliography** 163

## 7 The Behavior of Proteins: Enzymes, Mechanisms, and Control 165

### 7.1 The Behavior of Allosteric Enzymes 165

### 7.2 The Concerted and Sequential Models for Allosteric Enzymes 169

*Biochemical Connections:* Allosterism: Drug Companies Exploit the Concept 173

### 7.3 Control of Enzyme Activity by Phosphorylation 173

### 7.4 Zymogens 175

### 7.5 The Nature of the Active Site 177

### 7.6 Chemical Reactions Involved in Enzyme Mechanisms 182

### 7.7 The Active Site and Transition States 184

*Biochemical Connections:* Families of Enzymes: Proteases 185

### 7.8 Coenzymes 187

*Biochemical Connections:* Catalytic Antibodies against Cocaine 189

*Biochemical Connections:* Catalysts for Green Chemistry 190

**Summary** 190

**Review Exercises** 191

**Annotated Bibliography** 192

## 8 Lipids and Proteins Are Associated in Biological Membranes 193

### 8.1 The Definition of a Lipid 193

### 8.2 The Chemical Natures of the Lipid Types 193

*Biochemical Connections:* Ceramides, Oxygen, Cancer, and Strokes 200

### 8.3 Biological Membranes 200

*Biochemical Connections:* Butter versus Margarine—Which Is Healthier? 204

### 8.4 The Kinds of Membrane Proteins 205

### 8.5 The Fluid-Mosaic Model of Membrane Structure 207

*Biochemical Connections:* Membranes in Drug Delivery 208

### 8.6 The Functions of Membranes 209

*Biochemical Connections:* Lipid Droplets Are Not Just Great Balls of Fat 213

### 8.7 Lipid-Soluble Vitamins and Their Functions 214

Vitamin A 214

Vitamin D 214

*Biochemical Connections:* Vision Has Great Chemistry 217

Vitamin E 218

Vitamin K 218

### 8.8 Prostaglandins and Leukotrienes 219

*Biochemical Connections:* Why Should We Eat More Salmon? 222

**Summary** 222

**Review Exercises** 223

**Annotated Bibliography** 225

## 9 Nucleic Acids: How Structure Conveys Information 227

### 9.1 Levels of Structure in Nucleic Acids 227

### 9.2 The Covalent Structure of Polynucleotides 227

*Biochemical Connections:* The DNA Family Tree 232

### 9.3 The Structure of DNA 232

*Biochemical Connections:* Who Owns Your Genes? 239

*Biochemical Connections:* The Human Genome Project: Treasure or Pandora's Box? 240

### 9.4 Denaturation of DNA 241

### 9.5 The Principal Kinds of RNA and Their Structures 242

*Biochemical Connections:* Why Identical Twins Are Not Identical 248

*Biochemical Connections:* Synthetic Genome Created 249

**Summary** 249

**Review Exercises** 250

**Annotated Bibliography** 251

## 10 Biosynthesis of Nucleic Acids: Replication 253

### 10.1 The Flow of Genetic Information in the Cell 253

### 10.2 Replication of DNA 254

Semiconservative Replication 254

### 10.3 DNA Polymerase 256

Semidiscontinuous DNA Replication 256

DNA Polymerase from *E. coli* 258

### 10.4 Proteins Required for DNA Replication 261

Supercoiling and Replication 261

The Primase Reaction 262

Synthesis and Linking of New DNA Strands 263

### 10.5 Proofreading and Repair 263

*Biochemical Connections:* Why Does DNA Contain Thymine and Not Uracil? 267

### 10.6 DNA Recombination 268



**10.7 Eukaryotic DNA Replication 270**

**Biochemical Connections:** The SOS Response in *E. coli* 271

Eukaryotic DNA Polymerases 273

**Biochemical Connections:** Telomerase and Cancer 274

The Eukaryotic Replication Fork 274

**Biochemical Connections:** Self-Replicating RNAs 277

**Summary 277**

**Review Exercises 278**

**Annotated Bibliography 279**

## **11 Transcription of the Genetic Code: The Biosynthesis of RNA 281**

**11.1 Overview of Transcription 281****11.2 Transcription in Prokaryotes 282**

RNA Polymerase in *Escherichia coli* 282

Promoter Structure 283

Chain Initiation 284

Chain Elongation 284

Chain Termination 287

**11.3 Transcription Regulation in Prokaryotes 287**

Alternative  $\sigma$  Factors 287

Enhancers 289

Operons 289

Transcription Attenuation 294

**Biochemical Connections:** Riboswitches Provide Another Weapon Against Pathogens 295

**11.4 Transcription in Eukaryotes 296**

Structure of RNA Polymerase II 297

Pol II Promoters 298

Initiation of Transcription 299

Elongation and Termination 301

**11.5 Transcription Regulation in Eukaryotes 302**

Enhancers and Silencers 302

**Biochemical Connections:** TFIIF—Making the Most Out of the Genome 303

Response Elements 304

**11.6 Non-Coding RNAs 307**

**Biochemical Connections:** A MicroRNA Helps Regenerate Nerve Synapses After Injury 308

**Biochemical Connections:** CREB—The Most Important Protein You Have Never Heard Of? 309

**11.7 Structural Motifs in DNA-Binding Proteins 309**

DNA-Binding Domains 309

Helix–Turn–Helix Motifs 309

Zinc Fingers 311

Basic-Region Leucine Zipper Motif 311

Transcription-Activation Domains 311

**11.8 Posttranscriptional RNA Modification 312**

Transfer RNA and Ribosomal RNA 313

Messenger RNA 314

The Splicing Reaction: Lariats and Snurps 315

Alternative RNA Splicing 316

**11.9 Ribozymes 317**

**Biochemical Connections:** Proofreading in Transcription? RNA Fills In Another Missing Piece 319

**Summary 319**

**Review Exercises 321**

**Annotated Bibliography 322**

## **12 Protein Synthesis: Translation of the Genetic Message 323**

**12.1 Translating the Genetic Message 323****12.2 The Genetic Code 324**

Codon–Anticodon Pairing and Wobble 326

**12.3 Amino Acid Activation 329****12.4 Prokaryotic Translation 331**

Ribosomal Architecture 331

Chain Initiation 331

Chain Elongation 334

Chain Termination 336

The Ribosome Is a Ribozyme 336

**Biochemical Connections:** The 21st Amino Acid 339

Polysomes 340

**12.5 Eukaryotic Translation 341**

Chain Initiation 342

Chain Elongation 343

**Biochemical Connections:** Protein Synthesis Makes Memories 344

Chain Termination 345

Coupled Transcription and Translation in Eukaryotes? 345

**12.6 Posttranslational Modification of Proteins 345**

**Biochemical Connections:** Chaperones: Preventing Unsuitable Associations 346

**12.7 Protein Degradation 347**

*Biochemical Connections:* Silent Mutations Are Not Always Silent 348

*Biochemical Connections:* How Do We Adapt to High Altitude? 351

**Summary 352**

**Review Exercises 352**

**Annotated Bibliography 354**

**13 Nucleic Acid Biotechnology Techniques 355****13.1 Purification and Detection of Nucleic Acids 355**

Separation Techniques 355

Detection Methods 356

**13.2 Restriction Endonucleases 357**

Many Restriction Endonucleases Produce “Sticky Ends” 358

**13.3 Cloning 360**

Using “Sticky Ends” to Construct Recombinant DNA 360

**13.4 Genetic Engineering 367**

DNA Recombination Occurs in Nature 368

Bacteria as “Protein Factories” 368

*Biochemical Connections:* Genetic Engineering in Agriculture 369

Protein Expression Vectors 370

Genetic Engineering in Eukaryotes 372

*Biochemical Connections:* Human Proteins through Genetic Recombination Techniques 373

**13.5 DNA Libraries 374**

*Biochemical Connections:* Fusion Proteins and Fast Purifications 375

Finding an Individual Clone in a DNA Library 376

**13.6 The Polymerase Chain Reaction 377**

Quantitative PCR allows sensitive measurement of DNA samples 379

*Biochemical Connections:* CSI: Biochemistry—Forensic Uses of DNA Testing 379

**13.7 DNA Fingerprinting 381**

Restriction-Fragment Length Polymorphisms: A Powerful Method for Forensic Analysis 382

**13.8 Sequencing DNA 384**

*Biochemical Connections:* RNA Interference—The Newest Way to Study Genes 385

**13.9 Genomics and Proteomics 387**

The Power of Microarrays—Robotic Technology Meets Biochemistry 388

Protein Arrays 390

**Summary 391**

**Review Exercises 392**

**Annotated Bibliography 393**

**14 Viruses, Cancer, and Immunology 395****14.1 Viruses 395**

Families of Viruses 395

Virus Life Cycles 396

*Biochemical Connections:* A Little Swine Goes a Long Way 400

**14.2 Retroviruses 401**

*Biochemical Connections:* Viruses Are Used for Gene Therapy 403

**14.3 The Immune System 404**

Innate Immunity—The Front Lines of Defense 405

Acquired Immunity: Cellular Aspects 406

T-Cell Functions 406

T-Cell Memory 411

The Immune System: Molecular Aspects 411

*Biochemical Connections:* Modern Science Takes on the Flu Virus 413

Distinguishing Self from Nonself 414

*Biochemical Connections:* Viral RNAs Outwit the Immune System 415

**14.4 Cancer 416**

*Biochemical Connections:* Cancer: The Dark Side of the Human Genome 417

Oncogenes 418

Tumor Suppressors 419

Viruses and Cancer 421

Viruses Helping Cure Cancer 421

*Biochemical Connections:* Nanotech Tackles Cancer 422

*Biochemical Connections:* Attacking the Symptoms instead of the Disease? 424

**Summary 425**

**Review Exercises 426**

**Annotated Bibliography 427**

## 15 The Importance of Energy Changes and Electron Transfer in Metabolism 429

### 15.1 Standard States for Free-Energy Changes 429

### 15.2 A Modified Standard State for Biochemical Applications 430

*Biochemical Connections:* Living Things Need Energy—How Do They Use It? 431

### 15.3 The Nature of Metabolism 432

### 15.4 The Role of Oxidation and Reduction in Metabolism 432

*Biochemical Connections:* Living Things Are Unique Thermodynamic Systems 433

### 15.5 Coenzymes in Biologically Important Oxidation–Reduction Reactions 434

### 15.6 Coupling of Production and Use of Energy 438

*Biochemical Connections:* ATP in Cell Signaling 441

### 15.7 Coenzyme A in Activation of Metabolic Pathways 443

Summary 446

Review Exercises 447

Annotated Bibliography 449

## 16 Carbohydrates 451

### 16.1 Sugars: Their Structures and Stereochemistry 451

*Biochemical Connections:* Low-Carbohydrate Diets 457

### 16.2 Reactions of Monosaccharides 459

*Biochemical Connections:* Vitamin C Is Related to Sugars 461

### 16.3 Some Important Oligosaccharides 464

*Biochemical Connections:* Fruits, Flowers, Striking Colors, and Medicinal Uses Too 466

*Biochemical Connections:* Lactose Intolerance: Why Do So Many People Not Want to Drink Milk? 467

### 16.4 Structures and Functions of Polysaccharides 468

*Biochemical Connections:* Why Is Dietary Fiber So Good for You? 474

### 16.5 Glycoproteins 475

*Biochemical Connections:* Glycoproteins and Blood Transfusions 476

Summary 477

Review Exercises 477

Annotated Bibliography 479

## 17 Glycolysis 481

### 17.1 The Overall Pathway of Glycolysis 481

*Biochemical Connections:* Biofuels from Fermentation 484

### 17.2 Conversion of Six-Carbon Glucose to Three-Carbon Glyceraldehyde-3-Phosphate 485

*Biochemical Connections:* Dolphins as a Model for Humans with Diabetes 488

### 17.3 Glyceraldehyde-3-Phosphate Is Converted to Pyruvate 491

### 17.4 Anaerobic Metabolism of Pyruvate 499

*Biochemical Connections:* What Is the Connection between Anaerobic Metabolism and Dental Plaque? 502

*Biochemical Connections:* Fetal Alcohol Syndrome 503

### 17.5 Energy Production in Glycolysis 503

Summary 504

Review Exercises 505

Annotated Bibliography 506

## 18 Storage Mechanisms and Control in Carbohydrate Metabolism 507

### 18.1 How Glycogen Is Produced and Degraded 507

*Biochemical Connections:* Why Do Athletes Go In for Glycogen Loading? 514

### 18.2 Gluconeogenesis Produces Glucose from Pyruvate 514

### 18.3 Control of Carbohydrate Metabolism 519

*Biochemical Connections:* Using Pyruvate Kinase Isozymes to Treat Cancer 523

### 18.4 Glucose Is Sometimes Diverted through the Pentose Phosphate Pathway 524

*Biochemical Connections:* The Pentose Phosphate Pathway and Hemolytic Anemia 528

Summary 529

Review Exercises 530

Annotated Bibliography 531

## 19 The Citric Acid Cycle 533

### 19.1 The Central Role of the Citric Acid Cycle in Metabolism 533

### 19.2 The Overall Pathway of the Citric Acid Cycle 533

### 19.3 How Pyruvate Is Converted to Acetyl-CoA 535

### 19.4 The Individual Reactions of the Citric Acid Cycle 540

*Biochemical Connections:* Fluorine Compounds and Carbohydrate Metabolism 543

### 19.5 Energetics and Control of the Citric Acid Cycle 547

### 19.6 The Glyoxylate Cycle: A Related Pathway 550

### 19.7 The Citric Acid Cycle in Catabolism 551

### 19.8 The Citric Acid Cycle in Anabolism 552

*Biochemical Connections:* Why Can't Animals Use All the Same Energy Sources as Plants and Bacteria? 556

### 19.9 The Link to Oxygen 557

*Biochemical Connections:* Why Is It So Hard to Lose Weight? 558

Summary 559

Review Exercises 560

Annotated Bibliography 561

## 20 Electron Transport and Oxidative Phosphorylation 563

### 20.1 The Role of Electron Transport in Metabolism 563

### 20.2 Reduction Potentials in the Electron Transport Chain 564

### 20.3 Organization of Electron Transport Complexes 566

*Biochemical Connections:* Mito What . . . ? The Consequences of Mitochondrial Disease 572

### 20.4 The Connection between Electron Transport and Phosphorylation 574

### 20.5 The Mechanism of Coupling in Oxidative Phosphorylation 577

*Biochemical Connections:* What Does Brown Adipose Tissue Have to Do with Obesity? 580

### 20.6 Respiratory Inhibitors Can Be Used to Study Electron Transport 580

### 20.7 Shuttle Mechanisms 583

*Biochemical Connections:* Sports and Metabolism 585

### 20.8 The ATP Yield from Complete Oxidation of Glucose 586

Summary 586

Review Exercises 588

Annotated Bibliography 589

## 21 Lipid Metabolism 591

### 21.1 Lipids Are Involved in the Generation and Storage of Energy 591

### 21.2 Catabolism of Lipids 591

### 21.3 The Energy Yield from the Oxidation of Fatty Acids 596

### 21.4 Catabolism of Unsaturated Fatty Acids and Odd-Carbon Fatty Acids 598

### 21.5 Ketone Bodies 601

### 21.6 Fatty-Acid Biosynthesis 602

*Biochemical Connections:* Transcription Activators in Lipid Biosynthesis 602

### 21.7 Synthesis of Acylglycerols and Compound Lipids 609

Triacylglycerols 609

*Biochemical Connections:* A Gene for Obesity 609

*Biochemical Connections:* Acetyl-CoA Carboxylase—A New Target in the Fight against Obesity? 612

### 21.8 Cholesterol Biosynthesis 613

*Biochemical Connections:* Atherosclerosis 622

Summary 623

Review Exercises 624

Annotated Bibliography 625

## 22 Photosynthesis 627

### 22.1 Chloroplasts Are the Site of Photosynthesis 627

*Biochemical Connections:* The Relationship between Wavelength and Energy of Light 630

### 22.2 Photosystems I and II and the Light Reactions of Photosynthesis 631

Cyclic Electron Transport in Photosystem I 635

### 22.3 Photosynthesis and ATP Production 637

*Biochemical Connections:* Improving the Yield of Anti-Malarial Plants 638

### 22.4 Evolutionary Implications of Photosynthesis with and without Oxygen 639

*Biochemical Connections:* Plants Feed Animals—Plants Need Energy—Plants Can Produce Energy 640

### 22.5 Dark Reactions of Photosynthesis Fix CO<sub>2</sub> 642

### 22.6 CO<sub>2</sub> Fixation in Tropical Plants 646

*Biochemical Connections:* Chloroplast Genes 646

Summary 649  
 Review Exercises 650  
 Annotated Bibliography 651

## 23 The Metabolism of Nitrogen 653

- 23.1 Nitrogen Metabolism: An Overview 653
- 23.2 Nitrogen Fixation 653
  - Biochemical Connections:* Why Is the Nitrogen Content of Fertilizers So Important? 655
- 23.3 Feedback Inhibition in Nitrogen Metabolism 656
- 23.4 Amino Acid Biosynthesis 656
- 23.5 Essential Amino Acids 666
- 23.6 Amino Acid Catabolism 666
  - Excretion of Excess Nitrogen 667
  - Biochemical Connections:* Water and the Disposal of Nitrogen Wastes 668
  - Biochemical Connections:* Chemotherapy and Antibiotics—Taking Advantage of the Need for Folic Acid 671
- 23.7 Purine Biosynthesis 672
  - Anabolism of Inosine Monophosphate 672
- 23.8 Purine Catabolism 674
  - Biochemical Connections:* Cytoplasmic Defenses against Gout 677
- 23.9 Pyrimidine Biosynthesis and Catabolism 678
  - The Anabolism of Pyrimidine Nucleotides 678
  - Pyrimidine Catabolism 678
- 23.10 Conversion of Ribonucleotides to Deoxyribonucleotides 681
- 23.11 Conversion of dUDP to dTTP 682
  - Summary 683
  - Review Exercises 683
  - Annotated Bibliography 685

## 24 Integration of Metabolism: Cellular Signaling 687

- 24.1 Connections between Metabolic Pathways 687
- 24.2 Biochemistry and Nutrition 687
  - Biochemical Connections:* Alcohol Consumption and Addiction 689
  - Biochemical Connections:* Iron: An Example of a Mineral Requirement 692
  - The Food Pyramid 692
- 24.3 Hormones and Second Messengers 695
  - Hormones 695
  - Second Messengers 699
  - Cyclic AMP and G Proteins 699
  - Calcium Ion as a Second Messenger 700
  - Receptor Tyrosine Kinases 701
- 24.4 Hormones and the Control of Metabolism 703
- 24.5 Insulin and Its Effects 706
  - Insulin Receptors 707
  - Insulin's Effect on Glucose Uptake 707
  - Biochemical Connections:* Insulin and Low-Carbohydrate Diets 708
  - Insulin Affects Many Enzymes 708
  - Diabetes 708
  - Biochemical Connections:* A Workout a Day Keeps Diabetes Away? 709
  - Insulin and Sports 710
  - Biochemical Connections:* Aging and the Search for Longevity 710
- Summary 712
- Review Exercises 712
- Annotated Bibliography 714

Glossary  
 Answers to Questions  
 Index



# Preface

This text is intended for students in any field of science or engineering who want a one-semester introduction to biochemistry but who do not intend to be biochemistry majors. Our main goal in writing this book is to make biochemistry as clear and applied as possible and to familiarize science students with the major aspects of biochemistry. For students of biology, chemistry, physics, geology, nutrition, sports physiology, and agriculture, biochemistry impacts greatly on the content of their fields, especially in the areas of medicine and biotechnology. For engineers, studying biochemistry is especially important for those who hope to enter a career in biomedical engineering or some form of biotechnology.

Students who will use this text are at an intermediate level in their studies. A beginning biology course, general chemistry, and at least one semester of organic chemistry are assumed as preparation.

## New Integration with OWL



**Instructors** For the first time, this best-selling text is available with OWL, the #1 online homework and tutorial system for chemistry. With OWL, you can improve student learning outcomes, while saving time in grading homework and tracking student progress. Developed by chemistry instructors for teaching chemistry, OWL makes homework management a breeze and enables you to address students' different learning styles through tutorials, simulations, visualization exercises, and algorithmically generated homework questions with answer-specific feedback. With OWL's mastery learning approach, students work at their own pace until they understand each concept/skill. OWL includes an interactive, multimedia eBook.

## New to This Edition

All textbooks evolve to meet the interests and needs of students and instructors and to include the most current information. Several changes mark this edition.

**Biochemistry Hot Topics** A new supplement features up-to-date articles on new breakthroughs and topics in the area of biochemistry such as nanotechnology, the green fluorescent protein, single-molecule visualization, diabetes, blood doping, HPV, stem cells, and HIV.

**More Biochemical Connections Boxes** In response to customers' demand for more Biochemical Connection boxes, we have added several new boxes to the text, such as "Protein Synthesis Makes Memories," "Nanotech Tackles Cancer," and "ATP in Cell Signaling."

**Updated Coverage** Each chapter in the text has been updated with the current developments and scientific findings in the biochemistry field.

**Expanded and Updated Coverage of Selected Topics** A revised section on how buffers work appears in Chapter 2, with a new Apply Your Knowledge box on the same topic. The coverage of physiological roles of peptide hormones is expanded with a new box in Chapter 3. Material on methods such as HPLC and mass spectrometry has been added to Chapter 5, as has a new Apply Your Knowledge box on protein purification. Chapter 8 has new material on methods such as fluorescence spectroscopy and differential scanning calorimetry. In Chapter 9, new boxes cover legal aspects of genetic research and synthetic genomes. The new box in Chapter 12 on protein synthesis and memory reinforces the new material on enzymes and memory in Chapter 6. Chapter 13 concentrates on methods, with a new section on quantitative PCR, as well as more coverage of genomics and proteomics. New boxes on cancer and on nanotechnology appear in Chapter 14. The role of ATP in cell signaling is newly covered in Chapter 15. Genetic aspects of mitochondrial disease are discussed in a new box in Chapter 20. The topic of energy requirements for plant growth is new in Chapter 22, as is a discussion of plant sources of antimalarial agents. Chapter 24 has a new box on iron requirements, as well as expanded material on longevity.

**Table of Changes by Chapter**

Chapter 1	New material on origin of life	Chapter 12	Added material on elongation factor P. Expanded material on release factors. Added new box on protein synthesis and memories. Expanded material on silent mutations. Added 13 new EOC questions.
Chapter 2	Revised section on how buffers work, as well as new Apply your Knowledge box on same topic	Chapter 13	Added stronger warning about ethidium bromide and material on non-carcinogenic DNA stains. Enhanced graphics on several figures. Added new section on quantitative PCR. Expanded section on genomics and proteomics. Added two new EOC questions.
Chapter 3	New box on vasopressin	Chapter 14	Added new box on the swine flu. Expanded material in box on gene therapy. Added new box on flu research. Expanded material on autoimmune diseases. Added new box on cancer. Added new box on nanotechnology. Added eight new EOC questions.
Chapter 4	Expanded box on protein folding diseases. Five new EOC exercises.	Chapter 15	New box on role of ATP in cell signaling
Chapter 5	Improved figure graphics. Added brief section on HPLC and reverse phase HPLC. Added an Apply your Knowledge section on protein purification. Added new box on mass spectrometry. Added six new EOC questions.	Chapter 16	Expanded treatment of low-carbohydrate diets in weight loss
Chapter 6	Added new box on enzymes and memory. Added new box on carbonic anhydrase and enjoyment of champagne. Added seven new EOC questions.	Chapter 17	New box on dolphins as a model for human diabetes
Chapter 7	Added new box on allosterism in medicine. Added brief section on caspases. Added nine new EOC questions.	Chapter 18	New box on role of pyruvate kinase enzyme in metabolism of cancer cells
Chapter 8	New material on role of ceramides in health, new material on methods (fluorescence spectroscopy, DSC), new box on membranes in drug delivery (basis for student chemical engineering project)	Chapter 19	Expanded treatment of role of fluorine compound in metabolism
Chapter 9	Edited section on supercoiling for clarity. Added box on legal implications of genetic research. Added new material on nucleosomes. Added material to box on human genome project. Added box on synthetic genomes. Added four new EOC questions.	Chapter 20	New box on genetics of mitochondrial disease
Chapter 10	Added new numbered section on DNA recombination. Improved graphics in box on SOS response. Added new box on self-replicating RNAs. Added four new EOC questions.	Chapter 21	New box on transcription activators in lipid biosynthesis, additional material on effect of inflammation on lipid levels in bloodstream
Chapter 11	Added new box on riboswitches. Added new numbered section on non-coding RNAs. Added 12 new EOC questions.	Chapter 22	New box on anti-malarial plants, new box on energy requirements for plant production and plants as sources of energy
		Chapter 23	New box on gout
		Chapter 24	Updated box on iron requirements. Expanded box on longevity. Added two new EOC questions.



**New Design and Enhanced Labeling in Art** Updated labeling in the illustrations throughout the text increases readability, which in turn enhances students' ability to comprehend key concepts. As a corollary to the book's updated art program, the design and color palette have also been modernized.

## Proven Features

**Visual Impact** Ideal for visual learners, this book's state-of-the-art approach helps students visualize key processes and understand important topics.

**Biochemical Connections** The Biochemical Connections boxes highlight special topics of particular interest to students. Topics frequently have clinical implications, such as cancer, AIDS, and nutrition. These essays help students make the connection between biochemistry and the real world.

**Apply Your Knowledge** The Apply Your Knowledge boxes are interspersed within chapters and are designed to provide students with problem-solving experience. The topics chosen are areas of study where students usually have the most difficulty. *Solutions* and *problem-solving strategies* are included, giving examples of the problem-solving approach for specific material.

**Early Inclusion of Thermodynamics** Select material on thermodynamics appears much earlier in the text. Chapter 1 includes sections on *Energy and Change*, *Spontaneity in Biochemical Reactions*, and *Life and Thermodynamics*. Also, Chapter 4 contains an extended section on *Protein Folding Dynamics*. We feel it is critical that students understand the driving force of biological processes and see that so much of biology (protein folding, protein–protein interactions, small molecule binding, etc.) is driven by the favorable disordering of water molecules.

**Summaries and Questions** Each chapter closes with a concise summary, a broad selection of questions, and an annotated online bibliography. As stated previously, the summaries have been completely revised to reflect the in-text “Q & A” framework. The number of questions has been expanded to provide additional self-testing of content mastery and more homework material. These exercises fall into four categories: *Recall*, *Reflect and Apply*, *Biochemical Connections*, and *Mathematical*. The *Recall* questions are designed for students to quickly assess their mastery of the material, while the *Reflect and Apply* questions are for students to work through more thought-provoking questions. *Biochemical Connections* questions test students on the *Biochemical Connections* essays in that chapter. The *Mathematical* questions complete the selection of exercises. These questions are quantitative in nature and focus on calculations.

**Glossary and Answers** The book ends with a glossary of important terms and concepts (including the section number where the term was first introduced), an answer section, and a detailed index.

## Organization

Because biochemistry is a multidisciplinary science, the first task in presenting it to students of widely varying backgrounds is to put it in context. Chapters 1 and 2 provide the necessary background and connect biochemistry to the other sciences. Chapters 3 through 8 focus on the structure and dynamics of important cellular components. Molecular biology is covered in Chapters 9 through 14. The final part of the book is devoted to intermediary metabolism.

Some topics are discussed several times, such as the control of carbohydrate metabolism. Subsequent discussions make use of and build on information

students have already learned. It is particularly useful to return to a topic after students have had time to assimilate and reflect on it.

The first two chapters of the book relate biochemistry to other fields of science. Chapter 1 deals with some of the less obvious relationships, such as the connections of biochemistry with physics, astronomy, and geology, mostly in the context of the origins of life. Functional groups on organic molecules are discussed from the point of view of their role in biochemistry. This chapter goes on to the more readily apparent linkage of biochemistry with biology, especially with respect to the distinction between prokaryotes and eukaryotes, as well as the role of organelles in eukaryotic cells. Chapter 2 builds on material familiar from general chemistry, such as buffers and the solvent properties of water, but emphasizes the biochemical point of view toward such material.

Chapters 3 through 8, covering the structure of cellular components, focus on the structure and dynamics of proteins and membranes in addition to giving an introduction to some aspects of molecular biology. Chapters 3, 4, 6, and 7 deal with amino acids, peptides, and the structure and action of proteins, including enzyme catalysis. Chapter 4 includes more material on thermodynamics, such as hydrophobic interactions. Chapter 5 focuses on techniques for isolating and studying proteins. The discussion of enzymes is split into two chapters (Chapters 6 and 7) to give students more time to fully understand enzyme kinetics and enzyme mechanisms. Chapter 8 treats the structure of membranes and their lipid components.

Chapters 9 through 14 explore the topics of molecular biology. Chapter 9 introduces the structure of nucleic acids. In Chapter 10, the replication of DNA is discussed. Chapter 11 focuses on transcription and gene regulation. This material on the biosynthesis of nucleic acids is split into two chapters to give students ample time to appreciate the workings of these processes. Chapter 12 finishes the topic with translation of the genetic message and protein synthesis. Chapter 13 focuses on biotechnology techniques, and Chapter 14 deals with viruses, cancer, and immunology.

Chapters 15 through 24 explore intermediary metabolism. Chapter 15 opens the topic with chemical principles that provide some unifying themes. Thermodynamic concepts learned earlier in general chemistry and in Chapter 1 are applied specifically to biochemical topics such as coupled reactions. In addition, this chapter explicitly makes the connection between metabolism and electron transfer (oxidation–reduction) reactions.

Coenzymes are introduced in this chapter and are discussed in later chapters in the context of the reactions in which they play a role. Chapter 16 discusses carbohydrates. Chapter 17 begins the overview of the metabolic pathways by discussing glycolysis. Glycogen metabolism, gluconeogenesis, and the pentose phosphate pathway (Chapter 18) provide bases for treating control mechanisms in carbohydrate metabolism. Discussion of the citric acid cycle is followed by the electron transport chain and oxidative phosphorylation in Chapters 19 and 20. The catabolic and anabolic aspects of lipid metabolism are dealt with in Chapter 21. In Chapter 22, photosynthesis rounds out the discussion of carbohydrate metabolism. Chapter 23 completes the survey of the pathways by discussing the metabolism of nitrogen-containing compounds such as amino acids, porphyrins, and nucleobases. Chapter 24 is a summary chapter. It gives an integrated look at metabolism, including a treatment of hormones and second messengers. The overall look at metabolism includes a brief discussion of nutrition and a somewhat longer one of the immune system.

This text gives an overview of important topics of interest to biochemists and shows how the remarkable recent progress of biochemistry impinges on other sciences. The length is intended to provide instructors with a choice of favorite topics without being overwhelming for the limited amount of time available in one semester.

## Alternative Teaching Options

The order in which individual chapters are covered can be changed to suit the needs of specific groups of students. Although we prefer an early discussion of thermodynamics, the portions of Chapters 1 and 4 that deal with thermodynamics can be covered at the beginning of Chapter 15, “The Importance of Energy Changes and Electron Transfer in Metabolism.” All of the molecular biology chapters (Chapters 9 through 14) can precede metabolism or can follow it, depending on the instructor’s choice. The order in which the material on molecular biology is treated can be varied according to the preference of the instructor.

## Supplements

This seventh edition of Campbell and Farrell’s *Biochemistry* is accompanied by the following rich array of online, electronic, and print supplements.

### Student and Instructor Resources

#### OWL for Biochemistry

**OWL with eBook 6-Month Instant Access ISBN-13: 978-1-111-47168-2**

By David Gross of the University of Massachusetts, Amherst, Susan Young of Hartwick College, and Reginald H. Garrett and Charles M. Grisham of the University of Virginia. **OWL** Online Web Learning offers more assignable, gradable content (including end-of chapter questions specific to this textbook) and more reliability and flexibility than any other system. OWL’s powerful course management tools allow instructors to control due dates, number of attempts, and whether students see answers or receive feedback on how to solve problems. OWL includes the **YouBook**, a Flash-based eBook that is interactive and customizable. It features a text edit tool that allows instructors to modify the textbook narrative as needed. With YouBook, instructors can quickly re-order entire sections and chapters or hide any content they don’t teach to create an eBook that perfectly matches their syllabus. Instructors can further customize the YouBook by publishing web links. The YouBook includes animated figures, video clips, highlighting, notes, and more.

Developed by chemistry instructors for teaching chemistry, OWL is the only system specifically designed to support **mastery learning**, where students work as long as they need to master each chemical concept and skill. OWL has already helped hundreds of thousands of students master chemistry through a wide range of assignment types, including tutorials, interactive simulations, and algorithmically generated homework questions that provide instant, answer-specific feedback.

OWL is continually enhanced with online learning tools to address the various learning styles of today’s students such as:

- **Quick Prep** review courses that help students learn essential skills to succeed in General and Organic Chemistry
- **Jmol** molecular visualization program for rotating molecules and measuring bond distances and angles
- **Go Chemistry**® mini video lectures on key concepts that students can play on their computers or download to their video iPods, smart phones, or personal video players

In addition, when you become an OWL user, you can expect service that goes far beyond the ordinary. For more information or to see a demo, please contact your Cengage Learning representative or visit us at [www.cengage.com/owl](http://www.cengage.com/owl).





**CengageBrain.com App** Now, students can prepare for class anytime and anywhere using the CengageBrain.com application developed specifically for the Apple iPhone® and iPod touch®, which allows students to access free study materials—book-specific quizzes, flash cards, related Cengage Learning materials and more—so they can study the way they want, when they want to . . . even on the go. For more information about this complimentary application, please visit [www.cengagebrain.com](http://www.cengagebrain.com).

**Visit CengageBrain.com** To access these and additional course materials, please visit [www.cengagebrain.com](http://www.cengagebrain.com). At the CengageBrain.com home page, search for the ISBN (from the back cover of your book) using the search box at the top of the page. This will take you to the product page where these resources can be found. (Instructors can log in at [login.cengage.com](http://login.cengage.com).)

## Instructor Resources

**PowerLecture with ExamView® Instructor's CD/DVD Package** (ISBN 10: 1-111-42561-2; ISBN 13: 978-1-111-42561-6) This digital presentation tool contains prepared lecture slides, as well as customizable materials. The DVD contains art, photos, and tables from the text in a variety of electronic formats, as well as simulations, animations, and mini movies to supplement your lectures. PowerLecture also includes:

**Instructor's Manual** by co-author Shawn Farrell. Each chapter includes a chapter summary, lecture notes, lecture outlines, and answers to all of the end-of-chapter exercises in the text.

**ExamView® Computerized Testing** by co-author Shawn Farrell. ExamView® software allows you to create, deliver, and customize tests based on questions authored specifically for the text. A test bank is also available as multiple-choice exam questions, available as Word and PDF files.

**Lecture Slides** Prepared Microsoft® PowerPoint® lecture slides by Paul Adams at the University of Arkansas cover all key points from the text in a convenient format that you can enhance with your own materials or with the supplied interactive video animations for personalized media-enhanced lectures.

## Student Resources

**Student Companion Website** includes a glossary, flashcards, and an interactive periodic table, which are accessible from [www.cengagebrain.com](http://www.cengagebrain.com).

**Experiments in Biochemistry: A Hands-On Approach, Second Edition** This interactive manual, by Shawn O. Farrell and Lynn E. Taylor, is for the introductory biochemistry laboratory course. It offers a great selection of classroom-tested experiments, each designed to be completed in a normal laboratory period (ISBN 10: 0-495-01317-X; ISBN 13: 978-0-495-01317-4).

**Lecture Notebook** This printed notebook contains selected lecture slides printed in booklet format with note-taking space on each page. The lecture notebook is free when packaged with a new copy of the textbook (ISBN 10: 1-111-42566-3; ISBN 13: 978-1-111-42566-1).

---

Apple, iPhone, iPod touch, and iTunes are trademarks of Apple Inc., registered in the U.S. and other countries.

## Acknowledgments

The help of many made this book possible. A grant from the Dreyfus Foundation made possible the experimental introductory course that was the genesis of many of the ideas for this text. Edwin Weaver and Francis DeToma from Mount Holyoke College gave much of their time and energy in initiating that course. Many others at Mount Holyoke were generous with their support, encouragement, and good ideas, especially Anna Harrison, Lilian Hsu, Dianne Baranowski, Sheila Browne, Janice Smith, Jeffrey Knight, Sue Ellen Frederick Gruber, Peter Gruber, Marilyn Pryor, Craig Woodard, Diana Stein, and Sue Rusiecki. Particular thanks go to Sandy Ward, science librarian, and to Rosalia Tungaraza, a biochemistry major in the class of 2004. Three students, Nam Ho, Ben Long, and Alejandra Pesquiera, in Chemical Engineering 443 (Senior Design II) at the University of Arizona took the Biotechnology Biochemical Connection in Chapter 8 and turned it into an actual drug delivery process. Their mentor was Harry Patton, engineer and entrepreneur, and their course instructor was Professor Kimberly Ogden. Professor Todd Hoare of the Department of Chemical Engineering at McMaster University and Professor Daniel Kohane of Harvard Medical School provided many useful pointers about how to turn their original publication into the final process. Special thanks to Laurie Stargell, Marve Paule, and Steven McBryant at Colorado State University for their help and editorial assistance. We thank the many biochemistry students who have used and commented on early versions of this text.

We would like to acknowledge colleagues who contributed their ideas and critiques of the manuscript. Some reviewers responded to specific queries regarding the text itself. We thank them for their efforts and their helpful suggestions.

### Reviewers Acknowledgments

Paul D. Adams, *University of Kansas*

Dan Davis, *University of Arkansas*

Nick Flynn, *Angelo State University*

Denise Greathouse, *University of Arkansas*

James R. Paulson, *University of Wisconsin-Oshkosh*

Kerry Smith, *Clemson University*

Alexandre G. Volkov, *Oakwood University*

We would also like to thank the people at Brooks/Cole, a part of Cengage Learning, who were essential to the development of this book: Alyssa White, development editor, whose creative ideas added so much to this new edition; Teresa Trego, senior production manager, who directed production, making what could have been a chore so much easier; and Mary Finch, publisher, who was a source of so much encouragement. We also thank media editors Lisa Weber and Stephanie VanCamp, marketing manager Barb Bartoszek, assistant editor Jon Olafsson, and editorial assistant Laura Bowen.

Patrick Franzen of PreMediaGlobal served diligently as our production editor. Chris Altolf of Bill Studios, photo researcher, did wonders with challenging searches. We extend our most sincere gratitude to those listed here and to all others to whom we owe the opportunity to do this book. Instrumental in the direction given to this project was the late John Vondeling. John was a legend in the publishing field. His guidance and friendship shall be missed.

## **A Final Note from Mary Campbell**

I thank my family and friends, whose moral support has meant so much to me in the course of my work. When I started this project years ago, I did not realize that it would become a large part of my life. It has been a thoroughly satisfying one.

## **A Final Note from Shawn Farrell**

I cannot adequately convey how impossible this project would have been without my wonderful family, who put up with a husband and father who became a hermit in the back office. My wife, Courtney, knows the challenge of living with me when I am working on four hours of sleep per night. It isn't pretty, and few would have been so understanding. I would also like to thank David Hall, book representative, for starting me down this path, and the late John Vondeling for giving me an opportunity to expand into other types of books and projects.

Lastly, of course, I thank all of my students who have helped proofread the seventh edition.



# Biochemistry and the Organization of Cells



Biochemistry unlocks the mysteries of the human body.

## 1.1 Basic Themes

### How does biochemistry describe life processes?

Living organisms, such as humans, and even the individual cells of which they are composed, are enormously complex and diverse. Nevertheless, certain unifying features are common to all living things from the simplest bacterium to the human being. They all use the same types of *biomolecules*, and they all use energy. As a result, organisms can be studied via the methods of chemistry and physics. The belief in “vital forces” (forces thought to exist only in living organisms) held by 19th-century biologists has long since given way to awareness of an underlying unity throughout the natural world.

Disciplines that appear to be unrelated to biochemistry can provide answers to important biochemical questions. For example, the MRI (magnetic resonance imaging) tests that play an important role in the health sciences originated with physicists, became a vital tool for chemists, and currently play a large role in biochemical research. The field of biochemistry draws on many disciplines, and its multidisciplinary nature allows it to use results from many sciences to answer questions about the *molecular nature of life processes*. Important applications of this kind of knowledge are made in medically related fields; an understanding of health and disease at the molecular level leads to more effective treatment of illnesses of many kinds.

The activities within a cell are similar to the transportation system of a city. The cars, buses, and taxis correspond to the molecules involved in reactions (or series of reactions) within a cell. The routes traveled by vehicles likewise can be compared to the reactions that occur in the life of the cell. Note particularly that many vehicles travel more than one route—for instance, cars and taxis can go almost anywhere—whereas other, more specialized modes of transportation, such as subways and streetcars, are confined to single paths. Similarly, some molecules play multiple roles, whereas others take part only in specific series of reactions. Also, *the routes operate simultaneously*, and we shall see that this is true of the many reactions within a cell.

To continue the comparison, the transportation system of a large city has more kinds of transportation than does a smaller one. Whereas a small city may have only cars, buses, and taxis, a large city may have all of these plus others, such as streetcars or subways. Analogously, some reactions are found in all cells, and others are found only in specific kinds of cells. Also, more structural features are found in the larger, more complex cells of larger organisms than in the simpler cells of organisms such as bacteria.

An inevitable consequence of this complexity is the large quantity of terminology that is needed to describe it; learning considerable new vocabulary is an essential part of the study of biochemistry. You will also see many cross-references in this book, which reflect the many connections among the processes that take place in the cell.

## Chapter Outline

### 1.1 Basic Themes

- How does biochemistry describe life processes?
- How did living things originate?

### 1.2 Chemical Foundations of Biochemistry

- Can a chemist make the molecules of life in a laboratory?
- What makes biomolecules special?

### 1.3 The Beginnings of Biology: Origin of Life

- How and when did the Earth come to be?
- How were biomolecules likely to have formed on the early Earth?
- Which came first—the catalysts or the hereditary molecules?

### 1.4 The Biggest Biological Distinction—Prokaryotes and Eukaryotes

- What is the difference between a prokaryote and a eukaryote?

### 1.5 Prokaryotic Cells

- How is prokaryotic DNA organized without a nucleus?

### 1.6 Eukaryotic Cells

- What are the most important organelles?
- What are some other important components of cells?

### 1.7 Five Kingdoms, Three Domains

- How do scientists classify living organisms today?
- Is there a simpler basis for classifying organisms?

### 1.8 Common Ground for All Cells

- Did eukaryotes develop from prokaryotes?
- Did symbiosis play a role in the development of eukaryotes?

### 1.9 Biochemical Energetics

- What is the source of energy in life processes?
- How do we measure energy changes in biochemistry?

### 1.10 Energy and Change

- What kinds of energy changes take place in living cells?

### 1.11 Spontaneity in Biochemical Reactions

- How can we predict what reactions will happen in living cells?

### 1.12 Life and Thermodynamics

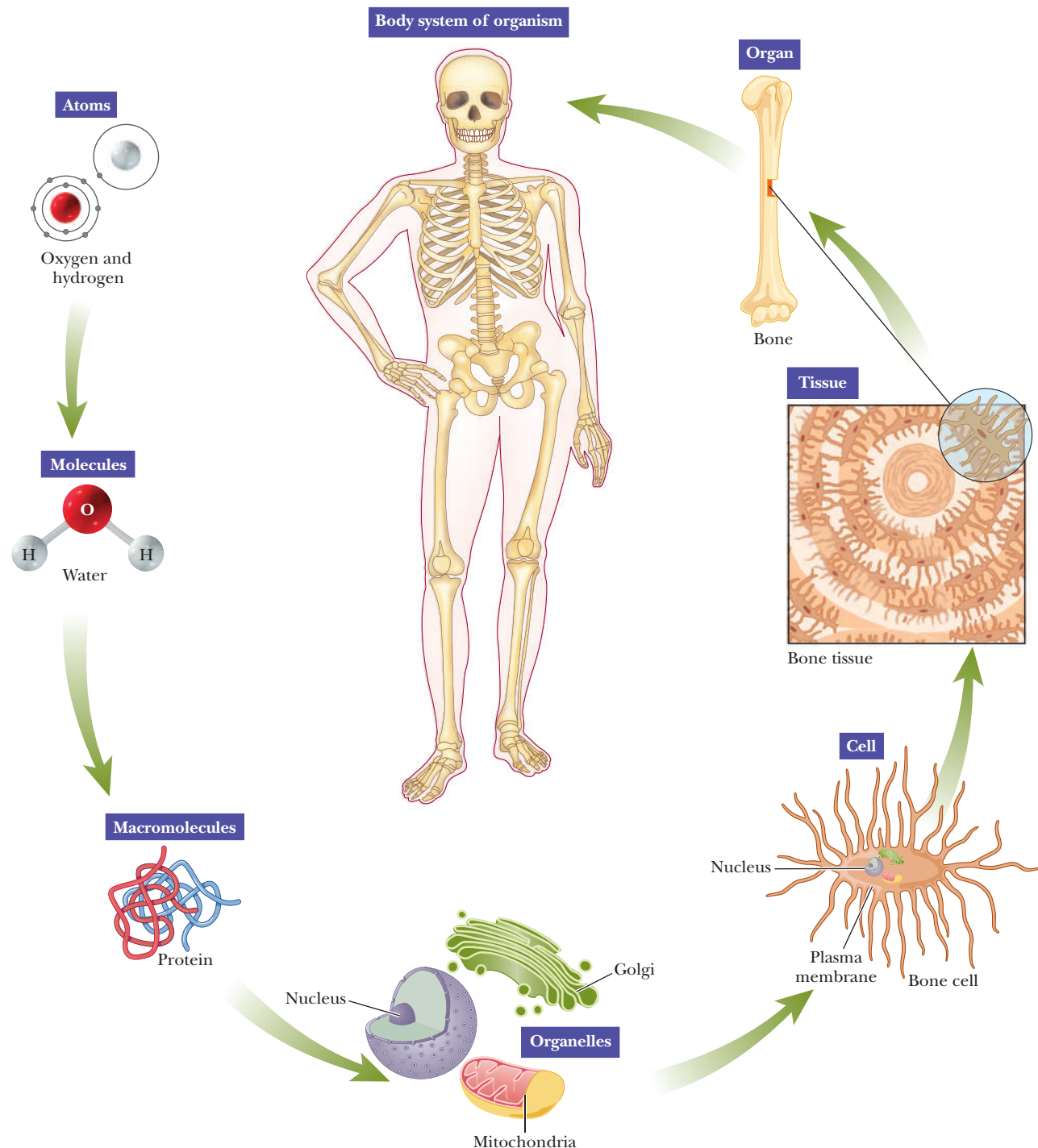
- Is life thermodynamically possible?

**OWL** Online homework for this chapter may be assigned in OWL.



### How did living things originate?

The fundamental similarity of cells of all types makes speculating on the origins of life a worthwhile question. How did the components of our bodies come to be and to do the things that they do? What are the molecules of life? Even the structures of comparatively small biomolecules consist of several parts. Large biomolecules, such as proteins and nucleic acids, have complex structures, and living cells are enormously more complex. Even so, *both molecules and cells must have arisen ultimately from very simple molecules*, such as water, methane, carbon dioxide, ammonia, nitrogen, and hydrogen (Figure 1.1). In turn, these simple



■ **FIGURE 1.1** Levels of structural organization in the human body. Note the hierarchy from simple to complex.

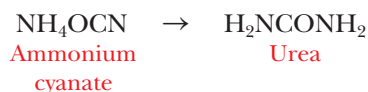
molecules must have arisen from atoms. The way in which the Universe itself, and the atoms of which it is composed, came to be is a topic of great interest to astrophysicists as well as other scientists. Simple molecules were formed by combining atoms, and reactions of simple molecules led in turn to more complex molecules. The molecules that play a role in living cells today are the same molecules as those encountered in organic chemistry; they simply operate in a different context.

## 1.2 Chemical Foundations of Biochemistry

**Organic chemistry** is the study of compounds of carbon and hydrogen and their derivatives. Because the cellular apparatus of living organisms is made up of carbon compounds, biomolecules are part of the subject matter of organic chemistry. Additionally, many carbon compounds are not found in any organism, and many topics of importance to organic chemistry have little connection with living things. We are going to concentrate on the aspects of organic chemistry that we need in order to understand what goes on in living cells.

### Can a chemist make the molecules of life in a laboratory?

Until the early part of the 19th century, there was a widely held belief in “vital forces,” forces presumably unique to living things. This belief included the idea that the compounds found in living organisms could not be produced in the laboratory. German chemist Friedrich Wöhler performed the critical experiment that disproved this belief in 1828. Wöhler synthesized urea, a well-known waste product of animal metabolism, from ammonium cyanate, a compound obtained from mineral (i.e., nonliving) sources.



It has subsequently been shown that any compound that occurs in a living organism can be synthesized in the laboratory, although in many cases the synthesis represents a considerable challenge to even the most skilled organic chemist.

The reactions of biomolecules can be described by the methods of organic chemistry, which requires the classification of compounds according to their **functional groups**. *The reactions of molecules are based on the reactions of their respective functional groups.*

### What makes biomolecules special?

Table 1.1 lists some biologically important functional groups. Note that most of these functional groups contain oxygen and nitrogen, which are among the most electronegative elements. As a result, many of these functional groups are polar, and their polar nature plays a crucial role in their reactivity. Some groups that are vitally important to organic chemists are missing from the table because molecules containing these groups, such as alkyl halides and acyl chlorides, do not have any particular applicability in biochemistry. Conversely, carbon-containing derivatives of phosphoric acid are mentioned infrequently in beginning courses on organic chemistry, but esters and anhydrides of phosphoric acid (Figure 1.2) are of vital importance in biochemistry. Adenosine triphosphate (ATP), a molecule that is the energy currency of the cell, contains both ester and anhydride linkages involving phosphoric acid.

Important classes of biomolecules have characteristic functional groups that determine their reactions. We shall discuss the reactions of the functional groups when we consider the compounds in which they occur.

**TABLE 1.1**

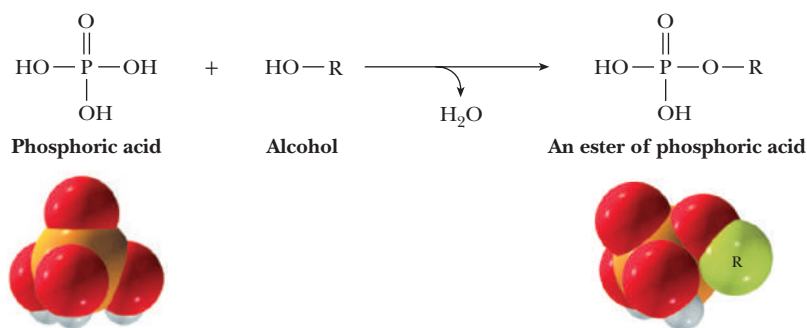
Functional Groups of Biochemical Importance				
Class of Compound	General Structure	Characteristic Functional Group	Name of Functional Group	Example
Alkenes	$\text{RCH}=\text{CH}_2$ $\text{RCH}=\text{CHR}$ $\text{R}_2\text{C}=\text{CHR}$ $\text{R}_2\text{C}=\text{CR}_2$	$\text{C}=\text{C}$	Double bond	$\text{CH}_2=\text{CH}_2$
Alcohols	$\text{ROH}$	$-\text{OH}$	Hydroxyl group	$\text{CH}_3\text{CH}_2\text{OH}$
Ethers	$\text{ROR}$	$-\text{O}-$	Ether group	$\text{CH}_3\text{OCH}_3$
Amines	$\text{RNH}_2$ $\text{R}_2\text{NH}$ $\text{R}_3\text{N}$	$-\text{N} \begin{smallmatrix} \diagup \\ \diagdown \end{smallmatrix}$	Amino group	$\text{CH}_3\text{NH}_2$
Thiols	$\text{RSH}$	$-\text{SH}$	Sulfhydryl group	$\text{CH}_3\text{SH}$
Aldehydes	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{H} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}- \end{array}$	Carbonyl group	$\begin{array}{c} \text{O} \\ \parallel \\ \text{CH}_3\text{CH} \end{array}$
Ketones	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{R} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}- \end{array}$	Carbonyl group	$\begin{array}{c} \text{O} \\ \parallel \\ \text{CH}_3\text{CCH}_3 \end{array}$
Carboxylic acids	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{OH} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}-\text{OH} \end{array}$	Carboxyl group	$\begin{array}{c} \text{O} \\ \parallel \\ \text{CH}_3\text{COH} \end{array}$
Esters	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{OR} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}-\text{OR} \end{array}$	Ester group	$\begin{array}{c} \text{O} \\ \parallel \\ \text{CH}_3\text{COCH}_3 \end{array}$
Amides	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{NR}_2 \\ \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{NHR} \\ \text{O} \\ \parallel \\ \text{R}-\text{C}-\text{NH}_2 \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{C}-\text{N} \begin{smallmatrix} \diagup \\ \diagdown \end{smallmatrix} \end{array}$	Amide group	$\begin{array}{c} \text{O} \\ \parallel \\ \text{CH}_3\text{CN}(\text{CH}_3)_2 \end{array}$
Phosphoric acid esters	$\begin{array}{c} \text{O} \\ \parallel \\ \text{R}-\text{O}-\text{P}-\text{OH} \\   \\ \text{OH} \end{array}$	$\begin{array}{c} \text{O} \\ \parallel \\ -\text{O}-\text{P}-\text{OH} \\   \\ \text{OH} \end{array}$	Phosphoric ester group	$\begin{array}{c} \text{O} \\ \parallel \\ \text{CH}_3-\text{O}-\text{P}-\text{OH} \\   \\ \text{OH} \end{array}$
Phosphoric acid anhydrides	$\begin{array}{c} \text{O} \qquad \text{O} \\ \parallel \qquad \parallel \\ \text{R}-\text{O}-\text{P}-\text{O}-\text{P}-\text{OH} \\   \qquad   \\ \text{OH} \qquad \text{OH} \end{array}$	$\begin{array}{c} \text{O} \qquad \text{O} \\ \parallel \qquad \parallel \\ -\text{P}-\text{O}-\text{P}- \\   \qquad   \\ \text{OH} \qquad \text{OH} \end{array}$	Phosphoric anhydride group	$\begin{array}{c} \text{O} \qquad \text{O} \\ \parallel \qquad \parallel \\ \text{HO}-\text{P}-\text{O}-\text{P}-\text{OH} \\   \qquad   \\ \text{OH} \qquad \text{OH} \end{array}$

1.3 The Beginnings of Biology: Origin of Life

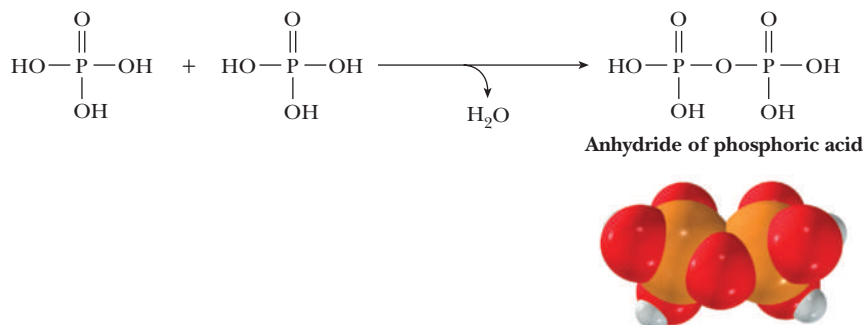
The Earth and Its Age

To date, we are aware of only one planet that unequivocally supports life: our own. The Earth and its waters are universally understood to be the source and mainstay of life as we know it. A natural first question is how the Earth, along with the Universe of which it is a part, came to be.

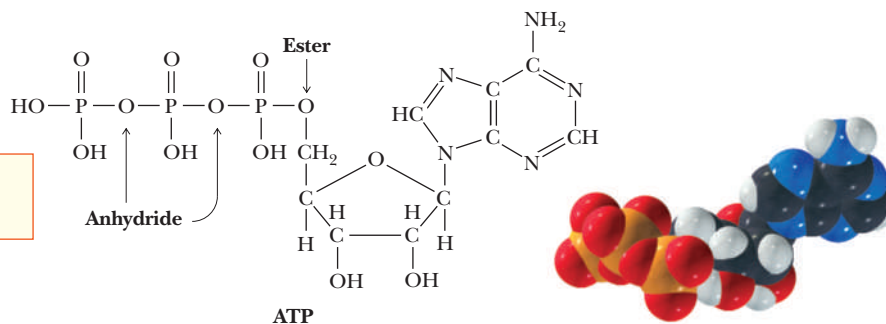
- 1 Reaction of phosphoric acid with a hydroxyl group to form an ester, which contains a P-O-R linkage. Phosphoric acid is shown in its nonionized form in this figure. Space-filling models of phosphoric acid and its methyl ester are shown. The red spheres represent oxygen; the white, hydrogen; the green, carbon; and the orange, phosphorus.



- 2 Reaction of two molecules of phosphoric acid to form an anhydride, which contains a P-O-P linkage. A space-filling model of the anhydride of phosphoric acid is shown.



- 3 The structure of ATP (adenosine triphosphate), showing two anhydride linkages and one ester.



■ **FIGURE 1.2** ATP and the reactions for its formation.

### How and when did the Earth come to be?

Currently, the most widely accepted cosmological theory for the origin of the Universe is the *big bang*, a cataclysmic explosion. According to big-bang cosmology, all the matter in the Universe was originally confined to a comparatively small volume of space. As a result of a tremendous explosion, this “primordial fireball” started to expand with great force. Immediately after the big bang, the Universe was extremely hot, on the order of 15 billion ( $15 \times 10^9$ ) K. (Note that Kelvin temperatures are written without a degree symbol.) The average temperature of the Universe has been decreasing ever since as a result of expansion, and the lower temperatures have permitted the formation of stars and planets. In its earliest stages, the Universe had a fairly simple composition. Hydrogen, helium, and some lithium (the three smallest and simplest elements on the periodic table) were present, having been formed in the original big-bang explosion. The rest of the chemical elements are thought to have been formed in three ways: (1) by thermonuclear reactions that normally take place in stars, (2) in explosions of stars, and (3) by the action of cosmic rays outside the stars since the formation of the galaxy. The process by which the elements are formed in stars is a topic of interest to chemists as well as to astrophysicists. For our purposes, note that the most abundant isotopes of biologically important

TABLE 1.2

Abundance of Important Elements Relative to Carbon*		
Element	Abundance in Organisms	Abundance in Universe
Hydrogen	80–250	10,000,000
Carbon	1,000	1,000
Nitrogen	60–300	1,600
Oxygen	500–800	5,000
Sodium	10–20	12
Magnesium	2–8	200
Phosphorus	8–50	3
Sulfur	4–20	80
Potassium	6–40	0.6
Calcium	25–50	10
Manganese	0.25–0.8	1.6
Iron	0.25–0.8	100
Zinc	0.1–0.4	0.12

\* Each abundance is given as the number of atoms relative to a thousand atoms of carbon.

elements such as carbon, oxygen, nitrogen, phosphorus, and sulfur have *particularly stable nuclei*. These elements were produced by nuclear reactions in first-generation stars, the original stars produced after the beginning of the Universe (Table 1.2). Many first-generation stars were destroyed by explosions called *supernovas*, and their stellar material was recycled to produce second-generation stars, such as our own Sun, along with our solar system. Radioactive dating, which uses the decay of unstable nuclei, indicates that the age of the Earth (and the rest of the solar system) is 4 billion to 5 billion ( $4 \times 10^9$  to  $5 \times 10^9$ ) years. The atmosphere of the early Earth was very different from the one we live in, and it probably went through several stages before reaching its current composition. The most important difference is that, according to most theories of the origins of the Earth, very little or no free oxygen ( $O_2$ ) existed in the early stages (Figure 1.3). The early Earth was constantly irradiated with ultraviolet light from the Sun because there was no ozone ( $O_3$ ) layer in the atmosphere to block it. Under these conditions, the chemical reactions that produced simple biomolecules took place.

The gases usually postulated to have been present in the atmosphere of the early Earth include  $NH_3$ ,  $H_2S$ ,  $CO$ ,  $CO_2$ ,  $CH_4$ ,  $N_2$ ,  $H_2$ , and (in both liquid and vapor forms)  $H_2O$ . However, there is no universal agreement on the relative amounts of these components, from which biomolecules ultimately arose. Many of the earlier theories of the origin of life postulated  $CH_4$  as the carbon source, but more recent studies have shown that appreciable amounts of  $CO_2$  must have existed in the atmosphere at least 3.8 billion ( $3.8 \times 10^9$ ) years ago.

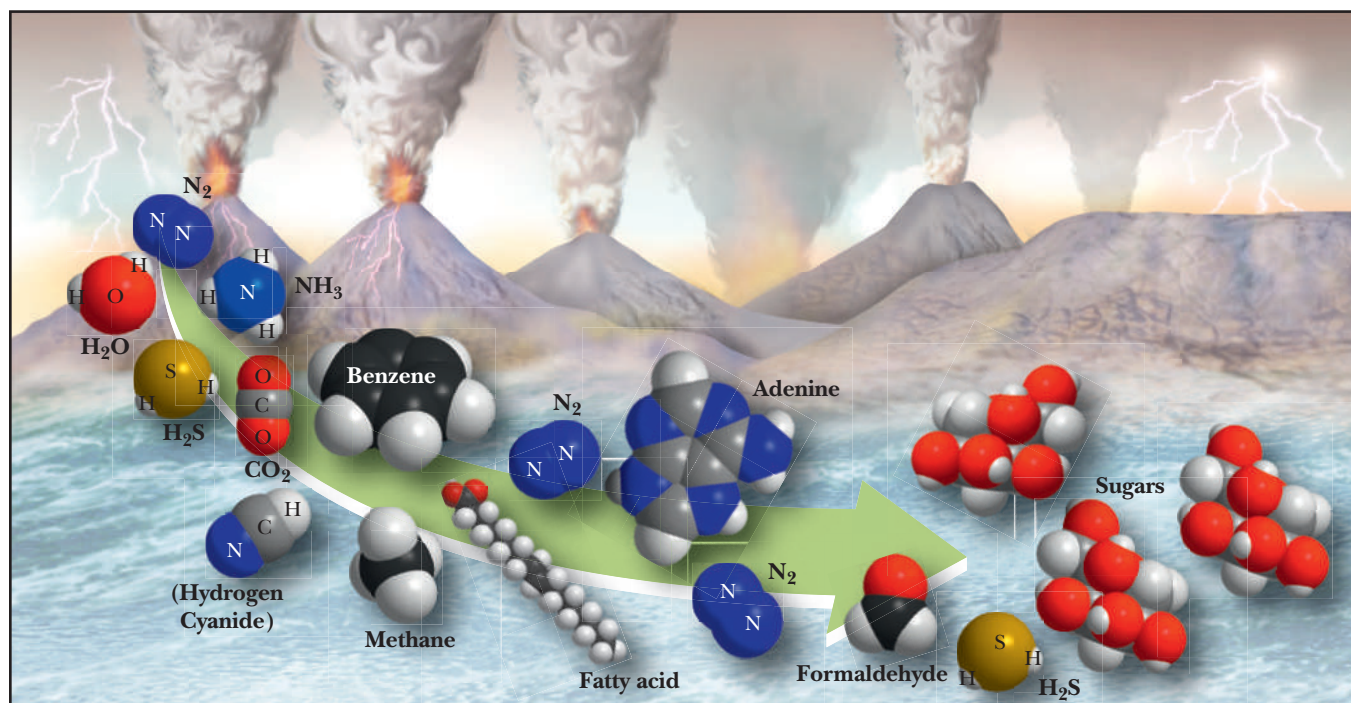
This conclusion is based on geological evidence: The earliest known rocks are 3.8 billion years old, and they are carbonates, which arise from  $CO_2$ . Any  $NH_3$  originally present must have dissolved in the oceans, leaving  $N_2$  in the atmosphere as the nitrogen source required for the formation of proteins and nucleic acids.

Biomolecules

How were biomolecules likely to have formed on the early Earth?

Experiments have been performed in which the simple compounds of the early atmosphere were allowed to react under the varied sets of conditions that might have been present on the early Earth. The results of such experiments indicate that these simple compounds react *abiotically* or, as the word indicates (*a*, “not,”



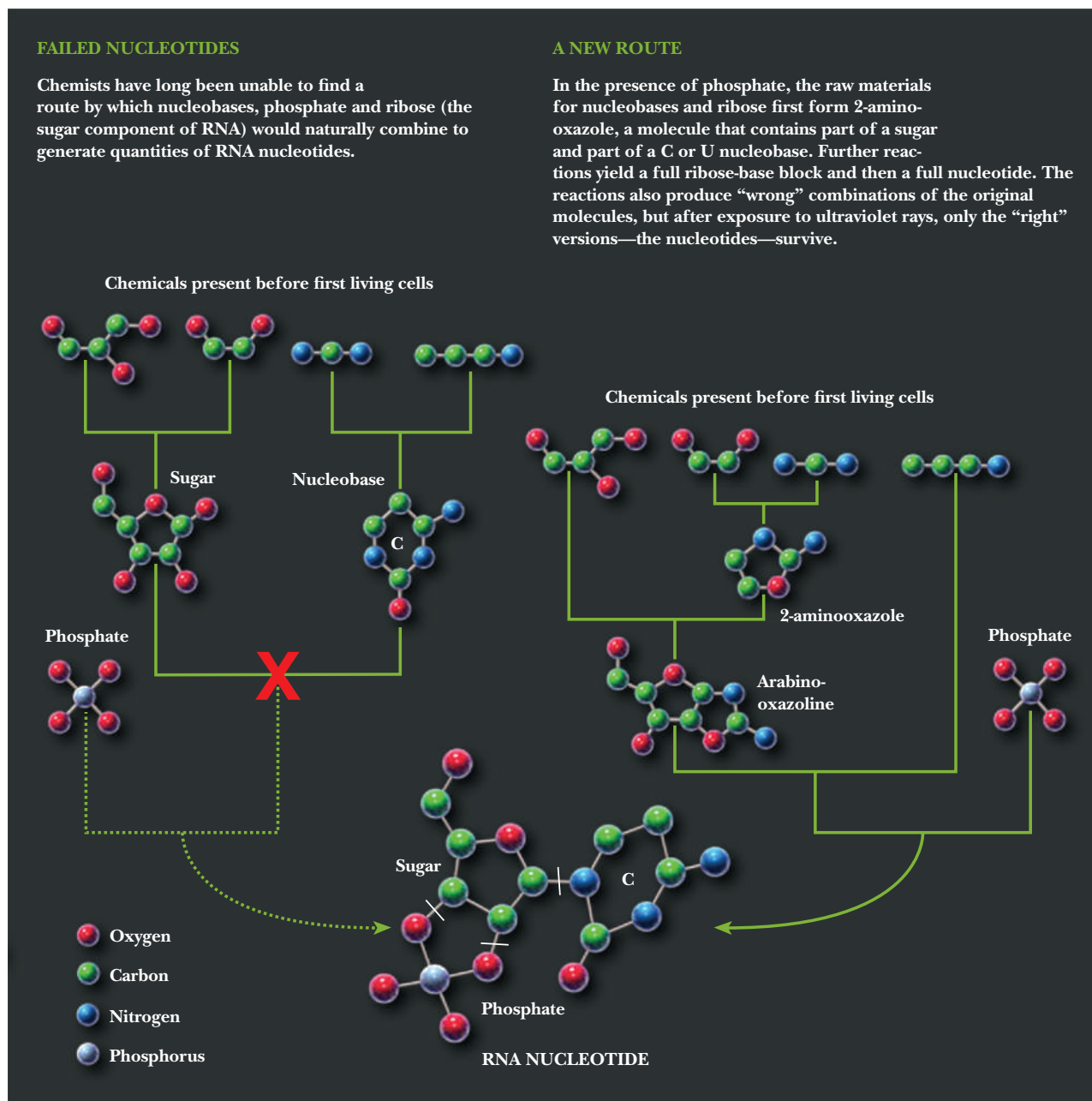


■ **FIGURE 1.3** Formation of biomolecules on the early Earth. Conditions on early Earth would have been inhospitable for most of today's life. Very little or no oxygen ( $O_2$ ) existed. Volcanoes erupted, spewing gases, and violent thunderstorms produced torrential rainfall that covered the Earth. The green arrow indicates the formation of biomolecules from simple precursors.

and *bios*, “life”), in the absence of life, to give rise to biologically important compounds such as the components of proteins and nucleic acids. Of historic interest is the well-known Miller–Urey experiment. In each trial, an electric discharge, simulating lightning, is passed through a closed system that contains  $H_2$ ,  $CH_4$ , and  $NH_3$ , in addition to  $H_2O$ . Simple organic molecules, such as formaldehyde ( $HCHO$ ) and hydrogen cyanide ( $HCN$ ), are typical products of such reactions, as are amino acids, the building blocks of proteins. According to one theory, reactions such as these took place in the Earth's early oceans; other researchers postulate that such reactions occurred on the surfaces of clay particles that were present on the early Earth. It is certainly true that mineral substances similar to clay can serve as catalysts in many types of reactions. Both theories have their proponents, and more research is needed to answer the many questions that remain. More to the point, recent theories of the origin of life focus of RNA, not proteins, as the first genetic molecules. Proteins are thought to have developed later in the evolution of the earliest cells. This point does not diminish the importance of this first experiment on abiotic synthesis of biomolecules.

Recent experiments have shown it is possible to synthesize nucleotides from simple molecules by a pathway that includes a precursor that is neither a sugar nor a nucleobase, but a fragment consisting of a sugar and a part of a base. This fragment, 2-aminooxazole, is highly volatile and can vaporize and condense so as to give rise to pockets of pure material in reasonably large amounts. In turn, phosphates released by volcanic action can react with the 2-aminooxazole to produce nucleotides (Figure 1.4). The products include nucleotides that are not part of present-day RNA, but intense ultraviolet light, which was present on the early Earth, destroyed those nucleotides, leaving those found in RNA today.

Living cells today are assemblages that include very large molecules, such as proteins, nucleic acids, and polysaccharides. These molecules are larger by many powers of ten than the smaller molecules from which they are built. Hundreds or thousands of these smaller molecules, or **monomers**, can be linked

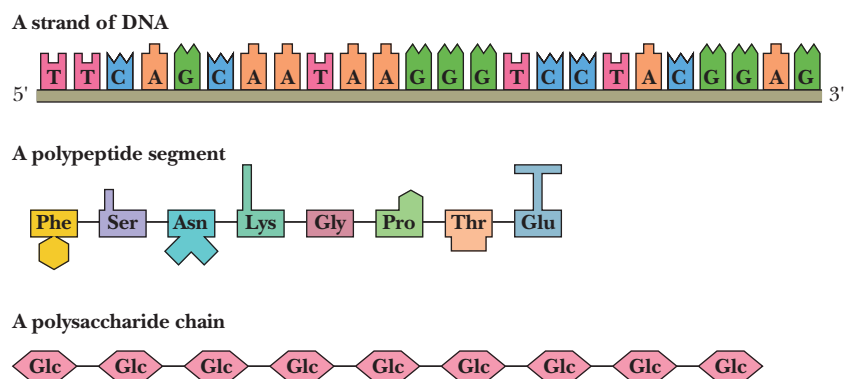


■ **FIGURE 1.4** Abiotic synthesis of nucleotides. The volatile compound 2-amino-oxazole is a key intermediate that eventually gives rise to nucleotides. (Copyright © Andrew Swift)

to produce macromolecules, which are also called **polymers**. The versatility of carbon is important here. Carbon is tetravalent and able to form bonds with itself and with many other elements, giving rise to different kinds of monomers, such as amino acids, nucleotides, and monosaccharides (sugar monomers).

*Proteins and nucleic acids play a key role in life processes.* In present-day cells, amino acids (the monomers) combine by polymerization to form **proteins**, nucleotides (also monomers) combine to form **nucleic acids**, and the polymerization of sugar monomers produces polysaccharides. Polymerization experiments with amino acids carried out under early-Earth conditions have produced proteinlike polymers. Similar experiments have been done on the abiotic polymerization of nucleotides and sugars, which tends to happen less readily than the polymerization of amino acids.



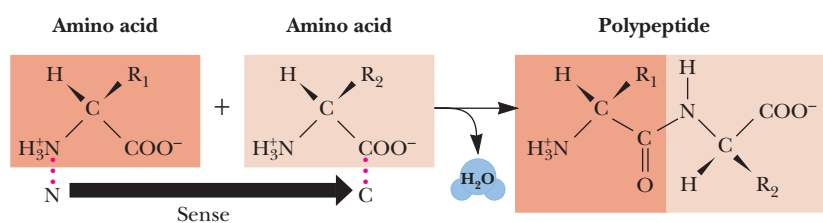


■ **FIGURE 1.5 Informational macromolecules.** Biological macromolecules are informational. The sequence of monomeric units in a biological polymer has the potential to contain information if the order of units is not overly repetitive. Nucleic acids and proteins are informational macromolecules; polysaccharides are not.

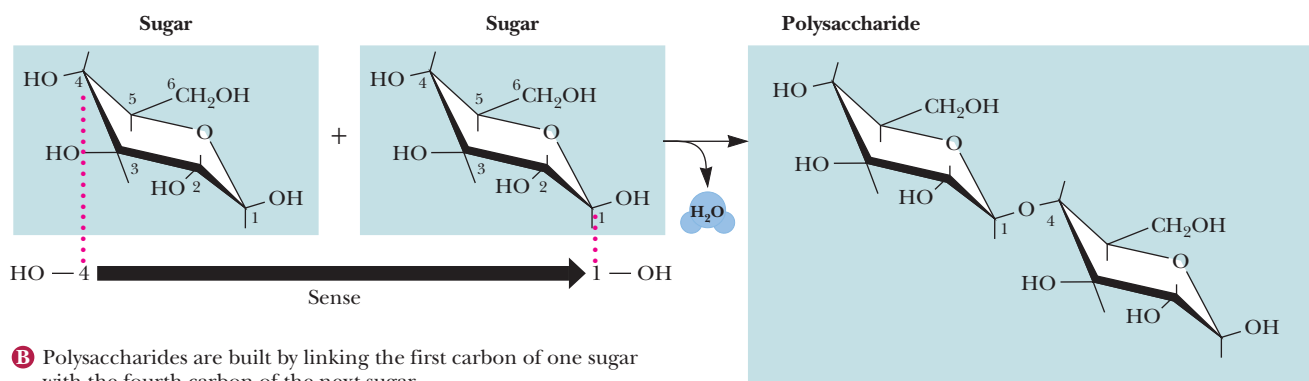
The several types of amino acids and nucleotides can easily be distinguished from one another. When amino acids form polymers, with the loss of water accompanying this spontaneous process, the sequence of amino acids determines the properties of the protein formed. Likewise, the genetic code lies in the sequence of monomeric nucleotides that polymerize to form nucleic acids, the molecules of heredity (Figure 1.5). In polysaccharides, however, the order of monomers rarely has an important effect on the properties of the polymer, nor does the order of the monomers carry any genetic information. (Other aspects of the *linkage* between monomers are important in polysaccharides, as we shall see when we discuss carbohydrates in Chapter 16.) Notice that all the building blocks have a “head” and a “tail,” giving a sense of direction even at the monomer level (Figure 1.6).

The effect of monomer sequence on the properties of polymers can be illustrated by another example. Proteins of the class called *enzymes* display **catalytic activity**, which means that they increase the rates of chemical reactions compared with uncatalyzed reactions. In the context of the origin of life, catalytic molecules can facilitate the production of large numbers of complex molecules, allowing for the accumulation of such molecules. When a large group of related molecules accumulates, a complex system arises with some of the characteristics of living organisms. Such a system has a nonrandom organization, tends to reproduce itself, and competes with other systems for the simple organic molecules present in the environment. One of the most important functions of proteins is **catalysis**, and the catalytic effectiveness of a given enzyme depends on its amino acid sequence. The specific sequence of the amino acids present ultimately determines the properties of all types of proteins, including enzymes. If not for protein catalysis, the chemical reactions that take place in our bodies would be so slow as to be useless for life processes. We are going to have a lot to say about this point in Chapters 6 and 7.

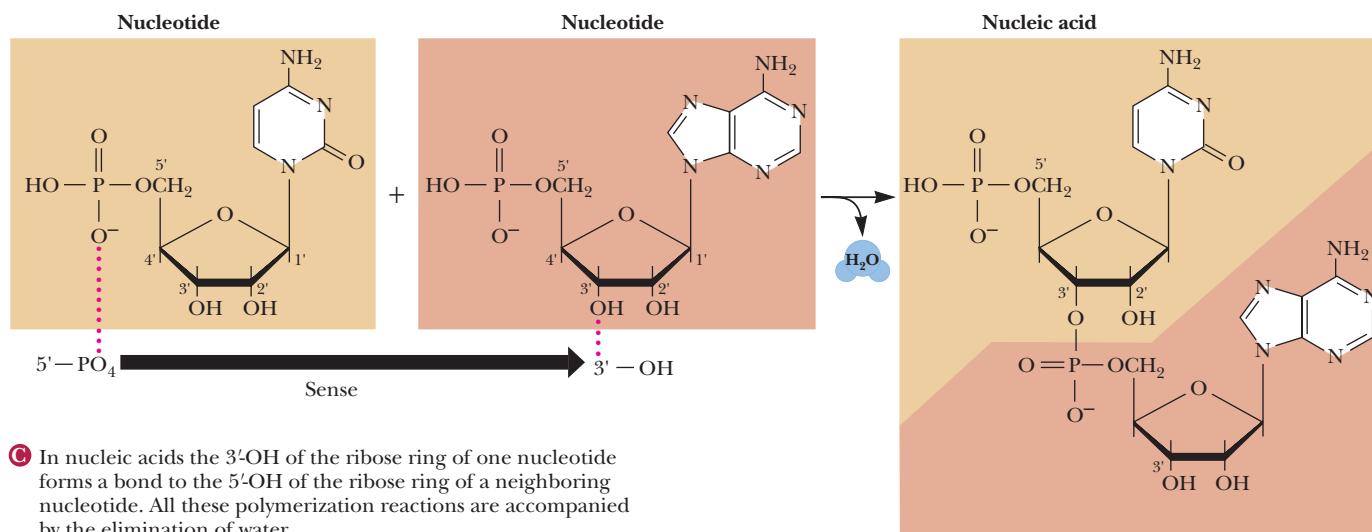
In present-day cells, the sequence of amino acids in proteins is determined by the sequence of nucleotides in nucleic acids. The process by which genetic information is translated into the amino acid sequence is very complex. *DNA* (*deoxyribonucleic acid*), one of the nucleic acids, serves as the coding material. The **genetic code** is the relationship between the nucleotide sequence in nucleic acids and the amino acid sequence in proteins. As a result of this relationship, the information for the structure and function of all living things is passed from one generation to the next. The workings of the genetic code are no longer completely mysterious, but they are far from completely understood. Theories on the origins of life consider how a coding system might have developed, and new insights in this area could shine some light on the present-day genetic code.



**A** Amino acids build proteins by connecting the carboxyl group of one amino acid with the amino group of the next amino acid.



**B** Polysaccharides are built by linking the first carbon of one sugar with the fourth carbon of the next sugar.



**C** In nucleic acids the 3'-OH of the ribose ring of one nucleotide forms a bond to the 5'-OH of the ribose ring of a neighboring nucleotide. All these polymerization reactions are accompanied by the elimination of water.

**FIGURE 1.6 Directionality in macromolecules.** Biological macromolecules and their building blocks have a “sense” or directionality.

## Molecules to Cells

### *Which came first—the catalysts or the hereditary molecules?*

A discovery with profound implications for discussions of the origin of life is that *RNA (ribonucleic acid)*, another nucleic acid, is capable of catalyzing its own processing. Until this discovery, catalytic activity was associated exclusively with proteins. RNA, rather than DNA, is now considered by many scientists to have been the original coding material, and it still serves this function in some viruses. The idea that catalysis and coding both occur in one molecule has provided a point of departure for more research on the origins of life. The “RNA

world” is the current conventional wisdom, but many unanswered questions exist regarding this point of view.

According to the RNA-world theory, the appearance of a form of RNA capable of coding for its own replication was the pivotal point in the origin of life. Polynucleotides can direct the formation of molecules whose sequence is an exact copy of the original. This process depends on a template mechanism (Figure 1.7), which is highly effective in producing exact copies but is a relatively slow process. A catalyst is required, which can be a polynucleotide, even the original molecule itself. Polypeptides, however, are more efficient catalysts than polynucleotides, but there is still the question of whether they can direct the formation of exact copies of themselves. Recall that in present-day cells, the genetic code is based on nucleic acids, and catalysis relies primarily on proteins. How did nucleic acid synthesis (which requires many protein enzymes) and protein synthesis (which requires the genetic code to specify the order of amino acids) come to be? According to this hypothesis, RNA (or a system of related kinds of RNA) originally played both roles, catalyzing and encoding its own replication. Eventually, the system evolved to the point of being able to encode the synthesis of more effective catalysts, namely proteins (Figure 1.8). Even later, DNA took over as the primary genetic material, relegating the more versatile RNA to an intermediary role in directing the synthesis of proteins under the direction of the genetic code residing in DNA. A certain amount of controversy surrounds this theory, but it has attracted considerable attention recently. Many unanswered questions remain about the role of RNA in the origin of life, but clearly that role must be important.

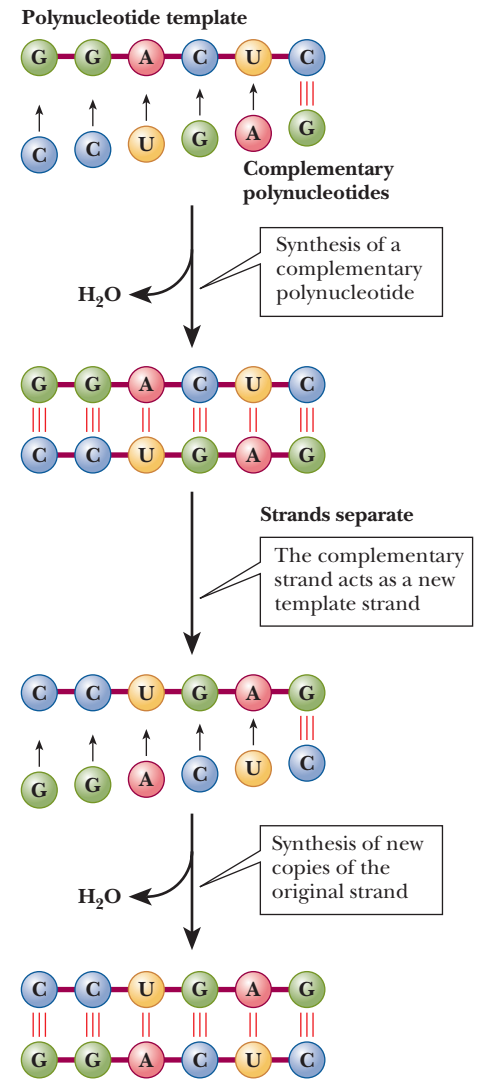
Another key point in the development of living cells is the formation of membranes that separate cells from their environment. The clustering of coding and catalytic molecules in a separate compartment brings molecules into closer contact with each other and excludes extraneous material. For reasons we shall explore in detail in Chapters 2 and 8, lipids are perfectly suited to form cell membranes (Figure 1.9).

Recently, attempts have been made to combine several lines of reasoning about the origin of life into a *double-origin theory*. According to this line of thought, the development of catalysis and the development of a coding system came about separately, and the combination of the two produced life as we know it. The rise of aggregates of molecules capable of catalyzing reactions was one origin of life, and the rise of a nucleic acid-based coding system was another origin.

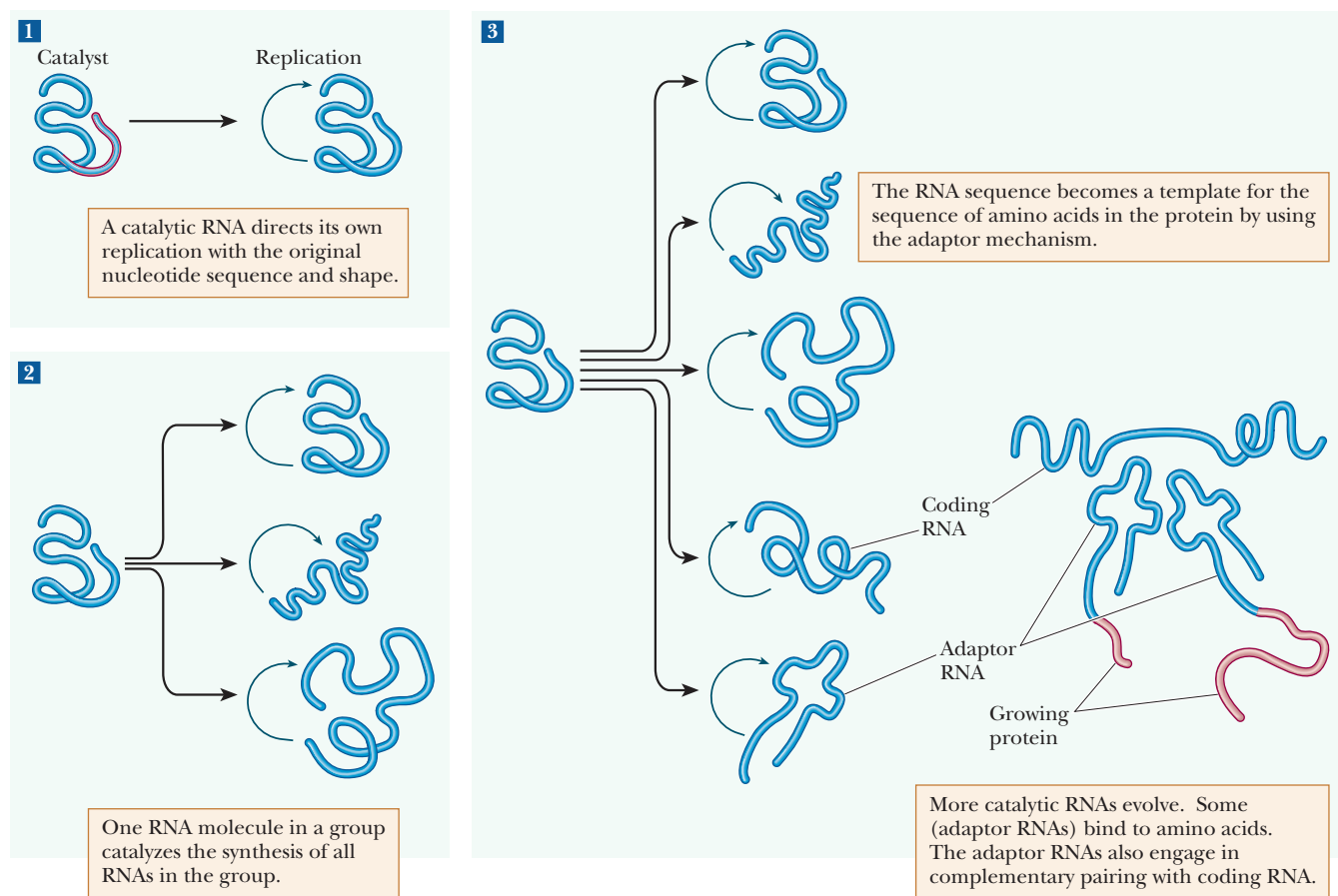
A theory that life began on clay particles is a form of the double-origin theory.

According to this point of view, coding arose first, but the coding material was the surface of naturally occurring clay. The pattern of ions on the clay surface is thought to have served as the code, and the process of crystal growth is thought to have been responsible for replication. Nucleotides, then RNA molecules, formed on the clay surface. The RNA molecules thus formed were released from the clay surface and enclosed in lipid sacs, forming protocells. In this scenario, protocells exist in a pond with a warm side and a cold side. Double-stranded polynucleotides are formed on the cold side of the pond on a single-stranded template (Figure 1.10). The protocell moves to the warm side of the pond, where the strands separate. The membrane incorporates more lipid molecules. The protocell divides, with a single-stranded RNA in each daughter cell, and the cycle repeats.

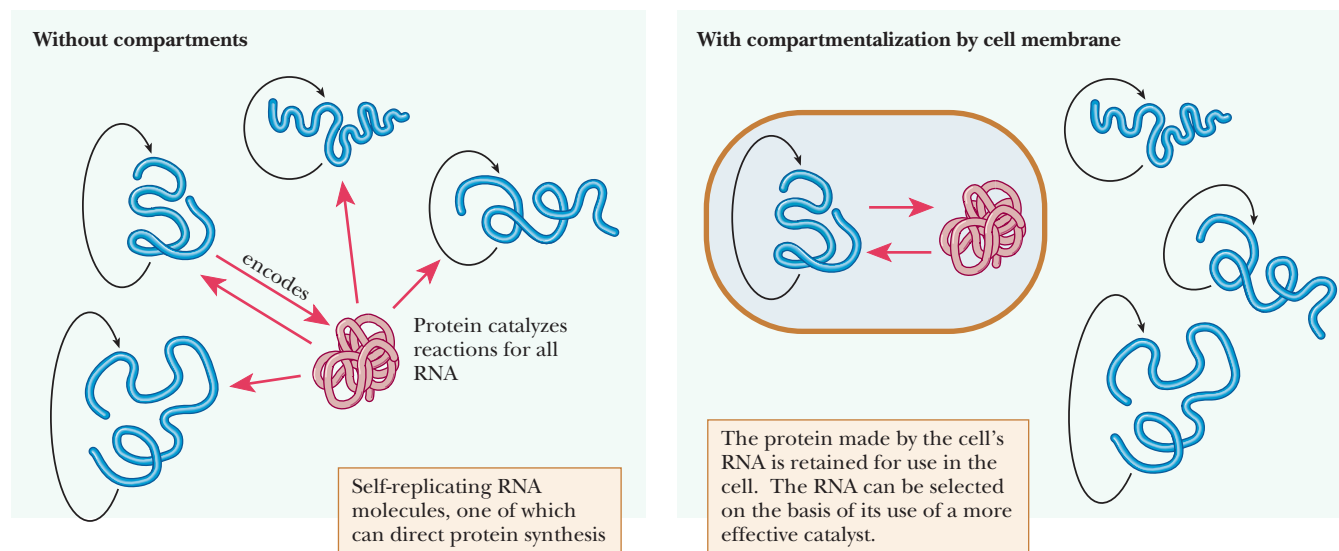
In the development from protocells to single cells similar to modern bacteria, proteins and then DNA enter the picture. In this scenario, ribozymes (catalytic RNA molecules) develop and direct the duplication of RNA. Other ribozymes catalyze metabolic reactions, eventually giving rise to proteins (Figure 1.11). Eventually, proteins rather than ribozymes catalyze most of the reactions in the cell. Still later, other enzymes catalyze the production of DNA, which takes over the primary role in coding. RNA now serves as an intermediary between DNA



**FIGURE 1.7 The role of templates in synthesis of polynucleotides.** Polynucleotides use a template mechanism to produce exact copies of themselves: G pairs with C, and A pairs with U by a relatively weak interaction. The original strand acts as a template to direct the synthesis of a complementary strand. The complementary strand then acts as a template for the production of copies of the original strand. Note that the original strand can be a template for a number of complementary strands, each of which in turn can produce a number of copies of the original strand. This process gives rise to a many-fold amplification of the original sequence. (Copyright © 1994 from *The Molecular Biology of the Cell*, 3rd Edition by A. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. Reproduced by permission of Garland Science/Taylor & Francis Books, Inc.)

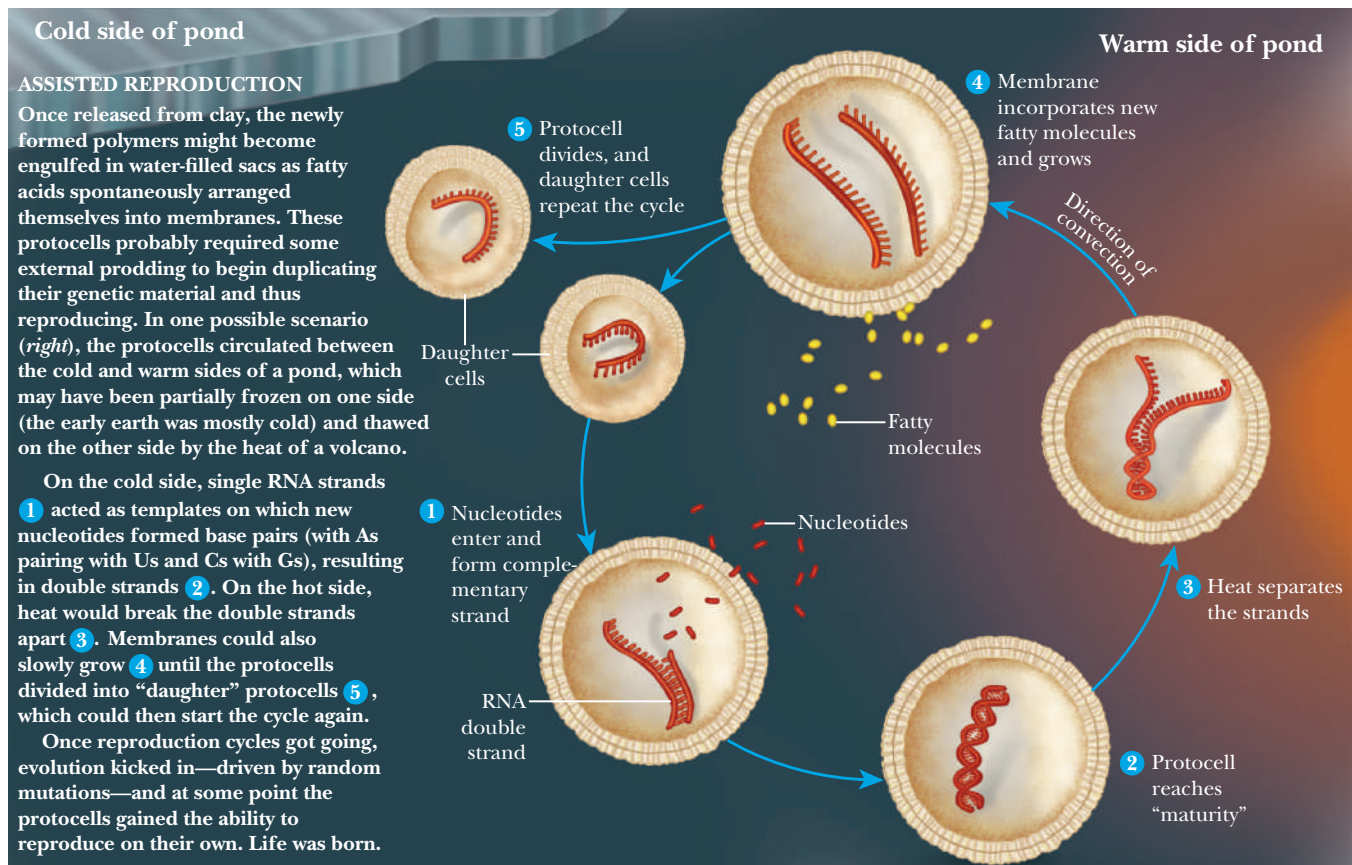


■ **FIGURE 1.8** Stages in the evolution of a system of self-replicating RNA molecules. At each stage, more complexity appears in the group of RNAs, leading eventually to the synthesis of proteins as more effective catalysts. (Copyright © 1994 from *The Molecular Biology of the Cell*, 3rd Edition by A. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. Reproduced by permission of Garland Science/Taylor & Francis Books, Inc.)



■ **FIGURE 1.9** The vital importance of a cell membrane in the origin of life. Without compartments, groups of RNA molecules must compete with others in their environment for the proteins they synthesize. With compartments, the RNAs have exclusive access to the more effective catalysts and are closer to each other, making it easier for reactions to take place. (Copyright © 1994 from *The Molecular Biology of the Cell*, 3rd Edition by A. Alberts, D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. Reproduced by permission of Garland Science/Taylor & Francis Books, Inc.)





■ **FIGURE 1.10 Replication and reproduction.** When RNA polymers first formed, they could replicate on the cold side of a pond, producing double-stranded RNA. The strands separate on the warm side and divide into new protocells. (Reprinted with permission. Copyright © 2009 Scientific American, a division of Nature America, Inc. All rights reserved.)

and proteins. This scenario assumes that time is not a limiting factor in the process. In an attempt to study the origins of life, scientists have also attempted to combine the best properties of proteins and nucleic acids and have created Peptide Nucleic Acids, PNA. Evidence shows that the building blocks of these hybrids could also have formed in the primordial world, and some theorize that PNA may have been the original molecule that allowed life to form. Currently scientists are attempting to create artificial living cells based on PNA. The goal is to demonstrate that under the conditions of the “primordial soup” simple molecules could form complex molecules possessing the critical functions of catalysis and replication, and that these could then form cells capable of dividing.

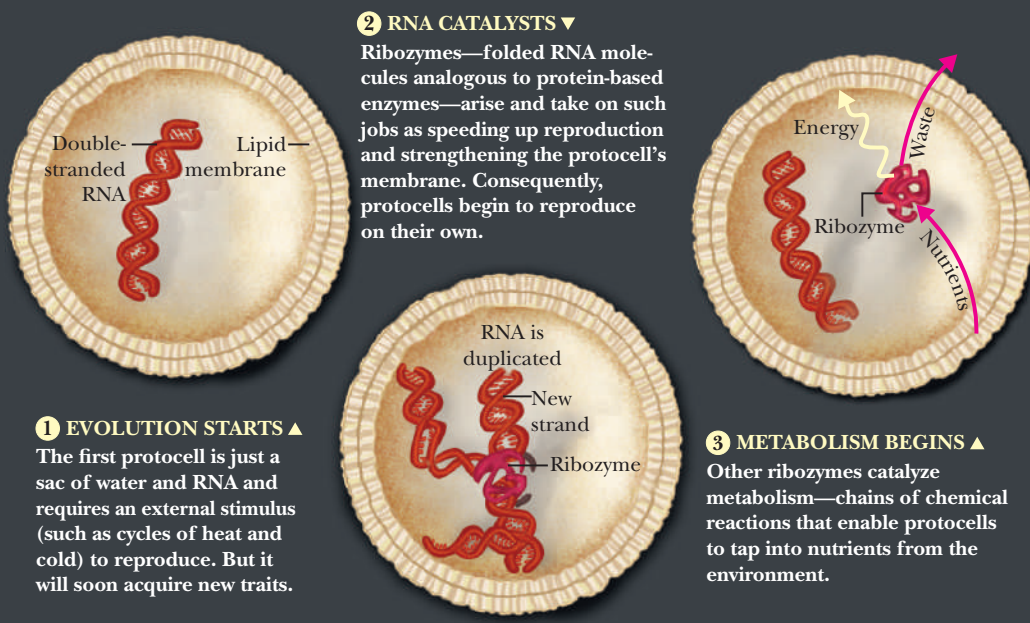
At this writing, none of the theories of the origin of life is definitely established, and none is definitely disproved. The topic is still under active investigation. It seems highly unlikely that we will ever know with certainty how life originated on this planet, but these conjectures allow us to ask some of the important questions, such as those about catalysis and coding, that we are going to see many times in this text.

## 1.4 The Biggest Biological Distinction—Prokaryotes and Eukaryotes

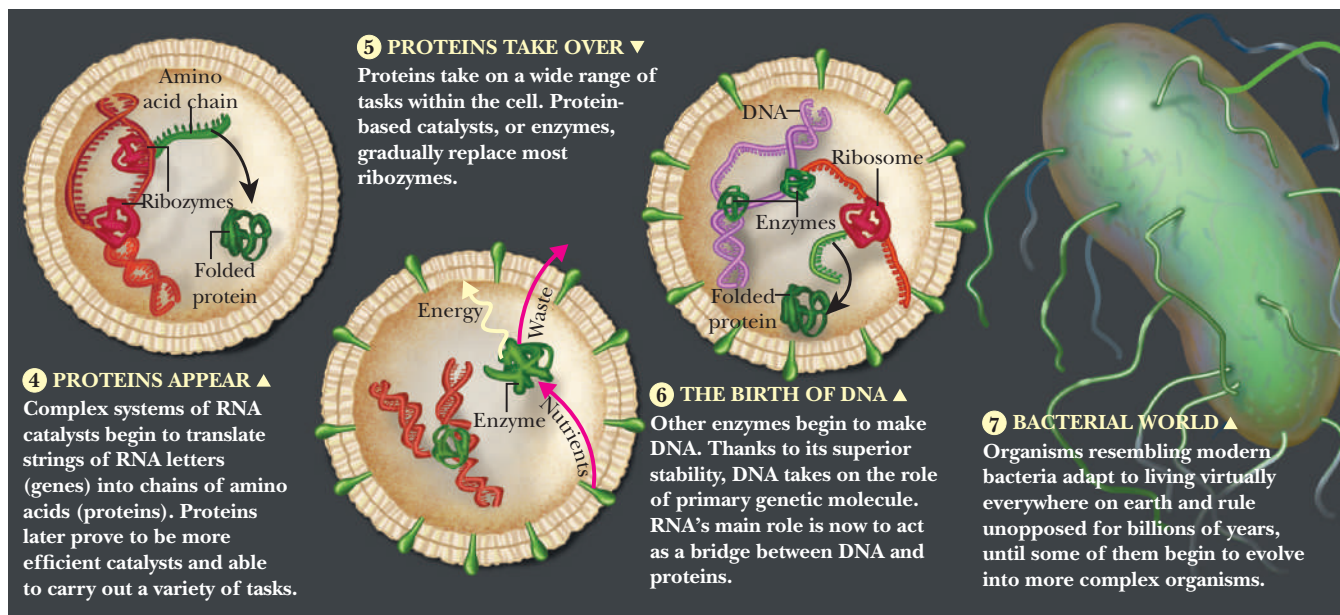
*All cells contain DNA.* The total DNA of a cell is called the **genome**. Individual units of heredity, controlling individual traits by coding for a functional protein or RNA, are **genes**.

## Journey to the Modern Cell

After life got started, competition among life-forms fueled the drive toward ever more complex organisms. We may never know the exact details of early evolution, but here is a plausible sequence of some of the major events that led from the first protocell to DNA-based cells such as bacteria.



A



B

■ **FIGURE 1.11** From membrane-coated RNA to bacteria. (A) Ribozymes start to catalyze a number of reactions, giving rise to metabolism. (B) Proteins eventually take over most of catalysis. DNA becomes the primary coding molecule. (Reprinted with permission. Copyright © 2009 Scientific American, a division of Nature America, Inc. All rights reserved.)

The earliest cells that evolved must have been very simple, having the minimum apparatus necessary for life processes. The types of organisms living today that probably most resemble the earliest cells are the **prokaryotes**. This word, of Greek derivation (*karyon*, “kernel, nut”), literally means “before the nucleus.” Prokaryotes include *bacteria* and *cyanobacteria*. (Cyanobacteria were formerly called blue-green algae; as the newer name indicates, they are more closely related to bacteria.) Prokaryotes are single-celled organisms, but groups of them can exist in association, forming colonies with some differentiation of cellular functions.



### What is the difference between a prokaryote and a eukaryote?

The word *eukaryote* means “true nucleus.” **Eukaryotes** are more complex organisms and can be multicellular or single-celled. A well-defined nucleus, set off from the rest of the cell by a membrane, is one of the chief features distinguishing a eukaryote from a prokaryote. A growing body of fossil evidence indicates that eukaryotes evolved from prokaryotes about 1.5 billion ( $1.5 \times 10^9$ ) years ago, about 2 billion years after life first appeared on Earth. Examples of single-celled eukaryotes include yeasts and *Paramecium* (an organism frequently discussed in beginning biology courses); all multicellular organisms (e.g., animals and plants) are eukaryotes. As might be expected, eukaryotic cells are more complex and usually much larger than prokaryotic cells. The diameter of a typical prokaryotic cell is on the order of 1 to 3  $\mu\text{m}$  ( $1 \times 10^{-6}$  to  $3 \times 10^{-6}$  m), whereas that of a typical eukaryotic cell is about 10 to 100  $\mu\text{m}$ . The distinction between prokaryotes and eukaryotes is so basic that it is now a key point in the classification of living organisms; it is far more important than the distinction between plants and animals.

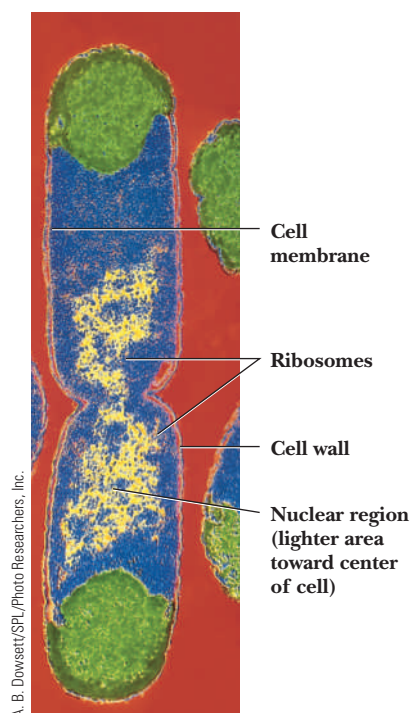
The main difference between prokaryotic and eukaryotic cells is the existence of organelles, especially the nucleus, in eukaryotes. An **organelle** is a part of the cell that has a distinct function; it is surrounded by its own membrane within the cell. In contrast, the structure of a prokaryotic cell is relatively simple, lacking membrane-enclosed organelles. Like a eukaryotic cell, however, a prokaryotic cell has a cell membrane, or plasma membrane, separating it from the outside world. The plasma membrane is the only membrane found in the prokaryotic cell. In both prokaryotes and eukaryotes, the cell membrane consists of a double layer (bilayer) of lipid molecules with a variety of proteins embedded in it.

Organelles have specific functions. A typical eukaryotic cell has a *nucleus* with a nuclear membrane. *Mitochondria* (respiratory organelles) and an internal membrane system known as the *endoplasmic reticulum* are also common to all eukaryotic cells. Energy-yielding oxidation reactions take place in eukaryotic mitochondria. In prokaryotes, similar reactions occur on the plasma membrane. *Ribosomes* (particles consisting of RNA and protein), which are the sites of protein synthesis in all living organisms, are frequently bound to the endoplasmic reticulum in eukaryotes. In prokaryotes, ribosomes are found free in the cytosol. A distinction can be made between the cytoplasm and the cytosol. *Cytoplasm* refers to the portion of the cell outside the nucleus, and the *cytosol* is the aqueous portion of the cell that lies outside the membrane-bounded organelles. *Chloroplasts*, organelles in which photosynthesis takes place, are found in plant cells and green algae. In prokaryotes that are capable of photosynthesis, the reactions take place in layers called *chromatophores*, which are extensions of the plasma membrane, rather than in chloroplasts.

Table 1.3 summarizes the basic differences between prokaryotic and eukaryotic cells.

**TABLE 1.3**

A Comparison of Prokaryotes and Eukaryotes		
Organelle	Prokaryotes	Eukaryotes
Nucleus	No definite nucleus; DNA present but not separate from rest of cell	Present
Cell membrane (plasma membrane)	Present	Present
Mitochondria	None; enzymes for oxidation reactions located on plasma membrane	Present
Endoplasmic reticulum	None	Present
Ribosomes	Present	Present
Chloroplasts	None; photosynthesis (if present) is localized in chromatophores	Present in green plants



■ **FIGURE 1.12** Electron micrograph of a **bacterium**. A colored electron microscope image of a typical prokaryote: the bacterium *Escherichia coli* (magnified 16,500 $\times$ ). The pair in the center shows that division into two cells is nearly complete.

## 1.5 Prokaryotic Cells

Although no well-defined nucleus is present in prokaryotes, the DNA of the cell is concentrated in one region called the **nuclear region**. This part of the cell directs the workings of the cell, much as the eukaryotic nucleus does.

### How is prokaryotic DNA organized without a nucleus?

The DNA of prokaryotes is not complexed with proteins in extensive arrays with specified architecture, as is the DNA of eukaryotes. In general, there is only a single, closed, circular molecule of DNA in prokaryotes. This circle of DNA, which is the genome, is attached to the cell membrane. Before a prokaryotic cell divides, the DNA replicates itself, and both DNA circles are bound to the plasma membrane. The cell then divides, and each of the two daughter cells receives one copy of the DNA (Figure 1.12).

In a prokaryotic cell, the cytosol (the fluid portion of the cell outside the nuclear region) frequently has a slightly granular appearance because of the presence of **ribosomes**. Because these consist of RNA and protein, they are also called *ribonucleoprotein particles*; they are the sites of protein synthesis in all organisms. The presence of ribosomes is the main visible feature of prokaryotic cytosol. (Membrane-bound organelles, characteristic of eukaryotes, are not found in prokaryotes.)

Every cell is separated from the outside world by a **cell membrane**, or plasma membrane, an assemblage of lipid molecules and proteins. In addition to the cell membrane and external to it, a prokaryotic bacterial cell has a **cell wall**, which is made up mostly of polysaccharide material, a feature it shares with eukaryotic plant cells. The chemical natures of prokaryotic and eukaryotic cell walls differ somewhat, but a common feature is that the polymerization of sugars produces the polysaccharides found in both. Because the cell wall is made up of rigid material, it presumably serves as protection for the cell.

## 1.6 Eukaryotic Cells

Multicellular plants and animals are eukaryotes, as are protista and fungi, but obvious differences exist among them. These differences are reflected on the cellular level. One of the biggest differences between eukaryotes and prokaryotes is the presence of subcellular organelles.

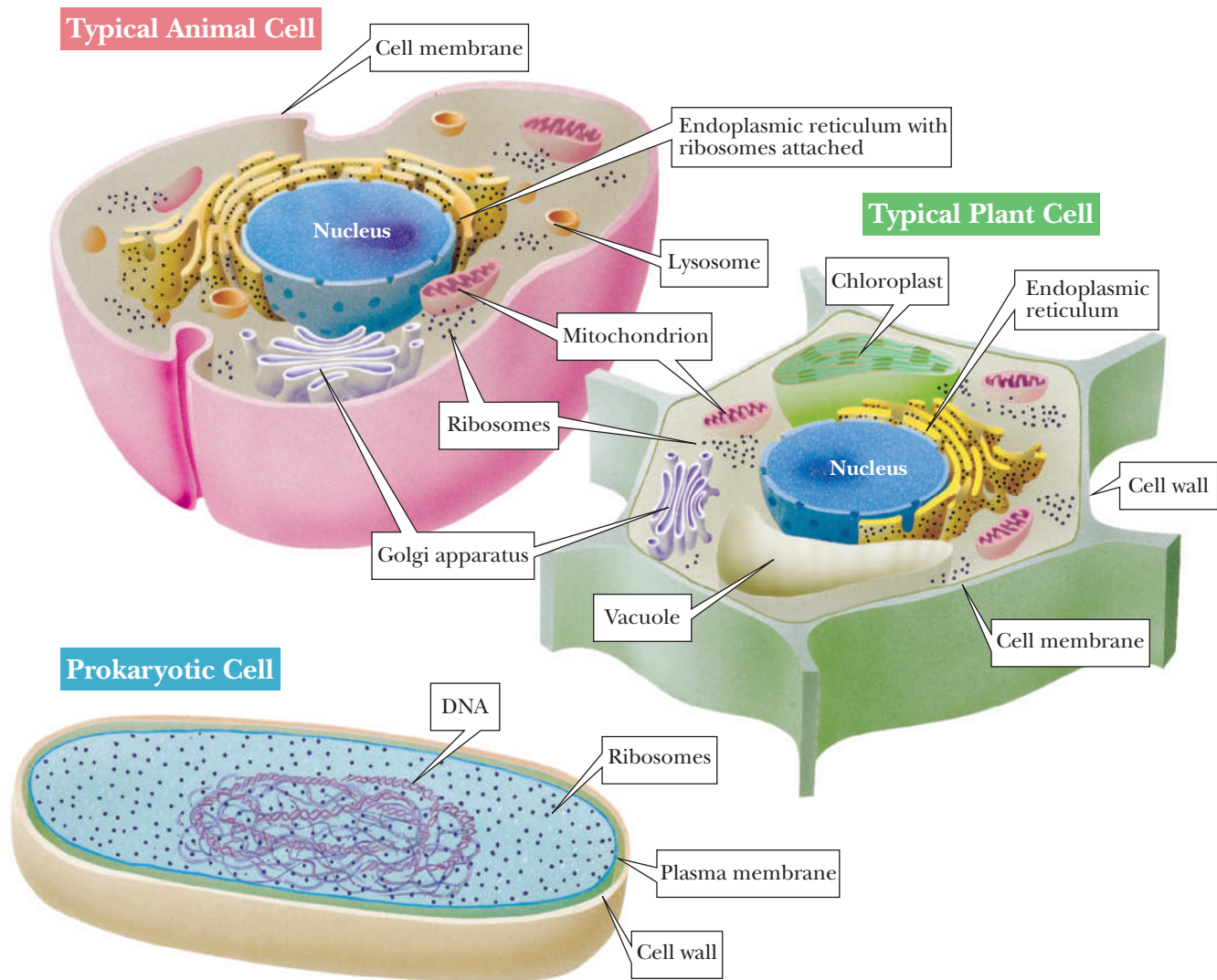
Three of the most important organelles in eukaryotic cells are the nucleus, the mitochondrion, and the chloroplast. Each is separated from the rest of the cell by a double membrane. The nucleus contains most of the DNA of the cell and is the site of RNA synthesis. The mitochondria contain enzymes that catalyze important energy-yielding reactions.

Chloroplasts, which are found in green plants and green algae, are the sites of photosynthesis. Both mitochondria and chloroplasts contain DNA that differs from that found in the nucleus, and both carry out transcription and protein synthesis distinct from that directed by the nucleus.

Plant cells, like bacteria, have cell walls. A plant cell wall is mostly made up of the polysaccharide cellulose, giving the cell its shape and mechanical stability. **Chloroplasts**, the photosynthetic organelles, are found in green plants and algae. Animal cells have neither cell walls nor chloroplasts; the same is true of some protists. Figure 1.13 shows some of the important differences between typical plant cells, typical animal cells, and prokaryotes.

### What are the most important organelles?

The **nucleus** is perhaps the most important eukaryotic organelle. A typical nucleus exhibits several important structural features (Figure 1.14). It is

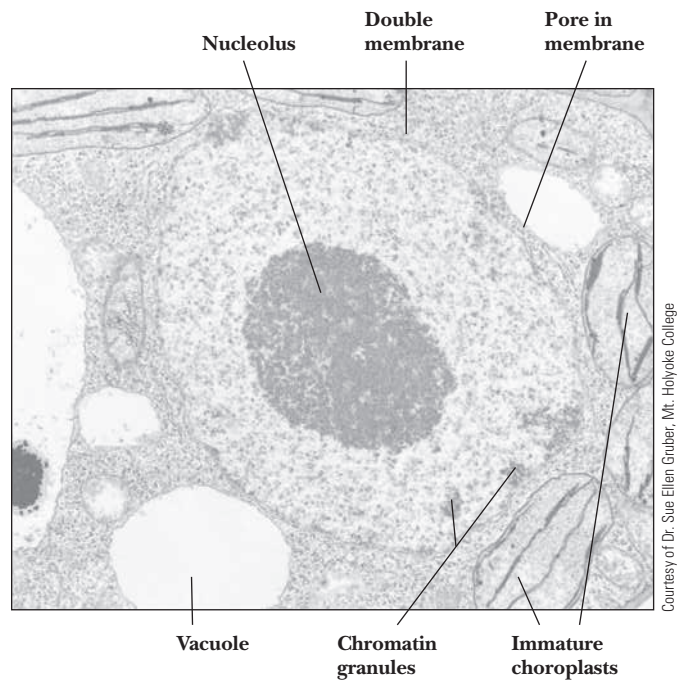


■ **FIGURE 1.13** A comparison of a typical animal cell, a typical plant cell, and a prokaryotic cell.

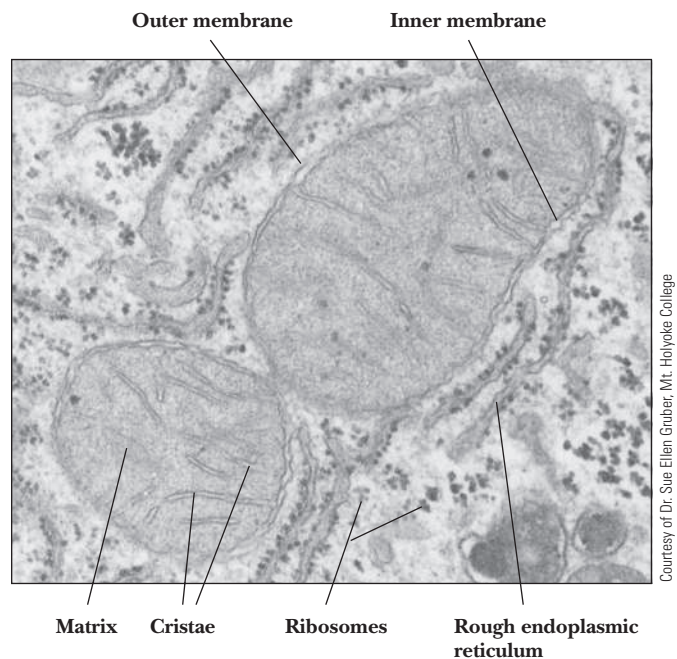
surrounded by a *nuclear double membrane* (usually called the nuclear envelope). One of its prominent features is the **nucleolus**, which is rich in RNA. The RNA of a cell (with the exception of the small amount produced in such organelles as mitochondria and chloroplasts) is synthesized on a DNA template in the nucleolus for export to the cytoplasm through pores in the nuclear membrane. This RNA is ultimately destined for the ribosomes. Also visible in the nucleus, frequently near the nuclear membrane, is **chromatin**, an aggregate of DNA and protein. The main eukaryotic genome (its nuclear DNA) is duplicated before cell division takes place, as in prokaryotes. In eukaryotes, both copies of DNA, which are to be equally distributed between the daughter cells, are associated with protein. When a cell is about to divide, the loosely organized strands of chromatin become tightly coiled, and the resulting **chromosomes** can be seen under a microscope. The genes, responsible for the transmission of inherited traits, are part of the DNA found in each chromosome.

A second very important eukaryotic organelle is the **mitochondrion**, which, like the nucleus, has a double membrane (Figure 1.15). The outer membrane has a fairly smooth surface, but the inner membrane exhibits many folds called **cristae**. The space within the inner membrane is called the **matrix**. Oxidation processes that occur in mitochondria yield energy for the cell. Most of the





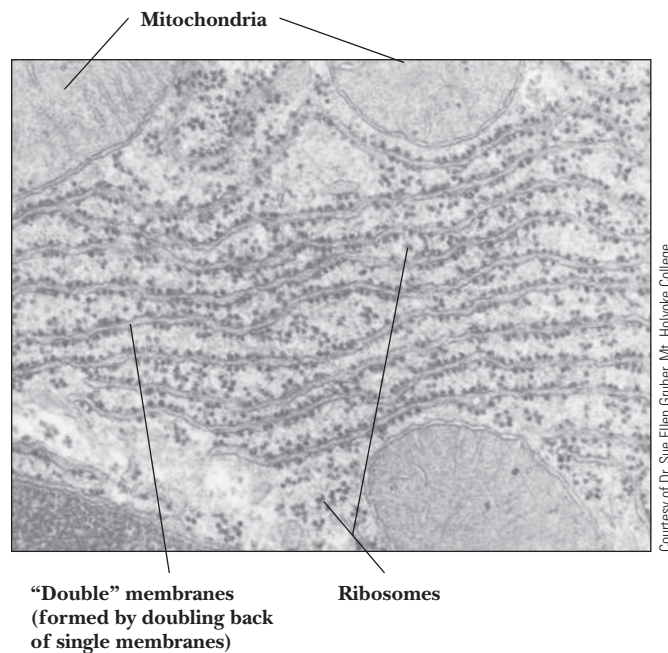
■ **FIGURE 1.14** The nucleus of a tobacco leaf cell (magnified 15,000 $\times$ ).



■ **FIGURE 1.15** Mouse liver mitochondria (magnified 50,000 $\times$ ).

enzymes responsible for these important reactions are associated with the inner mitochondrial membrane. Other enzymes needed for oxidation reactions, as well as DNA that differs from that found in the nucleus, are found in the internal mitochondrial matrix. Mitochondria also contain ribosomes similar to those found in bacteria. Mitochondria are approximately the size of many bacteria, typically about 1  $\mu\text{m}$  in diameter and 2 to 8  $\mu\text{m}$  in length. In theory, they may have arisen from the absorption of aerobic bacteria by larger host cells.

The **endoplasmic reticulum (ER)** is part of a continuous single-membrane system throughout the cell; the membrane doubles back on itself to give the appearance of a double membrane in electron micrographs. The endoplasmic reticulum is attached to the cell membrane and to the nuclear membrane. It occurs in two forms, rough and smooth. The *rough endoplasmic reticulum* is studded



■ **FIGURE 1.16** Rough endoplasmic reticulum from mouse liver cells (magnified 50,000 $\times$ ).

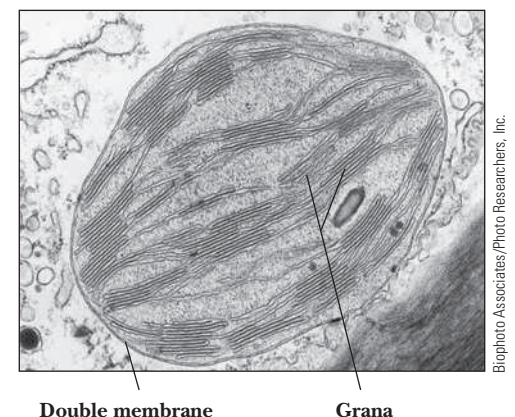
with ribosomes bound to the membrane (Figure 1.16). Ribosomes, which can also be found free in the cytosol, are the sites of protein synthesis in all organisms. The *smooth endoplasmic reticulum* does not have ribosomes bound to it.

Chloroplasts are important organelles found only in green plants and green algae. Their structure includes membranes, and they are relatively large, typically up to 2  $\mu\text{m}$  in diameter and 5 to 10  $\mu\text{m}$  in length. The photosynthetic apparatus is found in specialized structures called *grana* (singular *granum*), membranous bodies stacked within the chloroplast. Grana are easily seen through an electron microscope (Figure 1.17). Chloroplasts, like mitochondria, contain a characteristic DNA that is different from that found in the nucleus. Chloroplasts and mitochondria also contain ribosomes similar to those found in bacteria.

### What are some other important components of cells?

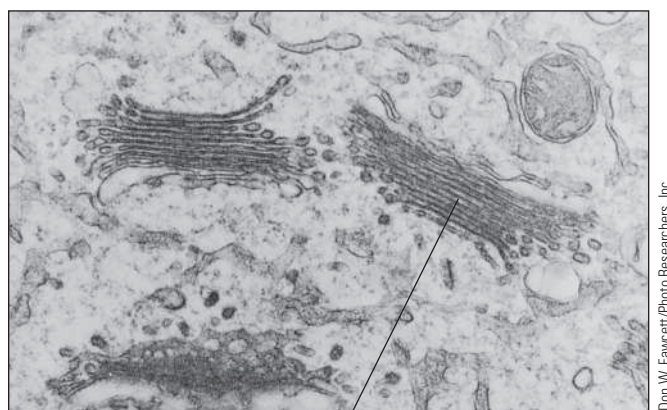
Membranes are important in the structures of some less well-understood organelles. One, the **Golgi apparatus**, is separate from the endoplasmic reticulum but is frequently found close to the smooth endoplasmic reticulum. It is a series of membranous sacs (Figure 1.18). The Golgi apparatus is involved in secretion of proteins from the cell, but it also appears in cells in which the primary function is not protein secretion. In particular, it is the site in the cell in which sugars are linked to other cellular components, such as proteins. The function of this organelle is still a subject of research.

Other organelles in eukaryotes are similar to the Golgi apparatus in that they involve single, smooth membranes and have specialized functions. **Lysosomes**, for example, are membrane-enclosed sacs containing hydrolytic enzymes that could cause considerable damage to the cell if they were not physically separated from the lipids, proteins, or nucleic acids that they are able to attack. Inside the lysosome, these enzymes break down target molecules, usually from outside sources, as a first step in processing nutrients for the cell. **Peroxisomes** are similar to lysosomes; their principal characteristic is that they contain enzymes involved in the metabolism of hydrogen peroxide ( $\text{H}_2\text{O}_2$ ), which is toxic to the cell. The enzyme *catalase*, which occurs in peroxisomes, catalyzes the conversion of  $\text{H}_2\text{O}_2$  to  $\text{H}_2\text{O}$  and  $\text{O}_2$ . **Glyoxysomes** are found in plant cells only. They contain the enzymes that catalyze the *glyoxylate cycle*, a pathway that converts some lipids to carbohydrate with glyoxylic acid as an intermediate.



■ **FIGURE 1.17** An electron microscope image of a chloroplast from the alga *Nitella* (magnified 60,000 $\times$ ).





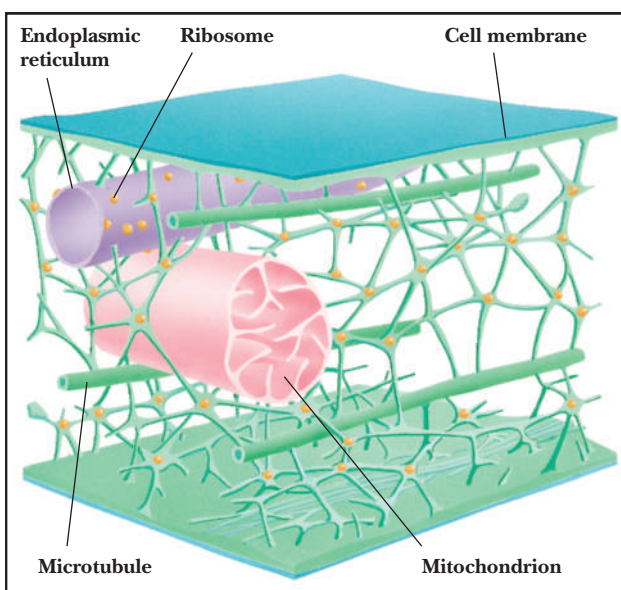
■ **FIGURE 1.18** Golgi apparatus from a mammalian cell (magnified 25,000 $\times$ ).

Stack of flattened  
membranous vesicles

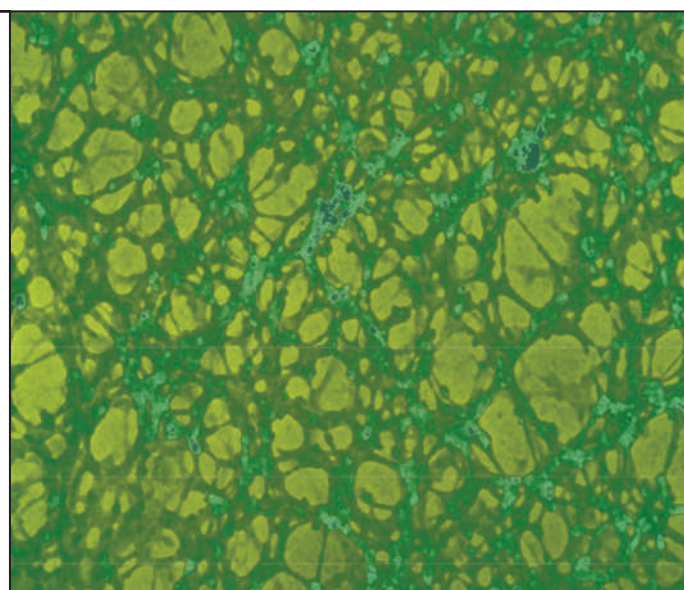
The **cytosol** was long considered nothing more than a viscous liquid, but recent studies by electron microscopy have revealed that this part of the cell has some internal organization. The organelles are held in place by a lattice of fine strands that seem to consist mostly of protein. This **cytoskeleton**, or *microtrabecular lattice*, is connected to all organelles (Figure 1.19). Many questions remain about its function in cellular organization, but its importance in maintaining the infrastructure of the cell is not doubted.

The cell membrane of eukaryotes separates the cell from the outside world. It consists of a double layer of lipids, with several types of proteins embedded in the lipid matrix. Some of the proteins transport specific substances across the membrane barrier. Transport can take place in both directions, with substances useful to the cell being taken in and others being exported.

Plant cells (and algae), but not animal cells, have cell walls external to the plasma membrane. The cellulose that makes up plant cell walls is a major



**A** This network of filaments, also called the cytoskeleton, pervades the cytosol. Some filaments, called microtubules, are known to consist of the protein tubulin. Organelles such as mitochondria are attached to the filaments.



**B** An electron micrograph of the microtrabecular lattice (magnified 87,450 $\times$ ).

■ **FIGURE 1.19** The microtrabecular lattice.



TABLE 1.4

A Summary of Organelles and Their Functions	
Organelle	Function
Nucleus	Location of main genome; site of most DNA and RNA synthesis
Mitochondrion	Site of energy-yielding oxidation reactions; has its own DNA
Chloroplast	Site of photosynthesis in green plants and algae; has its own DNA
Endoplasmic reticulum	Continuous membrane throughout the cell; rough part studded with <i>ribosomes</i> ( <i>the site of protein synthesis</i> )*
Golgi apparatus	Series of flattened membranes; involved in secretion of proteins from cells and in reactions that link sugars to other cellular components
Lysosomes	Membrane-enclosed sacs containing hydrolytic enzymes
Peroxisomes	Sacs that contain enzymes involved in the metabolism of hydrogen peroxide
Cell membrane	Separates the cell contents from the outside world; contents include organelles (held in place by the <i>cytoskeleton</i> *) and the <i>cytosol</i>
Cell wall	Rigid exterior layer of plant cells
Central vacuole	Membrane-enclosed sac (plant cells)

\* Because an organelle is defined as a portion of a cell enclosed by a membrane, ribosomes are not, strictly speaking, organelles. Smooth endoplasmic reticulum does not have ribosomes attached, and ribosomes also occur free in the cytosol. The definition of *organelle* also affects discussion of the cell membrane, cytosol, and cytoskeleton.

component of plant material; wood, cotton, linen, and most types of paper are mainly cellulose. Also present in plant cells are large central **vacuoles**, sacs in the cytoplasm surrounded by a single membrane. Although vacuoles sometimes appear in animal cells, those in plants are more prominent. They tend to increase in number and size as the plant cell ages. An important function of vacuoles is to isolate waste substances that are toxic to the plant and are produced in greater amounts than the plant can secrete to the environment.

These waste products may be unpalatable or even poisonous enough to discourage herbivores (plant-eating organisms) from ingesting them and may thus provide some protection for the plant.

Table 1.4 summarizes organelles and their functions.

## 1.7 Five Kingdoms, Three Domains

Living organisms can be classified in many possible ways. The original biological classification scheme, established in the 18th century, divided all organisms into two kingdoms: the plants and the animals. In this scheme, plants are organisms that obtain food directly from the Sun, and animals are organisms that move about to search for food. There are many other methods, however. For example, one could distinguish life based on whether the organism had a cell wall or not, or whether the organism was single-celled or not.

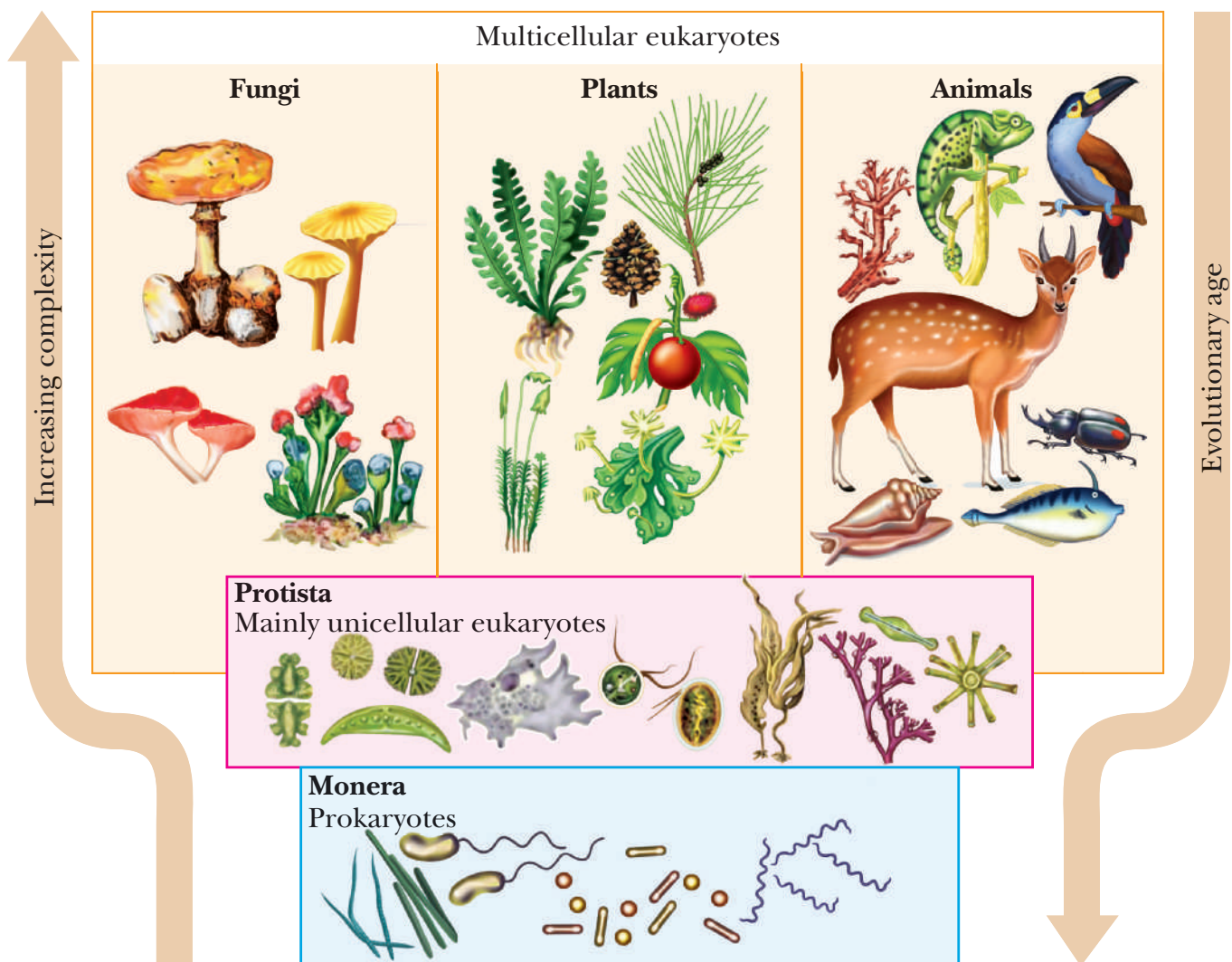
With the original classification, researchers discovered that some organisms, bacteria in particular, do not have an obvious relationship to either kingdom. It has also become clear that a more fundamental division of living organisms is actually not between plants and animals, but between prokaryotes and eukaryotes. In the 20th century, classification schemes that divide living organisms into more than the two traditional kingdoms have been introduced.

### How do scientists classify living organisms today?

A five-kingdom system takes into account the differences between prokaryotes and eukaryotes, and it also provides classifications for eukaryotes that appear to be neither plants nor animals.

The kingdom **Monera** consists only of prokaryotic organisms. Bacteria and cyanobacteria are members of this kingdom. The other four kingdoms are made up of eukaryotic organisms. The kingdom **Protista** includes unicellular organisms such as *Euglena*, *Volvox*, *Amoeba*, and *Paramecium*. Some protists, including algae, are multicellular. The three kingdoms that consist mainly of multicellular eukaryotes (with a few unicellular eukaryotes) are Fungi, Plantae, and Animalia. The kingdom Fungi includes yeasts, molds, and mushrooms. Fungi, plants, and animals must have evolved from simpler eukaryotic ancestors, but the major evolutionary change was the development of eukaryotes from prokaryotes (Figure 1.20).

One group of organisms can be classified as prokaryotes in the sense that the organisms lack a well-defined nucleus. These organisms are called **archaeobacteria** (early bacteria) to distinguish them from **eubacteria** (true bacteria) because there are marked differences between the two kinds of organisms. Archaeobacteria are found in extreme environments (see the Biochemical Connections box) and, for this reason, are also called extremophiles. Most of the differences between archaeobacteria and other organisms are biochemical



■ **FIGURE 1.20** The five-kingdom classification scheme.

# Biochemical Connections BIOTECHNOLOGY

## Extremophiles: The Toast of the Industry

Archaeobacteria live in extreme environments and, therefore, are sometimes called extremophiles. The three groups of archaeobacteria—methanogens, halophiles, and thermacidophiles—have specific preferences about the precise nature of their environment. *Methanogens* are strict anaerobes that produce methane ( $\text{CH}_4$ ) from carbon dioxide ( $\text{CO}_2$ ) and hydrogen ( $\text{H}_2$ ).

*Halophiles* require very high salt concentrations, such as those found in the Dead Sea, for growth. *Thermacidophiles* require high temperatures and acid conditions for growth—typically,  $80^\circ\text{C}$ – $90^\circ\text{C}$  and pH 2. These growth requirements may have resulted from adaptations to harsh conditions on the early Earth. Since these organisms can tolerate these conditions, the enzymes they produce must also be stable. Most enzymes isolated from eubacteria and eukaryotes are not stable under such conditions. Some of the reactions that are of greatest importance to the biotechnology industry are both enzyme-catalyzed and carried out under conditions that cause most enzymes to lose their catalytic ability in a short time. This difficulty can be avoided by using enzymes from extremophiles. An example is the DNA polymerase from *Thermus aquaticus* (Taq polymerase). Polymerase chain reaction (PCR) technology depends heavily on the properties of this enzyme (Section 13.6). Representatives of the biotechnology industry constantly search undersea thermal vents and hot springs for organisms that can provide such enzymes.



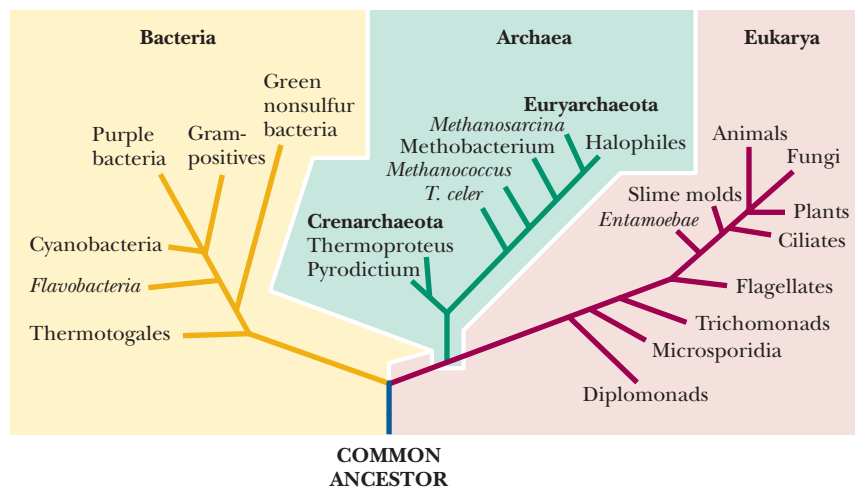
Sherrie Jones/Photo Researchers Inc.

■ **A hot spring in Yellowstone Park.** Some bacteria can thrive even in this inhospitable environment.

features, such as the molecular structure of the cell walls, membranes, and some types of RNA.

### Is there a simpler basis for classifying organisms?

Some biologists prefer a three-domain classification scheme—**Bacteria** (eubacteria), **Archaea** (archaeobacteria), and **Eukarya** (eukaryotes)—to the five-kingdom classification (Figure 1.21). The basis for this preference is the emphasis on biochemistry as the basis for classification.



■ **FIGURE 1.21 The three-domain classification scheme.** Two domains, Bacteria and Archaea, consist of prokaryotes. The third kingdom, Eukarya, consists of eukaryotes. All three domains have a common ancestor early in evolution. (Reprinted with permission from Science 273, 1044. Copyright © 1996 AAAS.)





James King-Holmes/SPL/Photo Researchers, Inc.

- **FIGURE 1.22** An organism whose genome has been determined. Like that of humans, the genome of *Caenorhabditis elegans* has been decoded. *C. elegans* is ideal for studying genetic blueprints because of its tendency to reproduce by self-fertilization. This results in offspring that are identical to the parent.



C. P. Vance/Visuals Unlimited

- **FIGURE 1.23** The root system of a leguminous plant. Leguminous plants live symbiotically with nitrogen-fixing bacteria in their root systems.

The three-domain classification scheme will certainly become more important as time goes on. A complete genome of the archaeobacterium *Methanococcus jannaschii* has been obtained. More than half the genes of this organism (56%) differ markedly from genes already known in both prokaryotes and eukaryotes, a piece of evidence that lends strong support to a three-domain classification scheme. Complete genomes are being obtained for organisms from all three domains. They include those of bacteria such as *Haemophilus influenzae* and *Escherichia coli*, the latter being a bacterium in which many biochemical pathways have been investigated. Complete sequences for eukaryotes such as *Saccharomyces cerevisiae* (brewer's yeast), *Arabidopsis thaliana* (mouse-ear cress), and *Caenorhabditis elegans* (a nematode) have been obtained (Figure 1.22). The sequencing of the genomes of the mouse (*Mus musculus*) and *Drosophila melanogaster* (a fruit fly) has also been completed, with genome sequences of many more organisms on the way. The most famous of all genome-sequencing projects, that for the human genome, has received wide publicity, with the results now available on the World Wide Web.

## 1.8 Common Ground for All Cells

### Did eukaryotes develop from prokaryotes?

The complexity of eukaryotes raises many questions about how such cells arose from simpler progenitors. Symbiosis plays a large role in current theories of the rise of eukaryotes; the symbiotic association between two organisms is seen as giving rise to a new organism that combines characteristics of both the original ones. The type of symbiosis called *mutualism* is a relationship that benefits both species involved, as opposed to *parasitic symbiosis*, in which one species gains at the other's expense. A classic example of mutualism (although it has been questioned from time to time) is the lichen, which consists of a fungus and an alga. The fungus provides water and protection for the alga; the alga is photosynthetic and provides food for both partners. Another example is the root-nodule system formed by a leguminous plant, such as alfalfa or beans, and anaerobic nitrogen-fixing bacteria (Figure 1.23). The plant gains useful compounds of nitrogen, and the bacteria are protected from oxygen, which is harmful to them. Still another example of mutualistic symbiosis, of great practical interest, is that between humans and bacteria, such as *Escherichia coli*, that live in the intestinal tract. The bacteria receive nutrients and protection from their immediate environment. In return, they aid our digestive process. Without beneficial intestinal bacteria, we would soon develop dysentery and other intestinal disorders. These bacteria are also a source of certain vitamins for us, because they can synthesize these vitamins and we cannot. The disease-causing strains of *E. coli* that have been in the news from time to time differ markedly from the ones that naturally inhabit the intestinal tract.

### Did symbiosis play a role in the development of eukaryotes?

In hereditary symbiosis, a larger host cell contains a genetically determined number of smaller organisms. An example is the protist *Cyanophora paradoxa*, a eukaryotic host that contains a genetically determined number of cyanobacteria (blue-green algae). This relationship is an example of *endosymbiosis*, because the cyanobacteria are contained within the host organism. The cyanobacteria are aerobic prokaryotes and are capable of photosynthesis (Figure 1.24). The host cell gains the products of photosynthesis; in return, the cyanobacteria are protected from the environment and still have access to oxygen and sunlight because of the host's small size. In this model, with the passage of many



©Stockphoto.com/Jouke van der Meer

■ **FIGURE 1.24 Stromatolite fossils.** Stromatolites are large, stony, cushionlike masses composed of numerous layers of cyanobacteria (blue-green algae) that have been preserved because of their ability to secrete calcium carbonate. They are among the oldest organic remains to have been found. This specimen dates from around 2400 million years ago. Stromatolite formation reached a peak during the late Precambrian period (4000 million to 570 million years ago) but is still occurring today. This specimen was found in Argentina.

generations, the cyanobacteria would have gradually lost the ability to exist independently and would have become organelles within a new and more complex type of cell. Such a situation in the past may well have given rise to chloroplasts, which are not capable of independent existence. Their autonomous DNA and their apparatus for synthesizing ribosomal proteins can no longer meet all their needs, but the fact that these organelles have their own DNA and are capable of protein synthesis suggests that they may have existed as independent organisms in the distant past.

A similar model can be proposed for the origin of mitochondria. Consider this scenario: A large anaerobic host cell assimilates a number of smaller aerobic bacteria. The larger cell protects the smaller ones and provides them with nutrients. As in the example we used for the development of chloroplasts, the smaller cells still have access to oxygen. The larger cell is not itself capable of aerobic oxidation of nutrients, but some of the end products of its anaerobic oxidation can be further oxidized by the more efficient aerobic metabolism of the smaller cells. As a result, the larger cell can get more energy out of a given amount of food than it could without the bacteria. In time, the two associated organisms evolve to form a new aerobic organism, which contains mitochondria derived from the original aerobic bacteria.

The fact that both mitochondria and chloroplasts have their own DNA is an important piece of biochemical evidence in favor of this model. Additionally, both mitochondria and chloroplasts have their own apparatus for synthesis of RNA and proteins. The genetic code in mitochondria differs slightly from that found in the nucleus, which supports the idea of an independent origin. Thus, the remains of these systems for synthesis of RNA and protein could reflect the organelles' former existence as free-living cells. It is reasonable to conclude that large unicellular organisms that assimilated aerobic bacteria went on to evolve mitochondria from the bacteria and eventually gave rise to animal cells. Other types of unicellular organisms assimilated both aerobic bacteria and cyanobacteria and evolved both mitochondria and chloroplasts; these organisms eventually gave rise to green plants.

The proposed connections between prokaryotes and eukaryotes are not established with complete certainty, and they leave a number of questions unanswered. Still, they provide an interesting frame of reference from which to consider evolution and the origins of the reactions that take place in cells.

## 1.9 Biochemical Energetics

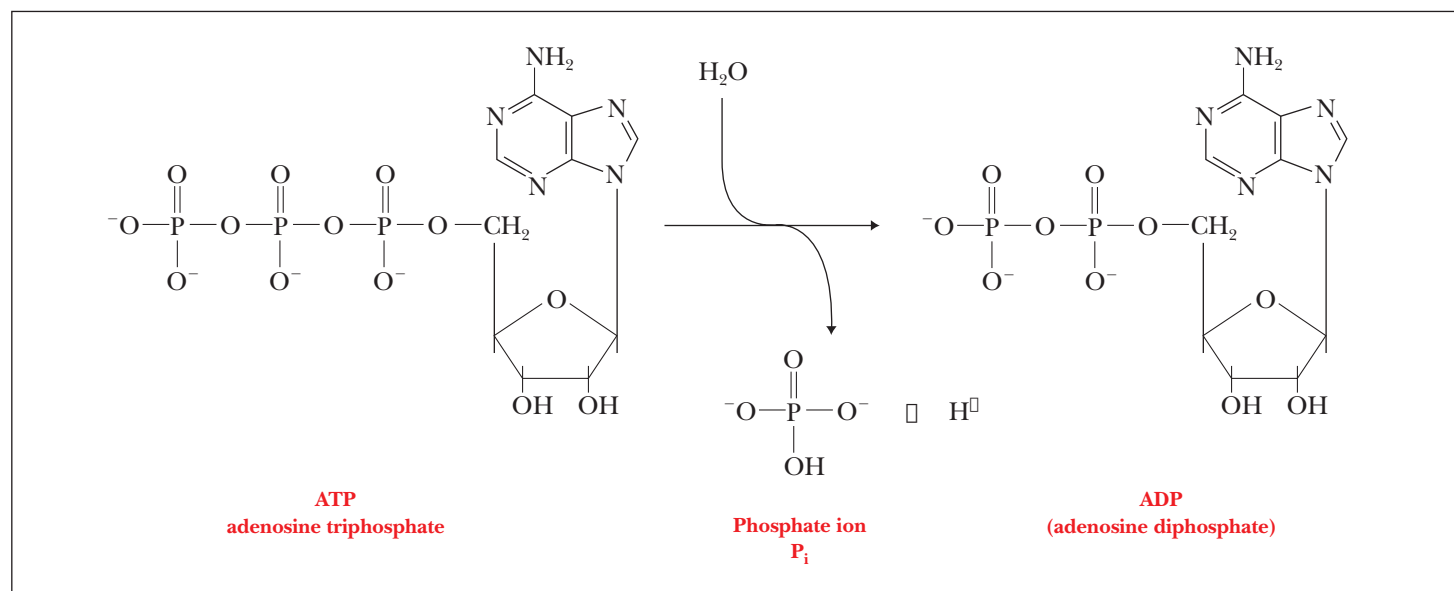
### What is the source of energy in life processes?

All cells require energy for a number of purposes. Many reactions that take place in the cell, particularly those involving synthesis of large molecules, cannot take place unless energy is supplied. The Sun is the ultimate source of energy for all life on Earth. Photosynthetic organisms trap light energy and use it to drive the energy-requiring reactions that convert carbon dioxide and water to carbohydrates and oxygen. (Note that these reactions involve the chemical process of **reduction**.) Nonphotosynthetic organisms, such as animals that consume these carbohydrates, use them as energy sources. (The reactions that release energy involve the chemical process of **oxidation**.) We are going to discuss the roles that oxidation and reduction reactions play in cellular processes in Chapter 15, and you will see many examples of such reactions in subsequent chapters. For the moment, it is useful and sufficient to recall from general chemistry that *oxidation is the loss of electrons and reduction is the gain of electrons*.

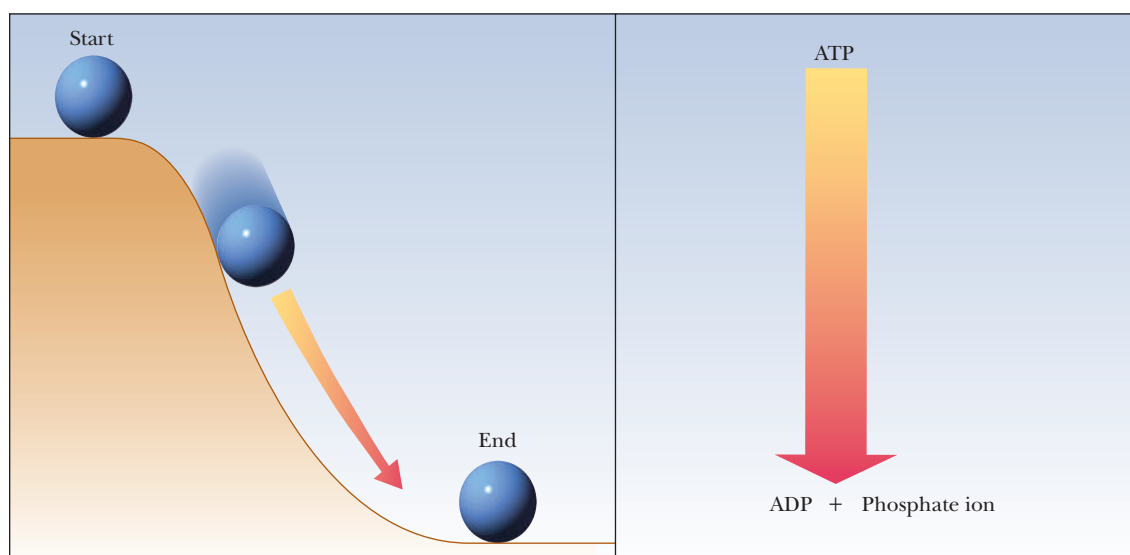
### How do we measure energy changes in biochemistry?

One of the most important questions about any process is whether energy changes favor the process. **Thermodynamics** is the branch of science that deals with this question. The key point is that *processes that release energy are favored*. Conversely, processes that require energy are disfavored. The change in energy depends only on the state of the molecules present at the start of the process and the state of those present at the end of the process. This is true whether the process in question is the formation or breaking of a bond, the formation or disruption of an intermolecular interaction, or any possible process that requires or can release energy. We are going to discuss these points in some detail when we look at protein folding in Chapter 4 and at energy considerations in metabolism in Chapter 15. This material is of central importance, and it tends to be challenging for many. What we say about it now will make it easier to apply in later chapters.

A reaction that takes place as a part of many biochemical processes is the hydrolysis of the compound adenosine triphosphate, or ATP (Section 1.2).







**A** A ball rolls down a hill, releasing potential energy.

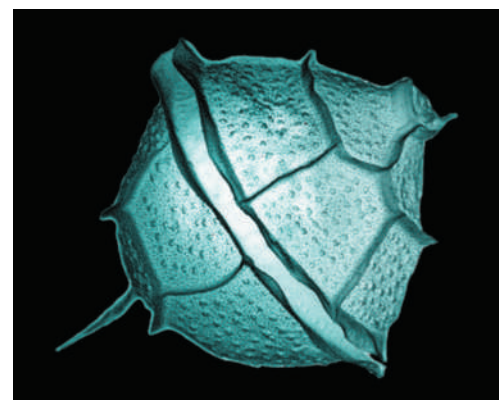
**B** ATP is hydrolyzed to produce ADP and phosphate ion, releasing energy. The release of energy when a ball rolls down a hill is analogous to the release of energy in a chemical reaction.

**FIGURE 1.25** Schematic representation of the lowering of energy.

This reaction releases energy ( $30.5 \text{ kJ mol}^{-1} \text{ ATP} = 7.3 \text{ kcal/mol ATP}$ ). More to the point, the energy released by this reaction allows energy-requiring reactions to proceed. Many ways are available to express energy transfer. One of the most common is the free energy,  $G$ , which is discussed in general chemistry. Also recall from general chemistry that a lowering (release) of energy leads to a more stable state of the system under consideration. The lowering of energy is frequently shown in pictorial form as analogous to an object rolling down a hill (Figure 1.25) or over a waterfall. This representation calls on common experience and aids understanding.



**A**



**B**

**FIGURE 1.26** Two examples of transformations of energy in biological systems. (a) This electric ray (a marine fish in the family Torpedinidae) converts chemical energy to electrical energy, and (b) the bioluminescent dinoflagellate *Pyrodinium* converts chemical energy into light energy.

## 1.10 Energy and Change

### What kinds of energy changes take place in living cells?

Energy can take several forms, and it can be converted from one form to another. All living organisms require and use energy in varied forms; for example, motion involves mechanical energy, and maintenance of body temperature uses thermal energy. Photosynthesis requires light energy from the Sun. Some organisms, such as several species of fish, are striking examples of the use of chemical energy to produce electrical energy (Figure 1.26). The formation and breakdown of biomolecules involve changes in chemical energy.

Any process that will actually take place with no outside intervention is **spontaneous** in the specialized sense used in thermodynamics. *Spontaneous does not mean “fast”; some spontaneous processes can take a long time to occur.* In the last section, we used the term *energetically favorable* to indicate spontaneous processes. The laws of thermodynamics can be used to predict whether any change involving transformations of energy will take place. An example of such a change is a chemical reaction in which covalent bonds are broken and new ones are formed. Another example is the formation of noncovalent interactions, such as hydrogen bonds, or hydrophobic interactions, when proteins fold to produce their characteristic three-dimensional structures. The tendency of polar and

nonpolar substances to exist in separate phases is a reflection of the energies of interaction between the individual molecules—in other words, a reflection of the thermodynamics of the interaction.

## 1.11 Spontaneity in Biochemical Reactions

### How can we predict what reactions will happen in living cells?

The most useful criterion for predicting the spontaneity of a process is the **free energy**, which is indicated by the symbol  $G$ . (Strictly speaking, the use of this criterion requires conditions of constant temperature and pressure, which are usual in biochemical thermodynamics.) It is not possible to measure absolute values of energy; only the *changes* in energy that occur during a process can be measured. The value of the change in free energy,  $\Delta G$  (where the symbol  $\Delta$  indicates change), gives the needed information about the spontaneity of the process under consideration.

The free energy of a system decreases in a spontaneous (energy-releasing) process, so  $\Delta G$  is negative ( $\Delta G < 0$ ). Such a process is called **exergonic**, meaning that energy is released. When the change in free energy is positive ( $\Delta G > 0$ ), the process is nonspontaneous. For a nonspontaneous process to occur, energy must be supplied. Nonspontaneous processes are also called **endergonic**, meaning that energy is absorbed. For a process at **equilibrium**, with no net change in either direction, the change in free energy is zero ( $\Delta G = 0$ ).

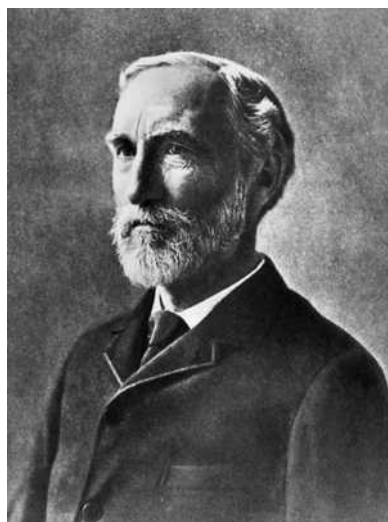
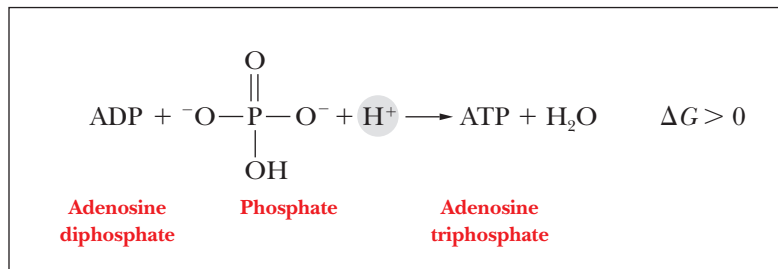
The sign of the change in free energy,  $\Delta G$ , indicates the direction of the reaction:

$\Delta G < 0$	Spontaneous exergonic—energy released
$\Delta G = 0$	Equilibrium
$\Delta G > 0$	Nonspontaneous endergonic—energy required

An example of a spontaneous process is the aerobic metabolism of glucose, in which glucose reacts with oxygen to produce carbon dioxide, water, and energy for the organism.



An example of a nonspontaneous process is the reverse of the reaction that we saw in Section 1.9—namely, the phosphorylation of ADP (adenosine diphosphate) to give ATP (adenosine triphosphate). This reaction takes place in living organisms because metabolic processes supply energy.



© Bettmann/CORBIS

#### ■ J. Willard Gibbs (1839–1903).

The symbol  $G$  is given to free energy in his honor. His work is the basis of biochemical thermodynamics, and he is considered by some to have been the greatest scientist born in the United States.

## 1.12 Life and Thermodynamics

From time to time, one encounters the statement that the existence of living things violates the laws of thermodynamics, specifically the second law. A look at the laws will clarify whether life is thermodynamically possible, and further discussion of thermodynamics will increase our understanding of this important topic.

### Is life thermodynamically possible?

The laws of thermodynamics can be stated in several ways. According to one formulation, the first law is, “You can’t win,” and the second is, “You can’t break even.” Put less flippantly, the first law states that it is impossible to convert energy from one form to another at greater than 100% efficiency. In other words, the first law of thermodynamics is the law of conservation of energy. The second law states that even 100% efficiency in energy transfer is impossible.

The two laws of thermodynamics can be related to the free energy by means of a well-known equation:

$$\Delta G = \Delta H - T\Delta S$$

In this equation,  $G$  is the free energy, as before;  $H$  stands for the **enthalpy**, and  $S$  for the entropy. Discussions of the first law focus on the change in enthalpy,  $\Delta H$ , which is the **heat of a reaction at constant pressure**. This quantity is relatively easy to measure. Enthalpy changes for many important reactions have been determined and are available in tables in textbooks of general chemistry. Discussions of the second law focus on changes in entropy,  $\Delta S$ , a concept that is less easily described and measured than changes in enthalpy. Entropy changes are particularly important in biochemistry.

One of the most useful definitions of entropy arises from statistical considerations. From a statistical point of view, an increase in the entropy of a system (the substance or substances under consideration) represents an increase in the number of possible arrangements of objects, such as individual molecules. Books have a higher entropy when they are scattered around the reading room of a library than when they are in their proper places on the shelves. Scattered books are clearly in a more dispersed state than books on shelves. The natural tendency of the Universe is in the direction of increasing dispersion of energy, and living organisms put a lot of energy into maintaining order against this tendency. As all parents know, they can spend hours cleaning up a two-year-old’s room, but the child can undo it all in seconds. Similarly, cells use a lot of energy to fight the natural tendency toward dispersion into many different arrangements and to keep the cell structure intact.



© Bettmann/CORBIS

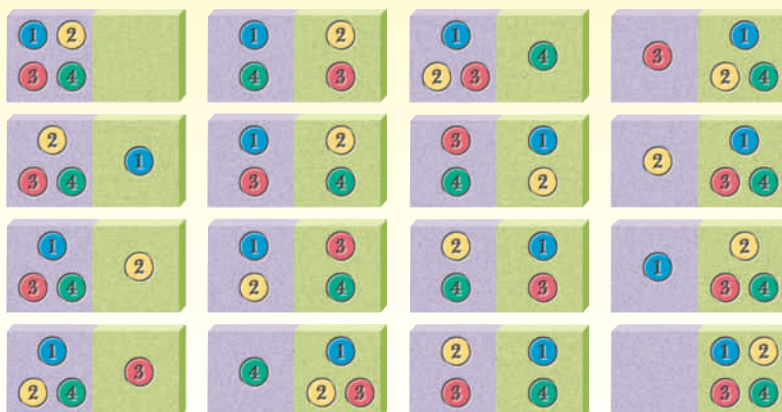
■ **Ludwig Boltzmann (1844–1906).** His equation for entropy in terms of the dispersion of objects in the Universe was one of his supreme achievements; his equation is carved on his tombstone.

## Biochemical Connections THERMODYNAMICS

### Predicting Reactions

Let us consider a very simple system to illustrate the concept of entropy. We place four molecules in a container. There is an equal chance that each molecule will be on the left or on the right side of the container. Mathematically stated, the *probability* of finding a given molecule on one side is  $1/2$ . We can express any probability as a fraction ranging from 0 (impossible) to 1 (completely certain). We can see that 16 possible ways exist to arrange the four molecules in the container. In only one of these will all four molecules lie on the left side, but six possible arrangements exist with the four molecules evenly distributed between the two sides. A *less ordered (more dispersed) arrangement is more probable than a highly ordered arrangement*. Entropy is defined in terms of the number of possible arrangements of molecules.

Boltzmann’s equation for entropy,  $S$ , is  $S = k \ln W$ . In this equation, the term  $W$  represents the number of possible arrangements of molecules,  $\ln$  is the logarithm to the base  $e$ , and  $k$  is the constant universally referred to as Boltzmann’s constant. It is equal to  $R/N$  where  $R$  is the gas constant and  $N$  is Avogadro’s number ( $6.02 \times 10^{23}$ ), the number of molecules in a mole.



■ **The 16 possible states for a system of four molecules that may occupy either side of a container.** In only one of these states are all four molecules on the left side.



Another statement of the second law is this: *in any spontaneous process, the entropy of the Universe increases ( $\Delta S_{\text{univ}} > 0$ )*. This statement is general, and it applies to any set of conditions. It is not confined to the special case of constant temperature and pressure, as is the statement that the free energy decreases in a spontaneous process. *Entropy changes are particularly important in determining the energetics of protein folding.*

## SUMMARY

**How does biochemistry describe life processes?** Biochemistry is a multidisciplinary field that asks questions about the molecular nature of life processes. Many chemical reactions take place simultaneously in living cells.

**How did living things originate?** The fundamental biochemical similarities observed in all living organisms have engendered speculation about the origins of life.

**Can a chemist make the molecules of life in a laboratory?** Both organic chemistry and biochemistry deal with the reactions of carbon-containing molecules. Since the structure of these molecules is the same whether they originate in a living organism or a laboratory, it is possible, but sometimes extremely difficult, to make the molecules of life in a laboratory.

**What makes biomolecules special?** Both organic chemistry and biochemistry base their approaches on the behavior of functional groups, but their emphases differ because some functional groups important to organic chemistry do not play a role in biochemistry, and vice versa. Functional groups of importance in biochemistry include carbonyl groups, hydroxyl groups, carboxyl groups, amines, amides, and esters; derivatives of phosphoric acid such as esters and anhydrides are also important.

**How and when did the Earth come to be?** The Earth, along with the rest of the solar system, was formed about 4 to 5 billion years ago from elements produced by first-generation stars.

**How were biomolecules likely to have formed on the early Earth?** It has been shown that important biomolecules can be produced under abiotic (nonliving) conditions from simple compounds postulated to have been present in the atmosphere of the early Earth. These simple biomolecules can polymerize, also under abiotic conditions, to give rise to compounds resembling proteins and others having a less marked resemblance to nucleic acids.

**Which came first—the catalysts or the hereditary molecules?** All cellular activity depends on the presence of catalysts, which increase the rates of chemical reactions, and on the genetic code, which directs the synthesis of the catalysts. In present-day cells, catalytic activity is associated with proteins,

and transmission of the genetic code is associated with nucleic acids, particularly with DNA. Both these functions may once have been carried out by a single biomolecule, RNA. It has been postulated that RNA was the original coding material, and it has recently been shown to have catalytic activity as well. The formation of peptide bonds in protein biosynthesis is catalyzed by the RNA portions of the ribosome.

**What is the difference between a prokaryote and a eukaryote?** Organisms are divided into two main groups based on their cell structures. *Prokaryotes* do not have internal membranes, whereas *eukaryotes* do. Organelles, which are membrane-enclosed portions of cells with specific functions, are characteristic of eukaryotes.

**How is prokaryotic DNA organized without a nucleus?** In *prokaryotes*, the cell lacks a well-defined nucleus and internal membrane; it has only a nuclear region, the portion of the cell that contains DNA, and a cell membrane that separates it from the outside world. The other principal feature of a prokaryotic cell's interior is the presence of ribosomes, the site of protein synthesis.

**What are the most important organelles?** A *eukaryotic* cell has a well-defined nucleus, internal membranes as well as a cell membrane, and a considerably more complex internal structure than a prokaryote. In eukaryotes, the nucleus is separated from the rest of the cell by a double membrane. Eukaryotic DNA in the nucleus is associated with proteins, particularly a class of proteins called histones. The combination of the two has specific structural motifs, which is not the case in prokaryotes. There is a continuous membrane system, called the endoplasmic reticulum, throughout the cell. Eukaryotic ribosomes are frequently bound to the endoplasmic reticulum, but some are also free in the cytosol. Membrane-enclosed organelles are characteristic of eukaryotic cells. Two of the most important are mitochondria, the sites of energy-yielding reactions, and chloroplasts, the sites of photosynthesis.

**What are some other important components of cells?** Other components of eukaryotic cells includes such features as the Golgi apparatus (involved in secretion of proteins from the cell), lysosomes (containers for hydrolytic enzymes), and the cytoskeleton (a framework for organization of various organelles).

**How do scientists classify living organisms today?** Two ways of classifying organisms depend on the distinction between prokaryotes and eukaryotes. In the five-kingdom scheme, prokaryotes occupy the kingdom Monera. The other four kingdoms consist of eukaryotes: Protista, Fungi, Plantae, and Animalia.

**Is there a simpler basis for classifying organisms?** In the three-domain scheme, prokaryotes occupy two domains—Bacteria and Archaea—based on biochemical differences, and all eukaryotes occupy a single domain, Eukarya.

**Did eukaryotes develop from prokaryotes?** A good deal of research has gone into the question of how eukaryotes may have arisen from prokaryotes.

**Did symbiosis play a role in the development of eukaryotes?** Much of the thinking depends on the idea of *endosymbiosis*, in which larger cells may have absorbed aerobic bacteria, eventually giving rise to mitochondria, or photosynthetic bacteria, eventually giving rise to chloroplasts.

**What is the source of energy in life processes?** All cells require energy to carry out life processes. The Sun is the ultimate source of energy on Earth. Photosynthetic organisms trap light energy from the Sun as the chemical energy of the

carbohydrates they produce. These carbohydrates serve as energy sources for other organisms in turn.

**How do we measure energy changes in biochemistry?** It is possible to measure amounts of energy released or absorbed in a process to see whether it is likely to occur or not. Reactions that release energy are energetically favored, whereas those that require energy are disfavored.

**What kinds of energy changes take place in living cells?** Energy can take several forms, which can be converted from one to another in cells. Thermodynamics deals with the changes in energy that determine whether a process will take place. A process that will take place without outside intervention is called *spontaneous*.

**How can we predict what reactions will happen in living cells?** In a spontaneous process, the free energy decreases ( $\Delta G$  is negative). In a nonspontaneous process, the free energy increases.

**Is life thermodynamically possible?** In addition to the free energy, entropy is an important quantity in thermodynamics. The entropy of the Universe increases in any spontaneous process. Local decreases in entropy can take place within an overall increase in entropy. Living organisms represent local decreases in entropy.

## REVIEW EXERCISES

**OWL** Interactive versions of these problems are assignable in OWL

The exercises at the end of each chapter are divided into two or more categories to provide the benefit of more than one approach to reviewing the material.

*Recall* questions will allow you to test yourself about having important facts readily available to you. In some chapters, the material lends itself to quantitative calculations, and in those chapters you will see a *Mathematical* category. *Reflect and Apply* questions ask you to put those facts to use in questions that require using the concepts in the chapter in moderately creative ways. Lastly, questions that relate specifically to Biochemical Connections boxes are labeled *Biochemical Connections*.

### 1.1 Basic Themes

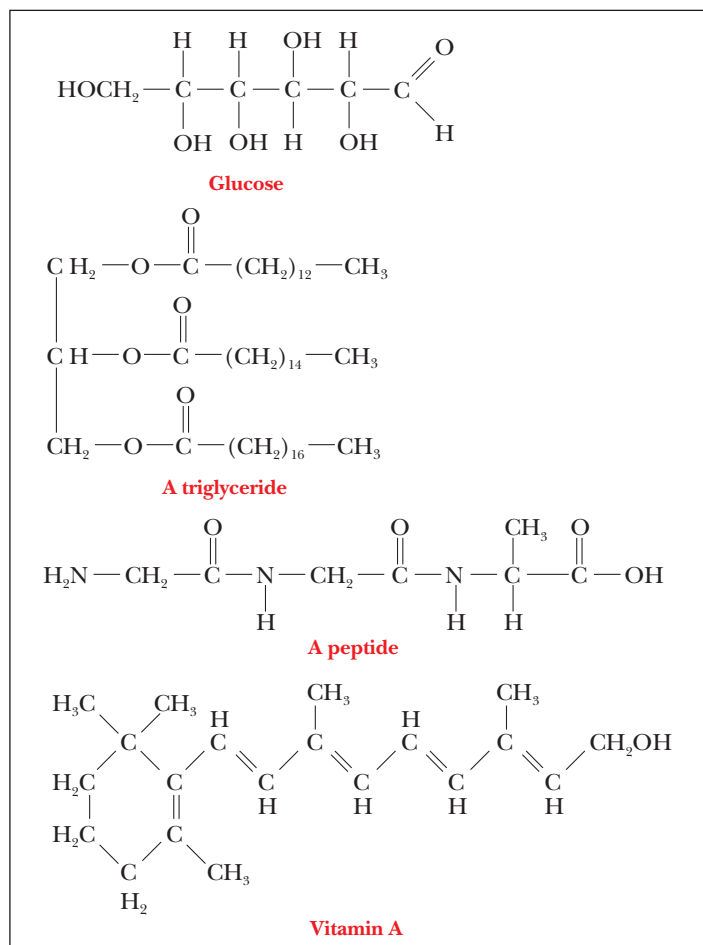
1. **Recall** State why the following terms are important in biochemistry: polymer, protein, nucleic acid, catalysis, genetic code.

### 1.2 Chemical Foundations of Biochemistry

2. **Recall** Match each entry in Column a with one in Column b; Column a shows the names of some important functional groups, and Column b shows their structures.

Column a	Column b
Amino group	$\text{CH}_3\text{SH}$
Carbonyl group (ketone)	$\text{CH}_3\text{CH}=\text{CHCH}_3$
	$\begin{array}{c} \text{O} \\    \\ \text{CH}_3\text{CH}_2\text{CH} \\ \text{CH}_3\text{CH}_2\text{NH}_2 \end{array}$
Hydroxyl group	
Carboxyl group	$\begin{array}{c} \text{O} \\    \\ \text{CH}_3\text{COCH}_2\text{CH}_3 \\ \text{CH}_3\text{CH}_2\text{OCH}_2\text{CH}_3 \end{array}$
	$\begin{array}{c} \text{O} \\    \\ \text{CH}_3\text{CCH}_3 \end{array}$
Carbonyl group (aldehyde)	
Thiol group	$\begin{array}{c} \text{O} \\    \\ \text{CH}_3\text{COH} \\ \text{CH}_3\text{OH} \end{array}$
	$\begin{array}{c} \text{O} \\    \\ \text{CH}_3\text{CN}(\text{CH}_3)_2 \end{array}$
Ester linkage	
Double bond	
Amide linkage	
Ether	

3. **Recall** Identify the functional groups in the following compounds.



4. **Reflect and Apply** In 1828, Wöhler was the first person to synthesize an organic compound (urea, from ammonium cyanate). How did this contribute, ultimately, to *biochemistry*?
5. **Reflect and Apply** A friend who is enthusiastic about health foods and organic gardening asks you whether urea is “organic” or “chemical.” How do you reply to this question?
6. **Reflect and Apply** Does biochemistry differ from organic chemistry? Explain your answer. (Consider such features as solvents, concentrations, temperatures, speed, yields, side reactions, and internal control.)

### 1.3 The Beginnings of Biology: Origin of Life

7. **Reflect and Apply** An earlier mission to Mars contained instruments that determined that amino acids were present on the surface of Mars. Why were scientists excited by this discovery?
8. **Reflect and Apply** Common proteins are polymers of 20 different amino acids. How big a protein (how many amino acid residues) would be necessary to have an Avogadro’s number of possible sequences?
9. **Reflect and Apply** Nucleic acids are polymers of just four different monomers in a linear arrangement. How many different sequences are available if one makes a polymer with only 40 monomers? How does this number compare with Avogadro’s number?
10. **Reflect and Apply** RNA is often characterized as being the first “biologically active” molecule. What two properties or activities does RNA display that are important to the evolution of life? *Hint:* Neither proteins nor DNA have *both* of these properties.
11. **Reflect and Apply** Why is the development of catalysis important to the development of life?

12. **Reflect and Apply** What are two major advantages of enzyme catalysts in living organisms when compared with other simple chemical catalysts such as acids or bases?
13. **Reflect and Apply** Why was the development of a coding system important to the development of life?
14. **Reflect and Apply** Comment on RNA’s role in catalysis and coding in theories of the origin of life.
15. **Reflect and Apply** Do you consider it a reasonable conjecture that cells could have arisen as bare cytoplasm without a cell membrane?

### 1.4 The Biggest Biological Distinction—Prokaryotes and Eukaryotes

16. **Recall** List five differences between prokaryotes and eukaryotes.
17. **Recall** Do the sites of protein synthesis differ in prokaryotes and eukaryotes?

### 1.5 Prokaryotic Cells

18. **Reflect and Apply** Assume that a scientist claims to have discovered mitochondria in bacteria. Is such a claim likely to prove valid?

### 1.6 Eukaryotic Cells

19. **Recall** Draw an idealized animal cell, and identify the parts by name and function.
20. **Recall** Draw an idealized plant cell, and identify the parts by name and function.
21. **Recall** What are the differences between the photosynthetic apparatus of green plants and photosynthetic bacteria?
22. **Recall** Which organelles are surrounded by a double membrane?
23. **Recall** Which organelles contain DNA?
24. **Recall** Which organelles are the sites of energy-yielding reactions?
25. **Recall** State how the following organelles differ from each other in terms of structure and function: Golgi apparatus, lysosomes, peroxisomes, glyoxysomes. How do they resemble each other?

### 1.7 Five Kingdoms, Three Domains

26. **Recall** List the five kingdoms into which living organisms are divided, and give at least one example of an organism belonging to each kingdom.
27. **Recall** Which of the five kingdoms consist of prokaryotes? Which consist of eukaryotes?
28. **Recall** List the three domains into which living organisms are divided, and indicate how this scheme differs from the five-kingdom classification scheme.

### 1.8 Common Ground for All Cells

29. **Reflect and Apply** What are the advantages of being eukaryotic (as opposed to prokaryotic)?
30. **Reflect and Apply** Mitochondria and chloroplasts contain some DNA, which more closely resembles prokaryotic DNA than (eukaryotic) nuclear DNA. Use this information to suggest how eukaryotes may have originated.
31. **Reflect and Apply** Fossil evidence indicates that prokaryotes have been around for about 3.5 billion years, whereas the origin of eukaryotes has been dated at only about 1.5 billion years ago. Suggest why, in spite of the lesser time for evolution, eukaryotes are much more diverse (have a much larger number of species) than prokaryotes.

### 1.9 Biochemical Energetics

32. **Recall** Which processes are favored: those that require energy or those that release energy?

### 1.10 Energy and Change

33. **Recall** Does the thermodynamic term *spontaneous* refer to a process that takes place quickly?



## 1.11 Spontaneity in Biochemical Reactions

### 34. Biochemical Connections

Nonpolar solute + H<sub>2</sub>O → Solution

what are the signs of  $\Delta S_{\text{univ}}$ ,  $\Delta S_{\text{sys}}$ , and  $\Delta S_{\text{surr}}$ ? What is the reason for each answer? ( $\Delta S_{\text{surr}}$  refers to the entropy change of the surroundings, all of the Universe but the system.)

### 35. Recall

Which of the following are spontaneous processes? Explain your answer for each process.

- The hydrolysis of ATP to ADP and P<sub>i</sub>
- The oxidation of glucose to CO<sub>2</sub> and H<sub>2</sub>O by an organism
- The phosphorylation of ADP to ATP
- The production of glucose and O<sub>2</sub> from CO<sub>2</sub> and H<sub>2</sub>O in photosynthesis

### 36. Reflect and Apply

In which of the following processes does the entropy increase? In each case, explain why it does or does not increase.

- A bottle of ammonia is opened. The odor of ammonia is soon apparent throughout the room.
- Sodium chloride dissolves in water.
- A protein is completely hydrolyzed to the component amino acids.

*Hint:* For Questions 37 through 39, consider the equation  $\Delta G = \Delta H - T(\Delta S)$ .

### 37. Reflect and Apply

Why is it necessary to specify the temperature when making a table listing  $\Delta G$  values?

### 38. Reflect and Apply

Why is the entropy of a system dependent on temperature?

### 39. Reflect and Apply

A reaction at 23°C has  $\Delta G = 1 \text{ kJ mol}^{-1}$ . Why might this reaction become spontaneous at 37°C?

### 40. Reflect and Apply

Urea dissolves very readily in water, but the solution becomes very cold as the urea dissolves. How is this possible? It appears that the solution is absorbing energy.

### 41. Reflect and Apply

Would you expect the reaction  $\text{ATP} \rightarrow \text{ADP} + \text{P}_i$  to be accompanied by a decrease or increase in entropy? Why?

## 1.12 Life and Thermodynamics

### 42. Reflect and Apply

The existence of organelles in eukaryotic cells represents a higher degree of organization than that found in prokaryotes. How does this affect the entropy of the Universe?

### 43. Reflect and Apply

Why is it advantageous for a cell to have organelles? Discuss this concept from the standpoint of thermodynamics.

### 44. Reflect and Apply

Which would you expect to have a higher entropy: DNA in its well-known double-helical form, or DNA with the strands separated?

### 45. Reflect and Apply

How would you modify your answer to Question 29 in light of the material on thermodynamics?

### 46. Reflect and Apply

Would it be more or less likely that cells of the kind we know would evolve on a gas giant such as the planet Jupiter?

### 47. Reflect and Apply

What thermodynamic considerations might enter into finding a reasonable answer to Question 46?

### 48. Reflect and Apply

If cells of the kind we know were to have evolved on any other planet in our solar system, would it be more likely to have happened on Mars or on Jupiter? Why?

### 49. Reflect and Apply

The process of protein folding is spontaneous in the thermodynamic sense. It gives rise to a highly ordered conformation that has a lower entropy than the unfolded protein. How can this be?

### 50. Reflect and Apply

In biochemistry, the exergonic process of converting glucose and oxygen to carbon dioxide and water in aerobic metabolism can be considered the reverse of photosynthesis, in which carbon dioxide and water are converted to glucose and oxygen. Do you expect both processes to be exergonic, both endergonic, or one exergonic and one endergonic? Why? Would you expect both processes to take place in the same way? Why?

# ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).

Research progress is very rapid in biochemistry, and the literature in the field is vast and growing. Many books appear each year, and a large number of primary research journals and review journals report on original research. References to this body of literature are available the website for each chapter. A particularly useful reference is *Scientific American*; its articles include general overviews of the topics discussed. It is available in most libraries. *Trends in Biochemical Sciences* and *Science* (a journal published weekly by the American Association for the Advancement of Science) are more advanced but can serve as primary sources of information about a given topic. In addition to material in print, a wealth

of information has become available in electronic form. *Science* regularly covers websites of interest and has its own website at <http://www.sciencemag.org>. Journals now appear on the Internet. Some require subscriptions, and many college and university libraries have subscriptions, making the journals available to students and faculty in this form. Others are free of charge. One, PubMed, is a service of the U.S. government. It lists articles in the biomedical sciences and has links to them. Its URL is <http://www.ncbi.nlm.nih.gov/PubMed>. Databases provide instant access to structures of proteins and nucleic acids. References will be given to electronic resources as well.



# Water: The Solvent for Biochemical Reactions

# 2

Life processes depend on the properties of water.

## 2.1 Water and Polarity

Water is the principal component of most cells. The geometry of the water molecule and its properties as a solvent play major roles in determining the properties of living systems.

The tendency of an atom to attract electrons to itself in a chemical bond (i.e., to become negative) is called **electronegativity**. Atoms of the same element, of course, share electrons equally in a bond—that is, they have equal electronegativity—but different elements do not necessarily have the same electronegativity. Oxygen and nitrogen are both highly electronegative, much more so than carbon and hydrogen (Table 2.1).

### What is polarity?

When two atoms with the same electronegativity form a bond, the electrons are shared equally between the two atoms. However, if atoms with differing electronegativity form a bond, the electrons are not shared equally and more of the negative charge is found closer to one of the atoms. In the O—H bonds in water, oxygen is more electronegative than hydrogen, so there is a higher probability that the bonding electrons are closer to the oxygen. The difference in electronegativity between oxygen and hydrogen gives rise to a *partial* positive and negative charge, usually pictured as  $\delta^+$  and  $\delta^-$ , respectively (Figure 2.1). Bonds such as this are called **polar bonds**. In situations in which the electronegativity difference is quite small, such as in the C—H bond in methane ( $\text{CH}_4$ ), the sharing of electrons in the bond is very nearly equal, and the bond is essentially **nonpolar**.

The bonds in a molecule may be polar, but the molecule itself can still be nonpolar because of its geometry. Carbon dioxide is an example. The two C=O bonds are polar, but because the  $\text{CO}_2$  molecule is linear, the attraction of the oxygen for the electrons in one bond is cancelled out by the equal and

## Chapter Outline

### 2.1 Water and Polarity

- What is polarity?
- Why do some chemicals dissolve in water while others do not?
- Why do oil and water mixed together separate into layers?

### 2.2 Hydrogen Bonds

- Why does water have such interesting and unique properties?

### 2.3 Acids, Bases, and pH

- What are acids and bases?
- What is pH?
- Why do we want to know the pH?

### 2.4 Titration Curves

### 2.5 Buffers

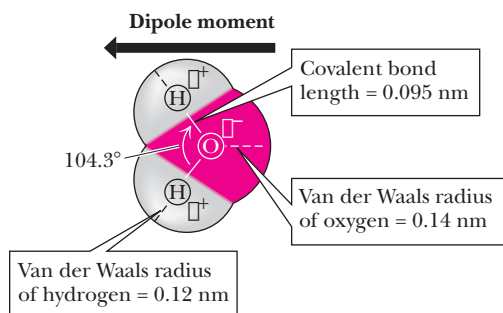
- How do buffers work?
- How do we choose a buffer?
- How do we make buffers in the laboratory?
- Are naturally occurring pH buffers present in living organisms?

**OWL** Online homework for this chapter may be assigned in OWL.

TABLE 2.1

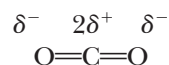
Electronegativities of Selected Elements	
Element	Electronegativity*
Oxygen	3.5
Nitrogen	3.0
Sulfur	2.6
Carbon	2.5
Phosphorus	2.2
Hydrogen	2.1

\* Electronegativity values are relative, and are chosen to be positive numbers ranging from less than 1 for some metals to 4 for fluorine.



■ **FIGURE 2.1** The structure of water. Oxygen has a partial negative charge, and the hydrogens have a partial positive charge. The uneven distribution of charge gives rise to the large dipole moment of water. The dipole moment in this figure points in the direction from negative to positive, the convention used by physicists and physical chemists; organic chemists draw it pointing in the opposite direction.

opposite attraction for the electrons by the oxygen on the other side of the molecule.



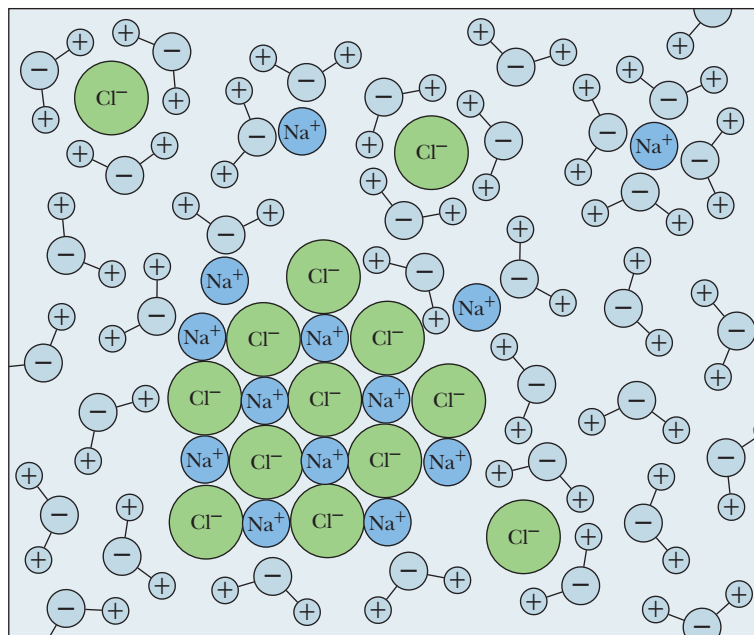
Water is a bent molecule with a bond angle of 104.3° (Figure 2.1), and the uneven sharing of electrons in the two bonds is not cancelled out as it is in CO<sub>2</sub>. The result is that the bonding electrons are more likely to be found at the oxygen end of the molecule than at the hydrogen end. Bonds with positive and negative ends are called **dipoles**.

## Solvent Properties of Water

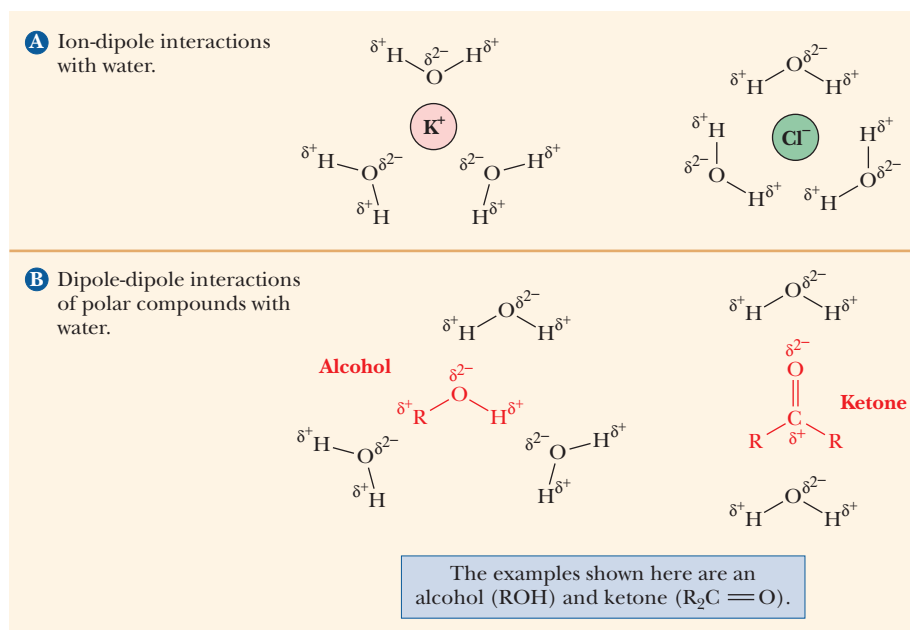
### Why do some chemicals dissolve in water while others do not?

The polar nature of water largely determines its solvent properties. *Ionic* compounds with full charges, such as potassium chloride (KCl, K<sup>+</sup> and Cl<sup>−</sup> in solution), and *polar* compounds with partial charges (i.e., dipoles), such as ethyl alcohol (C<sub>2</sub>H<sub>5</sub>OH) or acetone [(CH<sub>3</sub>)<sub>2</sub>C=O], tend to dissolve in water (Figures 2.2 and 2.3). The underlying physical principle is electrostatic attraction between unlike charges. The negative end of a water dipole attracts a positive ion or the positive end of another dipole. The positive end of a water molecule attracts a negative ion or the negative end of another dipole. The aggregate of unlike charges, held in proximity to one another because of electrostatic attraction, has a lower energy than would be possible if this interaction did not take place. The lowering of energy makes the system more stable and more likely to exist. These *ion–dipole* and *dipole–dipole* interactions are similar to the interactions between water molecules themselves in terms of the quantities of energy involved. Examples of polar compounds that dissolve easily in water are small organic molecules containing one or more electronegative atoms (e.g., oxygen or nitrogen), including alcohols, amines, and carboxylic acids. The attraction between the dipoles of these molecules and the water dipoles makes them tend to dissolve. Ionic and polar substances are referred to as **hydrophilic** (“water-loving,” from the Greek) because of this tendency.

■ **FIGURE 2.2** Hydration shells surrounding ions in solution. Unlike charges attract. The partial negative charge of water is attracted to positively charged ions. Likewise, the partial positive charge on the other end of the water molecule is attracted to negatively charged ions.







**FIGURE 2.3** Ion–dipole and dipole–dipole interactions. Ion–dipole and dipole–dipole interactions help ionic and polar compounds dissolve in water.

Hydrocarbons (compounds that contain only carbon and hydrogen) are nonpolar. The favorable ion–dipole and dipole–dipole interactions responsible for the solubility of ionic and polar compounds do not occur for nonpolar compounds, so these compounds tend not to dissolve in water. The interactions between nonpolar molecules and water molecules are weaker than dipolar interactions. The permanent dipole of the water molecule can induce a temporary dipole in the nonpolar molecule by distorting the spatial arrangements of the electrons in its bonds. Electrostatic attraction is possible between the induced dipole of the nonpolar molecule and the permanent dipole of the water molecule (a *dipole–induced dipole interaction*), but it is not as strong as that between permanent dipoles. Hence, its consequent lowering of energy is less than that produced by the attraction of the water molecules for one another. The association of nonpolar molecules with water is far less likely to occur than the association of water molecules with themselves.

A full discussion of why nonpolar substances are insoluble in water requires the thermodynamic arguments that we shall develop in Chapters 4 and 15. However, the points made here about intermolecular interactions will be useful background information for that discussion. For the moment, it is enough to know that it is less favorable thermodynamically for water molecules to be associated with nonpolar molecules than with other water molecules. As a result, nonpolar molecules do not dissolve in water and are referred to as **hydrophobic** (“water-hating,” from the Greek). Hydrocarbons in particular tend to sequester themselves from an aqueous environment. A nonpolar solid leaves undissolved material in water. A nonpolar liquid forms a two-layer system with water; an example is an oil slick. The interactions between nonpolar molecules are called **hydrophobic interactions** or, in some cases, **hydrophobic bonds**.

Table 2.2 gives examples of hydrophobic and hydrophilic substances.

### Why do oil and water mixed together separate into layers?

A single molecule may have both polar (hydrophilic) and nonpolar (hydrophobic) portions. Substances of this type are called **amphipathic**. A long-chain fatty acid having a polar carboxylic acid group and a long nonpolar hydrocarbon portion is a prime example of an amphipathic substance. The carboxylic acid

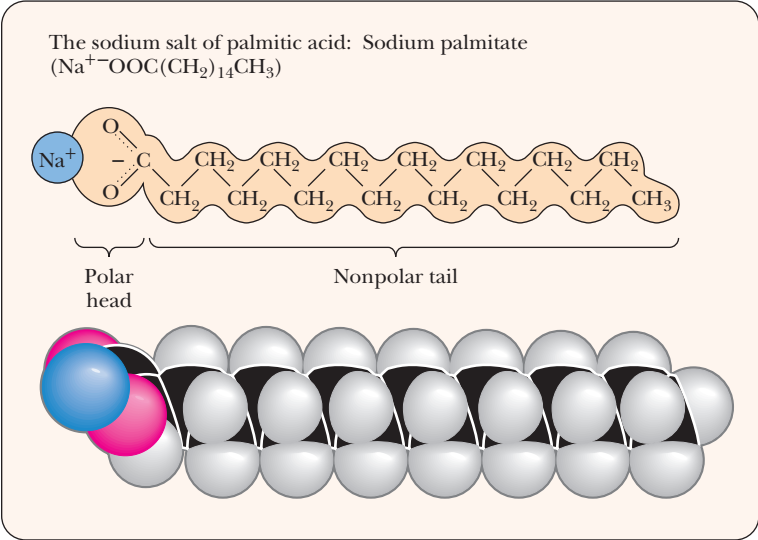
TABLE 2.2

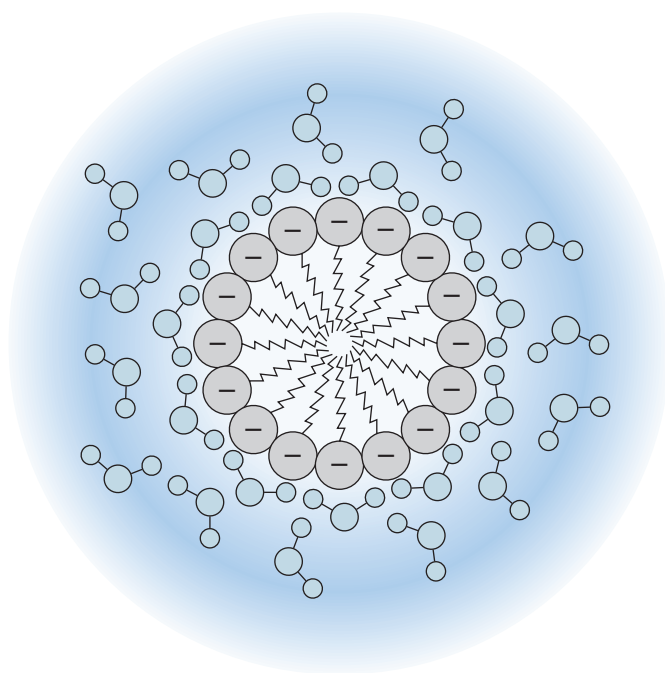
Examples of Hydrophobic and Hydrophilic Substances	
Hydrophilic	Hydrophobic
Polar covalent compounds (e.g., alcohols such as C <sub>2</sub> H <sub>5</sub> OH [ethanol] and ketones such as (CH <sub>3</sub> ) <sub>2</sub> C=O [acetone])	Nonpolar covalent compounds (e.g., hydrocarbons such as C <sub>6</sub> H <sub>14</sub> [hexane])
Sugars	Fatty acids, cholesterol
Ionic compounds (e.g., KCl)	
Amino acids, phosphate esters	

group, the “head” group, contains two oxygen atoms in addition to carbon and hydrogen; it is very polar and can form a carboxylate anion at neutral pH. The rest of the molecule, the “tail,” contains only carbon and hydrogen and is thus nonpolar (Figure 2.4). A compound such as this in the presence of water tends to form structures called *micelles*, in which the polar head groups are in contact with the aqueous environment and the nonpolar tails are sequestered from the water (Figure 2.5). A similar process is responsible for the separation of oil and water, such as you would see in Italian salad dressing. When shaken, initially the substances mix. Immediately thereafter you can see small spheres or oil droplets. As these float on water, they move to the top and coalesce into the oil layer.

Interactions between nonpolar molecules themselves are very weak and depend on the attraction between short-lived temporary dipoles and the dipoles they induce. A large sample of nonpolar molecules will always include some molecules with these temporary dipoles, which are caused by a momentary clumping of bonding electrons at one end of the molecule. A temporary dipole can induce another dipole in a neighboring molecule in the same way that a permanent dipole does. The interaction energy is low because the association is so short-lived. It is called a **van der Waals interaction** (also referred to as a *van der Waals bond*). The arrangement of molecules in cells strongly depends on the molecules’ polarity, as we saw with micelles.

■ **FIGURE 2.4** Sodium palmitate, an amphiphilic molecule. Amphiphilic molecules are frequently symbolized by a ball and zigzag line structure, where the ball represents the hydrophilic polar head and the zigzag line represents the nonpolar hydrophobic hydrocarbon tail.



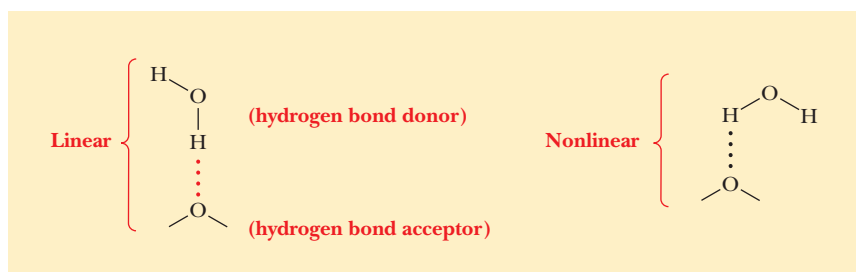


■ **FIGURE 2.5** Micelle formation by amphipathic molecules in aqueous solution. When micelles form, the ionized polar groups are in contact with the water, and the nonpolar parts of the molecule are protected from contact with the water.

## 2.2 Hydrogen Bonds

In addition to the interactions discussed in Section 2.1, another important type of noncovalent interaction exists: **hydrogen bonding**. Hydrogen bonding is of electrostatic origin and can be considered a special case of dipole–dipole interaction. When hydrogen is covalently bonded to a very electronegative atom such as oxygen or nitrogen, it has a partial positive charge due to the polar bond, a situation that does not occur when hydrogen is covalently bonded to carbon. This partial positive charge on hydrogen can interact with an unshared (nonbonding) pair of electrons (a source of negative charge) on another electronegative atom. All three atoms lie in a straight line, forming a hydrogen bond. This arrangement allows for the greatest possible partial positive charge on the hydrogen and, consequently, for the strongest possible interaction with the unshared pair of electrons on the second electronegative atom (Figure 2.6). The group comprising the electronegative atom that is covalently bonded to hydrogen is called the *hydrogen-bond donor*, and the electronegative atom that contributes the unshared pair of electrons to the interaction is the *hydrogen-bond acceptor*. The hydrogen is not covalently bonded to the acceptor in the usual description of hydrogen bonding.

Recent research has cast some doubt on this view, with experimental evidence to indicate some covalent character in the hydrogen bond.



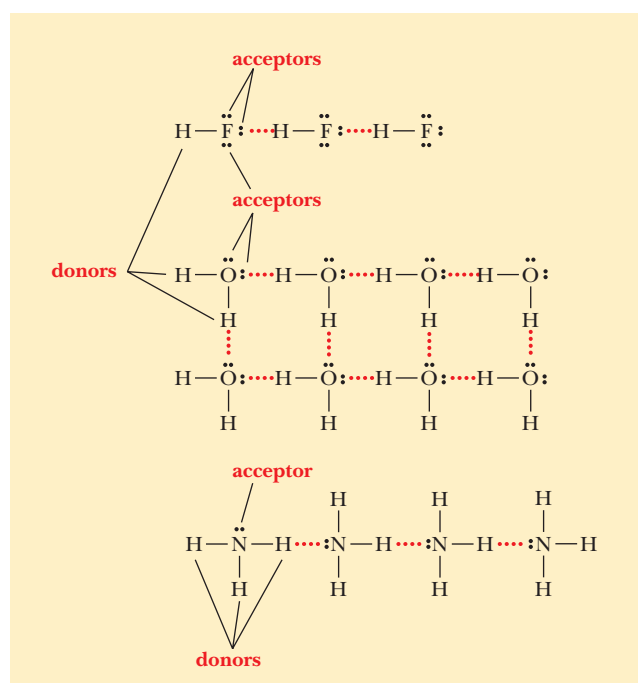
■ **FIGURE 2.6** A comparison of linear and nonlinear hydrogen bonds. Nonlinear bonds are weaker than bonds in which all three atoms lie in a straight line.

### Why does water have such interesting and unique properties?

A consideration of the hydrogen-bonding sites in HF, H<sub>2</sub>O, and NH<sub>3</sub> can yield some useful insights. Figure 2.7 shows that water constitutes an optimum situation in terms of the number of hydrogen bonds that each molecule can form. Water has two hydrogens to enter into hydrogen bonds and two unshared pairs of electrons on the oxygen to which other water molecules can be hydrogen-bonded. Each water molecule is involved in four hydrogen bonds—as a donor in two and as an acceptor in two. Hydrogen fluoride has only one hydrogen to enter into a hydrogen bond as a donor, but it has three unshared pairs of electrons on the fluorine that could bond to other hydrogens. Ammonia has three hydrogens to donate to a hydrogen bond but only one unshared pair of electrons, on the nitrogen.

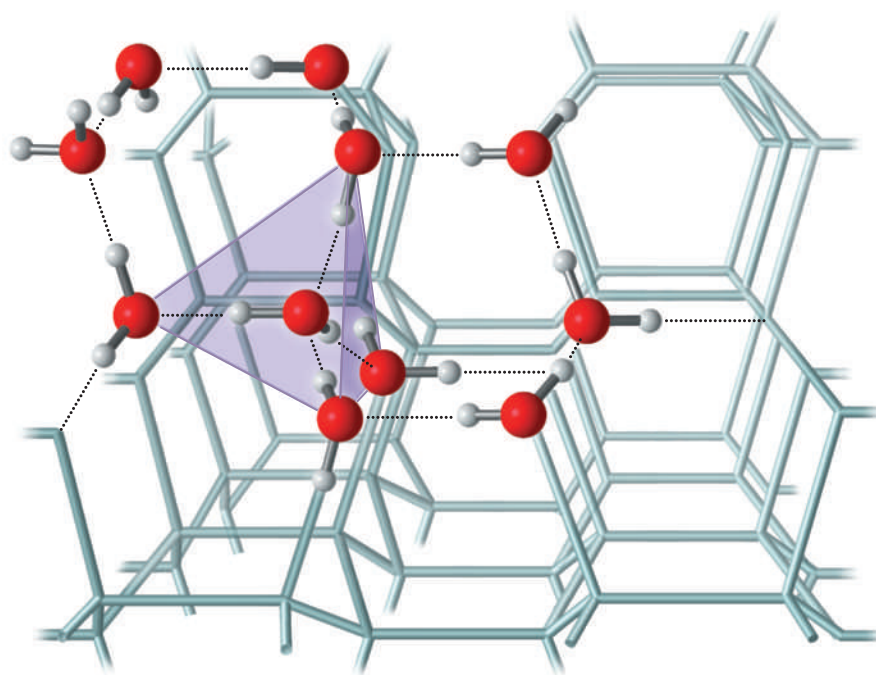
The geometric arrangement of hydrogen-bonded water molecules has important implications for the properties of water as a solvent. The bond angle in water is 104.3°, as was shown in Figure 2.1, and the angle between the unshared pairs of electrons is similar. The result is a tetrahedral arrangement of water molecules. Liquid water consists of hydrogen-bonded arrays that resemble ice crystals; each of these arrays can contain up to 100 water molecules. The hydrogen bonding between water molecules can be seen more clearly in the regular lattice structure of the ice crystal (Figure 2.8). There are several differences, however, between hydrogen-bonded arrays of this type in liquid water and the structure of ice crystals. In liquid water, hydrogen bonds are constantly breaking and new ones are constantly forming, with some molecules breaking off and others joining the cluster. A cluster can break up and re-form in 10<sup>-10</sup> to 10<sup>-11</sup> seconds in water at 25°C. An ice crystal, in contrast, has a more-or-less-stable arrangement of hydrogen bonds, and of course its number of molecules is many orders of magnitude greater than 100.

Hydrogen bonds are much weaker than normal covalent bonds. Whereas the energy required to break the O—H covalent bond is 460 kJ mol<sup>-1</sup> (110 kcal mol<sup>-1</sup>), the energy of hydrogen bonds in water is about 20 kJ mol<sup>-1</sup> (5 kcal mol<sup>-1</sup>) (Table 2.3). Even this comparatively small amount of energy is enough to affect the properties of water drastically, especially its melting point,



■ **FIGURE 2.7** Hydrogen-bonding sites. A comparison of the numbers of hydrogen-bonding sites in HF, H<sub>2</sub>O, and NH<sub>3</sub>. (Actual geometries are not shown.) Each HF molecule has one hydrogen-bond donor and three hydrogen-bond acceptors. Each H<sub>2</sub>O molecule has two donors and two acceptors. Each NH<sub>3</sub> molecule has three donors and one acceptor.





■ **FIGURE 2.8** Tetrahedral hydrogen bonding in  $\text{H}_2\text{O}$ . In an array of  $\text{H}_2\text{O}$  molecules in an ice crystal, each  $\text{H}_2\text{O}$  molecule is hydrogen-bonded to four others.

**TABLE 2.3**

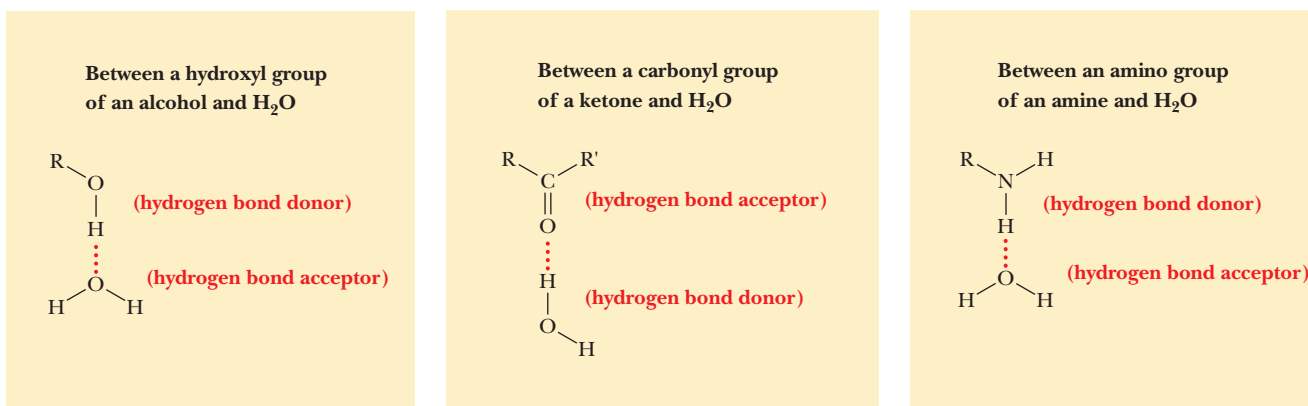
Some Bond Energies			
Type of Bond		Energy*	
		( $\text{kJ mol}^{-1}$ )	( $\text{kcal mol}^{-1}$ )
Covalent Bonds (Strong)	O—H	460	110
	H—H	416	100
	C—H	413	105
Noncovalent Bonds (Weaker)	Hydrogen bond	20	5
	Ion–dipole interaction	20	5
	Hydrophobic interaction	4–12	1–3
	Van der Waals interactions	4	1

\* Note that two units of energy are used throughout this text. The kilocalorie (kcal) is a commonly used unit in the biochemical literature. The kilojoule (kJ) is an SI unit and will come into wider use as time goes on. The kcal is the same as the “Calorie” reported on food labels.

**TABLE 2.4**

Comparison of Properties of Water, Ammonia, and Methane			
Substance	Molecular Weight	Melting Point ( $^{\circ}\text{C}$ )	Boiling Point ( $^{\circ}\text{C}$ )
Water ( $\text{H}_2\text{O}$ )	18.02	0.0	100.0
Ammonia ( $\text{NH}_3$ )	17.03	−77.7	−33.4
Methane ( $\text{CH}_4$ )	16.04	−182.5	−161.5

its boiling point, and its density relative to the density of ice. Both the melting point and the boiling point of water are significantly higher than would be predicted for a molecule of this size (Table 2.4). Other substances of about the same molecular weight, such as methane and ammonia, have much lower melting and boiling points. The forces of attraction between the molecules of these substances are weaker than the attraction between water molecules, because of



■ **FIGURE 2.9** Hydrogen bonding between polar groups and water.

the number and strength of their hydrogen bonds. The energy of this attraction must be overcome to melt ice or boil water.

Ice has a lower density than liquid water because the fully hydrogen bonded array in an ice crystal is less densely packed than that in liquid water. Liquid water is less extensively hydrogen-bonded and thus is denser than ice. Thus, ice cubes and icebergs float. Most substances contract when they freeze, but the opposite is true of water. In cold weather, the cooling systems of cars require antifreeze to prevent freezing and expansion of the water, which could crack the engine block. In laboratory procedures for cell fractionation, the same principle is used in a method of disrupting cells with several cycles of freezing and thawing. Finally, aquatic organisms can survive in cold climates because of the density difference between ice and liquid water; lakes and rivers freeze from top to bottom rather than vice versa.

Hydrogen bonding also plays a role in the behavior of water as a solvent. If a polar solute can serve as a donor or an acceptor of hydrogen bonds, not only can it form hydrogen bonds with water but it can also be involved in nonspecific dipole–dipole interactions. Figure 2.9 shows some examples. Alcohols, amines, carboxylic acids, and esters, as well as aldehydes and ketones, can all form hydrogen bonds with water, so they are soluble in water. It is difficult to overstate the importance of water to the existence of life on the Earth, and it is difficult to imagine life based on another solvent. The following Biochemical Connections box explores some of the implications of this statement.

### Other Biologically Important Hydrogen Bonds

Hydrogen bonds have a vital involvement in stabilizing the three-dimensional structures of biologically important molecules, including DNA, RNA, and proteins. The hydrogen bonds between complementary bases are one of the most striking characteristics of the double-helical structure of DNA (Section 9.3). Transfer RNA also has a complex three-dimensional structure characterized by hydrogen-bonded regions (Section 9.5). Hydrogen bonding in proteins gives rise to two important structures, the  $\alpha$ -helix and  $\beta$ -pleated sheet conformations. Both types of conformation are widely encountered in proteins (Section 4.3). Table 2.5 summarizes some of the most important kinds of hydrogen bonds in biomolecules.

TABLE 2.5

Examples of Major Types of Hydrogen Bonds Found in Biologically Important Molecules	
Bonding Arrangement	Molecules Where the Bond Occurs
$\text{—O—H} \cdots \text{O—}$ $\quad \quad \quad \text{H}$	H bond formed in $\text{H}_2\text{O}$
$\text{—O—H} \cdots \text{O}=\text{C} \diagup \diagdown$	Bonding of water to other molecules
$\text{N—H} \cdots \text{O—}$ $\quad \quad \quad \text{H}$	Important in protein and nucleic acid structures
$\text{N—H} \cdots \text{O}=\text{C} \diagup \diagdown$	
$\text{N—H} \cdots \text{N} \diagup \diagdown$	
$\text{N—H} \cdots \text{N} \text{ (in a ring structure)}$	

## Biochemical Connections CHEMISTRY

### How Basic Chemistry Affects Life: The Importance of the Hydrogen Bond

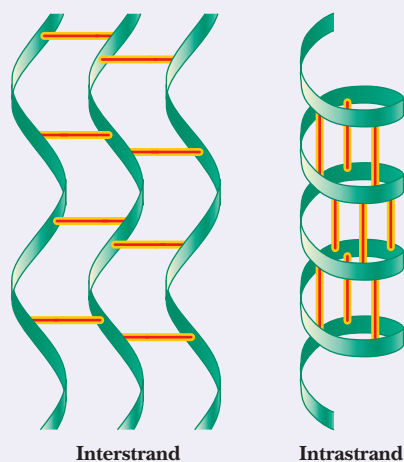
Many noted biochemists have speculated that the hydrogen bond is essential to the evolution of life. Just like carbon, polymers, and stereochemistry, it is one of the criteria that can be used to search for extraterrestrial life. Even though the individual hydrogen bond (H bond) is weak, the fact that so many H bonds can form means that collectively they can exert a *very* strong force. Virtually all the unique properties of water (high melting and boiling points, ice and density characteristics, and solvent potency) are a result of its ability to form many hydrogen bonds per molecule.

If we look at the solubility of a simple ion like  $\text{Na}^+$  or  $\text{Cl}^-$ , we find that water is attracted to these ions by polarity. In addition, other water molecules form H bonds with those surrounding water molecules, typically 20 or more water molecules per dissolved ion. When we consider a simple biomolecule such as glyceraldehyde, the H bonds start at the molecule itself. At least eight water molecules bind directly to the glyceraldehyde molecule, and then more water molecules bind to those eight.

The orderly and repetitive arrangement of hydrogen bonds in polymers determines their shape. The extended structures of cellulose and of peptides in a  $\beta$ -sheet allow for the formation of strong fibers through intrachain H bonding. Single helices (as in starch) and the  $\alpha$ -helices of proteins are stabilized by intrachain H bonds. Double and triple helices, as in DNA and collagen, involve H bonds between the two or three respective strands. Collagen contains several special amino acids that have an extra hydroxyl group; these allow for additional hydrogen bonds, which provide stability.

Hydrogen bonding is also fundamental to the specificity of transfer of genetic information. The complementary nature of the DNA double helix is assured by hydrogen bonds. The genetic code, both its specificity and its allowable variation, is a result of H bonds. Indeed, many compounds that cause genetic mutations work by altering the patterns of H bonding. For example, fluorouracil is often prescribed by dentists for cold sores (viral sores of the lip and mouth) because it causes mutations in the herpes simplex virus that causes the sores.

Types of hydrogen bonding in proteins



Hydrogen bonds between the strands of a DNA double helix



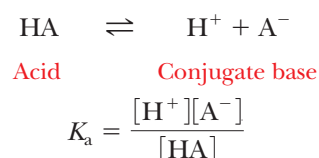
## 2.3 Acids, Bases, and pH

The biochemical behavior of many important compounds depends on their acid–base properties.

### What are acids and bases?

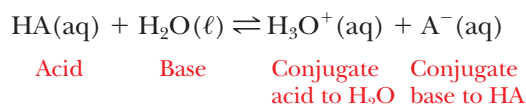
A biologically useful definition of an acid is a molecule that acts as a proton (hydrogen ion) donor. A base is similarly defined as a proton acceptor. How readily acids or bases lose or gain protons depends on the chemical nature of the compounds under consideration. The degree of dissociation of acids in water, for example, ranges from essentially complete dissociation for a strong acid to practically no dissociation for a very weak acid, and any intermediate value is possible.

It is useful to derive a numerical measure of **acid strength**, which is the amount of hydrogen ion released when a given amount of acid is dissolved in water. Such an expression, called the **acid dissociation constant**, or  $K_a$ , can be written for any acid, HA, that reacts according to the equation

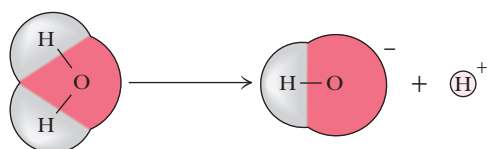


In this expression, the square brackets refer to molar concentration—that is, the concentration in moles per liter. For each acid, the quantity  $K_a$  has a fixed numerical value at a given temperature. This value is larger for more completely dissociated acids; thus the greater the  $K_a$ , the stronger the acid.

Strictly speaking, the preceding acid–base reaction is a proton-transfer reaction in which water acts as a base as well as the solvent. A more correct way to write the equation is as follows:



The notation (aq) refers to solutes in aqueous solution, whereas ( $\ell$ ) refers to water in the liquid state. It is well established that there are no “naked protons” (free hydrogen ions) in solution; even the hydronium ion ( $\text{H}_3\text{O}^+$ ) is an underestimate of the degree of hydration of hydrogen ion in aqueous solution. All solutes are extensively hydrated in aqueous solution. We will write the short form of equations for acid dissociation in the interest of simplicity, but the role of water should be kept in mind throughout our discussion.



■ FIGURE 2.10 The ionization of water.

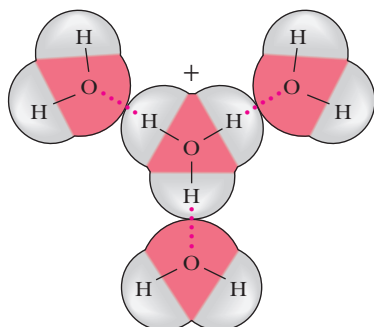
### What is pH?

The acid–base properties of water play an important part in biological processes because of the central role of water as a solvent. The extent of self-dissociation of water to hydrogen ion and hydroxide ion,



is small, but the fact that it takes place determines important properties of many solutes (Figure 2.10). Both the hydrogen ion ( $\text{H}^+$ ) and the hydroxide ion ( $\text{OH}^-$ ) are associated with several water molecules, as are all ions in aqueous solution, and the water molecule in the equation is itself part of a cluster of such molecules (Figure 2.11). It is especially important to have a quantitative estimate of the degree of dissociation of water. We can start with the expression

$$K_a = \frac{[\text{H}^+][\text{OH}^-]}{[\text{H}_2\text{O}]}$$



■ FIGURE 2.11 The hydration of hydrogen ion in water.

The molar concentration of pure water,  $[\text{H}_2\text{O}]$ , is quite large compared to any possible concentrations of solutes and can be considered a constant. (The numerical value is  $55.5\text{ M}$ , which can be obtained by dividing the number of grams of water in 1 L, 1000 g, by the molecular weight of water, 18 g/mol;  $1000/18 = 55.5\text{ M}$ .) Thus,

$$K_a = \frac{[\text{H}^+][\text{OH}^-]}{55.5}$$

$$K_a \times 55.5 = [\text{H}^+][\text{OH}^-] = K_w$$

A new constant,  $K_w$ , the **ion product constant for water**, has just been defined, where the concentration of water has been included in its value.

The numerical value of  $K_w$  can be determined experimentally by measuring the hydrogen ion concentration of pure water. The hydrogen ion concentration is also equal, by definition, to the hydroxide ion concentration because water is a monoprotic acid (one that releases a single proton per molecule). At  $25^\circ\text{C}$  in *pure* water,

$$[\text{H}^+] = 10^{-7}\text{ M} = [\text{OH}^-]$$

Thus, at  $25^\circ\text{C}$ , the numerical value of  $K_w$  is given by the expression

$$K_w = [\text{H}^+][\text{OH}^-] = (10^{-7})(10^{-7}) = 10^{-14}$$

This relationship, which we have derived for pure water, is valid for *any* aqueous solution, whether neutral, acidic, or basic.

The wide range of possible hydrogen ion and hydroxide ion concentrations in aqueous solution makes it desirable to define a quantity for expressing these concentrations more conveniently than by exponential notation. This quantity is called *pH* and is defined as

$$\text{pH} = -\log_{10} [\text{H}^+]$$

with the logarithm taken to the base 10. Note that, because of the logarithms involved, a difference of one pH unit implies a tenfold difference in hydrogen ion concentration,  $[\text{H}^+]$ . The pH values of some typical aqueous samples can be determined by a simple calculation.

### Apply Your Knowledge

#### pH Calculations

From time to time you will find practice sessions such as these in the chapters. These give you the opportunity to practice what was learned immediately.

Because in pure water  $[\text{H}^+] = 1 \times 10^{-7}\text{ M}$  and  $\text{pH} = 7.0$ , you can calculate the pH of the following aqueous solutions:

- $1 \times 10^{-3}\text{ M HCl}$
- $1 \times 10^{-4}\text{ M NaOH}$

Assume that the self-ionization of water makes a negligible contribution to the concentrations of hydronium ions and of hydroxide ions, which will typically be true unless the solutions are extremely dilute.

#### Solution

The key points in the approach to this problem are the definition of pH, which needs to be used in both parts, and the self-dissociation of water, needed in the second part.

- For  $1 \times 10^{-3}\text{ M HCl}$ ,  $[\text{H}_3\text{O}^+] = 1 \times 10^{-3}\text{ M}$ ; therefore,  $\text{pH} = 3$ .
- For  $1 \times 10^{-4}\text{ M NaOH}$ ,  $[\text{OH}^-] = 1 \times 10^{-4}\text{ M}$ . Because  $[\text{OH}^-][\text{H}_3\text{O}^+] = 1 \times 10^{-14}$ ,  $[\text{H}_3\text{O}^+] = 1 \times 10^{-10}\text{ M}$ ; therefore,  $\text{pH} = 10.0$ .

When a solution has a pH of 7, it is said to be *neutral*, like pure water. Acidic solutions have pH values lower than 7, and basic solutions have pH values higher than 7.



In biochemistry, most of the acids encountered are weak acids. These have a  $K_a$  well below 1. To avoid having to use numbers with large, negative exponents, a similar quantity,  $pK_a$ , has been defined by analogy with the definition of pH:

$$pK_a = -\log_{10} K_a$$

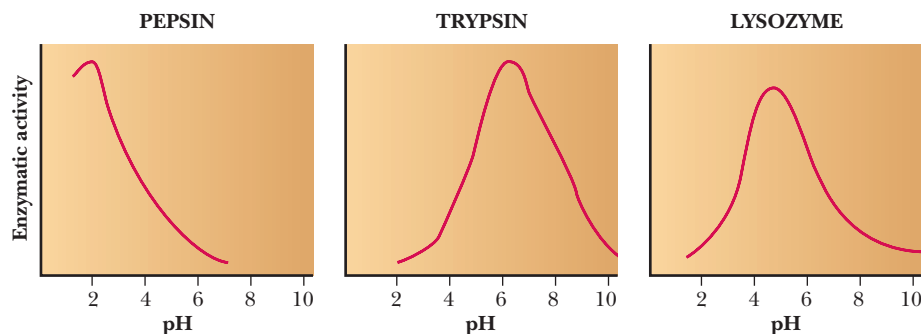
The  $pK_a$  is a more convenient numerical measure of acid strength. The smaller its value, the stronger the acid. This is the reverse of the situation with  $K_a$ , where larger values imply stronger acids (Table 2.6).

Why do we want to know the pH?

An equation connects the  $K_a$  of any weak acid with the pH of a solution containing both that acid and its conjugate base. This relationship has wide use in biochemistry, especially where it is necessary to control pH for optimum reaction conditions. Some reactions cannot take place if the pH varies from the optimum value. Important biological macromolecules lose activity at extremes of pH. Figure 2.12 shows how the activities of three enzymes are affected by pH. Note that each one has a peak activity that falls off rapidly as the pH is changed from the optimum. Also, some drastic physiological consequences can result from pH fluctuations in the body. Section 2.5 has more information about how pH can be controlled. To derive the involved

TABLE 2.6

Dissociation Constants of Some Acids				
Acid	HA	A <sup>−</sup>	$K_a$	$pK_a$
Pyruvic acid	CH <sub>3</sub> COCOOH	CH <sub>3</sub> C—COO <sup>−</sup>	$3.16 \times 10^{-3}$	2.50
Formic acid	HCOOH	HCOO <sup>−</sup>	$1.78 \times 10^{-4}$	3.75
Lactic acid	CH <sub>3</sub> CHOHCOOH	CH <sub>3</sub> CH—HCOO <sup>−</sup>	$1.38 \times 10^{-4}$	3.86
Benzoic acid	C <sub>6</sub> H <sub>5</sub> COOH	C <sub>6</sub> H <sub>5</sub> COO <sup>−</sup>	$6.46 \times 10^{-5}$	4.19
Acetic acid	CH <sub>3</sub> COOH	CH <sub>3</sub> COO <sup>−</sup>	$1.76 \times 10^{-5}$	4.76
Ammonium ion	NH <sub>4</sub> <sup>+</sup>	NH <sub>3</sub>	$5.6 \times 10^{-10}$	9.25
Oxalic acid (1)	HOOC—COOH	HOOC—COO <sup>−</sup>	$5.9 \times 10^{-2}$	1.23
Oxalic acid (2)	HOOC—COO <sup>−</sup>	<sup>−</sup> OOC—COO <sup>−</sup>	$6.4 \times 10^{-5}$	4.19
Malonic acid (1)	HOOC—CH <sub>2</sub> —COOH	HOOC—CH <sub>2</sub> —COO <sup>−</sup>	$1.49 \times 10^{-3}$	2.83
Malonic acid (2)	HOOC—CH <sub>2</sub> —COO <sup>−</sup>	<sup>−</sup> OOC—CH <sub>2</sub> —COO <sup>−</sup>	$2.03 \times 10^{-6}$	5.69
Malic acid (1)	HOOC—CH <sub>2</sub> —CHOH—COOH	HOOC—CH <sub>2</sub> —CHOH—COO <sup>−</sup>	$3.98 \times 10^{-4}$	3.40
Malic acid (2)	HOOC—CH <sub>2</sub> —CHOH—COO <sup>−</sup>	<sup>−</sup> OOC—CH <sub>2</sub> —CHOH—COO <sup>−</sup>	$5.5 \times 10^{-6}$	5.26
Succinic acid (1)	HOOC—CH <sub>2</sub> —CH <sub>2</sub> O—OOH	HOOC—CH <sub>2</sub> —CH <sub>2</sub> —COO <sup>−</sup>	$6.17 \times 10^{-5}$	4.21
Succinic acid (2)	HOOC—CH <sub>2</sub> —CH <sub>2</sub> —COO <sup>−</sup>	<sup>−</sup> OOC—CH <sub>2</sub> —CH <sub>2</sub> —COO <sup>−</sup>	$2.3 \times 10^{-6}$	5.63
Carbonic acid (1)	H <sub>2</sub> CO <sub>3</sub>	HCO <sub>3</sub> <sup>−</sup>	$4.3 \times 10^{-7}$	6.37
Carbonic acid (2)	HCO <sub>3</sub> <sup>−</sup>	CO <sub>3</sub> <sup>2−</sup>	$5.6 \times 10^{-11}$	10.20
Citric acid (1)	HOOC—CH <sub>2</sub> —C(OH)	HOOC—CH <sub>2</sub> —C(OH)	$8.14 \times 10^{-4}$	3.09
	(COOH) OCH <sub>2</sub> —COOH	(COOH) —CH <sub>2</sub> —COO <sup>−</sup>	$1.78 \times 10^{-5}$	4.75
Citric acid (2)	HOOC—CH <sub>2</sub> —C(OH) (COOH)	<sup>−</sup> OOC—CH <sub>2</sub> —C(OH)	$3.9 \times 10^{-6}$	5.41
	OCH <sub>2</sub> —COO <sup>−</sup>	(COOH) —CH <sub>2</sub> —COO <sup>−</sup>		
Citric acid (3)	<sup>−</sup> OOC—CH <sub>2</sub> —C(OH) (COOH)	<sup>−</sup> OOC—CH <sub>2</sub> —C(OH) (COO <sup>−</sup> )—		
	OCH <sub>2</sub> —COO <sup>−</sup>	CH <sub>2</sub> —COO <sup>−</sup>		
Phosphoric acid (1)	H <sub>3</sub> PO <sub>4</sub>	H <sub>2</sub> PO <sub>4</sub> <sup>−</sup>	$7.25 \times 10^{-3}$	2.14
Phosphoric acid (2)	H <sub>2</sub> PO <sub>4</sub> <sup>−</sup>	HPO <sub>4</sub> <sup>2−</sup>	$6.31 \times 10^{-8}$	7.20
Phosphoric acid (3)	HPO <sub>4</sub> <sup>2−</sup>	PO <sub>4</sub> <sup>3−</sup>	$3.98 \times 10^{-13}$	12.40



■ **FIGURE 2.12** pH versus enzymatic activity. Pepsin, trypsin, and lysozyme all have steep pH optimum curves. Pepsin has maximum activity under very acidic conditions, as would be expected for a digestive enzyme that is found in the stomach. Lysozyme has its maximum activity near pH 5, while trypsin is most active near pH 6.

equation, it is first necessary to take the logarithm of both sides of the  $K_a$  equation.

$$K_a = \frac{[\text{H}^+][\text{A}^-]}{[\text{HA}]}$$

$$\log K_a = \log [\text{H}^+] + \log \frac{[\text{A}^-]}{[\text{HA}]}$$

$$-\log [\text{H}^+] = -\log K_a + \log \frac{[\text{A}^-]}{[\text{HA}]}$$

We then use the definitions of pH and  $\text{p}K_a$ :

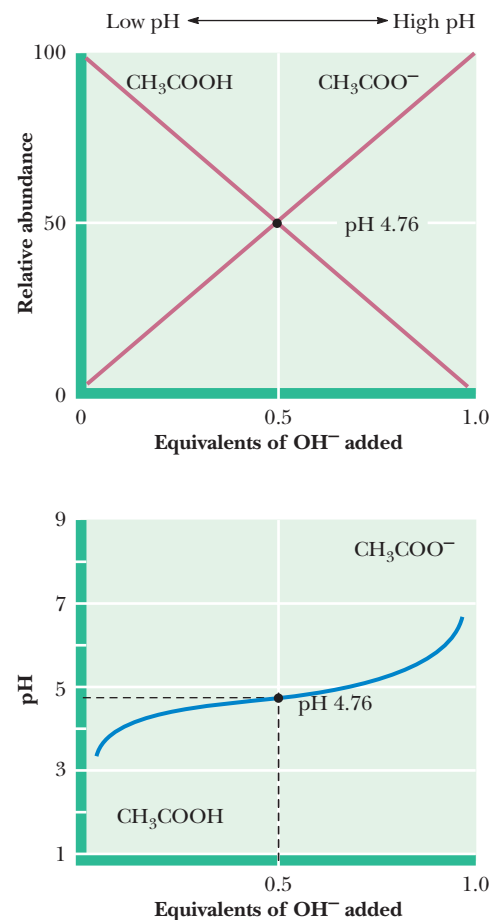
$$\text{pH} = \text{p}K_a + \log \frac{[\text{A}^-]}{[\text{HA}]}$$

This relationship is known as the **Henderson–Hasselbalch equation** and is useful in predicting the properties of buffer solutions used to control the pH of reaction mixtures. When buffers are discussed in Section 2.5, we will be interested in the situation in which the concentration of acid,  $[\text{HA}]$ , and the concentration of the conjugate base,  $[\text{A}^-]$ , are equal ( $[\text{HA}] = [\text{A}^-]$ ). The ratio  $[\text{A}^-]/[\text{HA}]$  is then equal to 1, and the logarithm of 1 is equal to zero. Therefore, when a solution contains equal concentrations of a weak acid and its conjugate base, the pH of that solution equals the  $\text{p}K_a$  value of the weak acid.

## 2.4 Titration Curves

When base is added to a sample of acid, the pH of the solution changes. A **titration** is an experiment in which measured amounts of base are added to a measured amount of acid. It is convenient and straightforward to follow the course of the reaction with a pH meter. The point in the titration at which the acid is exactly neutralized is called the **equivalence point**.

If the pH is monitored as base is added to a sample of acetic acid in the course of a titration, an inflection point in the titration curve is reached when the pH equals the  $\text{p}K_a$  of acetic acid (Figure 2.13). As we saw in our discussion of the Henderson–Hasselbalch equation, a pH value equal to the  $\text{p}K_a$  corresponds to a mixture with equal concentrations of the weak acid and its conjugate base—in this case, acetic acid and acetate ion, respectively. The pH at the inflection point is 4.76, which is the  $\text{p}K_a$  of acetic acid. The inflection point occurs when 0.5 mol of base has been added for each mole of acid present. Near the inflection point, the pH changes very little as more base is added.



■ **FIGURE 2.13** Titration curve for acetic acid. Note that there is a region near the  $\text{p}K_a$  at which the titration curve is relatively flat. In other words, the pH changes very little as base is added in this region of the titration curve.

When 1 mol of base has been added for each mole of acid, the equivalence point is reached, and essentially all the acetic acid has been converted to acetate ion. (See Question 44 at the end of this chapter.) Figure 2.13 also plots the relative abundance of acetic acid and acetate ion with increasing additions of NaOH. Notice that the percentage of acetic acid plus the percentage of acetate ion adds up to 100%. The acid (acetic acid) is progressively converted to its conjugate base (acetate ion) as more NaOH is added and the titration proceeds. It can be helpful to keep track of the percentages of a conjugate acid and base in this way to understand the full significance of the reaction taking place in a titration. The form of the curves in Figure 2.13 represents the behavior of any monoprotic weak acid, but the value of the  $pK_a$  for each individual acid determines the pH values at the inflection point and at the equivalence point.

### Apply Your Knowledge

#### Calculating pH Values for Weak Acids and Bases

Calculate the relative amounts of acetic acid and acetate ion present at the following points when 1 mol of acetic acid is titrated with sodium hydroxide. Also use the Henderson–Hasselbalch equation to calculate the values of the pH at these points. Compare your results with Figure 2.13.

- 0.1 mol of NaOH is added
- 0.3 mol of NaOH is added
- 0.5 mol of NaOH is added
- 0.7 mol of NaOH is added
- 0.9 mol of NaOH is added

#### Solution

We approach this problem as an exercise in stoichiometry. There is a 1:1 ratio of moles of acid reacted to moles of base added. The difference between the original number of moles of acid and the number reacted is the number of moles of acid remaining. These are the values to be used in the numerator and denominator, respectively, of the Henderson–Hasselbalch equation.

- When 0.1 mol of NaOH is added, 0.1 mol of acetic acid reacts with it to form 0.1 mol of acetate ion, leaving 0.9 mol of acetic acid. The composition is 90% acetic acid and 10% acetate ion.

$$\text{pH} = \text{p}K_a + \log \frac{0.1}{0.9}$$

$$\text{pH} = 4.76 + \log \frac{0.1}{0.9}$$

$$\text{pH} = 4.76 - 0.95$$

$$\text{pH} = 3.81$$

- When 0.3 mol of NaOH is added, 0.3 mol of acetic acid reacts with it to form 0.3 mol of acetate ion, leaving 0.7 mol of acetic acid. The composition is 70% acetic acid and 30% acetate ion.

$$\text{pH} = \text{p}K_a + \log \frac{0.3}{0.7}$$

$$\text{pH} = 4.39$$

- When 0.5 mol of NaOH is added, 0.5 mol of acetic acid reacts with it to form 0.5 mol of acetate ion, leaving 0.5 mol of acetic acid. The composition is 50% acetic acid and 50% acetate ion.

$$\text{pH} = \text{p}K_a + \log \frac{0.5}{0.5}$$

$$\text{pH} = 4.76$$

Note that this one is possible without doing much math. We know that when the  $[HA] = [A^-]$ , the  $pH = pK_a$ . Therefore, the minute we saw that we added 0.5 mol of NaOH to 1 mol of acetic acid, we knew that we had added enough NaOH to convert half of the acid to the conjugate base form. Therefore the pH has to be equal to the  $pK_a$ .

- d. When 0.7 mol of NaOH is added, 0.7 mol of acetic acid reacts with it to form 0.7 mol of acetate ion, leaving 0.3 mol of acetic acid. The composition is 30% acetic acid and 70% acetate ion.

$$pH = pK_a + \log \frac{0.7}{0.3}$$

$$pH = 5.13$$

- e. When 0.9 mol of NaOH is added, 0.9 mol of acetic acid reacts with it to form 0.9 mol of acetate ion, leaving 0.1 mol of acetic acid. The composition is 10% acetic acid and 90% acetate ion.

$$pH = pK_a + \log \frac{0.9}{0.1}$$

$$pH = 5.71$$

Table 2.6 lists values for the acid dissociation constant,  $K_a$ , and for the  $pK_a$  for a number of acids. Note that these acids are categorized into three groups. The first group consists of monoprotic acids, which release one hydrogen ion and have a single  $K_a$  and  $pK_a$ . The second group consists of diprotic acids, which can release two hydrogen ions and have two  $K_a$  values and two  $pK_a$  values. The third group consists of polyprotic acids, which can release more than two hydrogen ions. The two examples of polyprotic acids given here, citric acid and phosphoric acid, can release three hydrogen ions and have three  $K_a$  values and three  $pK_a$  values. Amino acids and peptides, the subject of Chapter 3, behave as diprotic and polyprotic acids; we shall see examples of their titration curves later. Here is a way to keep track of protonated and deprotonated forms of acids and their conjugate bases, and this can be particularly useful with diprotic and polyprotic acids. When the pH of a solution is less than the  $pK_a$  of an acid, the protonated form predominates. (Remember that the definition of pH includes a negative logarithm.) When the pH of a solution is greater than the  $pK_a$  of an acid, the deprotonated (conjugate base) form predominates.

$$pH < pK_a$$

$H^+$  on, substance protonated

$$pH > pK_a$$

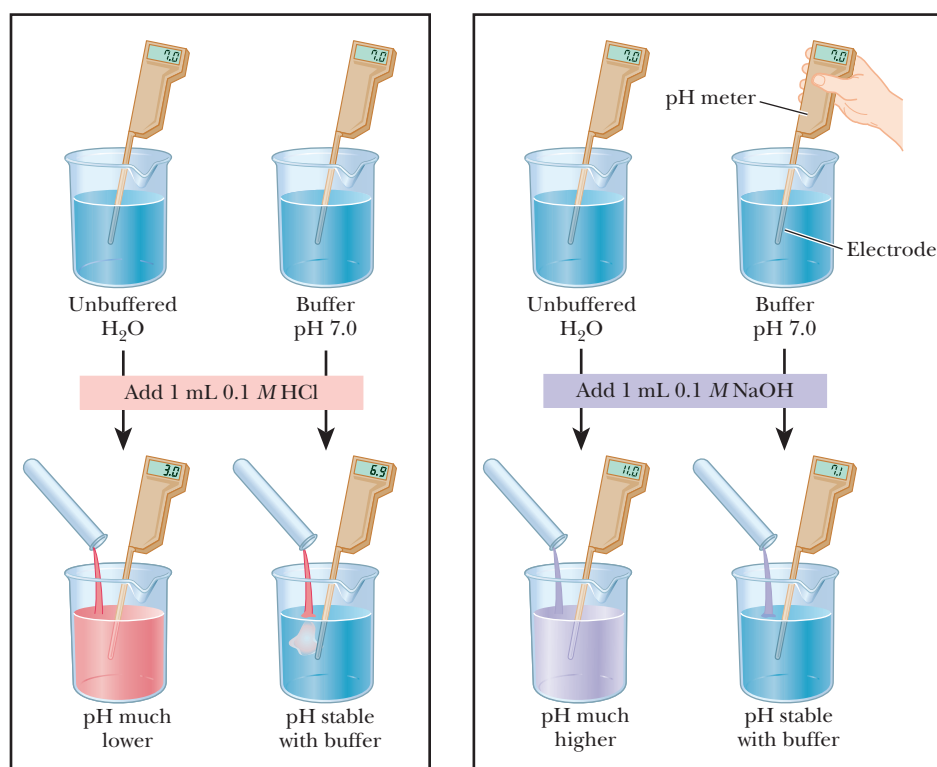
$H^+$  off, substance deprotonated

## 2.5 Buffers

A **buffer** is something that resists change. In terms of acid and base chemistry, a **buffer solution** tends to resist change in pH when small to moderate amounts of a strong acid or strong base are added. A buffer solution consists of a mixture of a weak acid and its conjugate base.

### How do buffers work?

Let us compare the changes in pH that occur on the addition of equal amounts of strong acid or strong base to pure water at pH 7 and to a buffer solution at pH 7. If 1.0 mL of 0.1 M HCl is added to 99.0 mL of pure water, the pH drops



■ **FIGURE 2.14 Buffering.** Acid is added to the two beakers on the left. The pH of unbuffered H<sub>2</sub>O drops dramatically while that of the buffer remains stable. Base is added to the two beakers on the right. The pH of the unbuffered water rises drastically while that of the buffer remains stable.

drastically. If the same experiment is conducted with 0.1 M NaOH instead of 0.1 M HCl, the pH rises drastically (Figure 2.14).

Let's start by calculating what happens when we add 1 mL of 0.1 M HCl to 99.0 mL of pure water.

Because HCl is a strong acid, we will assume that 0.1 M HCl dissociates completely to give 0.1 M H<sub>3</sub>O<sup>+</sup>. If we have 1 mL of the acid, we calculate the amount of H<sub>3</sub>O<sup>+</sup> as follows:

$$1 \text{ mL} = 1 \times 10^{-3} \text{ L}$$

$$1 \times 10^{-3} \text{ L} \times 0.1 \text{ mol/L} = 1 \times 10^{-4} \text{ mol H}_3\text{O}^+$$

Therefore,  $1 \times 10^{-4}$  mol of H<sub>3</sub>O<sup>+</sup> are diluted to a final volume of 100 mL or 0.1 L, because 1 mL was added to 99 mL. The final concentration of H<sub>3</sub>O<sup>+</sup> is calculated as follows:

$$1 \times 10^{-4} \text{ mol H}_3\text{O}^+ / 0.1 \text{ L} = 1 \times 10^{-3} \text{ M}$$

The pH is then calculated based on the definition:

$$\text{pH} = -\log [\text{H}_3\text{O}^+] = -\log (1 \times 10^{-3}) = 3$$

If we add 1 mL of 0.1 M NaOH, the calculations are done similarly to generate the concentration of [OH<sup>−</sup>], which also equals  $1 \times 10^{-3}$  M because we used the same concentration and same volume of base.

The [H<sub>3</sub>O<sup>+</sup>] is then calculated using the concentration of OH<sup>−</sup> and the water equation:

$$[\text{OH}^-][\text{H}_3\text{O}^+] = 1 \times 10^{-14}$$

$$[\text{H}_3\text{O}^+] = 1 \times 10^{-14} / [\text{OH}^-] = 1 \times 10^{-14} / 1 \times 10^{-3} = 1 \times 10^{-11}$$



Finally, the pH is calculated:

$$\text{pH} = -\log (1 \times 10^{-11}) = 11$$

The results are different when 99.0 mL of buffer solution is used instead of pure water. A solution that contains the monohydrogen phosphate and dihydrogen phosphate ions,  $\text{HPO}_4^{2-}$  and  $\text{H}_2\text{PO}_4^-$ , in suitable proportions can serve as such a buffer. The Henderson–Hasselbalch equation can be used to calculate the  $\text{HPO}_4^{2-}/\text{H}_2\text{PO}_4^-$  ratio that corresponds to pH 7.0, as shown in the Apply Your Knowledge box below.

### Apply Your Knowledge

#### Using the Henderson–Hasselbalch Equation

First, convince yourself that the proper ratio of  $\text{A}^-/\text{HA}$  for pH 7.00 is 0.63 parts  $\text{HPO}_4^{2-}$  to 1 part  $\text{H}_2\text{PO}_4^-$  by doing the calculation.

#### Solution

Use the Henderson–Hasselbalch equation with pH = 7.00 and  $\text{p}K_a = 7.20$ .

$$\begin{aligned}\text{pH} &= \text{p}K_a + \log \frac{[\text{A}^-]}{[\text{HA}]} \\ 7.00 &= 7.20 + \log \frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]} \\ -0.20 &= \log \frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]} \\ \frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]} &= \text{antilog } -0.20 = 0.63\end{aligned}$$

For purposes of illustration, let us consider a solution in which the concentrations are  $[\text{HPO}_4^{2-}] = 0.063 \text{ M}$  and  $[\text{H}_2\text{PO}_4^-] = 0.10 \text{ M}$ ; this gives the conjugate base/weak acid ratio of 0.63 seen previously. If 1.0 mL of 0.10 M HCl is added to 99.0 mL of the buffer, the reaction



takes place, and almost all the added  $\text{H}^+$  will be used up. The concentrations of  $[\text{HPO}_4^{2-}]$  and  $[\text{H}_2\text{PO}_4^-]$  will change, and the new concentrations can be calculated.

Concentrations (mol/L)			
	$[\text{HPO}_4^{2-}]$	$[\text{H}^+]$	$[\text{H}_2\text{PO}_4^-]$
Before addition of HCl	0.063	$1 \times 10^{-7}$	0.10
HCl added—no reaction yet	0.063	$1 \times 10^{-3}$	0.10
After HCl reacts with $\text{HPO}_4^{2-}$	0.062	To be found	0.101

The new pH can then be calculated using the Henderson–Hasselbalch equation and the phosphate ion concentrations. The appropriate  $\text{p}K_a$  is 7.20 (Table 2.6).

$$\begin{aligned}\text{pH} &= \text{p}K_a + \log \frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]} \\ \text{pH} &= 7.20 + \log \frac{0.062}{0.101} \\ \text{pH} &= 6.99\end{aligned}$$

The new pH is 6.99, a much smaller change than in the unbuffered pure water (Figure 2.14). Similarly, if 1.0 mL of 0.1 M NaOH is used, the same reaction takes place as in a titration:



Almost all the added  $\text{OH}^-$  is used up, but a small amount remains. Because this buffer is an aqueous solution, it is still true that  $K_w = [\text{H}^+][\text{OH}^-]$ . The increase in hydroxide ion concentration implies that the hydrogen ion concentration decreases and that the pH increases. Use the Henderson–Hasselbalch equation to calculate the new pH and to convince yourself that the result is  $\text{pH} = 7.01$ , again a much smaller change in pH than took place in pure water (Figure 2.14). Many biological reactions will not take place unless the pH remains within fairly narrow limits, and, as a result, buffers have great practical importance in the biochemistry laboratory.

### Apply Your Knowledge

#### How do Buffers Work?

Calculate the pH if you added 3 mL of 0.1 M HCl to (a) 97 mL of pure water at pH 7.0 or (b) 100 mL of the same phosphate buffer at pH 7.0 from above.

#### Solution

(a) 3 mL = 0.003 L of 0.1 M HCl.

$$0.003 \text{ L} \times 0.1 \text{ mol/L} = 0.0003 \text{ moles } \text{H}^+$$

The 0.0003 moles  $\text{H}^+$  are in a final volume of 100 mL or 0.1 L

$$0.0003 \text{ moles}/0.1 \text{ L} = 0.003 \text{ M } \text{H}^+$$

$$\text{pH} = -\log 0.003 = 2.52$$

(b) From the description previously, we know that the phosphate buffer at pH 7.0 is composed of  $\text{HPO}_4^{2-} = 0.063 \text{ M}$  and  $\text{H}_2\text{PO}_4^- = 0.1 \text{ M}$ . If there are 100 mL of the buffer, then the moles  $\text{HA} = 0.01$  and the moles of  $\text{A}^- = 0.0063$ .

When we add 0.0003 moles of  $\text{H}^+$  then the following reaction occurs:

$\text{HPO}_4^{2-} + \text{H}^+ \rightarrow \text{H}_2\text{PO}_4^-$ , until one of the reagents runs out. In this case, it will be the  $\text{H}^+$ .

The new amount of  $\text{H}_2\text{PO}_4^-$  will be  $0.01 + 0.0003 = 0.0103$

The new amount of  $\text{HPO}_4^{2-}$  will be  $0.0063 - 0.0003 = 0.006$

The pH is calculated using the Henderson–Hasselbalch equation:

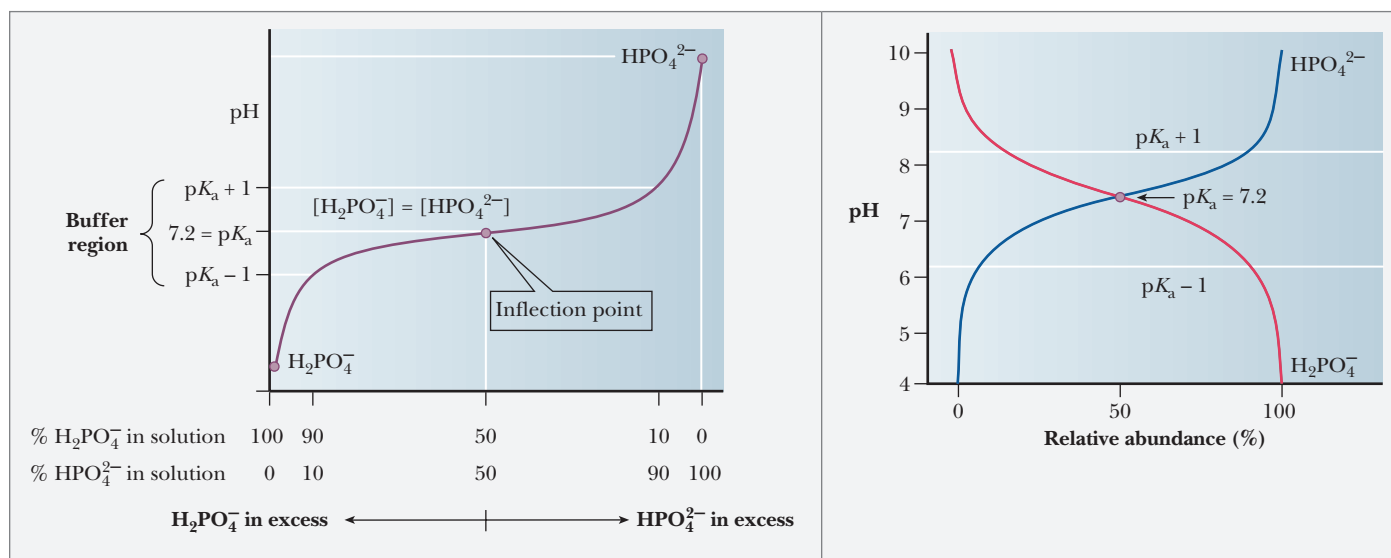
$$\text{pH} = 7.2 + \log (0.006/0.0103) = 6.97$$

### How do we choose a buffer?

A consideration of titration curves can give insight into how buffers work (Figure 2.15a). The pH of a sample being titrated changes very little in the vicinity of the inflection point of a titration curve. Also, at the inflection point, half the amount of acid originally present has been converted to the conjugate base. The second stage of ionization of phosphoric acid,



was the basis of the buffer just used as an example. The pH at the inflection point of the titration is 7.20, a value numerically equal to the  $\text{p}K_a$  of the dihydrogen phosphate ion. At this pH, the solution contains equal concentrations of the dihydrogen phosphate ions and monohydrogen phosphate ions, the acid and base forms. Using the Henderson–Hasselbalch equation, we can calculate the ratio of the conjugate base form to the conjugate acid form for any pH when we know the  $\text{p}K_a$ . For example, if we choose a pH of 8.2 for a buffer composed of  $\text{H}_2\text{PO}_4^-$  and  $\text{HPO}_4^{2-}$ , we can solve for the ratio



**A** The titration curve of  $\text{H}_2\text{PO}_4^-$ , showing the buffer region for the  $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}$  pair.

**B** Relative abundance of  $\text{H}_2\text{PO}_4^-$  and  $\text{HPO}_4^{2-}$ .

**FIGURE 2.15** The relationship between the titration curve and buffering action in  $\text{H}_2\text{PO}_4^-$ .

$$\text{pH} = \text{p}K_a + \log \frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]}$$

$$8.2 = 7.2 + \log \frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]}$$

$$1 = \log \frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]}$$

$$\frac{[\text{HPO}_4^{2-}]}{[\text{H}_2\text{PO}_4^-]} = 10$$

Thus, when the pH is one unit higher than the  $\text{p}K_a$ , the ratio of the conjugate base form to the conjugate acid form is 10. When the pH is two units higher than the  $\text{p}K_a$ , the ratio is 100, and so on. Table 2.7 shows this relationship for several increments of pH value.

A buffer solution can maintain the pH at a relatively constant value because of the presence of appreciable amounts of both the acid and its conjugate base. This condition is met at pH values at or near the  $\text{p}K_a$  of the acid. If  $\text{OH}^-$  is added, an appreciable amount of the acid form of the buffer is present in solution to react with the added base. If  $\text{H}^+$  is added, an appreciable amount of the basic form of the buffer also is present to react with the added acid.

The  $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}$  pair is suitable as a buffer near pH 7.2, and the  $\text{CH}_3\text{COOH}/\text{CH}_3\text{COO}^-$  pair is suitable as a buffer near pH 4.76. At pH values below the  $\text{p}K_a$ , the acid form predominates, and at pH values above the  $\text{p}K_a$ , the basic form predominates. The plateau region in a titration curve, where the pH does not change rapidly, covers a pH range extending approximately one pH unit on each side of the  $\text{p}K_a$ . Thus, the buffer is effective within a range of about two pH units (Figure 2.15b).

In many biochemical studies, a strict pH range must be maintained in order for the experiment to be successful. Using our knowledge of the range of an effective buffer compared to its  $\text{p}K_a$ , we can select an appropriate buffer. If we were doing an experiment and needed the pH to be 7.2, we might select the  $\text{H}_2\text{PO}_4^-/\text{HPO}_4^{2-}$  pair to be our buffer. If we wanted a pH near 9.0, we would look at tables of buffers to find one with a  $\text{p}K_a$  close to nine. The following Biochemical Connections box goes into greater detail on buffer selection.

TABLE 2.7

pH Values and Base/Acid Ratios for Buffers	
If the pH equals	The ratio of base form/acid form equals
$pK_a - 3$	1/1000
$pK_a - 2$	1/100
$pK_a - 1$	1/10
$pK_a$	1/1
$pK_a + 1$	10/1
$pK_a + 2$	100/1
$pK_a + 3$	1000/1

The condition that a buffer contains appreciable amounts of both a weak acid and its conjugate base applies both to the ratio of the two forms and to the absolute amount of each present in a given solution. If a buffer solution contained a suitable ratio of acid to base, but very low concentrations of both, it would take very little added acid to use up all of the base form, and vice versa. A buffer solution with low concentrations of both the acid and base forms is said to have a low **buffering capacity**. A buffer that contains greater amounts of both acid and base has a higher buffering capacity.

### How do we make buffers in the laboratory?

When we study buffers in theory, we often use the Henderson–Hasselbalch equation and do many calculations concerning ratios of conjugate base form to conjugate acid form. In practice, however, making a buffer is much easier. To have a buffer, all that is necessary are the two forms of the buffer present in the solution at reasonable quantities. This situation can be obtained by adding predetermined amounts of the conjugate base form ( $A^-$ ) to the acid form (HA), or we could start with one and create the other, which is how it is usually done in practice. Remember that HA and  $A^-$  are interconverted by adding strong acid or strong base (Figure 2.16). To make a buffer, we could start with the HA form and add NaOH until the pH is correct, as determined by a pH meter. We could also start with  $A^-$  and add HCl until the pH is correct. Depending on the relationship of the pH we desire to the  $pK_a$  of the buffer, it may be more convenient to start with one than the other. For example, if we are making an acetic acid/acetate buffer at pH 5.7, it would make more sense to start with the  $A^-$  form and to add a small amount of HCl to bring the pH down to 5.7, rather than to start with HA and to add much more NaOH to bring the pH up past the  $pK_a$ .

### Are naturally occurring pH buffers present in living organisms?

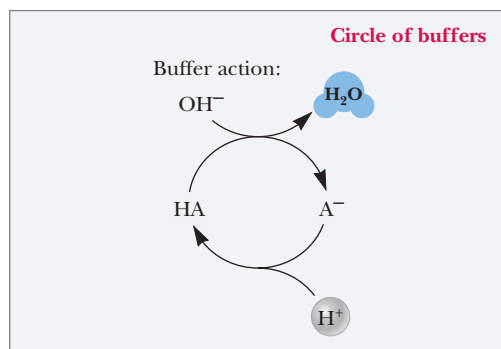
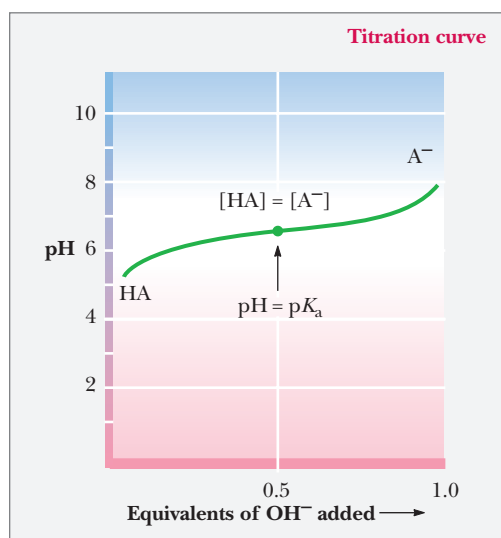
Up until now, we have been considering buffers from the perspective of a chemist trying to control an experiment. However, the real importance of buffers is that they are critical to life. Buffer systems in living organisms and in the laboratory are based on many types of compounds. Because physiological pH in most organisms stays around 7, it might be expected that the phosphate buffer system would be widely used in living organisms. This is the case where phosphate ion concentrations are high enough for the buffer to be effective, as in most intracellular fluids.

The  $H_2PO_4^-/HPO_4^{2-}$  pair is the principal buffer in cells. In blood, phosphate ion levels are inadequate for buffering, and a different system operates.

The buffering system in blood is based on the dissociation of carbonic acid ( $H_2CO_3$ ):



where the  $pK_a$  of  $H_2CO_3$  is 6.37. The pH of human blood, 7.4, is near the end of the buffering range of this system, but another factor enters into the situation.



■ **FIGURE 2.16** Two ways of looking at buffers. In the titration curve, we see that the pH varies only slightly near the region in which  $[HA] = [A^-]$ . In the circle of buffers, we see that adding  $OH^-$  to the buffer converts HA to  $A^-$ . Adding  $H^+$  converts  $A^-$  to HA.

# Biochemical Connections BUFFER CHEMISTRY

## Buffer Selection

Much of biochemistry is studied by carrying out enzymatic reactions in a test tube or *in vitro* (literally, in glass). Such reactions are usually buffered to maintain a constant pH. Similarly, virtually all methods for enzyme isolation, and even for growth of cells in tissue culture, use buffered solutions. The following criteria are typical for selecting a buffer for a biochemical reaction.

1. Suitable  $pK_a$  for the buffer.
2. No interference with the reaction or with the assay.
3. Suitable ionic strength of the buffer.
4. No precipitation of reactants or products due to presence of the buffer.
5. Nonbiological nature of the buffer.

The rule of thumb is that the  $pK_a$  should be  $\pm 1$  pH unit from the pH of the reaction;  $\pm \frac{1}{2}$  pH unit is even better. Although the perfect generic buffer would have a pH equal to its  $pK_a$ , if the reaction is known to produce an acidic product, it is advantageous if the  $pK_a$  is below the reaction pH, because then the buffer capacity increases as the reaction proceeds.

Sometimes a buffer can interfere with a reaction or with the assay method. For example, a reaction that requires or produces phosphate or  $\text{CO}_2$  may be inhibited if too much phosphate or carbonate is present in the reaction mixture. Even the counterion

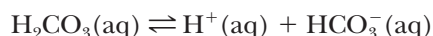
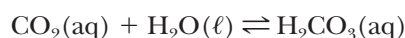
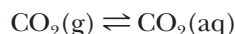
may be important. Typically a phosphate or carbonate buffer is prepared from the  $\text{Na}^+$  or  $\text{K}^+$  salt. Because many enzymes that react with nucleic acids are activated by one of these two ions and inhibited by the other, the choice of  $\text{Na}^+$  or  $\text{K}^+$  for a counterion could be critical. A buffer can also affect the spectrophotometric determination of a colored assay product.

If a buffer has a poor buffering capacity at the desired pH, its efficiency can often be increased by increasing the concentration; however, many enzymes are sensitive to high salt concentration. Beginning students in biochemistry often have difficulty with enzyme isolations and assays because they fail to appreciate the sensitivity of many enzymes. Fortunately, to minimize this problem, most beginning biochemistry laboratory manuals call for the use of enzymes that are very stable.

A buffer may cause precipitation of an enzyme or even of a metallic ion that may be a cofactor for the reaction. For example, many phosphate salts of divalent cations are only marginally soluble.

Finally, it is often desirable to use a buffer that has no biological activity at all, so it can never interfere with the system being studied. TRIS is a very desirable buffer, because it rarely interferes with a reaction. Special buffers, such as HEPES and PIPES (Table 2.8), have been developed for growing cells in tissue culture.

Carbon dioxide can dissolve in water and in water-based fluids, such as blood. The dissolved carbon dioxide forms carbonic acid, which, in turn, reacts to produce bicarbonate ion:



At the pH of blood, which is about one unit higher than the  $pK_a$  of carbonic acid, most of the dissolved  $\text{CO}_2$  is present as  $\text{HCO}_3^-$ . The  $\text{CO}_2$  being transported to the lungs to be expired takes the form of bicarbonate ion. A direct relationship exists between the pH of the blood and the pressure of carbon dioxide gas in the lungs. The properties of hemoglobin, the oxygen-carrying protein in the blood, also enter into the situation (see the Biochemical Connections box in Chapter 4).

The phosphate buffer system is common in the laboratory (*in vitro*, outside the living body) as well as in living organisms (*in vivo*). The buffer system based on TRIS [*tris*(hydroxymethyl)aminomethane] is also widely used *in vitro*. Other buffers that have come into wide use more recently are **zwitterions**, which are compounds that have both a positive charge and a negative charge. Zwitterions are usually considered less likely to interfere with biochemical reactions than some of the earlier buffers (Table 2.8).

Most living systems operate at pH levels close to 7. The  $pK_a$  values of many functional groups, such as the carboxyl and amino groups, are well above or well below this value. As a result, under physiological conditions, many important biomolecules exist as charged species to one extent or another. The practical consequences of this fact are explored in the following Biochemical Connections box.



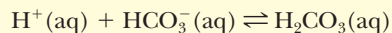
TABLE 2.8

Acid and Base Form of Some Useful Biochemical Buffers			
Acid Form		Base Form	pK <sub>a</sub>
$\text{TRIS} - \text{H}^+$ (protonated form) $(\text{HOCH}_2)_3\text{CNH}_3^+$	$\text{N} - \text{tris}[\text{hydroxymethyl}]\text{aminomethane (TRIS)}$ $\rightleftharpoons$	$\text{TRIS}$ (free amine) $(\text{HOCH}_2)_3\text{CNH}_2$	8.3
$^- \text{TES} - \text{H}^+$ (zwitterionic form) $(\text{HOCH}_2)_3\text{CNH}_2^+\text{CH}_2\text{CH}_2\text{SO}_3^-$	$\text{N} - \text{tris}[\text{hydroxymethyl}]\text{methyl-2-aminoethane sulfonate (TES)}$ $\rightleftharpoons$	$^- \text{TES}$ (anionic form) $(\text{HOCH}_2)_3\text{CNHCH}_2\text{CH}_2\text{SO}_3^-$	7.55
$^- \text{HEPES} - \text{H}^+$ (zwitterionic form)	$\text{N} - 2\text{-hydroxyethylpiperazine-N}'\text{-2-ethane sulfonate (HEPES)}$ $\rightleftharpoons$	$^- \text{HEPES}$ (anionic form)	7.55
$\text{HOCH}_2\text{CH}_2\text{N}^+\text{CH}_2\text{CH}_2\text{SO}_3^-$ $\text{H}$ $^- \text{MOPS} - \text{H}^+$ (zwitterionic form)	$3\text{-[N-morpholino]propane-sulfonic acid (MOPS)}$ $\rightleftharpoons$	$^- \text{MOPS}$ (anionic form)	7.2
$\text{O}^+\text{NCH}_2\text{CH}_2\text{CH}_2\text{SO}_3^-$ $\text{H}$ $^{2-} \text{PIPES} - \text{H}^+$ (protonated dianion)	$\text{Piperazine-N,N}'\text{-bis[2-ethanesulfonic acid] (PIPES)}$ $\rightleftharpoons$	$^{2-} \text{PIPES}$ (dianion)	6.8
$^- \text{O}_3\text{SCH}_2\text{CH}_2\text{N}^+\text{CH}_2\text{CH}_2\text{SO}_3^-$ $\text{H}$		$^- \text{O}_3\text{SCH}_2\text{CH}_2\text{NCH}_2\text{CH}_2\text{SO}_3^-$	

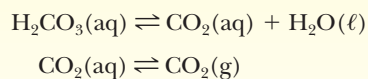
## Biochemical Connections CHEMISTRY OF THE BLOOD

### Some Physiological Consequences of Blood Buffering

The process of respiration plays an important role in the buffering of blood. In particular, an increase in  $\text{H}^+$  concentration can be dealt with by raising the rate of respiration. Initially, the added hydrogen ion binds to bicarbonate ion, forming carbonic acid.



An increased level of carbonic acid raises the levels of dissolved carbon dioxide and, ultimately, gaseous carbon dioxide in the lungs.



A high respiration rate removes this excess carbon dioxide from the lungs, starting a shift in the equilibrium positions of all the foregoing reactions. The removal of gaseous  $\text{CO}_2$  decreases the amount of dissolved  $\text{CO}_2$ . Hydrogen ion reacts with  $\text{HCO}_3^-$ , and, in the process, lowers the  $\text{H}^+$  concentration of blood back to its original level. In this way, the blood pH is kept constant.

In contrast, *hyperventilation* (excessively deep and rapid breathing) removes such large amounts of carbon dioxide from the lungs that it raises the pH of blood, sometimes to dan-

gerously high levels that bring on weakness and fainting. Athletes, however, have learned how to use the increase in blood pH caused by hyperventilation. Short bursts of strenuous exercise produce high levels of lactic acid in the blood as a result of the anaerobic breakdown of glycogen. The presence of so much lactic acid tends to lower the pH of the blood, but a brief (30-second) period of hyperventilation before a short-distance event (say, a 400-m dash, 100-m swim, 1-km bicycle race, or any event that lasts between 30 seconds and about a minute) counteracts the effects of the added lactic acid and maintains the pH balance.

An increase in  $\text{H}^+$  in blood can be caused by large amounts of any acid entering the bloodstream. Aspirin, like lactic acid, is an acid, and extreme acidity resulting from the ingestion of large doses of aspirin can cause *aspirin poisoning*. Exposure to *high altitudes* has an effect similar to hyperventilation at sea level. In response to the tenuous atmosphere, the rate of respiration increases. As with hyperventilation, more carbon dioxide is expired from the lungs, ultimately lowering the  $\text{H}^+$  level in blood and raising the pH. When people who normally live at sea level are suddenly placed at a high elevation, their blood pH rises temporarily, until they become acclimated.

# Biochemical Connections ACIDS AND SPORTS

## Lactic Acid—Not Always the Bad Guy

If you ask anybody who knows something of sports about lactic acid, you will likely hear that it is the acid that causes muscle pain and muscle fatigue. This has been the dogma since 1929, when the first papers were written about accumulation of lactic acid in muscle tissue under anaerobic conditions. However, lactic acid is not biochemically all bad and, in fact, recent evidence suggests that there are some benefits we did not previously know about.

In the first place, we should distinguish between the hydrogen ion dissociated from the lactic acid and the conjugate base, lactate. The  $H^+$  ion released is the ion that is the reactive species and most likely accounts for the pain we associate with lactic acid buildup in the muscles. The lowering of the cellular pH would have effects on a variety of enzyme and muscular systems. However, the conjugate base, lactate, is removed by the blood and makes its way to the liver. Once in the liver, it is converted to glucose via a process called gluconeogenesis, which we will see in Chapter 18. Hospital patients are often given intravenous lactate solutions to indirectly help keep their blood glucose levels up.

Until recently, any athlete would have proclaimed the evils of lactic acid on muscle performance. However, recent evidence suggests that lactic acid actually has a positive effect on fatiguing muscles. A recent study suggested that in a situation where muscles were becoming fatigued, the lactic acid actually maintained



Mango Productions/Corbis

the muscle membrane's ability to depolarize and repolarize longer, allowing the muscles to continue to contract even though they were fatigued. When isolated muscle cells were stripped of lactic acid, they fatigued even faster.

The same evidence was seen with cases of humans that suffered from a disease where they lacked the enzyme that breaks down muscle glycogen. Without the muscle glycogen breakdown, anaerobic metabolism is impossible and no lactic acid accumulates. The results—these patients' muscle fibers fatigued even faster. Some of the "knowledge" about the undesirable effect of lactic acid on muscle fatigue may be explained by the fact that muscle pain seems linked to decreased muscle performance. Lactic acid, as the putative cause of the muscle pain, was therefore assumed to also be the cause of the fatigue. This is an active area of research and we still have a lot to learn about it. Despite decades of study and popular myth, we still do not really know exactly what causes muscle fatigue.

## SUMMARY

**What is polarity?** When two atoms with the same electronegativity form a bond, the electrons are shared equally between the two atoms. However, if atoms with differing electronegativity form a bond, the electrons are not shared equally and more of the negative charge is found closer to one of the atoms.

**Why do some chemicals dissolve in water while others do not?**

The polar nature of water largely determines its solvent properties. Ionic compounds with full charges and polar compounds with partial charges tend to dissolve in water. The underlying physical principle is electrostatic attraction between unlike charges. The negative end of a water dipole attracts a positive ion or the positive end of another dipole. The positive end of a water molecule attracts a negative ion or the negative end of another dipole.

**Why do oil and water mixed together separate into layers?**

Oil molecules are amphipathic—having both polar (hydrophilic) heads and nonpolar (hydrophobic) tail portions. When oil and water separate into layers, the polar head groups of the oil molecules are in contact with the aqueous environment and the nonpolar tails are sequestered from the water. Van der Waals interactions between nonpolar molecules provide the energetic basis for this spontaneous molecular arrangement.

**Why does water have such interesting and unique properties?**

Water has unique properties for a molecule its size, such as a very high boiling point and melting point. This is due to the extensive hydrogen bonding possible between water molecules. Each water molecule has two sources of partial positive charge and two of partial negative charge. This allows water to form an array in a solid form and to bond with many other water molecules in liquid form. The extensive hydrogen bonding requires large amounts of energy to disrupt, and therefore it melts and boils at higher temperatures than other molecules of its relative size.

**What are acids and bases?** Acids are compounds that release hydrogen ions (protons) when dissolved in aqueous solution. In other words, they are proton donors. Bases are compounds that are proton acceptors.

**What is pH?** The mathematical definition of pH is the negative of the logarithm of the hydrogen ion concentration. It is a measure of the acidity of the solution. The lower the pH, the more acidic the solution. Because of the log term, a pH change of one unit means a tenfold change in hydrogen ion concentration.

**Why do we want to know the pH?** It is important to know the pH because many biological reactions require a very

tight range of pH values. For example, an enzyme that is active at pH 7.0 may be completely inactive at pH 8.0. Solutions used in science often must have their pH controlled in order to have an experiment function correctly. While local variations in pH may occur in certain subcellular organelles, a cell must maintain a pH near neutrality in order to stay alive.

**How do buffers work?** Buffers work based on the nature of weak acids and their conjugate bases that compose the buffer. If a source of extra hydrogen ion is added to a buffer solution, it reacts with the conjugate base to form the weak acid. If a source of hydroxide ion is added to the buffer, it reacts with the weak acid to form water and the conjugate base. In this way, either added  $H^+$  or  $OH^-$  is “used up” by adding it to a buffer. This keeps the pH much more stable than if the same acid or base had been added to an unbuffered system.

**How do we choose a buffer?** We choose a buffer primarily by knowing the pH that we wish to maintain. For example, if

we are performing an experiment and we want the solution to stay at pH 7.5, we look for a buffer that has a  $pK_a$  of 7.5 because buffers are most effective when the pH is close to the buffer  $pK_a$ .

**How do we make buffers in the laboratory?** The most efficient way to make a buffer in the laboratory is to add either the weak acid form or the weak base form of the buffer compound to a container, add water, and then measure the pH with a pH meter. The pH will be either too low or too high. We then add strong acid or strong base until the pH is the desired buffer pH. Then we bring the solution up to the final volume so that concentration is correct.

**Are naturally occurring pH buffers present in living organisms?** Buffers are not just an artificial system used in the laboratory. Living systems are buffered by naturally occurring compounds. Naturally occurring phosphate and carbonate buffers help maintain physiological pH near 7.0.

## REVIEW EXERCISES

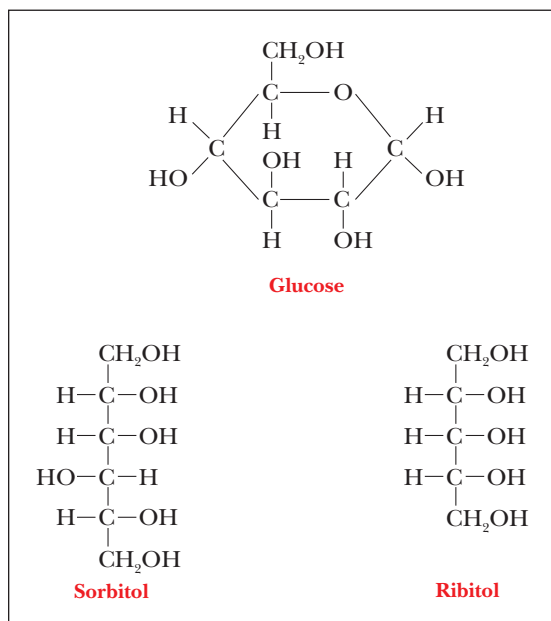
**OWL** Interactive versions of these problems are assignable in OWL

### 2.1 Water and Polarity

1. **Reflect and Apply** Why is water necessary for life?
2. **Reflect and Apply** Contemplate biochemistry if atoms did not differ in electronegativity.

### 2.2 Hydrogen Bonds

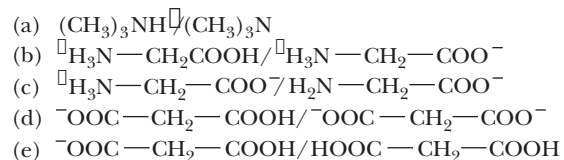
3. **Recall** What are some macromolecules that have hydrogen bonds as a part of their structures?
4. **Biochemical Connections** How are hydrogen bonds involved in the transfer of genetic information?
5. **Reflect and Apply** Rationalize the fact that hydrogen bonding has not been observed between  $CH_4$  molecules.
6. **Reflect and Apply** Draw three examples of types of molecules that can form hydrogen bonds.
7. **Recall** What are the requirements for molecules to form hydrogen bonds? (What atoms must be present and involved in such bonds?)
8. **Reflect and Apply** Many properties of acetic acid can be rationalized in terms of a hydrogen-bonded dimer. Propose a structure for such a dimer.
9. **Reflect and Apply** How many water molecules could hydrogen-bond *directly* to the molecules of glucose, sorbitol, and ribitol, shown here?



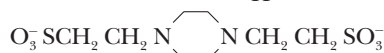
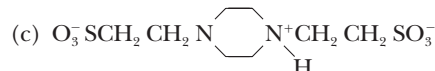
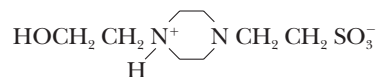
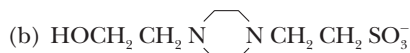
10. **Reflect and Apply** Both RNA and DNA have negatively charged phosphate groups as part of their structure. Would you expect ions that bind to nucleic acids to be positively or negatively charged? Why?

## 2.3 Acids, Bases, and pH

11. **Recall** Identify the conjugate acids and bases in the following pairs of substances:



12. **Recall** Identify conjugate acids and bases in the following pairs of substances:



13. **Reflect and Apply** Aspirin is an acid with a  $\text{p}K_a$  of 3.5; its structure includes a carboxyl group. To be absorbed into the bloodstream, it must pass through the membrane lining the stomach and the small intestine. Electrically neutral molecules can pass through a membrane more easily than can charged molecules. Would you expect more aspirin to be absorbed in the stomach, where the pH of gastric juice is about 1, or in the small intestine, where the pH is about 6? Explain your answer.
14. **Recall** Why does the pH change by one unit if the hydrogen ion concentration changes by a factor of 10?
15. **Mathematical** Calculate the hydrogen ion concentration,  $[\text{H}^+]$ , for each of the following materials:
- (a) Blood plasma, pH 7.4  
 (b) Orange juice, pH 3.5  
 (c) Human urine, pH 6.2  
 (d) Household ammonia, pH 11.5  
 (e) Gastric juice, pH 1.8
16. **Mathematical** Calculate the hydrogen ion concentration,  $[\text{H}^+]$ , for each of the following materials:
- (a) Saliva, pH 6.5  
 (b) Intracellular fluid of liver, pH 6.9  
 (c) Tomato juice, pH 4.3  
 (d) Grapefruit juice, pH 3.2
17. **Mathematical** Calculate the hydroxide ion concentration,  $[\text{OH}^-]$ , for each of the materials used in Question 16.

## 2.4 Titration Curves

18. **Recall** Define the following:

- (a) Acid dissociation constant  
 (b) Acid strength  
 (c) Amphipathic  
 (d) Buffering capacity  
 (e) Equivalence point  
 (f) Hydrophilic  
 (g) Hydrophobic  
 (h) Nonpolar  
 (i) Polar  
 (j) Titration

19. **Reflect and Apply** Look at Figure 2.15 and Table 2.8. Which compound in the table would give a titration curve the most similar to the one shown in the figure? Why?
20. **Reflect and Apply** Look at Figure 2.15. If you did this titration using TRIS instead of phosphate, how would the titration curve look compared to the figure? Explain.

## 2.5 Buffers

21. **Biochemical Connections** List the criteria used to select a buffer for a biochemical reaction.
22. **Biochemical Connections** What is the relationship between  $\text{p}K_a$  and the useful range of a buffer?
23. **Mathematical** What is the  $[\text{CH}_3\text{COO}^-]/[\text{CH}_3\text{COOH}]$  ratio in an acetate buffer at pH 5.00?
24. **Mathematical** What is the  $[\text{CH}_3\text{COO}^-]/[\text{CH}_3\text{COOH}]$  ratio in an acetate buffer at pH 4.00?
25. **Mathematical** What is the ratio of TRIS/TRIS- $\text{H}^+$  in a TRIS buffer at pH 8.7?
26. **Mathematical** What is the ratio of HEPES/HEPES- $\text{H}^+$  in a HEPES buffer at pH 7.9?
27. **Mathematical** How would you prepare 1 L of a 0.050 M phosphate buffer at pH 7.5 using crystalline  $\text{K}_2\text{HPO}_4$  and a solution of 1.0 M HCl?
28. **Mathematical** The buffer needed for Question 27 can also be prepared using crystalline  $\text{NaH}_2\text{PO}_4$  and a solution of 1.0 M NaOH. How would you do this?
29. **Mathematical** Calculate the pH of a buffer solution prepared by mixing 75 mL of 1.0 M lactic acid (see Table 2.6) and 25 mL of 1.0 M sodium lactate.
30. **Mathematical** Calculate the pH of a buffer solution prepared by mixing 25 mL of 1.0 M lactic acid and 75 mL of 1.0 M sodium lactate.
31. **Mathematical** Calculate the pH of a buffer solution that contains 0.10 M acetic acid (Table 2.6) and 0.25 M sodium acetate.
32. **Mathematical** A catalog in the lab has a recipe for preparing 1 L of a TRIS buffer at 0.0500 M and with pH 8.0: dissolve 2.02 g of TRIS (free base, MW = 121.1 g/mol) and 5.25 g of TRIS hydrochloride (the acidic form, MW = 157.6 g/mol) in a total volume of 1 L. Verify that this recipe is correct.
33. **Mathematical** If you mix equal volumes of 0.1 M HCl and 0.20 M TRIS (free amine form; see Table 2.8), is the resulting solution a buffer? Why or why not?
34. **Mathematical** What would be the pH of the solution described in Question 33?
35. **Mathematical** If you have 100 mL of a 0.10 M TRIS buffer at pH 8.3 (Table 2.8) and you add 3.0 mL of 1 M HCl, what will be the new pH?
36. **Mathematical** What would be the pH of the solution in Question 35 if you were to add 3.0 mL more of 1 M HCl?
37. **Mathematical** Show that, for a pure weak acid in water,  $\text{pH} = (\text{p}K_a - \log [\text{HA}])/2$ .
38. **Mathematical** What is the ratio of concentrations of acetate ion and undissociated acetic acid in a solution that has a pH of 5.12?
39. **Biochemical Connections** You need to carry out an enzymatic reaction at pH 7.5. A friend suggests a weak acid with a  $\text{p}K_a$  of 3.9 as the basis of a buffer. Will this substance and its conjugate base make a suitable buffer? Why or why not?
40. **Mathematical** If the buffer suggested in Question 39 were made, what would be the ratio of the conjugate base/conjugate acid?
41. **Biochemical Connections** Suggest a suitable buffer range for each of the following substances:
- (a) Lactic acid ( $\text{p}K_a = 3.86$ ) and its sodium salt  
 (b) Acetic acid ( $\text{p}K_a = 4.76$ ) and its sodium salt  
 (c) TRIS ( $\text{p}K_a = 8.3$ ; see Table 2.8) in its protonated form and its free amine form  
 (d) HEPES ( $\text{p}K_a = 7.55$ ; see Table 2.8) in its zwitterionic form and its anionic form

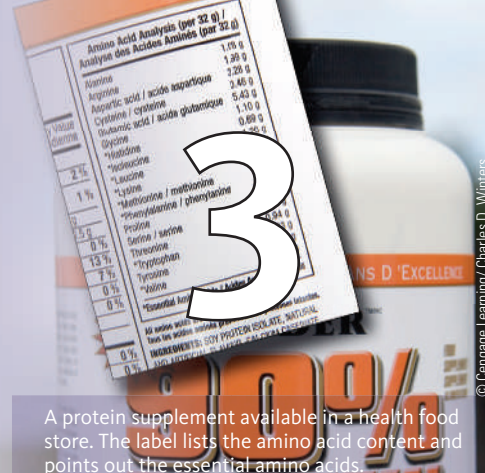
42. **Biochemical Connections** Which of the buffers shown in Table 2.8 would you choose to make a buffer with a pH of 7.3? Explain why.
43. **Mathematical** The solution in Question 27 is called 0.050 *M*, even though the concentration of neither the free base nor the conjugate acid is 0.050 *M*. Why is 0.050 *M* the correct concentration to report?
44. **Reflect and Apply** In Section 2.4 we said that at the equivalence point of a titration of acetic acid, *essentially all* the acid has been converted to acetate ion. Why do we not say that *all* the acetic acid has been converted to acetate ion?
45. **Mathematical** Define buffering capacity. How do the following buffers differ in buffering capacity? How do they differ in pH?  
 Buffer a: 0.01 *M* Na<sub>2</sub>HPO<sub>4</sub> and 0.01 *M* NaH<sub>2</sub>PO<sub>4</sub>  
 Buffer b: 0.10 *M* Na<sub>2</sub>HPO<sub>4</sub> and 0.10 *M* NaH<sub>2</sub>PO<sub>4</sub>  
 Buffer c: 1.0 *M* Na<sub>2</sub>HPO<sub>4</sub> and 1.0 *M* NaH<sub>2</sub>PO<sub>4</sub>
46. **Biochemical Connections** If you wanted to make a HEPES buffer at pH 8.3, and you had both HEPES acid and HEPES base available, which would you start with, and why?
47. **Biochemical Connections** We usually say that a perfect buffer has its pH equal to its p*K*<sub>a</sub>. Give an example of a situation in which it would be advantageous to have a buffer with a pH 0.5 unit higher than its p*K*<sub>a</sub>.
48. **Recall** What quality of zwitterions makes them desirable buffers?
49. **Reflect and Apply** Many of the buffers used these days, such as HEPES and PIPES, were developed because they have desirable characteristics, such as resisting pH change with dilution. Why would resisting pH change with dilution be advantageous?
50. **Reflect and Apply** Another characteristic of modern buffers such as HEPES is that their pH changes little with changes in temperature. Why is this desirable?
51. **Reflect and Apply** Identify the zwitterions in the list of substances in Question 11.
52. **Biochemical Connections** A frequently recommended treatment for hiccups is to hold one's breath. The resulting condition, hypoventilation, causes buildup of carbon dioxide in the lungs. Predict the effect on the pH of blood.

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# Amino Acids and Peptides



© Cengage Learning/Charles D. Winters

## 3.1 Amino Acids Exist in a Three-Dimensional World

### Why is it important to specify the three-dimensional structure of amino acids?

Among all the possible amino acids, only 20 are usually found in proteins. The general structure of amino acids includes an **amino group** and a **carboxyl group**, both of which are bonded to the  $\alpha$ -carbon (the one next to the carboxyl group). The  $\alpha$ -carbon is also bonded to a hydrogen and to the **side chain group**, which is represented by the letter R. The R group determines the identity of the particular amino acid (Figure 3.1). The two-dimensional formula shown here can only partially convey the common structure of amino acids because one of the most important properties of these compounds is their three-dimensional shape, or **stereochemistry**.

Every object has a mirror image. Many pairs of objects that are mirror images can be superimposed on each other; two identical solid-colored coffee mugs are an example. In other cases, the mirror-image objects cannot be superimposed on one another but are related to each other as the right hand is to the left. Such nonsuperimposable mirror images are said to be **chiral** (from the Greek *cheir*, “hand”); many important biomolecules are chiral. A frequently encountered chiral center in biomolecules is a carbon atom with four different groups bonded to it (Figure 3.1). Such a center occurs in all amino acids except glycine. Glycine has two hydrogen atoms bonded to the  $\alpha$ -carbon; in other words, the side chain (R group) of glycine is hydrogen. Glycine is not chiral (or, alternatively, is **achiral**) because of this symmetry. In all the other commonly occurring amino acids, the  $\alpha$ -carbon has four different groups bonded to it, giving rise to two nonsuperimposable mirror-image forms. Figure 3.2 shows perspective drawings of these two possibilities, or **stereoisomers**, for alanine, where the R group is  $-\text{CH}_3$ . The dashed wedges represent bonds directed away from the observer, and the solid triangles represent bonds directed out of the plane of the paper in the direction of the observer.

The two possible stereoisomers of another chiral compound, L- and D-glyceraldehyde, are shown for comparison with the corresponding forms of alanine. These two forms of glyceraldehyde are the basis of the classification of amino acids into L and D forms. The terminology comes from the Latin *laevus* and *dexter*, meaning “left” and “right,” respectively, which comes from the ability of optically active compounds to rotate polarized light to the left or the right. The two stereoisomers of each amino acid are designated as **L- and D-amino acids** on the basis of their similarity to the glyceraldehyde standard. When drawn in a certain orientation, the L form of glyceraldehyde has the hydroxyl group on the left side of the molecule, and the D form has it on the right side, as shown in perspective in Figure 3.2 (a Fischer projection). To determine the L or D designation for an amino acid, it is drawn as shown. The position of the amino group on the left or right side of the  $\alpha$ -carbon determines

## Chapter Outline

### 3.1 Amino Acids Exist in a Three-Dimensional World

- Why is it important to specify the three-dimensional structure of amino acids?

### 3.2 Individual Amino Acids: Their Structures and Properties

- Why are amino acid side chains so important?
- Which amino acids have nonpolar side chains? (Group 1)
- Which amino acids have electrically neutral polar side chains? (Group 2)
- Which amino acids have carboxyl groups in their side chains? (Group 3)
- Which amino acids have basic side chains? (Group 4)
- Which amino acids are found less commonly in proteins?

### 3.3 Amino Acids Can Act as Both Acids and Bases

- What happens when we titrate an amino acid?

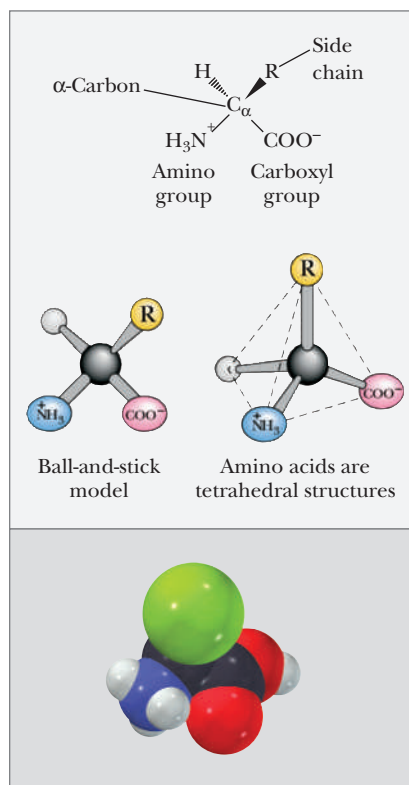
### 3.4 The Peptide Bond

- Which groups on amino acids react to form a peptide bond?

### 3.5 Small Peptides with Physiological Activity

- What are some biological functions of small peptides?

**OWL** Online homework for this chapter may be assigned in OWL.



■ **FIGURE 3.1** The general formula of amino acids, showing the ionic forms that predominate at pH 7.

the L or D designation. The amino acids that occur in proteins are all of the L form. Although D-amino acids occur in nature, most often in bacterial cell walls and in some antibiotics, they are not found in proteins.

## 3.2 Individual Amino Acids: Their Structures and Properties

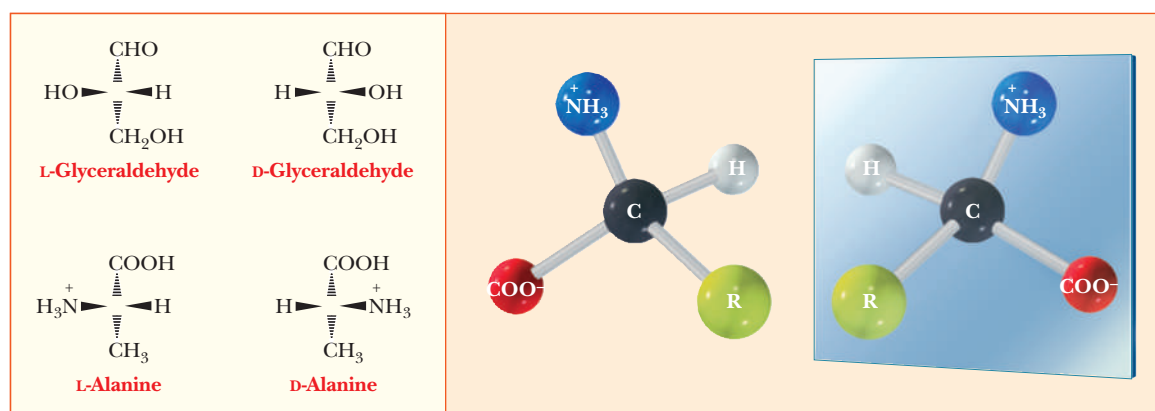
### Why are amino acid side chains so important?

The R groups, and thus the individual amino acids, are classified according to several criteria, two of which are particularly important. The first of these is the polar or nonpolar nature of the side chain. The second depends on the presence of an acidic or basic group in the side chain. Other useful criteria include the presence of functional groups other than acidic or basic ones in the side chains and the nature of those groups.

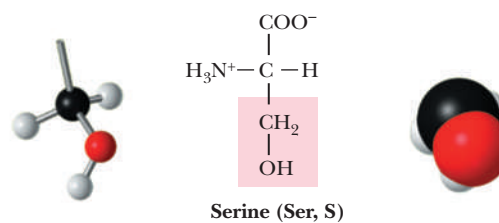
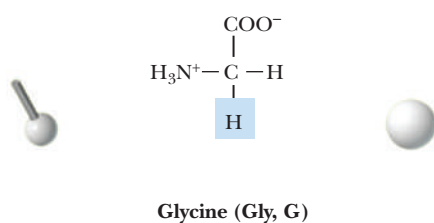
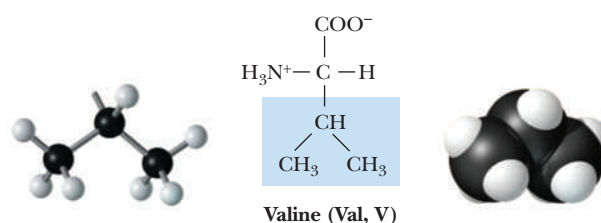
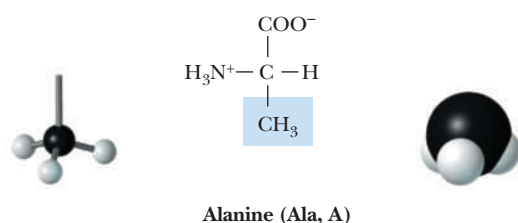
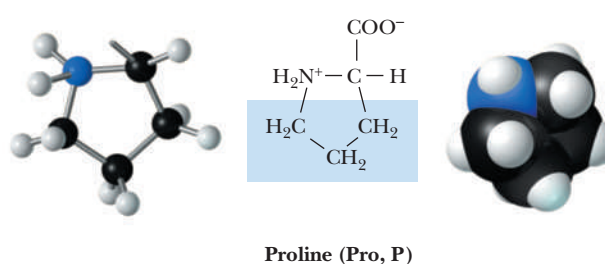
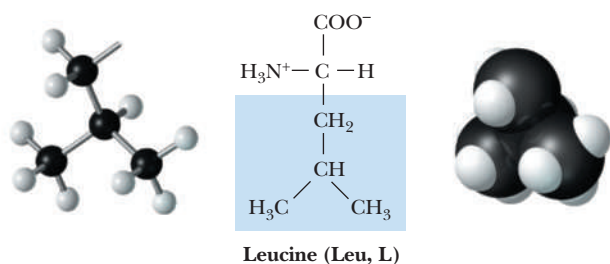
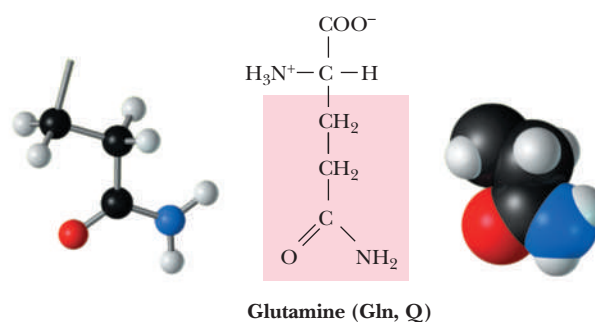
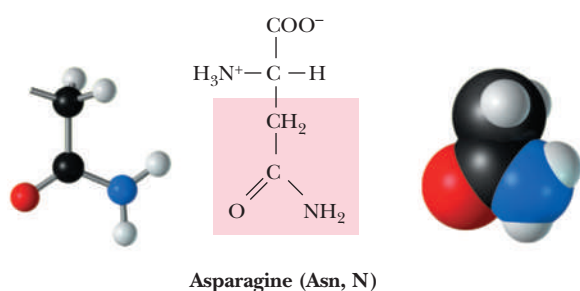
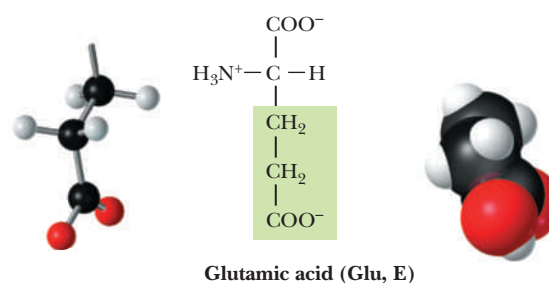
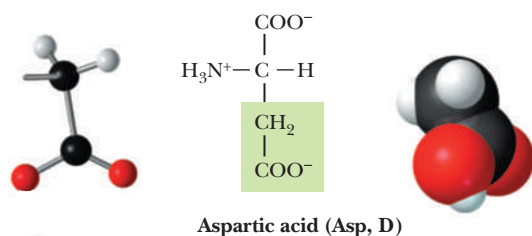
As mentioned, the side chain of the simplest amino acid, glycine, is a hydrogen atom, and in this case alone two hydrogen atoms are bonded to the  $\alpha$ -carbon. In all other amino acids, the side chain is larger and more complex (Figure 3.3). Side-chain carbon atoms are designated with letters of the Greek alphabet, counting from the  $\alpha$ -carbon. These carbon atoms are, in turn, the  $\beta$ -,  $\gamma$ -,  $\delta$ -, and  $\epsilon$ -carbons (see lysine in Figure 3.3); a terminal carbon atom is referred to as the  $\omega$ -carbon, from the name of the last letter of the Greek alphabet. We frequently refer to amino acids by three-letter or one-letter abbreviations of their names, with the one-letter designations becoming much more prevalent these days; Table 3.1 lists these abbreviations.

### Which amino acids have nonpolar side chains? (Group 1)

One group of amino acids has nonpolar side chains. This group consists of glycine, alanine, valine, leucine, isoleucine, proline, phenylalanine, tryptophan, and methionine. In several members of this group—namely alanine, valine, leucine, and isoleucine—each side chain is an aliphatic hydrocarbon group. (In organic chemistry, the term *aliphatic* refers to the absence of a benzene ring or related structure.) Proline has an aliphatic cyclic structure, and the nitrogen is bonded to two carbon atoms. In the terminology of organic chemistry, the amino group of proline is a secondary amine, and proline is often called an *imino acid*. In contrast, the amino groups of all the other common amino acids are primary amines. In phenylalanine, the hydrocarbon group is aromatic (it contains a cyclic group similar to a benzene ring) rather than aliphatic.



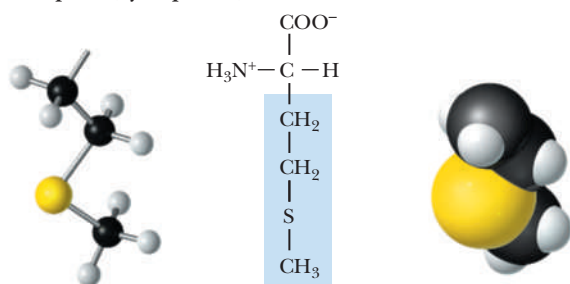
■ **FIGURE 3.2** Stereochemistry of alanine and glycine. The amino acids found in proteins have the same chirality as L-glyceraldehyde, which is opposite to that of D-glyceraldehyde.

**A Non-polar (hydrophobic)****B Polar, uncharged****C Acidic**

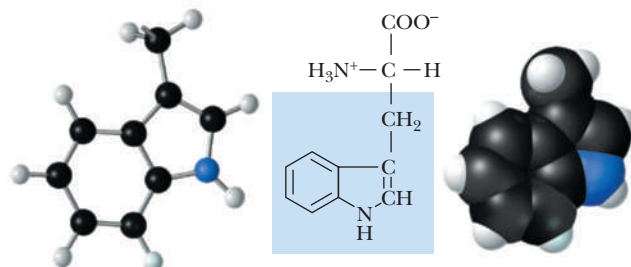
(continued)

**FIGURE 3.3 Structures of the amino acids commonly found in proteins.** The 20 amino acids that are the building blocks of proteins can be classified as (a) nonpolar (hydrophobic), (b) polar, (c) acidic, or (d) basic. Also shown are the one-letter and three-letter codes used to denote amino acids. For each amino acid, the ball-and-stick model (left) and the space-filling model (right) show only the side chain. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

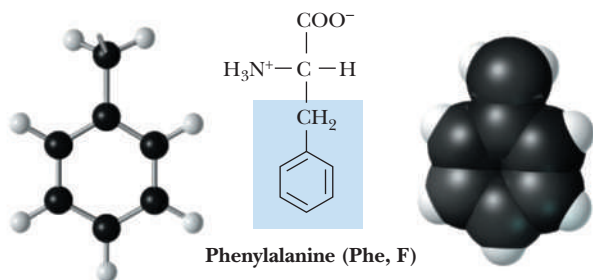
(continued)

**A** Non-polar (hydrophobic)

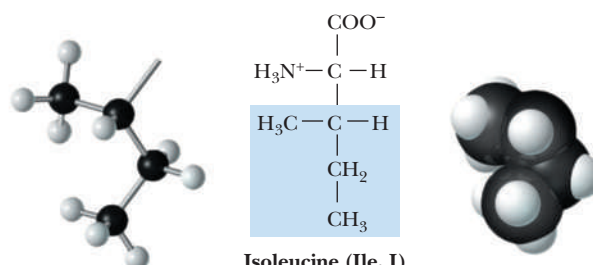
Methionine (Met, M)



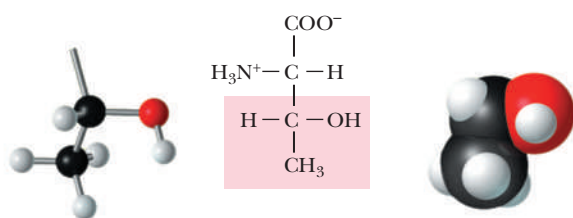
Tryptophan (Trp, W)



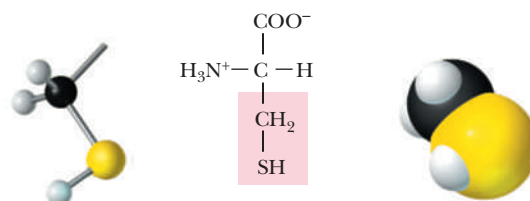
Phenylalanine (Phe, F)



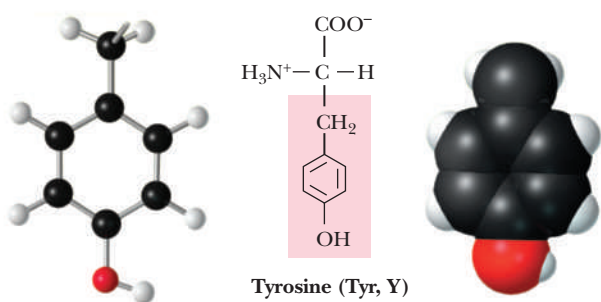
Isoleucine (Ile, I)

**B** Polar, uncharged

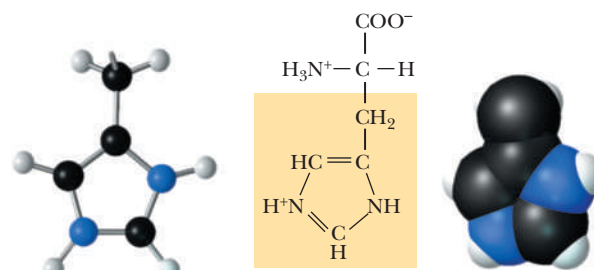
Threonine (Thr, T)



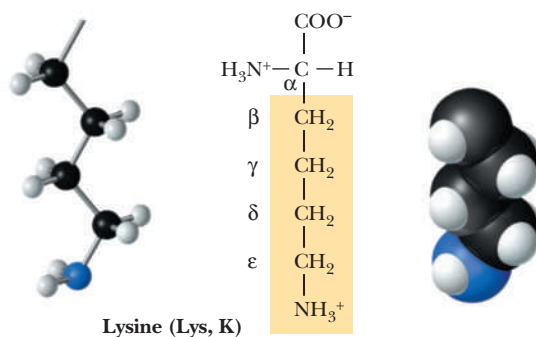
Cysteine (Cys, C)



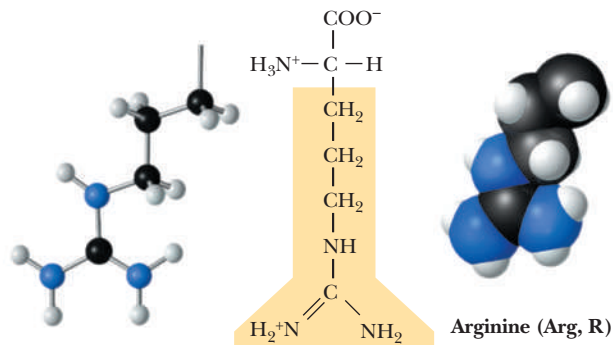
Tyrosine (Tyr, Y)



Histidine (His, H)

**D** Basic

Lysine (Lys, K)



Arginine (Arg, R)

TABLE 3.1

Names and Abbreviations of the Common Amino Acids		
Amino Acid	Three-Letter Abbreviation	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamic acid	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Note: One-letter abbreviations start with the same letter as the name of the amino acid where this is possible. When the names of several amino acids start with the same letter, phonetic names (occasionally facetious ones) are used, such as Rginine, asparDic, Fenylalanine, tWryptophan. Where two or more amino acids start with the same letter, it is the smallest one whose one-letter abbreviation matches its first letter.

In tryptophan, the side chain contains an indole ring, which is also aromatic. In methionine, the side chain contains a sulfur atom in addition to aliphatic hydrocarbon groupings. (See Figure 3.3.)

### Which amino acids have electrically neutral polar side chains? (Group 2)

Another group of amino acids has polar side chains that are electrically neutral (uncharged) at neutral pH. This group includes serine, threonine, tyrosine, cysteine, glutamine, and asparagine. Glycine is sometimes included here for convenience because it lacks a nonpolar side chain.

In serine and threonine, the polar group is a hydroxyl ( $\text{—OH}$ ) bonded to aliphatic hydrocarbon groups. The hydroxyl group in tyrosine is bonded to an aromatic hydrocarbon group, which eventually loses a proton at higher pH. (The hydroxyl group in tyrosine is a phenol, which is a stronger acid than an aliphatic alcohol. As a result, the side chain of tyrosine can lose a proton in a titration, whereas those of serine and threonine would require such a high pH that  $\text{pK}_a$  values are not normally listed for these side chains.) In cysteine, the polar side chain consists of a thiol group ( $\text{—SH}$ ), which can react with other cysteine thiol groups to form disulfide ( $\text{—S—S—}$ ) bridges in proteins in an oxidation reaction (Section 1.9). The thiol group can also lose a proton. The amino acids glutamine and asparagine have amide groups, which are derived from carboxyl groups, in their side chains. Amide bonds do not ionize in the range of pH usually encountered in biochemistry. Glutamine and asparagine can be considered derivatives of the Group 3 amino acids, glutamic acid and aspartic acid, respectively; those two amino acids have carboxyl groups in their side chains.



### Which amino acids have carboxyl groups in their side chains? (Group 3)

Two amino acids, glutamic acid and aspartic acid, have carboxyl groups in their side chains in addition to the one present in all amino acids. A carboxyl group can lose a proton, forming the corresponding carboxylate anion (Section 2.5)—glutamate and aspartate, respectively, in the case of these two amino acids. Because of the presence of the carboxylate, the side chain of each of these two amino acids is negatively charged at neutral pH.

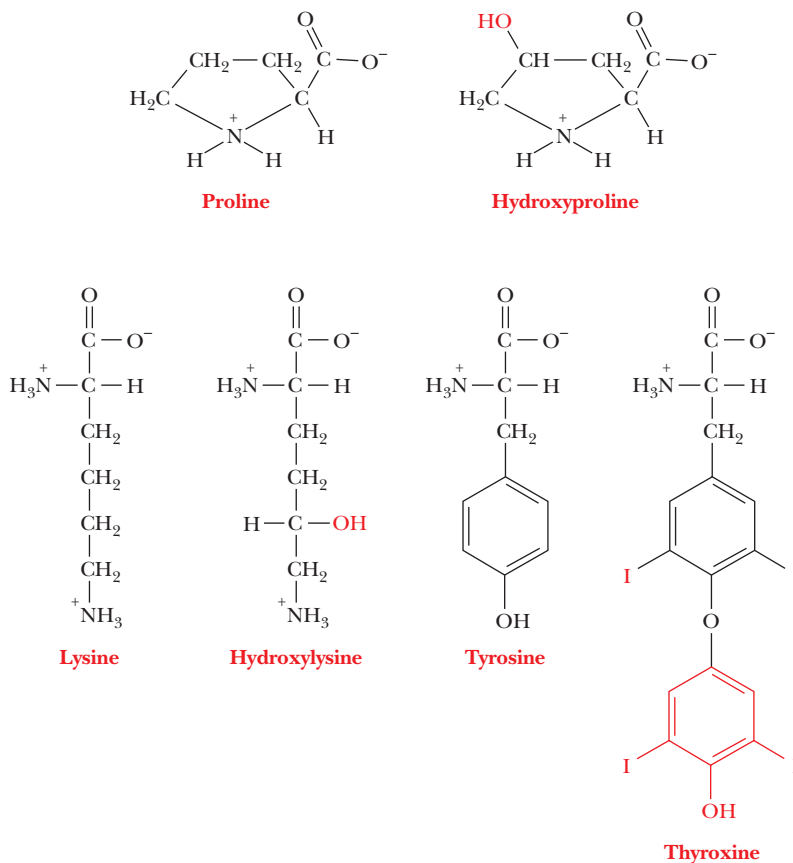
### Which amino acids have basic side chains? (Group 4)

Three amino acids—histidine, lysine, and arginine—have basic side chains, and the side chain in all three is positively charged at or near neutral pH. In lysine, the side-chain amino group is attached to an aliphatic hydrocarbon tail. In arginine, the side-chain basic group, the guanidino group, is more complex in structure than the amino group, but it is also bonded to an aliphatic hydrocarbon tail. In free histidine, the  $pK_a$  of the side-chain imidazole group is 6.0, which is not far from physiological pH. The  $pK_a$  values for amino acids depend on the environment and can change significantly within the confines of a protein. Histidine can be found in the protonated or unprotonated forms in proteins, and the properties of many proteins depend on whether individual histidine residues are or are not charged.

### Uncommon Amino Acids

#### Which amino acids are found less commonly in proteins?

Many other amino acids, in addition to the ones listed here, are known to exist. They occur in some, but by no means all, proteins. Figure 3.4 shows



■ **FIGURE 3.4** Structures of hydroxyproline, hydroxylysine, and thyroxine. The structures of the parent amino acids—proline for hydroxyproline, lysine for hydroxylysine, and tyrosine for thyroxine—are shown for comparison. All amino acids are shown in their predominant ionic forms at pH 7.

some examples of the many possibilities. They are derived from the common amino acids and are produced by modification of the parent amino acid after the protein is synthesized by the organism in a process called posttranslational modification. Hydroxyproline and hydroxylysine differ from the parent amino acids in that they have hydroxyl groups on their side chains; they are found only in a few connective-tissue proteins, such as collagen. Thyroxine differs from tyrosine in that it has an extra iodine-containing aromatic group on the side chain; it is produced only in the thyroid gland, formed by posttranslational modification of tyrosine residues in the protein thyroglobulin. Thyroxine is then released as a hormone by proteolysis of thyroglobulin.

### Apply Your Knowledge

#### Amino Acids, Their Structures and Properties

1. In the following group, identify the amino acids with nonpolar side chains and those with basic side chains: alanine, serine, arginine, lysine, leucine, and phenylalanine.
2. The  $pK_a$  of the side-chain imidazole group of histidine is 6.0. What is the ratio of uncharged to charged side chains at pH 7.0?

#### Solution

Notice that in the first part of this exercise in applying your knowledge, you are asked to do a fact check on material from this chapter, and in the second part you are asked to recall and apply concepts from an earlier chapter.

1. See Figure 3.3. Nonpolar: alanine, leucine, and phenylalanine; basic: arginine and lysine. Serine is not in either category because it has a polar side chain.
2. The ratio is 10:1 because the pH is one unit higher than the  $pK_a$ .

## 3.3 Amino Acids Can Act as Both Acids and Bases

In a free amino acid, the carboxyl group and amino group of the general structure are charged at neutral pH—the carboxylate portion negatively and the amino group positively. Amino acids without charged groups on their side chains exist in neutral solution as zwitterions with no net charge. A zwitterion has equal positive and negative charges; in solution, it is electrically neutral.

Neutral amino acids do not exist in the form  $\text{NH}_2\text{—CHR—COOH}$  (that is, without charged groups).

### What happens when we titrate an amino acid?

When an amino acid is titrated, its titration curve indicates the reaction of each functional group with hydrogen ion. In alanine, the carboxyl and amino groups are the two titratable groups. At very low pH, alanine has a protonated (and thus uncharged) carboxyl group and a positively charged amino group that is also protonated. Under these conditions, the alanine has a net positive charge of 1. As base is added, the carboxyl group loses its proton to become a negatively charged carboxylate group (Figure 3.5a), and the pH of the solution increases. Alanine now has no net charge.

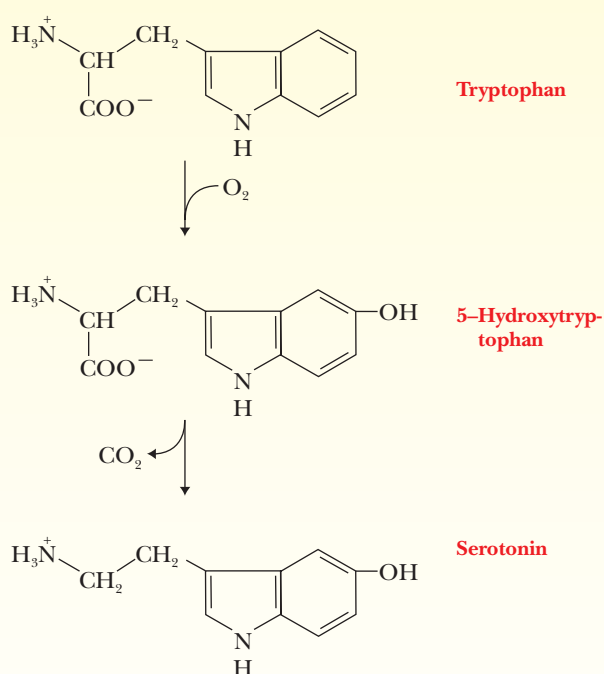
# Biochemical Connections NEUROPHYSIOLOGY

## Amino Acids to Calm Down and Pep Up

Two amino acids deserve some special notice because both are key precursors to many hormones and neurotransmitters (substances involved in the transmission of nerve impulses). The study of neurotransmitters is work in progress, but we do recognize that certain key molecules appear to be involved. Because many neurotransmitters have very short biological half-lives and function at very low concentrations, we also recognize that other derivatives of these molecules may be the actual biologically active forms.

Two of the neurotransmitter classes are simple derivatives of the two amino acids **tyrosine** and **tryptophan**. The active products are monoamine derivatives, which are themselves degraded or deactivated by monoamine oxidases (MAOs).

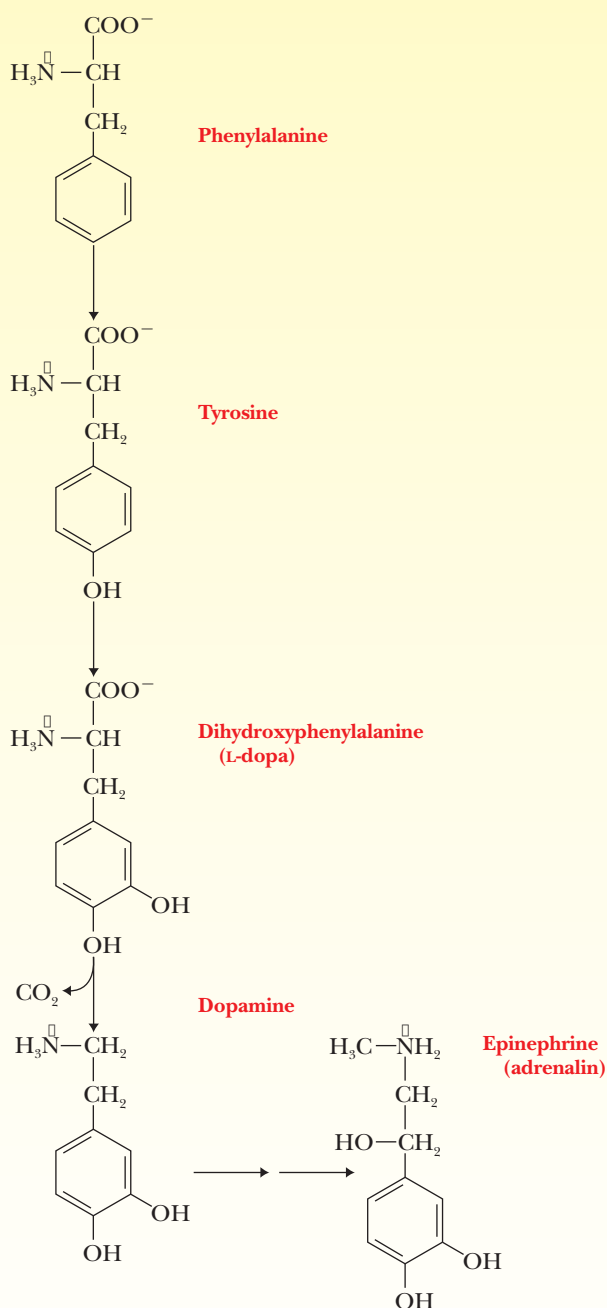
Tryptophan is converted to serotonin, more properly called 5-hydroxytryptamine.



Tyrosine, itself normally derived from phenylalanine, is converted to the class called catecholamines, which includes epinephrine, commonly known by its proprietary name, **adrenalin**.

Note that L-dihydroxyphenylalanine (L-dopa) is an intermediate in the conversion of tyrosine. Lower-than-normal levels of L-dopa are involved in Parkinson's disease. Tyrosine or phenylalanine supplements might increase the levels of dopamine, though L-dopa, the immediate precursor, is usually prescribed because L-dopa passes into the brain quickly through the blood-brain barrier.

Tyrosine and phenylalanine are precursors to norepinephrine and epinephrine, both of which are stimulatory. Epinephrine is commonly known as the "flight or fight" hormone. It causes



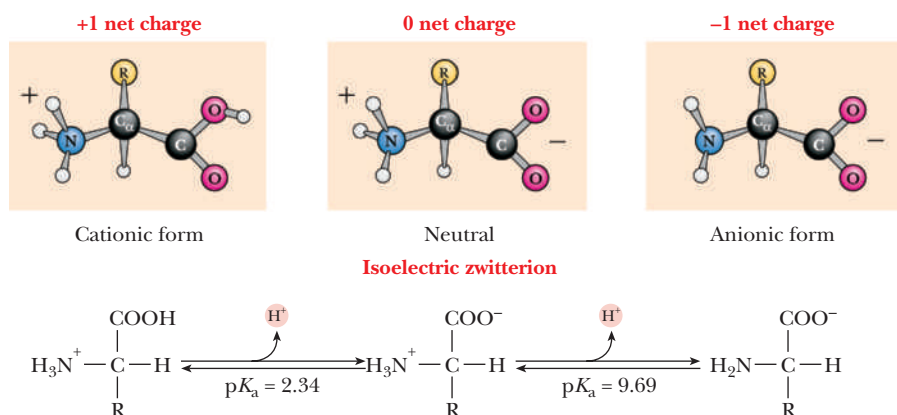
the release of glucose and other nutrients into the blood and also stimulates brain function. People taking MAO inhibitors stay in a relatively high mental state, sometimes too high, because the epinephrine is not metabolized rapidly. Tryptophan is a precursor to serotonin, which has a sedative effect, giving a pleasant feeling. Very low levels of serotonin are associated with depression, while extremely high levels actually produce a manic state. Manic-depressive illness (also called bipolar disorder) can be managed by controlling the levels of serotonin and its further metabolites.

## Biochemical Connections (CONTINUED)

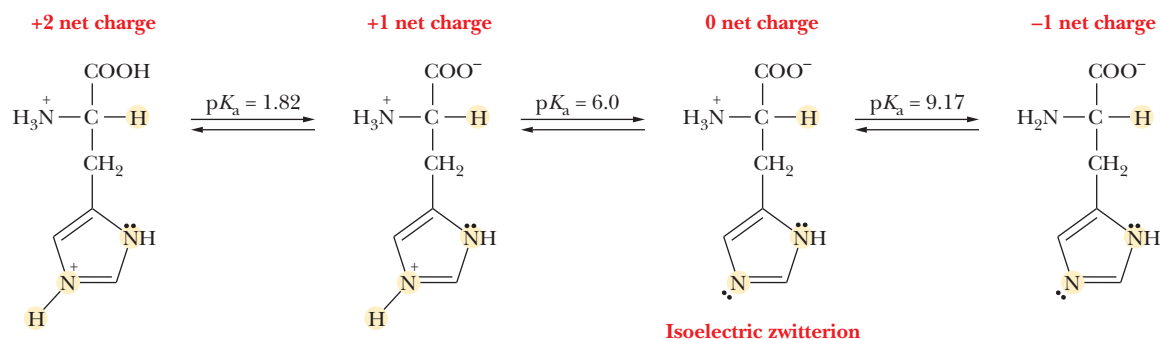
It has been suggested that tyrosine and phenylalanine may have unexpected effects in some people. For example, there is increasing evidence that some people get headaches from the phenylalanine in aspartame (a low-calorie sweetener). It is also likely that many illegal psychedelic drugs, such as mescaline and psilocine, mimic and interfere with the effects of neurotransmitters. The Oscar-winning film *A Beautiful Mind* focused on the disturbing problems associated with schizophrenia. Until recently, the neurotransmitter dopamine was a major focus in the study of schizophrenia. More

recently, it has been suggested that irregularities in the metabolism of glutamate, a neurotransmitter, can lead to the disease.

Some people insist that supplements of tyrosine give them a morning lift and that tryptophan helps them sleep at night. Milk proteins have high levels of tryptophan; a glass of warm milk before bed is widely believed to be an aid in inducing sleep. Cheese and red wines contain high amounts of tyramine, which mimics epinephrine; for many people a cheese omelet in the morning is a favorite way to start the day.

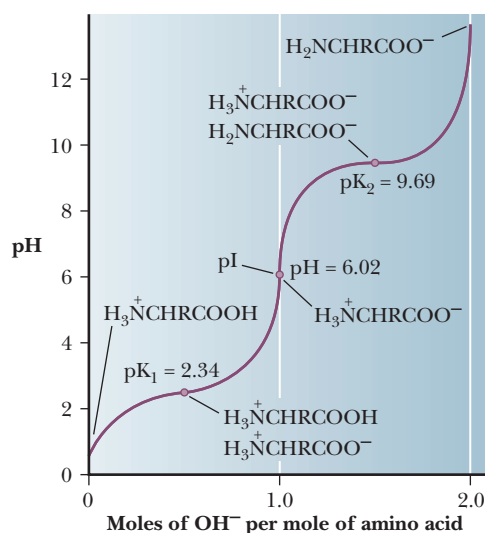


- A** The ionic forms of the amino acids, shown without consideration of any ionizations on the side chain. The cationic form is the low-pH form, and the titration of the cationic species with base yields the zwitterions and finally the anionic form.



- B** The ionization of histidine (an amino acid with a titratable side chain).

**FIGURE 3.5** The ionization of amino acids.



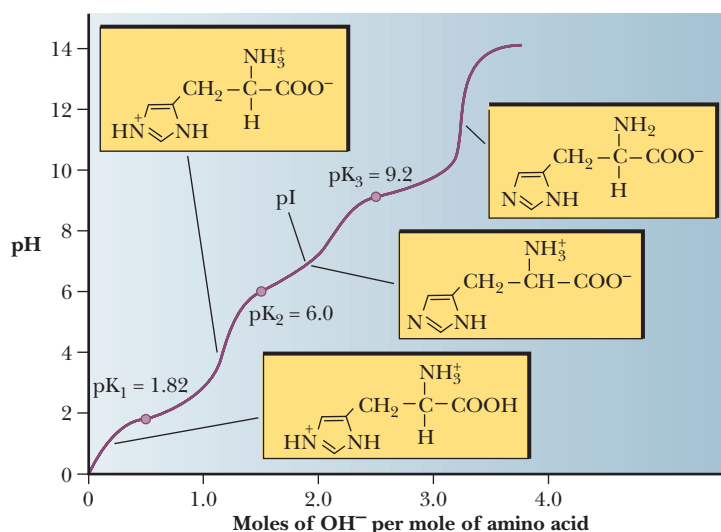
■ **FIGURE 3.6** The titration curve of alanine.

As the pH increases still further with addition of more base, the protonated amino group (a weak acid) loses its proton, and the alanine molecule now has a negative charge of 1. The titration curve of alanine is that of a diprotic acid (Figure 3.6).

In histidine, the imidazole side chain also contributes a titratable group. At very low pH values, the histidine molecule has a net positive charge of 2 because both the imidazole and amino groups have positive charges. As base is added and the pH increases, the carboxyl group loses a proton to become a carboxylate as before, and the histidine now has a positive charge of 1 (Figure 3.5b). As still more base is added, the charged imidazole group loses its proton, and this is the point at which the histidine has no net charge. At still higher values of pH, the amino group loses its proton, as was the case with alanine, and the histidine molecule now has a negative charge of 1. The titration curve of histidine is that of a triprotic acid (Figure 3.7).

Like the acids we discussed in Chapter 2, the titratable groups of each of the amino acids have characteristic  $pK_a$  values. The  $pK_a$  values of  $\alpha$ -carboxyl groups are fairly low, around 2. The  $pK_a$  values of amino groups are much higher, with values ranging from 9 to 10.5. The  $pK_a$  values of side-chain groups, including side-chain carboxyl and amino groups, depend on the groups' chemical nature. Table 3.2 lists the  $pK_a$  values of the titratable groups of the amino acids. The classification of an amino acid as acidic or basic depends on the  $pK_a$  of the side chain as well as the chemical nature of the group. Histidine, lysine, and arginine are considered basic amino acids because each of their side chains has a nitrogen-containing group that can exist in either a protonated or deprotonated form. However, histidine has a  $pK_a$  in the acidic range. Aspartic acid and glutamic acid are considered acidic because each has a carboxylic acid side chain with a low  $pK_a$  value. These groups can still be titrated after the amino acid is incorporated into a peptide or protein, but the  $pK_a$  of the titratable group on the side chain is not necessarily the same in a protein as it is in a free amino acid. In fact, it can be very different. For example, a  $pK_a$  of 9 has been reported for an aspartate side chain in the protein thioredoxin.

The fact that amino acids, peptides, and proteins have different  $pK_a$  values gives rise to the possibility that they can have different charges at a given pH. Alanine and histidine, for example, both have net charges of  $-1$  at



■ **FIGURE 3.7** The titration curve of histidine. The isoelectric pH (pI) is the value at which positive and negative charges are the same. The molecule has no net charge.



TABLE 3.2

pK <sub>a</sub> Values of Common Amino Acids			
Acid	$\alpha$ -COOH	$\alpha$ -NH <sub>3</sub> <sup>+</sup>	RH or RH <sup>+</sup>
Gly	2.34	9.60	
Ala	2.34	9.69	
Val	2.32	9.62	
Leu	2.36	9.68	
Ile	2.36	9.68	
Ser	2.21	9.15	
Thr	2.63	10.43	
Met	2.28	9.21	
Phe	1.83	9.13	
Trp	2.38	9.39	
Asn	2.02	8.80	
Gln	2.17	9.13	
Pro	1.99	10.6	
Asp	2.09	9.82	3.86*
Glu	2.19	9.67	4.25*
His	1.82	9.17	6.0*
Cys	1.71	10.78	8.33*
Tyr	2.20	9.11	10.07
Lys	2.18	8.95	10.53
Arg	2.17	9.04	12.48

\* For these amino acids, the R group ionization occurs before the  $\alpha$ -NH<sub>3</sub><sup>+</sup> ionization.

high pH, above 10; the only charged group is the carboxylate anion. At lower pH, around 5, alanine is a zwitterion with no net charge, but histidine has a net charge of 1 at this pH because the imidazole group is protonated. This property is useful in **electrophoresis**, a common method for separating molecules in an electric field. This method is extremely useful in determining the important properties of proteins and nucleic acids. We shall see the applications to proteins in Chapter 5 and to nucleic acids in Chapter 14. The pH at which a molecule has no net charge is called the **isoelectric pH**, or isoelectric point (given the symbol **pI**). At its isoelectric pH, a molecule will not migrate in an electric field. This property can be put to use in separation methods. The pI of an amino acid can be calculated by the following equation:

$$\text{pI} = \frac{\text{p}K_{\text{a}1} + \text{p}K_{\text{a}2}}{2}$$

Most of the amino acids have only two pK<sub>a</sub> values, so this equation is easily used to calculate the pI. For the acidic and basic amino acids, however, we must be sure to average the correct pK<sub>a</sub> values. The pK<sub>a1</sub> is for the functional group that has dissociated at its isoelectric point. If two groups are dissociated at isoelectric pH, the pK<sub>a1</sub> is the higher pK<sub>a</sub> of the two. Therefore, pK<sub>a2</sub> is for the group that has not dissociated at isoelectric pH. If there are two groups that are not dissociated, the one with the lower pK<sub>a</sub> is used. See the following apply your knowledge exercise.

### Apply Your Knowledge

#### Amino Acid Titrations

1. Which of the following amino acids has a net charge of +2 at low pH? Which has a net charge of -2 at high pH? Aspartic acid, alanine, arginine, glutamic acid, leucine, lysine.
2. What is the pI for histidine?

#### Solution

Notice that the first part of this exercise deals only with the qualitative description of the successive loss of protons by the titratable groups on the individual amino acids. In the second part, you need to refer to the titration curve as well to do a numerical calculation of pH values.

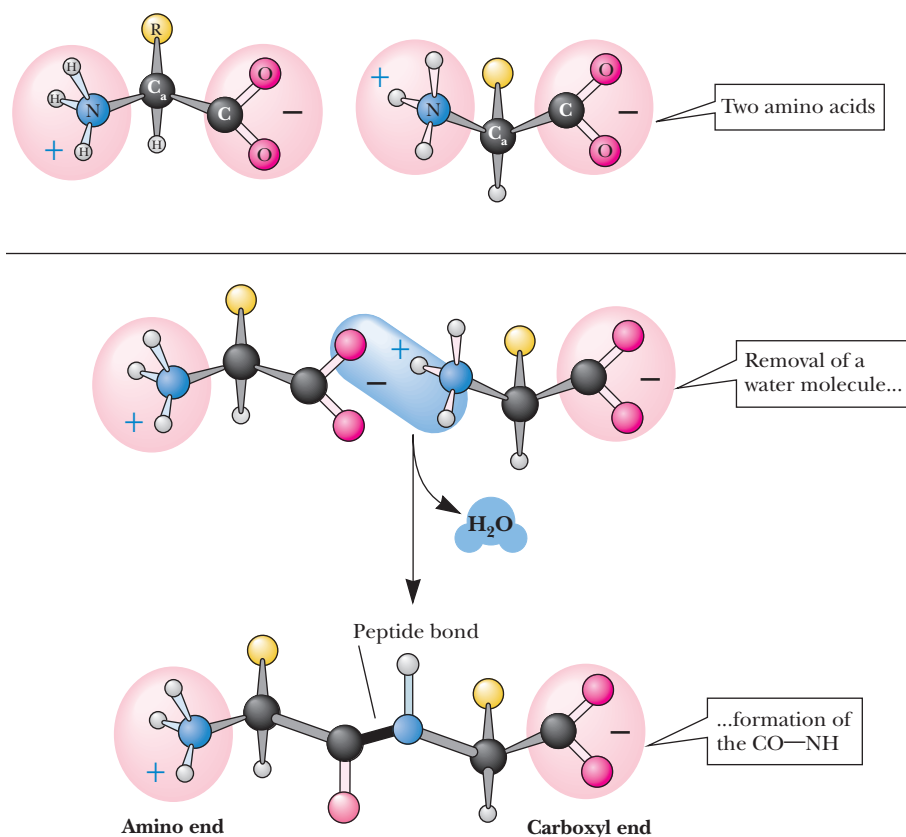
1. Arginine and lysine have net charges of +2 at low pH because of their basic side chains; aspartic acid and glutamic acid have net charges of -2 at high pH because of their carboxylic acid side chains. Alanine and leucine do not fall into either category because they do not have titratable side chains.
2. Draw or picture histidine at very low pH. It will have the formula shown in Figure 3.5b on the far left side. This form has a net charge of +2. To arrive at the isoelectric point, we must add some negative charge or remove some positive charge. This will happen in solution in order of increasing  $pK_a$ . Therefore, we begin by taking off the hydrogen from the carboxyl group because it has the lowest  $pK_a$  (1.82). This leaves us with the form shown second from the left in Figure 3.5b. This form has a charge of +1, so we must remove yet another hydrogen to arrive at the isoelectric form. This hydrogen would come from the imidazole side chain because it has the next highest  $pK_a$  (6.0); this is the isoelectric form (second from right). Now we average the  $pK_a$  from the highest  $pK_a$  group that lost a hydrogen with that of the lowest  $pK_a$  group that still retains its hydrogen. In the case of histidine, the numbers to substitute in the equation for the pI are 6.0 [ $pK_{a1}$ ] and 9.17 [ $pK_{a2}$ ], which gives a pI of 7.58.

## 3.4 The Peptide Bond

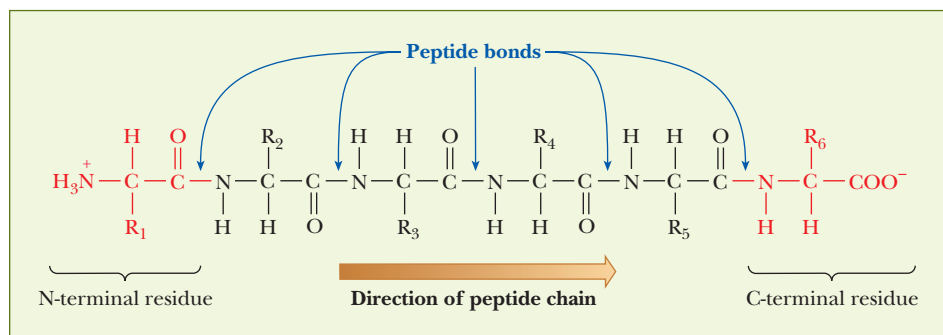
### Which groups on amino acids react to form a peptide bond?

Individual amino acids can be linked by forming covalent bonds. The bond is formed between the  $\alpha$ -carboxyl group of one amino acid and the  $\alpha$ -amino group of the next one. Water is eliminated in the process, and the linked amino acid **residues** remain after water is eliminated (Figure 3.8). A bond formed in this way is called a **peptide bond**. **Peptides** are compounds formed by linking small numbers of amino acids, ranging from two to several dozen. In a protein, many amino acids (usually more than a hundred) are linked by peptide bonds to form a **polypeptide chain** (Figure 3.9). Another name for a compound formed by the reaction between an amino group and a carboxyl group is an *amide*.

The carbon–nitrogen bond formed when two amino acids are linked in a peptide bond is usually written as a single bond, with one pair of electrons shared between the two atoms. With a simple shift in the position of a pair of electrons, it is quite possible to write this bond as a double bond. This shifting of electrons is well known in organic chemistry and results in **resonance structures**, structures that differ from one another only in the positioning of electrons. The positions of double and single bonds in one resonance structure are different from their positions in another resonance structure of the same compound. No single resonance structure actually represents the bonding in the compound; instead all resonance structures contribute to the bonding situation.



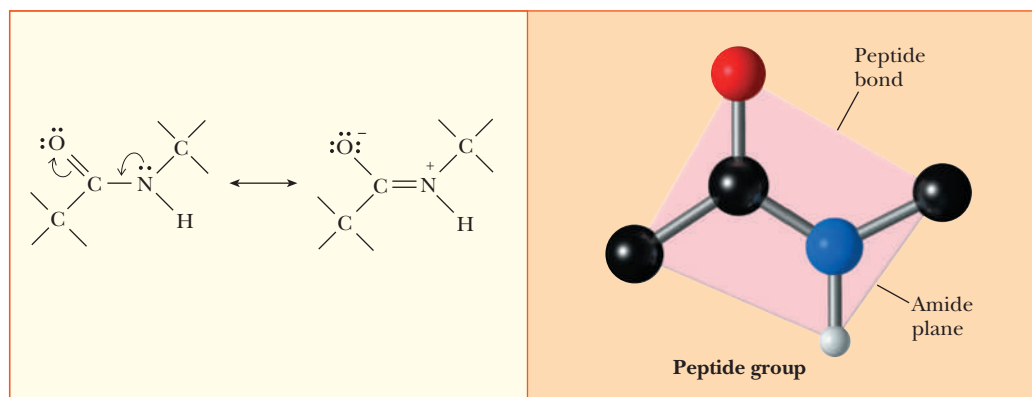
■ **FIGURE 3.8** Formation of the peptide bond. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)



■ **FIGURE 3.9** A small peptide showing the direction of the peptide chain (N-terminal to C-terminal).

The peptide bond can be written as a resonance hybrid of two structures (Figure 3.10), one with a single bond between the carbon and nitrogen and the other with a double bond between the carbon and nitrogen. The peptide bond has partial double bond character. As a result, the peptide group that forms the link between the two amino acids is planar. The peptide bond is also stronger than an ordinary single bond because of this resonance stabilization.

This structural feature has important implications for the three-dimensional conformations of peptides and proteins. There is free rotation around the bonds between the  $\alpha$ -carbon of a given amino acid residue and the amino nitrogen and carbonyl carbon of that residue, but there is no significant rotation



**A** Resonance structures of the peptide group.

**B** The planar peptide group.

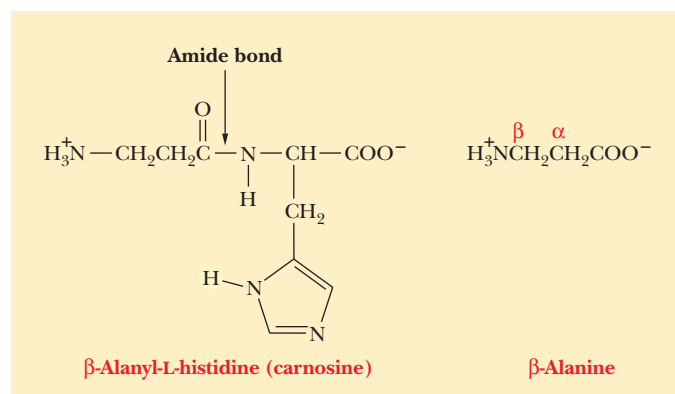
■ **FIGURE 3.10** The resonance structures of the peptide bond lead to a planar group.  
(Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

around the peptide bond. This stereochemical constraint plays an important role in determining how the protein backbone can fold.

## 3.5 Small Peptides with Physiological Activity

### What are some biological functions of small peptides?

The simplest possible covalently bonded combination of amino acids is a dipeptide, in which two amino acid residues are linked by a peptide bond. An example of a naturally occurring dipeptide is carnosine, which is found in muscle tissue. This compound, which has the alternative name  $\beta$ -alanyl-L-histidine, has an interesting structural feature. (In the systematic nomenclature of peptides, the **N-terminal** amino acid residue—the one with the free amino group—is given first; then other residues are given as they occur in sequence. The **C-terminal** amino acid residue—the one with the free carboxyl group—is given last.) The N-terminal amino acid residue,  $\beta$ -alanine, is structurally different from the  $\alpha$ -amino acids we have seen up to now. As the name implies, the amino group is bonded to the third or  $\beta$ -carbon of the alanine (Figure 3.11).



■ **FIGURE 3.11** Structures of carnosine and its component amino acid  $\beta$ -alanine.

# Biochemical Connections ORGANIC CHEMISTRY

## Amino Acids Go Many Different Places

### Why Are Amino Acids Featured in Health Food Stores?

Amino acids have biological functions other than as parts of proteins and oligopeptides. The following examples illustrate some of these functions for a few of the amino acids and their derivatives.

### Branched-Chain Amino Acids

Some products sold in health food stores feature the presence of the branched-chain amino acids isoleucine, leucine, and valine. These are *essential* amino acids in the sense that the body cannot synthesize them. Under normal circumstances, a diet with adequate protein intake provides enough of all the essential amino acids. Athletes involved in intensive training want to prevent muscle loss and to increase muscle mass. As a result, they take protein supplements and pay particular attention to branched-chain amino acids. (These three amino acids are by no means the only essential ones, but they are mentioned specifically here.)

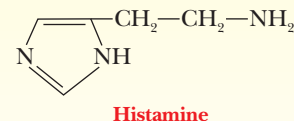
Even a very quick search of the Internet will turn up many articles about branched-chain amino acids, with numerous opportunities to order a supply.

### Glutamic Acid

Monosodium glutamate, or MSG, is a derivative of glutamic acid that finds wide use as a flavor enhancer. MSG causes a physiological reaction in some people, with chills, headaches, and dizziness resulting. Because many Asian foods contain significant amounts of MSG, this problem is often referred to as *Chinese restaurant syndrome*.

### Histidine

If the acid group of histidine is removed, it is converted to histamine, which is a potent vasodilator, increasing the diameter of blood vessels. Histamine, which is released as part of the immune response, increases the localized blood volume for white blood cells. This results in the swelling and stuffiness that are associated with a cold. Most cold medications contain antihistamines to overcome this stuffiness.



### Serotonin

The tryptophan derivative serotonin (see the Biochemical Connection on page 68), plays a key role in sudden infant death syndrome (SIDS). In the developed world, more newborns die as a result of SIDS than any other cause. It is known from autopsies that brainstem cells involved in processing serotonin are altered in infants who die of SIDS. It is also known that a gene that is overexpressed in these cells encodes a receptor for serotonin on the cell surface. The receptor mediates the transmission of serotonin among cells. The autonomic nervous system, controlling heart rate, body temperature, and respiration, is affected. This information provides the basis for research on the prevention of SIDS.

## Apply Your Knowledge

### Sequence of Peptides

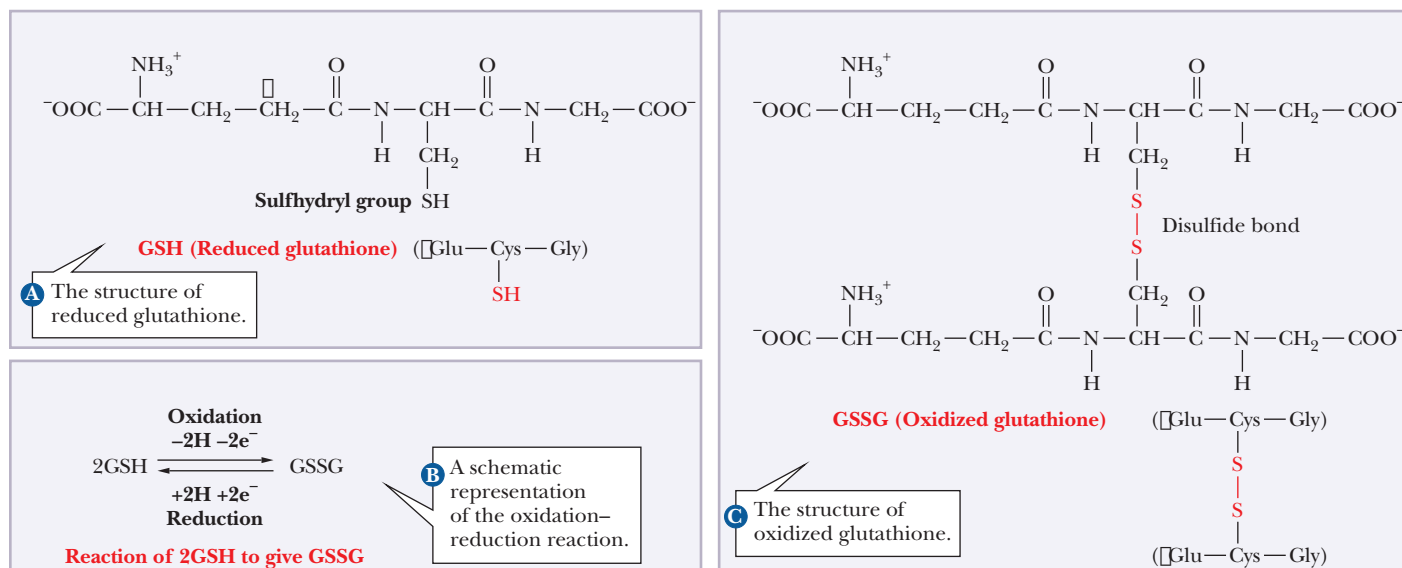
Write an equation with structures for the formation of a dipeptide when alanine reacts with glycine to form a peptide bond. Is there more than one possible product for this reaction?

### Solution

The main point here is to be aware of the possibility that amino acids can be linked together in more than one order when they form peptide bonds. Thus, there are two possible products when alanine and glycine react: alanyl-glycine, in which alanine is at the N-terminal end and glycine is at the C-terminal end, and glycyl-alanine, in which glycine is at the N-terminal end and alanine is at the C-terminal end.

Glutathione is a commonly occurring tripeptide; it has considerable physiological importance because it is a scavenger for oxidizing agents. Recall from Section 1.9 that oxidation is the loss of electrons; an oxidizing agent causes another substance to lose electrons. (It is thought that some oxidizing agents are harmful to organisms and play a role in the development of cancer.) In terms of its amino acid composition and bonding order, it is  $\gamma$ -glutamyl-L-cysteinylglycine (Figure 3.12a). The letter  $\gamma$  (gamma) is the third letter in





■ **FIGURE 3.12** The oxidation and reduction of glutathione. (a) The structure of reduced glutathione. (b) A schematic representation of the oxidation–reduction reaction. (c) The structure of oxidized glutathione.

the Greek alphabet; in this notation, it refers to the third carbon atom in the molecule, counting the one bonded to the amino group as the first. Once again, the N-terminal amino acid is given first. In this case, the  $\gamma$ -carboxyl group (the side-chain carboxyl group) of the glutamic acid is involved in the peptide bond; the amino group of the cysteine is bonded to it. The carboxyl group of the cysteine is bonded, in turn, to the amino group of the glycine. The carboxyl group of the glycine forms the other end of the molecule, the C-terminal end. The glutathione molecule shown in Figure 3.12a is the reduced form. It scavenges oxidizing agents by reacting with them. The oxidized form of glutathione is generated from two molecules of the reduced peptide by forming a disulfide bond between the  $\text{—SH}$  groups of the two cysteine residues (Figure 3.12b). The full structure of oxidized glutathione is shown in Figure 3.12c.

Two pentapeptides found in the brain are known as enkephalins, naturally occurring analgesics (pain relievers). For molecules of this size, abbreviations for the amino acids are more convenient than structural formulas. The same notation is used for the amino acid sequence, with the N-terminal amino acid listed first and the C-terminal listed last. The two peptides in question, leucine enkephalin and methionine enkephalin, differ only in their C-terminal amino acids.

Tyr—Gly—Gly—Phe—Leu (three-letter abbreviations)

Y—G—G—F—L (one-letter abbreviations)

Leucine enkephalin

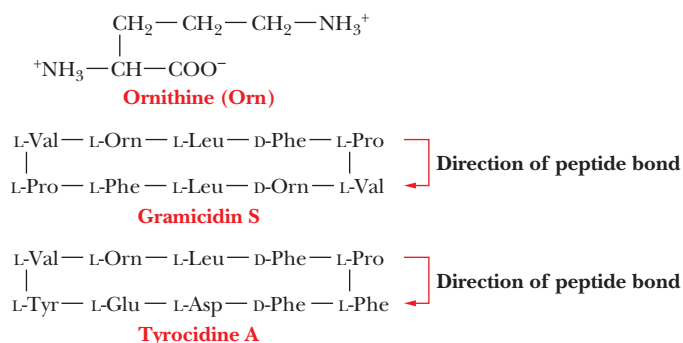
Tyr—Gly—Gly—Phe—Met

Y—G—G—F—M

Methionine enkephalin

It is thought that the aromatic side chains of tyrosine and phenylalanine in these peptides play a role in their activities. It is also thought that there are similarities between the three-dimensional structures of opiates, such as morphine, and those of the enkephalins. As a result of these structural similarities,





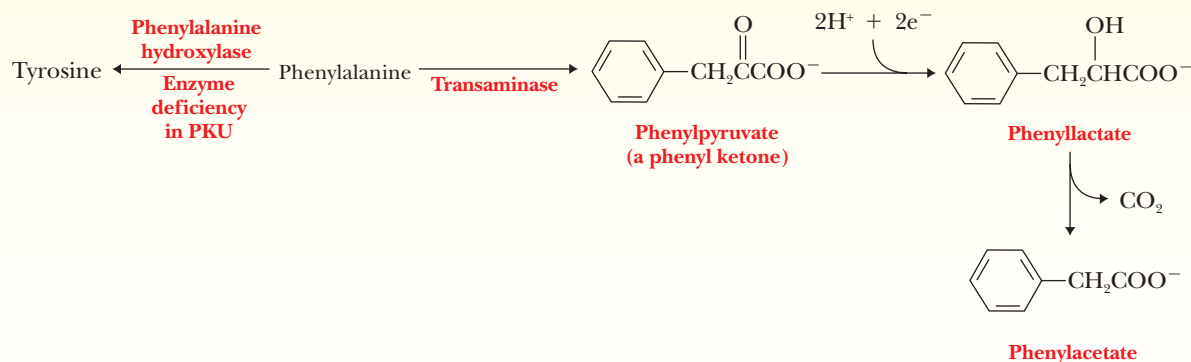
■ **FIGURE 3.14** Structures of ornithine, gramicidin S, and tyrocidine A.

## Biochemical Connections ALLIED HEALTH

### Phenylketonuria—Little Molecules Have Big Effects

Mutations leading to deficiencies in enzymes are usually referred to as “inborn errors of metabolism,” because they involve defects in the DNA of the affected individual. Errors in enzymes that catalyze reactions of amino acids frequently have disastrous consequences, many of them leading to severe forms of mental retardation. Phenylketonuria (PKU) is a well-known example. In this condition, phenylalanine, phenylpyruvate, phenyllactate, and phenylacetate all accumulate in the blood and urine. Available evidence suggests that phenylpyruvate, which is a phenylketone, causes mental retardation by interfering with the conversion of pyruvate to acetyl-CoA (an important intermediate in many biochemical reactions) in the brain. It is also likely that the accumulation of these products in the brain cells results in an osmotic imbalance in which water flows into the brain cells. These cells expand in size until they crush each other in the developing brain. In either case, the brain is not able to develop normally.

Fortunately, PKU can be easily detected in newborns, and all 50 states and the District of Columbia mandate that such a test be performed because it is cheaper to treat the disease with a modified diet than to cope with the costs of a mentally retarded individual who is usually institutionalized for life. The dietary changes are relatively simple. Phenylalanine must be limited to the amount needed for protein synthesis, and tyrosine must now be supplemented, because phenylalanine is no longer a source. You may have noticed that foods containing aspartame carry a warning about the phenylalanine portion of that artificial sweetener. A substitute for aspartame, which carries the trade name Alatame, contains alanine rather than phenylalanine. It has been introduced to retain the benefits of aspartame without the dangers associated with phenylalanine.



■ Reactions involved in the development of phenylketonuria (PKU). A deficiency in the enzyme that catalyzes the conversion of phenylalanine to tyrosine leads to the accumulation of phenylpyruvate, a phenyl ketone.

# Biochemical Connections NEUROSCIENCE

## Peptide Hormones Revisited

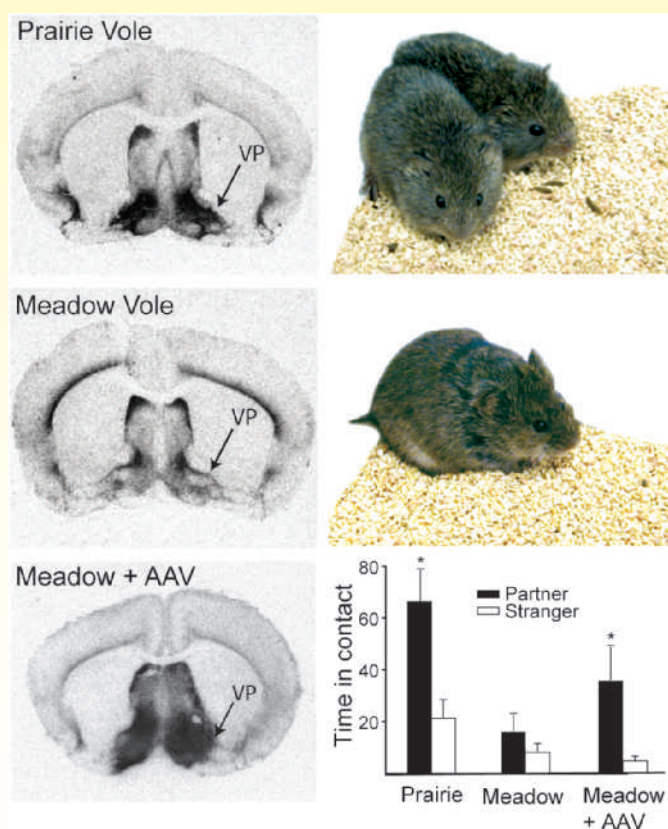
The fundamental physiological roles of the peptides oxytocin and vasopressin have been known for decades. More recently, it has been becoming clear that they also play roles in behavior, even quite complex social behavior in mammalian species. In the Biochemical Connection box on page 77, we concentrated on the role of oxytocin, with some mention of vasopressin. In this box, we are going to concentrate on vasopressin, with some mention of oxytocin.

Vasopressin binds to two kinds of receptor sites in the brain, one of which has profound effects on social behavior. Studies on small animals such as voles indicate that mating behavior is affected when the receptor level is artificially elevated. Genetic variations known as polymorphisms are widely distributed in the gene that encodes the receptor in these animals. The level of polymorphisms correlates with activation of areas of the brain where vasopressin receptors occur. The figure shows that receptor levels differ in monogamous prairie voles and polygamous meadow voles. Using a viral vector (AAV in the figure), researchers transferred the prairie vole gene to meadow voles using standard genetic engineering techniques (Chapter 13). Receptor patterns in the brain were different in the genetically altered meadow voles. In addition, the bar chart shows that these meadow voles tended to adopt the monogamous behavior pattern of the prairie voles.

Other researchers correlated genetic analysis of human subjects with studies of social behavior. A clear connection was found between mutations and socialization skills. Marital difficulties in particular were more frequently encountered with one polymorphism that has also attracted the attention of researchers who study autism. No firm connection with autism has been established, however.

In experiments with human volunteers, oxytocin and vasopressin were both administered by nasal spray. Psychological tests then measured trust, using various scenarios based on investment. Subjects who received oxytocin tended to have higher levels of trust than controls.

Behavior is a complex topic, especially in humans. Studies such as these raise at least as many questions as they answer. A full understanding of the molecular basis of behavior is far in the future, but studies like these are first steps in that direction.



■ Left: Vasopressin receptor patterns in the brains of prairie voles, meadow voles, and meadow voles genetically altered to express the prairie vole gene. Right: Time in minutes spent in contact with a partner or with a stranger in the three groups of voles. (From *Oxytocin, Vasopressin, and the Neurogenetics of Sociality* by Zoe R. Donaldson and Larry J. Young (7 November 2008) *Science* 322 (5903), 900. Reprinted with permission from AAAS.)

## SUMMARY

### Why is it important to specify the three-dimensional structure of amino acids?

The amino acids that are the monomer units of proteins have a general structure in common, with an amino group and a carboxyl group bonded to the same carbon atom. The nature of the side chains, which are referred to as R groups, is the basis of the differences among amino acids. Except for glycine, amino acids can exist in two forms, designated L and D. These two stereoisomers are nonsuperimposable mirror images of each other. The amino acids found in proteins are of the L form, but some D-amino acids occur in nature.

**Why are amino acid side chains so important?** A classification scheme for amino acids can be based on the properties of their side chains. Two particularly important criteria are the polar or nonpolar nature of the side chain and the presence of an acidic or basic group in the side chain.

**Which amino acids have nonpolar side chains? (Group 1)** One group of amino acids has nonpolar side chains. The side chains are mostly aliphatic or aromatic hydrocarbons or their derivatives.



**Which amino acids have electrically neutral polar side chains? (Group 2)** A second group of amino acids has side chains that contain electronegative atoms such as oxygen, nitrogen, and sulfur.

**Which amino acids have carboxyl groups in their side chains? (Group 3)** Two amino acids, glutamic acid and aspartic acid, have carboxyl groups in their side chains.

**Which amino acids have basic side chains? (Group 4)** Three amino acids—histidine, lysine, and arginine—have basic side chains.

**Which amino acids are found less commonly in proteins?** Some amino acids are found only in a few proteins. They are formed from the common ones after the protein has been synthesized in the cell.

**What happens when we titrate an amino acid?** In free amino acids at neutral pH, the carboxylate group is negatively charged (acidic) and the amino group is positively charged

(basic). Amino acids without charged groups on their side chains exist in neutral solution as zwitterions, with no net charge. Titration curves of amino acids indicate the pH ranges in which titratable groups gain or lose a proton. Side chains of amino acids can also contribute titratable groups; the charge (if any) on the side chain must be taken into consideration in determining the net charge on the amino acid.

**Which groups on amino acids react to form a peptide bond?** Peptides are formed by reacting the carboxyl group of one amino acid with the amino group of another amino acid in a covalent (amide) bond. Proteins consist of polypeptide chains; the number of amino acids in a protein is usually 100 or more. The peptide group is planar; this stereochemical constraint plays an important role in determining the three-dimensional structures of peptides and proteins.

**What are some biological functions of small peptides?** Small peptides, containing two to several dozen amino acid residues, can have marked physiological effects in organisms.

## REVIEW EXERCISES

**OWL** Interactive versions of these problems are assignable in OWL

### 3.1 Amino Acids Exist in a Three-Dimensional World

1. **Recall** How do D-amino acids differ from L-amino acids? What biological roles are played by peptides that contain D-amino acids?

### 3.2 Individual Amino Acids: Their Structures and Properties

2. **Recall** Which amino acid is technically *not* an amino acid? Which amino acid contains no chiral carbon atoms?
3. **Recall** For each of the following, name an amino acid in which the R group contains it: a hydroxyl group, a sulfur atom, a second chiral carbon atom, an amino group, an amide group, an acid group, an aromatic ring, and a branched side chain.
4. **Recall** Identify the polar amino acids, the aromatic amino acids, and the sulfur-containing amino acids, given a peptide with the following amino acid sequence:

Val—Met—Ser—Ile—Phe—Arg—Cys—Tyr—Leu

5. **Recall** Identify the nonpolar amino acids and the acidic amino acids in the following peptide:

Glu—Thr—Val—Asp—Ile—Ser—Ala

6. **Recall** Are amino acids other than the usual 20 amino acids found in proteins? If so, how are such amino acids incorporated into proteins? Give an example of such an amino acid and a protein in which it occurs.

### 3.3 Amino Acids Can Act as Both Acids and Bases

7. **Mathematical** Predict the predominant ionized forms of the following amino acids at pH 7: glutamic acid, leucine, threonine, histidine, and arginine.
8. **Mathematical** Draw structures of the following amino acids, indicating the charged form that exists at pH 4: histidine, asparagine, tryptophan, proline, and tyrosine.
9. **Mathematical** Predict the predominant forms of the amino acids from Question 8 at pH 10.

10. **Mathematical** Calculate the isoelectric point of each of the following amino acids: glutamic acid, serine, histidine, lysine, tyrosine, and arginine.
11. **Mathematical** Sketch a titration curve for the amino acid cysteine, and indicate the  $pK_a$  values for all titratable groups. Also indicate the pH at which this amino acid has no net charge.
12. **Mathematical** Sketch a titration curve for the amino acid lysine, and indicate the  $pK_a$  values for all titratable groups. Also indicate the pH at which the amino acid has no net charge.
13. **Mathematical** An organic chemist is generally happy with 95% yields. If you synthesized a polypeptide and realized a 95% yield with each amino acid residue added, what would be your overall yield after adding 10 residues (to the first amino acid)? After adding 50 residues? After 100 residues? Would these low yields be biochemically “satisfactory”? How are low yields avoided, biochemically?
14. **Mathematical** Sketch a titration curve for aspartic acid, and indicate the  $pK_a$  values of all titratable groups. Also indicate the pH range in which the conjugate acid–base pair +1 Asp and 0 Asp will act as a buffer.
15. **Reflect and Apply** Suggest a reason why amino acids are usually more soluble at pH extremes than they are at neutral pH. (Note that this does not mean that they are insoluble at neutral pH.)
16. **Reflect and Apply** Write equations to show the ionic dissociation reactions of the following amino acids: aspartic acid, valine, histidine, serine, and lysine.
17. **Reflect and Apply** Based on the information in Table 3.2, is there any amino acid that could serve as a buffer at pH 8? If so, which one?
18. **Reflect and Apply** If you were to have a mythical amino acid based on glutamic acid, but one in which the hydrogen that is attached to the  $\gamma$ -carbon were replaced by another amino group, what would be the predominant form of this amino acid at pH 4, 7, and 10, if the  $pK_a$  value were 10 for the unique amino group?



19. **Reflect and Apply** What would be the pI for the mythical amino acid described in Question 18?
  20. **Reflect and Apply** Identify the charged groups in the peptide shown in Question 4 at pH 1 and at pH 7. What is the net charge of this peptide at these two pH values?
  21. **Reflect and Apply** Consider the following peptides: Phe—Glu—Ser—Met and Val—Trp—Cys—Leu. Do these peptides have different net charges at pH 1? At pH 7? Indicate the charges at both pH values.
  22. **Reflect and Apply** In each of the following two groups of amino acids, which amino acid would be the easiest to distinguish from the other two amino acids in the group, based on a titration?
    - (a) gly, leu, lys
    - (b) glu, asp, ser
  23. **Reflect and Apply** Could the amino acid glycine serve as the basis of a buffer system? If so, in what pH range would it be useful?
- ### 3.4 The Peptide Bond
24. **Recall** Sketch resonance structures for the peptide group.
  25. **Recall** How do the resonance structures of the peptide group contribute to the planar arrangement of this group of atoms?
  26. **Biochemical Connections** Which amino acids or their derivatives are neurotransmitters?
  27. **Biochemical Connections** What is a monoamine oxidase, and what function does it serve?
  28. **Reflect and Apply** Consider the peptides Ser—Glu—Gly—His—Ala and Gly—His—Ala—Glu—Ser. How do these two peptides differ?
  29. **Reflect and Apply** Would you expect the titration curves of the two peptides in Question 28 to differ? Why or why not?
  30. **Reflect and Apply** What are the sequences of all the possible tripeptides that contain the amino acids aspartic acid, leucine, and phenylalanine? Use the three-letter abbreviations to express your answer.
  31. **Reflect and Apply** Answer Question 30 using one-letter designations for the amino acids.
  32. **Reflect and Apply** Most proteins contain more than 100 amino acid residues. If you decided to synthesize a “100-mer,” with 20 different amino acids available for each position, how many different molecules could you make?
  33. **Biochemical Connections** What are the structural differences between oxytocin and vasopressin? How do they differ in physiological function?
  34. **Biochemical Connections** Why might a glass of warm milk help you sleep at night?
  35. **Biochemical Connections** Which would be better to eat before an exam, a glass of milk or a piece of cheese? Why?
  36. **Reflect and Apply** What might you infer (or know) about the stability of amino acids, when compared with that of other building-block units of biopolymers (sugars, nucleotides, fatty acids, etc.)?
  37. **Reflect and Apply** If you knew everything about the properties of the 20 common (proteinous) amino acids, would you be able to predict the properties of a protein (or large peptide) made from them?
  38. **Reflect and Apply** Suggest a reason why the amino acids thyroxine and hydroxyproline are produced by posttranslational modification of the amino acids tyrosine and proline, respectively.
  39. **Reflect and Apply** Consider the peptides Gly—Pro—Ser—Glu—Thr (open chain) and Gly—Pro—Ser—Glu—Thr with a peptide bond linking the threonine and the glycine. Are these peptides chemically the same?
  40. **Reflect and Apply** Can you expect to separate the peptides in Question 39 by electrophoresis?
  41. **Reflect and Apply** Suggest a reason why biosynthesis of amino acids and of proteins would eventually cease in an organism with carbohydrates as its only food source.
  42. **Reflect and Apply** You are studying with a friend who draws the structure of alanine at pH 7. It has a carboxyl group ( $-\text{COOH}$ ) and an amino group ( $-\text{NH}_2$ ). What suggestions would you make?
  43. **Reflect and Apply** Suggest a reason (or reasons) why amino acids polymerize to form proteins that have comparatively few covalent crosslinks in the polypeptide chain.
  44. **Reflect and Apply** Suggest the effect on the structure of peptides if the peptide group were not planar.
  45. **Reflect and Apply** Speculate on the properties of proteins and peptides if none of the common amino acids contained sulfur.
  46. **Reflect and Apply** Speculate on the properties of proteins that would be formed if amino acids were not chiral.
- ### 3.5 Small Peptides with Physiological Activity
47. **Recall** What are the structural differences between the peptide hormones oxytocin and vasopressin? How do they differ in function?
  48. **Recall** How do the oxidized and reduced forms of glutathione differ from each other?
  49. **Recall** What is an enkephalin?
  50. **Reflect and Apply** The enzyme D-amino acid oxidase, which converts D-amino acids to their  $\alpha$ -keto form, is one of the most potent enzymes in the human body. Suggest a reason why this enzyme should have such a high rate of activity.

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# The Three-Dimensional Structure of Proteins

# 4

Red blood cells contain hemoglobin, a classic example of protein structure.

© Dr. Philippa Uwins, Whistler Research Pty/  
Photo Researchers, Inc.

## 4.1 Protein Structure and Function

Biologically active proteins are polymers consisting of amino acids linked by covalent peptide bonds. Many different conformations (three-dimensional structures) are possible for a molecule as large as a protein. Of these many structures, one or (at most) a few have biological activity; these are called the **native conformations**. Many proteins have no obvious regular repeating structure. As a consequence, these proteins are frequently described as having large segments of “random structure” (also referred to as *random coil*). The term *random* is really a misnomer, because the same nonrepeating structure is found in the native conformation of all molecules of a given protein, and this conformation is needed for its proper function. Because proteins are complex, they are defined in terms of four levels of structure.

### What are the levels of protein structure?

**Primary structure** is the order in which the amino acids are covalently linked together. The peptide Leu—Gly—Thr—Val—Arg—Asp—His (recall that the N-terminal amino acid is listed first) has a different primary structure from the peptide Val—His—Asp—Leu—Gly—Arg—Thr, even though both have the same number and kinds of amino acids. Note that the order of amino acids can be written on one line. The primary structure is the one-dimensional first step in specifying the three-dimensional structure of a protein. Some biochemists define primary structure to include all covalent interactions, including the disulfide bonds that can be formed by cysteines; however, we shall consider the disulfide bonds to be part of the tertiary structure, which will be considered later.

Two three-dimensional aspects of a single polypeptide chain, called the secondary and tertiary structure, can be considered separately. **Secondary structure** is the arrangement in space of the atoms in the peptide backbone. The  $\alpha$ -helix and  $\beta$ -pleated sheet arrangements are two different types of secondary structure. Secondary structures have repetitive interactions resulting from hydrogen bonding between the amide N—H and the carbonyl groups of the peptide backbone. The conformations of the side chains of the amino acids are not part of the secondary structure. In many proteins, the folding of parts of the chain can occur independently of the folding of other parts. Such independently folded portions of proteins are referred to as **domains** or **super-secondary structure**.

**Tertiary structure** includes the three-dimensional arrangement of all the atoms in the protein, including those in the side chains and in any **prosthetic groups** (groups of atoms other than amino acids).

A protein can consist of multiple polypeptide chains called **subunits**. The arrangement of subunits with respect to one another is the **quaternary structure**. Interaction between subunits is mediated by noncovalent interactions, such as hydrogen bonds, electrostatic attractions, and hydrophobic interactions.

## Chapter Outline

### 4.1 Protein Structure and Function

- What are the levels of protein structure?

### 4.2 Primary Structure of Proteins

- Why is it important to know the primary structure?

### 4.3 Secondary Structure of Proteins

- Why is the  $\alpha$ -helix so prevalent?
- How is the  $\beta$ -sheet different from the  $\alpha$ -helix?

### 4.4 Tertiary Structure of Proteins

- How can the three-dimensional structure of a protein be determined?
- Why does oxygen have imperfect binding to the heme group?

### 4.5 Quaternary Structure of Proteins

- How does hemoglobin work?

### 4.6 Protein Folding Dynamics

- Can we predict the tertiary structure of a protein if we know its amino acid sequence?
- What makes hydrophobic interactions favorable?

**OWL** Online homework for this chapter may be assigned in OWL.

## 4.2 Primary Structure of Proteins

The amino acid sequence (the primary structure) of a protein determines its three-dimensional structure, which, in turn, determines its properties. In every protein, the correct three-dimensional structure is needed for correct functioning.

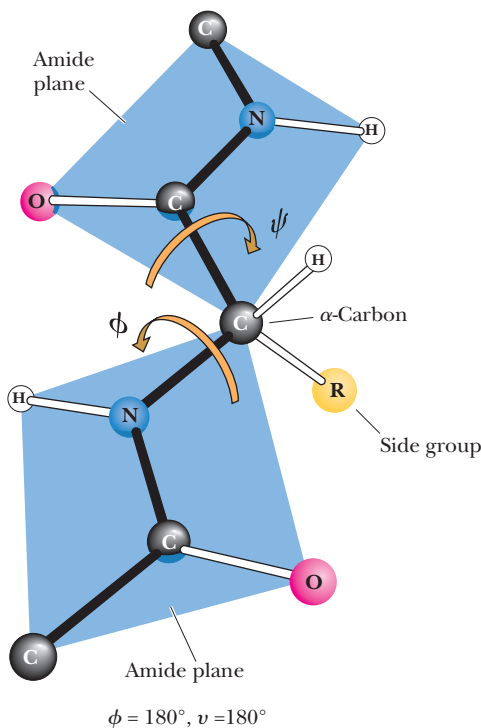
### Why is it important to know the primary structure?

One of the most striking demonstrations of the importance of primary structure is found in the hemoglobin associated with *sickle-cell anemia*. In this genetic disease, red blood cells cannot bind oxygen efficiently. The red blood cells also assume a characteristic sickle shape, giving the disease its name. The sickled cells tend to become trapped in small blood vessels, cutting off circulation and thereby causing organ damage. These drastic consequences stem from a change in one amino acid residue in the sequence of the primary structure.

Considerable research is being done to determine the effects of changes in primary structure on the functions of proteins. Using molecular-biology techniques, such as site-directed mutagenesis, it is possible to replace any chosen amino acid residue in a protein with another specific amino acid residue. The conformation of the altered protein, as well as its biological activity, can then be determined. The results of such amino acid substitutions range from negligible effects to complete loss of activity, depending on the protein and the nature of the altered residue.

Determining the sequence of amino acids in a protein is a routine, but not trivial, operation in classical biochemistry. It consists of several steps, which must be carried out carefully to obtain accurate results (Section 5.4).

The following Biochemical Connections box describes an important practical aspect of the amino acid composition of proteins. This property can differ markedly, depending on the source of the protein (plant or animal), with important consequences for human nutrition.



■ **FIGURE 4.1** Phi and psi angles of the peptide backbone. Definition of the angles that determine the conformation of a polypeptide chain. The rigid planar peptide groups (called “playing cards” in the text) are shaded. The angle of rotation around the C<sup>α</sup>—N bond is designated  $\phi$  (phi), and the angle of rotation around the C—C bond is designated  $\psi$  (psi). These two bonds are the ones around which there is freedom of rotation. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

## 4.3 Secondary Structure of Proteins

The secondary structure of proteins is the hydrogen-bonded arrangement of the backbone of the protein, the polypeptide chain. The nature of the bonds in the peptide backbone plays an important role here. Within each amino acid residue are two bonds with reasonably free rotation: (1) the bond between the  $\alpha$ -carbon and the amino nitrogen of that residue and (2) the bond between the  $\alpha$ -carbon and the carboxyl carbon of that residue. The combination of the planar peptide group and the two freely rotating bonds has important implications for the three-dimensional conformations of peptides and proteins. A peptide-chain backbone can be visualized as a series of playing cards, each card representing a planar peptide group. The cards are linked at opposite corners by swivels, representing the bonds about which there is considerable freedom of rotation (Figure 4.1). The side chains also play a vital role in determining the three-dimensional shape of a protein, but only the backbone is considered in the secondary structure. The angles  $\phi$  (phi) and  $\psi$  (psi), frequently called Ramachandran angles (after their originator, G. N. Ramachandran), are used to designate rotations around the C—N and C—C bonds, respectively. The conformation of a protein backbone can be described by specifying the values of  $\phi$  and  $\psi$  for each residue ( $-180^\circ$  to  $180^\circ$ ). Two kinds of secondary structures that occur frequently in proteins are the repeating  $\alpha$ -helix and  $\beta$ -pleated sheet (or  $\beta$ -sheet) hydrogen-bonded structures. The  $\phi$  and  $\psi$  angles repeat themselves in contiguous amino acids in regular secondary structures. The  $\alpha$ -helix and  $\beta$ -pleated sheet are not the only possible secondary structures, but they are by far the most important and deserve a closer look.

# Biochemical Connections NUTRITION

## Complete Proteins and Nutrition

A **complete protein** is one that provides all essential amino acids (Section 23.5) in appropriate amounts for human survival. These amino acids cannot be synthesized by humans, but they are needed for the biosynthesis of proteins. Lysine and methionine are two essential amino acids that are frequently in short supply in plant proteins.

Because grains such as rice and corn are usually poor in lysine, and because beans are usually poor in methionine, vegetarians are at risk for malnutrition unless they eat grains and beans together. This leads to the concept of *complementary proteins*, mixtures that provide all the essential amino acids—for example, corn and beans in succotash, or a bean burrito made with a corn tortilla. The specific recommended dietary allowances for adult males follow. Adult females who are neither pregnant nor lactating need 20% less than the amounts indicated for adult males.

RDA		RDA	
Arg*	Unknown	Met	0.70 g
His*	Unknown	Phe	1.12 (includes Tyr)
He	0.84 g	Thr	0.56 g
Leu	1.12 g	Trp	0.21 g
Lys	0.84 g	Val	0.96 g

\*The inclusion of His and Arg is controversial. They appear to be required only by growing children and for the repair of injured tissue. Arg is required to maintain fertility in males.

The *protein efficiency ratio* (PER) describes how well a protein supplies essential amino acids. This parameter is useful for deciding how much of a food you need to eat. Most college-age, nonpregnant females require 46 g (or about 1.6 oz) of complete protein, and males require 58 g (or about 2 oz) of complete protein per day. If one chooses only a *single* source of protein for the

diet, eggs are perhaps the best choice because they contain high-quality protein. For a female, the need for 1.6 oz of complete protein could be met with 10.7 oz of eggs, or about four whole extra-large eggs. For a male, 13.6 oz of eggs, or a little more than five eggs, would be needed. The same requirement could be met with a lean beef steak, but it would require 345 g, or about 0.75 lb, for a female (or 431 g, or nearly a full pound, for a male) because beef steak has a lower PER. If one ate only corn, it would require 1600 g/day for women and 2000 g/day for men (1600 g is about 3.6 pounds of fresh corn kernels—something in excess of 160 eight-inch ears per day). However, if you simply combine a small amount of beans or peas with the corn, it complements the low amount of lysine in the corn, and the protein is now complete. This can easily be done with normal food portions.

Protein	PER	% Protein
Whole egg	100	15
Beef muscle	84	16
Cow's milk	66	4 (largely H <sub>2</sub> O)
Peanuts	45	28
Corn	32	9
Wheat	26	12

In an attempt to increase the nutritional value of certain crops that are grown as food for livestock, scientists have used genetic techniques to create strains of corn that are much higher in lysine than the wild-type corn. This has proven effective in increasing growth rates in pigs. Many vegetable crops are now being produced using biotechnology to increase shelf life, decrease spoilage, and give crops defenses against insects. These genetically modified foods are currently a hot spot of debate and controversy.

## Periodic Structures in Protein Backbones

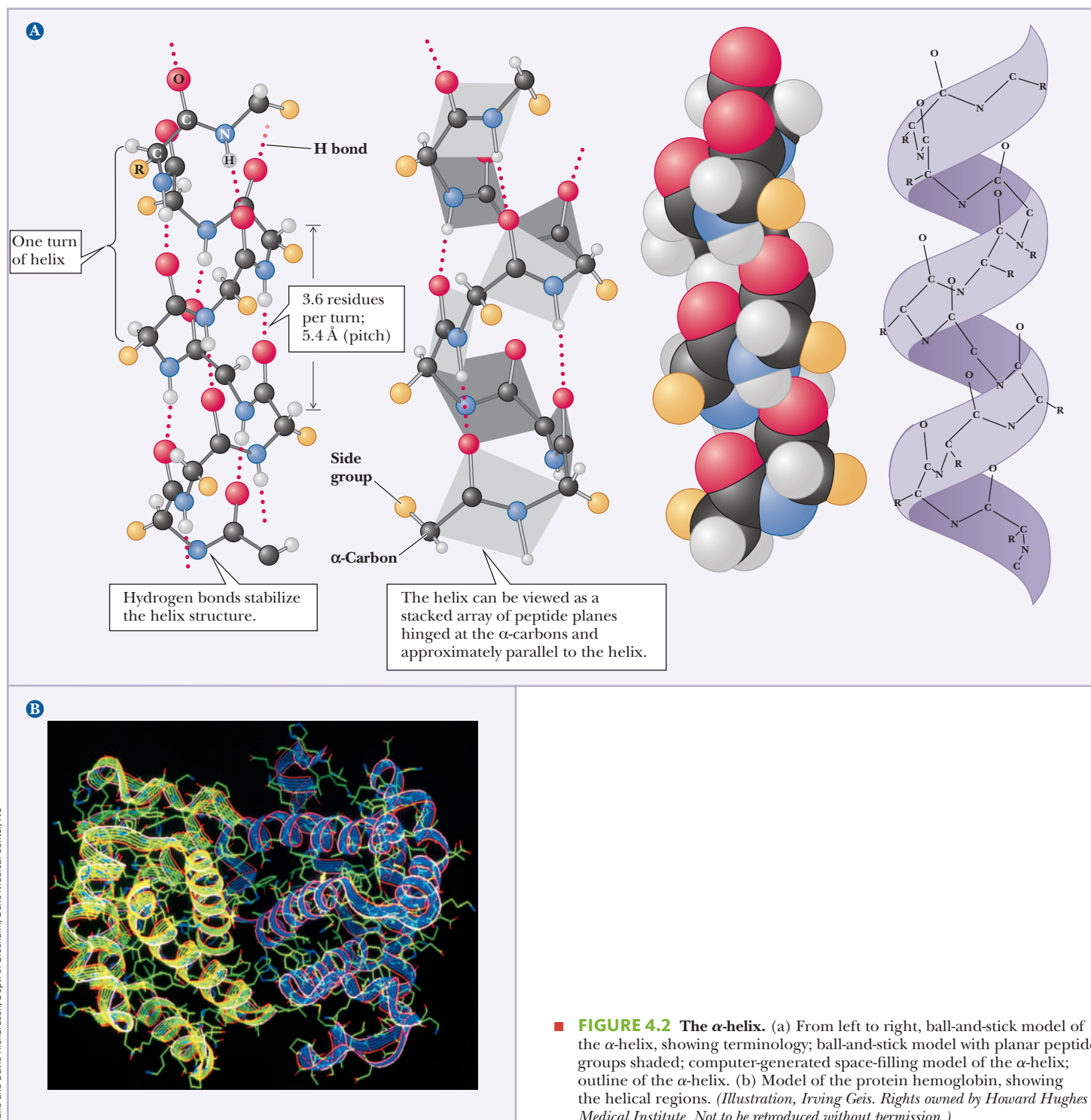
The  $\alpha$ -helix and  $\beta$ -pleated sheet are periodic structures; their features repeat at regular intervals. The  $\alpha$ -helix is rodlike and involves only one polypeptide chain. The  $\beta$ -pleated sheet structure can give a two-dimensional array and can involve one or more polypeptide chains.

### Why is the $\alpha$ -helix so prevalent?

The  $\alpha$ -helix is stabilized by hydrogen bonds parallel to the helix axis within the backbone of a single polypeptide chain. Counting from the N-terminal end, the C—O group of each amino acid residue is hydrogen bonded to the N—H group of the amino acid four residues away from it in the covalently bonded sequence. The helical conformation allows a linear arrangement of the atoms involved in the hydrogen bonds, which gives the bonds maximum strength and thus makes the helical conformation very stable (Section 2.2). There are 3.6 residues for each turn of the helix, and the *pitch* of the helix (the linear distance between corresponding points on successive turns) is 5.4 Å (Figure 4.2).

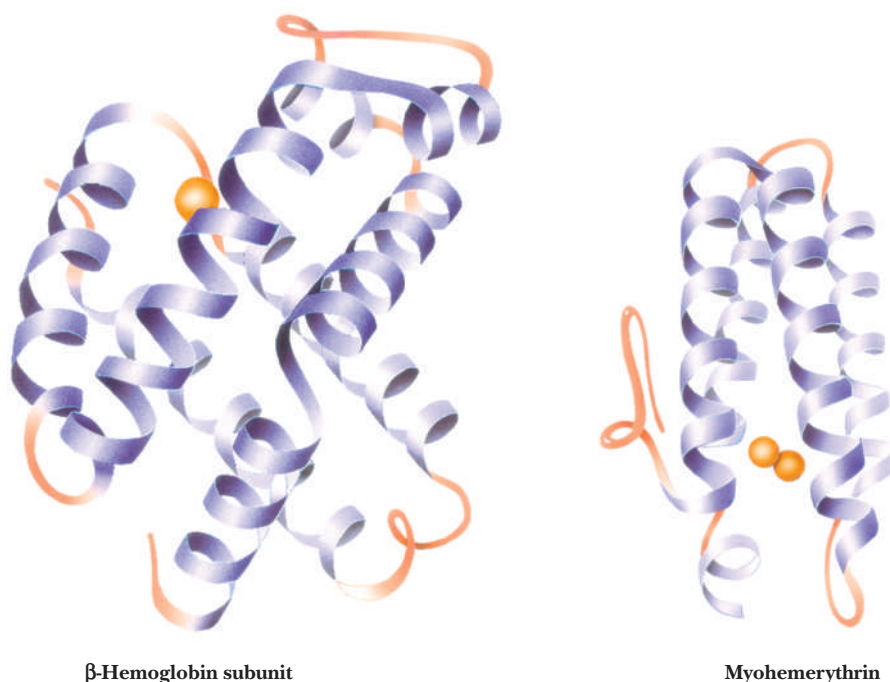
The angstrom unit,  $1 \text{ Å} = 10^{-8} \text{ cm} = 10^{-10} \text{ m}$ , is convenient for interatomic distances in molecules, but it is not a Système International (SI) unit. Nanometers ( $1 \text{ nm} = 10^{-9} \text{ m}$ ) and picometers ( $1 \text{ pm} = 10^{-12} \text{ m}$ ) are the SI units used for interatomic distances. In SI units, the pitch of the  $\alpha$ -helix is





0.54 nm or 540 pm. Figure 4.3 shows the structures of two proteins with a high degree of  $\alpha$ -helical content.

Proteins have varying amounts of  $\alpha$ -helical structures, varying from a few percent to nearly 100%. Several factors can disrupt the  $\alpha$ -helix. The amino acid proline creates a bend in the backbone because of its *cyclic* structure. It cannot fit into the  $\alpha$ -helix because (1) rotation around the bond between the nitrogen and the  $\alpha$ -carbon is severely restricted, and (2) proline's  $\alpha$ -amino group cannot participate in intrachain hydrogen bonding. Other localized factors involving the side chains include strong electrostatic repulsion owing to the proximity of



■ **FIGURE 4.3** Three-dimensional structure of two proteins with substantial amounts of  $\alpha$ -helix in their structures. The helices are represented by the regularly coiled sections of the ribbon diagram. Myohemerythrin is an oxygen-carrying protein in invertebrates.

several charged groups of the same sign, such as groups of positively charged lysine and arginine residues or groups of negatively charged glutamate and aspartate residues. Another possibility is crowding (steric repulsion) caused by the proximity of several bulky side chains. In the  $\alpha$ -helical conformation, all the side chains lie outside the helix; there is not enough room for them in the interior. The  $\alpha$ -carbon is just outside the helix, and crowding can occur if it is bonded to two atoms other than hydrogen, as is the case with valine, isoleucine, and threonine.

### How is the $\beta$ -sheet different from the $\alpha$ -helix?

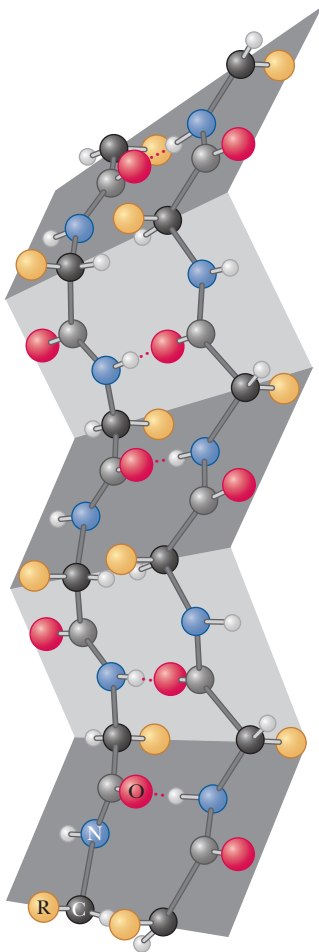
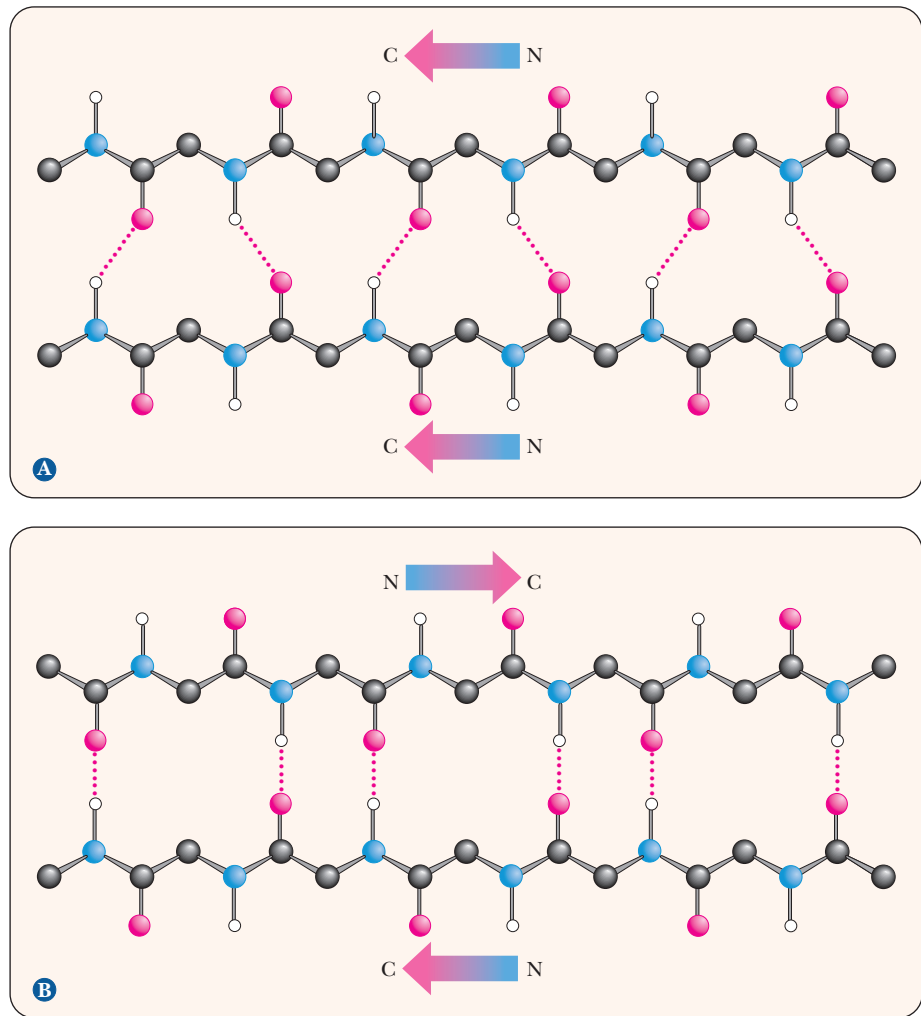
The arrangement of atoms in the  $\beta$ -pleated sheet conformation differs markedly from that in the  $\alpha$ -helix. The peptide backbone in the  $\beta$ -sheet is almost completely extended. Hydrogen bonds can be formed between different parts of a single chain that is doubled back on itself (*intrachain bonds*) or between different chains (*interchain bonds*). If the peptide chains run in the same direction (i.e., if they are all aligned in terms of their N-terminal and C-terminal ends), a parallel pleated sheet is formed. When alternating chains run in opposite directions, an antiparallel pleated sheet is formed (Figure 4.4). The hydrogen bonding between peptide chains in the  $\beta$ -pleated sheet gives rise to a repeated zigzag structure; hence, the name “pleated sheet” (Figure 4.5). Note that the hydrogen bonds are perpendicular to the direction of the protein chain, not parallel to it as in the  $\alpha$ -helix.

### Irregularities in Regular Structures

Other helical structures are found in proteins. These are often found in shorter stretches than with the  $\alpha$ -helix, and they sometimes break up the regular nature of the  $\alpha$ -helix. The most common is the  $3_{10}$  helix, which has three residues per turn and 10 atoms in the ring formed by making the hydrogen bond. Other common helices are designated  $2_7$  and  $4_{16}$ , following the same nomenclature as the  $3_{10}$  helix.

A  **$\beta$ -bulge** is a common nonrepetitive irregularity found in antiparallel  $\beta$ -sheets. It occurs between two normal  $\beta$ -structure hydrogen bonds and involves two residues on one strand and one on the other. Figure 4.6 shows typical  $\beta$ -bulges.

■ **FIGURE 4.4** Hydrogen bonding in  $\beta$ -pleated sheets. Ball-and-stick diagram showing the arrangement of hydrogen bonds in (a) parallel and (b) antiparallel  $\beta$ -pleated sheets.

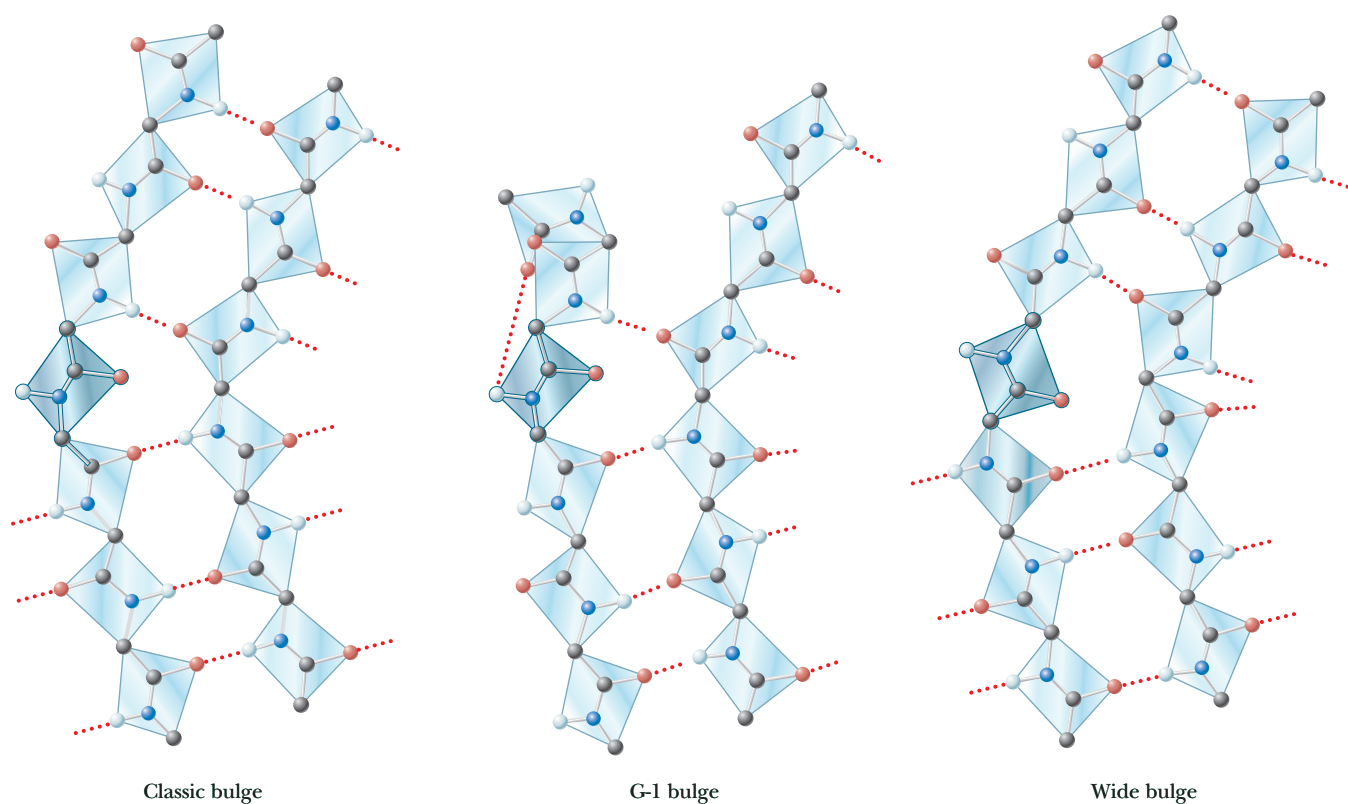


■ **FIGURE 4.5** The three-dimensional form of the antiparallel  $\beta$ -pleated sheet arrangement. The chains do not fold back on each other but are in a fully extended conformation. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

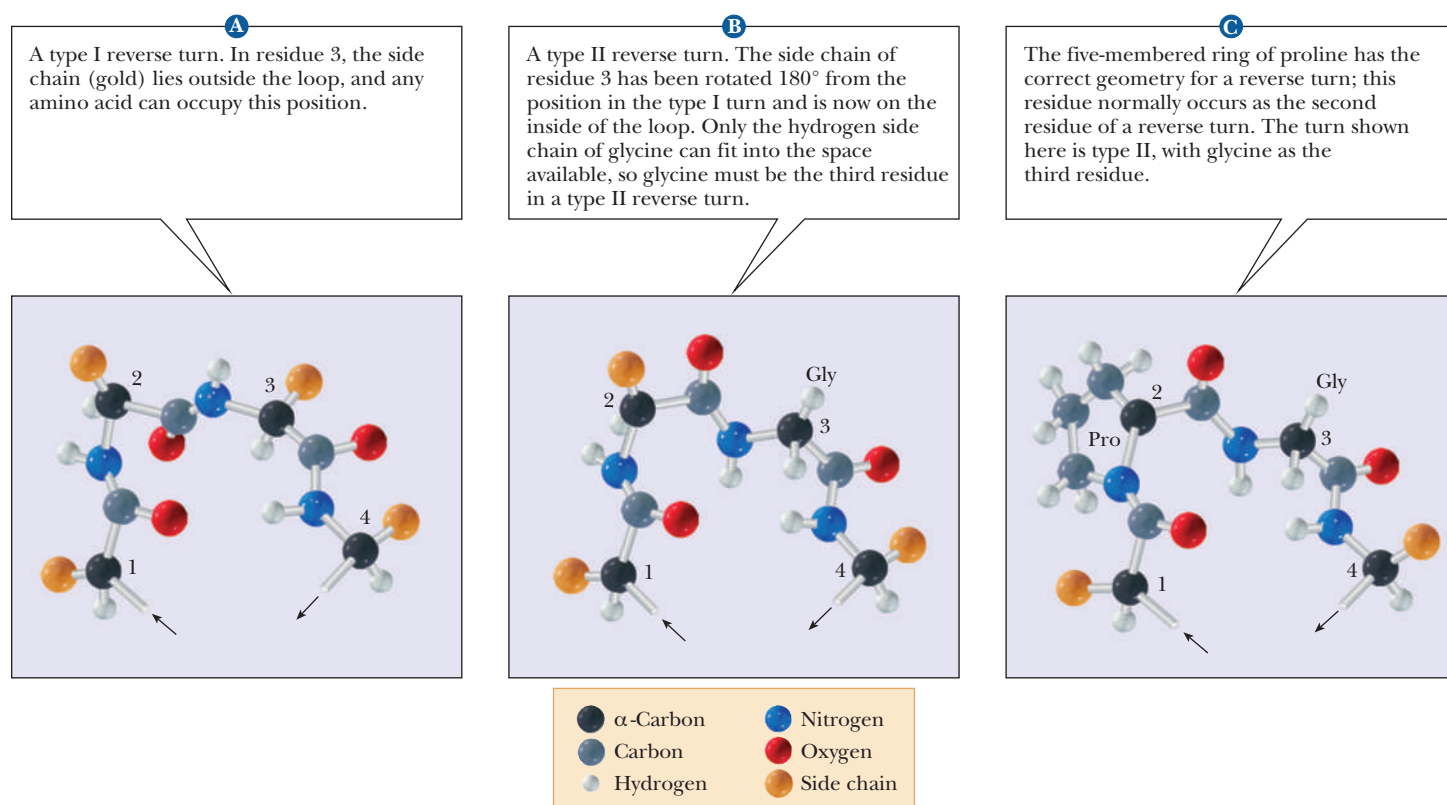
Protein folding requires that the peptide backbones and the secondary structures be able to change directions. Often a reverse turn marks a transition between one secondary structure and another. For steric (spatial) reasons, glycine is frequently encountered in **reverse turns**, at which the polypeptide chain changes direction; the single hydrogen of the side chain prevents crowding (Figures 4.7a and 4.7b). Because the cyclic structure of proline has the correct geometry for a reverse turn, this amino acid is also frequently encountered in such turns (Figure 4.7c).

### Supersecondary Structures and Domains

The  $\alpha$ -helix,  $\beta$ -pleated sheet, and other secondary structures are combined in many ways as the polypeptide chain folds back on itself in a protein. The combination of  $\alpha$ - and  $\beta$ -strands produces various kinds of supersecondary structures in proteins. The most common feature of this sort is the  $\beta\alpha\beta$  unit, in which two parallel strands of  $\beta$ -sheet are connected by a stretch of  $\alpha$ -helix (Figure 4.8a). An  $\alpha\alpha$  unit (helix-turn-helix) consists of two antiparallel  $\alpha$ -helices (Figure 4.8b). In such an arrangement, energetically favorable contacts exist between the side chains in the two stretches of helix. In a  $\beta$ -meander, an antiparallel sheet is formed by a series of tight reverse turns connecting stretches of the polypeptide chain (Figure 4.8c). Another kind of antiparallel sheet is formed when the polypeptide chain doubles back on itself in a pattern known as the *Greek key*, named for a decorative design found on pottery from the classical



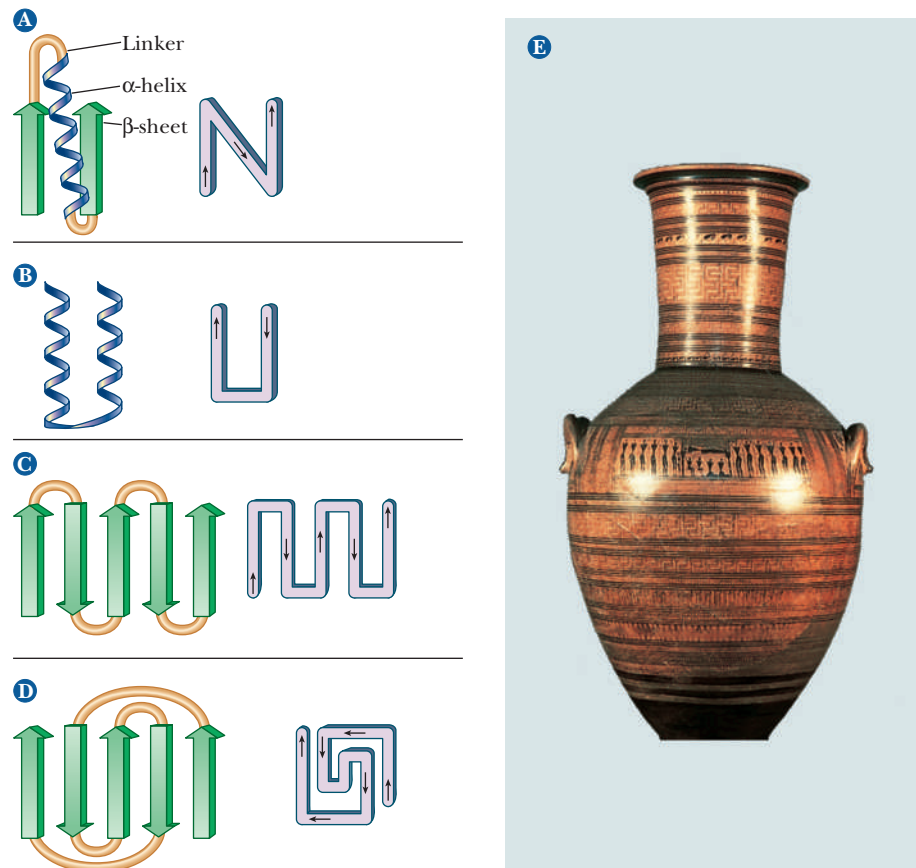
■ **FIGURE 4.6**  $\beta$ -bulges. Ball-and-stick model of three different  $\beta$ -bulge structures. Hydrogen bonds are shown as red dots.



■ **FIGURE 4.7** Structures of reverse turns. Arrows indicate the directions of the polypeptide chains.



■ **FIGURE 4.8** Schematic diagrams of supersecondary structures. Arrows indicate the directions of the polypeptide chains. (a) A  $\beta\alpha\beta$  unit, (b) an  $\alpha\alpha$  unit, (c) a  $\beta$ -meander, and (d) the Greek key. (e) The Greek key motif in protein structure resembles the geometric patterns on this ancient Greek vase, giving rise to the name.



period (Figure 4.8e). A **motif** is a repetitive supersecondary structure. Some of the common smaller motifs are shown in Figure 4.9. These smaller motifs can often be repeated and organized into larger motifs. Protein sequences that allow for a  $\beta$ -meander or Greek key can often be found arranged into a  $\beta$ -barrel in the tertiary structure of the protein (Figure 4.10). Motifs are important and tell us much about the folding of proteins. However, these motifs do not allow us to predict anything about the biological function of the protein because they are found in proteins and enzymes with very dissimilar functions.

Many proteins that have the same type of function have similar protein sequences; consequently, domains with similar conformations are associated with the particular function. Many types of domains have been identified, including three different types of domains by which proteins bind to DNA. In addition, short polypeptide sequences within a protein direct the posttranslational modification and subcellular localization. For example, several sequences play a role in the formation of glycoproteins (ones that contain sugars in addition to the polypeptide chain). Other specific sequences indicate that a protein is to be bound to a membrane or secreted from the cell. Still other specific sequences mark a protein for phosphorylation by a specific enzyme.

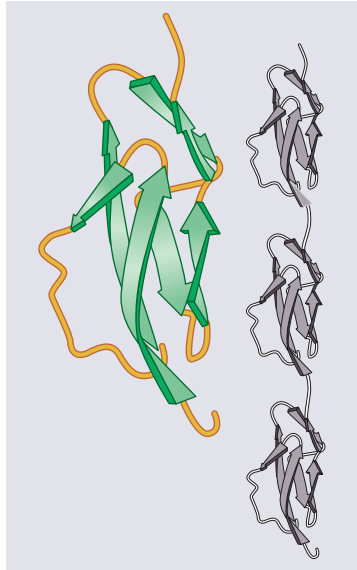
### The Collagen Triple Helix

Collagen, a component of bone and connective tissue, is the most abundant protein in vertebrates. It is organized in water-insoluble fibers of great strength. A collagen fiber consists of three polypeptide chains wrapped around each other in a ropelike twist, or triple helix. Each of the three chains has, within limits, a repeating sequence of three amino acid residues, X—Pro—Gly or X—Hyp—Gly, where Hyp stands for hydroxyproline, and any amino acid can occupy the first position, designated by X.

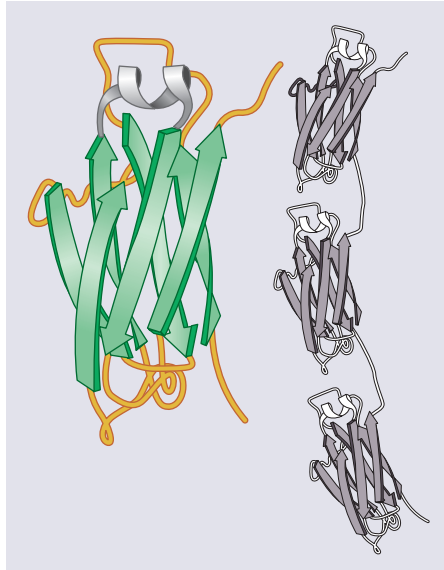
Proline and hydroxyproline can constitute up to 30% of the residues in collagen. Hydroxyproline is formed from proline by a specific hydroxylating



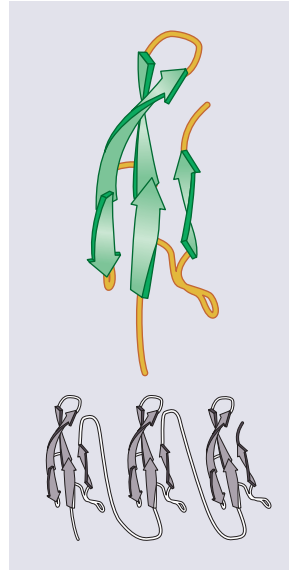
**1** The complement-control protein module.



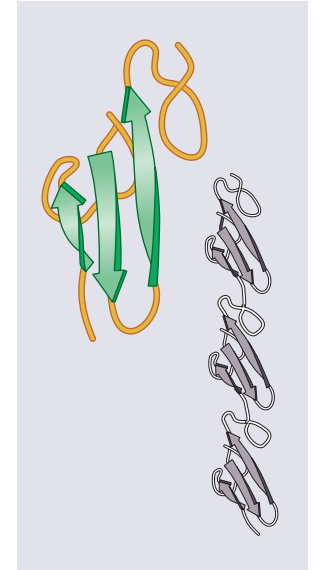
**2** The immunoglobulin module.



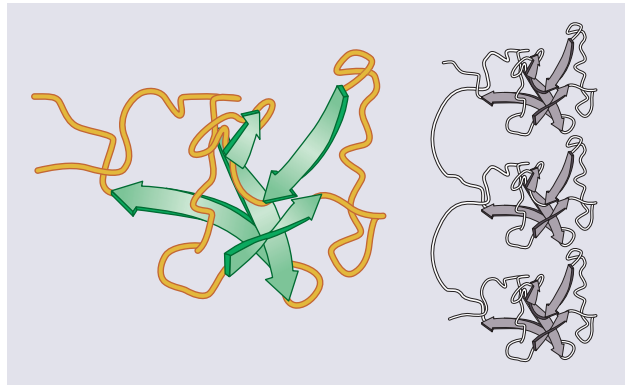
**3** The fibronectin type I module.



**4** The growth-factor module.

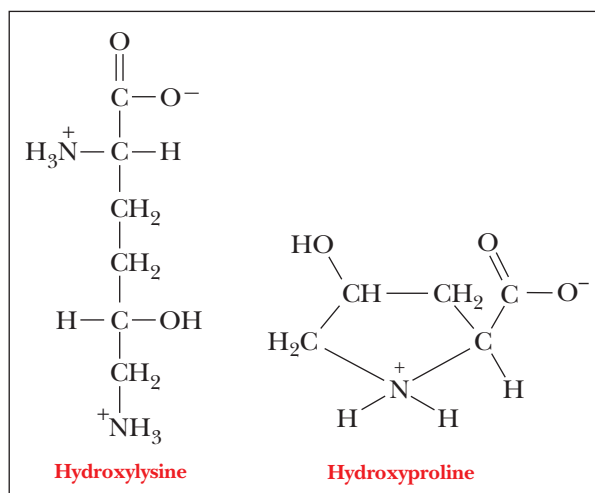


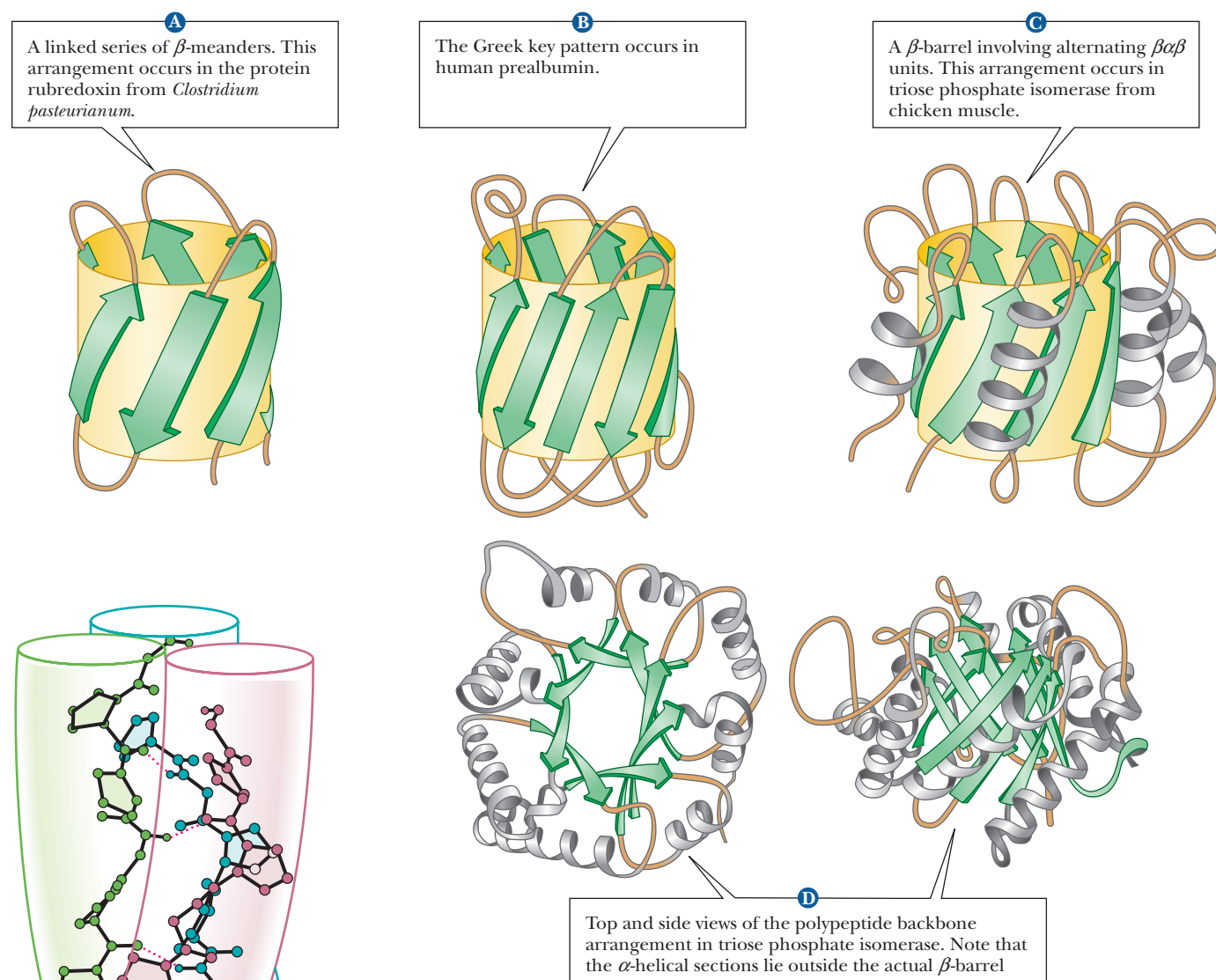
**5** The kringle module.



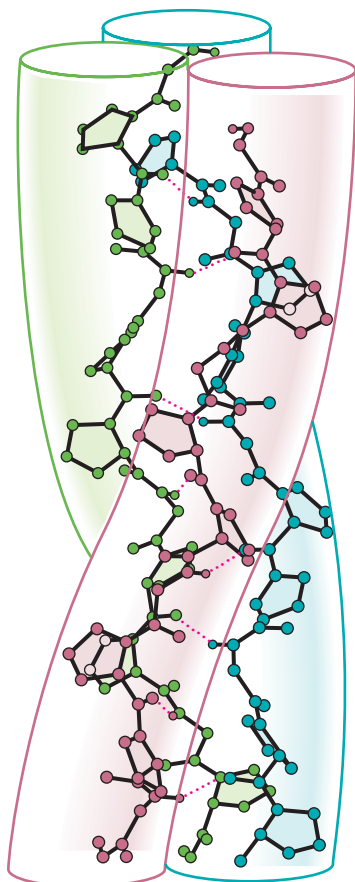
**FIGURE 4.9 Motifs and modules.** Motifs are repeated supersecondary structures, sometimes called modules. All of these have a particular secondary structure that is repeated in the protein. (Reprinted from "Protein Modules," Trends in Biochemical Sciences, Vol. 16, pp. 13–17, Copyright © 1991, with permission from Elsevier.)

enzyme after the amino acids are linked together. Hydroxylysine also occurs in collagen. In the amino acid sequence of collagen, every third position must be occupied by glycine. The triple helix is arranged so that every third residue on each chain is inside the helix. Only glycine is small enough to fit into the space available (Figure 4.11).





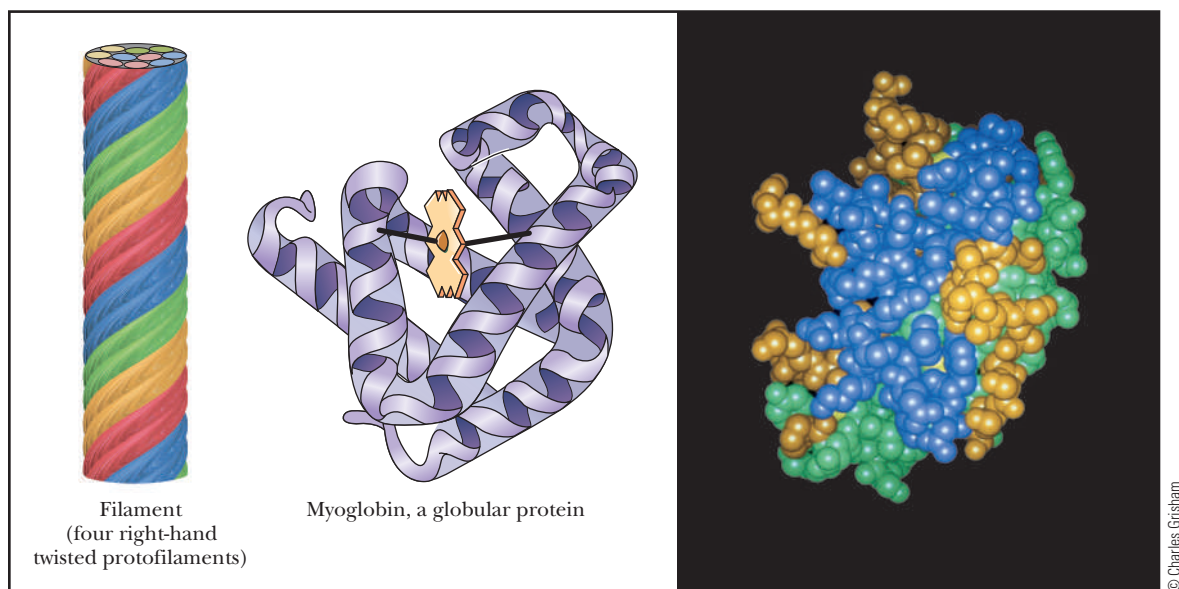
■ **FIGURE 4.10** Some  $\beta$ -barrel arrangements.



■ **FIGURE 4.11** A triple helix. Poly (Gly—Pro—Pro) is a collagen-like right-handed triple helix composed of three left-handed helical chains. (Adapted from M. H. Miller and H. A. Scheraga, 1976, *Calculation of the structures of collagen models. Role of interchain interactions in determining the triple-helical coiled-coil conformations. I. Poly(glycyl-prolyl-prolyl)*. Journal of Polymer Science Symposium 54:171–200. © 1976 John Wiley & Sons, Inc. Reprinted by permission.)

The three individual collagen chains are themselves helices that differ from the  $\alpha$ -helix. They are twisted around each other in a superhelical arrangement to form a stiff rod. This triple helical molecule is called *tropocollagen*; it is 300 nm (3000 Å) long and 1.5 nm (15 Å) in diameter. The three strands are held together by hydrogen bonds involving the hydroxyproline and hydroxylysine residues. The molecular weight of the triple-stranded array is about 300,000; each strand contains about 800 amino acid residues. Collagen is both intramolecularly and intermolecularly linked by covalent bonds formed by reactions of lysine and histidine residues. The amount of cross-linking in a tissue increases with age. That is why meat from older animals is tougher than meat from younger animals.

Collagen in which the proline is not hydroxylated to hydroxyproline to the usual extent is less stable than normal collagen. Symptoms of scurvy, such as bleeding gums and skin discoloration, are the results of fragile collagen. The enzyme that hydroxylates proline and thus maintains the normal state of collagen requires ascorbic acid (vitamin C) to remain active. Scurvy is ultimately caused by a dietary deficiency of vitamin C. See the Biochemical Connections box in Chapter 16.



**A** Schematic diagrams of a portion of a fibrous protein and of a globular protein.

**B** Computer-generated model of a globular protein. Alpha-helices are shown in blue, beta-sheets are green, and random coil is gold.

**FIGURE 4.12** A comparison of the shapes of fibrous and globular proteins.

## Two Types of Protein Conformations: Fibrous and Globular

It is difficult to draw a clear separation between secondary and tertiary structures. The nature of the side chains in a protein (part of the tertiary structure) can influence the folding of the backbone (the secondary structure). Comparing collagen with silk and wool fibers can be illuminating. Silk fibers consist largely of the protein fibroin, which, like collagen, has a fibrous structure, but which, unlike collagen, consists largely of  $\beta$ -sheets. Fibers of wool consist largely of the protein keratin, which is largely  $\alpha$ -helical. The amino acids of which collagen, fibroin, and keratin are composed determine which conformation they will adopt, but all are **fibrous proteins** (Figure 4.12a).

In other proteins, the backbone folds back on itself to produce a more or less spherical shape. These are called **globular proteins** (Figure 4.12b), and we shall see many examples of them. Their helical and pleated-sheet sections can be arranged so as to bring the ends of the sequence close to each other in three dimensions. Globular proteins, unlike fibrous proteins, are water-soluble and have compact structures; their tertiary and quaternary structures can be quite complex.

## 4.4 Tertiary Structure of Proteins

The tertiary structure of a protein is the three-dimensional arrangement of all the atoms in the molecule. The conformations of the side chains and the positions of any prosthetic groups are parts of the tertiary structure, as is the arrangement of helical and pleated-sheet sections with respect to one another. In a fibrous protein, the overall shape of which is a long rod, the secondary structure also provides much of the information about the tertiary structure. The helical backbone of the protein does not fold back on itself, and the only important aspect of the tertiary structure that is not specified by the secondary structure is the arrangement of the atoms of the side chains.

For a globular protein, considerably more information is needed. It is necessary to determine the way in which the helical and pleated-sheet sections fold back on each other, in addition to the positions of the side-chain atoms and any prosthetic groups. The interactions between the side chains play an important role in the folding of proteins. The folding pattern frequently brings residues that are separated in the amino acid sequence into proximity in the tertiary structure of the native protein.

### Forces Involved in Tertiary Structures

Many types of forces and interactions play a role in holding a protein together in its correct, native conformation. Some of these forces are covalent, but many are not. The primary structure of a protein—the order of amino acids in the polypeptide chain—depends on the formation of peptide bonds, which are covalent. Higher-order levels of structure, such as the conformation of the backbone (secondary structure) and the positions of all the atoms in the protein (tertiary structure), depend on noncovalent interactions. If the protein consists of several subunits, the interaction of the subunits (quaternary structure, Section 4.5) also depends on noncovalent interactions. Noncovalent stabilizing forces contribute to the most stable structure for a given protein, the one with the lowest energy.

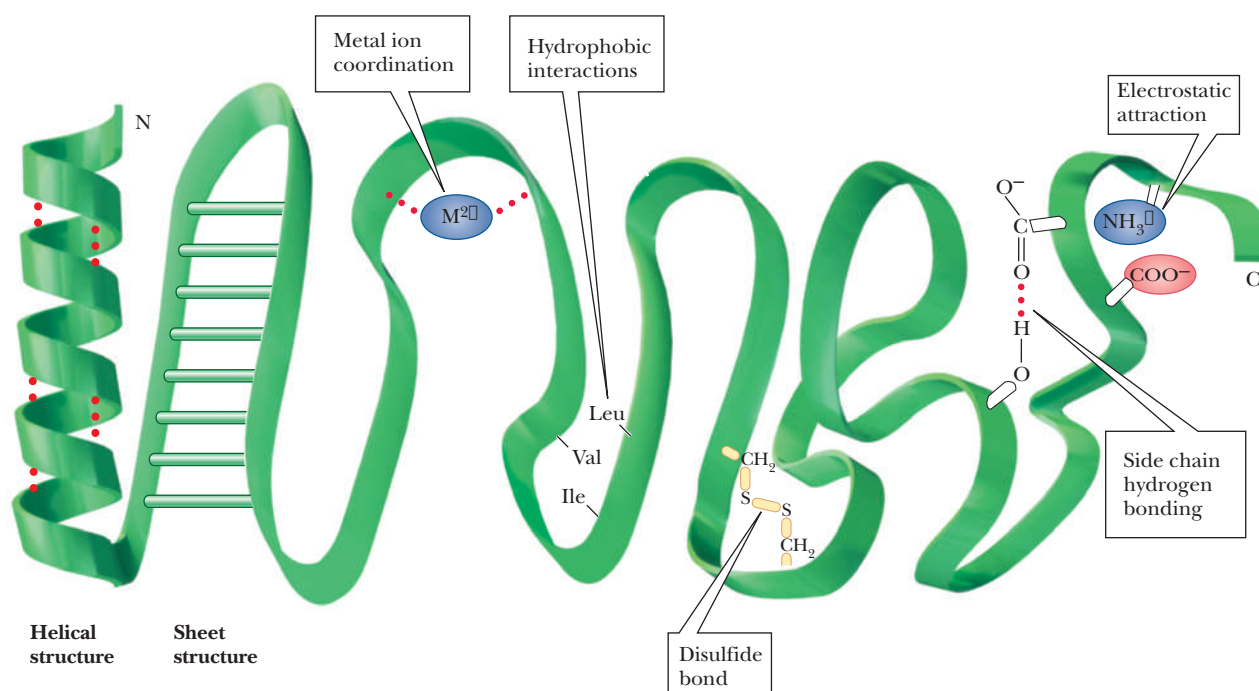
Several types of hydrogen bonding occur in proteins. *Backbone* hydrogen bonding is a major determinant of secondary structure; hydrogen bonds *between the side chains of amino acids* are also possible in proteins. Nonpolar residues tend to cluster together in the interior of protein molecules as a result of *hydrophobic* interactions. *Electrostatic* attraction between oppositely charged groups, which frequently occurs on the surface of the molecule, results in such groups being close to one another. Several side chains can be *complexed* to a single metal ion. (Metal ions also occur in some prosthetic groups.)

In addition to these noncovalent interactions, *disulfide bonds* form covalent links between the side chains of cysteines. When such bonds form, they restrict the folding patterns available to polypeptide chains. There are specialized laboratory methods for determining the number and positions of disulfide links in a given protein. Information about the locations of disulfide links can then be combined with knowledge of the primary structure to give the *complete covalent structure* of the protein. Note the subtle difference here: The primary structure is the order of amino acids, whereas the complete covalent structure also specifies the positions of the disulfide bonds (Figure 4.13).

Not every protein necessarily exhibits all possible structural features of the kinds just described. For instance, there are no disulfide bridges in myoglobin and hemoglobin, which are oxygen-storage and transport proteins and classic examples of protein structure, but they both contain Fe(II) ions as part of a prosthetic group. In contrast, the enzymes trypsin and chymotrypsin do not contain complexed metal ions, but they do have disulfide bridges. Hydrogen bonds, electrostatic interactions, and hydrophobic interactions occur in most proteins.

The three-dimensional conformation of a protein is the result of the interplay of all the stabilizing forces. It is known, for example, that proline does not fit into an  $\alpha$ -helix and that its presence can cause a polypeptide chain to turn a corner, ending an  $\alpha$ -helical segment. The presence of proline is not, however, a *requirement* for a turn in a polypeptide chain. Other residues are routinely encountered at bends in polypeptide chains. The segments of proteins at bends in the polypeptide chain and in other portions of the protein that are not involved in helical or pleated-sheet structures are frequently referred to as “random” or “random coil.” In reality, the forces that stabilize each protein are responsible for its conformation.





■ **FIGURE 4.13** Forces that stabilize the tertiary structure of proteins. Note that the helical structure and sheet structure are two kinds of backbone hydrogen bonding. Although backbone hydrogen bonding is part of secondary structure, the conformation of the backbone constrains the possible arrangement of the side chains.

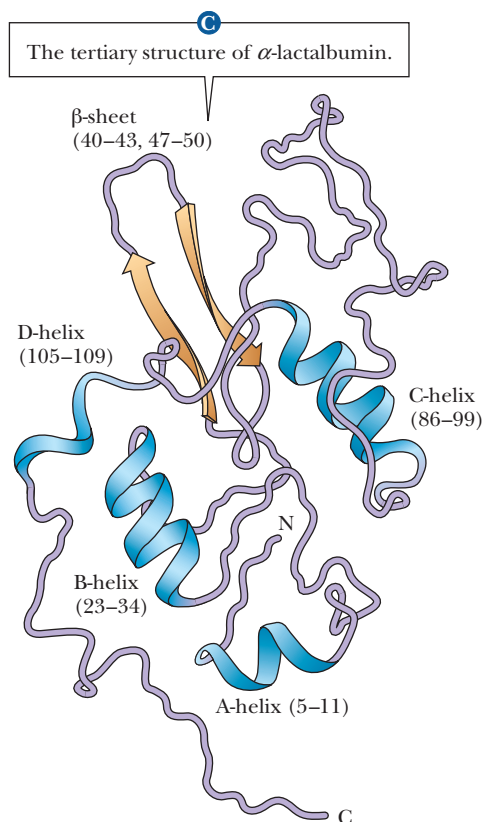
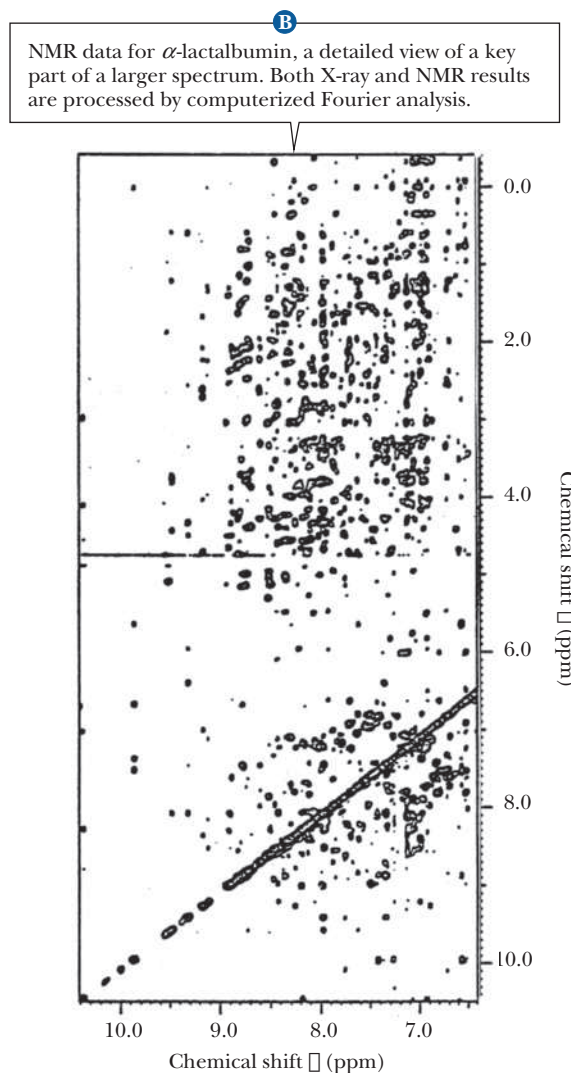
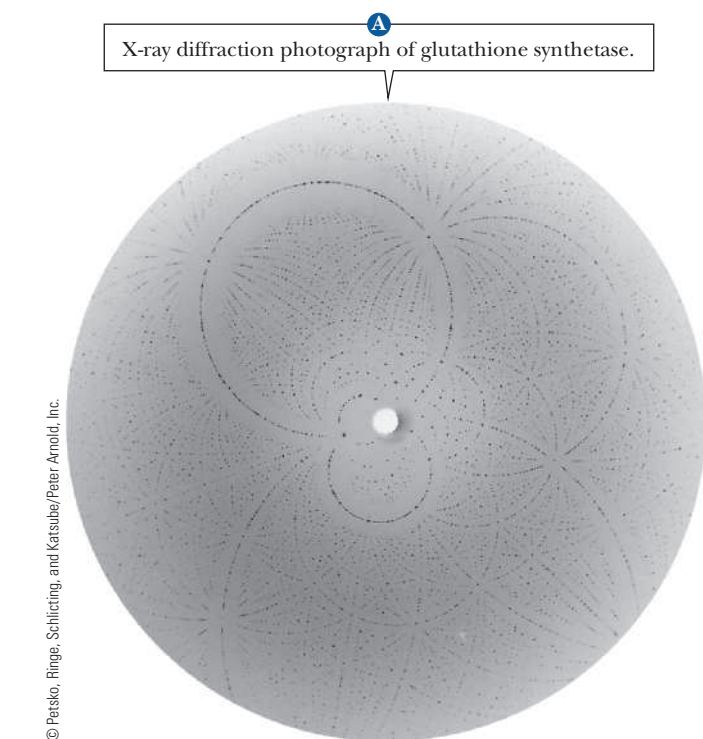
### How can the three-dimensional structure of a protein be determined?

The experimental technique used to determine the tertiary structure of a protein is **X-ray crystallography**. Perfect crystals of some proteins can be grown under carefully controlled conditions. In such a crystal, all the individual protein molecules have the same three-dimensional conformation and the same orientation. Crystals of this quality can be formed only from proteins of very high purity, and it is not possible to obtain a structure if the protein cannot be crystallized.

When a suitably pure crystal is exposed to a beam of X rays, a *diffraction pattern* is produced on a photographic plate (Figure 4.14a) or a radiation counter. The pattern is produced when the electrons in each atom in the molecule scatter the X rays. The number of electrons in the atom determines the intensity of its scattering of X rays; heavier atoms scatter more effectively than lighter atoms. The scattered X rays from the individual atoms can reinforce each other or cancel each other (set up constructive or destructive interference), giving rise to the characteristic pattern for each type of molecule. A series of diffraction patterns taken from several angles contains the information needed to determine the tertiary structure. The information is extracted from the diffraction patterns through a mathematical analysis known as a *Fourier series*. Many thousands of such calculations are required to determine the structure of a protein, and even though they are performed by computer, the process is a fairly long one. Improving the calculation procedure is a subject of active research.

Another technique that supplements the results of X-ray diffraction has come into wide use in recent years. It is a form of **nuclear magnetic resonance (NMR) spectroscopy**. In this particular application of NMR, called *2-D* (two-dimensional) *NMR*, large collections of data points are subjected to computer analysis (Figure 4.14b). Like X-ray diffraction, this method uses a Fourier series



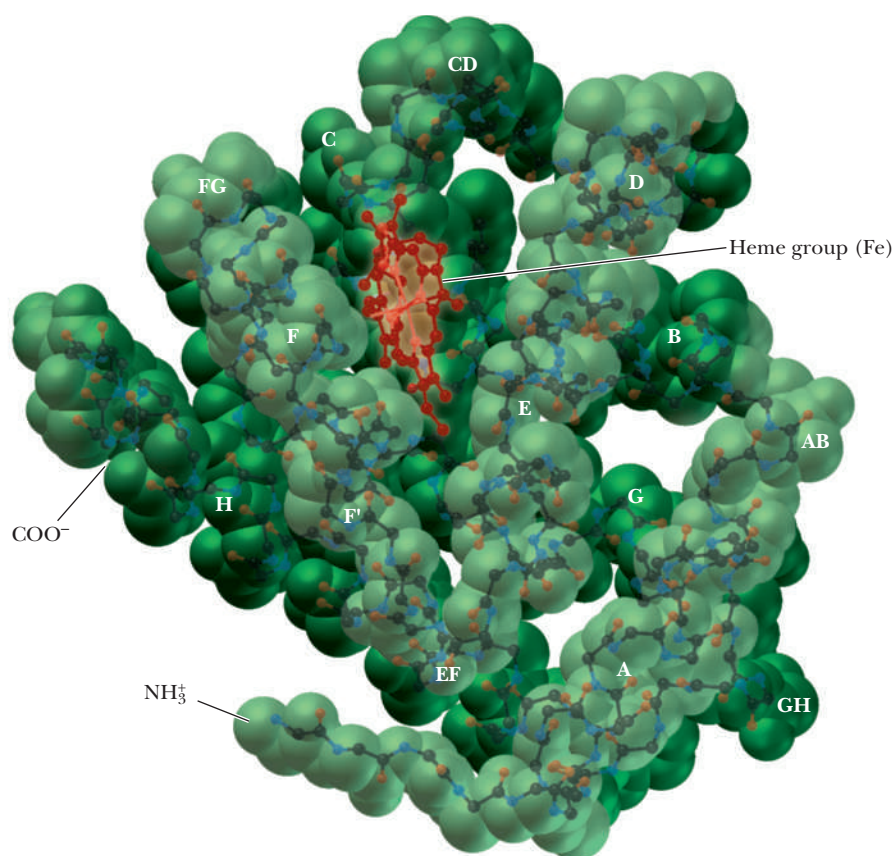


**FIGURE 4.14** Large numbers of data points are needed to determine the tertiary structure of a protein. (See Figure 4.15 for the structure of myoglobin as determined by X-ray crystallography.) (b, courtesy of Professor C. M. Dobson, University of Oxford.)

to analyze results. It is similar to X-ray diffraction in other ways: It is a long process, and it requires considerable amounts of computing power and milligram quantities of protein. One way in which 2-D NMR differs from X-ray diffraction is that it uses protein samples in aqueous solution rather than crystals. This environment is closer to that of proteins in cells, and thus it is one of the main advantages of the method. The NMR method most widely used in the determination of protein structure ultimately depends on the distances between hydrogen atoms, giving results independent of those obtained by X-ray crystallography. The NMR method is undergoing constant improvement and is being applied to larger proteins as these improvements progress.

### Myoglobin: An Example of Protein Structure

In many ways, myoglobin is the classic example of a globular protein. We shall use it here as a case study in tertiary structure. (We shall see the tertiary structures of many other proteins in context when we discuss their roles in biochemistry.) Myoglobin was the first protein for which the complete tertiary structure (Figure 4.15) was determined by X-ray crystallography. The complete myoglobin molecule consists of a single polypeptide chain of 153 amino acid

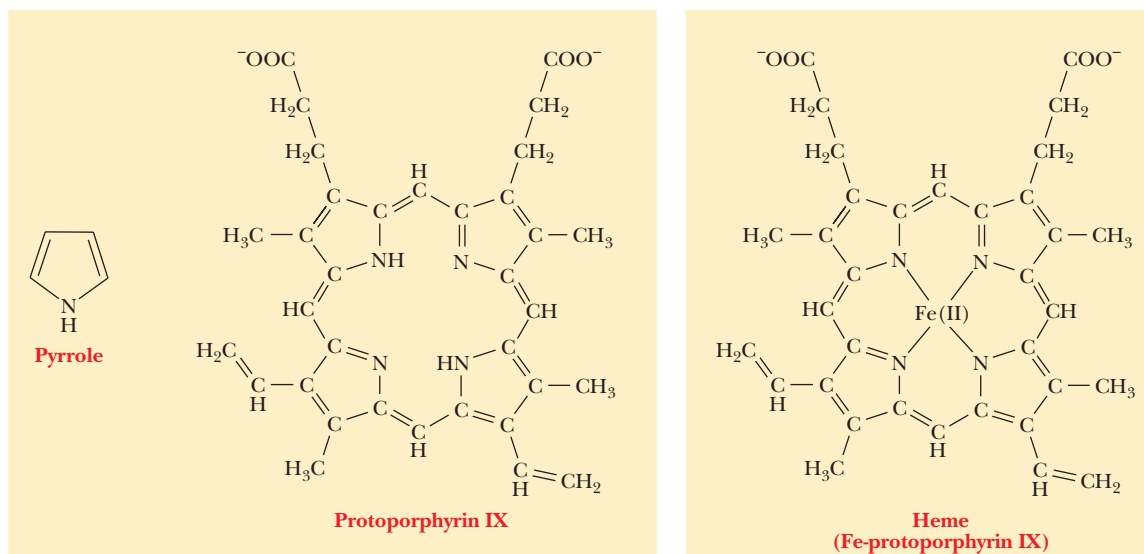


■ **FIGURE 4.15** The structure of myoglobin. The peptide backbone and the heme group are shown overlain on the space-filling model. The helical segments are designated by the letters A through H. The terms  $\text{NH}_3^+$  and  $\text{COO}^-$  indicate the N-terminal and C-terminal ends, respectively.

residues and includes a prosthetic group, the **heme** group, which also occurs in hemoglobin. The myoglobin molecule (including the heme group) has a compact structure, with the interior atoms very close to each other. This structure provides examples of many of the forces responsible for the three-dimensional shapes of proteins.

Myoglobin has eight  $\alpha$ -helical regions and no  $\beta$ -pleated sheet regions. Approximately 75% of the residues in myoglobin are found in these helical regions, which are designated by the letters A through H. Hydrogen bonding in the polypeptide backbone stabilizes the  $\alpha$ -helical regions; amino acid side chains are also involved in hydrogen bonds. The polar residues are on the exterior of the molecule. The interior of the protein contains almost exclusively nonpolar amino acid residues. Two polar histidine residues are found in the interior; they are involved in interactions with the heme group and bound oxygen, and thus play an important role in the function of the molecule. The planar heme group fits into a hydrophobic pocket in the protein portion of the molecule and is held in position by hydrophobic attractions between heme's porphyrin ring and the nonpolar side chains of the protein. The presence of the heme group drastically affects the conformation of the polypeptide: The apoprotein (the polypeptide chain alone, without the prosthetic heme group) is not as tightly folded as the complete molecule.

The heme group consists of a metal ion,  $\text{Fe(II)}$ , and an organic part, protoporphyrin IX (Figure 4.16). (The notation  $\text{Fe(II)}$  is preferred to  $\text{Fe}^{2+}$  when metal ions occur in complexes.) The porphyrin part consists of four five-membered rings based on the pyrrole structure; these four rings are linked by bridging methine ( $-\text{CH}=\text{}$ ) groups to form a square planar structure. The  $\text{Fe(II)}$  ion has six coordination sites, and it forms six metal-ion complexation bonds. Four of the six sites are occupied by the nitrogen atoms of the four



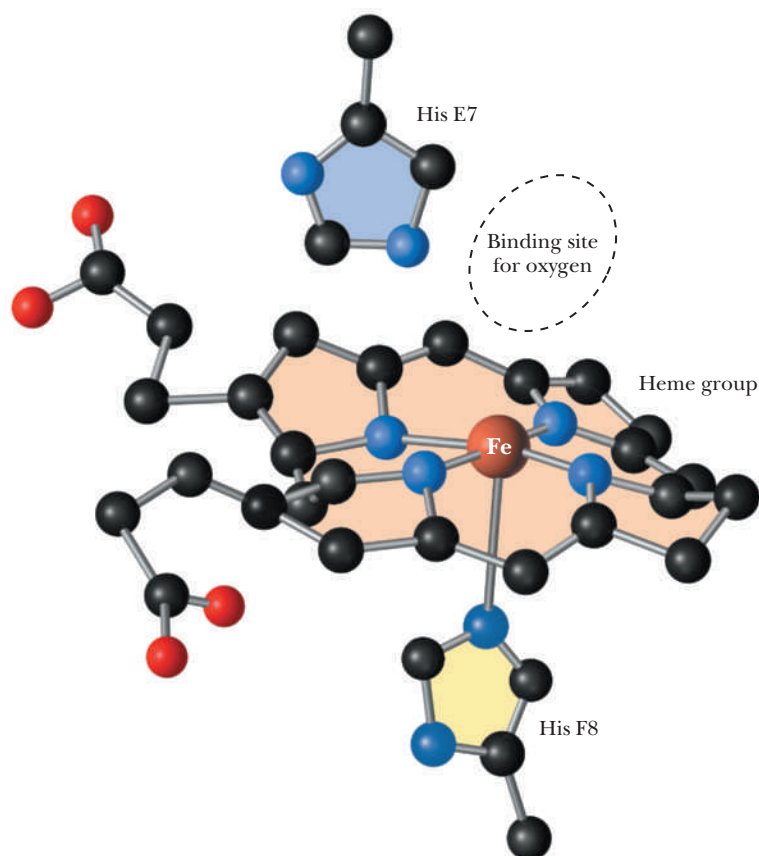
■ **FIGURE 4.16** The structure of the heme group. Four pyrrole rings are linked by bridging groups to form a planar porphyrin ring. Several isomeric porphyrin rings are possible, depending on the nature and arrangement of the side chains. The porphyrin isomer found in heme is protoporphyrin IX. Addition of iron to protoporphyrin IX produces the heme group.

pyrrole-type rings of the porphyrin to give the complete heme group. The presence of the heme group is required for myoglobin to bind oxygen.

The fifth coordination site of the Fe(II) ion is occupied by one of the nitrogen atoms of the imidazole side chain of histidine residue F8 (the eighth residue in helical segment F). This histidine residue is one of the two in the interior of the molecule. The oxygen is bound at the sixth coordination site of the iron. The fifth and sixth coordination sites lie perpendicular to, and on opposite sides of, the plane of the porphyrin ring. The other histidine residue in the interior of the molecule, residue E7 (the seventh residue in helical segment E), lies on the same side of the heme group as the bound oxygen (Figure 4.17). This second histidine is not bound to the iron, or to any part of the heme group, but it acts as a gate that opens and closes as oxygen enters the hydrophobic pocket to bind to the heme. The E7 histidine sterically inhibits oxygen from binding perpendicularly to the heme plane, with biologically important ramifications.

### Why does oxygen have imperfect binding to the heme group?

At first, it would seem counterintuitive that oxygen would bind imperfectly to the heme group. After all, the job of both myoglobin and hemoglobin is to bind to oxygen. Wouldn't it make sense that oxygen should bind strongly? The answer lies in the fact that more than one molecule can bind to heme. Besides oxygen, carbon monoxide also binds to heme. The affinity of free heme for carbon monoxide (CO) is 25,000 times greater than its affinity for oxygen. When carbon monoxide is forced to bind at an angle in myoglobin because of the steric block by His E7, its advantage over oxygen drops by two orders of magnitude (Figure 4.18). This guards against the possibility that traces of CO produced during metabolism would occupy all the oxygen-binding sites on the hemes. Nevertheless, CO is a potent poison in larger quantities because of its effect both on oxygen binding to hemoglobin and on the final step of the electron transport chain (Section 20.5). It is also important to remember that although our metabolism requires that hemoglobin and myoglobin bind oxygen, it would be equally disastrous if the heme never let the oxygen go. Thus having binding be too perfect would defeat the purpose of having the oxygen-carrying proteins.

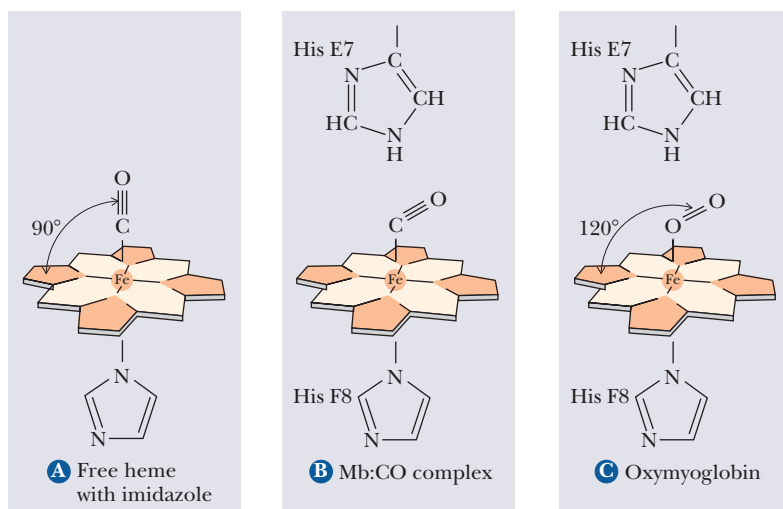


■ **FIGURE 4.17** The oxygen-binding site of myoglobin. The porphyrin ring occupies four of the six coordination sites of the Fe(II). Histidine F8 (His F8) occupies the fifth coordination site of the iron (see text). Oxygen is bound at the sixth coordination site of the iron, and histidine E7 lies close to the oxygen. (*Leonard Lessin/Waldo Feng/Mt. Sinai CORE.*)

In the absence of the protein, the iron of the heme group can be oxidized to Fe(III); the oxidized heme will not bind oxygen. Thus, the combination of both heme and protein is needed to bind  $O_2$  for oxygen storage.

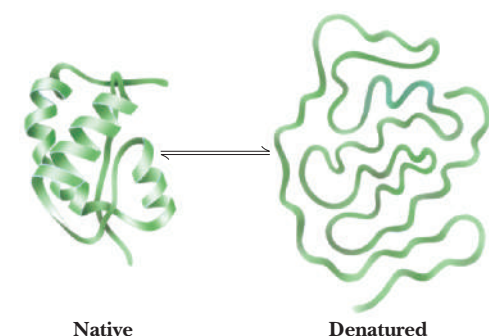
### Denaturation and Refolding

The noncovalent interactions that maintain the three-dimensional structure of a protein are weak, and it is not surprising that they can be disrupted easily. The unfolding of a protein (i.e., disruption of the tertiary structure) is called **denaturation**. Reduction of disulfide bonds (Section 3.5) leads to even more extensive unraveling of the tertiary structure. Denaturation and reduction of



■ **FIGURE 4.18** Oxygen and carbon monoxide binding to the heme group of myoglobin. The presence of the E7 histidine forces a  $120^\circ$  angle to the oxygen or CO.





■ **FIGURE 4.19** Denaturation of a protein. The native conformation can be recovered when denaturing conditions are removed.

disulfide bonds are frequently combined when complete disruption of the tertiary structure of proteins is desired. Under proper experimental conditions, the disrupted structure can then be completely recovered. This process of denaturation and refolding is a dramatic demonstration of the relationship between the primary structure of the protein and the forces that determine the tertiary structure. For many proteins, various other factors are needed for complete refolding, but the important point is that the primary structure determines the tertiary structure.

Proteins can be denatured in several ways. One is *heat*. An increase in temperature favors vibrations within the molecule, and the energy of these vibrations can become great enough to disrupt the tertiary structure. At either high or low *extremes of pH*, at least some of the charges on the protein are missing, and so the electrostatic interactions that would normally stabilize the native, active form of the protein are drastically reduced. This leads to denaturation.

The binding of *detergents*, such as sodium dodecyl sulfate (SDS), also denatures proteins. Detergents tend to disrupt hydrophobic interactions. If a detergent is charged, it can also disrupt electrostatic interactions within the protein. Other reagents, such as *urea* and *guanidine hydrochloride*, form hydrogen bonds with the protein that are stronger than those within the protein itself. These two reagents can also disrupt hydrophobic interactions in much the same way as detergents (Figure 4.19).

$\beta$ -Mercaptoethanol ( $\text{HS}-\text{CH}_2-\text{CH}_2-\text{OH}$ ) is frequently used to reduce disulfide bridges to two sulfhydryl groups. Urea is usually added to the reaction mixture to facilitate unfolding of the protein and to increase the accessibility of the disulfides to the reducing agent. If experimental conditions are properly chosen, the native conformation of the protein can be recovered when both mercaptoethanol and urea are removed (Figure 4.20). Experiments of this type provide some of the strongest evidence that the amino acid sequence of the protein contains all the information required to produce the complete three-dimensional structure. Protein researchers are pursuing with some interest the conditions under which a protein can be denatured—including reduction of disulfides—and its native conformation later recovered.

## 4.5 Quaternary Structure of Proteins

Quaternary structure is the final level of protein structure and pertains to proteins that consist of more than one polypeptide chain. Each chain is called a *subunit*. The number of chains can range from two to more than a dozen, and the chains may be identical or different. Commonly occurring examples are **dimers**, **trimers**, and **tetramers**, consisting of two, three, and four polypeptide chains, respectively. (The generic term for such a molecule, made up of a small number of subunits, is **oligomer**.) The chains interact with one another noncovalently via electrostatic attractions, hydrogen bonds, and hydrophobic interactions.

As a result of these noncovalent interactions, subtle changes in structure at one site on a protein molecule may cause drastic changes in properties at a distant site. Proteins that exhibit this property are called **allosteric**. Not all multisubunit proteins exhibit allosteric effects, but many do.

A classic illustration of the quaternary structure and its effect on protein properties is a comparison of hemoglobin, an allosteric protein, with myoglobin, which consists of a single polypeptide chain.

### Hemoglobin

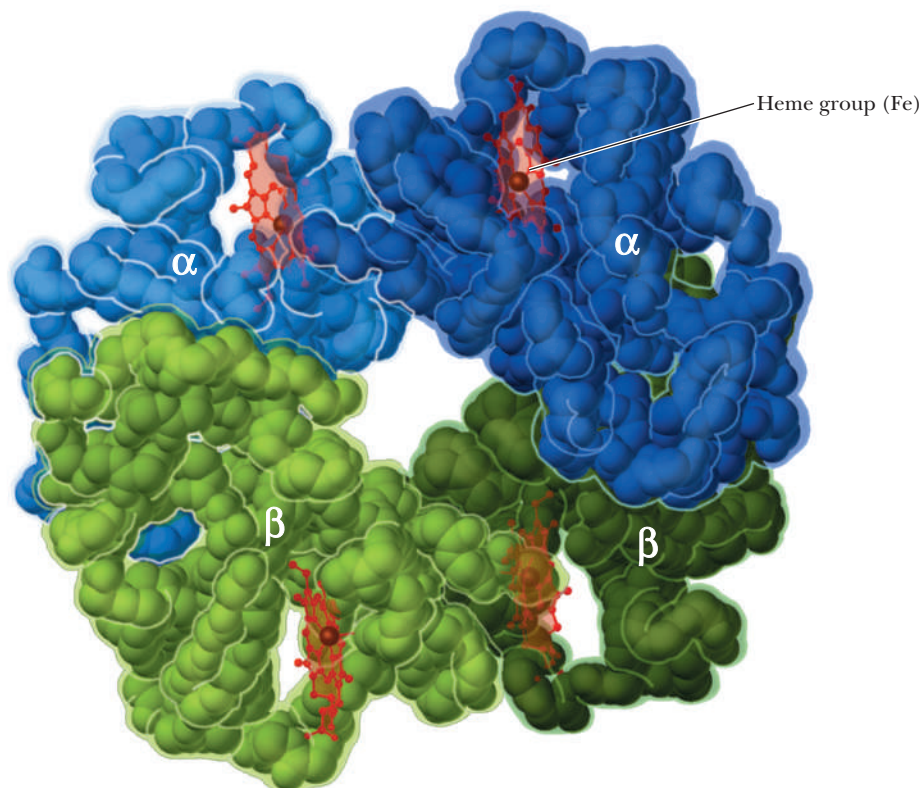
Hemoglobin is a tetramer, consisting of four polypeptide chains, two  $\alpha$ -chains, and two  $\beta$ -chains (Figure 4.21). (In oligomeric proteins, the types of polypeptide



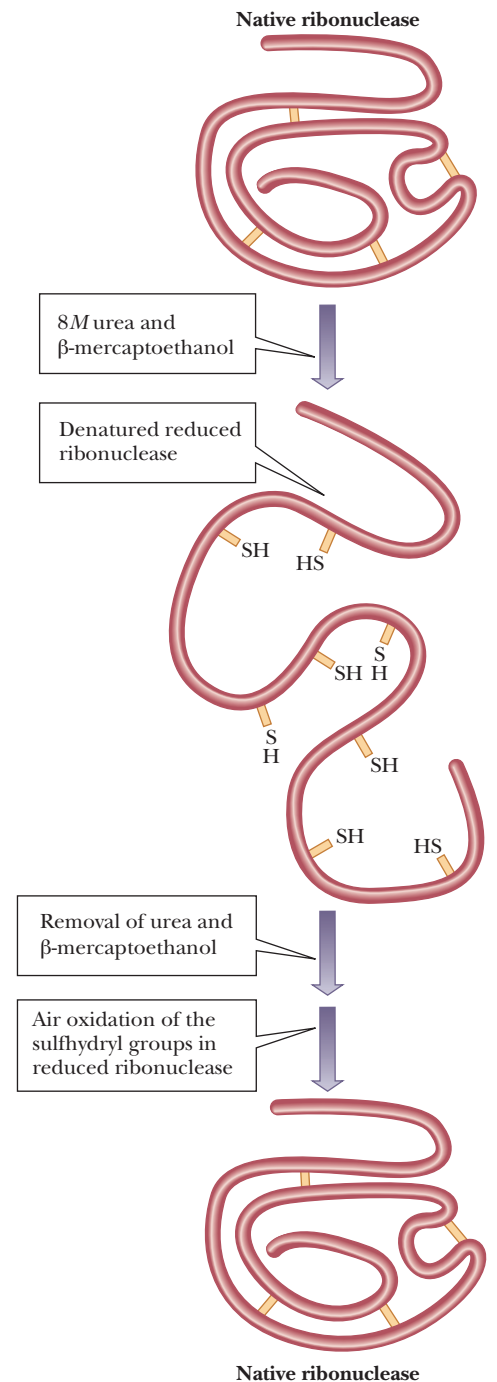
chains are designated with Greek letters. In this case, the terms  $\alpha$  and  $\beta$  have nothing to do with the  $\alpha$ -helix and the  $\beta$ -pleated sheet; rather they just refer to two different polypeptide chain subunits.) The two  $\alpha$ -chains of hemoglobin are identical, as are the two  $\beta$ -chains. The overall structure of hemoglobin is  $\alpha_2\beta_2$  in Greek-letter notation. Both the  $\alpha$ - and  $\beta$ -chains of hemoglobin are very similar to the myoglobin chain. The  $\alpha$ -chain is 141 residues long, and the  $\beta$ -chain is 146 residues long; for comparison, the myoglobin chain is 153 residues long. Many of the amino acids of the  $\alpha$ -chain, the  $\beta$ -chain, and myoglobin are *homologous*; that is, the same amino acid residues are in the same positions. The heme group is the same in myoglobin and hemoglobin.

We have already seen that one molecule of myoglobin binds one oxygen molecule. Four molecules of oxygen can therefore bind to one hemoglobin molecule. Both hemoglobin and myoglobin bind oxygen reversibly, but the binding of oxygen to hemoglobin exhibits **positive cooperativity**, whereas oxygen binding to myoglobin does not. Positive cooperativity means that when one oxygen molecule is bound, it becomes easier for the next to bind. A graph of the oxygen-binding properties of hemoglobin and myoglobin is one of the best ways to illustrate this point (Figure 4.22).

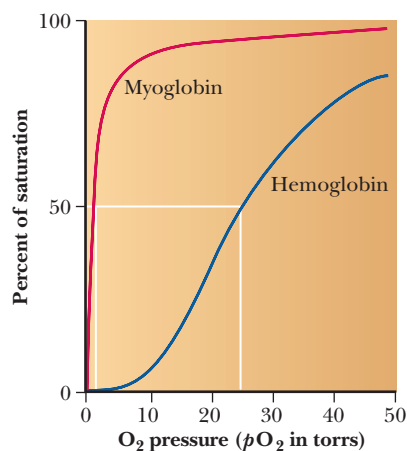
When the degree of saturation of myoglobin with oxygen is plotted against oxygen pressure, a steady rise is observed until complete saturation is approached and the curve levels off. The oxygen-binding curve of myoglobin is thus said to be **hyperbolic**. In contrast, the shape of the oxygen-binding curve for hemoglobin is **sigmoidal**. This shape indicates that the binding of the first oxygen molecule facilitates the binding of the second oxygen, which facilitates the binding of the third, which in turn facilitates the binding of the fourth. This is precisely what is meant by the term *cooperative binding*. However, note that even though cooperative binding means that binding of each subsequent



■ **FIGURE 4.21** The structure of hemoglobin. Hemoglobin ( $\alpha_2\beta_2$ ) is a tetramer consisting of four polypeptide chains (two  $\alpha$ -chains and two  $\beta$ -chains).



■ **FIGURE 4.20** Denaturation and refolding in ribonuclease. The protein ribonuclease can be completely denatured by the actions of urea and mercaptoethanol. When denaturing conditions are removed, activity is recovered.



■ **FIGURE 4.22** A comparison of the oxygen-binding behavior of myoglobin and hemoglobin. The oxygen-binding curve of myoglobin is hyperbolic, whereas that of hemoglobin is sigmoidal. Myoglobin is 50% saturated with oxygen at 1 torr partial pressure; hemoglobin does not reach 50% saturation until the partial pressure of oxygen reaches 26 torr.

oxygen is easier than the previous one, the binding curve is still lower than that of myoglobin at any oxygen pressure. In other words, at any oxygen pressure, myoglobin will have a higher percentage of saturation than hemoglobin.

### How does hemoglobin work?

The two different types of behavior exhibited by myoglobin and hemoglobin are related to the functions of these proteins. Myoglobin has the function of oxygen *storage* in muscle. It must bind strongly to oxygen at very low pressures, and it is 50% saturated at 1 torr partial pressure of oxygen. (The **torr** is a widely used unit of pressure, but it is not an SI unit. One torr is the pressure exerted by a column of mercury 1 mm high at 0°C. One atmosphere is equal to 760 torr.) The function of hemoglobin is oxygen *transport*, and it must be able both to bind strongly to oxygen and to release oxygen easily, depending on conditions. In the alveoli of lungs (where hemoglobin must bind oxygen for transport to the tissues), the oxygen pressure is 100 torr. At this pressure, hemoglobin is 100% saturated with oxygen. In the capillaries of active muscles, the pressure of oxygen is 20 torr, corresponding to less than 50% saturation of hemoglobin, which occurs at 26 torr. In other words, hemoglobin gives up oxygen easily in capillaries, where the need for oxygen is great.

Structural changes during binding of small molecules are characteristic of allosteric proteins such as hemoglobin. Hemoglobin has different quaternary structures in the bound (oxygenated) and unbound (deoxygenated) forms. The two  $\beta$ -chains are much closer to each other in oxygenated hemoglobin than in deoxygenated hemoglobin. The change is so marked that the two forms of hemoglobin have different crystal structures (Figure 4.23).

### Conformational Changes That Accompany Hemoglobin Function

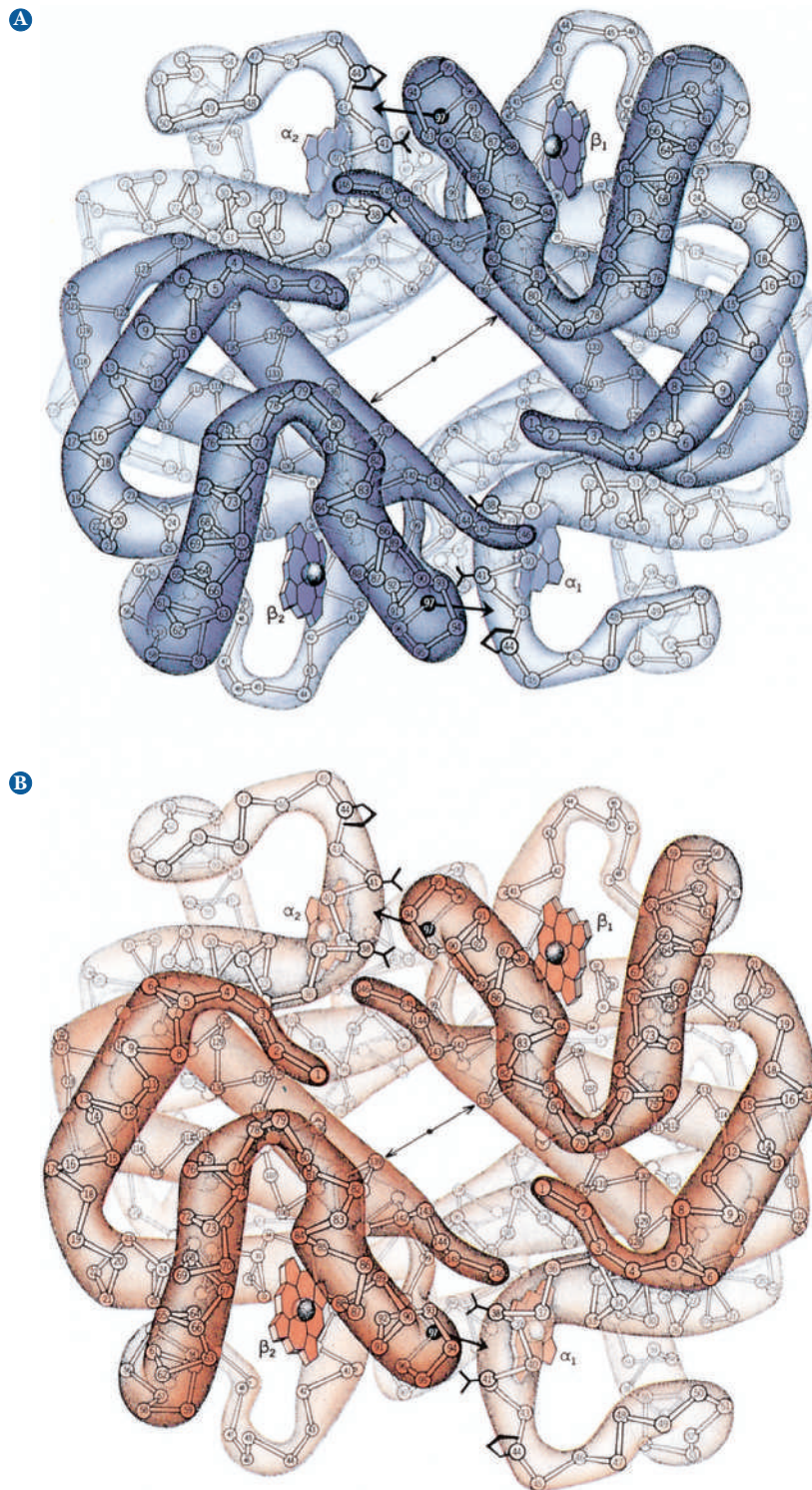
Other ligands are involved in cooperative effects when oxygen binds to hemoglobin. Both  $H^+$  and  $CO_2$ , which themselves bind to hemoglobin, affect the affinity of hemoglobin for oxygen by altering the protein's three-dimensional structure in subtle but important ways. The effect of  $H^+$  (Figure 4.24) is called the *Bohr effect*, after its discoverer, Christian Bohr (the father of physicist Niels Bohr). The oxygen-binding ability of myoglobin is not affected by the presence of  $H^+$  or of  $CO_2$ .

An increase in the concentration of  $H^+$  (i.e., a lowering of the pH) reduces the oxygen affinity of hemoglobin. Increasing  $H^+$  causes the protonation of key amino acids, including the N-terminals of the  $\alpha$ -chains and His<sup>146</sup> of the  $\beta$ -chains. The protonated histidine is attracted to, and stabilized by, a salt bridge to Asp<sup>94</sup>. This favors the deoxygenated form of hemoglobin. Actively metabolizing tissue, which requires oxygen, releases  $H^+$ , thus acidifying its local environment. Hemoglobin has a lower affinity for oxygen under these conditions, and it releases oxygen where it is needed (Figure 4.25). Hemoglobin's acid-base properties affect, and are affected by, its oxygen-binding properties. The oxygenated form of hemoglobin is a stronger acid (has a lower  $pK_a$ ) than the deoxygenated form. In other words, deoxygenated hemoglobin has a higher affinity for  $H^+$  than does the oxygenated form. Thus, changes in the quaternary structure of hemoglobin can modulate the buffering of blood through the hemoglobin molecule itself.

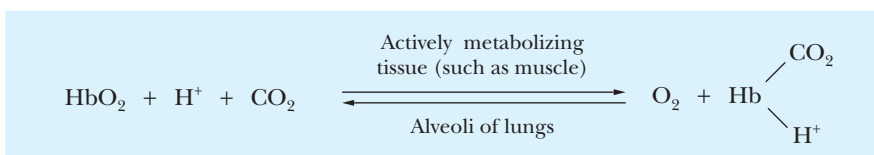
Table 4.1 summarizes the important features of the Bohr effect.

Large amounts of  $CO_2$  are produced by metabolism. The  $CO_2$ , in turn, forms carbonic acid,  $H_2CO_3$ . The  $pK_a$  of  $H_2CO_3$  is 6.35; the normal pH of blood is 7.4. As a result, about 90% of dissolved  $CO_2$  will be present as the bicarbonate ion,  $HCO_3^-$ , releasing  $H^+$ . (The Henderson–Hasselbalch equation can be used to confirm this point.) The *in vivo* buffer system involving  $H_2CO_3$  and  $HCO_3^-$  in blood was discussed in Section 2.5. The presence of larger amounts of  $H^+$  as a result of  $CO_2$  production favors the quaternary structure that is characteristic

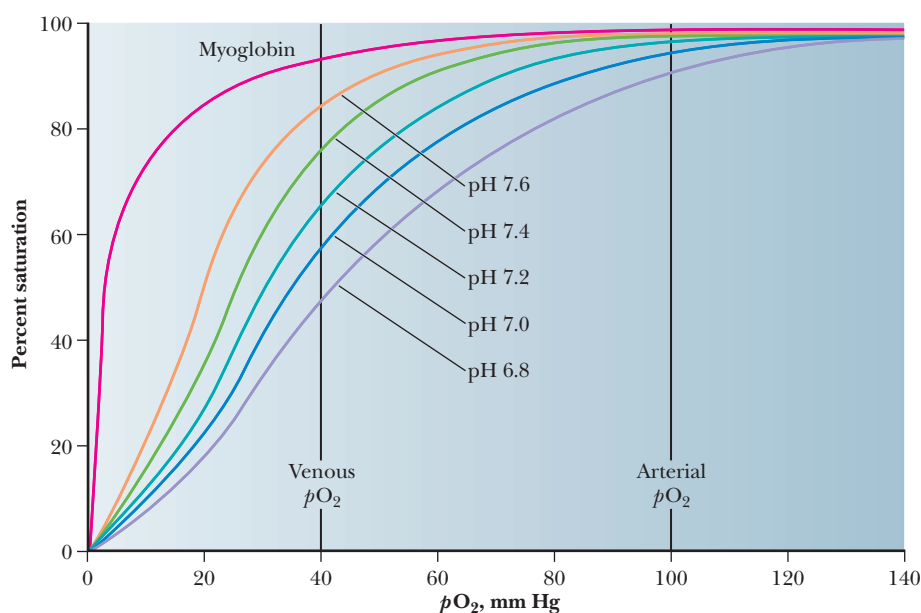




■ **FIGURE 4.23** The structures of (a) deoxyhemoglobin and (b) oxyhemoglobin. Note the motions of subunits with respect to one another. There is much less room at the center of oxyhemoglobin. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)



■ **FIGURE 4.24** The general features of the Bohr effect. In actively metabolizing tissue, hemoglobin releases oxygen and binds both  $\text{CO}_2$  and  $\text{H}^+$ . In the lungs, hemoglobin releases both  $\text{CO}_2$  and  $\text{H}^+$  and binds oxygen.

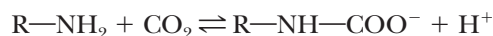


■ **FIGURE 4.25** The oxygen saturation curves for myoglobin and for hemoglobin at five different pH values.

**TABLE 4.1**

A Summary of the Bohr Effect	
Lungs	Actively Metabolizing Muscle
Higher pH than actively metabolizing tissue	Lower pH due to production of $H^+$
Hemoglobin binds $O_2$	Hemoglobin releases $O_2$
Hemoglobin releases $H^+$	Hemoglobin binds $H^+$

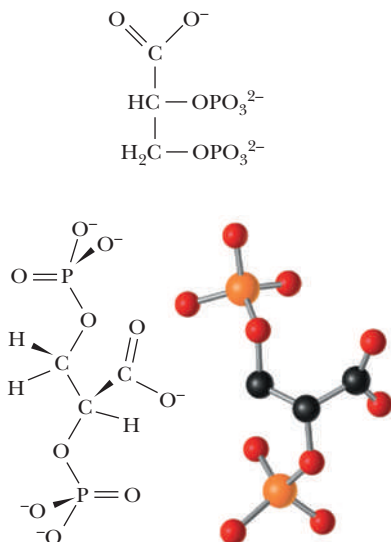
of deoxygenated hemoglobin. Hence, the affinity of hemoglobin for oxygen is lowered. The  $HCO_3^-$  is transported to the lungs, where it combines with  $H^+$  released when hemoglobin is oxygenated, producing  $H_2CO_3$ . In turn,  $H_2CO_3$  liberates  $CO_2$ , which is then exhaled. Hemoglobin also transports some  $CO_2$  directly. When the  $CO_2$  concentration is high, it combines with the free  $\alpha$ -amino groups to form carbamate:



This reaction turns the  $\alpha$ -amino terminals into anions, which can then interact with the  $\alpha$ -chain  $Arg^{141}$ , also stabilizing the deoxygenated form.

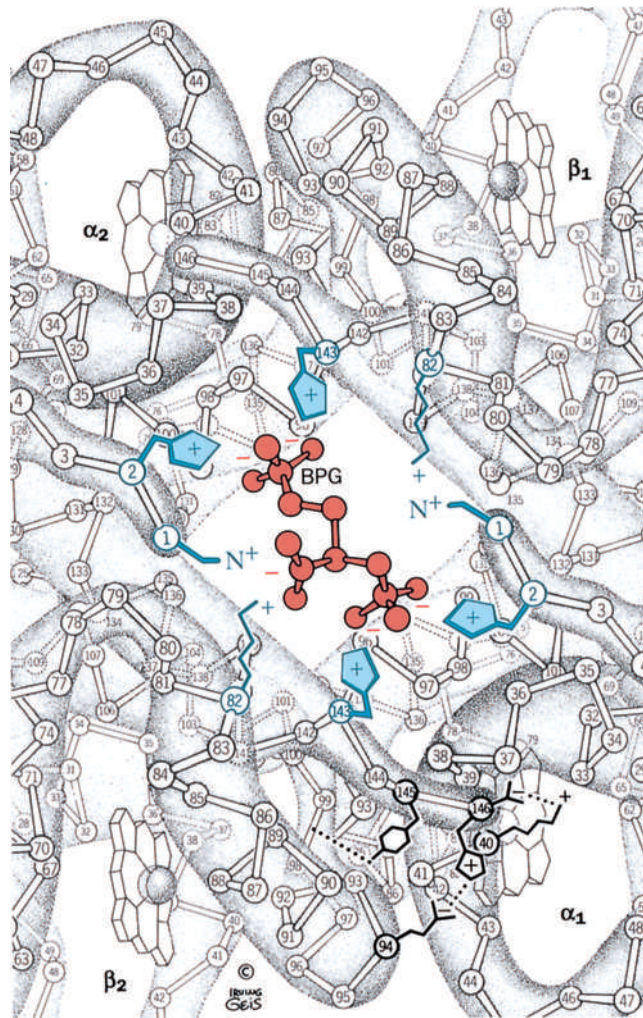
In the presence of large amounts of  $H^+$  and  $CO_2$ , as in respiring tissue, hemoglobin releases oxygen. The presence of large amounts of oxygen in the lungs reverses the process, causing hemoglobin to bind  $O_2$ . The oxygenated hemoglobin can then transport oxygen to the tissues. The process is complex, but it allows for fine-tuning of pH as well as levels of  $CO_2$  and  $O_2$ .

Hemoglobin in blood is also bound to another ligand, **2,3-bisphosphoglycerate (BPG)** (Figure 4.26), with drastic effects on its oxygen-binding capacity. The binding of BPG to hemoglobin is electrostatic; specific interactions take place between the negative charges on BPG and the positive charges on the protein (Figure 4.27). In the presence of BPG, the partial pressure at which 50% of hemoglobin is bound to oxygen is 26 torr. If BPG were not present in blood, the oxygen-binding capacity of hemoglobin would be much higher (50% of hemoglobin bound to oxygen at about 1 torr), and little oxygen would be released in the capillaries. “Stripped” hemoglobin, which is isolated from blood and from which the endogenous BPG has been removed, displays this behavior (Figure 4.28).

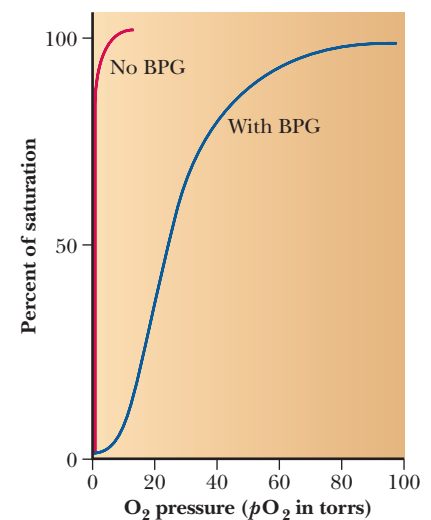


■ **FIGURE 4.26** The structure of BPG. BPG (2,3-bisphosphoglycerate) is an important allosteric effector of hemoglobin.





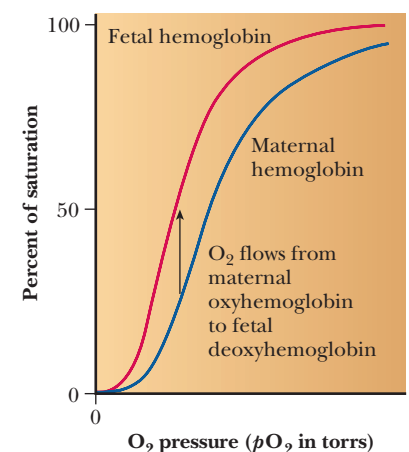
**FIGURE 4.27** The binding of BPG to deoxyhemoglobin. Note the electrostatic interactions between the BPG and the protein. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)



**FIGURE 4.28** A comparison of the oxygen-binding properties of hemoglobin in the presence and absence of BPG. Note that the presence of the BPG markedly decreases the affinity of hemoglobin for oxygen.

BPG also plays a role in supplying a growing fetus with oxygen. The fetus obtains oxygen from the mother's bloodstream via the placenta. Fetal hemoglobin (Hb F) has a higher affinity for oxygen than does maternal hemoglobin, allowing for efficient transfer of oxygen from the mother to the fetus (Figure 4.29). Two features of fetal hemoglobin contribute to this higher oxygen-binding capacity. One is the presence of two different polypeptide chains. The subunit structure of Hb F is  $\alpha_2\gamma_2$ , where the  $\beta$ -chains of adult hemoglobin (Hb A), the usual hemoglobin, have been replaced by the  $\gamma$ -chains, which are similar but not identical in structure. The second feature is that Hb F binds less strongly to BPG than does Hb A. In the  $\beta$ -chain of adult hemoglobin, His<sup>143</sup> makes a salt bridge to BPG. In the fetal hemoglobin, the  $\gamma$ -chain has an amino acid substitution of a serine for His<sup>143</sup>. This change of a positively charged amino acid for a neutral one diminishes the number of contacts between the hemoglobin and the BPG, effectively reducing the allosteric effect enough to give fetal hemoglobin a higher binding curve than adult hemoglobin.

Another type of hemoglobin that has been studied extensively is sickle-cell hemoglobin, Hb S. In Hb S, the  $\beta$ -chains have a single amino acid substitution of a glutamic acid for a valine. This substitution of a nonpolar amino acid for a polar one causes the characteristic effects of the disease. The nonpolar amino acid is on the surface and leads to aggregation of the molecules through nonpolar interactions. These aggregations lead to the sickling of the blood cells.



**FIGURE 4.29** A comparison of the oxygen-binding capacity of fetal and maternal hemoglobins. Fetal hemoglobin binds less strongly to BPG and, consequently, has a greater affinity for oxygen than does maternal hemoglobin.



### Apply Your Knowledge

#### Oxygen Response to pH in Hemoglobin

Assume that during a 400-m running race, the pH decreases in muscle cells from 7.6 to 7.0 while the  $pO_2$  remains constant at 40 mm Hg. What is the effect on the oxygen binding of hemoglobin in the muscle cells? What is the implication of this? What is the effect on myoglobin?

#### Solution

Using Figure 4.25, we can see that at pH 7.6 and 40 mm Hg, the hemoglobin is about 82% saturated. At pH 7.0, this drops to about 58% saturated. This means that hemoglobin will bind less oxygen at lower pH or, in other words, it will release more oxygen to the muscle cells. Myoglobin does not have a Bohr effect, however, and there is no effect caused by lowering the pH.

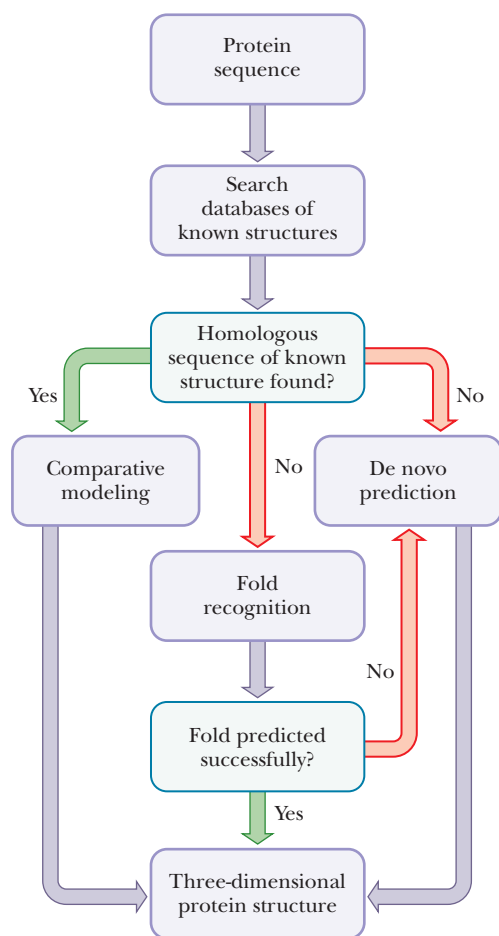
## 4.6 Protein Folding Dynamics

We know that the sequence of amino acids ultimately determines the three-dimensional structure of a protein. We also know that proteins can spontaneously adopt their native conformations, be denatured, and be renatured back into their native conformations, as was shown in Figure 4.20. These facts can lead us to the following question:

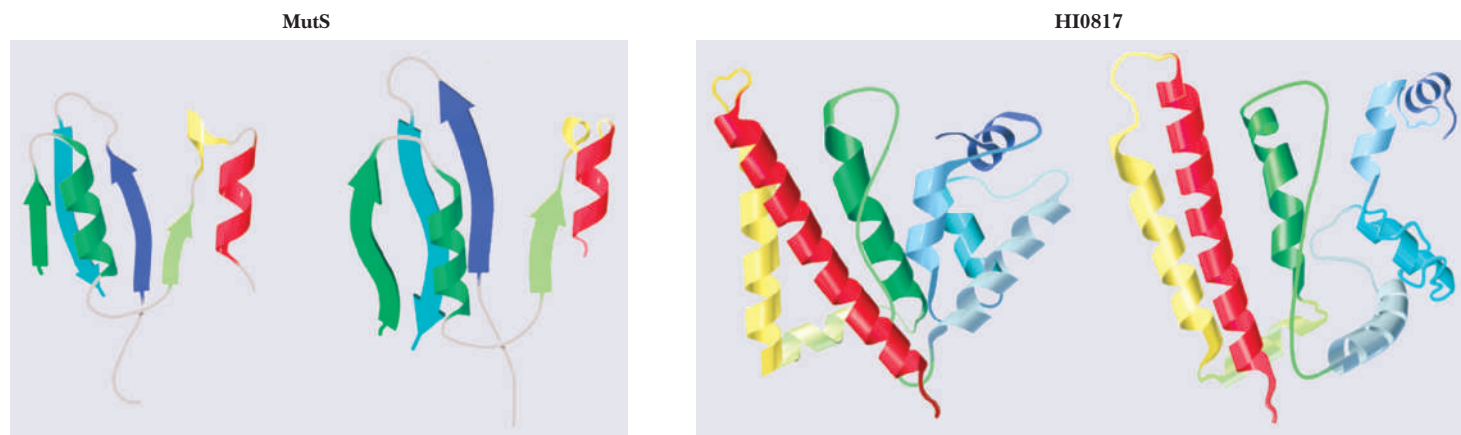
### Can we predict the tertiary structure of a protein if we know its amino acid sequence?

With modern computing techniques, we are able to predict protein structure. This is becoming more and more possible as more powerful computers allow the processing of large amounts of information. The encounter of biochemistry and computing has given rise to the burgeoning field of **bioinformatics**. Prediction of protein structure is one of the principal applications of bioinformatics. Another important application is the comparison of base sequences in nucleic acids, a topic we shall discuss in Chapter 14, along with other methods for working with nucleic acids. As we shall see, we can now predict protein structure and function by knowing the nucleotide sequence of the gene that eventually leads to the final protein.

The first step in predicting protein architecture is a search of databases of known structures for *sequence homology* between the protein whose structure is to be determined and proteins of known architecture, where the term **homology** refers to similarity of two or more sequences. If the sequence of the known protein is similar enough to that of the protein being studied, the known protein's structure becomes the point of departure for *comparative modeling*. Use of modeling algorithms that compare the protein being studied with known structures leads to a structure prediction. This method is most useful when the sequence homology is greater than 25%–30%. If the sequence homology is less than 25%–30%, other approaches are more useful. *Fold recognition* algorithms allow comparison with known folding motifs common to many secondary structures. We saw a number of these motifs in Section 4.3. Here is an application of that information. Yet another method is *de novo prediction*, based on first principles from chemistry, biology, and physics. This method too can give rise to structures subsequently confirmed by X-ray crystallography. The flow chart in Figure 4.30 shows how prediction techniques use existing information from databases. Figure 4.31 shows a comparison of the predicted structures of two proteins (on the right side) for the DNA repair protein MutS and the bacterial protein HI0817. The crystal structures of the two proteins are shown on the left.



■ **FIGURE 4.30** Predicting protein conformation. A flow chart showing the use of existing information from databases to predict protein conformation. (Courtesy of Rob Russell, EMBL.)



■ **FIGURE 4.31** Predicted versus actual protein structures. A comparison of the predicted structures of two proteins (on the right side) for the DNA repair protein MutS and the bacterial protein HI0817. The crystal structures of the two proteins are shown on the left. (Courtesy of University of Washington, Seattle.)

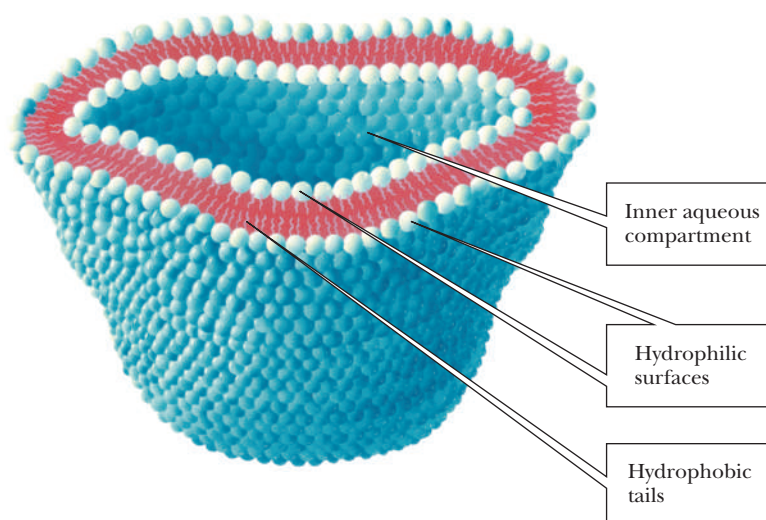
A considerable amount of information about protein sequences and architecture is available on the World Wide Web. One of the most important resources is the Protein Data Bank operated under the auspices of the Research Collaboratory for Structural Bioinformatics (RCSB). Its URL is <http://www.rcsb.org/pdb>. This site, which has a number of mirror sites around the world, is the single repository of structural information about large molecules. It includes material about nucleic acids as well as proteins. Its home page has a button with links specifically geared to educational applications.

Results of structure prediction using the methods discussed in this section are available on the Web as well. One of the most useful URLs is <http://predictioncenter.gc.ucdavis.edu>. Other excellent sources of information are available through the National Institutes of Health (<http://pubmedcentral.nih.gov/tocrender.fcgi?iid=1005> and <http://www.ncbi.nlm.nih.gov>), and through the ExPASy (Expert Protein Analysis System) server (<http://us.expasy.org>).

### Hydrophobic Interactions: A Case Study in Thermodynamics

We briefly introduced the notion of hydrophobic interactions in Section 4.4. Hydrophobic interactions have important consequences in biochemistry and play a major role in protein folding. Large arrays of molecules can take on definite structures as a result of hydrophobic interactions. We have already seen the way in which phospholipid bilayers can form one such array. Recall (Chapter 2, Section 2.1) that phospholipids are molecules that have polar head groups and long nonpolar tails of hydrocarbon chains. These bilayers are less complex than a folded protein, but the interactions that lead to their formation also play a vital role in protein folding. Under suitable conditions, a double-layer arrangement is formed so that the polar head groups of many molecules face the aqueous environment, while the nonpolar tails are in contact with each other and are kept away from the aqueous environment. These bilayers form three-dimensional structures called **liposomes** (Figure 4.32). Such structures are useful model systems for biological membranes, which consist of similar bilayers with proteins embedded in them. The interactions between the bilayer and the embedded proteins are also examples of hydrophobic interactions. The very existence of membranes depends on hydrophobic interactions. The same hydrophobic interactions play a crucial role in protein folding.

Hydrophobic interactions are a major factor in the folding of proteins into the specific three-dimensional structures required for their functioning as enzymes, oxygen carriers, or structural elements. It is known experimentally that



■ **FIGURE 4.32** Schematic diagram of a liposome. This three-dimensional structure is arranged so that hydrophilic head groups of lipids are in contact with the aqueous environment. The hydrophobic tails are in contact with each other and are kept away from the aqueous environment.

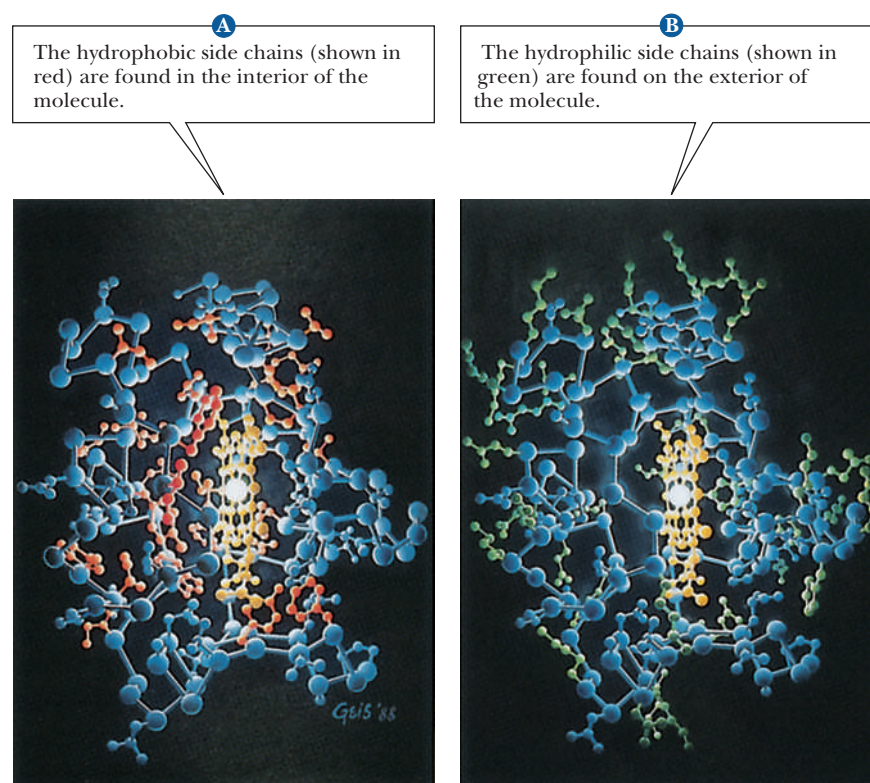
proteins tend to be folded so that the nonpolar hydrophobic side chains are sequestered from water in the interior of the protein, while the polar hydrophilic side chains lie on the exterior of the molecule and are accessible to the aqueous environment (Figure 4.33).

### What makes hydrophobic interactions favorable?

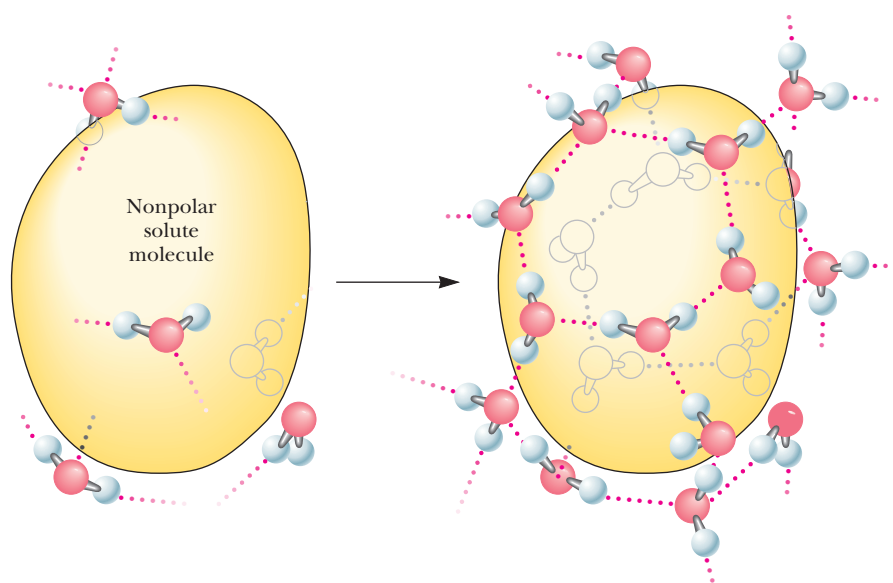
Hydrophobic interactions are spontaneous processes. The entropy of the Universe increases when hydrophobic interactions occur.

$$\Delta S_{\text{univ}} > 0$$

As an example, let us assume that we have tried to mix the liquid hydrocarbon hexane ( $\text{C}_6\text{H}_{14}$ ) with water and have obtained not a solution but a two-layer



■ **FIGURE 4.33** The three-dimensional structure of the protein cytochrome c. (Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)



■ **FIGURE 4.34** A “cage” of water molecules forms around a nonpolar solute.

system, one layer of hexane and one of water. Formation of a mixed solution is nonspontaneous, and the formation of two layers is spontaneous. Unfavorable entropy terms enter into the picture if solution formation requires the creation of ordered arrays of solvent, in this case water (Figure 4.34). The water molecules surrounding the nonpolar molecules can hydrogen bond with each other, but they have fewer possible orientations than if they were surrounded by other water molecules on all sides. This introduces a higher degree of order, preventing the dispersion of energy, more like the lattice of ice than liquid water, and thus a lower entropy. The required entropy decrease is too large for the process to take place. Therefore, nonpolar substances do not dissolve in water; rather, nonpolar molecules associate with one another by hydrophobic interactions and are excluded from water.

Many people think of hydrophobic interactions between amino acids backward. For example, if we look at Figure 4.13 and see the indication of hydrophobic interactions between leucine, valine, and isoleucine, we might conclude that hydrophobic interactions refer to an attraction for these amino acids for each other. However, we now know that in reality it is not so much the attraction of the nonpolar amino acids for each other, but rather it is more that they are forced together so that water can avoid having to interact with them.

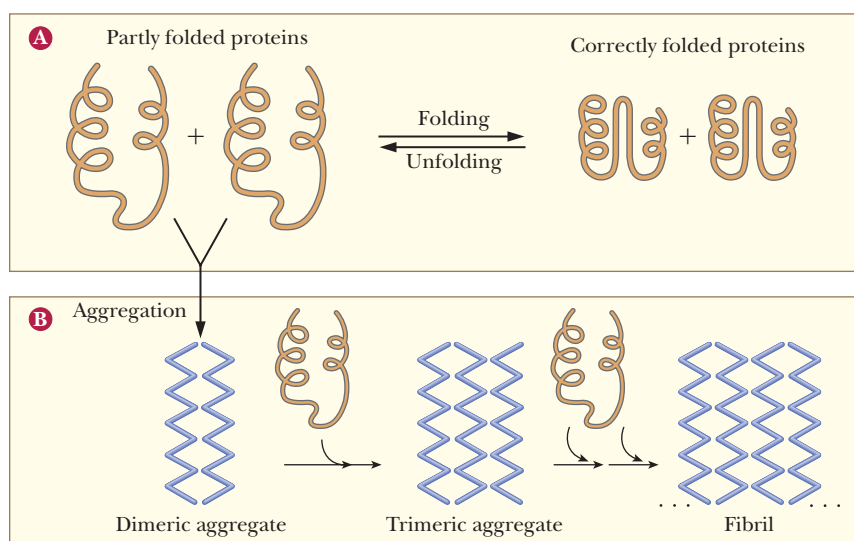
## The Importance of Correct Folding

The primary structure conveys all the information necessary to produce the correct tertiary structure, but the folding process *in vivo* can be a bit trickier. In the protein-dense environment of the cell, proteins may begin to fold incorrectly as they are produced, or they may begin to associate with other proteins before completing their folding process. In eukaryotes, proteins may need to remain unfolded long enough to be transported across the membrane of a subcellular organelle.

Correctly folded proteins are usually soluble in the aqueous cell environment, or they are correctly attached to membranes. However, when proteins do not fold correctly, they may interact with other proteins and form aggregates as shown in Figure 4.35. This occurs because hydrophobic regions that should be buried inside the protein remain exposed and interact with other hydrophobic regions on other molecules. Several neurodegenerative disorders, such as Alzheimer's, Parkinson's, and Huntington's diseases, are caused by accumulation of protein deposits from such aggregates. See the following Biochemical Connections box for a description of some of the deadly diseases caused by protein misfolding.



■ **FIGURE 4.35** The problem of protein aggregation. (a) Partly folded polypeptide chains, released from ribosomes (the protein-synthesizing machines), normally form correctly folded, functional proteins. (b) However, partly folded proteins may sometimes associate with similar chains to form aggregates. Both soluble and insoluble aggregates can be toxic to cells. [Reprinted by permission from “Danger—Misfolding Proteins” by R. J. Ellis and T. J. T. Pinheiro, *Nature* 416, 483–484 (2002).]

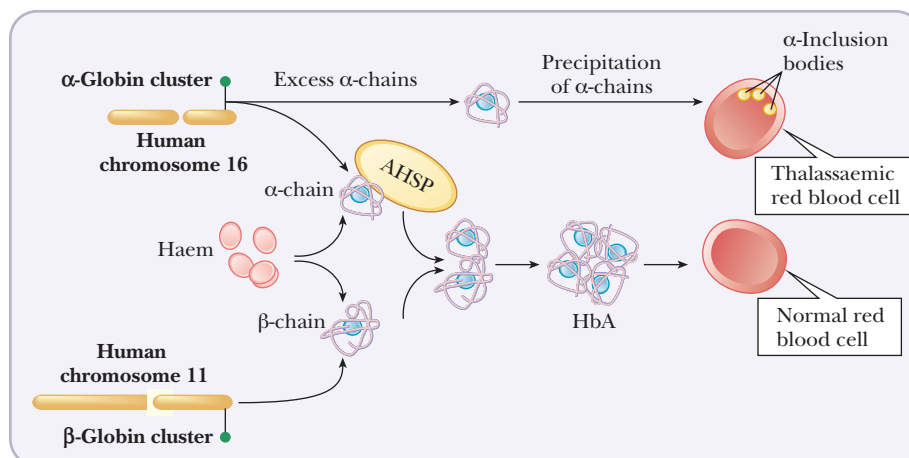


### Protein-Folding Chaperones

To help avoid the protein misfolding problem, special proteins called chaperones aid in the correct and timely folding of many proteins. The chaperone protein gets its name from the old-fashioned notion of sending a young person on a date with a “protector,” called a chaperone, who would make sure the date did not stray from socially acceptable behavior. In other words, the chaperone prevents “unsuitable” liaisons. In protein folding dynamics, the chaperone does the same thing. It either prevents a protein from associating with another protein with which it should not associate or keeps it from associating with itself in inappropriate ways. The first such proteins discovered were a family called hsp70 (for 70,000 MW *heat-shock protein*), which are proteins produced in *E. coli* grown above optimal temperatures. Chaperones exist in organisms from prokaryotes through humans, and their mechanisms of action are currently being studied. It is becoming more and more evident that protein folding dynamics are crucial to protein function in vivo. To conclude this chapter and finish our study of protein structure, we will look at a chaperone that aids the proper formation of hemoglobin.

In the blood, hemoglobin accumulates to a level of 340 g/L, which is a very large amount of a single protein. The control of globin gene expression is complicated and made more so by the fact that there are separate genes for the  $\alpha$ -chain and the  $\beta$ -chain, and they are found on different chromosomes. There are also two  $\alpha$ -globin genes for every  $\beta$ -globin gene, so there is always an excess of the  $\alpha$ -chain. Excess  $\alpha$ -chains can form aggregates as shown in Figure 4.36,

■ **FIGURE 4.36** Balancing the components of hemoglobin. The  $\alpha$ - and  $\beta$ -globin genes are on different chromosomes. Excess  $\alpha$ -chain is produced. If excess  $\alpha$ -chains can interact, they form aggregates called  $\alpha$ -inclusion bodies that damage red blood cells. The globin chaperone (AHSP) binds to  $\alpha$ -globin and both keeps it from aggregating with itself and delivers it to the  $\beta$ -globin so that the  $\alpha$ -globin and  $\beta$ -globin can bind together to form the active tetramer. [Reprinted by permission from “Haemoglobin’s Chaperone” by L. Luzzatto and R. Notaro, *Nature* 417, 703–705 (2002).]





which could lead to damaged red blood cells and a disease called *thalassemia*. The  $\alpha$ -chains can also form aggregates among themselves, leading to a useless form of hemoglobin. The secret to success for hemoglobin production is to maintain the proper stoichiometry between the two types of globin chains. The  $\alpha$ -chains must be kept from aggregating together so that there will be enough  $\alpha$ -chain to complex with the  $\beta$ -chain. In this way the  $\alpha$ -chains will be occupied with  $\beta$ -chains and will not form  $\alpha$ -chain aggregates. Fortunately, there is a specific chaperone for the  $\alpha$ -chain, called  $\alpha$ -hemoglobin stabilizing protein (AHSP). This chaperone prevents the  $\alpha$ -chains from causing the damage to blood cells as well as delivering them to the  $\beta$ -chains.

Protein folding is a very hot topic in biochemistry today. The following Biochemical Connections box describes a particularly striking examples of the importance of protein folding.

## Biochemical Connections MEDICINE

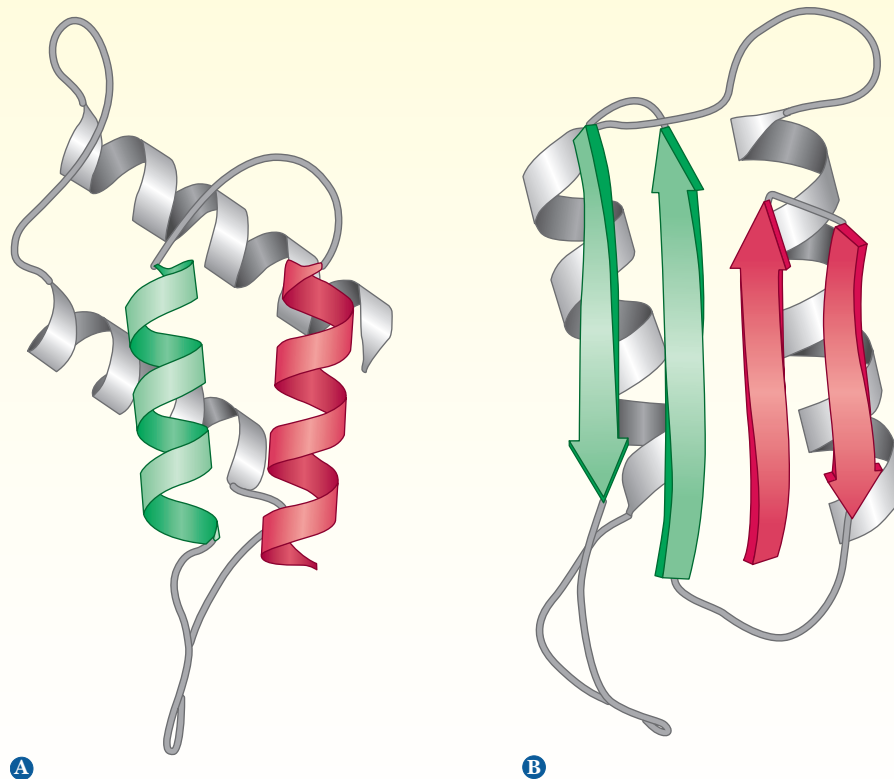
### Protein Folding Diseases

There are several well-known diseases that are caused by misfolded proteins, including Creutzfeldt-Jakob disease, Alzheimer's disease, Parkinson's disease, and Huntington's disease. We will look at a couple of these in detail here.

#### Prion Diseases

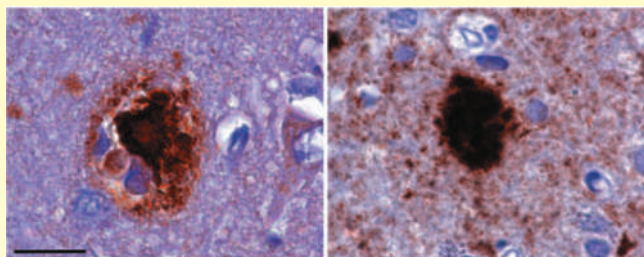
It has been established that the causative agent of mad-cow disease (also known as bovine spongiform encephalopathy or BSE), as well as the related diseases scrapie in sheep, chronic wasting disease (CWD) in deer and elk, and human spongiform

encephalopathy (kuru and Creutzfeldt-Jakob disease) in humans, is a small (28-kDa) protein called a prion. (Note that biochemists tend to call the unit of atomic mass the *dalton*, abbreviated Da.) Prions are natural glycoproteins found in the cell membranes of nerve tissue. Recently the prion protein has been found in the cell membrane of hematopoietic stem cells, precursors to the cells of the bloodstream, and there is some evidence that the prion helps guide cell maturation. The disease state comes about when the normal form of the prion protein, PrP (Figure a), folds into an incorrect form called PrP<sup>sc</sup> (Figure b). Like a bad role



■ (a) Normal prion structure (PrP). (b) Abnormal prion (PrP<sup>sc</sup>).

## Biochemical Connections (CONTINUED)

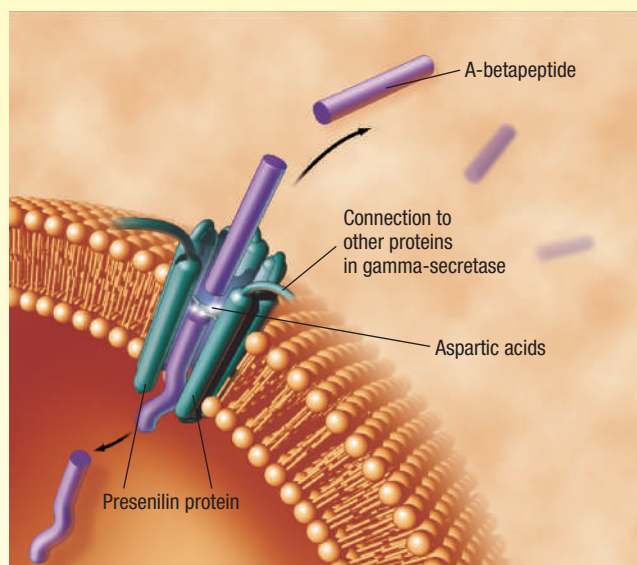


**(c)** Shows a tissue sample from a patient with Alzheimer's disease. **(d)** Shows one from a patient with Creutzfeldt-Jakob disease. (From *Games Played by Rogue Proteins in Prion Disorders and Alzheimer's Disease* by Adrian Aguzzi and Christian Haass (31 October 2003) *Science* 302 (5646), 814. Reprinted with permission from AAAS.)

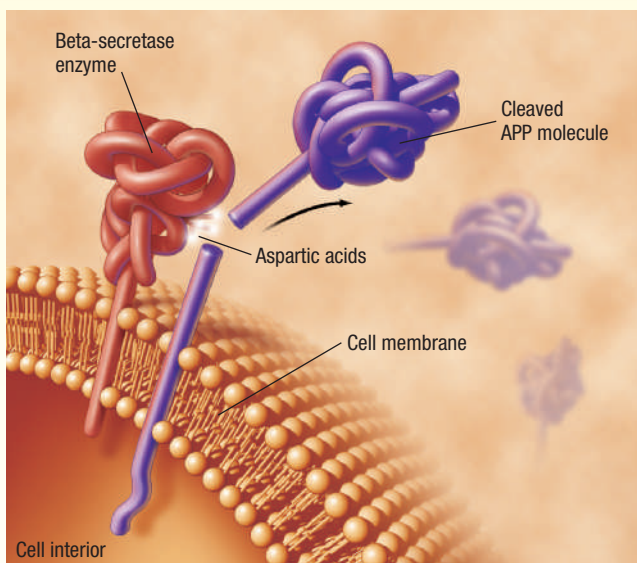
model, these abnormal forms of the prion protein are able to convert other, normal forms into abnormal forms. This change can be propagated in nervous tissue. Scrapie had been known for years, but it had not been known to cross species barriers. Then an outbreak of mad-cow disease was shown to have followed the inclusion of sheep remains in cattle feed. It is now known that eating tainted beef from animals with mad-cow disease can cause spongiform encephalopathy, now known as new variant Creutzfeldt-Jakob disease (vCJD), in humans. The normal prions have a large percentage of  $\alpha$ -helix, but the abnormal forms have more  $\beta$ -pleated sheets. Notice that in this case the same protein (a single, well-defined sequence) can exist in alternative forms. These  $\beta$ -pleated sheets in the abnormal proteins interact between protein molecules and form insoluble plaques, a fate also seen in Alzheimer's disease and several other neurological diseases. The presence of these plaques can be seen with immuno-stained tissue samples from the brains of people inflicted with the diseases. Figure c shows a tissue sample from a patient with Alzheimer's disease, and Figure d shows one from a patient with Creutzfeldt-Jakob disease. [Reprinted by permission from "Games Played by Rogue Proteins in Prion Disorders and Alzheimer's Disease." by A. Aguzzi and C. Haass. *Science* 302, fig. 1, 814A, (2003).] Ingested abnormal prions use macrophages from the immune system to travel in the body until they come in contact with nerve tissue. They can then propagate up the nerves until they reach the brain.

This mechanism was a subject of considerable controversy when it was first proposed. A number of scientists expected that a slow-acting virus would be found to be the ultimate cause of these neurological diseases. A susceptibility to these diseases can be inherited, so some involvement of DNA (or RNA) was also expected. Some went so far as to talk about "heresy" when Stanley Prusiner received the 1997 Nobel Prize in Medicine for his discovery of prions, but substantial evidence shows that prions are themselves the infectious agent and that no virus or bacteria are involved. It now appears that genes for susceptibility to the incorrect form exist in all vertebrates, giving rise to the observed pattern of disease transmission, but many individuals with the genetic susceptibility never develop the disease if they do not come in contact with abnormal prions from another source. This combination of genetic predisposition combined with transmission by an infectious agent makes prion diseases unique.

Further studies have shown that all of the humans who showed symptoms of vCJD had the same amino acid substitution in their prions, a substitution of a methionine at position 129, now known to confer extreme sensitivity to the disease.



**(E)**



**(F)**

■ In the first step (e), the enzyme  $\beta$ -secretase cuts APP outside the cell membrane. Then the  $\gamma$ -secretase enzyme, itself located in the membrane, cuts the remaining portion of the APP inside the membrane, releasing  $A\beta$  (f).

### Alzheimer's and Other Neurodegenerative Diseases

Direct prion diseases like BSE or Creutzfeldt-Jakob disease are not the only diseases to involve prion-like pathologies. Although, unlike BSE, diseases such as Alzheimer's, Parkinson's, and Huntington's are not transmissible, they do have pathologies involving plaque-forming proteins, and they do affect brain functioning, often with similar effects. They also seem to have the same progression of abnormal proteins propagating in nerve tissue.

## Biochemical Connections (CONTINUED)

About a third of people in their 80s will show signs of Alzheimer's, a disease that erases memories, usually starting with more recent ones and proceeding in later stages to remove long-held memories. Eventually patients don't recognize their closest family members. At the heart of the disease is the destruction of neurons in the brain. The destruction is believed to be caused by plaques formed by two different proteins, one called amyloid  $\beta$ , or  $A\beta$ , and the other one called Tau.  $A\beta$  is a short peptide first isolated in 1984. It is derived from a larger protein called amyloid beta precursor protein (APP). APP is a membrane-bound protein, partly found inside a neuron and partly outside it. Two protein-cutting enzymes called  $\beta$ -secretase and  $\gamma$ -secretase cut out  $A\beta$  from APP, a normal process that occurs in most cells of the body. Scientists believe that the metabolism of the  $A\beta$  is what causes the disease. At high enough concentrations, the  $A\beta$  proteins form insoluble fibers and form plaques on the nerve cells.

Mutations in the genes that produce  $\gamma$ -secretase have been shown to cause an early and aggressive form of Alzheimer's.

Other mutations of APP itself are also implicated. Combining the evidence of the importance of the  $A\beta$  peptide and the enzymes that create it has led to the "amyloid-cascade hypothesis." According to this hypothesis, Alzheimer's begins with the buildup of  $A\beta$ , which is cut from the APP. In the first step (figure e), the enzyme  $\beta$ -secretase cuts APP outside the cell membrane. Then the  $\gamma$ -secretase enzyme, itself located in the membrane, cuts the remaining portion of the APP inside the membrane, releasing  $A\beta$  (figure f).

Researchers are focusing their search for an Alzheimer's cure on this process. Drugs have been developed that inhibit the activity of the  $\gamma$ -secretase. Others cause the  $\beta$ -secretase enzyme to cut APP at a different location, which produces a shorter and less harmful form of  $A\beta$ . Unfortunately, just killing off the activity of the enzymes comes with its own dangers, as the enzymes do have a natural function within nerve tissue too. For example,  $\beta$ -secretase is also involved in proper myelination of nerves.

## SUMMARY

**What are the levels of protein structure?** There are four levels of protein structure: primary, secondary, tertiary, and quaternary. Not all proteins have all four levels. For example, only proteins with multiple polypeptide chains have quaternary structure.

**Why is it important to know the primary structure?** Primary structure is the order in which the amino acids are covalently linked. The primary structure of a protein can be determined by chemical methods. The amino acid sequence (the primary structure) of a protein determines its three-dimensional structure, which in turn determines its properties. A striking example of the importance of primary structure is sickle-cell anemia, a disease caused by a change in one amino acid in each of two of the four chains of hemoglobin.

**Why is the  $\alpha$ -helix so prevalent?** The  $\alpha$ -helix is stabilized by hydrogen bonds parallel to the helix axis within the backbone of a single polypeptide chain. The helical conformation allows a linear arrangement of the atoms involved in the hydrogen bonds, which gives the bonds maximum strength and thus makes the helical conformation very stable.

**How is the  $\beta$ -sheet different from the  $\alpha$ -helix?** The arrangement of atoms in the  $\beta$ -pleated sheet conformation differs markedly from that in the  $\alpha$ -helix. The peptide backbone in the  $\beta$ -sheet is almost completely extended. Hydrogen bonds can be formed between different parts of a single chain that is doubled back on itself (intrachain bonds) or between different chains (interchain bonds). The hydrogen bonding between peptide chains in the  $\beta$ -pleated sheet gives rise to a repeated zigzag structure. The hydrogen bonds are

perpendicular to the direction of the protein chain, not parallel to it as in the  $\alpha$ -helix.

**How can the three-dimensional structure of a protein be determined?** The experimental technique used to determine the tertiary structure of a protein is X-ray crystallography. Perfect crystals of some proteins can be grown under carefully controlled conditions. When a suitably pure crystal is exposed to a beam of X-rays, a diffraction pattern is produced on a photographic plate or a radiation counter. The pattern is produced when the electrons in each atom in the molecule scatter the X-rays. The scattered X-rays from the individual atoms can reinforce each other or cancel each other (set up constructive or destructive interference), giving rise to a characteristic pattern for each type of molecule.

**Why does oxygen have imperfect binding to the heme group?** More than one type of molecule can bind to heme. Besides oxygen, carbon monoxide also binds to heme. The affinity of free heme for carbon monoxide (CO) is 25,000 times greater than its affinity for oxygen. When carbon monoxide is forced to bind at an angle in myoglobin, its advantage over oxygen drops by two orders of magnitude. This guards against the possibility that traces of CO produced during metabolism would occupy all the oxygen-binding sites on the hemes.

**How does hemoglobin work?** The function of hemoglobin is oxygen transport, and it must be able both to bind strongly to oxygen and to release oxygen easily, depending on conditions. In hemoglobin, the binding of oxygen is cooperative



(as each oxygen is bound, it becomes easier for the next one to bind) and is modulated by such ligands as  $H^+$ ,  $CO_2$ , and BPG. The binding of oxygen to myoglobin is not cooperative.

**Can we predict the tertiary structure of a protein if we know its amino acid sequence?** It is possible, to some extent, to predict the three-dimensional structure of a protein from its amino acid sequence. Computer algorithms are based on two approaches, one of which is based on comparison of sequences

with those of proteins whose folding pattern is known. Another one is based on the folding motifs that occur in many proteins.

**What makes hydrophobic interactions favorable?** Hydrophobic interactions are spontaneous processes. The entropy of the Universe increases when hydrophobic interactions occur. Hydrophobic interactions, which depend on the unfavorable entropy of the water of hydration surrounding nonpolar solutes, are particularly important determinants of protein folding.

## REVIEW EXERCISES

 Interactive versions of these problems are assignable in OWL

### 4.1 Protein Structure and Function

- Recall** Match the following statements about protein structure with the proper levels of organization. (i) Primary structure (ii) Secondary structure (iii) Tertiary structure (iv) Quaternary structure
  - The three-dimensional arrangement of all atoms
  - The order of amino acid residues in the polypeptide chain
  - The interaction between subunits in proteins that consist of more than one polypeptide chain
  - The hydrogen-bonded arrangement of the polypeptide backbone.
- Recall** Define denaturation in terms of the effects of secondary, tertiary, and quaternary structure.
- Recall** What is the nature of “random” structure in proteins?

### 4.2 Primary Structure of Proteins

- Reflect and Apply** Suggest an explanation for the observation that when proteins are chemically modified so that specific side chains have a different chemical nature, these proteins cannot be denatured reversibly.
- Reflect and Apply** Rationalize the following observations.
  - Serine is the amino acid residue that can be replaced with the least effect on protein structure and function.
  - Replacement of tryptophan causes the greatest effect on protein structure and function.
  - Replacements such as  $Lys \rightarrow Arg$  and  $Leu \rightarrow Ile$  usually have very little effect on protein structure and function.
- Reflect and Apply** Glycine is a highly conserved amino acid residue in proteins (i.e., it is found in the same position in the primary structure of related proteins). Suggest a reason why this might occur.
- Reflect and Apply** A mutation that changes an alanine residue in a protein to an isoleucine leads to a loss of activity. Activity is regained when a further mutation at the same site changes the isoleucine to a glycine. Why?
- Reflect and Apply** A biochemistry student characterizes the process of cooking meat as an exercise in denaturing proteins. Comment on the validity of this remark.
- Biochemical Connections** Severe combined immunodeficiency disease (SCID) is characterized by the complete lack of an immune system. Strains of mice have been developed that have SCID. When SCID mice that carry genetic predisposition to prion diseases are infected with  $PrP^{Sc}$ , they do not develop prion diseases. How do these facts relate to the transmission of prion diseases?

- Biochemical Connections** An isolated strain of sheep was found in New Zealand. Most of these sheep carried the gene for predisposition to scrapie, yet none of them ever came down with the disease. How do these facts relate to the transmission of prion diseases?

### 4.3 Secondary Structure of Proteins

- Recall** List three major differences between fibrous and globular proteins.
- Biochemical Connections** What is a protein efficiency ratio?
- Biochemical Connections** Which food has the highest PER?
- Biochemical Connections** What are the essential amino acids?
- Biochemical Connections** Why are scientists currently trying to create genetically modified foods?
- Recall** What are Ramachandran angles?
- Recall** What is a  $\beta$ -bulge?
- Recall** What is a reverse turn? Draw two types of reverse turns.
- Recall** List some of the differences between the  $\alpha$ -helix and  $\beta$ -sheet forms of secondary structure.
- Recall** List some of the possible combinations of  $\alpha$ -helices and  $\beta$ -sheets in supersecondary structures.
- Recall** Why is proline frequently encountered at the places in the myoglobin and hemoglobin molecules where the polypeptide chain turns a corner?
- Recall** Why must glycine be found at regular intervals in the collagen triple helix?
- Reflect and Apply** You hear the comment that the difference between wool and silk is the difference between helical and pleated-sheet structures. Do you consider this a valid point of view? Why or why not?
- Reflect and Apply** Woolen clothing shrinks when washed in hot water, but items made of silk do not. Suggest a reason, based on information from this chapter.

### 4.4 Tertiary Structure of Proteins

- Recall** Draw two hydrogen bonds, one that is part of a secondary structure and another that is part of a tertiary structure.
- Recall** Draw a possible electrostatic interaction between two amino acids in a polypeptide chain.
- Recall** Draw a disulfide bridge between two cysteines in a polypeptide chain.
- Recall** Draw a region of a polypeptide chain showing a hydrophobic pocket containing nonpolar side chains.

29. **Reflect and Apply** The terms *configuration* and *conformation* appear in descriptions of molecular structure. How do they differ?
30. **Reflect and Apply** Theoretically, a protein could assume a virtually infinite number of configurations and conformations. Suggest several features of proteins that drastically limit the actual number.
31. **Reflect and Apply** What is the highest level of protein structure found in collagen?

#### 4.5 Quaternary Structure of Proteins

32. **Recall** List two similarities and two differences between hemoglobin and myoglobin.
33. **Recall** What are the two critical amino acids near the heme group in both myoglobin and hemoglobin?
34. **Recall** What is the highest level of organization in myoglobin? In hemoglobin?
35. **Recall** Suggest a way in which the difference between the functions of hemoglobin and myoglobin is reflected in the shapes of their respective oxygen-binding curves.
36. **Recall** Describe the Bohr effect.
37. **Recall** Describe the effect of 2,3-bisphosphoglycerate on the binding of oxygen by hemoglobin.
38. **Recall** How does the oxygen-binding curve of fetal hemoglobin differ from that of adult hemoglobin?
39. **Recall** What is the critical amino acid difference between the  $\beta$ -chain and the  $\gamma$ -chain of hemoglobin?
40. **Reflect and Apply** In oxygenated hemoglobin,  $pK_a = 6.6$  for the histidines at position 146 on the  $\beta$ -chain. In deoxygenated hemoglobin, the  $pK_a$  of these residues is 8.2. How can this piece of information be correlated with the Bohr effect?
41. **Reflect and Apply** You are studying with a friend who is describing the Bohr effect. She tells you that in the lungs, hemoglobin binds oxygen and releases hydrogen ion; as a result, the pH increases. She goes on to say that in actively metabolizing muscle tissue, hemoglobin releases oxygen and binds hydrogen ion and, as a result, the pH decreases. Do you agree with her reasoning? Why or why not?
42. **Reflect and Apply** How does the difference between the  $\beta$ -chain and the  $\gamma$ -chain of hemoglobin explain the differences in oxygen binding between Hb A and Hb F?
43. **Reflect and Apply** Suggest a reason for the observation that people with sickle-cell trait sometimes have breathing problems during high-altitude flights.
44. **Reflect and Apply** Does a fetus homozygous for sickle-cell hemoglobin (HbS) have normal Hb F?
45. **Reflect and Apply** Why is fetal Hb essential for the survival of placental animals?

46. **Reflect and Apply** Why might you expect to find some Hb F in adults who are afflicted with sickle-cell anemia?
47. **Reflect and Apply** When deoxyhemoglobin was first isolated in crystalline form, the researcher who did so noted that the crystals changed color from purple to red and also changed shape as he observed them under a microscope. What is happening on the molecular level? *Hint:* The crystals were mounted on a microscope slide with a *loosely* fitting cover slip.

#### 4.6 Protein Folding Dynamics

48. **Reflect and Apply** You have discovered a new protein, one whose sequence has about 25% homology with ribonuclease A. How would you go about predicting, rather than experimentally determining, its tertiary structure?
49. **Reflect and Apply** Comment on the energetics of protein folding in light of the information in this chapter.
50. **Reflect and Apply** Go to the RCSB site for the Protein Data Bank (<http://www.rcsb.org/pdb>). Give a brief description of the molecule prefoldin, which can be found under *chaperones*.
51. **Recall** What is a chaperone?
52. **Biochemical Connections** What is a prion?
53. **Biochemical Connections** What are the known diseases caused by abnormal prions?
54. **Biochemical Connections** What are the protein secondary structures that differ between a normal prion and an infectious one?
55. **Recall** What are some diseases caused by misfolded proteins?
56. **Recall** What causes protein aggregates to form?
57. **Reflect and Apply** What other possible organizations of the globin gene could exist if there were no need for a globin chaperone?
58. **Biochemical Connections** What is the nature of the prion mutation that leads to extreme sensitivity to prion disease?
59. **Biochemical Connections** What is the most significant difference between prion diseases and other diseases caused by amyloid type plaques, such as Alzheimer's?
60. **Biochemical Connections** What aspects of the transmission of scrapie or other spongiform encephalopathies act like genetic diseases? What aspects act like transmissible diseases?
61. **Biochemical Connections** What enzymes are known to be involved in Alzheimer's disease?
62. **Biochemical Connections** What proteins are involved in the formation of the destructive plaques found in Alzheimer's disease?
63. **Biochemical Connections** Describe what happens according to the "amyloid cascade hypothesis."
64. **Biochemical Connections** Why would doctors not want to just completely inhibit  $\beta$ -secretase in a patient with Alzheimer's disease?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





# Protein Purification and Characterization Techniques

## 5

Column chromatography is widely used to purify proteins.

## 5.1 Extracting Pure Proteins from Cells

Many different proteins exist in a single cell. A detailed study of the properties of any one protein requires a homogeneous sample consisting of only one kind of molecule. The separation and isolation, or purification, of proteins constitutes an essential first step to further experimentation. In general, separation techniques focus on size, charge, and polarity—the sources of differences between molecules. Many techniques are performed to eliminate contaminants and to arrive at a pure sample of the protein of interest. As the purification steps are followed, we make a table of the recovery and purity of the protein to gauge our success. Table 5.1 shows a typical purification for an enzyme. The **percent recovery** column tracks how much of the protein of interest has been retained at each step. This number usually drops steadily during the purification, and we hope that by the time the protein is pure, sufficient product will be left for study and characterization. The specific activity column compares the purity of the protein at each step, and this value should go up if the purification is successful.

### How do we get the proteins out of the cells?

Before the real purification steps can begin, the protein must be released from the cells and subcellular organelles. The first step, called **homogenization**, involves breaking open the cells. This can be done with a wide variety of techniques. The simplest approach is grinding the tissue in a blender with a suitable buffer. The cells are broken open, releasing soluble proteins. This process also breaks many of the subcellular organelles, such as mitochondria, peroxisomes, and endoplasmic reticulum. A gentler technique is to use a Potter–Elvehjem homogenizer, a thick-walled test tube through which a tight-fitting plunger is passed. The squeezing of the homogenate around the plunger breaks open cells, but it leaves many of the organelles intact. Another technique, called sonication, involves using sound waves to break

## Chapter Outline

### 5.1 Extracting Pure Proteins from Cells

- How do we get the proteins out of the cells?

### 5.2 Column Chromatography

- What are the different types of chromatography?

### 5.3 Electrophoresis

- What is the difference between agarose gels and polyacrylamide gels?

### 5.4 Determining the Primary Structure of a Protein

- Why are the proteins cleaved into small fragments for protein sequencing?

**OWL** Online homework for this chapter may be assigned in OWL.

TABLE 5.1

Example of a Protein Purification Scheme:  
Purification of the Enzyme Xanthine Dehydrogenase from a Fungus

Fraction	Volume (mL)	Total Protein (mg)	Total Activity	Specific Activity	Percent Recovery
1. Crude extract	3,800	22,800	2,460	0.108	100
2. Salt precipitate	165	2,800	1,190	0.425	48
3. Ion-exchange chromatography	65	100	720	7.2	29
4. Molecular-sieve chromatography	40	14.5	555	38.3	23
5. Immunoaffinity chromatography	6	1.8	275	152.108	11

open the cells. Cells can also be ruptured by cycles of freezing and thawing. If the protein of interest is solidly attached to a membrane, detergents may have to be added to detach the proteins. After the cells are homogenized, they are subjected to **differential centrifugation**.

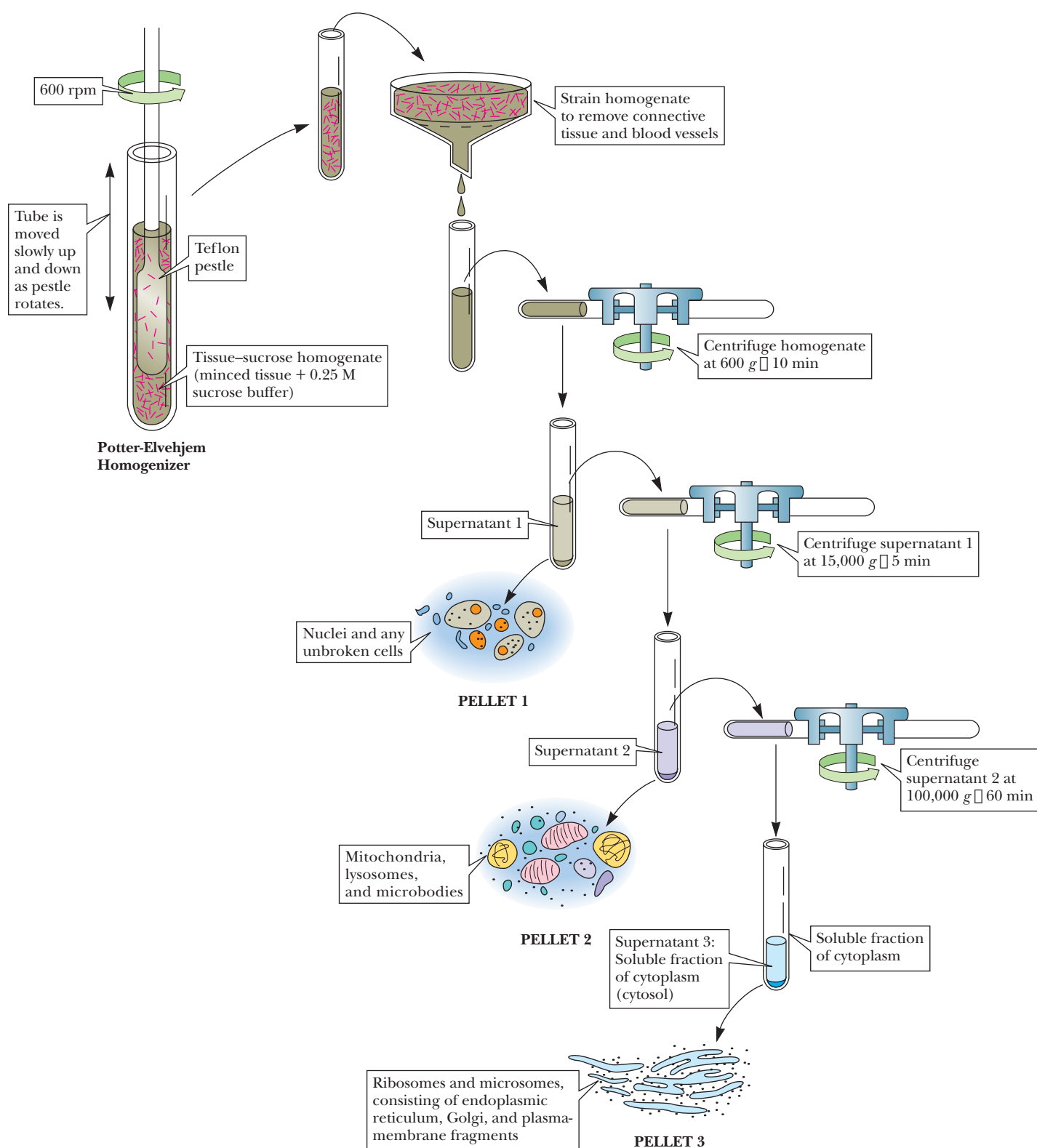
Spinning the sample at 600 times the force of gravity ( $600 \times g$ ) results in a pellet of unbroken cells and nuclei. If the protein of interest is not found in the nuclei, this precipitate is discarded. The supernatant can then be centrifuged at higher speed, such as  $15,000 \times g$ , to bring down the mitochondria. Further centrifugation at  $100,000 \times g$  brings down the microsomal fraction, consisting of ribosomes and membrane fragments. If the protein of interest is soluble, the supernatant from this spin will be collected and will already be partially purified because the nuclei and mitochondria will have been removed. Figure 5.1 shows a typical separation via differential centrifugation.

After the proteins are solubilized, they are often subjected to a crude purification based on solubility. Ammonium sulfate is the most common reagent to use at this step, and this procedure is referred to as **salting out**. Proteins have varying solubilities in polar and ionic compounds. Proteins remain soluble because of their interactions with water. When ammonium sulfate is added to a protein solution, some of the water is taken away from the protein to make ion–dipole bonds with the salts. With less water available to hydrate the proteins, they begin to interact with each other through hydrophobic bonds. At a defined amount of ammonium sulfate, a precipitate that contains contaminating proteins forms. These proteins are centrifuged down and discarded. Then more salt is added, and a different set of proteins, which usually contains the protein of interest, precipitates. This precipitate is collected by centrifugation and saved. The quantity of ammonium sulfate is usually measured in comparison with a 100% saturated solution. A common procedure involves bringing the solution to around 40% saturation and then spinning down the precipitate that forms. Next, more ammonium sulfate is added to the supernatant, often to a level of 60%–70% saturation. The precipitate that forms often contains the protein of interest. These preliminary techniques do not generally give a sample that is very pure, but they serve the important task of preparing the crude homogenate for the more effective procedures that follow.

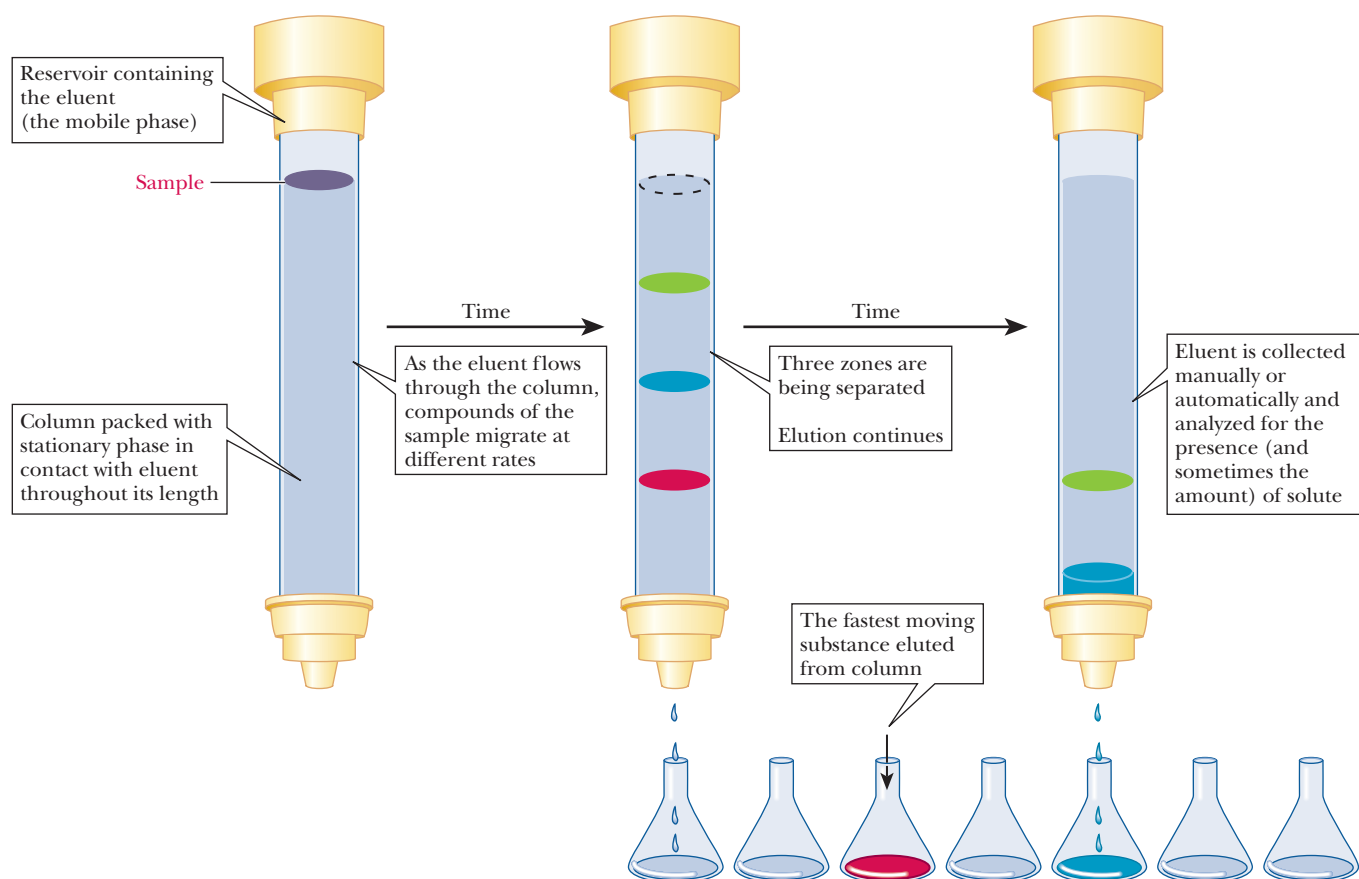
## 5.2 Column Chromatography

The word *chromatography* comes from the Greek *chroma*, “color,” and *graphein*, “to write”; the technique was first used around the beginning of the 20th century to separate plant pigments with easily visible colors. It has long since been possible to separate colorless compounds, as long as methods exist for detecting them. Chromatography is based on the fact that different compounds can distribute themselves to varying extents between different phases, or separable portions of matter. One phase is the **stationary phase**, and the other is the **mobile phase**. The mobile phase flows over the stationary material and carries the sample to be separated along with it. The components of the sample interact with the stationary phase to different extents. Some components interact relatively strongly with the stationary phase and are therefore carried along more slowly by the mobile phase than are those that interact less strongly. The differing mobilities of the components are the basis of the separation.

Many chromatographic techniques used for research on proteins are forms of **column chromatography**, in which the material that makes up the stationary phase is packed in a column. The sample is a small volume of concentrated solution that is applied to the top of the column; the mobile phase, called the *eluent*, is passed through the column. The sample is diluted by the eluent, and the separation process also increases the volume occupied by the sample.



■ **FIGURE 5.1 Differential centrifugation.** Differential centrifugation is used to separate cell components. As a cell homogenate is subjected to increasing  $g$  forces, smaller and smaller cell components end up in the pellet.

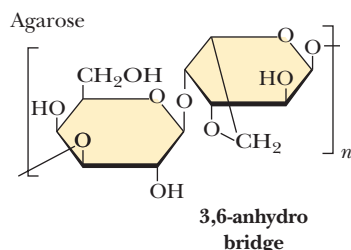


■ **FIGURE 5.2** Column chromatography. A sample containing several components is applied to the column. The various components travel at different rates and can be collected individually.

In a successful experiment, the entire sample eventually comes off the column. Figure 5.2 diagrams an example of column chromatography.

### What are the different types of chromatography?

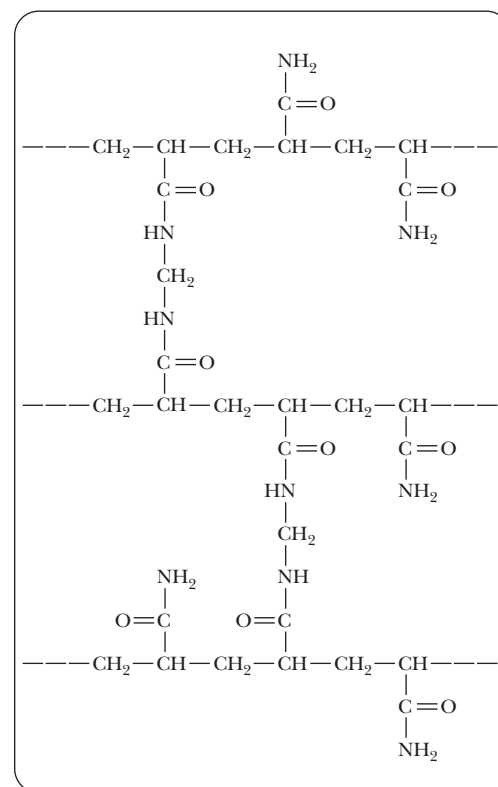
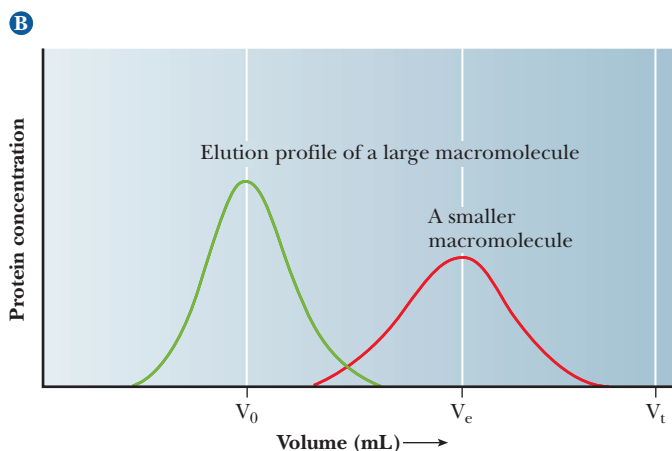
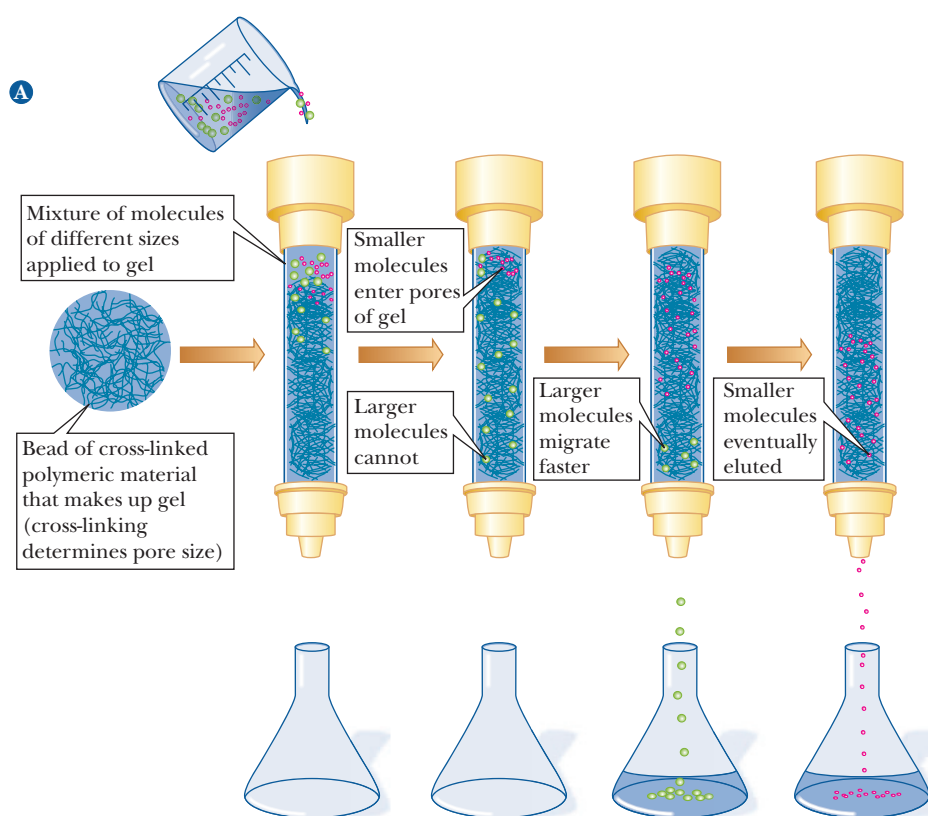
**Size-exclusion chromatography**, also called **gel-filtration chromatography**, separates molecules on the basis of size, making it a useful way to sort proteins of varied molecular weights. It is a form of column chromatography in which the stationary phase consists of cross-linked gel particles. The gel particles are usually in bead form and consist of one of two kinds of polymers. The first is a carbohydrate polymer, such as **dextran** or **agarose**; these two polymers are often referred to by the trade names Sephadex and Sepharose, respectively (Figure 5.3). The second is based on **polyacrylamide** (Figure 5.4), which is sold under the trade name Bio-Gel. The cross-linked structure of these polymers produces pores in the material. The extent of cross-linking can be controlled to select a desired pore size. When a sample is applied to the column, smaller molecules, which are able to enter the pores, tend to be delayed in their progress down the column, unlike the larger molecules. As a result, the larger molecules are eluted first, followed later by the smaller ones, after escaping from the pores. Molecular-sieve chromatography is represented schematically in Figure 5.5. The advantages of this type of chromatography are (1) its convenience as a way to separate molecules on the basis of size and (2) the fact that it can be used to estimate molecular weight by comparing the sample with a set of standards. Each type of gel used has a specific range of sizes that separate linearly with the log of the molecular weight. Each gel also has an exclusion limit, a size of protein that is too large to fit inside the pores. All proteins that size or larger elute first and simultaneously.



■ **FIGURE 5.3** The repeating disaccharide unit of agarose that is used for column chromatography.

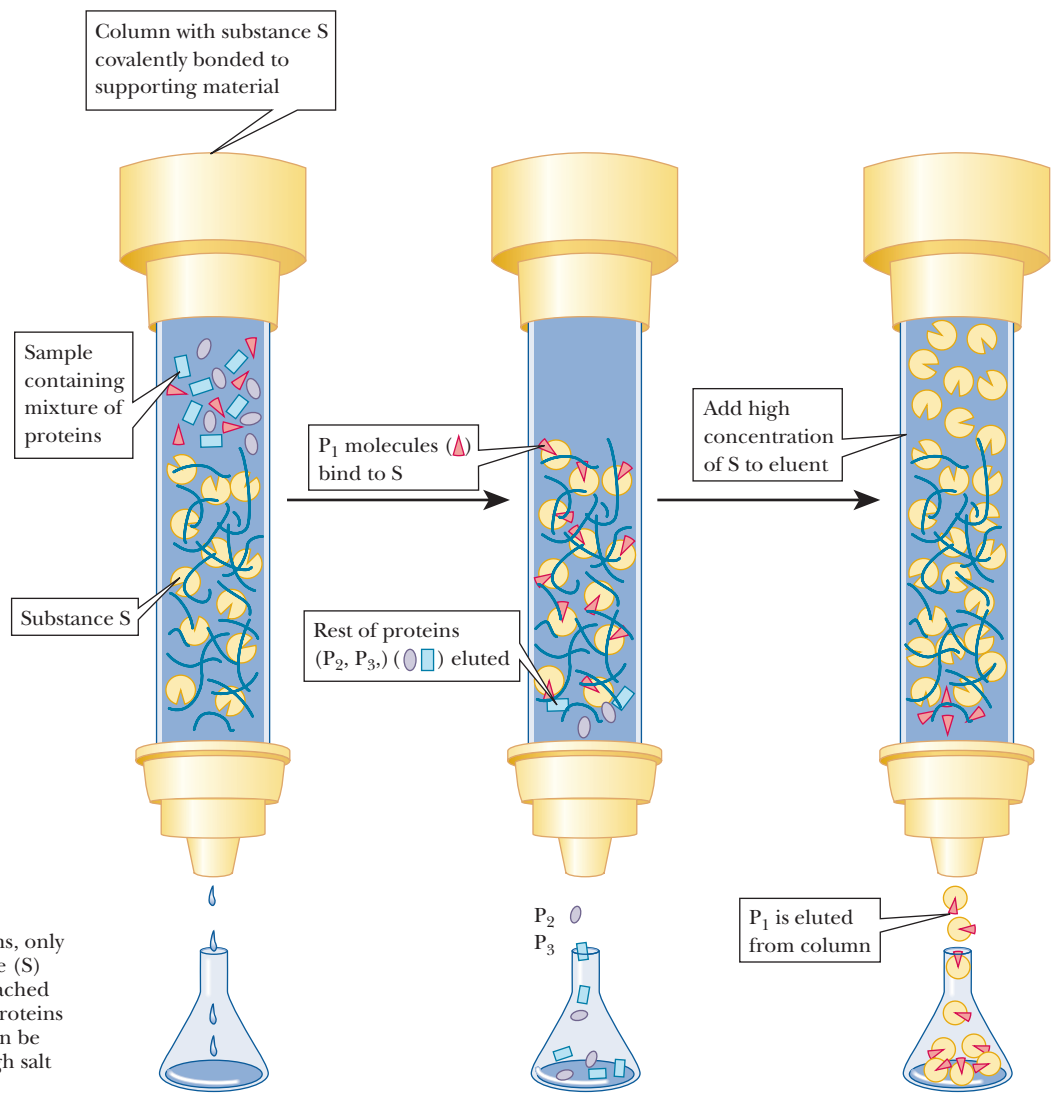


**Affinity chromatography** uses the specific binding properties of many proteins. It is another form of column chromatography with a polymeric material used as the stationary phase. The distinguishing feature of affinity chromatography is that the polymer is covalently linked to some compound, called a *ligand*, that binds specifically to the desired protein (Figure 5.6). The other proteins in the sample do not bind to the column and can easily be eluted with buffer, while the bound protein remains on the column. The bound protein can then be eluted from the column by adding high concentrations of the ligand in soluble form, thus competing for the binding of the protein with the stationary phase. The protein binds to the ligand in the mobile phase and is recovered from the column. This protein–ligand interaction can also be disrupted with a change in pH or ionic strength. Affinity chromatography is a convenient separation method and has the advantage of producing very pure proteins. Some affinity ligands are designed to be completely specific



**FIGURE 5.4** The structure of cross-linked polyacrylamide.

**FIGURE 5.5** Gel-filtration chromatography. Larger molecules are excluded from the gel and move more quickly through the column. Small molecules have access to the interior of the gel beads, so they take a longer time to elute. In column chromatography procedures such as this, the protein concentration is usually measured by UV absorption as the samples elute from the column.



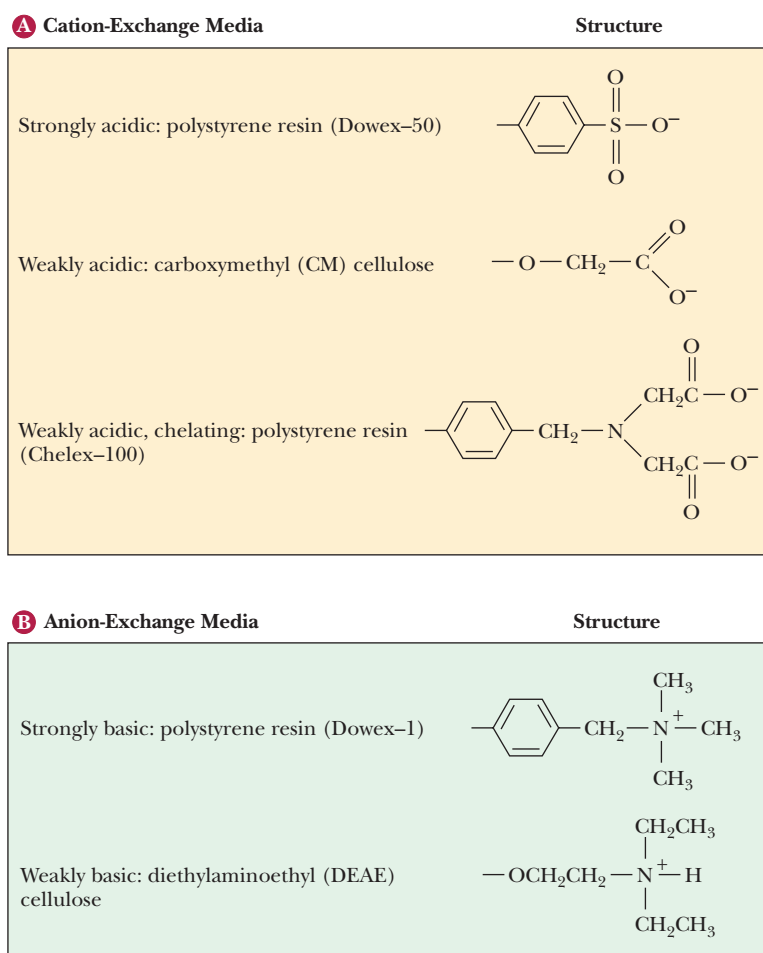
■ **FIGURE 5.6** The principle of affinity chromatography. In a mixture of proteins, only one (designated P<sub>1</sub>) binds to a substance (S) called the substrate. The substrate is attached to the column matrix. Once the other proteins (P<sub>2</sub> and P<sub>3</sub>) have been washed out, P<sub>1</sub> can be eluted either by adding a solution of high salt concentration or by adding free S.

**TABLE 5.2**

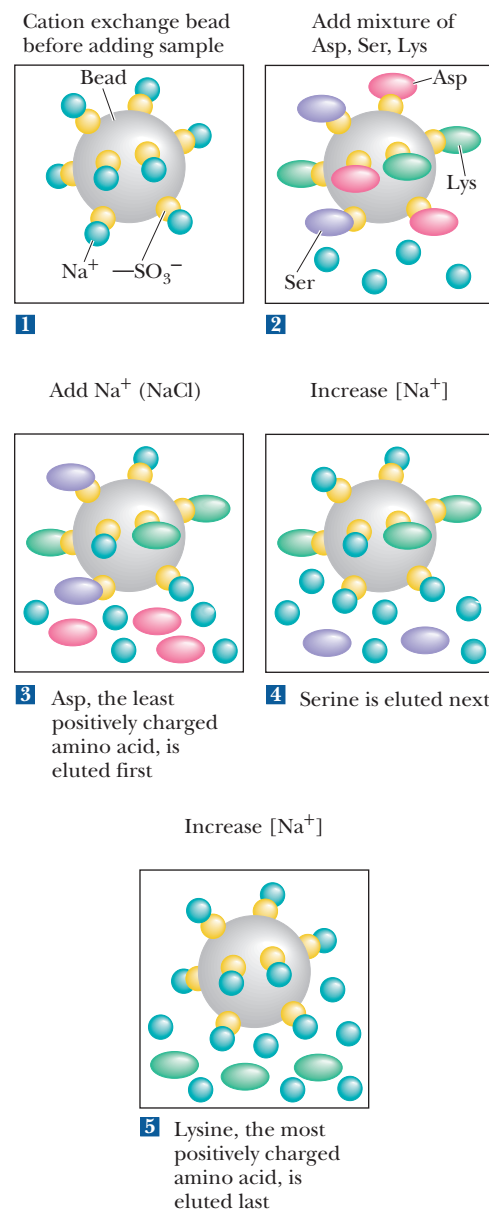
Group-Specific Affinity Resins	
Group-Specific Adsorbent	Group Specificity
Concanavalin A–agarose	Glycoproteins and glycolipids
Cibacron Blue–agarose	Enzymes with nucleotide cofactors
Boronic acid–agarose	Compounds with cis-diol groups
Protein A–agarose	IgG-type antibodies
Poly(U)–agarose	Nucleic acids containing poly(A) sequences
Poly(A)–agarose	Nucleic acids containing poly(U) sequences
Iminodiacetate–agarose	Proteins with heavy metal affinity
AMP–agarose	Enzymes with NAD <sup>+</sup> cofactors, ATP-dependent kinases

for a molecule the scientist is interested in purifying. However, this is often very expensive. There are other ligands that are specific for groups of compounds. Table 5.2 lists some group-specific affinity resins. The Biochemical Connections box in Chapter 13 describes an interesting way in which affinity chromatography can be combined with molecular biological techniques to offer a one-step purification of a protein.

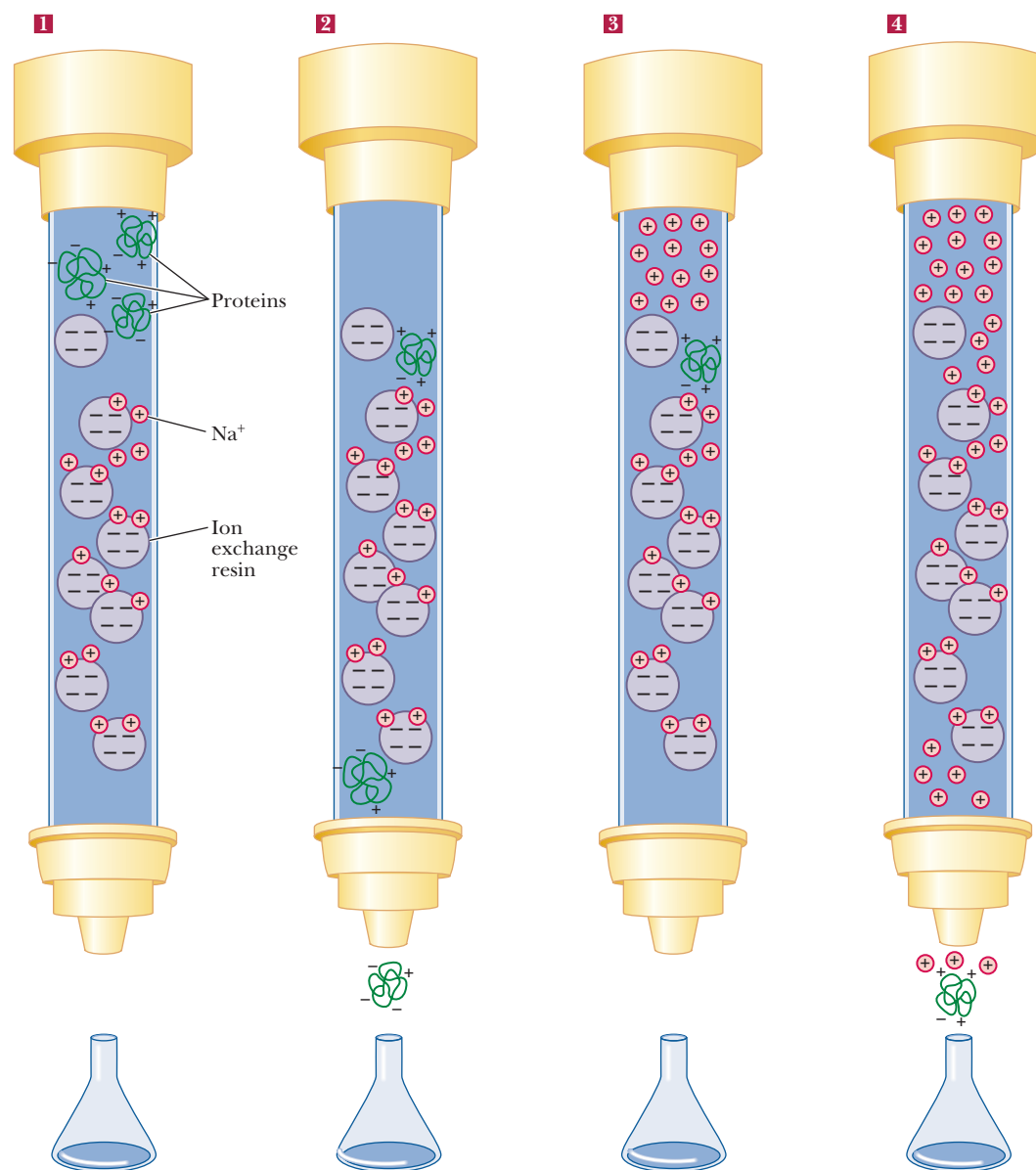
**Ion-exchange chromatography** is logistically similar to affinity chromatography. Both use a column resin that binds the protein of interest. With ion-exchange chromatography, however, the interaction is less specific and is based on net charge. An ion-exchange resin has a ligand with a positive charge or a negative charge. A negatively charged resin is a **cation exchanger**, and a positively charged one is an **anion exchanger**. Figure 5.7 shows some typical ion-exchange ligands. Figure 5.8 illustrates their principle of operation with three amino acids of different charge. Figure 5.9 shows how cation exchange chromatography would separate proteins. The column is initially equilibrated with a buffer of suitable pH and ionic strength. The exchange resin is bound to counterions. A cation-exchange resin is usually bound to  $\text{Na}^+$  or  $\text{K}^+$  ions, and an anion exchanger is usually bound to  $\text{Cl}^-$  ions. A mixture of proteins is loaded on the column and allowed to flow through it. Proteins that have a net charge opposite to that of the exchanger stick to the column, *exchanging* places with the bound counterions. Proteins that have no net charge or have the same charge as the exchanger elute. After all the nonbinding proteins are eluted, the eluent is changed either to a buffer that has a pH that removes the charge on the bound proteins or to one with a higher salt concentration. The latter outcompetes the bound proteins for the limited binding space on the column. The once-bound molecules then elute, having been separated from many of the contaminating ones.



■ **FIGURE 5.7** Resins used in ion-exchange chromatography. (a) Cation-exchange resins and (b) anion-exchange resins commonly used for biochemical separations.



■ **FIGURE 5.8** Operation of a cation-exchange column to separate a mixture of aspartate, serine, and lysine. (1) The cation-exchange resin in the initial  $\text{Na}^+$  form. (2) A mixture of aspartate, serine, and lysine is added to the column containing the resin. (3) A gradient of the eluting salt (for example,  $\text{NaCl}$ ) is added to the column. Aspartate, the least positively charged amino acid, is eluted first. (4) As the salt concentration increases, serine is eluted. (5) As the salt concentration is increased further, lysine, the most positively charged of the three amino acids, is eluted last.



■ **FIGURE 5.9** Ion-exchange chromatography using a cation exchanger. (1) At the beginning of the separation, various proteins are applied to the column. The column resin is bound to Na<sup>+</sup> counterions (small red spheres). (2) Proteins that have no net charge or a net negative charge pass through the column. Proteins that have a net positive charge stick to the column, displacing the Na<sup>+</sup>. (3) An excess of Na<sup>+</sup> ion is then added to the column. (4) The Na<sup>+</sup> ions outcompete the bound proteins for the binding sites on the resin, and the proteins elute.

**High Performance Liquid Chromatography (HPLC)** exploits the same principles seen with other chromatographic techniques, but very high resolution columns that can be run under high pressures are used. High resolution separations can be effected very quickly using automated instrumentation. A separation that might take hours on a standard column can be done in minutes with HPLC. **Reverse Phase HPLC** is a widely used technique for the separation of nonpolar molecules. In reverse phase HPLC, a solution of nonpolar compounds is put through a column that has a nonpolar liquid immobilized on an inert matrix. A more polar liquid serves as the mobile phase and is passed over the matrix. The solute molecules are eluted in proportion to their solubility in the more polar liquid.

## Apply Your Knowledge

### Protein Purification

The table below shows some typical results for a protein purification. The protein being purified is the enzyme Lactate Dehydrogenase, which catalyzes a reaction between lactic acid and  $\text{NAD}^+$  to give NADH and pyruvate. The purification of an enzyme is monitored by comparing the specific activity of the enzyme at various points in the purification. The specific activity is a measure of the activity of the enzyme divided by the weight of protein in the sample. The higher the number, the more pure the sample.

Purification Step	Total Activity of Enzyme ( $\mu\text{mol product/min}$ )	Specific Activity of Enzyme ( $\mu\text{mol product/min/mg protein}$ )
Crude Homogenate	100,000	0.15
20,000 $\times g$ supernatant	75,000	0.24
Salt precipitation	36,000	0.75
Ion Exchange Chromatography	12,000	3.4
Cibacron Blue Agarose Affinity Chromatography	6,000	42
Sephadex Gel Filtration Chromatography	500	90

Which step was the most effective at purifying the enzyme? Which step was the most costly in terms of enzyme recovery?

### Solution

If we divide the specific activity of any purification step (fraction) by the one before it, we will get what is called the “fold purification” for that step. The larger the number, the more effective that step was. Adding that calculation to the table gives the following results:

Purification Step	Total Activity of Enzyme ( $\mu\text{mol product/min}$ )	Specific Activity of Enzyme ( $\mu\text{mol product/min/mg protein}$ )	Fold Purification
Crude Homogenate	100,000	0.15	n/a
20,000 $\times g$ supernatant	75,000	0.24	1.6
Salt precipitation	36,000	0.75	3.1
Ion Exchange Chromatography	12,000	3.4	4.5
Cibacron Blue Agarose Affinity Chromatography	6,000	42	12.4
Sephadex Gel Filtration Chromatography	500	90	2.1

The results show that the affinity chromatography step gave the highest single purification of 12.4 for a single step. This is often the case due to the power of this technique.

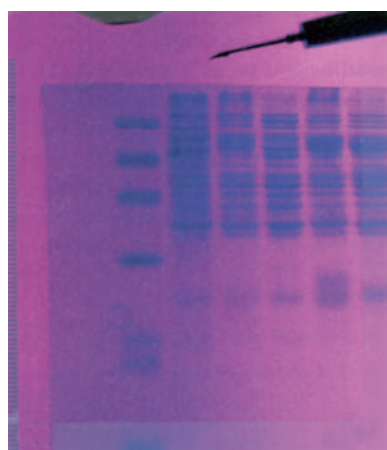
Similar calculations can show which step was the most costly by comparing the total activity of each fraction. The last step of the purification, the gel filtration chromatography, caused the loss of over 90% of the activity that was applied to it. This was the most costly single step. Scientists doing a purification must weigh the benefits and costs of each step in terms of purification and loss of product.





Gustoinimages/Photo Researchers, Inc.

■ **FIGURE 5.10** The experimental setup for gel electrophoresis. The samples are placed on the top of the gel. When the current is applied, the negatively charged proteins migrate towards the positive electrode at the bottom.



Michael Gabrielge/Visuals Unlimited

■ **FIGURE 5.11** Separation of proteins by gel electrophoresis. Each band seen in the gel represents a different protein. In the SDS-PAGE technique, the sample is treated with detergent before being applied to the gel. In isoelectric focusing, a pH gradient runs the length of the gel. The proteins in the gel have been stained with Coomassie Blue.

## 5.3 Electrophoresis

**Electrophoresis** is based on the motion of charged particles in an electric field toward an electrode of opposite charge. Macromolecules have differing mobilities based on their charge, shape, and size. Although many supporting media have been used for electrophoresis, including paper and liquid, the most common support is a polymer of agarose or acrylamide that is similar to those used for column chromatography. A sample is applied to wells that are formed in the supporting medium. An electric current is passed through the medium at a controlled voltage to achieve the desired separation (Figure 5.10). After the proteins are separated on the gel, the gel is stained to reveal the protein locations, as shown in Figure 5.11.

### What is the difference between agarose gels and polyacrylamide gels?

Agarose-based gels are most often used to separate nucleic acids and will be discussed in Chapter 13. For proteins, the most common electrophoretic support is polyacrylamide (Figure 5.4), although sometimes agarose is used. A polyacrylamide gel is prepared and cast as a continuous cross-linked matrix, rather than being produced in the bead form employed in column chromatography. In one variation of polyacrylamide-gel electrophoresis, the protein sample is treated with the detergent sodium dodecyl sulfate (SDS) before it is applied to the gel. The structure of SDS is  $\text{CH}_3(\text{CH}_2)_{10}\text{CH}_2\text{OSO}_3\text{Na}^+$ . The anion binds strongly to proteins via nonspecific adsorption. The larger the protein, the more of the anion it adsorbs. SDS completely denatures proteins, breaking all the noncovalent interactions that determine tertiary and quaternary structure. This means that multisubunit proteins can be analyzed as the component polypeptide chains. All the proteins in a sample have a negative charge as a result of adsorption of the anionic  $\text{SO}_3^-$ . The proteins also have roughly the same shape, which is a random coil. In **SDS-polyacrylamide-gel electrophoresis (SDS-PAGE)**, the acrylamide offers more resistance to large molecules than to small molecules. Because the shape and charge are approximately the same for all the proteins in the sample, the size of the protein becomes the determining factor in the separation: small proteins move faster than large ones. Like molecular-sieve chromatography, SDS-PAGE can be used to estimate the molecular weights of proteins by comparing the sample with standard samples. For most proteins, the log of the molecular weight is linearly related to its mobility on SDS-PAGE, as shown in Figure 5.12. Proteins can also be separated on acrylamide without the SDS, in which case the gel is called a **native gel**. This is useful for times when the study calls for a protein in its native conformation. In this case, however, the mobility is not correlated with size specifically, as three variables control the movement down the gel—size, shape, and charge.

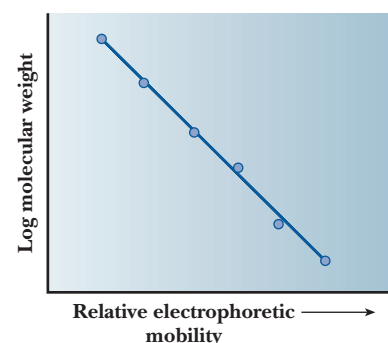
**Isoelectric focusing** is another variation of gel electrophoresis. Because different proteins have different titratable groups, they also have different isoelectric points. Recall (Section 3.3) that the isoelectric pH (pI) is the pH at which a protein (or amino acid or peptide) has no net charge. At the pI, the number of positive charges exactly balances the number of negative charges. In an isoelectric focusing experiment, the gel is prepared with a pH gradient that parallels the electric-field gradient. As proteins migrate through the gel under the influence of the electric field, they encounter regions of different pH, so the charge on the protein changes. Eventually each protein reaches the point at which it has no net charge—its isoelectric point—and no longer migrates. Each protein remains at the position on the gel corresponding to its pI, allowing for an effective method of separation.

An ingenious combination, known as two-dimensional gel electrophoresis (2-D gels), allows for enhanced separation by using isoelectric focusing in one dimension and SDS-PAGE run at 90° to the first (Figure 5.13).

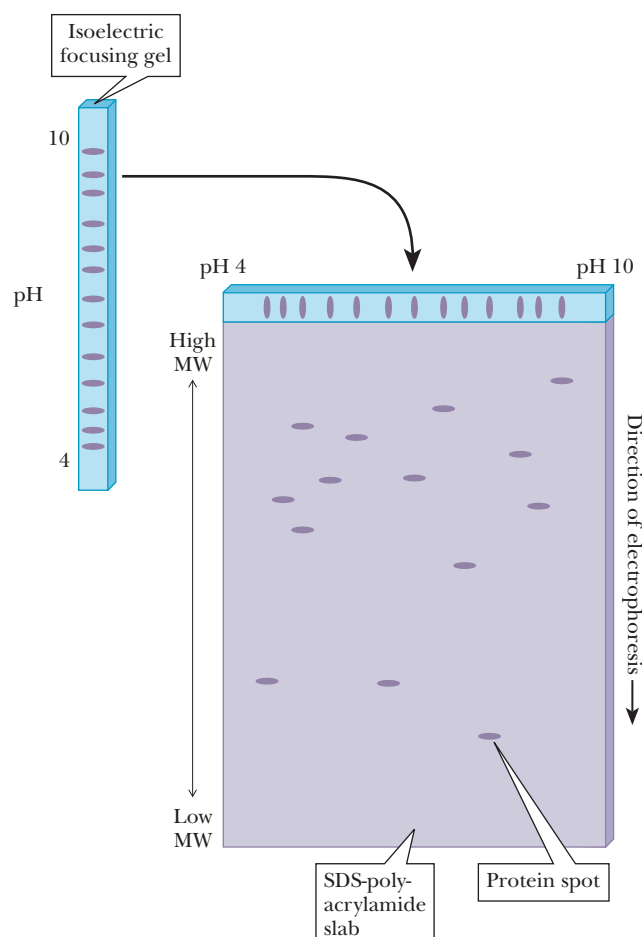
## 5.4 Determining the Primary Structure of a Protein

Determining the sequence of amino acids in a protein is a routine, but not trivial, operation in classical biochemistry. Its several parts must be carried out carefully to obtain accurate results (Figure 5.14).

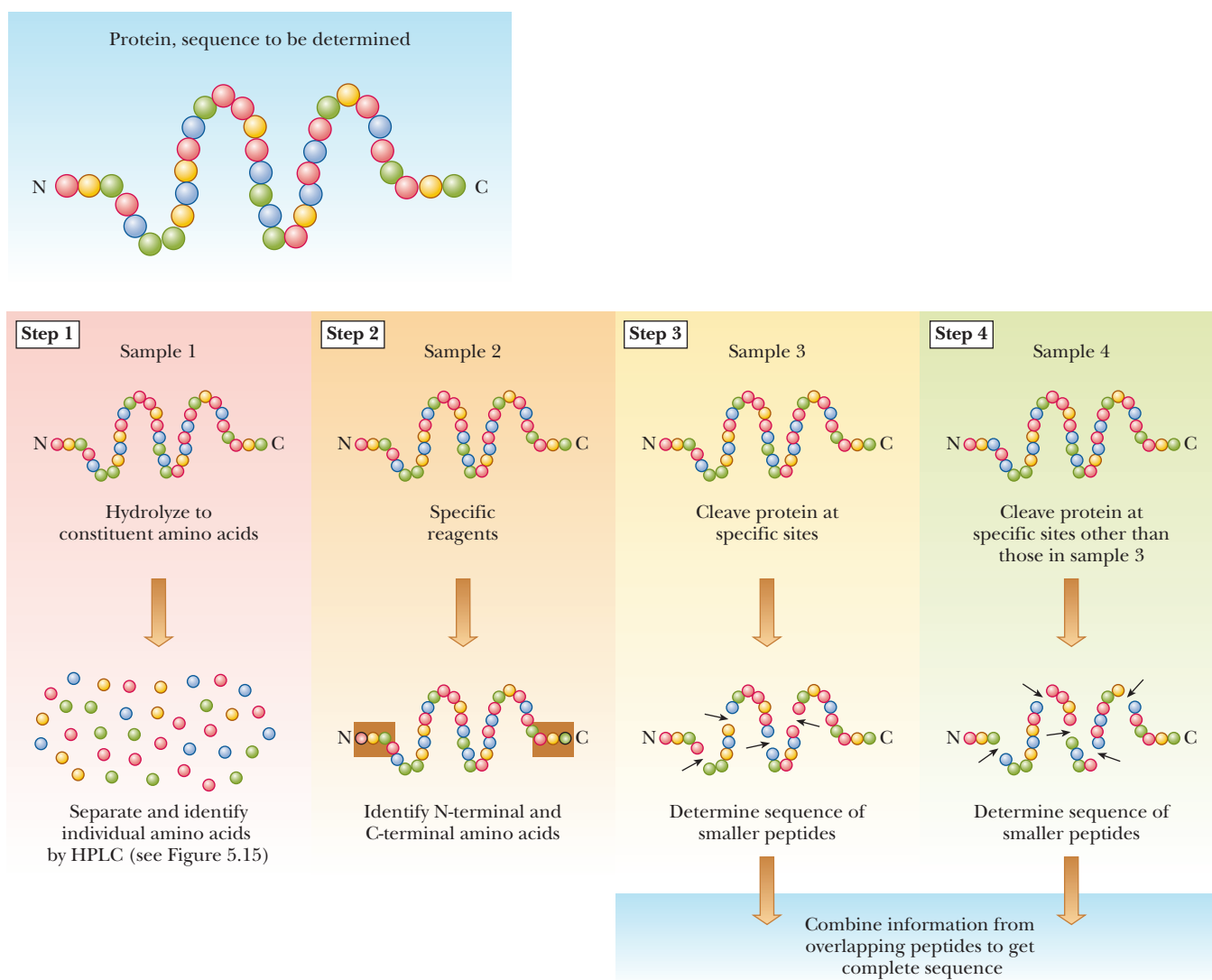
Step 1 in determining the primary structure of a protein is to establish which amino acids are present and in what proportions. Breaking a protein down to its component amino acids is relatively easy: heat a solution of the protein in acid, usually 6 M HCl, at 100°C to 110°C for 12 to 36 hours to hydrolyze the peptide bonds. Separation and identification of the products are somewhat more difficult and are best done by an amino acid analyzer. This automated instrument gives both qualitative information about the identities of the amino acids present and quantitative information about the relative amounts of those amino acids. Not only does it analyze amino acids, but it also allows informed decisions to be made about which procedures to choose later in the sequencing (see Steps 3 and 4 in Figure 5.14). An amino acid analyzer separates the mixture of amino acids either by ion-exchange chromatography or by **high-performance liquid chromatography (HPLC)**. Figure 5.15 shows a typical result of amino acid separation with this technique.



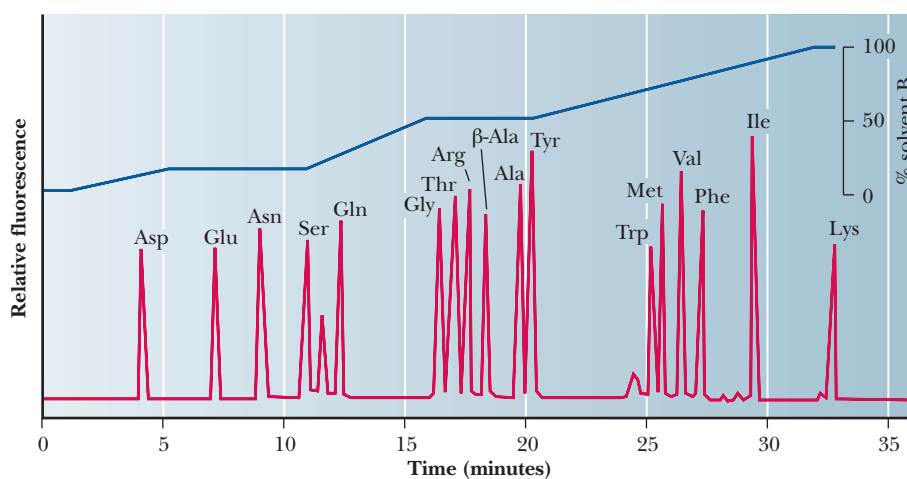
■ **FIGURE 5.12** Relationship between molecular weight and mobility. A plot of the relative electrophoretic mobility of proteins in SDS-PAGE versus the log of the molecular weights of the individual polypeptides approximates a straight line.



■ **FIGURE 5.13** Two-dimensional electrophoresis. A mixture of proteins is separated by isoelectric focusing in one direction. The focused proteins are then run using SDS-PAGE perpendicular to the direction of the isoelectric focusing. Thus the bands that appear on the gel have been separated first by their isoelectric points and then by size.



■ **FIGURE 5.14** The strategy for determining the primary structure of a given protein. The amino acid can be determined by four different analyses performed on four separate samples of the same protein.



■ **FIGURE 5.15** HPLC chromatogram of amino acid separation.

In Step 2, the identities of the N-terminal and C-terminal amino acids in a protein sequence are determined. This procedure is becoming less and less necessary as the sequencing of individual peptides improves, but it can be used to check whether a protein consists of one or two polypeptide chains.

In Steps 3 and 4, the protein is cleaved into smaller fragments, and the amino acid sequence is determined. Automated instruments can perform a stepwise modification starting from the N-terminal end, followed by cleavage of each amino acid in the sequence and the subsequent identification of each modified amino acid as it is removed. This process is called the **Edman degradation**.

### **Why are the proteins cleaved into small fragments for protein sequencing?**

The Edman degradation method becomes more difficult as the number of amino acids increases. In most proteins, the chain is more than 100 residues long. For sequencing, it is usually necessary to break a long polypeptide chain into fragments, ranging from 20 to 50 residues for reasons that will be explained later.

### **Cleavage of the Protein into Peptides**

Proteins can be cleaved at specific sites by enzymes or by chemical reagents. The enzyme **trypsin** cleaves peptide bonds preferentially at amino acids that have positively charged R groups, such as lysine and arginine. The cleavage takes place in such a way that the amino acid with the charged side chain ends up at the C-terminal end of one of the peptides produced by the reaction (Figure 5.16). The C-terminal amino acid of the original protein can be any one of the 20 amino acids and is not necessarily one at which cleavage takes place. A peptide can be automatically identified as the C-terminal end of the original chain if its C-terminal amino acid is not a site of cleavage.

Another enzyme, **chymotrypsin**, cleaves peptide bonds preferentially at the aromatic amino acids: tyrosine, tryptophan, and phenylalanine. The aromatic amino acid ends up at the C-terminal ends of the peptides produced by the reaction (Figure 5.17).

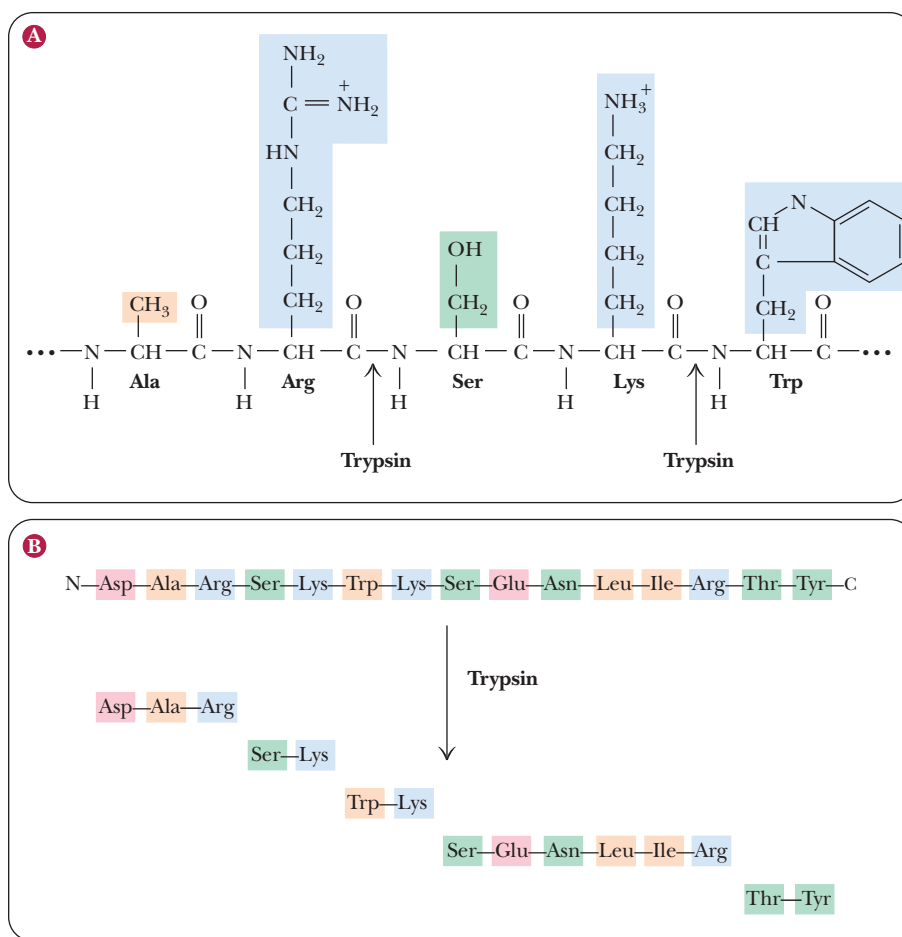
In the case of the chemical reagent **cyanogen bromide** (CNBr), the sites of cleavage are at internal methionine residues. The sulfur of the methionine reacts with the carbon of the cyanogen bromide to produce a homoserine lactone at the C-terminal end of the fragment (Figure 5.18).

The cleavage of a protein by any of these reagents produces a mixture of peptides, which are then separated by high-performance liquid chromatography. The use of several such reagents on different samples of a protein to be sequenced produces different mixtures. The sequences of a set of peptides produced by one reagent overlap the sequences produced by another reagent (Figure 5.19). As a result, the peptides can be arranged in the proper order after their own sequences have been determined.

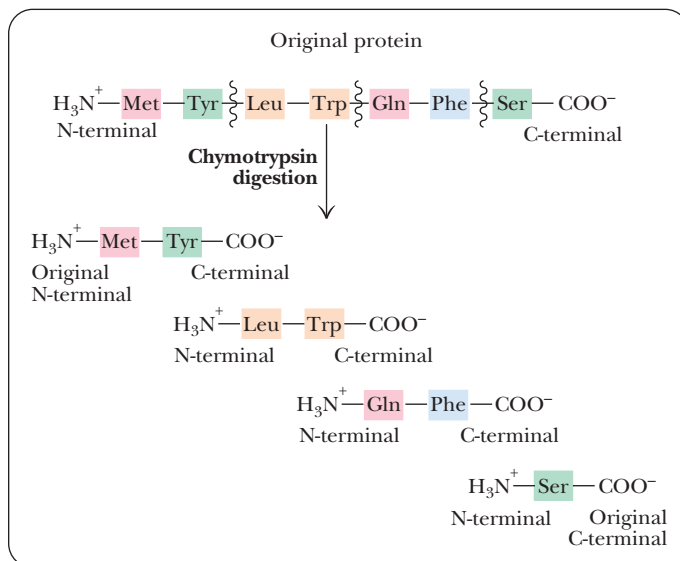
### **Sequencing of Peptides: The Edman Method**

The actual sequencing of each peptide produced by specific cleavage of a protein is accomplished by repeated application of the Edman degradation. The sequence of a peptide containing 10 to 40 residues can be determined by this method in about 30 minutes using as little as 10 picomoles of material, with the range being based on the amount of purified fragment and the complexity of the sequence. For example, proline is more difficult to sequence than serine because of its chemical reactivity. (The amino acid sequences of the

- **FIGURE 5.16** Peptide digestion with trypsin. (a) Trypsin is a proteolytic enzyme, or protease, that specifically cleaves only those peptide bonds in which arginine or lysine contributes the carbonyl function. (b) The products of the reaction are a mixture of peptide fragments with C-terminal Arg or Lys residues and a single peptide derived from the polypeptide's C-terminal end.

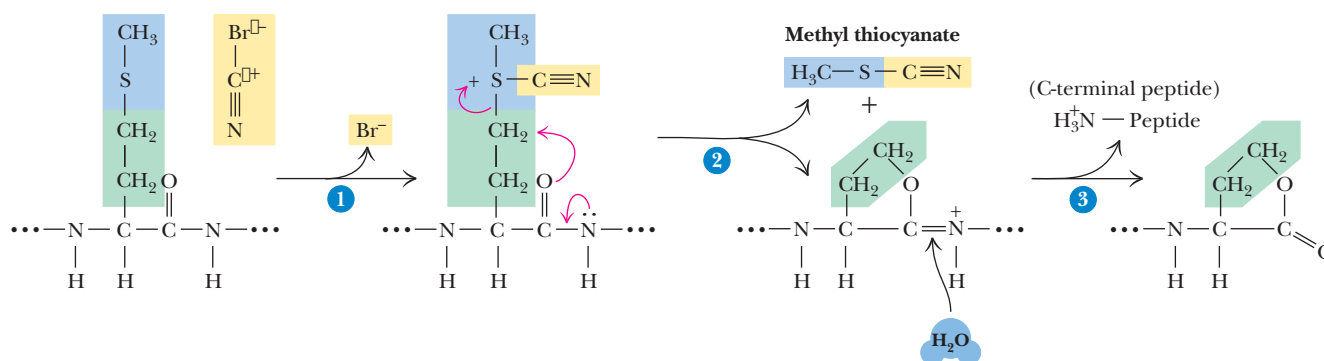


- **FIGURE 5.17** Cleavage of proteins by chymotrypsin. Chymotrypsin hydrolyzes proteins at aromatic amino acids.

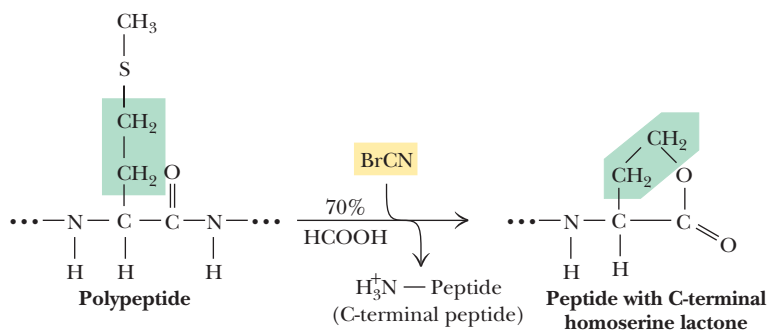


individual peptides in Figure 5.19 are determined by the Edman method after the peptides are separated from one another.) The overlapping sequences of peptides produced by different reagents provide the key to solving the puzzle.





OVERALL REACTION:

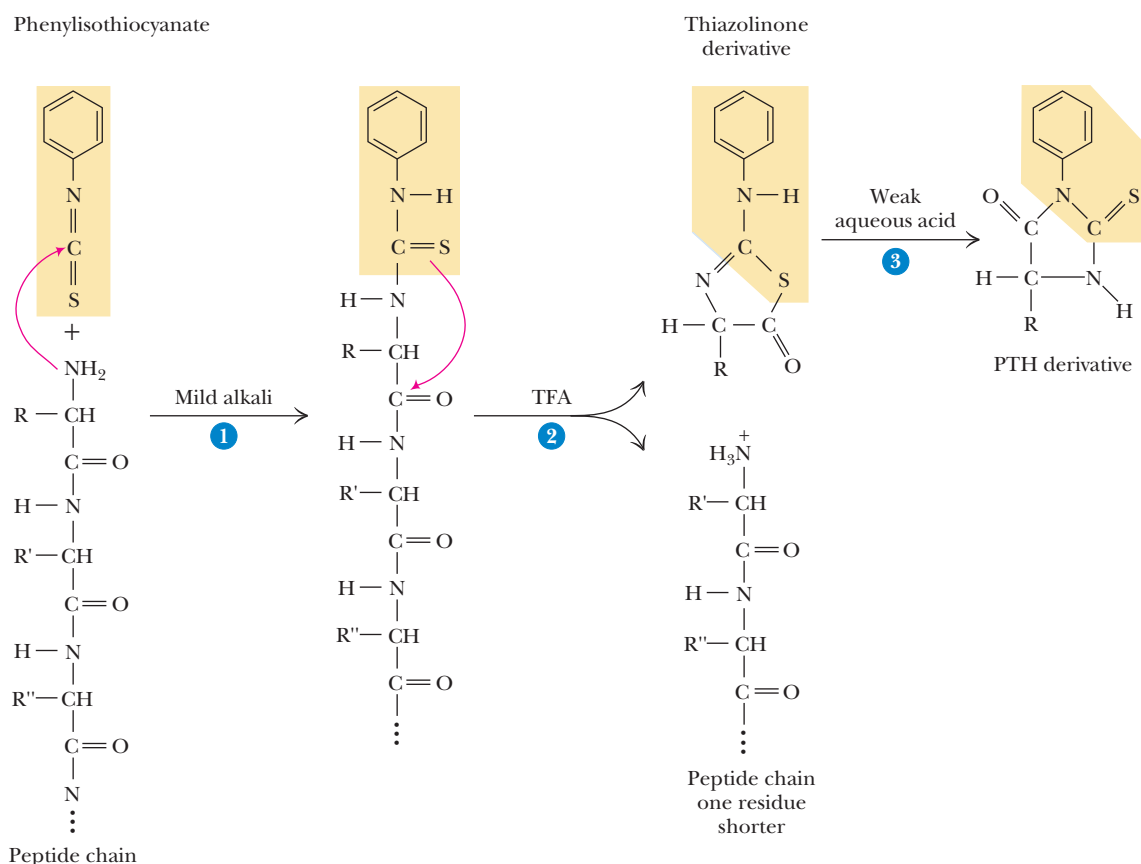


■ **FIGURE 5.18** Cleavage of proteins at internal methionine residues by cyanogen bromide.

Chymotrypsin	H <sub>3</sub> N <sup>+</sup> —Leu—Asn—Asp—Phe
Cyanogen bromide	H <sub>3</sub> N <sup>+</sup> —Leu—Asn—Asp—Phe—His—Met
Chymotrypsin	His—Met—Thr—Met—Ala—Trp
Cyanogen bromide	Thr—Met
Cyanogen bromide	Ala—Trp—Val—Lys—COO <sup>-</sup>
Chymotrypsin	Val—Lys—COO <sup>-</sup>
Overall sequence	H <sub>3</sub> N <sup>+</sup> —Leu—Asn—Asp—Phe—His—Met—Thr—Met—Ala—Trp—Val—Lys—COO <sup>-</sup>

■ **FIGURE 5.19** Use of overlapping sequences to determine protein sequence. Partial digestion was effected using chymotrypsin and cyanogen bromide. For clarity, only the original N-terminus and C-terminus of the complete peptide are shown.

The alignment of like sequences on different peptides makes deducing the overall sequence possible. The Edman method has become so efficient that it is no longer considered necessary to identify the N-terminal and C-terminal ends of a protein by chemical or enzymatic methods. While interpreting results, however, it is necessary to keep in mind that a protein may consist of more than one polypeptide chain.



■ **FIGURE 5.20 Sequencing of peptides by the Edman method.** (1) Phenylisothiocyanate combines with the N-terminus of a peptide under mildly alkaline conditions to form a phenylthiocarbamoyl substitution. (2) Upon treatment with TFA (trifluoroacetic acid), this cyclizes to release the N-terminal amino acid residue as a thiazolinone derivative, but the other peptide bonds are not hydrolyzed. (3) Organic extraction and treatment with aqueous acid yield the N-terminal amino acid as a phenylthiohydantoin (PTH) derivative. The process is repeated with the remainder of the peptide chain to determine the N-terminus exposed at each stage until the entire peptide is sequenced.

In the sequencing of a peptide, the Edman reagent, *phenyl isothiocyanate*, reacts with the peptide's N-terminal residue. The modified amino acid can be cleaved off, *leaving the rest of the peptide intact*, and can be detected as the phenylthiohydantoin derivative of the amino acid. The second amino acid of the original peptide can then be treated in the same way, as can the third. With an automated instrument called a **sequencer** (Figure 5.20), the process is repeated until the whole peptide is sequenced.

Another sequencing method uses the fact that the amino acid sequence of a protein reflects the base sequence of the DNA in the gene that coded for that protein. Using currently available methods, it is sometimes easier to obtain the sequence of the DNA than that of the protein. (See Section 13.11 for a discussion of sequencing methods for nucleic acids.) Using the genetic code (Section 12.2), one can immediately determine the amino acid sequence of the protein. Convenient though this method may be, it does not determine the positions of disulfide bonds or detect amino acids, such as hydroxyproline, that are modified after translation, nor does it take into account the extensive processing that occurs with eukaryotic genomes before the final protein is synthesized (Chapters 11 and 12).

### Apply Your Knowledge

#### Peptide Sequencing

A solution of a peptide of unknown sequence was divided into two samples. One sample was treated with trypsin, and the other was treated with chymotrypsin. The smaller peptides obtained by trypsin treatment had the following sequences:

Leu—Ser—Tyr—Ala—Ile—Arg  
**LSYAIR**

and

Asp—Gly—Met—Phe—Val—Lys  
**DGMFVK**

The smaller peptides obtained by chymotrypsin treatment had the following sequences:

Val—Lys—Leu—Ser—Tyr  
**VKLSY**

Ala—Ile—Arg  
**AIR**

and

Asp—Gly—Met—Phe  
**DGMF**

Deduce the sequence of the original peptide.

#### Solution

The key point here is that the fragments produced by treatment with the two different enzymes have overlapping sequences. These overlapping sequences can be compared to give the complete sequence. The results of the trypsin treatment indicate that there are two basic amino acids in the peptide, arginine and lysine. One of them must be the C-terminal amino acid, because no fragment was generated with a C-terminal amino acid other than these two. If there had been an amino acid other than a basic residue at the C-terminal position, trypsin treatment alone would have provided the sequence. Treatment with chymotrypsin gives the information needed. The sequence of the peptide Val—Lys—Leu—Ser—Tyr (VKLSY) indicates that lysine is an internal residue. The complete sequence is Asp—Gly—Met—Phe—Val—Lys—Leu—Ser—Tyr—Ala—Ile—Arg (DGMFVKLSYAIR).

To finish this section, let's go back to why we needed to cut the protein into pieces. Because the amino acid analyzer is giving us the sequence, it is easy to think that we could analyze a 100-amino-acid protein in one step with the analyzer and get the sequence without having to digest the protein with trypsin, chymotrypsin, or other chemicals. However, we must consider the logistical reality of doing the Edman degradation. As shown in step 1 of Figure 5.20, we react the peptide with the Edman reagent, phenylisothiocyanate (PITC). The stoichiometry of this reaction is that one molecule of the peptide reacts with one molecule of PITC. This yields one molecule of the PTH derivative

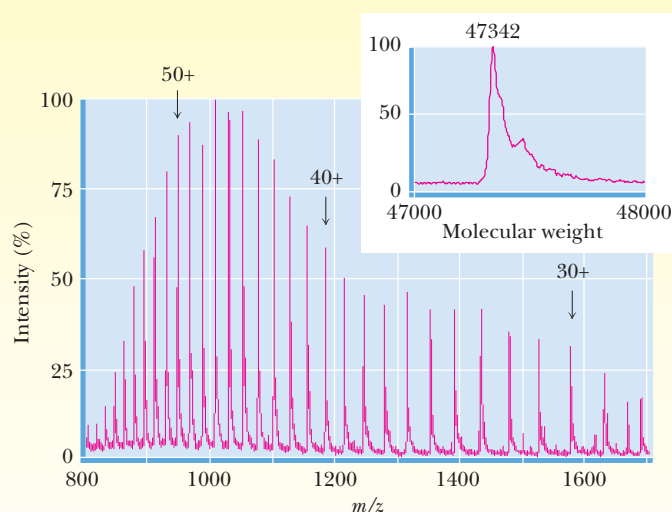
# Biochemical Connections INSTRUMENTATION

## The Power of Mass Spectrometry

While there are many techniques that allow a subtle approach to determining protein content and structure, none of them has the raw power of **mass spectrometry (MS)**. A mass spectrometer exploits the difference in the mass-to-charge ratio ( $m/z$ ) of ionized atoms or molecules in order to separate them from each other. The  $m/z$  ratio is such a characteristic property that it can be used to get structural and chemical information about the molecules and identify them.

When the charged particles are separated on the basis of their  $m/z$  ratio, they arrive at the detector at different times. The original application of MS in the early 20th century led to the discovery of isotopes. The isotopes of the noble gas argon were detected using apparatus that would seem quite simple now. For many years, the detection methods were based on having the substance to be analyzed in gaseous form or one that was easy to volatilize. As time went on, MS methods were developed that allow molecules as large as proteins, which are usually thought of as nonvolatile, to be analyzed.

One common type of MS is **electrospray Ionization (ESI-MS)**. A solution of macromolecules is sprayed in the form of droplets from a capillary under a strong electric field. The droplets pick up positive charges as they leave the capillary. Evaporation of the solvent leaves multiply charged molecules. A typical 20 k-Da protein will pick up 10 to 30 positive charges. The MS spectrum of this protein reveals all of the differently charged species as a series of sharp peaks whose consecutive  $m/z$  values differ by the charge and mass of a single proton, as shown to the right: Decreasing  $m/z$  values indicate an increasing number of charges per molecule. **Tandem Mass Spectrometry** uses another spectrometer downstream from the ESI source that can analyze complex protein mixtures, such as tryptic digests or proteins emerging from an HPLC column.



Another type of MS is called **Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF MS)**. A protein sample is mixed with a chemical matrix that includes a light-absorbing substance. A laser pulse is used to excite the chemical matrix, creating a microplasma that transfers the energy to protein molecules in the sample, ionizing them and ejecting them into the gas phase. Among the products are protein molecules that have picked up a single proton. These positively charged species can be selected by the MS for mass analysis. MALDI-TOF MS is very sensitive and very accurate. Attomole ( $10^{-18}$ ) quantities of a molecule can be detected.

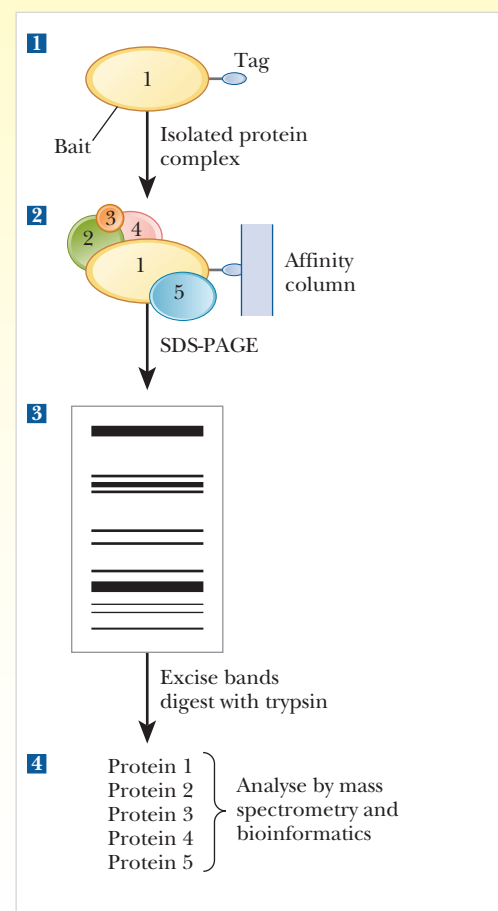
in step 3 that is then analyzed. Unfortunately, it is very difficult to get an exact stoichiometric match. For example, let's say we are analyzing a peptide with the sequence Asp—Leu—Tyr, etc. For simplicity, assume we add 100 molecules of the peptide to 98 molecules of the PITC because we cannot measure the quantities perfectly accurately. What happens then? In step 1, the PITC is limiting, so we eventually end up with 98 PTH derivatives of aspartate, which are analyzed correctly, and we know the N-terminus is aspartate. In the second round of the reaction, we add more PITC, but now there are two peptides; 98 of them begin with leucine and 2 of them begin with aspartate. When we analyze the PTH derivatives of round 2, we get two signals, one saying the derivative is leucine and the other saying aspartate. In round 2, the small amount of PTH derivative of aspartate does not interfere with our ability to recognize the true second amino acid. However, with every round, this situation gets worse and worse as more of the by-products show up. At some point, we get an analysis of the PTH derivatives that cannot be identified. For this reason, we have to start with smaller fragments so that we can analyze their sequences before the signal degrades.

# Biochemical Connections PROTEOMICS

## Pulling It All Together

The techniques introduced in this chapter are the backbone of the modern biologically based sciences, and they will be seen often throughout this book along with the information gleaned from their use. This could not be any more true than for the current trend known as **proteomics**. Proteomics is the systematic analysis of an organism's complete complement of proteins, its **proteome**, and it is one of the fastest-growing fields. Kumar et al. describe an elegant system involving three of the techniques we have seen to determine interactions between proteins in a cellular system. They created proteins they called "the bait," shown as protein 1 in the figure. These were tagged with an affinity label and allowed to react with the other cell components. The tagged bait proteins were then allowed to bind to an affinity column. In binding to the column, they took any other bound proteins with them. The bound complex was eluted from the column, then purified with SDS-PAGE. The bands were excised and digested with trypsin. After digestion, the pieces were identified with mass spectrometry. In this way the identities of the proteins associated with the bait protein were established. In the course of this book you will see many examples of interactions of proteins. This example demonstrates one of the ways that such information is gathered.

■ **Analyzing protein interactions.** In the method shown, an affinity tag is first attached to a target protein (the bait) in step (1) and allowed to react with other cell proteins. In step (2) the bait protein is bound to an affinity column. Any proteins that were interacting with the bait protein bind as well. In step (3), these proteins are purified using SDS-PAGE. In step (4), the proteins are excised from the gel and digested with trypsin, and the pieces are identified by mass spectrometry. *Redrawn by permission from Kumar, Anuj & Snyder, Michael, Proteomics: Protein complexes take the bait. Nature 415, fig 1, p123–124 (10 Jan 2002).*



## SUMMARY

**How do we get the proteins out of the cells?** Disruption of cells is the first step in protein purification. The various parts of cells can be separated by centrifugation. This is a useful step because proteins tend to occur in given organelles. High salt concentrations precipitate groups of proteins, which are then further separated by chromatography and electrophoresis.

**What are the different types of chromatography?** Gel-filtration chromatography separates proteins based on size. Ion-exchange chromatography separates proteins based on net charge. Affinity chromatography separates proteins based on their affinity for specific ligands. To purify a protein, many techniques are used and often several different chromatography steps are used.

**What is the difference between agarose gels and polyacrylamide gels?** Agarose gel electrophoresis is mainly used for separating nucleic acids, although it can also be used for native gel separation of proteins. Acrylamide is the usual medium for protein separation. When acrylamide gels are run with the chemical SDS, then the proteins separate based on size alone.

**Why are the proteins cleaved into small fragments for protein sequencing?** The Edman degradation has practical limits to how many amino acids can be cleaved from a protein and analyzed before the resulting data become confusing. To avoid this problem, the proteins are cut into small fragments using enzymes and chemicals, and these fragments are sequenced by the Edman degradation.



## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

### 5.1 Extracting Pure Proteins from Cells

- Recall** What types of homogenization techniques are available for solubilizing a protein?
- Recall** When would you choose to use a Potter–Elvehjem homogenizer instead of a blender?
- Recall** What is meant by “salting out”? How does it work?
- Recall** What differences between proteins are responsible for their differential solubility in ammonium sulfate?
- Recall** How could you isolate mitochondria from liver cells using differential centrifugation?
- Recall** Can you separate mitochondria from peroxisomes using only differential centrifugation?
- Recall** Give an example of a scenario in which you could partially isolate a protein with differential centrifugation using only one spin.
- Recall** Describe a procedure for isolating a protein that is strongly embedded in the mitochondrial membrane.
- Reflect and Apply** You are purifying a protein for the first time. You have solubilized it with homogenization in a blender followed by differential centrifugation. You wish to try ammonium sulfate precipitation as the next step. Knowing nothing beforehand about the amount of ammonium sulfate to add, design an experiment to find the proper concentration (% saturation) of ammonium sulfate to use.
- Reflect and Apply** If you had a protein X, which is a soluble enzyme found inside the peroxisome, and you wished to separate it from a similar protein Y, which is an enzyme found embedded in the mitochondrial membrane, what would be your initial techniques for isolating those proteins?

### 5.2 Column Chromatography

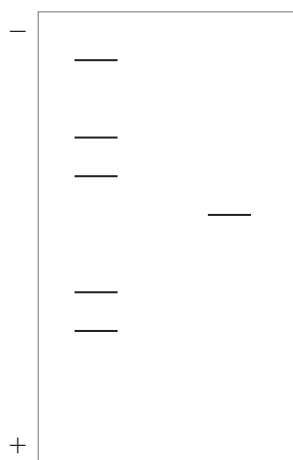
- Recall** What is the basis for the separation of proteins by the following techniques?
  - gel-filtration chromatography
  - affinity chromatography
  - ion-exchange chromatography
  - reverse phase HPLC
- Recall** What is the order of elution of proteins on a gel-filtration column? Why is this so?
- Recall** What are two ways that a compound can be eluted from an affinity column? What could be the advantages or disadvantages of each?
- Recall** What are two ways that a compound can be eluted from an ion-exchange column? What could be the advantages or disadvantages of each?
- Recall** Why do most people elute bound proteins from an ion-exchange column by raising the salt concentration instead of changing the pH?
- Recall** What are two types of compounds that make up the resin for column chromatography?
- Recall** Draw an example of a compound that would serve as a cation exchanger. Draw one for an anion exchanger.
- Recall** How can gel-filtration chromatography be used to arrive at an estimate of the molecular weight of a protein?
- Reflect and Apply** Sephadex G-75 has an exclusion limit of 80,000 molecular weight for globular proteins. If you tried to use this column material to separate alcohol dehydrogenase (MW 150,000) from  $\beta$ -amylase (MW 200,000), what would happen?
- Reflect and Apply** Referring to Question 19, could you separate  $\beta$ -amylase from bovine serum albumin (MW 66,000) using this column?

- Recall** What is the main difference between reverse phase HPLC and standard ion-exchange or gel filtration chromatography?
- Recall** How does HPLC differ from ion-exchange chromatography?
- Reflect and Apply** Design an experiment to purify protein X on an anion-exchange column. Protein X has an isoelectric point of 7.0.
- Reflect and Apply** Referring to Question 21, how would you purify protein X using ion-exchange chromatography if it turns out the protein is only stable at a pH between 6 and 6.5?
- Reflect and Apply** What could be an advantage of using an anion exchange column based on a quaternary amine [i.e., resin- $\text{N}^+(\text{CH}_2\text{CH}_3)_3$ ] as opposed to a tertiary amine [resin- $\text{NH}^+(\text{CH}_2\text{CH}_3)_2$ ]?
- Reflect and Apply** You wish to separate and purify enzyme A from contaminating enzymes B and C. Enzyme A is found in the matrix of the mitochondria. Enzyme B is embedded in the mitochondrial membrane, and enzyme C is found in the peroxisome. Enzymes A and B have molecular weights of 60,000 Da. Enzyme C has a molecular weight of 100,000 Da. Enzyme A has a pI of 6.5. Enzymes B and C have pI values of 7.5. Design an experiment to separate enzyme A from the other two enzymes.
- Reflect and Apply** An amino acid mixture consisting of lysine, leucine, and glutamic acid is to be separated by ion-exchange chromatography, using a cation-exchange resin at pH 3.5, with the eluting buffer at the same pH. Which of these amino acids will be eluted from the column first? Will any other treatment be needed to elute one of these amino acids from the column?
- Reflect and Apply** An amino acid mixture consisting of phenylalanine, glycine, and glutamic acid is to be separated by HPLC. The stationary phase is aqueous and the mobile phase is a solvent less polar than water. Which of these amino acids will move the fastest? Which one will move the slowest?
- Reflect and Apply** In reverse-phase HPLC, the stationary phase is nonpolar and the mobile phase is a polar solvent at neutral pH. Which of the three amino acids in Question 28 will move fastest on a reverse-phase HPLC column? Which one will move the slowest?
- Reflect and Apply** Gel-filtration chromatography is a useful method for removing salts, such as ammonium sulfate, from protein solutions. Describe how such a separation is accomplished.

### 5.3 Electrophoresis

- Recall** What physical parameters of a protein control its migration on electrophoresis?
- Recall** What types of compounds make up the gels used in electrophoresis?
- Recall** Of the two principal polymers used in column chromatography and electrophoresis, which one would be most immune to contamination by bacteria and other organisms?
- Recall** What types of macromolecules are usually separated on agarose electrophoresis gels?
- Recall** If you had a mixture of proteins with different sizes, shapes, and charges and you separated them with electrophoresis, which proteins would move fastest toward the anode (positive electrode)?
- Recall** What does SDS–PAGE stand for? What is the benefit of doing SDS–PAGE?
- Recall** How does the addition of sodium dodecylsulfate to proteins affect the basis of separation on electrophoresis?
- Recall** Why is the order of separation based on size opposite for gel filtration and gel electrophoresis, even though they often use the same compound to form the matrix?

39. **Recall** The accompanying figure is from an electrophoresis experiment using SDS-PAGE. The left lane has the following standards: bovine serum albumin (MW 66,000), ovalbumin (MW 45,000), glyceraldehyde 3-phosphate dehydrogenase (MW 36,000), carbonic anhydrase (MW 24,000), and trypsinogen (MW 20,000). The right lane is an unknown. Calculate the MW of the unknown.



### 5.4 Determining the Primary Structure of a Protein

40. **Recall** Why is it no longer considered necessary to determine the N-terminal amino acid of a protein as a separate step?
41. **Recall** What useful information might you get if you did determine the N-terminal amino acid as a separate step?
42. **Reflect and Apply** Show by a series of equations (with structures) the first stage of the Edman method applied to a peptide that has leucine as its N-terminal residue.
43. **Reflect and Apply** Why can the Edman degradation not be used effectively with very long peptides? *Hint:* Think about the stoichiometry of the peptides and the Edman reagent and the percent yield of the organic reactions involving them.
44. **Reflect and Apply** What would happen during an amino acid sequencing experiment using the Edman degradation if you accidentally added twice as much Edman reagent (on a per-mole basis) as the peptide you were sequencing?
45. **Reflect and Apply** A sample of an unknown peptide was divided into two aliquots. One aliquot was treated with trypsin; the other was treated with cyanogen bromide. Given the following sequences (N-terminal to C-terminal) of the resulting fragments, deduce the sequence of the original peptide.

#### Trypsin treatment

Asn—Thr—Trp—Met—Ile—Lys

Gly—Tyr—Met—Gln—Phe

Val—Leu—Gly—Met—Ser—Arg

#### Cyanogen bromide treatment

Gln—Phe

Val—Leu—Gly—Met

Ile—Lys—Gly—Tyr—Met

Ser—Arg—Asn—Thr—Trp—Met

46. **Reflect and Apply** A sample of a peptide of unknown sequence was treated with trypsin; another sample of the same peptide was treated with chymotrypsin. The sequences (N-terminal to C-terminal) of the smaller peptides produced by trypsin digestion were as follows:

Met—Val—Ser—Thr—Lys

Val—Ile—Trp—Thr—Leu—Met—Ile

Leu—Phe—Asn—Glu—Ser—Arg

The sequences of the smaller peptides produced by chymotrypsin digestion were as follows:

Asn—Glu—Ser—Arg—Val—Ile—Trp

Thr—Leu—Met—Ile

Met—Val—Ser—Thr—Lys—Leu—Phe

Deduce the sequence of the original peptide.

47. **Reflect and Apply** You are in the process of determining the amino acid sequence of a protein and must reconcile contradictory results. In one trial, you determine a sequence with glycine as the N-terminal amino acid and asparagine as the C-terminal amino acid. In another trial, your results indicate phenylalanine as the N-terminal amino acid and alanine as the C-terminal amino acid. How do you reconcile this apparent contradiction?
48. **Reflect and Apply** You are in the process of determining the amino acid sequence of a peptide. After trypsin digestion followed by the Edman degradation, you see the following peptide fragments:

Leu—Gly—Arg

Gly—Ser—Phe—Tyr—Asn—His

Ser—Glu—Asp—Met—Cys—Lys

Thr—Tyr—Glu—Val—Cys—Met—His

What is abnormal concerning these results? What might have been the problem that caused it?

49. **Reflect and Apply** Amino acid compositions can be determined by heating a protein in 6 M HCl and running the hydrolysate through an ion-exchange column. If you were going to do an amino acid sequencing experiment, why would you want to get an amino acid composition first?
50. **Reflect and Apply** Assume that you are getting ready to do an amino acid sequencing experiment on a protein containing 100 amino acids, and amino acid analysis shows the following data:

Amino Acid	Number of Residues
Ala	7
Arg	23.7
Asn	5.6
Asp	4.1
Cys	4.7
Gln	4.5
Glu	2.2
Gly	3.7
His	3.7
Ile	1.1
Leu	1.7
Lys	11.4
Met	0
Phe	2.4
Pro	4.5
Ser	8.2
Thr	4.7
Trp	0
Tyr	2.0
Val	5.1

Which of the chemicals or enzymes normally used for cutting proteins into fragments would be the least useful to you?

51. **Reflect and Apply** Which enzymes or chemicals would you choose to use to cut the protein from Question 50? Why?
52. **Reflect and Apply** With which amino acid sequences would chymotrypsin be an effective reagent for sequencing the protein from Question 50? Why?
53. **Biochemical Connections** What are the two principal types of mass spectrometry?
54. **Biochemical Connections** What is the advantage of MALDI-TOF MS?
55. **Biochemical Connections** What is proteomics?
56. **Biochemical Connections** What is the purpose of the tag on the bait protein described in the Biochemical Connections box on page 135?
57. **Biochemical Connections** What are some of the assumptions behind the logic of the experiment described in the Biochemical Connections box on page 135?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).

# The Behavior of Proteins: Enzymes

# 6

Traveling over a mountain pass is an analogy frequently used to describe the progress of a chemical reaction. Catalysts speed up the process.

## 6.1 Enzymes Are Effective Biological Catalysts

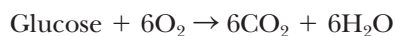
Of all the functions of proteins, **catalysis** is probably the most important. In the absence of catalysis, most reactions in biological systems would take place far too slowly to provide products at an adequate pace for a metabolizing organism. The catalysts that serve this function in organisms are called **enzymes**. With the exception of some RNAs (ribozymes) that have catalytic activity (described in Sections 11.7 and 12.4), all other enzymes are globular proteins (section 4.3). Enzymes are the most efficient catalysts known; they can increase the rate of a reaction by a factor of up to  $10^{20}$  over uncatalyzed reactions. Nonenzymatic catalysts, in contrast, typically enhance the rate of reaction by factors of  $10^2$  to  $10^4$ .

As we shall see in the next two chapters, enzymes are characterized by being highly specific, even to the point of being able to distinguish stereoisomers of a given compound, and by greatly increasing the speed of a reaction. In many cases, the actions of enzymes are fine-tuned by regulatory processes.

## 6.2 Kinetics versus Thermodynamics

The rate of a reaction and its thermodynamic favorability are two different topics, although they are closely related. This is true of all reactions, whether or not a catalyst is involved. The difference between the energies of the reactants (the initial state) and the energies of the products (the final state) of a reaction gives the energy change for that reaction, expressed as the **standard free energy change**, or  $\Delta G^\circ$ . Energy changes can be described by several related thermodynamic quantities. We shall use standard free energy changes for our discussion; the question of whether a reaction is favored depends on  $\Delta G^\circ$  (see Sections 1.9 and 15.2). Enzymes, like all catalysts, speed up reactions, but they cannot alter the equilibrium constant or the free energy change. The reaction rate depends on the free energy of activation or **activation energy** ( $\Delta G^\ddagger$ ), the energy input required to initiate the reaction. The activation energy for an uncatalyzed reaction is higher than that for a catalyzed reaction; in other words, an uncatalyzed reaction requires more energy to get started. For this reason, its rate is slower than that of a catalyzed reaction.

The reaction of glucose and oxygen gas to produce carbon dioxide and water is an example of a reaction that requires a number of enzymatic catalysts:



This reaction is thermodynamically favorable (spontaneous in the thermodynamic sense) because its free energy change is negative ( $\Delta G^\circ = -2880 \text{ kJ mol}^{-1} = -689 \text{ kcal mol}^{-1}$ ).

## Chapter Outline

### 6.1 Enzymes Are Effective Biological Catalysts

### 6.2 Kinetics versus Thermodynamics

- If a reaction is spontaneous, does that mean it will be fast?
- Will a reaction go faster if you raise the temperature?

### 6.3 Enzyme Kinetic Equations

- Is the rate of a reaction always based on the concentration of reactants?

### 6.4 Enzyme–Substrate Binding

- Why do enzymes bind to substrates?

### 6.5 Examples of Enzyme-Catalyzed Reactions


- Why do chymotrypsin and ATCase have different velocity curves?

### 6.6 The Michaelis–Menten Approach to Enzyme Kinetics

- How do we calculate  $K_M$  and  $V_{\max}$  from a graph?
- What is the significance of  $K_M$  and  $V_{\max}$ ?

### 6.7 Enzyme Inhibition

- How can we identify a competitive inhibitor?
- How can we identify a noncompetitive inhibitor?

 **OWL** Online homework for this chapter may be assigned in OWL.

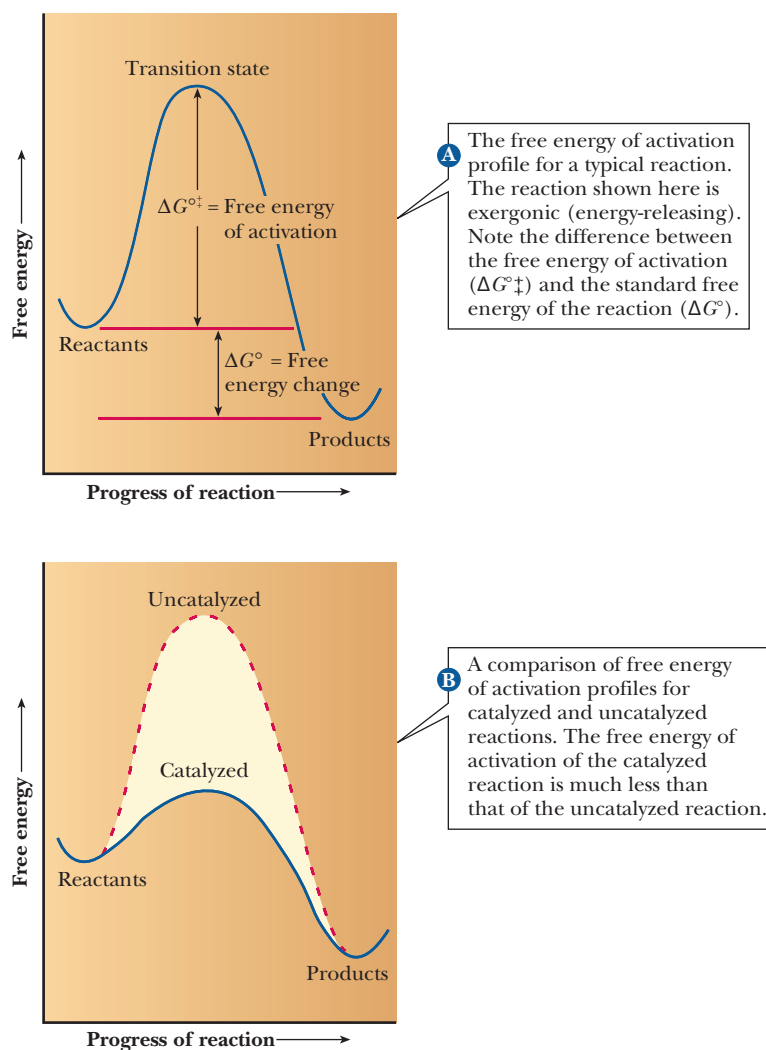
### If a reaction is spontaneous, does that mean it will be fast?

Note that the term *spontaneous* does not mean “instantaneous.” Glucose is stable in air with an unlimited supply of oxygen. The energy that must be supplied to start the reaction (which then proceeds with a release of energy)—the activation energy—is conceptually similar to the act of pushing an object to the top of a hill so that it can then slide down the other side.

Activation energy and its relationship to the free energy change of a reaction can best be shown graphically. In Figure 6.1a, the  $x$  coordinate shows the extent to which the reaction has taken place, and the  $y$  coordinate indicates free energy for an idealized reaction. The *activation energy profile* shows the intermediate stages of a reaction, those between the initial and final states. Activation energy profiles are essential in the discussion of catalysts. The activation energy directly affects the rate of reaction, and the presence of a catalyst speeds up a reaction by changing the mechanism and thus lowering the activation energy. Figure 6.1a plots the energies for an exergonic, spontaneous reaction, such as the complete oxidation of glucose. At the maximum of the curve connecting the reactants and the products lies the **transition state** with the necessary amount of energy and the correct arrangement of atoms to produce products. The activation energy can also be seen as the amount of free energy required to bring the reactants to the transition state.

The analogy of traveling over a mountain pass between two valleys is frequently used in discussions of activation energy profiles. The change in

■ **FIGURE 6.1** Free energy profiles.





energy corresponds to the change in elevation, and the progress of the reaction corresponds to the distance traveled. The analogue of the transition state is the top of the pass. Considerable effort has gone into elucidating the intermediate stages in reactions of interest to chemists and biochemists and determining the pathway or reaction mechanism that lies between the initial and final states. Reaction dynamics, the study of the intermediate stages of reaction mechanisms, is currently a very active field of research. In Chapter 7, we shall look at the use of molecules that mimic the transition state, called transition-state analogues, which are used to study the specific mechanisms of enzyme catalysis.

The most important effect of a catalyst on a chemical reaction is apparent from a comparison of the activation energy profiles of the same reaction, catalyzed and uncatalyzed, as shown in Figure 6.1b. The standard free energy change for the reaction,  $\Delta G^\circ$ , remains unchanged when a catalyst is added, but the activation energy,  $\Delta G^\ddagger$ , is lowered. In the hill-and-valley analogy, the catalyst is a guide that finds an easier path between the two valleys. A similar comparison can be made between two routes from San Francisco to Los Angeles. The highest point on Interstate 5 is Tejon Pass (elevation 4400 feet) and is analogous to the uncatalyzed path. The highest point on U.S. Highway 101 is not much over 1000 feet. Thus, Highway 101 is an easier route and is analogous to the catalyzed pathway. The initial and final points of the trip are the same, but the paths between them are different, as are the mechanisms of catalyzed and uncatalyzed reactions. The presence of an enzyme lowers the activation energy needed for substrate molecules to reach the transition state. The concentration of the transition state increases markedly. As a result, the rate of the catalyzed reaction is much greater than the rate of the uncatalyzed reaction. Enzymatic catalysts enhance a reaction rate by many powers of 10.

The biochemical reaction in which hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) is converted to water and oxygen provides an example of the effect of catalysts on activation energy.



The activation energy of this reaction is lowered if the reaction is allowed to proceed on platinum surfaces, but it is lowered even more by the enzyme catalase. Table 6.1 summarizes the energies involved.

### Will a reaction go faster if you raise the temperature?

Raising the temperature of a reaction mixture increases the energy available to the reactants to reach the transition state. Consequently, the rate of a chemical reaction increases with temperature. One might be tempted to assume that this is universally true for biochemical reactions. In fact, increase of reaction rate with temperature occurs only to a limited extent with biochemical reactions. It is helpful to raise the temperature at first, but eventually there comes a point at which heat denaturation of the enzyme (Section 4.4) is reached. Above this

**TABLE 6.1**

Lowering of the Activation Energy of Hydrogen Peroxide Decomposition by Catalysts			
Reaction Conditions	Activation Free Energy		Relative Rate
	$\text{kJmol}^{-1}$	$\text{kcal mol}^{-1}$	
No catalyst	75.2	18.0	1
Platinum surface	48.9	11.7	$2.77 \times 10^4$
Catalase	23.0	5.5	$6.51 \times 10^8$

Rates are given in arbitrary units relative to a value of 1 for the uncatalyzed reaction at 37°C.

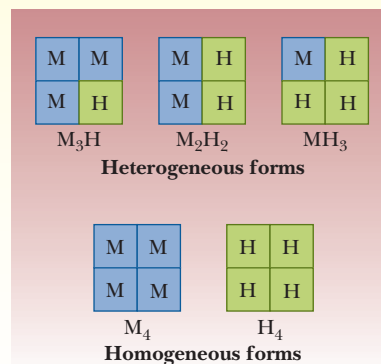
# Biochemical Connections HEALTH SCIENCES

## Enzymes as Markers for Disease

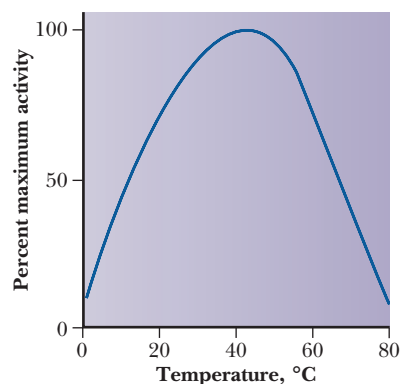
Some enzymes are found only in specific tissues or in a limited number of such tissues. The enzyme lactate dehydrogenase (LDH) has two different types of subunits—one found primarily in heart muscle (H), and another found in skeletal muscle (M). The two different subunits differ slightly in amino acid composition; consequently, they can be separated electrophoretically or chromatographically on the basis of charge. Because LDH is a tetramer of four subunits, and because the H and M subunits can combine in all possible combinations, LDH can exist in five different forms, called **isozymes**, depending on the source. An increase of any form of LDH in the blood indicates some kind of tissue damage. A heart attack used to be diagnosed by an increase of LDH from heart muscle. Similarly, there are different forms of creatine kinase (CK), an enzyme that occurs in the brain, heart, and skeletal muscle. Appearance of the brain type can indicate a stroke or a brain tumor, whereas the heart type indicates a heart attack. After a heart attack, CK shows up more rapidly in the blood than LDH. Monitoring the presence of both enzymes extends the possibility of diagnosis, which is useful, because a very mild heart attack might be difficult to diagnose. An elevated level of the isozyme from heart muscle in blood is a definite indication of damage to the heart tissue.

A particularly useful enzyme to assay is acetylcholinesterase (ACE), which is important in controlling certain nerve impulses. Many pesticides interfere with this enzyme, so farm workers are often tested to be sure that they have not received inappropriate exposure to these important agricultural toxins. In fact, more

than 20 enzymes are typically used in the clinical lab to diagnose disease. There are highly specific markers for enzymes active in the pancreas, red blood cells, liver, heart, brain, and prostate gland, and many of the endocrine glands. Because these enzymes are relatively easy to assay, even using automated techniques, they are part of the “standard” blood test your doctor is likely to request.



■ The possible isozymes of lactate dehydrogenase. The symbol M refers to the dehydrogenase form that predominates in skeletal muscle, and the symbol H refers to the form that predominates in heart (cardiac) muscle.



■ **FIGURE 6.2** The effect of temperature on enzyme activity. The relative activity of an enzymatic reaction as a function of temperature. The decrease in activity above 50°C is due to thermal denaturation.

temperature, adding more heat denatures more enzyme and slows down the reaction. Figure 6.2 shows a typical curve of temperature effect on an enzyme-catalyzed reaction. The preceding Biochemical Connections box describes another way in which the specificity of enzymes is of great use.

## 6.3 Enzyme Kinetic Equations

The rate of a chemical reaction is usually expressed in terms of a change in the concentration of a reactant or of a product in a given time interval. Any convenient experimental method can be used to monitor changes in concentration. In a reaction of the form  $A + B \rightarrow P$ , where A and B are reactants and P is the product, the rate of the reaction can be expressed either in terms of the rate of disappearance of one of the reactants or in terms of the rate of appearance of the product. The rate of disappearance of A is  $-\Delta[A]/\Delta t$ , where  $\Delta$  symbolizes change, [A] is the concentration of A in moles per liter, and  $t$  is time. Likewise, the rate of disappearance of B is  $-\Delta[B]/\Delta t$ , and the rate of appearance of P is  $\Delta[P]/\Delta t$ . The rate of the reaction can be expressed in terms of any of these changes because the rates of appearance of product and disappearance of reactant are related by the stoichiometric equation for the reaction.

$$\text{Rate} = \frac{-\Delta[A]}{\Delta t} = \frac{-\Delta[B]}{\Delta t} = \frac{\Delta[P]}{\Delta t}$$

The negative signs for the changes in concentration of A and B indicate that A and B are being used up in the reaction, while P is being produced.

It has been established that the rate of a reaction at a given time is proportional to the product of the concentrations of the reactants raised to the appropriate powers,

$$\text{Rate} \propto [\text{A}]^f [\text{B}]^g$$

or, as an equation,

$$\text{Rate} = k[\text{A}]^f [\text{B}]^g$$

where  $k$  is a proportionality constant called the **rate constant**. The exponents  $f$  and  $g$  must be determined experimentally. They are *not necessarily* equal to the coefficients of the balanced equation, but frequently they are. The square brackets, as usual, denote molar concentration. When the exponents in the rate equation have been determined experimentally, a mechanism for the reaction—a description of the detailed steps along the path between reactants and products—can be proposed.

The exponents in the rate equation are usually small whole numbers, such as 1 or 2. (There are also some cases in which the exponent 0 occurs.) The values of the exponents are related to the number of molecules involved in the detailed steps that constitute the mechanism. The *overall order* of a reaction is the sum of all the exponents. If, for example, the rate of a reaction  $\text{A} \rightarrow \text{P}$  is given by the rate equation

$$\text{Rate} = k[\text{A}]^1 \quad (6.1)$$

where  $k$  is the rate constant and the exponent for the concentration of A is 1, then the reaction is **first order** with respect to A and first order overall. The rate of radioactive decay of the widely used tracer isotope phosphorus 32 ( $^{32}\text{P}$ ; atomic weight = 32) depends only on the concentration of  $^{32}\text{P}$  present. Here we have an example of a first-order reaction. Only the  $^{32}\text{P}$  atoms are involved in the mechanism of the radioactive decay, which, as an equation, takes the form

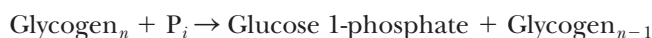


$$\text{Rate} = k[^{32}\text{P}]^1 = k[^{32}\text{P}]$$

If the rate of a reaction  $\text{A} + \text{B} \rightarrow \text{C} + \text{D}$  is given by

$$\text{Rate} = k[\text{A}]^1 [\text{B}]^1 \quad (6.2)$$

where  $k$  is the rate constant, the exponent for the concentration of A is 1, and the exponent for the concentration of B is 1, then the reaction is said to be first order with respect to A, first order with respect to B, and **second order** overall. In the reaction of glycogen<sub>*n*</sub> (a polymer of glucose with  $n$  glucose residues) with inorganic phosphate,  $\text{P}_i$  to form glucose 1-phosphate + glycogen<sub>*n-1*</sub>, the rate of reaction depends on the concentrations of both reactants.



$$\text{Rate} = k[\text{Glycogen}]^1 [\text{P}_i]^1 = k[\text{Glycogen}][\text{P}_i]$$

where  $k$  is the rate constant. Both the glycogen and the phosphate take part in the reaction mechanism. The reaction of glycogen with phosphate is first order with respect to glycogen, first order with respect to phosphate, and second order overall.

Many common reactions are first or second order. After the order of the reaction is determined experimentally, proposals can be made about the mechanism of a reaction.

### Is the rate of a reaction always based on the concentration of reactants?

Exponents in a rate equation may be equal to zero, with the rate for a reaction  $\text{A} \rightarrow \text{B}$  given by the equation

$$\text{Rate} = k[\text{A}]^0 = k \quad (6.3)$$

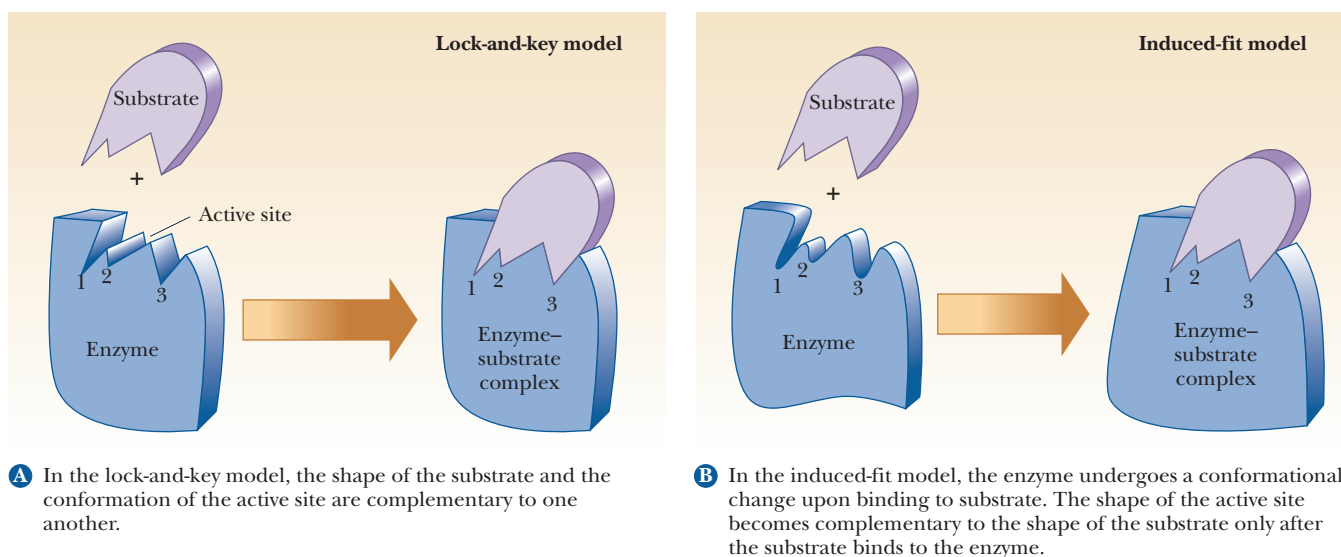
Such a reaction is called **zero order**, and its rate, which is constant, depends not on concentrations of reactants but on other factors, such as the presence of catalysts. Enzyme-catalyzed reactions can exhibit zero-order kinetics when the concentrations of reactants are so high that the enzyme is completely saturated with reactant molecules. This point will be discussed in more detail later in this chapter but, for the moment, we can consider the situation analogous to a traffic bottleneck in which six lanes of cars are trying to cross a two-lane bridge. The rate at which the cars cross is not affected by the number of waiting cars, only by the number of lanes available on the bridge.

## 6.4 Enzyme–Substrate Binding

In an enzyme-catalyzed reaction, the enzyme binds to the **substrate** (one of the reactants) to form a complex. The formation of the complex leads to the formation of the transition-state species, which then forms the product. The nature of transition states in enzymatic reactions is a large field of research in itself, but some general statements can be made on the subject. A substrate binds, usually by noncovalent interactions, to a small portion of the enzyme called the **active site**, frequently situated in a cleft or crevice in the protein and consisting of certain amino acids that are essential for enzymatic activity (Figure 6.3). The catalyzed reaction takes place at the active site, usually in several steps.

### Why do enzymes bind to substrates?

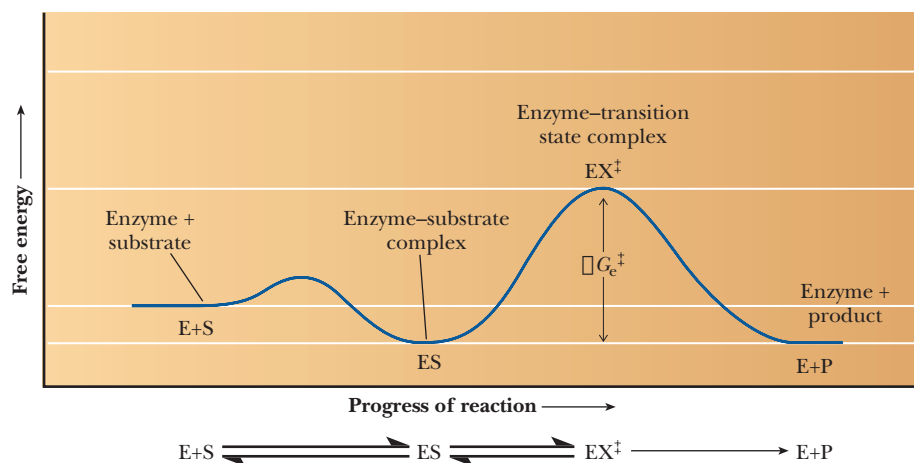
The first step is the binding of substrate to the enzyme, which occurs because of highly specific interactions between the substrate and the side chains and backbone groups of the amino acids making up the active site. Two important models have been developed to describe the binding process. The first, the **lock-and-key model**, assumes a high degree of similarity between the shape of the substrate and the geometry of the binding site on the enzyme (Figure 6.3a). The substrate binds to a site whose shape complements its own, like a key in a lock or the correct piece in a three-dimensional jigsaw puzzle. This model has intuitive appeal but is now largely of historical interest because it does not take into account an important property of proteins, namely their conformational flexibility. The second model takes into account the fact that proteins have some three-dimensional flexibility. According to this **induced-fit model**, the binding of the substrate induces a conformational change in the enzyme that results in a complementary



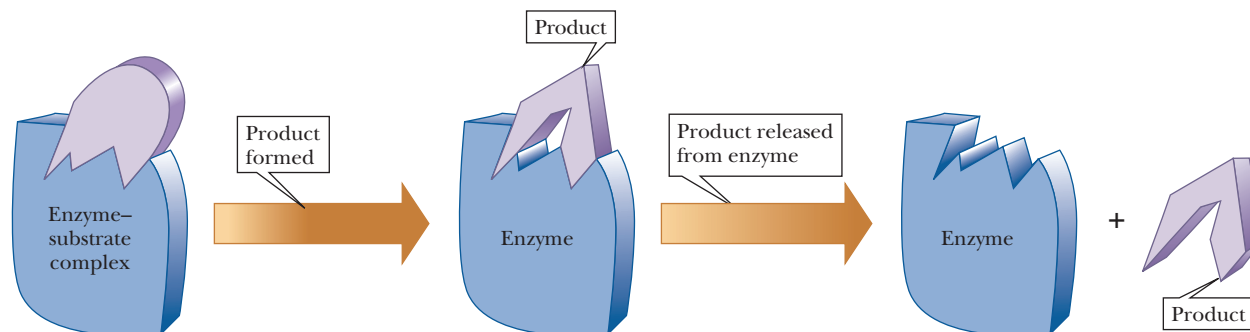
■ **FIGURE 6.3** Two models for the binding of a substrate to an enzyme.

fit after the substrate is bound (Figure 6.3b). The binding site has a different three-dimensional shape before the substrate is bound. The induced-fit model is also more attractive when we consider the nature of the transition state and the lowered activation energy that occurs with an enzyme-catalyzed reaction. The enzyme and substrate must bind to form the ES complex before anything else can happen. What would happen if this binding were too perfect? Figure 6.4 shows what happens when E and S bind. An attraction must exist between E and S for them to bind. This attraction causes the ES complex to be lower on an energy diagram than the  $E + S$  at the start. Then the bound ES must attain the conformation of the transition state  $EX^\ddagger$ . If the binding of E and S to form ES were a perfect fit, the ES would be at such a low energy that the difference between ES and  $EX^\ddagger$  would be very large. This would slow down the rate of reaction. Many studies have shown that enzymes increase the rate of reaction by lowering the energy of the transition state,  $EX^\ddagger$ , while raising the energy of the ES complex. The induced-fit model certainly supports this last consideration better than the lock-and-key model; in fact, the induced-fit model mimics the transition state.

After the substrate is bound and the transition state is subsequently formed, catalysis can occur. This means that bonds must be rearranged. In the transition state, the substrate is bound close to atoms with which it is to react. Furthermore, the substrate is placed in the correct orientation with respect to those atoms. Both effects, proximity and orientation, speed up the reaction. As bonds are broken and new bonds are formed, the substrate is transformed into product. The product is released from the enzyme, which can then catalyze the reaction of more substrate to form more product (Figure 6.5). Each enzyme has its own unique mode of catalysis, which is not surprising in view of enzymes' great specificity. Even so, some general modes of catalysis exist in enzymatic reactions. Two enzymes, chymotrypsin and aspartate transcarbamoylase, are good examples of these general principles.



■ **FIGURE 6.4** The free energy of activation profile of a reaction with strong binding of the substrate to the enzyme to form an enzyme–substrate complex.

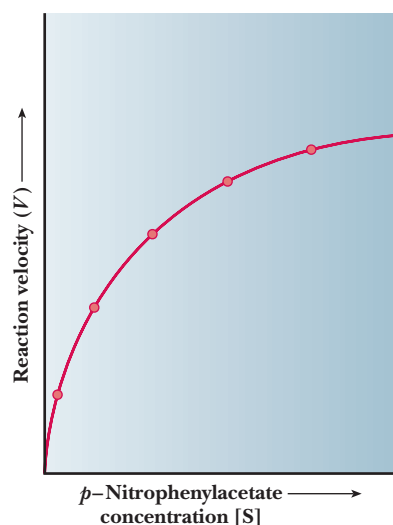
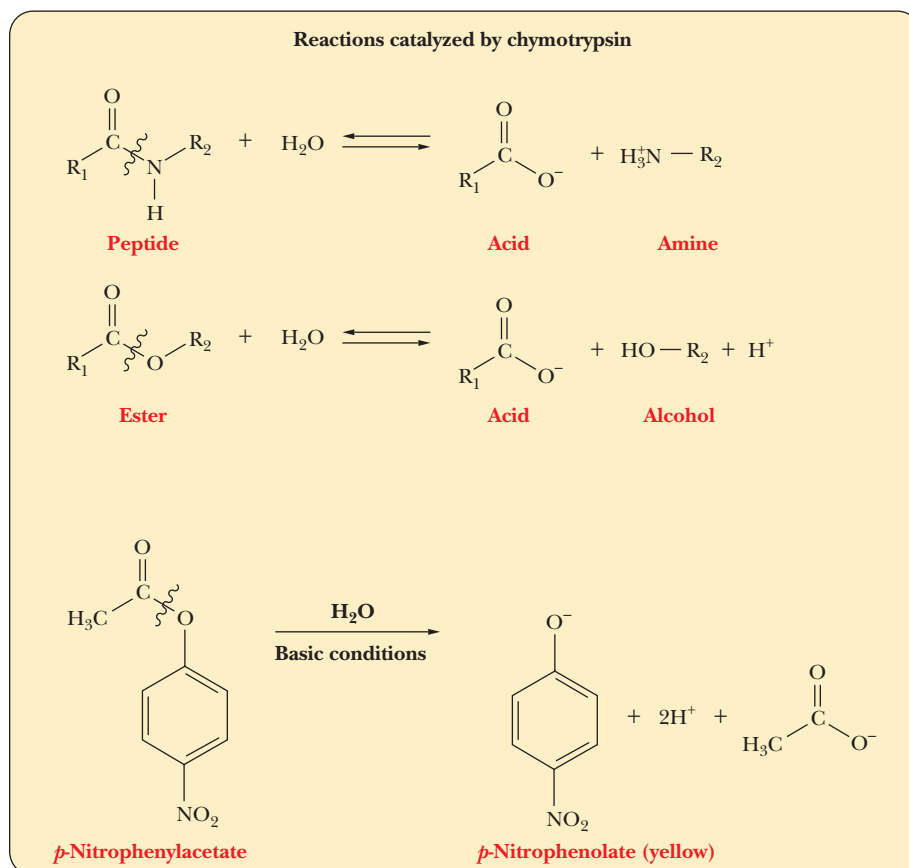


■ **FIGURE 6.5** Formation of product from substrate (bound to the enzyme), followed by release of the product.



## 6.5 Examples of Enzyme-Catalyzed Reactions

**Chymotrypsin** is an enzyme that catalyzes the hydrolysis of peptide bonds, with some specificity for residues containing aromatic side chains. Chymotrypsin also cleaves peptide bonds at other sites, such as leucine, histidine, and glutamine, but with a lower frequency than at aromatic amino acid residues. It also catalyzes the hydrolysis of ester bonds.

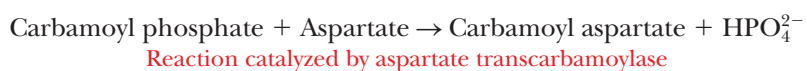


■ **FIGURE 6.6** Dependence of reaction velocity,  $V$ , on *p*-nitrophenylacetate concentration,  $[S]$ , in a reaction catalyzed by chymotrypsin. The shape of the curve is hyperbolic.

Although ester hydrolysis is not important to the physiological role of chymotrypsin in the digestion of proteins, it is a convenient model system for investigating the enzyme's catalysis of hydrolysis reactions. The usual laboratory procedure is to use *p*-nitrophenyl esters as the substrate and to monitor the progress of the reaction by the appearance of a yellow color in the reaction mixture caused by the production of *p*-nitrophenolate ion.

In a typical reaction in which a *p*-nitrophenyl ester is hydrolyzed by chymotrypsin, the experimental rate of the reaction depends on the concentration of the substrate—in this case, the *p*-nitrophenyl ester. At low substrate concentrations, the rate of reaction increases as more substrate is added. At higher substrate concentrations, the rate of the reaction changes very little with the addition of more substrate, and a maximum rate is reached. When these results are presented in a graph, the curve is hyperbolic (Figure 6.6).

Another enzyme-catalyzed reaction is the one catalyzed by the enzyme **aspartate transcarbamoylase** (ATCase). This reaction is the first step in a pathway leading to the formation of cytidine triphosphate (CTP) and uridine triphosphate (UTP), which are ultimately needed for the biosynthesis of RNA and DNA. In this reaction, carbamoyl phosphate reacts with aspartate to produce carbamoyl aspartate and phosphate ion.

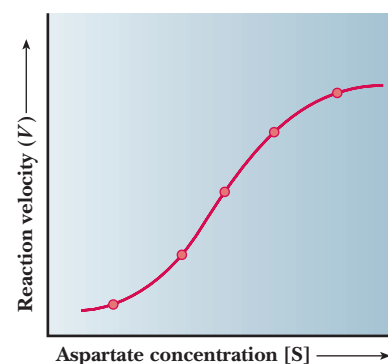


The rate of this reaction also depends on substrate concentration—in this case, the concentration of aspartate (the carbamoyl phosphate concentration is kept constant). Experimental results show that, once again, the rate of the reaction depends on substrate concentration at low and moderate concentrations, and, once again, a maximum rate is reached at high substrate concentrations.

There is, however, one very important difference. For this reaction, a graph showing the dependence of reaction rate on substrate concentration has a sigmoidal rather than hyperbolic shape (Figure 6.7).

### Why do chymotrypsin and ATCase have different velocity curves?

The results of experiments on the reaction kinetics of chymotrypsin and aspartate transcarbamoylase are representative of experimental results obtained with many enzymes. The overall kinetic behavior of many enzymes resembles that of chymotrypsin, while other enzymes behave similarly to aspartate transcarbamoylase. We can use this information to draw some general conclusions about the behavior of enzymes. The comparison between the kinetic behaviors of chymotrypsin and ATCase is reminiscent of the relationship between the oxygen-binding behaviors of myoglobin and hemoglobin, discussed in Chapter 4. ATCase and hemoglobin are allosteric proteins; chymotrypsin and myoglobin are not. (Recall from Section 4.5 that allosteric proteins are the ones in which subtle changes at one site affect structure and function at another site. Cooperative effects, such as the fact that the binding of the first oxygen molecule to hemoglobin makes it easier for other oxygen molecules to bind, are a hallmark of allosteric proteins.) The differences in behavior between allosteric and nonallosteric proteins can be understood in terms of models based on structural differences between the two kinds of proteins. When we encounter the mechanisms of the many enzyme-catalyzed reactions in subsequent chapters, we shall need a model that explains the hyperbolic plot of kinetic data for nonallosteric enzymes and another model that explains the sigmoidal plot for allosteric enzymes. The Michaelis–Menten model is widely used for nonallosteric enzymes, and several models are used for allosteric enzymes.



■ **FIGURE 6.7** Dependence of reaction velocity,  $V$ , on aspartate concentration,  $[S]$ , in a reaction catalyzed by aspartate transcarbamoylase. The shape of the curve is sigmoidal.

## Biochemical Connections NEUROSCIENCE

### Enzymes and Memory

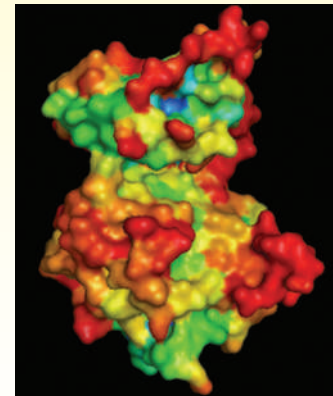
There are thousands of different enzymes in a cell and we will study many of them in the following chapters. New information about the importance of enzymes is published every week in the scientific literature. One class of enzyme important in many metabolic processes is **kinases**, which we will see many examples of in the next chapters. One kinase, protein Kinase M $\zeta$  (PKM $\zeta$ ), has been implicated in the maintenance of long-term memory. Scientists created a drug called ZIP that blocks this enzyme. They gave rats saccharine-laced water and then artificially induced nausea shortly afterward. Control rats then had an aversion to saccharine-laced water for weeks. Humans have the same response; normally a person who vomits shortly after eating a specific type of food will remember the experience and not want to consume the same food.

Researchers then injected the cerebral cortex of test rats and found that they lost their aversion to saccharine within 2 hours. Since blocking the PKM $\zeta$  eliminated the memory, this was a first indication that this specific enzyme is required for long-term memory retention, a novel finding. The next step will be to determine whether the drug eliminates all learning past a certain

point or whether it could be used selectively. Researchers have been looking for ways to selectively block memories, such as the painful memories of trauma survivors.



■ **Memory molecule.** PKM $\zeta$  sustains long-term memory in the cerebral cortex of rats.



© Kevin Bolding/Sacktor Laboratory

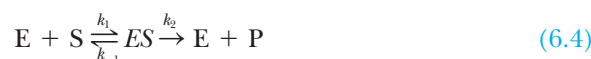
## 6.6 The Michaelis–Menten Approach to Enzyme Kinetics

A particularly useful model for the kinetics of enzyme-catalyzed reactions was devised in 1913 by Leonor Michaelis and Maud Menten. It is still the basic model for nonallosteric enzymes and is widely used, even though it has undergone many modifications.

A typical reaction might be the conversion of some substrate, S, to a product, P. The stoichiometric equation for the reaction is



The mechanism for an enzyme-catalyzed reaction can be summarized in the form



Note the assumption that the product is not converted to substrate to any appreciable extent. In this equation,  $k_1$  is the rate constant for the formation of the enzyme–substrate complex, ES, from the enzyme, E, and the substrate, S;  $k_{-1}$  is the rate constant for the reverse reaction, dissociation of the ES complex to free enzyme and substrate; and  $k_2$  is the rate constant for the conversion of the ES complex to product P and the subsequent release of product from the enzyme. The enzyme appears explicitly in the mechanism, and the concentrations of both free enzyme, E, and enzyme–substrate complex, ES, therefore, appear in the rate equations. Catalysts characteristically are regenerated at the end of the reaction, and this is true of enzymes.

When we measure the rate (also called the velocity) of an enzymatic reaction at varying substrate concentrations, we see that the rate depends on the substrate concentration, [S]. We measure the initial rate of the reaction (the rate measured immediately after the enzyme and substrate are mixed) so that we can be certain that the product is not converted to substrate to any appreciable extent. This velocity is sometimes written  $V_{\text{init}}$  or  $V_0$  to indicate this initial velocity, but it is important to remember that all the calculations involved in enzyme kinetics assume that the velocity measured is the initial velocity. We can graph our results as in Figure 6.8. In the lower region of the curve (at low levels of substrate), the reaction is first order (Section 6.3), implying that the velocity,  $V$ , depends on substrate concentration [S]. In the upper portion of the curve (at higher levels of substrate), the reaction is zero order; the rate is independent of concentration. The active sites of all of the enzyme molecules are saturated. At infinite substrate concentration, the reaction would proceed at its maximum velocity, written  $V_{\text{max}}$ .

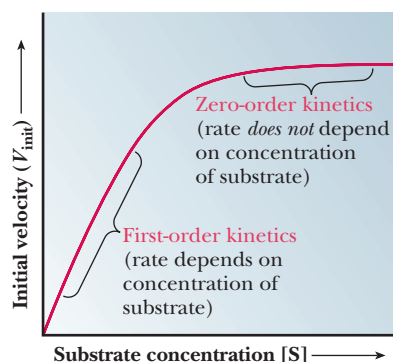
The substrate concentration at which the reaction proceeds at one-half its maximum velocity has a special significance. It is given the symbol  $K_M$ , which can be considered an inverse measure of the affinity of the enzyme for the substrate. The lower the  $K_M$ , the higher the affinity.

Let us examine the mathematical relationships among the quantities [E], [S],  $V_{\text{max}}$ , and  $K_M$ . The general mechanism of the enzyme-catalyzed reaction involves binding of the enzyme, E, to the substrate to form a complex, ES, which then forms the product. The rate of formation of the enzyme–substrate complex, ES, is

$$\text{Rate of formation} = \frac{\Delta[\text{ES}]}{\Delta t} = k_1[\text{E}][\text{S}] \quad (6.5)$$

where  $\Delta[\text{ES}]/\Delta t$  means the change in the concentration of the complex,  $\Delta[\text{ES}]$ , during a given time  $\Delta t$ , and  $k_1$  is the rate constant for the formation of the complex.

The ES complex breaks down in two reactions, by returning to enzyme and substrate or by giving rise to product and releasing enzyme. The rate of disappearance of complex is the sum of the rates of the two reactions.



■ **FIGURE 6.8** The rate and the observed kinetics of an enzymatic reaction depend on substrate concentration. The concentration of enzyme, [E], is constant.

$$\text{Rate of breakdown} = \frac{-\Delta[\text{ES}]}{\Delta t} = k_{-1}[\text{ES}] + k_2[\text{ES}] \quad (6.6)$$

The negative sign in the term  $-\Delta[\text{ES}]/\Delta t$  means that the concentration of the complex decreases as the complex breaks down. The term  $k_{-1}$  is the rate constant for the dissociation of complex to regenerate enzyme and substrate, and  $k_2$  is the rate constant for the reaction of the complex to give product and enzyme.

Enzymes are capable of processing the substrate very efficiently, and a **steady state** is soon reached in which the rate of formation of the enzyme–substrate complex equals the rate of its breakdown. Very little complex is present, and it turns over rapidly, but its concentration stays the same with time. According to the *steady-state theory*, then, the rate of formation of the enzyme–substrate complex equals the rate of its breakdown,

$$\frac{\Delta[\text{ES}]}{\Delta t} = \frac{-\Delta[\text{ES}]}{\Delta t} \quad (6.7)$$

and

$$k_1[\text{E}][\text{S}] = k_{-1}[\text{ES}] + k_2[\text{ES}] \quad (6.8)$$

To solve for the concentration of the complex, ES, it is necessary to know the concentration of the other species involved in the reaction. The initial concentration of substrate is a known experimental condition and does not change significantly during the initial stages of the reaction. The substrate concentration is much greater than the enzyme concentration. The total concentration of the enzyme,  $[\text{E}]_{\text{T}}$ , is also known, but a large proportion of it may be involved in the complex. The concentration of free enzyme,  $[\text{E}]$ , is the difference between  $[\text{E}]_{\text{T}}$ , the total concentration, and  $[\text{ES}]$ , which can be written as an equation:

$$[\text{E}] = [\text{E}]_{\text{T}} - [\text{ES}] \quad (6.9)$$

Substituting for the concentration of free enzyme,  $[\text{E}]$ , in Equation 6.8,

$$k_1([\text{E}]_{\text{T}} - [\text{ES}])[\text{S}] = k_{-1}[\text{ES}] + k_2[\text{ES}] \quad (6.10)$$

Collecting all the rate constants for the individual reactions,

$$\frac{([\text{E}]_{\text{T}} - [\text{ES}])[\text{S}]}{[\text{ES}]} = \frac{k_{-1} + k_2}{k_1} = K_{\text{M}} \quad (6.11)$$

where  $K_{\text{M}}$  is called the **Michaelis constant**. It is now possible to solve Equation 6.11 for the concentration of enzyme–substrate complex:

$$\begin{aligned} \frac{([\text{E}]_{\text{T}} - [\text{ES}])[\text{S}]}{[\text{ES}]} &= K_{\text{M}} \\ [\text{E}]_{\text{T}}[\text{S}] - [\text{ES}][\text{S}] &= K_{\text{M}}[\text{ES}] \\ [\text{E}]_{\text{T}}[\text{S}] &= [\text{ES}](K_{\text{M}} + [\text{S}]) \end{aligned}$$

or

$$[\text{ES}] = \frac{[\text{E}]_{\text{T}}[\text{S}]}{K_{\text{M}} + [\text{S}]} \quad (6.12)$$

In the initial stages of the reaction, so little product is present that no reverse reaction of product to complex need be considered. Thus the initial rate determined in enzymatic reactions depends on the rate of breakdown of the enzyme–substrate complex into product and enzyme. In the Michaelis–Menten model, the initial rate,  $V$ , of the formation of product depends only on the rate of the breakdown of the ES complex,

$$V = k_2[\text{ES}] \quad (6.13)$$

and on the substitution of the expression for [ES] from Equation 6.12,

$$V = \frac{k_2[E]_T[S]}{K_M + [S]} \quad (6.14)$$

If the substrate concentration is so high that the enzyme is completely saturated with substrate ( $[ES] = [E]_T$ ), the reaction proceeds at its maximum possible rate ( $V_{\max}$ ). Substituting  $[E]_T$  for [ES] in Equation 6.13,

$$V = V_{\max} = k_s[E]_T \quad (6.15)$$

The total concentration of enzyme is a constant, which means that

$$V_{\max} = \text{Constant}$$

This expression for  $V_{\max}$  resembles that for a zero-order reaction given in Equation 6.3:

$$\text{Rate} = k[A]_0 = k$$

Note that the concentration of substrate, [A], appears in Equation 6.3 rather than the concentration of enzyme, [E], as in Equation 6.15. When the enzyme is saturated with substrate, zero-order kinetics with respect to substrate are observed.

Substituting the expression for  $V_{\max}$  into Equation 6.14 enables us to relate the observed velocity at any substrate concentration to the maximum rate of an enzymatic reaction:

$$V = \frac{V_{\max}[S]}{K_M + [S]} \quad (6.16)$$

Figure 6.8 shows the effect of increasing substrate concentration on the observed rate. In such an experiment, the reaction is run at several substrate concentrations, and the rate is determined by following the disappearance of reactant, or the appearance of product, by way of any convenient method. At low-substrate concentrations, first-order kinetics are observed. At higher substrate concentrations (well beyond  $10 \times K_M$ ), when the enzyme is saturated, the constant reaction rate characteristic of zero-order kinetics is observed.

This constant rate, when the enzyme is saturated with substrate, is the  $V_{\max}$  for the enzyme, a value that can be roughly estimated from the graph. The value of  $K_M$  can also be estimated from the graph. From Equation 6.16,

$$V = \frac{V_{\max}[S]}{K_M + [S]}$$

When experimental conditions are adjusted so that  $[S] = K_M$ ,

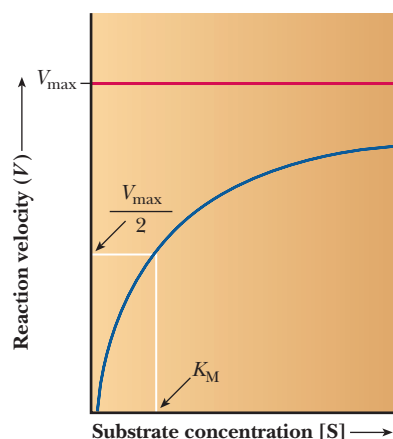
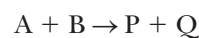
$$V = \frac{V_{\max}[S]}{[S] + [S]}$$

and

$$V = \frac{V_{\max}}{2}$$

In other words, when the rate of the reaction is half its maximum value, the substrate concentration is equal to the Michaelis constant (Figure 6.9). This fact is the basis of the graphical determination of  $K_M$ .

Note that the reaction used to generate the Michaelis–Menten equation was the simplest enzyme equation possible, that with a single substrate going to a single product. Most enzymes catalyze reactions containing two or more substrates. This does not invalidate our equations, however. For enzymes with multiple substrates, the same equations can be used, but only one substrate can be studied at a time. If, for example, we had the enzyme-catalyzed reaction



■ **FIGURE 6.9** Graphical determination of  $V_{\max}$  and  $K_M$  from a plot of reaction velocity,  $V$ , against substrate concentration,  $[S]$ .  $V_{\max}$  is the constant rate reached when the enzyme is completely saturated with substrate, a value that frequently must be estimated from such a graph.



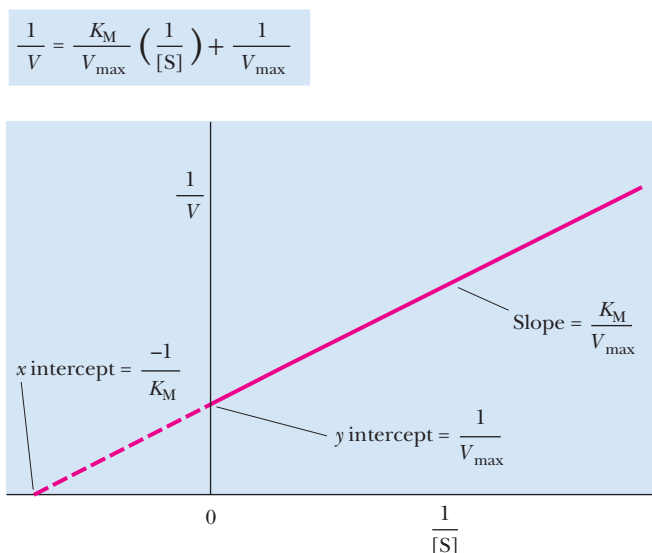
we could still use the Michaelis–Menten approach. If we hold A at saturating levels and then vary the amount of B over a broad range, the curve of velocity versus [B] will still be a hyperbola, and we can still calculate the  $K_M$  for B. Conversely, we could hold the level of B at saturating levels and vary the amount of A to determine the  $K_M$  for A. There are even enzymes that have two substrates where, if we plot  $V$  versus [substrate A], we see the Michaelis–Menten hyperbola, but, if we plot  $V$  versus [substrate B], we see the sigmoidal curve shown for aspartate transcarbamoylase in Figure 6.7. Technically the term  $K_M$  is appropriate only for enzymes that exhibit a hyperbolic curve of velocity versus [substrate].

### How do we calculate $K_M$ and $V_{max}$ from a graph?

The curve that describes the rate of a nonallosteric enzymatic reaction is hyperbolic. It is quite difficult to estimate  $V_{max}$  because it is an asymptote, and the value is never reached with any finite substrate concentration that we could use in the lab. This, in turn, makes it difficult to determine the  $K_M$  of the enzyme. It is considerably easier to work with a straight line than a curve. One can transform the equation for a hyperbola (Equation 6.16) into an equation for a straight line by taking the reciprocals of both sides:

$$\begin{aligned}\frac{1}{V} &= \frac{K_M + [S]}{V_{max} [S]} \\ \frac{1}{V} &= \frac{K_M}{V_{max} [S]} + \frac{[S]}{V_{max} [S]} \\ \frac{1}{V} &= \frac{K_M}{V_{max}} \times \frac{1}{[S]} + \frac{1}{V_{max}}\end{aligned}\quad (6.17)$$

The equation now has the form of a straight line,  $y = mx + b$ , where  $1/V$  takes the place of the  $y$  coordinate and  $1/[S]$  takes the place of the  $x$  coordinate. The slope of the line,  $m$ , is  $K_M/V_{max}$ , and the  $y$  intercept,  $b$ , is  $1/V_{max}$ . Figure 6.10 presents this information graphically as a **Lineweaver–Burk double-reciprocal plot**. It is usually easier to draw the best straight line through a set of points than to estimate the best fit of points to a curve. Convenient computer methods exist for drawing the best straight line through a series of experimental points. Such a line can be extrapolated to high values of [S], ones that might be unattainable because of solubility limits or the cost of the substrate. The extrapolated line can be used to obtain  $V_{max}$ .



■ **FIGURE 6.10** A Lineweaver–Burk double-reciprocal plot of enzyme kinetics. The reciprocal of reaction velocity,  $1/V$ , is plotted against the reciprocal of the substrate concentration,  $1/[S]$ . The slope of the line is  $K_M/V_{max}$ , and the  $y$  intercept is  $1/V_{max}$ . The  $x$  intercept is  $-1/K_M$ .

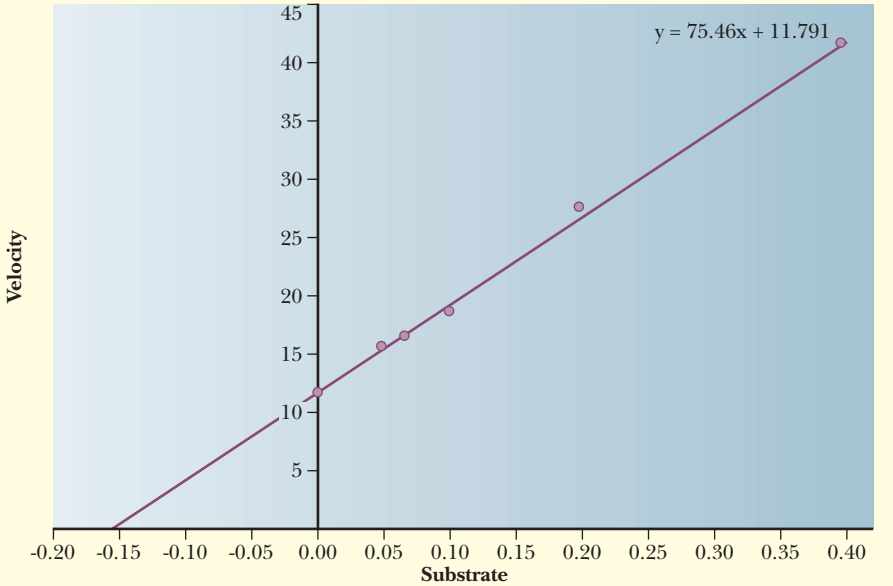
Apply Your Knowledge

The following data describe an enzyme-catalyzed reaction. Plot these results using the Lineweaver–Burk method, and determine values for  $K_M$  and  $V_{\max}$ . The symbol mM represents millimoles per liter;  $1\text{ mM} = 1 \times 10^{-3}\text{ mol L}^{-1}$ . (The concentration of the enzyme is the same in all experiments.)

Substrate Concentration (mM)	Velocity (mM sec <sup>-1</sup> )
2.5	0.024
5.0	0.036
10.0	0.053
15.0	0.060
20.0	0.061

**Solution**  
The reciprocal of substrate concentration and of velocity gives the following results:

1/[S] (mM <sup>-1</sup> )	1/V (mM sec <sup>-1</sup> ) <sup>-1</sup>
0.400	41.667
0.200	27.778
0.100	18.868
0.067	16.667
0.050	15.625



Plotting the results gives a straight line. Visually from the graph, the  $y$  intercept is 12 and the  $x$  intercept is  $-0.155$ . The reciprocal of the  $y$  intercept is  $V_{\max}$ , which equals  $0.083\text{ mM sec}^{-1}$ . The reciprocal of the negative of the  $x$  intercept =  $K_M = 6.45\text{ mM}$ . We can also use the exact equation for the line of best fit to the experimental points, which is  $1/V = 75.46 (1/[S]) + 11.8$ . Using the equation generates the following:  $K_M = 6.39\text{ mM}$  and  $V_{\max} = 0.0847\text{ mM sec}^{-1}$ .

### What is the significance of $K_M$ and $V_{max}$ ?

We have already seen that when the rate of a reaction,  $V$ , is equal to half the maximum rate possible,  $V = V_{max}/2$ , then  $K_M = [S]$ . One interpretation of the Michaelis constant,  $K_M$ , is that it equals the concentration of substrate at which 50% of the enzyme active sites are occupied by substrate. The Michaelis constant has the units of concentration.

Another interpretation of  $K_M$  relies on the assumptions of the original Michaelis–Menten model of enzyme kinetics. Recall Equation 6.4:



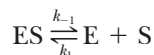
As before,  $k_1$  is the rate constant for the formation of the enzyme–substrate complex, ES, from the enzyme and substrate;  $k_{-1}$  is the rate constant for the reverse reaction, dissociation of the ES complex to free enzyme and substrate; and  $k_2$  is the rate constant for the formation of product P and the subsequent release of product from the enzyme. Also recall from Equation 6.11 that

$$K_M = \frac{k_{-1} + k_2}{k_1}$$

Consider the case in which the reaction  $E + S \rightarrow ES$  takes place more frequently than  $ES \rightarrow E + P$ . In kinetic terms, this means that the dissociation rate constant  $k_{-1}$  is greater than the rate constant for the formation of product,  $k_2$ . If  $k_{-1}$  is *much* larger than  $k_2$  ( $k_{-1} \gg k_2$ ), as was originally assumed by Michaelis and Menten, then approximately

$$K_M = \frac{k_{-1}}{k_1}$$

It is informative to compare the expression for the Michaelis constant with the equilibrium constant expression for the dissociation of the ES complex,



The  $k$  values are the rate constants, as before. The equilibrium constant expression is

$$K_{eq} = \frac{[E][S]}{[ES]} = \frac{k_{-1}}{k_1}$$

This expression is the same as that for  $K_M$  and makes the point that, when the assumption that  $k_{-1} \gg k_2$  is valid,  $K_M$  is simply the dissociation constant for the ES complex.  $K_M$  is a measure of how tightly the substrate is bound to the enzyme. The greater the value of  $K_M$ , the less tightly the substrate is bound to the enzyme. Note that in the steady-state approach,  $k_2$  is not assumed to be small compared with  $k_{-1}$ ; therefore,  $K_M$  is not technically a dissociation constant, even though it is often used to estimate the affinity of the enzyme for the substrate.

$V_{max}$  is related to the **turnover number** of an enzyme, a quantity equal to the catalytic constant,  $k_2$ . This constant is also referred to as  $k_{cat}$  or  $k_p$ :

$$\frac{V}{[E_T]} = \text{turnover number} = k_{cat}$$

The turnover number is the number of moles of substrate that react to form product per mole of enzyme per unit time. This statement assumes that the enzyme is fully saturated with substrate and thus that the reaction is proceeding at the maximum rate. Table 6.2 lists turnover numbers for typical enzymes, where the units are *per second*.

Turnover numbers are a particularly dramatic illustration of the efficiency of enzymatic catalysis. Catalase is an example of a particularly efficient enzyme. In Section 6.1, we encountered catalase in its role in converting hydrogen

TABLE 6.2

Turnover Numbers and $K_M$ for Some Typical Enzymes			
Enzyme	Function	$k_{\text{cat}}$ = Turnover Number*	$K_M$ **
Catalase	Conversion of $\text{H}_2\text{O}_2$ to $\text{H}_2\text{O}$ and $\text{O}_2$	$4 \times 10^7$	25
Carbonic Anhydrase	Hydration of $\text{CO}_2$	$1 \times 10^6$	12
Acetylcholinesterase	Regenerates acetylcholine, an important substance in transmission of nerve impulses, from acetate and choline	$1.4 \times 10^4$	$9.5 \times 10^{-2}$
Chymotrypsin	Proteolytic enzyme	$1.9 \times 10^2$	$6.6 \times 10^{-1}$
Lysozyme	Degrades bacterial cell-wall polysaccharides	0.5	$6 \times 10^{-3}$

\*The definition of turnover number is the moles of substrate converted to product per mole of enzyme per second. The units are  $\text{sec}^{-1}$ .

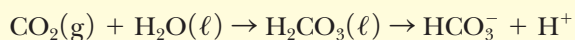
\*\*The units of  $K_M$  are millimolar.

peroxide to water and oxygen. As Table 6.2 indicates, it can transform 40 million moles of substrate to product every second. The following Biochemical Connections box describes some practical information available from the kinetic parameters we have discussed in this section.

## Biochemical Connections NEUROSCIENCE

### Enzyme Lets You Enjoy Champagne

As shown in the table above, the enzyme carbonic anhydrase acts very fast, turning over a million product molecules per second per enzyme molecule. This enzyme is very important physiologically, since it is responsible for how we transport  $\text{CO}_2$  to and from the lungs as part of our metabolism. It catalyzes the reaction:



Carbon dioxide is a molecule common to many metabolic processes, but it travels in the blood in the form of the much more soluble carbonic acid. Blood is unable to carry enough dissolved  $\text{CO}_2$  directly to support our metabolism, making carbonic anhydrase critical. Carbonic acid is also in equilibrium with bicarbonate and  $\text{H}^+$ , which helps maintain blood pH.

Carbon dioxide is also what gives carbonated beverages their fizz, which greatly affects our experience with the drink. Nobody likes flat beer or soda. However, until recently nobody really understood why. A few years ago, two physicians climbed a high mountain while taking the drug, acetazolamide, commonly used to prevent altitude sickness. They brought along a six-pack of beer in anticipation of celebrating reaching the top. Unfortunately, the beer was flat and tasted terrible. Further investigation showed that using the drug ruined the taste of soda and champagne, but not whisky or other noncarbonated beverages.

In 2009, a team of neuroscientists led by Dr. Charles Zuker did studies to explain this phenomenon. They identified taste receptor cells on the tongue that respond to  $\text{CO}_2$ . These cells also respond to sour taste. He determined that the molecular sensor

of these cells was, in fact, a type of carbonic anhydrase called carbonic anhydrase 4. Carbonic anhydrase is inhibited by acetazolamide. This work showed that carbonic anhydrase is responsible for how we perceive carbonated beverages. In the past, people had thought that our taste perception of carbonation was due to the popping of the bubbles, which triggered mechanoreceptors in the mouth.



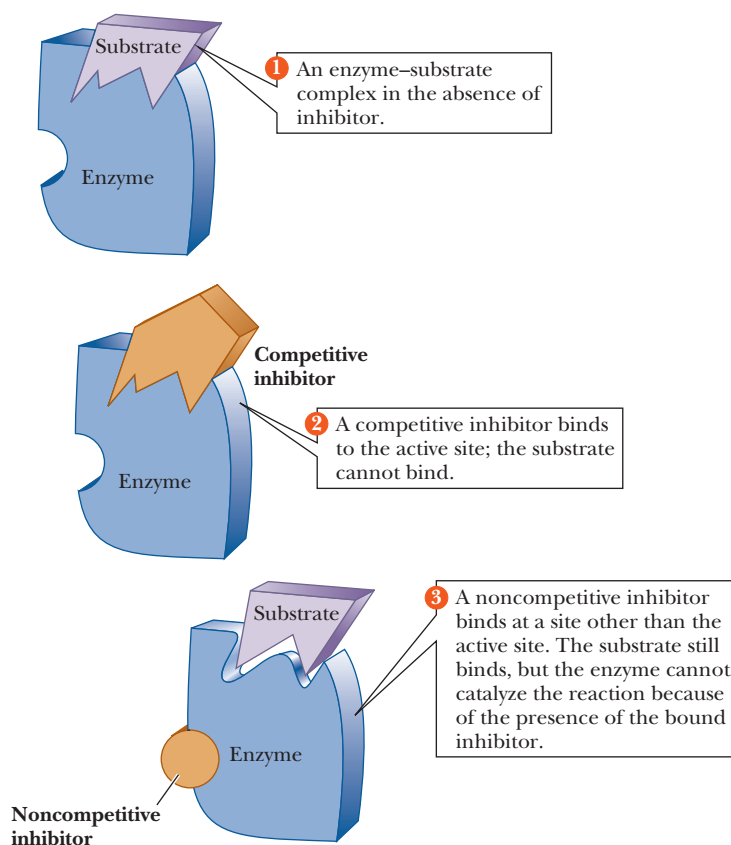
© 2010 Jupiter Images

## 6.7 Enzyme Inhibition

An **inhibitor**, as the name implies, is a substance that interferes with the action of an enzyme and slows the rate of a reaction. A good deal of information about enzymatic reactions can be obtained by observing the changes in the reaction caused by the presence of inhibitors. Inhibitors can affect an enzymatic reaction in two ways. A reversible inhibitor can bind to the enzyme and subsequently be released, leaving the enzyme in its original condition. An irreversible inhibitor reacts with the enzyme to produce a protein that is not enzymatically active and from which the original enzyme cannot be regenerated.

Two major classes of reversible inhibitors can be distinguished on the basis of the sites on the enzyme to which they bind. One class consists of compounds very similar in structure to the substrate. In this case, the inhibitor can bind to the active site and block the substrate's access to it. This mode of action is called **competitive inhibition** because the inhibitor competes with the substrate for the active site on the enzyme. Another major class of reversible inhibitors includes any inhibitor that binds to the enzyme at a site other than the active site and, as a result of binding, causes a change in the structure of the enzyme, especially around the active site. The substrate is still able to bind to the active site, but the enzyme cannot catalyze the reaction when the inhibitor is bound to it. This mode of action is called **noncompetitive inhibition** (Figure 6.11).

The two kinds of inhibition can be distinguished from one another in the laboratory. The reaction is carried out in the presence of inhibitor at several substrate concentrations, and the rates obtained are compared with those of the uninhibited reaction. The differences in the Lineweaver–Burk plots for the inhibited and uninhibited reactions provide the basis for the comparison.



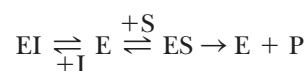
■ **FIGURE 6.11** Modes of action of inhibitors. The distinction between competitive and noncompetitive inhibitors is that a competitive inhibitor prevents binding of the substrate to the enzyme, whereas a noncompetitive inhibitor does not.



### How can we identify a competitive inhibitor?

In the presence of a competitive inhibitor, the slope of the Lineweaver–Burk plot changes, but the  $y$  intercept does not. (The  $x$  intercept also changes.)  $V_{\max}$  is unchanged, but  $K_M$  increases. More substrate is needed to get to a given rate in the presence of inhibitor than in its absence. This point specifically applies to the specific value  $V_{\max}/2$  (recall that at  $V_{\max}/2$ , the substrate concentration,  $[S]$ , equals  $K_M$ ) (Figure 6.12). Competitive inhibition can be overcome by a sufficiently high substrate concentration.

In the presence of a competitive inhibitor, the equation for an enzymatic reaction becomes



where EI is the enzyme–inhibitor complex. The dissociation constant for the enzyme–inhibitor complex can be written

$$EI \rightleftharpoons E + I$$

$$K_I = \frac{[E][I]}{[EI]}$$

It can be shown algebraically (although we shall not do so here) that, in the presence of inhibitor the value of  $K_M$  increases by the factor

$$1 + \frac{[I]}{K_I}$$

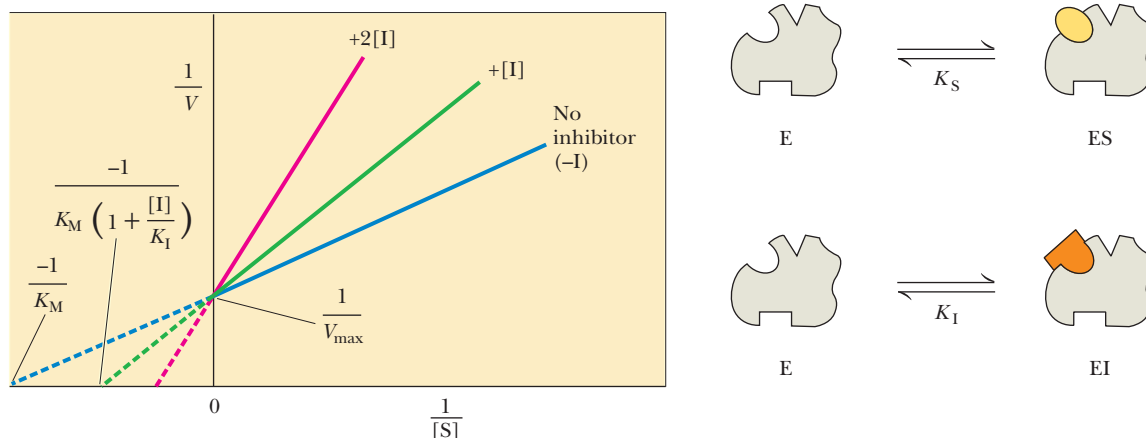
If we substitute  $K_M (1 + [I]/K_I)$  for  $K_M$  in Equation 6.17, we obtain

$$\frac{1}{V} = \frac{k_M}{V_{\max}} \times \frac{1}{[S]} + \frac{1}{V_{\max}}$$

$$\frac{1}{V} = \frac{k_M}{V_{\max}} \left( 1 + \frac{[I]}{K_I} \right) \times \frac{1}{[S]} + \frac{1}{V_{\max}}$$

$$y = m \times x + b \quad (6.18)$$

Here the term  $1/V$  takes the place of the  $y$  coordinate, and the term  $1/[S]$  takes the place of the  $x$  coordinate, as was the case in Equation 6.17. The intercept  $1/V_{\max}$ , the  $b$  term in the equation for a straight line, has not changed from the



■ **FIGURE 6.12** A Lineweaver–Burk double-reciprocal plot of enzyme kinetics for competitive inhibition.

# Biochemical Connections

## PHYSICAL ORGANIC CHEMISTRY

### Practical Information from Kinetic Data

The mathematics of enzyme kinetics can certainly look challenging. In fact, an understanding of kinetic parameters can often provide key information about the role of an enzyme within a living organism. Many of the ways of doing kinetic plots of this sort were developed by physical organic chemists, who then went on to propose mechanisms for reactions of all sorts based on kinetic data (see Section 7.6). Three aspects are useful: comparison of  $K_M$ , comparison of  $k_{\text{cat}}$  or turnover number, and comparison of  $k_{\text{cat}}/K_M$  ratios.

#### Comparison of $K_M$

Let us start by comparing the values of  $K_M$  for two enzymes that catalyze an early step in the breakdown of sugars: hexokinase and glucokinase. Both enzymes catalyze the formation of a phosphate ester linkage to a hydroxyl group of a sugar. Hexokinase can use any one of several six-carbon sugars, including glucose and fructose, the two components of sucrose (common table sugar), as substrates. Glucokinase is an isozyme of hexokinase that is primarily involved in glucose metabolism. The  $K_M$  for hexokinase is 0.15 mM for glucose and 1.5 mM for fructose.

The  $K_M$  for glucokinase, a liver-specific enzyme, is 20 mM. (We shall use the expression  $K_M$  here, even though some hexokinases studied do not follow Michaelis–Menten kinetics, and the term  $[S]_{0.5}$  might be more appropriate. Not all enzymes have a  $K_M$ , but they do all have a substrate concentration that gives rise to  $V_{\text{max}}/2$ .)

Comparison of these numbers tells us a lot about sugar metabolism. Because the resting level for blood glucose is about 5 mM, hexokinase would be expected to be fully active for all body cells. The liver would not be competing with the other cells for glucose. However, after a carbohydrate-rich meal, blood glucose levels often exceed 10 mM, and, at that concentration, the liver glucokinase would have reasonable activity. Furthermore, because the enzyme is found only in the liver, the excess glucose will be preferentially taken into the liver, where it can be stored as glycogen until it is needed. Also, the comparison of the two sugars for hexokinase indicates clearly that glucose is preferred over fructose as a nutrient.

#### Comparison of Turnover Number

As can be seen from Table 6.2, the first two enzymes are very reactive; catalase has one of the highest turnover numbers of all known enzymes. These high numbers allude to their importance

in detoxifying hydrogen peroxide and preventing formation of  $\text{CO}_2$  bubbles in the blood; these are their respective reactions. The values for chymotrypsin and acetylcholinesterase are within the range for “normal” metabolic enzymes. Lysozyme is an enzyme that degrades certain polysaccharide components of bacterial cell walls. It is present in many body tissues. Its low catalytic efficiency indicates that it operates well enough to catalyze polysaccharide degradation under normal conditions.

#### Comparison of $k_{\text{cat}}/K_M$

Even though  $k_{\text{cat}}$  alone is indicative of the catalytic efficiency under saturating substrate conditions,  $[S]$  is rarely saturating under physiological conditions for many enzymes. The *in vivo* ratio of  $[S]/K_M$  is often in the range of 0.01 to 1, meaning that active sites are not filled with substrate. Under these conditions, the level of substrate is small, and the amount of free enzyme approximates the level of total enzyme, because most of it is not bound to substrate. The Michaelis–Menten equation can be rewritten in the following form:

$$V = \frac{V_{\text{max}} [S]}{K_M + [S]} = \frac{k_{\text{cat}} [E_T][S]}{K_M + [S]}$$

If we then replace  $E_T$  with  $E$  and assume that  $[S]$  is negligible compared with  $K_M$ , we can rewrite the equation as follows:

$$V = (k_{\text{cat}}/K_M) [E][S]$$

Thus, under these conditions, the ratio of  $k_{\text{cat}}$  to  $K_M$  is a second-order rate constant and provides a measure of the catalytic efficiency of the enzyme under nonsaturating conditions. The ratio of  $k_{\text{cat}}$  to  $K_M$  is much more constant between different enzymes than either  $K_M$  or  $k_{\text{cat}}$  alone. Looking at the first three enzymes in Table 6.2, we can see that the  $k_{\text{cat}}$  values vary over a range of nearly 3000. The  $K_M$  values vary over a range of nearly 300. When the ratio of  $k_{\text{cat}}$  to  $K_M$  is compared, however, the range is only 4. The upper limit of a second-order rate constant is dependent on the diffusion-controlled limit of how fast the  $E$  and  $S$  can come together. The diffusion limit in an aqueous environment is in the range of  $10^8$  to  $10^9$ . Many enzymes have evolved to have  $k_{\text{cat}}$  to  $K_M$  ratios that allow reactions to proceed at these limiting rates. This is referred to as being catalytically perfect.

earlier equation, but the slope  $K_M/V_{\text{max}}$  in Equation 6.17 has increased by the factor  $(1 + [I]/K_I)$ . The slope, the  $m$  term in the equation for a straight line, is now

$$\frac{K_M}{V_{\text{max}}} \left( 1 + \frac{[I]}{K_I} \right)$$

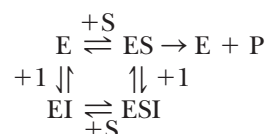
accounting for the changes in the slope of the Lineweaver–Burk plot. Note that the  $y$  intercept does not change. This algebraic treatment of competitive inhibition agrees with experimental results, validating the model, just as experimental results validate the underlying Michaelis–Menten model for enzyme action. It is important to remember that the most distinguishing characteristic of a competitive inhibitor is that substrate or inhibitor can bind the enzyme, but not both. Because both are vying for the same location, sufficiently high

substrate will “outcompete” the inhibitor. This is why  $V_{\max}$  does not change; it is a measure of the velocity at infinite [substrate].

### How can we identify a noncompetitive inhibitor?

The kinetic results of noncompetitive inhibition differ from those of competitive inhibition. The Lineweaver–Burk plots for a reaction in the presence and absence of a noncompetitive inhibitor show that both the slope and the y intercept change for the inhibited reaction (Figure 6.13), without changing the x intercept. The value of  $V_{\max}$  decreases, but that of  $K_M$  remains the same; the inhibitor does not interfere with the binding of substrate to the active site. Increasing the substrate concentration cannot overcome noncompetitive inhibition because the inhibitor and substrate are not competing for the same site.

The reaction pathway has become considerably more complicated, and several equilibria must be considered.



In the presence of a noncompetitive inhibitor, I, the maximum velocity of the reaction,  $V_{\max}^I$ , has the form (we shall not do the derivation here)

$$V_{\max}^I = \frac{V_{\max}}{1 + [I]/K_I}$$

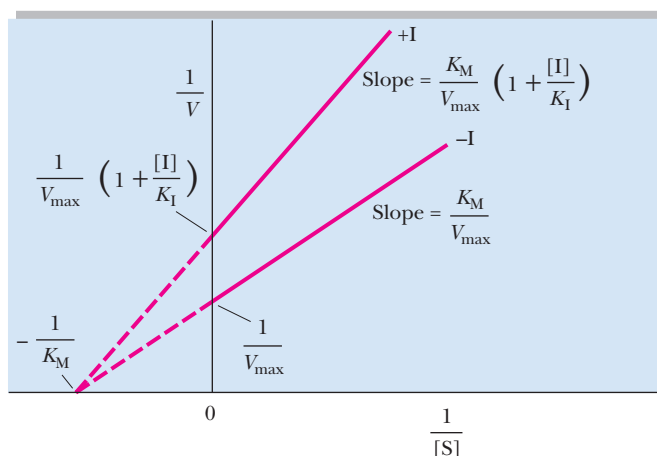
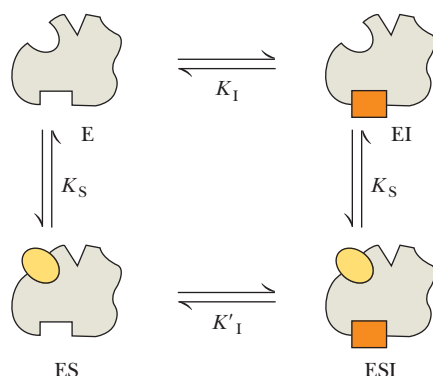
where  $K_I$  is again the dissociation constant for the enzyme–inhibitor complex, EI. Recall that the maximum rate,  $V_{\max}$ , appears in the expressions for both the slope and the intercept in the equation for the Lineweaver–Burk plot (Equation 6.17):

$$\begin{aligned} \frac{1}{V} &= \frac{K_M}{V_{\max}} \times \frac{1}{[S]} + \frac{1}{V_{\max}} \\ y &= m \times x + b \end{aligned}$$

In noncompetitive inhibition, we replace the term  $V_{\max}$  with the expression for  $V_{\max}^I$ , to obtain

$$\begin{aligned} \frac{1}{V} &= \frac{K_M}{V_{\max}} \left( 1 + \frac{[I]}{K_I} \right) \times \frac{1}{[S]} + \frac{1}{V_{\max}} \left( 1 + \frac{[I]}{K_I} \right) \\ y &= m \times x + b \end{aligned} \quad (6.19)$$

Noncompetitive inhibition



■ **FIGURE 6.13** A Lineweaver–Burk plot of enzyme kinetics for noncompetitive inhibition.

The expressions for both the slope and the intercept in the equation for a Lineweaver–Burk plot of an uninhibited reaction have been replaced by more complicated expressions in the equation that describes noncompetitive inhibition. This interpretation is borne out by the observed results. With a pure, noncompetitive inhibitor, the binding of substrate does not affect the binding of inhibitor, and vice versa. Because the  $K_M$  is a measure of the affinity of the enzyme and substrate, and because the inhibitor does not affect the binding, the  $K_M$  does not change with noncompetitive inhibition.

The two types of inhibition presented here are the two extreme cases. There are many other types of inhibition. **Uncompetitive inhibition** is seen when an inhibitor can bind to the ES complex but not to free E. A Lineweaver–Burk plot of an uncompetitive inhibitor shows parallel lines. The  $V_{\max}$  decreases and the apparent  $K_M$  decreases as well. Noncompetitive inhibition is actually a limiting case of a more general inhibition type called **mixed inhibition**. With a mixed inhibitor, the same binding diagram is seen as in the preceding equilibrium equations but, in this case, the binding of inhibitor does affect the binding of substrate and vice versa. A Lineweaver–Burk plot of an enzyme plus mixed inhibitor gives lines that intersect in the left-hand quadrant of the graph. The  $K_M$  increases, and the  $V_{\max}$  decreases.

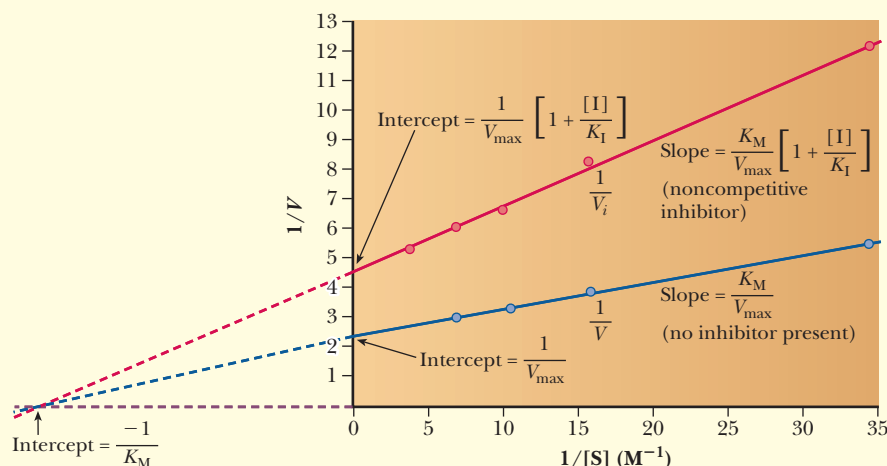
### Apply Your Knowledge

Sucrose (common table sugar) is hydrolyzed to glucose and fructose (Section 16.3) in a classic experiment in kinetics. The reaction is catalyzed by the enzyme invertase. Using the following data, determine, by the Lineweaver–Burk method, whether the inhibition of this reaction by 2 M urea is competitive or noncompetitive.

Sucrose Concentration (mol L <sup>-1</sup> )	V, No Inhibitor (arbitrary units)	V, Inhibitor Present (same arbitrary units)
0.0292	0.182	0.083
0.0584	0.265	0.119
0.0876	0.311	0.154
0.117	0.330	0.167
0.175	0.372	0.192

### Solution

Plot the data with the reciprocal of the sucrose concentration on the  $x$  axis and the reciprocals of the two reaction velocities on the  $y$  axis. Note that the two plots have different slopes and different  $y$  intercepts, typical of noncompetitive inhibition. Note the same intercept on the negative  $x$  axis, which gives  $-1/K_M$ .



# Biochemical Connections MEDICINE

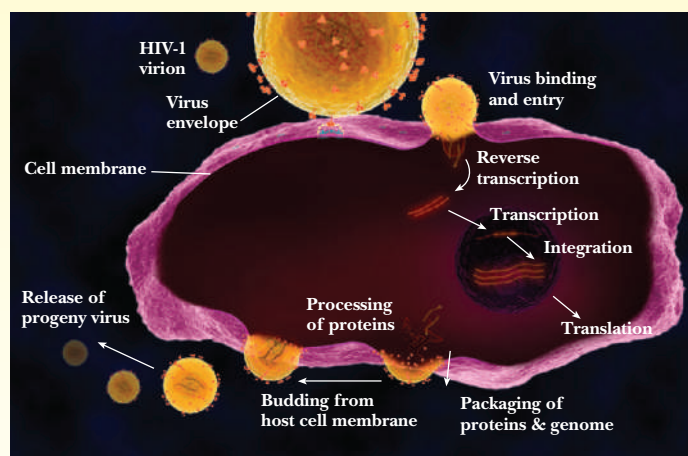
## Enzyme Inhibition in the Treatment of AIDS

A key strategy in the treatment of acquired immunodeficiency syndrome (AIDS) has been to develop specific inhibitors that selectively block the actions of enzymes unique to the human immunodeficiency virus (HIV), which causes AIDS. Many laboratories are working on this approach to the development

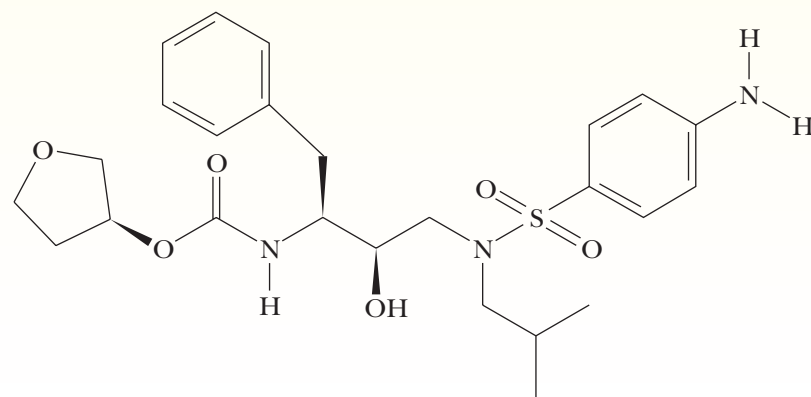
of therapeutic agents. Three key enzymes are current targets for AIDS therapy—reverse transcriptase, integrase, and protease.

One of the most important target enzymes is HIV protease, an enzyme essential to the production of new virus particles in infected cells. HIV protease is unique to this virus. It catalyzes the processing of viral proteins in an infected cell. Without these proteins, viable virus particles cannot be released to cause further infection. The structure of HIV protease, including its active site, was known from the results of X-ray crystallography. With this structure in mind, scientists have designed and synthesized compounds to bind to the active site. Improvements were made in the drug design by obtaining structures of a series of inhibitors bound to the active site of HIV protease. These structures were also elucidated by X-ray crystallography. This process eventually led to several compounds marketed by different pharmaceutical companies. These HIV protease inhibitors include saquinavir from Hoffman-LaRoche, ritonavir from Abbott Laboratories, indinavir from Merck, Viracept from Pfizer, and amprenavir from Vertex Pharmaceuticals. (These companies maintain highly informative home pages on the World Wide Web.)

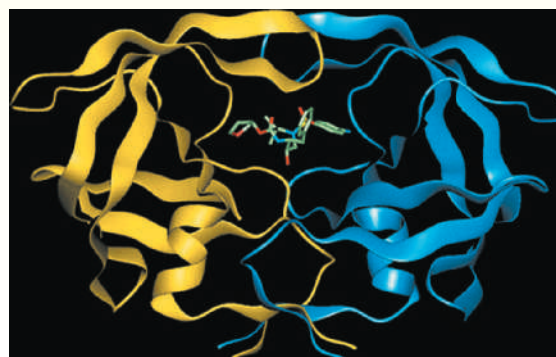
The most recent target is the viral enzyme called integrase, which is needed for the virus to copy itself in the host cell. A recent drug made by Merck, called MK-0518, inhibits the integrase enzyme. Treatment of AIDS is most effective when a combination of drug therapies is used, and HIV protease, integrase, and reverse transcriptase inhibitors play an important role. The use of multiple inhibitors for the key viral enzymes allows levels of each to remain below toxic levels to the cell.



■ From the time a cell is infected with HIV, three key enzymes are involved in the replication of the virus—reverse transcriptase, integrase, and protease. Reprinted by permission from *Science* 311, 943 (2006).



■ Structure of amprenavir (VX-478), an HIV protease inhibitor developed by Vertex Pharmaceuticals. (Vertex Pharmaceuticals, Inc.)



■ Active site of VX-478 complexed with HIV-1 protease.

## SUMMARY

### If a reaction is spontaneous, does that mean it will be fast?

Thermodynamic spontaneity cannot tell us whether a reaction will be fast. The speed of a reaction is a kinetic property controlled by the nature of the energy state of the ES complex and the transition state. Enzymes speed up the reaction rate by creating a situation where the distance between the

transition state and the ES complex on an energy diagram is reduced.

**Will a reaction go faster if you raise the temperature?** A chemical reaction may go faster at higher temperatures. However, when the reaction is catalyzed by an enzyme, this is true only



for a specific range of temperatures. If the temperature is raised too much, it denatures the enzyme and the rate of reaction is reduced significantly, perhaps to zero.

**Is the rate of a reaction always based on the concentration of reactants?** In many situations the concentration of the reactants does influence the rate of an enzyme-catalyzed reaction. However, if there is very little enzyme and a saturating amount of substrate, then all of the enzyme molecules are bound to substrate. Adding additional substrate under this condition will not increase the rate of reaction. When this happens, the enzyme is already working at its  $V_{\max}$  and is exhibiting zero-order kinetics.

**Why do enzymes bind to substrates?** Enzymes and substrates are attracted to each other via noncovalent interactions, such as electrostatic attractions. The active site of an enzyme has amino acids in a specific orientation where they can bind to the substrate. The energy diagram will show that the energy of the ES complex is less than the energy of the  $E + S$  alone.

**Why do chymotrypsin and ATCase have different velocity curves?** Chymotrypsin and aspartate transcarbamoylase exhibit different types of kinetics. Chymotrypsin is a nonallosteric enzyme and exhibits hyperbolic kinetics. ATCase is an

allosteric enzyme. It has multiple subunits, and the binding of one molecule of substrate affects the binding of the next molecule of substrate. It exhibits sigmoidal kinetics.

**How do we calculate  $K_M$  and  $V_{\max}$  from a graph?**  $K_M$  and  $V_{\max}$  can be estimated by plotting the velocity versus  $[S]$ . However, a more accurate way is to make a Lineweaver–Burk plot of  $1/V$  versus  $1/[S]$ . With such a graph, the  $y$  intercept yields  $1/V_{\max}$ , which can then be converted to  $V_{\max}$ . The  $x$  intercept is  $-1/K_M$ , which can also be converted to  $K_M$ .

**What is the significance of  $K_M$  and  $V_{\max}$ ?** Mathematically,  $K_M$  is equal to the substrate concentration that yields a velocity of  $V_{\max}/2$ . It is also a crude measure of the affinity between the enzyme and substrate, where a low  $K_M$  indicates a high affinity.  $V_{\max}$  tells us how fast the enzyme can generate product under saturating substrate conditions.

**How can we identify a competitive inhibitor?** Comparing a Lineweaver–Burk plot of an uninhibited reaction to one for an inhibited reaction, one can identify the inhibitor as competitive if the curves intersect on the  $y$  axis.

**How can we identify a noncompetitive inhibitor?** With a noncompetitive inhibitor, a Lineweaver–Burk plot shows lines that intersect on the  $x$  axis.

## REVIEW EXERCISES

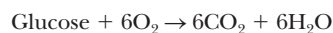
**OWL** Interactive versions of these problems are assignable in OWL

### 6.1 Enzymes Are Effective Biological Catalysts

- Recall** How does the catalytic effectiveness of enzymes compare with that of nonenzymatic catalysts?
- Recall** Are all enzymes proteins?
- Mathematical** Catalase breaks down hydrogen peroxide about  $10^7$  times faster than the uncatalyzed reaction. If the latter required one year, how much time would be needed by the catalase-catalyzed reaction?
- Reflect and Apply** Give two reasons why enzyme catalysts are  $10^3$  to  $10^5$  more effective than reactions that are catalyzed by, for example, simple  $H^+$  or  $OH^-$ .

### 6.2 Kinetics versus Thermodynamics

- Recall** For the reaction of glucose with oxygen to produce carbon dioxide and water,



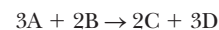
the  $\Delta G^\circ$  is  $-2880 \text{ kJ mol}^{-1}$ , a strongly exergonic reaction. However, a sample of glucose can be maintained indefinitely in an oxygen-containing atmosphere. Reconcile these two statements.

- Reflect and Apply** Would nature rely on the same enzyme to catalyze a reaction either way (forward or backward) if the  $\Delta G^\circ$  were  $-0.8 \text{ kcal mol}^{-1}$ ? If it were  $-5.3 \text{ kcal mol}^{-1}$ ?
- Reflect and Apply** Suggest a reason why heating a solution containing an enzyme markedly decreases its activity. Why is the decrease of activity frequently much less when the solution contains high concentrations of the substrate?
- Reflect and Apply** A model is proposed to explain the reaction catalyzed by an enzyme. Experimentally obtained rate data fit the model to within experimental error. Do these findings prove the model?

- Reflect and Apply** Does the presence of a catalyst alter the standard free energy change of a chemical reaction?
- Reflect and Apply** What effect does a catalyst have on the activation energy of a reaction?
- Reflect and Apply** An enzyme catalyzes the formation of ATP from ADP and phosphate ion. What is its effect on the rate of hydrolysis of ATP to ADP and phosphate ion?
- Reflect and Apply** Can the presence of a catalyst increase the amount of product obtained in a reaction?

### 6.3 Enzyme Kinetic Equations

- Recall** For the hypothetical reaction



the rate was experimentally determined to be

$$\text{Rate} = k[A]^1[B]^1$$

What is the order of the reaction with respect to A? With respect to B? What is the overall order of the reaction? Suggest how many molecules each of A and B are likely to be involved in the detailed mechanism of the reaction.

- Reflect and Apply** The enzyme lactate dehydrogenase catalyzes the reaction



NADH absorbs light at 340 nm in the near-ultraviolet region of the electromagnetic spectrum, but  $\text{NAD}^+$  does not. Suggest an experimental method for following the rate of this reaction, assuming that you have available a spectrophotometer capable of measuring light at this wavelength.

15. **Reflect and Apply** Would you use a pH meter to monitor the progress of the reaction described in Question 14? Why or why not?
16. **Reflect and Apply** Suggest a reason for carrying out enzymatic reactions in buffer solutions.

### 6.4 Enzyme–Substrate Binding

17. **Recall** Distinguish between the lock-and-key and induced-fit models for binding of a substrate to an enzyme.
18. **Recall** Using an energy diagram, show why the lock-and-key model could lead to an inefficient enzyme mechanism. *Hint:* Remember that the distance to the transition state must be minimized for an enzyme to be an effective catalyst.
19. **Reflect and Apply** Other things being equal, what is a potential disadvantage of an enzyme having a very high affinity for its substrate?
20. **Reflect and Apply** Amino acids that are far apart in the amino acid sequence of an enzyme can be essential for its catalytic activity. What does this suggest about its active site?
21. **Reflect and Apply** If only a few of the amino acid residues of an enzyme are involved in its catalytic activity, why does the enzyme need such a large number of amino acids?
22. **Reflect and Apply** A chemist synthesizes a new compound that may be structurally analogous to the transition-state species in an enzyme-catalyzed reaction. The compound is experimentally shown to inhibit the enzymatic reaction strongly. Is it likely that this compound is indeed a transition-state analogue?

### 6.5 Examples of Enzyme-Catalyzed Reactions

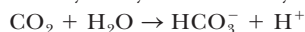
23. **Recall** Show graphically the dependence of reaction velocity on substrate concentration for an enzyme that follows Michaelis–Menten kinetics and for an allosteric enzyme.
24. **Recall** Do all enzymes display kinetics that obey the Michaelis–Menten equation? Which ones do not?
25. **Recall** How can you recognize an enzyme that does not display Michaelis–Menten kinetics?

### 6.6 The Michaelis–Menten Approach to Enzyme Kinetics

26. **Recall** Show graphically how the reaction velocity depends on the enzyme concentration. Can a reaction be saturated with enzyme?
27. **Recall** Define *steady state*, and comment on the relevance of this concept to theories of enzyme reactivity.
28. **Recall** How is the turnover number of an enzyme related to  $V_{\max}$ ?
29. **Mathematical** For an enzyme that displays Michaelis–Menten kinetics, what is the reaction velocity,  $V$  (as a percentage of  $V_{\max}$ ), observed at the following values?
- $[S] = K_M$
  - $[S] = 0.5K_M$
  - $[S] = 0.1K_M$
  - $[S] = 2K_M$
  - $[S] = 10K_M$
30. **Mathematical** Determine the values of  $K_M$  and  $V_{\max}$  for the decarboxylation of a  $\beta$ -keto acid given the following data.

Substrate Concentration (mol L <sup>-1</sup> )	Velocity (mM min <sup>-1</sup> )
2.500	0.588
1.000	0.500
0.714	0.417
0.526	0.370
0.250	0.256

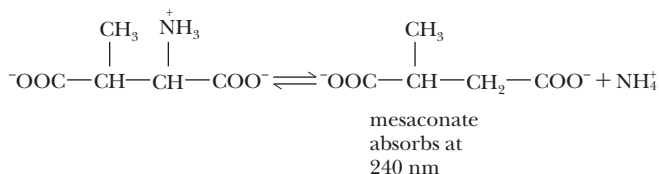
31. **Mathematical** The kinetic data in the following table were obtained for the reaction of carbon dioxide and water to produce bicarbonate and hydrogen ion catalyzed by carbonic anhydrase:



[H. De Voe and G. B. Kistiakowsky, *J. Am. Chem. Soc.* **83**, 274 (1961)]. From these data, determine  $K_M$  and  $V_{\max}$  for the reaction.

Carbon Dioxide Concentration (mmol L <sup>-1</sup> )	1/Velocity (M <sup>-1</sup> sec)
1.25	$36 \times 10^3$
2.5	$20 \times 10^3$
5.0	$12 \times 10^3$
20.0	$6 \times 10^3$

32. **Mathematical** The enzyme  $\beta$ -methylaspartase catalyzes the deamination of  $\beta$ -methylaspartate



[V. Williams and J. Selbin, *J. Biol. Chem.* **239**, 1636 (1964)]. The rate of the reaction was determined by monitoring the absorbance of the product at 240 nm ( $A_{240}$ ). From the data in the following table, determine  $K_M$  for the reaction. How does the method of calculation differ from that in Questions 30 and 31?

Substrate Concentration (mmol L <sup>-1</sup> )	Velocity ( $\Delta A_{240} \text{ min}^{-1}$ )
0.002	0.045
0.005	0.115
0.020	0.285
0.040	0.380
0.060	0.460
0.080	0.475
0.100	0.505

33. **Mathematical** The hydrolysis of a phenylalanine-containing peptide is catalyzed by  $\alpha$ -chymotrypsin with the following results. Calculate  $K_M$  and  $V_{\max}$  for the reaction.

Peptide Concentration (M)	Velocity (M min <sup>-1</sup> )
$2.5 \times 10^{-4}$	$2.2 \times 10^{-6}$
$5.0 \times 10^{-4}$	$5.8 \times 10^{-6}$
$10.0 \times 10^{-4}$	$5.9 \times 10^{-6}$
$15.0 \times 10^{-4}$	$7.1 \times 10^{-6}$

34. **Mathematical** For the  $V_{\max}$  obtained in Question 30, calculate the turnover number (catalytic rate constant) assuming that  $1 \times 10^{-4}$  mol of enzyme were used.
35. **Mathematical** You do an enzyme kinetic experiment and calculate a  $V_{\max}$  of 100  $\mu\text{mol}$  of product per minute. If each assay used 0.1 mL of an enzyme solution that had a concentration of 0.2 mg/mL, what would be the turnover number if the enzyme had a molecular weight of 128,000 g/mol?
36. **Reflect and Apply** The enzyme D-amino acid oxidase has a very high turnover number because the D-amino acids are potentially toxic. The  $K_M$  for the enzyme is in the range of 1 to 2 mM for the aromatic amino acids and in the range of 15 to 20 mM for such amino acids as serine, alanine, and the acidic amino acids. Which of these amino acids are the preferred substrates for the enzyme?
37. **Reflect and Apply** Why is it useful to plot rate data for enzymatic reactions as a straight line rather than as a curve?
38. **Reflect and Apply** Under what conditions can we assume that  $K_M$  indicates the binding affinity between substrate and enzyme?

39. **Biochemical Connections** Why does acetazolamide make beer taste flat?
40. **Biochemical Connections** How did scientists determine that carbonic anhydrase is a chemical sensor for  $\text{CO}_2$ ?
41. **Biochemical Connections** How do the  $K_M$  values for glucokinase and hexokinase reflect their roles in sugar metabolism?
42. **Biochemical Connections** When does the  $k_{\text{cat}}/K_M$  value approximate the catalytic efficiency of an enzyme?

## 6.7 Enzyme Inhibition

43. **Recall** How can competitive and noncompetitive inhibition be distinguished in terms of  $K_M$ ?
44. **Recall** Why does a competitive inhibitor not change  $V_{\text{max}}$ ?
45. **Recall** Why does a noncompetitive inhibitor not change the observed  $K_M$ ?
46. **Recall** Distinguish between the molecular mechanisms of competitive and noncompetitive inhibition.
47. **Recall** Can enzyme inhibition be reversed in all cases?
48. **Recall** Why is a Lineweaver–Burk plot useful in analyzing kinetic data from enzymatic reactions?
49. **Recall** Where do lines intersect on a Lineweaver–Burk plot showing competitive inhibition? On a Lineweaver–Burk plot showing noncompetitive inhibition?
50. **Mathematical** Draw Lineweaver–Burk plots for the behavior of an enzyme for which the following experimental data are available.

[S] (mM)	$V$ , No Inhibitor (mmol min <sup>-1</sup> )	$V$ , Inhibitor Present (mmol min <sup>-1</sup> )
3.0	4.58	3.66
5.0	6.40	5.12
7.0	7.72	6.18
9.0	8.72	6.98
11.0	9.50	7.60

What are the  $K_M$  and  $V_{\text{max}}$  values for the inhibited and uninhibited reactions? Is the inhibitor competitive or noncompetitive?

51. **Mathematical** For the following aspartase reaction (see Question 32) in the presence of the inhibitor hydroxymethylaspartate, determine  $K_M$  and whether the inhibition is competitive or noncompetitive.

[S] (molarity)	$V$ , No Inhibitor (arbitrary units)	$V$ , Inhibitor Present (same arbitrary units)
$1 \times 10^{-4}$	0.026	0.010
$5 \times 10^{-4}$	0.092	0.040
$1.5 \times 10^{-3}$	0.136	0.086
$2.5 \times 10^{-3}$	0.150	0.120
$5 \times 10^{-3}$	0.165	0.142

52. **Reflect and Apply** Is it good (or bad) that enzymes can be reversibly inhibited? Why?
53. **Reflect and Apply** Noncompetitive inhibition is a limiting case in which the effect of binding inhibitor has no effect on the affinity for the substrate and vice versa. Suggest what a Lineweaver–Burk plot would look like for an inhibitor that had a reaction scheme similar to that on page 158 (noncompetitive inhibition reaction), but where binding inhibitor lowered the affinity of EI for the substrate.
54. **Biochemical Connections** You have been hired by a pharmaceutical company to work on development of drugs to treat AIDS. What information from this chapter will be useful to you?
55. **Reflect and Apply** Would you expect an irreversible inhibitor of an enzyme to be bound by covalent or by noncovalent interactions? Why?
56. **Reflect and Apply** Would you expect the structure of a noncompetitive inhibitor of a given enzyme to be similar to that of its substrate?
57. **Biochemical Connections** What part of the HIV lifecycle is disrupted by the drugs indinavir and amprenavir?
58. **Biochemical Connections** What part of the HIV lifecycle is disrupted by MK-0518?
59. **Biochemical Connections** How did scientists show that protein kinase M $\zeta$  was involved with memory?

# ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# The Behavior of Proteins: Enzymes, Mechanisms, and Control

# 7

Signals regulate the flow of traffic in much the same fashion as control mechanisms in chemical reactions.

## 7.1 The Behavior of Allosteric Enzymes

The behavior of many well-known enzymes can be described quite adequately by the Michaelis–Menten model, but allosteric enzymes behave very differently. In the last chapter, we saw similarities between the reaction kinetics of an enzyme such as chymotrypsin, which does not display allosteric behavior, and the binding of oxygen by myoglobin, which is also an example of non-allosteric behavior. The analogy extends to show the similarity in the kinetic behavior of an allosteric enzyme such as aspartate transcarbamoylase (ATCase) and the binding of oxygen by hemoglobin. Both ATCase and hemoglobin are allosteric proteins; the behaviors of both exhibit cooperative effects caused by subtle changes in quaternary structure. (Recall that *quaternary structure* is the arrangement in space that results from the interaction of subunits through noncovalent forces, and that *positive cooperativity* refers to the fact that the binding of low levels of substrate facilitates the action of the protein at higher levels of substrate, whether the action is catalytic or some other kind of binding.) In addition to displaying cooperative kinetics, allosteric enzymes have a different response to the presence of inhibitors from that of nonallosteric enzymes.

### How are allosteric enzymes controlled?

ATCase catalyzes the first step in a series of reactions in which the end product is cytidine triphosphate (CTP), a nucleoside triphosphate needed to make RNA and DNA (Chapter 9). The pathways that produce nucleotides are energetically costly and involve many steps. The reaction catalyzed by aspartate transcarbamoylase is a good example of how such a pathway is controlled to avoid overproduction of such compounds. For DNA and RNA synthesis, the levels of several nucleotide triphosphates are controlled. CTP is an inhibitor of ATCase, the enzyme that catalyzes the first reaction in the pathway. This behavior is an example of **feedback inhibition** (also called end-product inhibition), in which the end product of the sequence of reactions inhibits the first reaction in the series (Figure 7.1). Feedback inhibition is an efficient control mechanism because the entire series of reactions can be shut down when an excess of the final product exists, thus preventing the accumulation of intermediates in the pathway. Feedback inhibition is a general feature of metabolism and is not confined to allosteric enzymes. However, the observed kinetics of the ATCase reaction, including the mode of inhibition, are typical of allosteric enzymes.

When ATCase catalyzes the condensation of aspartate and carbamoyl phosphate to form carbamoyl aspartate, the graphical representation of the rate as a function of increasing substrate concentration (aspartate) is a sigmoidal curve rather than the hyperbola obtained with nonallosteric enzymes (Figure 7.2a). The sigmoidal curve indicates the cooperative behavior of allosteric enzymes. In this two-substrate reaction, aspartate is the substrate for which the concentration is varied, while the concentration of carbamoyl phosphate is kept constant at high levels.

## Chapter Outline

### 7.1 The Behavior of Allosteric Enzymes

- How are allosteric enzymes controlled?

### 7.2 The Concerted and Sequential Models for Allosteric Enzymes

- What is the concerted model for allosteric behavior?
- What is the sequential model for allosteric behavior?

### 7.3 Control of Enzyme Activity by Phosphorylation

- Does phosphorylation always increase enzyme activity?

### 7.4 Zymogens

### 7.5 The Nature of the Active Site

- How do we determine the essential amino acid residues?
- How does the architecture of the active site affect catalysis?
- How do the critical amino acids catalyze the chymotrypsin reaction?

### 7.6 Chemical Reactions Involved in Enzyme Mechanisms

- What are the most common types of reactions?

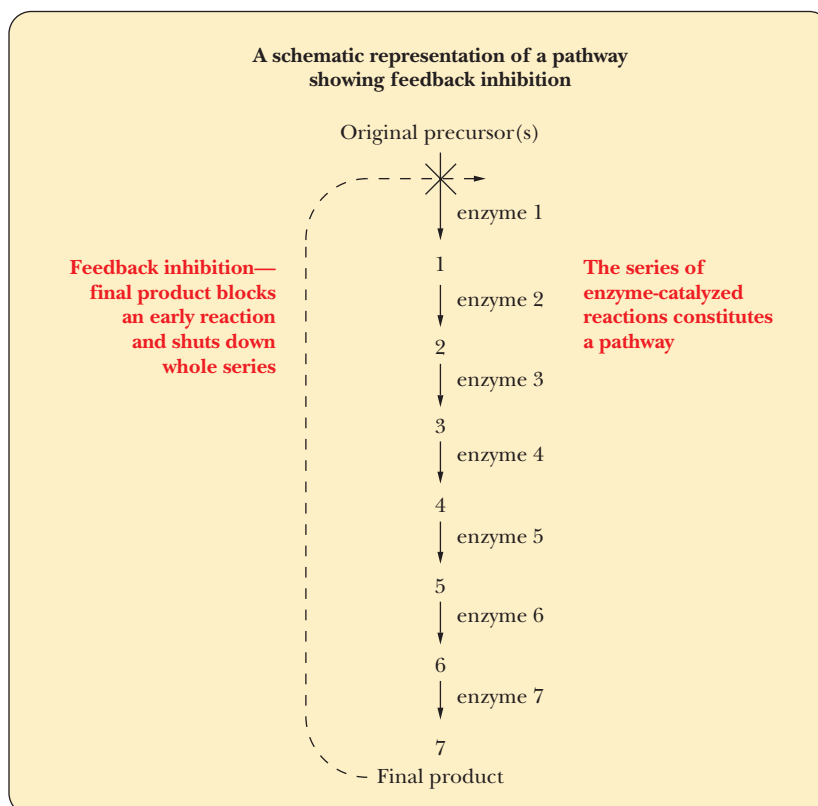
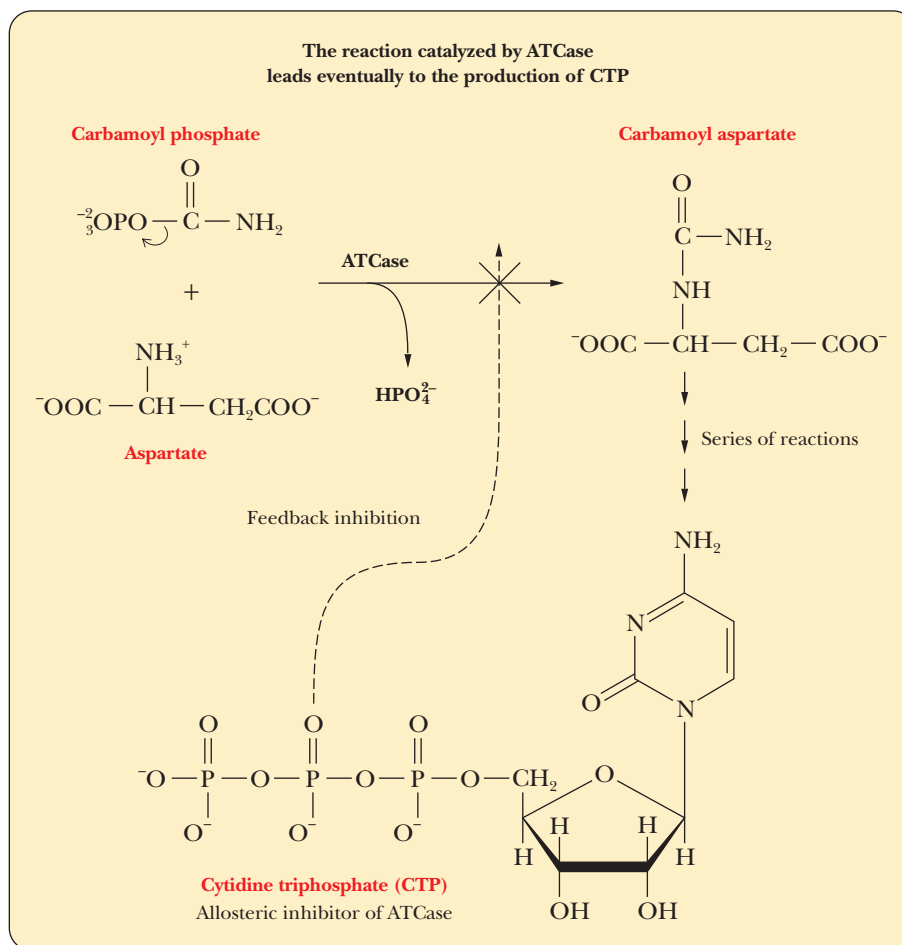
### 7.7 The Active Site and Transition States

- How do we determine the nature of the transition state?

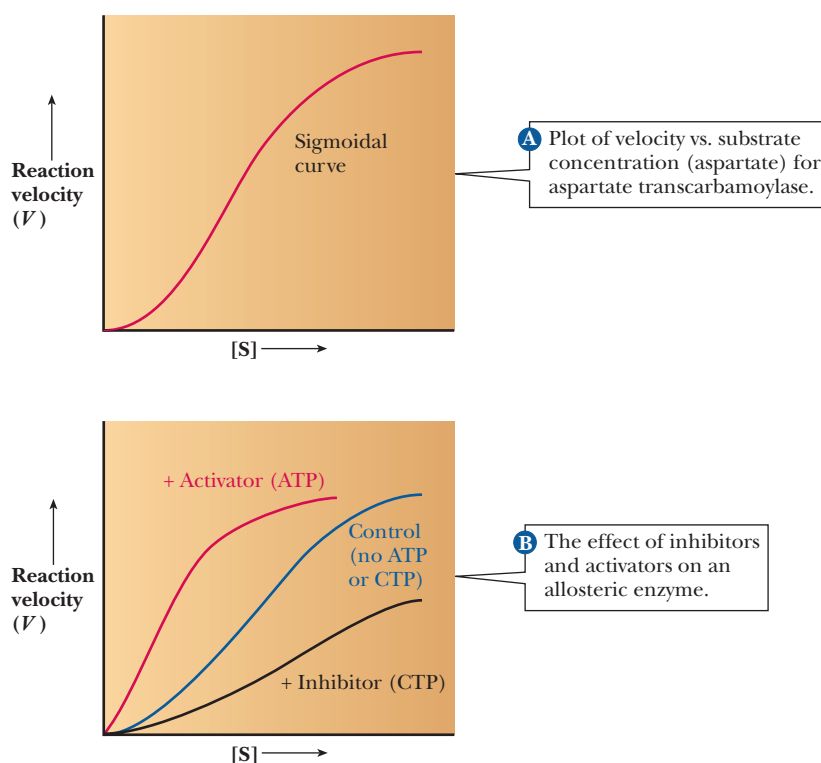
### 7.8 Coenzymes

**OWL** Online homework for this chapter may be assigned in OWL.





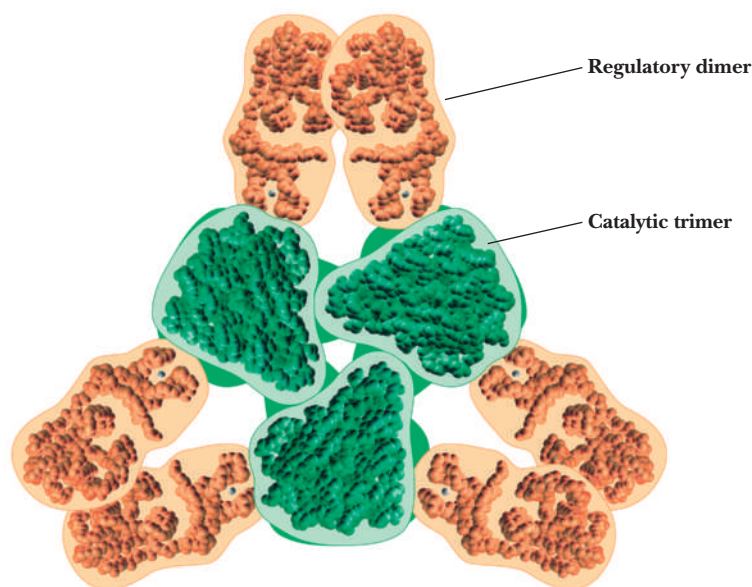
■ **FIGURE 7.1** Schematic representation of a pathway, showing feedback inhibition.



■ FIGURE 7.2

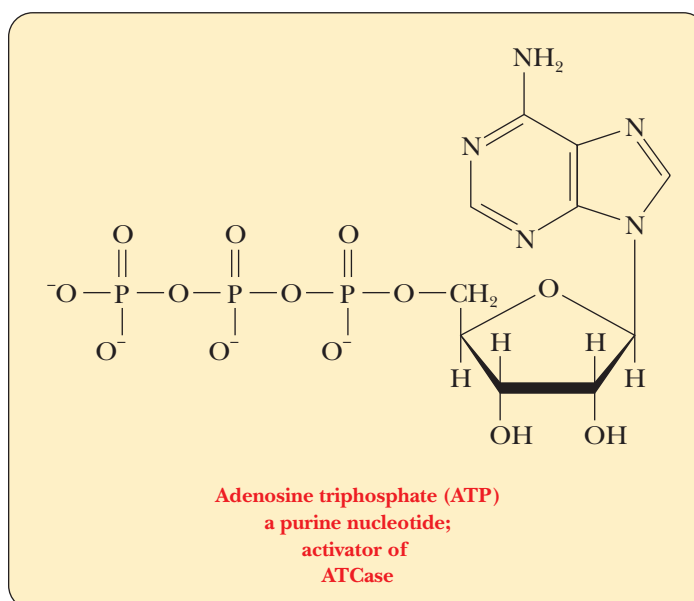
Figure 7.2b compares the rate of the uninhibited reaction of ATCase with the reaction rate in the presence of CTP. In the latter case, a sigmoidal curve still describes the rate behavior of the enzyme, but the curve is shifted to higher substrate levels; a higher concentration of aspartate is needed for the enzyme to achieve the same rate of reaction. At high substrate concentrations, the same maximal rate,  $V_{\max}$ , is observed in the presence and absence of **inhibitor**. (Recall this from Section 6.7.) Because in the Michaelis–Menten scheme  $V_{\max}$  changes when a reaction takes place in the presence of a noncompetitive inhibitor, noncompetitive inhibition cannot be the case here. The same Michaelis–Menten model associates this sort of behavior with competitive inhibition, but that part of the model still does not provide a reasonable picture. Competitive inhibitors bind to the same site as the substrate because they are very similar in structure. The CTP molecule is very *different* in structure from the substrate, aspartate, and it is bound to a different site on the ATCase molecule. ATCase is made up of two different types of subunits. One of them is the catalytic subunit, which consists of six protein subunits organized into two trimers. The other is the regulatory subunit, which also consists of six protein subunits organized into three dimers (Figure 7.3). The catalytic subunits can be separated from the regulatory subunits by treatment with *p*-hydroxymercuribenzoate, which reacts with the cysteines in the protein. When so treated, ATCase still catalyzes the reaction, but it loses its allosteric control by CTP, and the curve becomes hyperbolic.

The situation becomes “curiouser and curiouser” when the ATCase reaction takes place not in the presence of CTP, a pyrimidine nucleoside triphosphate, but in the presence of adenosine triphosphate (ATP), a purine nucleoside triphosphate. The structural similarities between CTP and ATP are apparent, but ATP is not a product of the pathway that includes the reaction of ATCase and that produces CTP. Both ATP and CTP are needed for the synthesis of RNA and DNA. The relative proportions of ATP and CTP are specified by the needs of the organism. If there is not enough CTP relative to the amount of ATP, the enzyme requires a signal to produce more. In the presence of ATP, the rate of



■ **FIGURE 7.3** Organization of aspartate transcarbamoylase, showing the two catalytic trimers and the three regulatory dimers.

the enzymatic reaction is increased at lower levels of aspartate, and the shape of the rate curve becomes less sigmoidal and more hyperbolic (Figure 7.2b). In other words, there is less cooperativity in the reaction. The binding site for ATP on the enzyme molecule is the same as that for CTP (which is not surprising in view of their structural similarity), but ATP is an activator rather than an inhibitor like CTP. When CTP is in short supply in an organism, the ATCase reaction is not inhibited, and the binding of ATP increases the activity of the enzyme still more.



Even though it is tempting to consider inhibition of allosteric enzymes in the same fashion as nonallosteric enzymes, much of the terminology is not appropriate. *Competitive inhibition* and *noncompetitive inhibition* are terms reserved for the enzymes that behave in line with Michaelis–Menten kinetics. With allosteric enzymes, the situation is more complex. In general, two types of enzyme systems exist, called **K systems** and **V systems**. A K system is an enzyme for which the substrate concentration that yields one-half  $V_{\max}$  is altered by the presence

of inhibitors or activators. ATCase is an example of a K system. Because we are not dealing with a Michaelis–Menten type of enzyme, the term  $K_M$  is not applicable. For an allosteric enzyme, the substrate level at one-half  $V_{\max}$  is called the  $K_{0.5}$ . In a V system, the effect of inhibitors and activators changes the  $V_{\max}$ , but not the  $K_{0.5}$ .

The key to allosteric behavior, including cooperativity and modifications of cooperativity, is the existence of multiple forms for the quaternary structures of allosteric proteins. The word *allosteric* is derived from *allo*, “other,” and *steric*, “shape,” referring to the fact that the possible conformations affect the behavior of the protein. The binding of substrates, inhibitors, and activators changes the quaternary structure of allosteric proteins, and the changes in structure are reflected in the behavior of those proteins. A substance that modifies the quaternary structure, and thus the behavior, of an allosteric protein by binding to it is called an **allosteric effector**. The term *effector* can apply to substrates, inhibitors, or activators. Several models for the behavior of allosteric enzymes have been proposed, and it is worthwhile to compare them.

Let us first define two terms. **Homotropic** effects are allosteric interactions that occur when several identical molecules are bound to a protein. The binding of substrate molecules to different sites on an enzyme, such as the binding of aspartate to ATCase, is an example of a homotropic effect. **Heterotropic** effects are allosteric interactions that occur when different substances (such as inhibitor and substrate) are bound to the protein. In the ATCase reaction, inhibition by CTP and activation by ATP are both heterotropic effects.

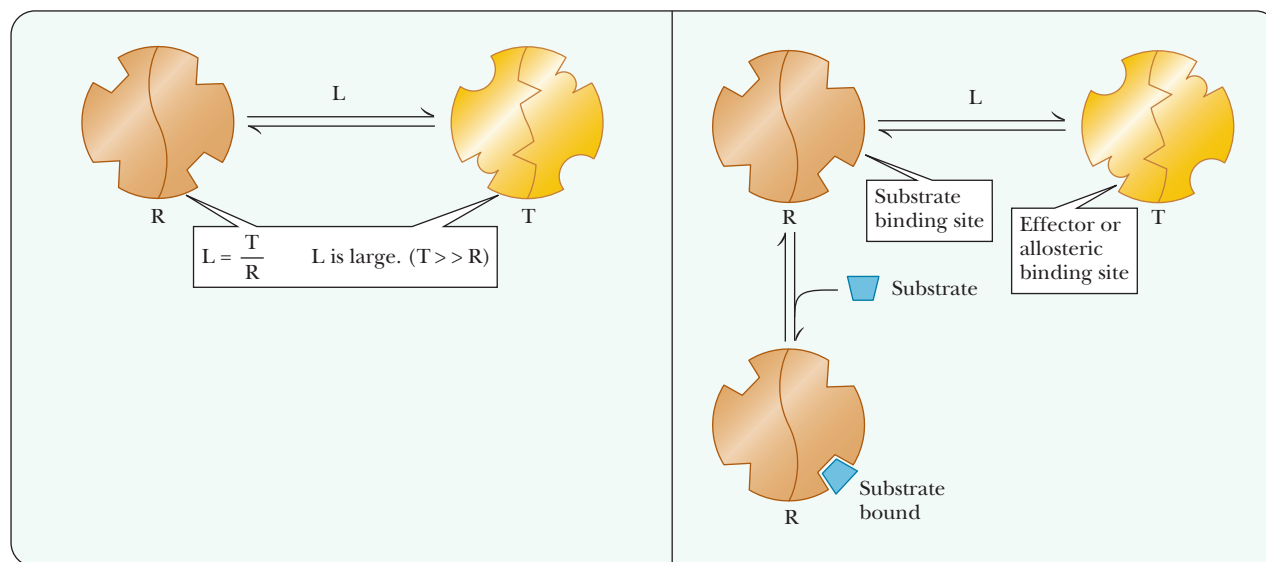
## 7.2 The Concerted and Sequential Models for Allosteric Enzymes

The two principal models for the behavior of allosteric enzymes are the concerted model and the sequential model. They were proposed in 1965 and 1966, respectively, and both are currently used as a basis for interpreting experimental results. The concerted model has the advantage of comparative simplicity, and it describes the behavior of some enzyme systems very well.

The sequential model sacrifices a certain amount of simplicity for a more realistic picture of the structure and behavior of proteins; it also deals very well with the behavior of some enzyme systems.

### What is the concerted model for allosteric behavior?

In 1965, Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeux proposed the **concerted model** for the behavior of allosteric proteins in a paper that has become a classic in the biochemical literature. In this picture, the protein has two conformations, the active R (relaxed) conformation, which binds substrate tightly, and the inactive T (tight, also called taut) conformation, which binds substrate less tightly. The distinguishing feature of this model is that the conformations of *all* subunits change simultaneously. Figure 7.4a shows a hypothetical protein with two subunits. Both subunits change conformation from the inactive T conformation to the active R conformation at the same time; that is, a concerted change of conformation occurs. The equilibrium ratio of the T/R forms is called  $L$  and is assumed to be high—that is, more enzyme is present in the unbound T form than in the unbound R form. The binding of substrate to either form can be described by the dissociation constant of the enzyme and substrate,  $K$ , with the affinity for substrate higher in the R form than in the T form. Thus,  $K_R \ll K_T$ . The ratio of  $K_R/K_T$  is called  $c$ . Figure 7.4b shows a limiting case in which  $K_T$  is infinitely greater than  $K_R$  ( $c = 0$ ).



**A** A dimeric protein can exist in either of two conformational states at equilibrium, the T (taut) form or the R (relaxed) form.  $L$  is the ratio of the T form to the R form. With most allosteric systems,  $L$  is large, so there is more enzyme present in the T form than in the R form.

**B** By Le Chatelier's principle, substrate binding shifts the equilibrium in favor of the relaxed state (R) by removing unbound R. The dissociation constant for the enzyme-substrate complex is  $K_R$  for the relaxed form and  $K_T$  for the taut form.  $K_R < K_T$ , so the substrate binds better to the relaxed form. The ratio of  $K_R/K_T$  is called  $c$ . This figure shows a limiting case in which the taut form does not bind substrate at all, in which case  $K_T$  is infinite and  $c = 0$ .

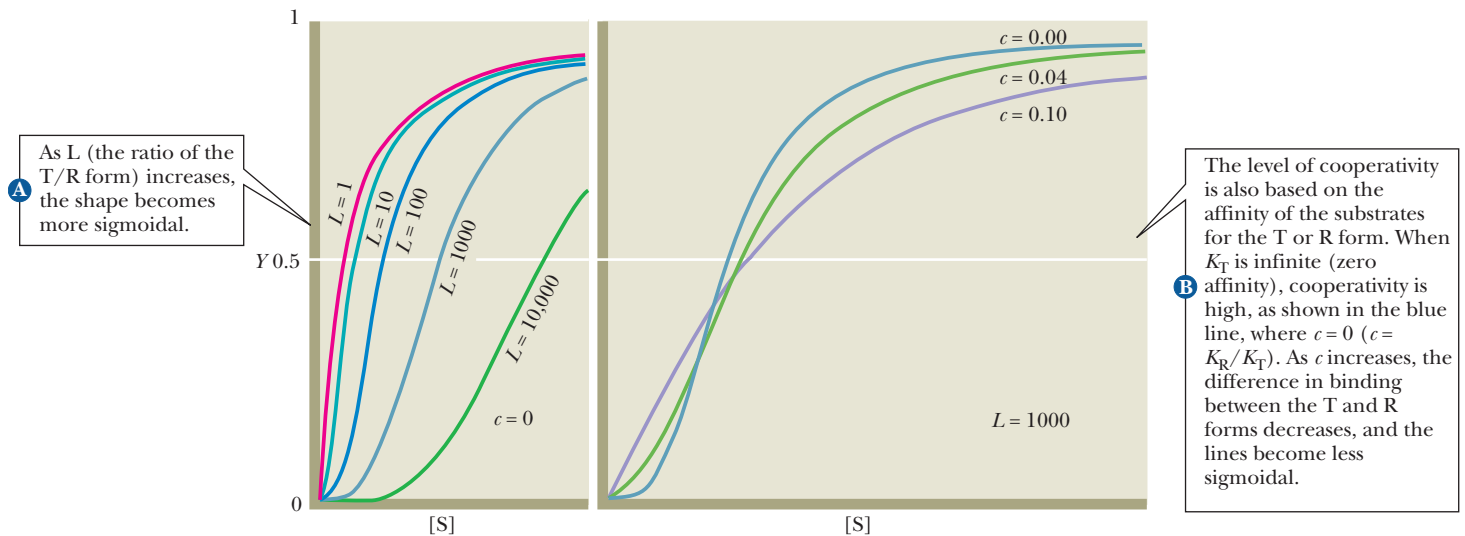
■ **FIGURE 7.4** Monod-Wyman-Changeux (MWC) model for allosteric transitions, also called the concerted model.

In other words, substrate will not bind to the T form at all. The allosteric effect is explained by this model based on perturbing the equilibrium between the T and R forms. Although initially the amount of enzyme in the R form is small, when substrate binds to the R form, it removes free R form. This causes the production of more R form to reestablish the equilibrium, which makes binding more substrate possible. This shifting of the equilibrium is responsible for the observed allosteric effects. The Monod-Wyman-Changeux model has been shown mathematically to explain the sigmoidal effects seen with allosteric enzymes. The shape of the curve will be based on the  $L$  and  $c$  values. As  $L$  increases (free T form more highly favored), the shape becomes more sigmoidal (Figure 7.5). As the value for  $c$  decreases (higher affinity between substrate and R form), the shape also becomes more sigmoidal.

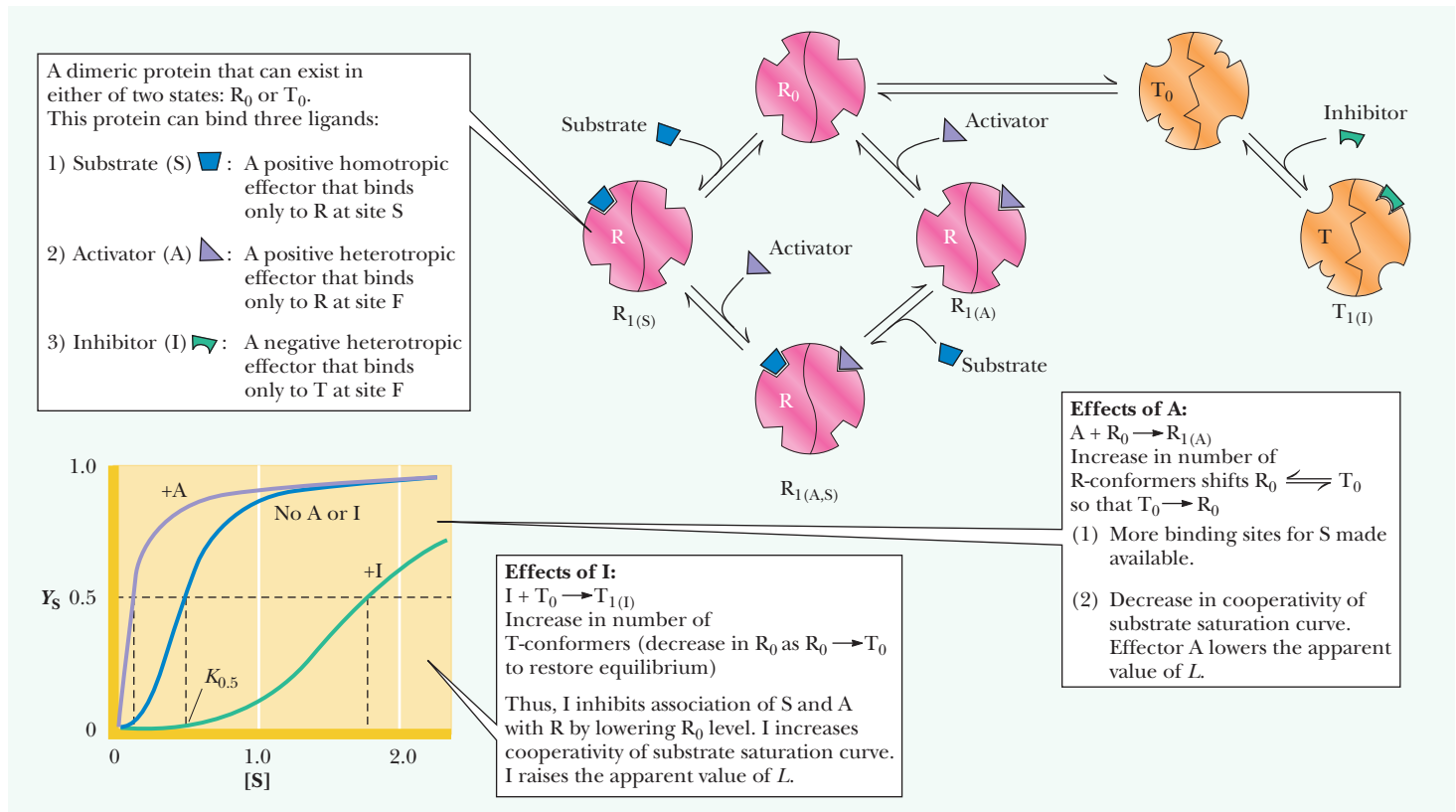
In the concerted model, the effects of inhibitors and activators can also be considered in terms of shifting the equilibrium between the T and R forms of the enzyme. The binding of inhibitors to allosteric enzymes is cooperative; allosteric inhibitors bind to and stabilize the T form of the enzyme. The binding of activators to allosteric enzymes is also cooperative; allosteric activators bind to and stabilize the R form of the enzyme. When an activator,  $A$ , is present, the cooperative binding of  $A$  shifts the equilibrium between the T and R forms, with the R form favored (Figure 7.6). As a result, there is less need for substrate,  $S$ , to shift the equilibrium in favor of the R form, and less cooperativity in the binding of  $S$  is seen.

When an inhibitor,  $I$ , is present, the cooperative binding of  $I$  also shifts the equilibrium between the T and R forms, but this time the T form is favored (Figure 7.6). More substrate is needed to shift the T-to-R equilibrium in favor of the R form. A greater degree of cooperativity is seen in the binding of  $S$ .





**FIGURE 7.5** The Monod-Wyman-Changeux (or concerted) model. (Adapted from Monod, J., Wyman, J., and Changeux, J.-P., 1965. On the nature of allosteric transitions: A plausible model. *Journal of Molecular Biology* 12:92.)



**FIGURE 7.6** Effects of binding activators and inhibitors with the concerted model. An activator is a molecule that stabilizes the R form. An inhibitor stabilizes the T form.

### What is the sequential model for allosteric behavior?

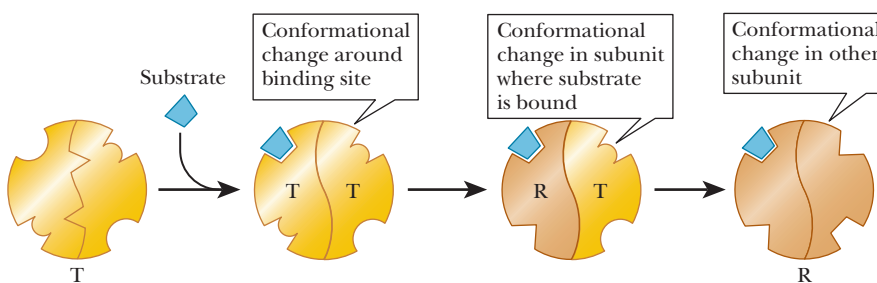
The name Daniel Koshland is associated with the direct **sequential model** of allosteric behavior. The distinguishing feature of this model is that the binding of substrate induces the conformational change from the T form to the R form—the type of behavior postulated by the induced-fit theory of substrate binding. A conformational change from T to R in one subunit makes the same conformational change easier in another subunit, and this is the form in which cooperative binding is expressed in this model (Figure 7.7a).

In the sequential model, the binding of activators and inhibitors also takes place by the induced-fit mechanism. The conformational change that begins with binding of inhibitor or activator to one subunit affects the conformations of other subunits. The net result is to favor the R state when activator is present and to favor the T form when inhibitor, I, is present (Figure 7.7b). Binding I to one subunit causes a conformational change such that the T form is even less likely to bind substrate than before. This conformational change is passed along to other subunits, making them also more likely to bind inhibitor and less likely to bind substrate. This is an example of cooperative behavior that leads to more inhibition of the enzyme. Likewise, binding an activator causes a conformational change that favors substrate binding, and this effect is passed from one subunit to another.

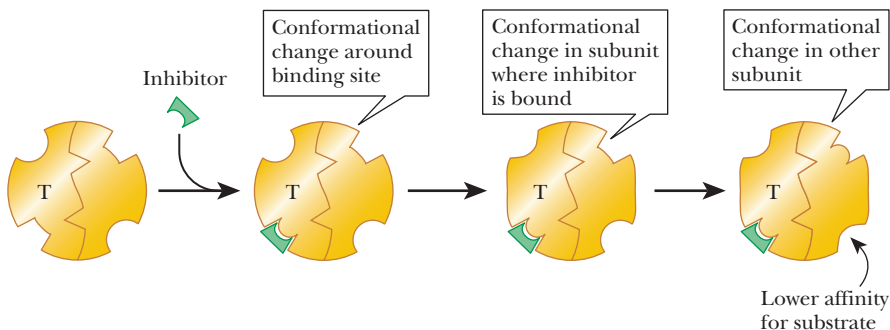
The sequential model for binding effectors of all types, including substrates, to allosteric enzymes has a unique feature not seen in the concerted model. The conformational changes thus induced can make the enzyme less likely to bind more molecules of the same type. This phenomenon, called **negative cooperativity**, has been observed in a few enzymes. One is tyrosyl tRNA synthetase, which plays a role in protein synthesis. In the reaction catalyzed by this enzyme, the amino acid tyrosine forms a covalent bond to a molecule of transfer RNA (tRNA). In subsequent steps, the tyrosine is passed along to its place in the sequence of the growing protein. The tyrosyl tRNA synthetase consists of two subunits. Binding of the first molecule of substrate to one of the subunits inhibits binding of a second molecule to the other subunit.

The sequential model has successfully accounted for the negative cooperativity observed in the behavior of tyrosyl tRNA synthetase. The concerted model makes no provision for negative cooperativity.

**A** Sequential model of cooperative binding of substrate S to an allosteric enzyme. Binding substrate to one subunit induces the other subunit to adopt the R state, which has a higher affinity for substrate.



**B** Sequential model of cooperative binding of inhibitor I to an allosteric enzyme. Binding inhibitor to one subunit induces a change in the other subunit to a form that has a lower affinity for substrate.



■ FIGURE 7.7

# Biochemical Connections MEDICINE

## Allosterism: Drug Companies Exploit the Concept

While researchers have been looking for definitive medicinal cures for many diseases, such as AIDS, for decades, pharmaceutical companies are having a harder and harder time finding novel treatments using classical drug design techniques. Traditionally, scientists looked for drugs that would mimic the behavior of signaling molecules, such as hormones and neurotransmitters. The idea was that these “fake” hormones would attach to the cell receptors for

their natural counterparts and activate or inhibit the receptor accordingly. However, one of the problems with this technique is that billions of years of evolution have led to an efficient system in which many different receptor types bind to the same signaling molecule. Therefore side effects occur because the drug that is meant to affect one type of receptor will likely affect several others unintentionally.

More recently, pharmaceutical companies have moved beyond attempting to mimic the exact signaling molecule, and even are moving away from making drugs that bind to the receptor’s active site at all. Instead, they are designing drugs that bind to other sites on the receptor, **allosteric sites**.

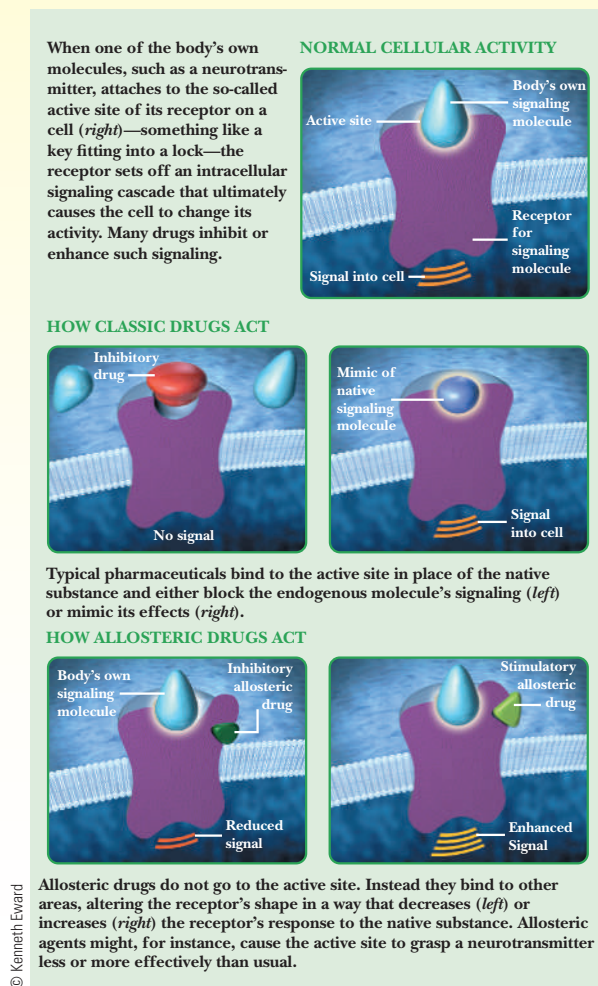
Allosteric drugs have several advantages over their more traditional “orthosteric (same site)” counterparts. First, binding a molecule to the actual receptor active site will give rise to an on/off or yes/no response. As we have seen in this chapter, allosteric effectors modulate the response in a more subtle way, like a dimmer switch opposed to a strict on/off switch.

Second, using allosteric drugs allows the drug to be more specific for one or a few receptor types. If there are 50 different receptors that bind the neurotransmitter glutamate, for example, evolution will have made the binding site very similar for all of the receptors, otherwise they could not all bind glutamate. However, there would not be evolutionary pressure to make the rest of the receptors so similar. This allows scientists to find allosteric binding sites that are much more specific.

Third, allosteric drugs can be safer because they have no effect at all unless the natural ligand is present. As an example, we can compare the effects of two central nervous system depressants—Phenobarbital and Valium. Phenobarbital binds to the active site of a neurotransmitter receptor and is deadly if taken in large quantities. Valium is now known to be an allosteric drug that binds to a different site on the receptors for  $\gamma$ -aminobutyric acid (GABA), the body’s principal inhibitory neurotransmitter. Valium turns up the response of the receptor for the GABA. When Valium is bound, the response of the receptor to GABA goes up several fold, but taking too much Valium is not as deadly as taking too much Phenobarbital, as the Valium is not causing a direct effect.

In the last few years, two new allosteric drugs hit the market. One is Amgen’s Cincalcet, a drug designed to fight chronic kidney failure by improving the action of calcium receptors. The other is an HIV medication by Pfizer called Maraviroc. It interferes with HIV entry into the cells.

In addition, several drugs are in or will be beginning clinical trials, including drugs that will combat Alzheimer’s disease, gastroesophageal reflux disease, schizophrenia, and Parkinson’s disease.



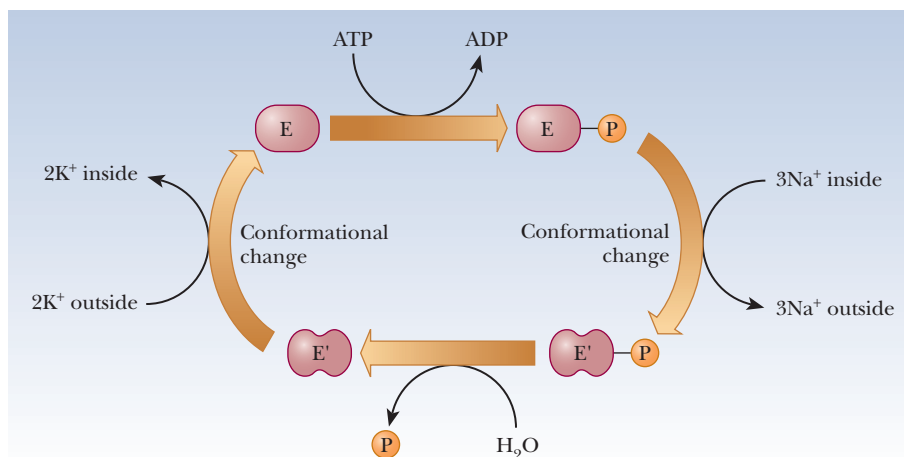
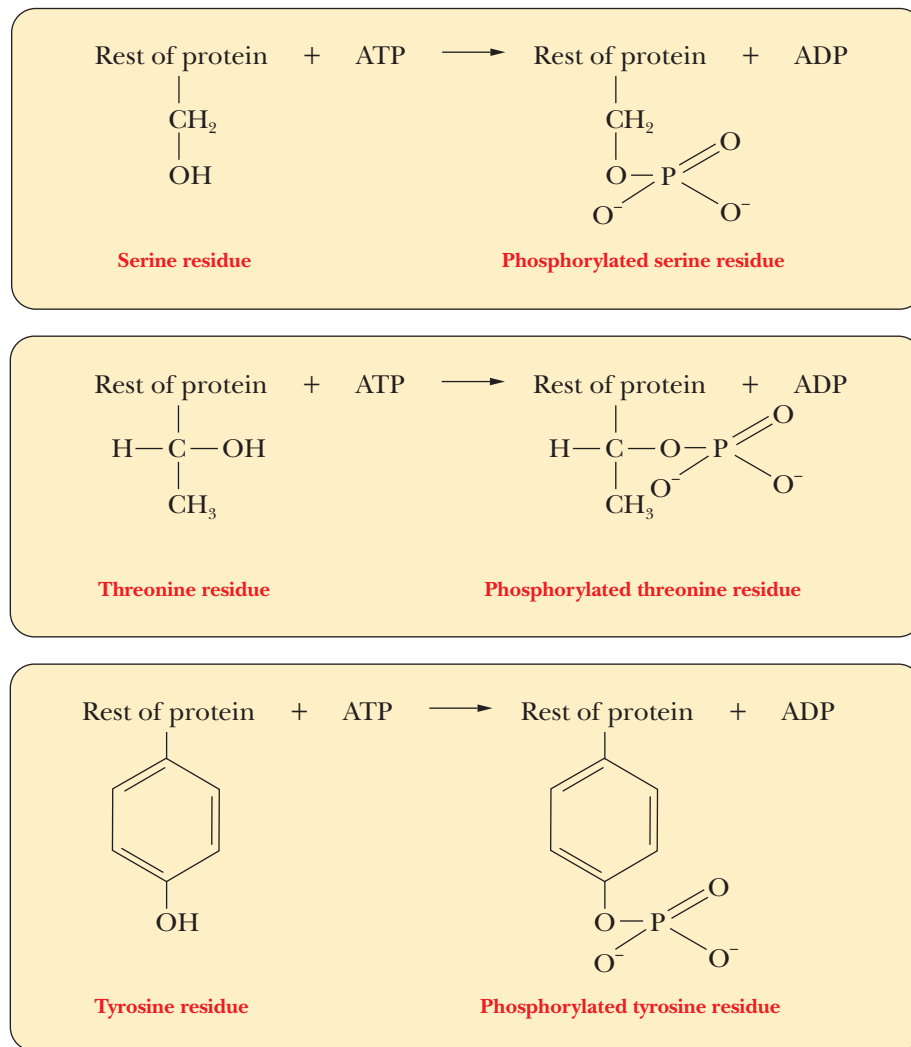
## 7.3 Control of Enzyme Activity by Phosphorylation

One of the most common control mechanisms for enzymes is by phosphorylation. The side-chain hydroxyl groups of serine, threonine, and tyrosine can all form phosphate esters. Transport across membranes provides an important example, such as the sodium–potassium ion pump, which moves potassium into the cell and sodium out (Section 8.6). The source of the phosphate group for the protein component of the sodium–potassium ion pump and for many enzyme phosphorylations is the ubiquitous ATP. When ATP is hydrolyzed to adenosine diphosphate (ADP), enough energy is released to allow a number of

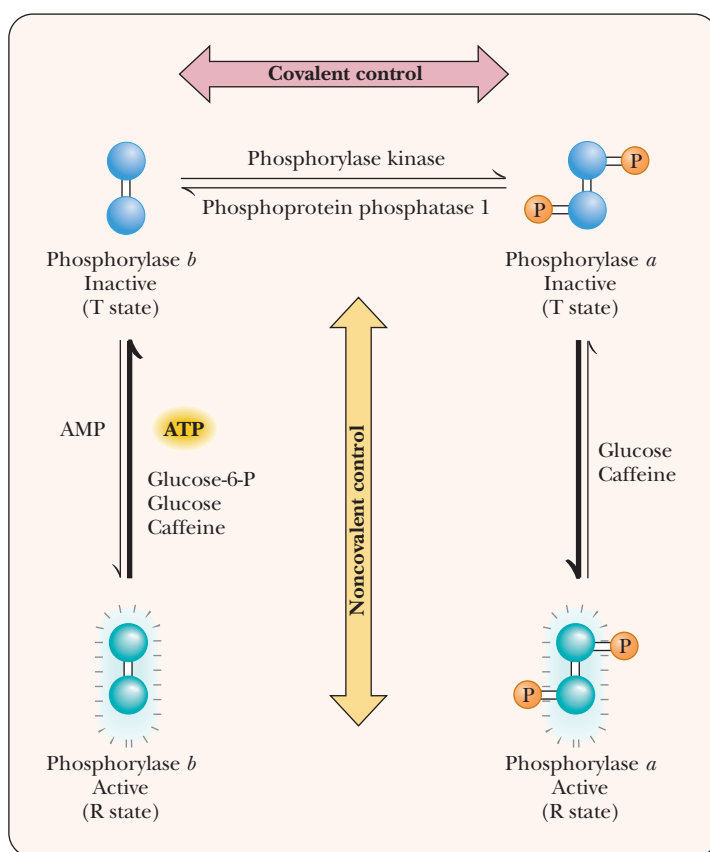


■ Edwin G. Krebs (1918–2009) received the 1992 Nobel Prize in Physiology or Medicine along with Edmond H. Fischer for their work demonstrating that phosphorylation of key proteins is a critical regulatory mechanism.

otherwise energetically unfavorable reactions to take place. In the case of the  $\text{Na}^+/\text{K}^+$  pump, ATP donates a phosphate to aspartate 369 as part of the mechanism, causing a conformational change in the enzyme (Figure 7.8). Proteins that catalyze these phosphorylation reactions are called **protein kinases**. *Kinase* refers to an enzyme that catalyzes transfer of a phosphate group, almost always from ATP, to some substrate. These enzymes play an important role in metabolism.



■ **FIGURE 7.8** Phosphorylation of the sodium–potassium pump is involved in cycling the membrane protein between the form that binds to sodium and the form that binds to potassium.



■ **FIGURE 7.9** Glycogen phosphorylase activity is subject to allosteric control and covalent modification via phosphorylation. The phosphorylated form is more active. The enzyme that puts a phosphate group on phosphorylase is called phosphorylase kinase.

Many examples appear in processes involved in generating energy, as is the case in carbohydrate metabolism. Glycogen phosphorylase, which catalyzes the initial step in the breakdown of stored glycogen (Section 18.1), exists in two forms—the phosphorylated glycogen phosphorylase *a* and the dephosphorylated glycogen phosphorylase *b* (Figure 7.9). The *a* form is more active than the *b* form, and the two forms of the enzyme respond to different allosteric effectors, depending on tissue type. Glycogen phosphorylase is thus subject to two kinds of control—allosteric regulation and covalent modification. The net result is that the *a* form is more abundant and active when phosphorylase is needed to break down glycogen to provide energy.

### Does phosphorylation always increase enzyme activity?

Although it would be convenient to have a model in which phosphorylation always increases the activity of an enzyme, biochemistry is not so kind to us. In reality, we cannot predict whether phosphorylation will increase or decrease the activity of an enzyme. In some systems, the effects on two opposing enzymes are coordinated. For example, a key enzyme in a catabolic pathway may be activated by phosphorylation while its counterpart in an anabolic opposing pathway is inhibited by phosphorylation.

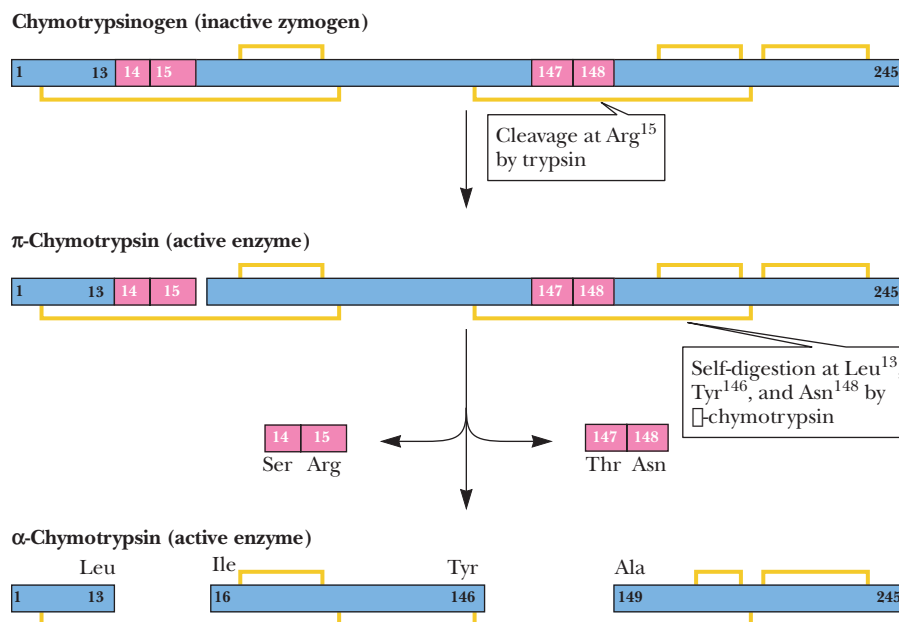
## 7.4 Zymogens

Allosteric interactions control the behavior of proteins through reversible changes in quaternary structure, but this mechanism, effective though it may be, is not the only one available. A **zymogen**, an inactive precursor of an enzyme, can be irreversibly transformed into an active enzyme by cleavage of covalent bonds.



The proteolytic enzymes trypsin and chymotrypsin (Chapter 5) provide a classic example of zymogens and their activation. Their inactive precursor molecules, trypsinogen and chymotrypsinogen, respectively, are formed in the pancreas, where they would do damage if they were in an active form. In the small intestine, where their digestive properties are needed, they are activated by cleavage of specific peptide bonds. The conversion of chymotrypsinogen to chymotrypsin is catalyzed by trypsin, which in turn arises from trypsinogen as a result of a cleavage reaction catalyzed by the enzyme enteropeptidase. Chymotrypsinogen consists of a single polypeptide chain 245 residues long, with five disulfide ( $\text{—S—S—}$ ) bonds. When chymotrypsinogen is secreted into the small intestine, trypsin present in the digestive system cleaves the peptide bond between arginine 15 and isoleucine 16, counting from the N-terminal end of the chymotrypsinogen sequence (Figure 7.10). The cleavage produces active  $\pi$ -chymotrypsin. The 15-residue fragment remains bound to the rest of the protein by a disulfide bond. Although  $\pi$ -chymotrypsin is fully active, it is not the end product of this series of reactions. It acts on itself to remove two dipeptide fragments, producing  $\alpha$ -chymotrypsin, which is also fully active. The two dipeptide fragments cleaved off are Ser 14—Arg 15 and Thr 147—Asn 148; the final form of the enzyme,  $\alpha$ -chymotrypsin, has three polypeptide chains held together by two of the five original, and still intact, disulfide bonds. (The other three disulfide bonds remain intact as well; they link portions of single polypeptide chains.) When the term *chymotrypsin* is used without specifying the  $\alpha$  or the  $\pi$  form, the final  $\alpha$  form is meant.

The changes in primary structure that accompany the conversion of chymotrypsinogen to  $\alpha$ -chymotrypsin bring about changes in the tertiary structure. The enzyme is active because of its tertiary structure, just as the zymogen is inactive because of its tertiary structure. The three-dimensional structure of chymotrypsin has been determined by X-ray crystallography. The protonated amino group of the isoleucine residue exposed by the first cleavage reaction is involved in an ionic bond with the carboxylate side chain of aspartate residue 194. This ionic bond is necessary for the active conformation of the enzyme because it is near the active site. Chymotrypsinogen lacks this bond; therefore, it does not have the active conformation and cannot bind substrate.



■ **FIGURE 7.10** The proteolytic activation of chymotrypsinogen.

Another important class of proteases is the **caspases**, which are a family of homodimer cysteine proteases responsible for many processes in cell biology, including programmed cell death, or **apoptosis**. They are also involved in signaling within the immune system and in stem cell differentiation. Apoptosis is a natural phenomenon, in that cells are always being turned over. Disruption of apoptosis can lead to forms of cancer. However, apoptosis can also cause unwanted cell death, as with cells surrounding neurons that have died from a stroke. The immediate cells that were oxygen-deprived die quickly, but surrounding cells can die more slowly due to apoptosis.

The caspases are first produced in an inactive form called procaspases, which are later activated by proteolysis of their immature forms. Once activated, the caspases launch a series of attacks against specific targets, leading to the death of the cells. Caspases are currently a hot topic of study, as scientists try to find ways to exploit their relationship to cancer and strokes in order to combat both.

## 7.5 The Nature of the Active Site

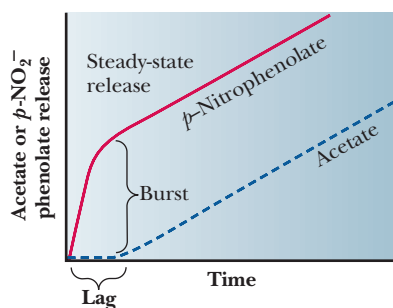
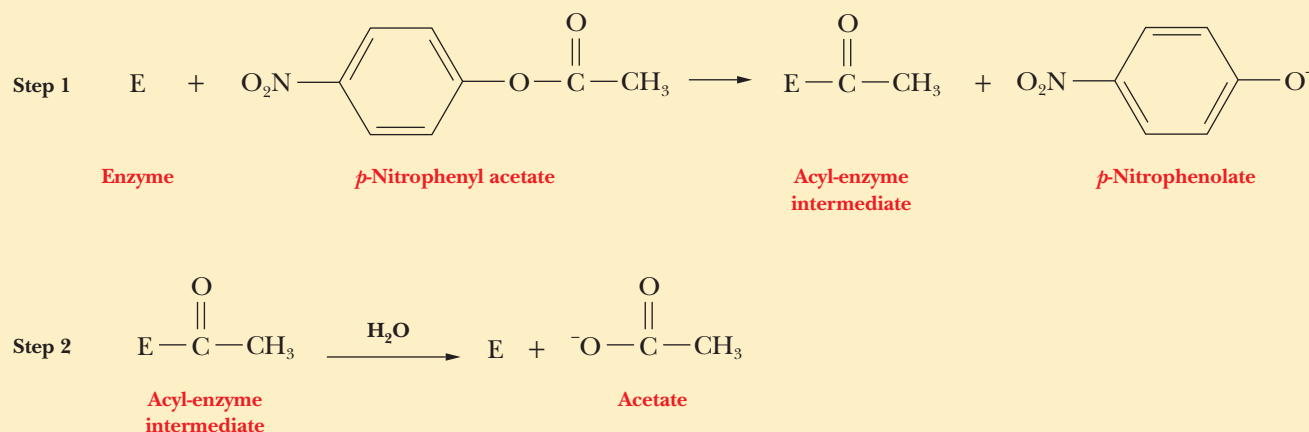
In this section we will look at the specific mechanism by which an enzyme is able to increase the rate of a chemical reaction. This mechanism is based on the exact three-dimensional arrangement of the amino acids in the active site. We can ask several questions about the mode of action of an enzyme. Here are some of the most important:

1. Which amino acid residues on the enzyme are in the active site (recall this term from Chapter 6) and catalyze the reaction? In other words, which are the critical amino acid residues?
2. What is the spatial relationship of the critical amino acid residues in the active site?
3. What is the mechanism by which the critical amino acid residues catalyze the reaction?

Answers to these questions are available for chymotrypsin, and we shall use its mechanism as an example of enzyme action. Information on well-known systems such as chymotrypsin can lead to general principles that are applicable to all enzymes. Enzymes catalyze chemical reactions in many ways, but all reactions have in common the requirement that some reactive group on the enzyme interact with the substrate. In proteins, the  $\alpha$ -carboxyl and  $\alpha$ -amino groups of the amino acids are no longer free because they have formed peptide bonds. Thus, the side-chain reactive groups are the ones involved in the action of the enzyme. Hydrocarbon side chains do not contain reactive groups and are not involved in the process. Functional groups that can play a catalytic role include the imidazole group of histidine, the hydroxyl group of serine, the carboxyl side chains of aspartate and glutamate, the sulfhydryl group of cysteine, the amino side chain of lysine, and the phenol group of tyrosine. If the  $\alpha$ -carboxyl or the  $\alpha$ -amino group of the peptide chain are positioned in the active site, then they too can play a role.

Chymotrypsin catalyzes the hydrolysis of peptide bonds adjacent to aromatic amino acid residues in the protein being hydrolyzed; other residues are attacked at a lower frequency. In addition, chymotrypsin catalyzes the hydrolysis of esters in model studies in the laboratory. The use of model systems is common in biochemistry because a model provides the essential features of a reaction in a simple form that is easier to work with than the one found in nature. The amide (peptide) bond and the ester bond are similar enough that the enzyme can accept both types of compounds as substrates. Model systems based on the hydrolysis of esters are frequently used to study the peptide hydrolysis reaction.

A typical model compound is *p*-nitrophenyl acetate, which is hydrolyzed in two stages. The acetyl group is covalently attached to the enzyme at the end of the first stage (Step 1) of the reaction, but the *p*-nitrophenolate ion is released. In the second stage (Step 2), the acyl-enzyme intermediate is hydrolyzed, releasing acetate and regenerating the free enzyme. The kinetics observed when *p*-nitrophenyl acetate is first mixed with chymotrypsin shows an initial burst and then a slower phase (Figure 7.11). This reaction is consistent with an enzyme that has two phases, one often forming an acylated enzyme intermediate.



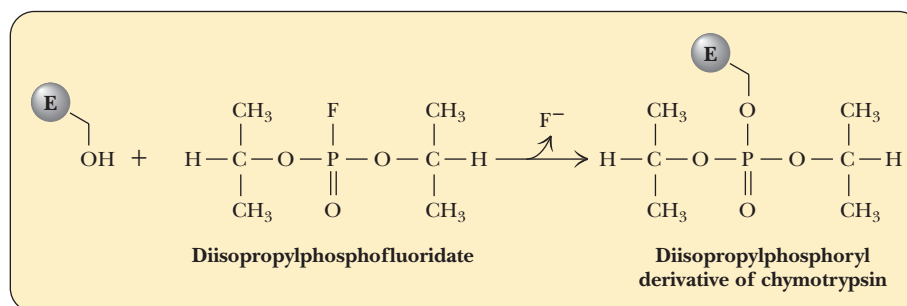
■ **FIGURE 7.11** The kinetics observed in the chymotrypsin reaction. An initial burst of *p*-nitrophenolate is seen, followed by a slower, steady-state release that matches the appearance of the other product, acetate.

### How do we determine the essential amino acid residues?

The serine residue at position 195 is required for the activity of chymotrypsin; in this respect, chymotrypsin is typical of a class of enzymes known as **serine proteases**. Trypsin and thrombin, mentioned previously, are also serine proteases (see the Biochemical Connections box on page 193). The enzyme is completely inactivated when this serine reacts with diisopropylphosphorfluoridate (DIPF), forming a covalent bond that links the serine side chain with DIPF. The formation of covalently modified versions of specific side chains on proteins is called **labeling**; it is widely used in laboratory studies. The other serine residues of chymotrypsin are far less reactive and are not labeled by DIPF (Figure 7.12).

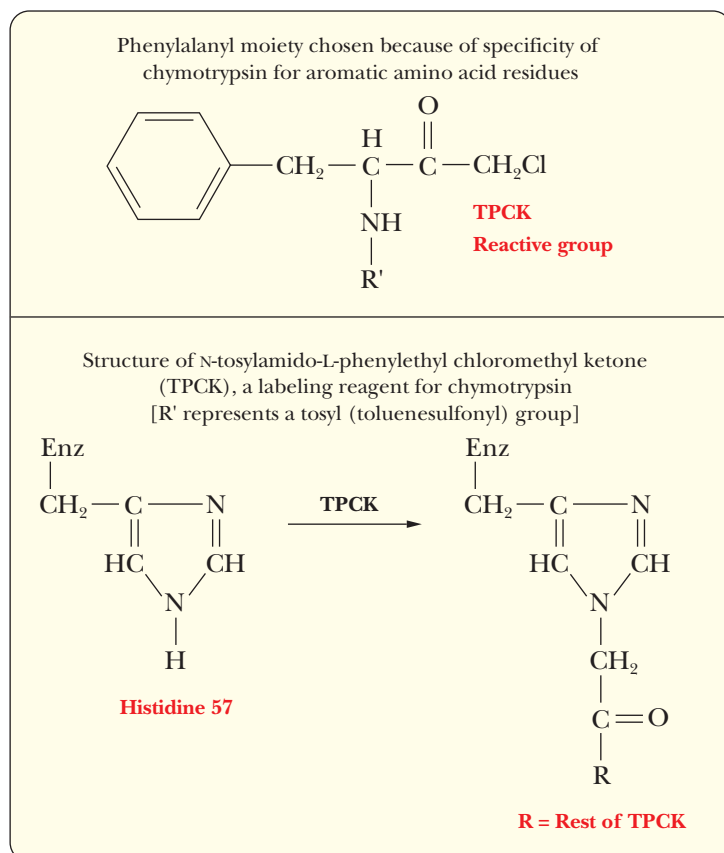
Histidine 57 is another critical amino acid residue in chymotrypsin. Chemical labeling again provides the evidence for involvement of this residue in the activity of chymotrypsin. In this case, the reagent used to label the critical amino

■ **FIGURE 7.12** Diisopropylphosphorfluoridate (DIPF) labels the active-site serine of chymotrypsin.



acid residue is *N*-tosylamido-L-phenylethyl chloromethyl ketone (TPCK), also called tosyl-L-phenylalanine chloromethyl ketone. The phenylalanine moiety is bound to the enzyme because of the specificity for aromatic amino acid residues at the active site, and the active site histidine residue reacts because the labeling reagent is similar to the usual substrate.

**The labeling of the active-site histidine of chymotrypsin by TPCK**

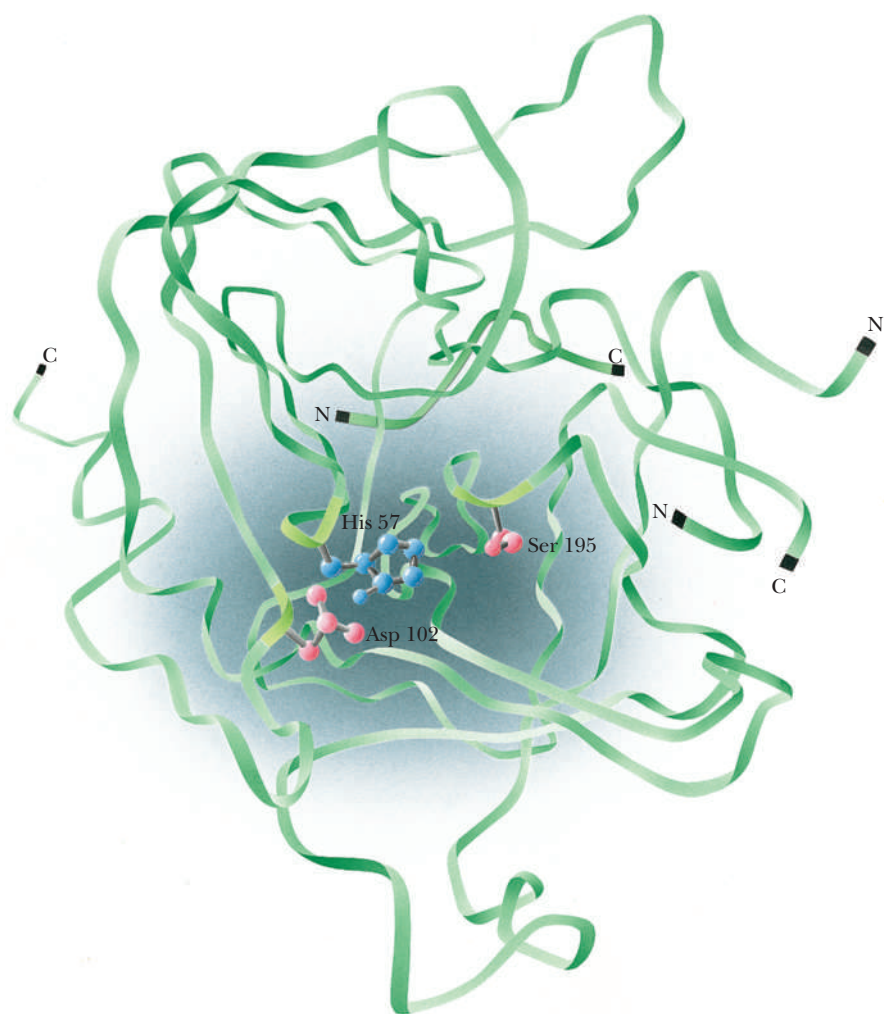


### How does the architecture of the active site affect catalysis?

Both serine 195 and histidine 57 are required for the activity of chymotrypsin; therefore, they must be close to each other in the active site. The determination of the three-dimensional structure of the enzyme by X-ray crystallography provides evidence that the active-site residues do indeed have a close spatial relationship. The folding of the chymotrypsin backbone, mostly in an antiparallel pleated-sheet array, positions the essential residues around an active-site pocket (Figure 7.13). Only a few residues are directly involved in the active site, but the whole molecule is necessary to provide the correct three-dimensional arrangement for those critical residues.

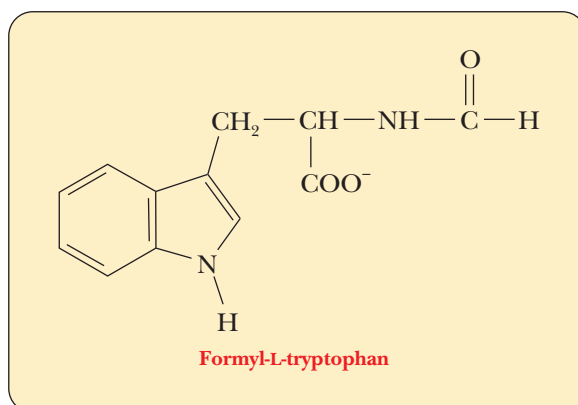
Other important pieces of information about the three-dimensional structure of the active site emerge when a complex is formed between chymotrypsin and a substrate analogue. When one such substrate analog, formyl-L-tryptophan, is bound to the enzyme, the tryptophan side chain fits into a hydrophobic pocket near serine 195. This type of binding is not surprising, in view of the specificity of the enzyme for aromatic amino acid residues at the cleavage site.

The results of X-ray crystallography show, in addition to the binding site for aromatic amino acid side chains of substrate molecules, a definite arrangement of the



■ **FIGURE 7.13** The tertiary structure of chymotrypsin places the essential amino acid residues close to one another. They are shown in blue and red. (Abeles, R., Frey, P., Jencks, W. *Biochemistry* © Boston: Jones and Bartlett, Publishers, 1992, reprinted by permission.)

amino acid side chains that are responsible for the catalytic activity of the enzyme. The residues involved in this arrangement are serine 195 and histidine 57.



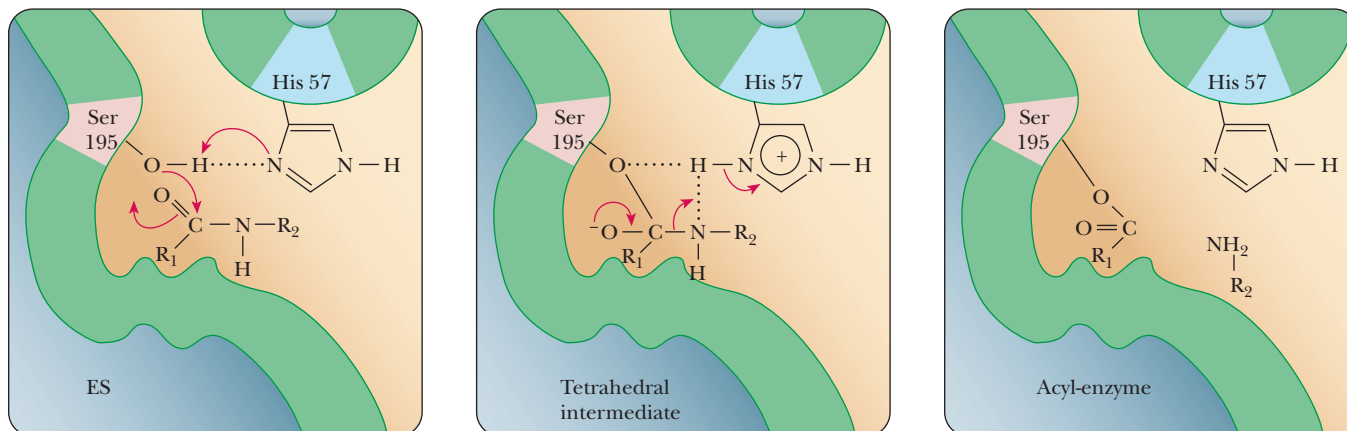
### How do the critical amino acids catalyze the chymotrypsin reaction?

Any postulated reaction mechanism must be modified or discarded if it is not consistent with experimental results. There is consensus, but not total agreement, on the main features of the mechanism discussed in this section.

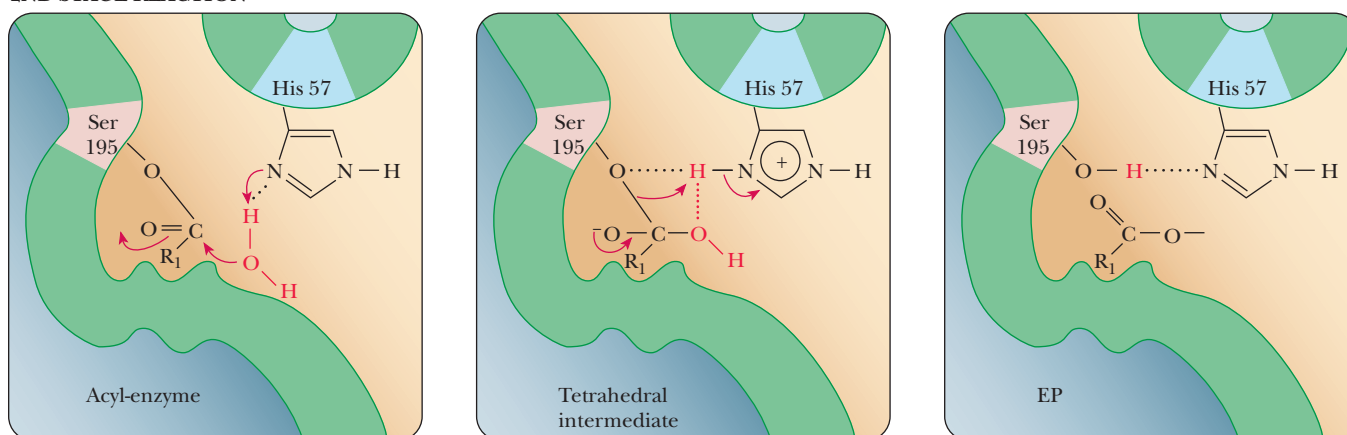
The critical amino acid residues, serine 195 and histidine 57, are involved in the mechanism of catalytic action. In the terminology of organic chemistry, the oxygen of the serine side chain is a **nucleophile**, or nucleus-seeking



## 1ST STAGE REACTION



## 2ND STAGE REACTION



■ **FIGURE 7.14** The mechanism of chymotrypsin action. In the first stage of the reaction, the nucleophile serine 195 attacks the carbonyl carbon of the substrate. In the second stage, water is the nucleophile that attacks the acyl-enzyme intermediate. Note the involvement of histidine 57 in both stages of the reaction. (From Hammes, G.: Enzyme Catalysis and Regulation, New York: Academic Press, 1982. Used by permission of Elsevier.)

substance. A nucleophile tends to bond to sites of positive charge or polarization (electron-poor sites), in contrast to an **electrophile**, or electron-seeking substance, which tends to bond to sites of negative charge or polarization (electron-rich sites). The nucleophilic oxygen of the serine attacks the carbonyl carbon of the peptide group. The carbon now has four single bonds, and a tetrahedral intermediate is formed; the original  $\text{—C=O}$  bond becomes a single bond, and the carbonyl oxygen becomes an oxyanion. The acyl-enzyme intermediate is formed from the tetrahedral species (Figure 7.14). The histidine and the amino portion of the original peptide group are involved in this part of the reaction as the amino group hydrogen bonds to the imidazole portion of the histidine. Note that the imidazole is already protonated and that the proton came from the hydroxyl group of the serine. The histidine behaves as a base in abstracting the proton from the serine; in the terminology of the physical organic chemist, the histidine acts as a general base catalyst. The carbon–nitrogen bond of the original peptide group breaks, leaving the acyl-enzyme intermediate. The proton abstracted by the histidine has been donated to the leaving amino group. In donating the proton, the histidine has acted as an acid in the breakdown of the tetrahedral intermediate, although it acted as a base in its formation.

In the deacylation phase of the reaction, the last two steps are reversed, with water acting as the attacking nucleophile. In this second phase, the water is hydrogen-bonded to the histidine. The oxygen of water now performs the

nucleophilic attack on the acyl carbon that came from the original peptide group. Once again, a tetrahedral intermediate is formed. In the final step of the reaction, the bond between the serine oxygen and the carbonyl carbon breaks, releasing the product with a carboxyl group where the original peptide group used to be and regenerating the original enzyme. Note that the serine is hydrogen-bonded to the histidine. This hydrogen bond increases the nucleophilicity of the serine, whereas in the second part of the reaction, the hydrogen bond between the water and the histidine increased the nucleophilicity of the water.

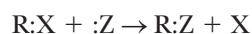
The mechanism of chymotrypsin action is particularly well studied and, in many respects, typical. Numerous types of reaction mechanisms for enzyme action are known, and we shall discuss them in the contexts of the reactions catalyzed by the enzymes in question. To lay the groundwork, it is useful to discuss some general types of catalytic mechanisms and how they affect the specificity of enzymatic reactions.

## 7.6 Chemical Reactions Involved in Enzyme Mechanisms

The overall mechanism for a reaction may be fairly complex, as we have seen in the case of chymotrypsin, but the individual parts of a complex mechanism can themselves be fairly simple. Concepts such as nucleophilic attack and acid catalysis commonly enter into discussions of enzymatic reactions. We can draw quite a few general conclusions from these two general descriptions.

### What are the most common types of reactions?

**Nucleophilic substitution reactions** play a large role in the study of organic chemistry, and they are excellent illustrations of the importance of kinetic measurements in determining the mechanism of a reaction. A nucleophile is an electron-rich atom that attacks an electron-deficient atom. A general equation for this type of reaction is



where :Z is the nucleophile and X is called a *leaving group*. In biochemistry, the carbon of a carbonyl group (C=O) is often the atom attacked by the nucleophile. Common nucleophiles are the oxygens of serine, threonine, and tyrosine. If the rate of the reaction shown here is found to depend solely on the concentration of the R:X, then the nucleophilic reaction is called an **S<sub>N</sub>1** (substitution nucleophilic unimolecular). Such a mechanism would mean that the slow part of the reaction is the breaking of the bond between R and X, and that the addition of the nucleophile Z happens very quickly compared to that. An S<sub>N</sub>1 reaction follows first-order kinetics (Chapter 6). If the nucleophile attacks the R:X while the X is still attached, then both the concentration of R:X and the concentration of :Z will be important. This reaction will follow second-order kinetics and is called an **S<sub>N</sub>2** reaction (substitution nucleophilic bimolecular). The difference between S<sub>N</sub>1 and S<sub>N</sub>2 is very important to biochemists because it explains much about the stereospecificity of the products formed. An S<sub>N</sub>1 reaction often leads to loss of stereospecificity. Because the leaving group is gone before the attacking group enters, the attacking group can often end up in one of two orientations, although the specificity of the active site can also limit this. With an S<sub>N</sub>2 reaction, the fact that the leaving group is still attached forces the nucleophile to attack from a particular side of the bond, leading to only one possible stereospecificity in the product. The chymotrypsin nucleophilic attacks were examples of S<sub>N</sub>2 reactions, although no stereochemistry is noted because the carbonyl that was attacked became a carbonyl group again at the end of the reaction and was, therefore, not chiral.

To discuss acid–base catalysis, it is helpful to recall the definitions of acids and bases. In the Brønsted–Lowry definition, an acid is a proton donor and a

base is a proton acceptor. The concept of **general acid–base catalysis** depends on donation and acceptance of protons by groups such as the imidazole, hydroxyl, carboxyl, sulfhydryl, amino, and phenolic side chains of amino acids; all these functional groups can act as acids or bases. The donation and acceptance of protons gives rise to the bond breaking and re-formation that constitute the enzymatic reaction.

If the enzyme mechanism involves an amino acid donating a hydrogen ion, as in the reaction

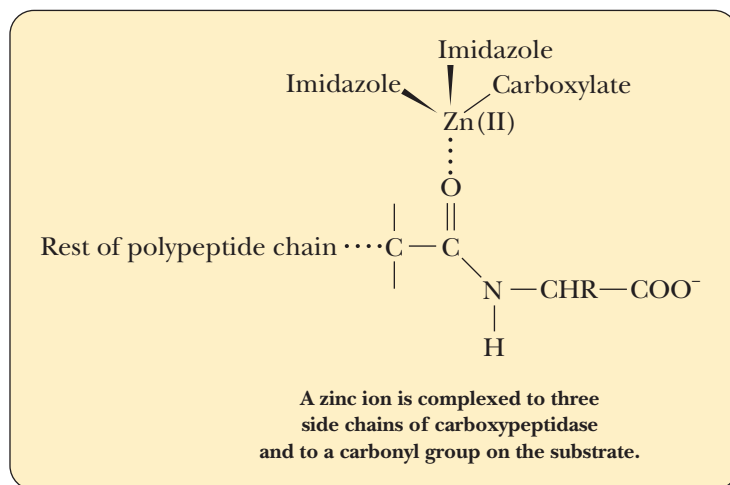


then that part of the mechanism would be called general acid catalysis. If an amino acid takes a hydrogen ion from one of the substrates, such as in the reaction



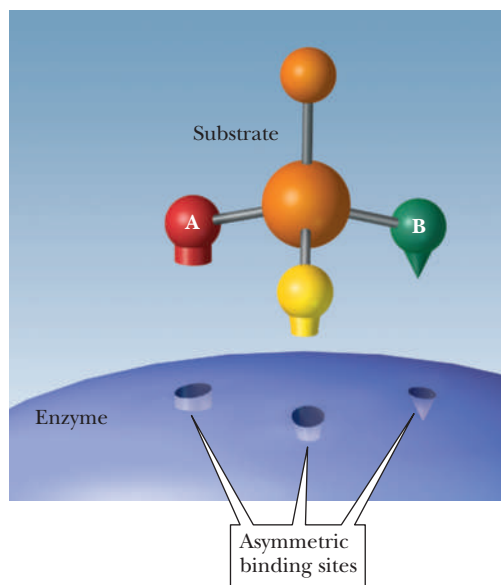
then that part is called general base catalysis. Histidine is an amino acid that often takes part in both reactions, because it has a reactive hydrogen on the imidazole side chain that dissociates near physiological pH. In the chymotrypsin mechanism, we saw both acid and base catalysis by histidine.

A second form of acid–base catalysis reflects another, more general definition of acids and bases. In the Lewis formulation, an acid is an electron-pair acceptor, and a base is an electron-pair donor. Metal ions, including such biologically important ones as  $\text{Mn}^{2+}$ ,  $\text{Mg}^{2+}$ , and  $\text{Zn}^{2+}$ , are Lewis acids. Thus, they can play a role in **metal–ion catalysis** (also called Lewis acid–base catalysis). The involvement of  $\text{Zn}^{2+}$  in the enzymatic activity of carboxypeptidase A is an example of this type of behavior. This enzyme catalyzes the hydrolysis of C terminal peptide bonds of proteins. The  $\text{Zn}(\text{II})$ , which is required for the activity of the enzyme, is complexed to the imidazole side chains of histidines 69 and 196 and to the carboxylate side chain of glutamate 72. The zinc ion is also complexed to the substrate.



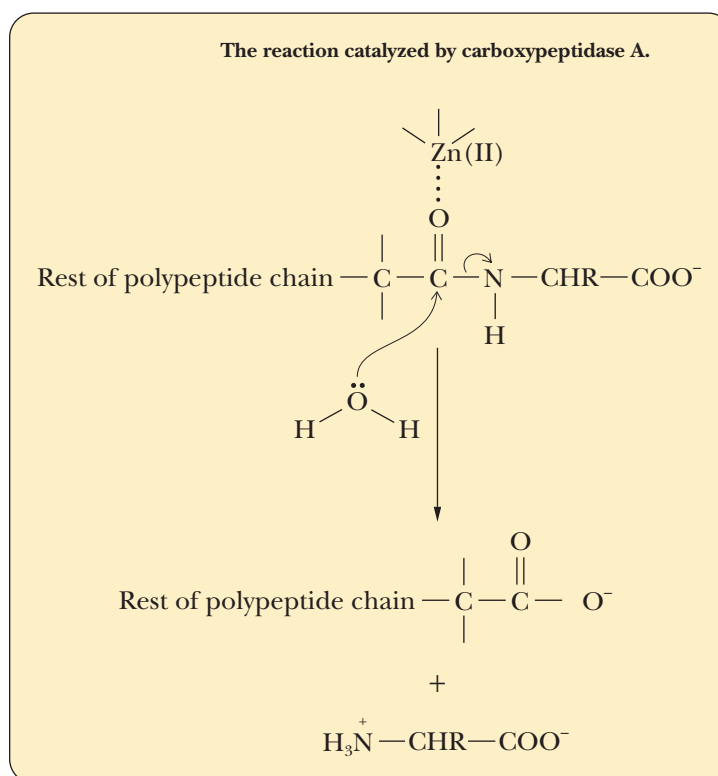
The type of binding involved in the complex is similar to the binding that links iron to the large ring involved in the heme group. Binding the substrate to the zinc ion polarizes the carbonyl group, making it susceptible to attack by water and allowing the hydrolysis to proceed more rapidly than it does in the uncatalyzed reaction.

A definite connection exists between the concepts of acids and bases and the idea of nucleophiles and their complementary substances, electrophiles. A Lewis acid is an electrophile, and a Lewis base is a nucleophile. Catalysis by enzymes, including their remarkable specificity, is based on these well-known chemical principles operating in a complex environment.



■ **FIGURE 7.15** An asymmetric binding site on an enzyme can distinguish between identical groups, such as A and B. Note that the binding site consists of three parts, giving rise to asymmetric binding because one part is different from the other two.

The nature of the active site plays a particularly important role in the specificity of enzymes. An enzyme that displays *absolute specificity*, catalyzing the reaction of one, and only one, substrate to a particular product, is likely to have a fairly rigid active site that is best described by the lock-and-key model of substrate binding. The many enzymes that display *relative specificity*, catalyzing the reactions of structurally related substrates to related products, apparently have more flexibility in their active sites and are better characterized by the induced-fit model of enzyme–substrate binding; chymotrypsin is a good example. Finally, there are *stereospecific* enzymes with specificity in which optical activity plays a role. The binding site itself must be asymmetric in this situation (Figure 7.15). If the enzyme is to bind specifically to an optically active substrate, the binding site must have the shape of the substrate and not its mirror image. There are even enzymes that introduce a center of optical activity into the product. The substrate itself is not optically active in this case. There is only one product, which is one of two possible isomers, not a mixture of optical isomers.



## 7.7 The Active Site and Transition States

Now that we have spent some time looking at mechanisms and the active site, it is worth revisiting the nature of enzyme catalysis. Recall that an enzyme lowers the activation energy by lowering the energy necessary to reach the transition state (Figure 6.1). The true nature of the transition state is a chemical species that is intermediate in structure between the substrate and the product. This transition state often has a very different shape from either the substrate or the product. In the case of chymotrypsin, the substrate has the carbonyl group that is attacked by the reactive serine. The carbon of the carbonyl group has three bonds, and the orientation is planar. After the serine performs the nucleophilic attack, the carbon has four bonds and a tetrahedral arrangement. This tetrahedral shape is the transition state of the reaction, and the active site must make this change more likely.

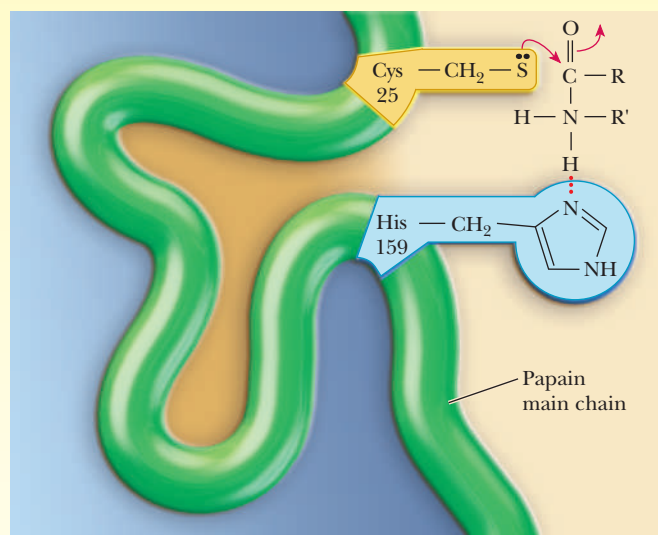
### How do we determine the nature of the transition state?

The fact that the enzyme stabilizes the transition state has been shown many times by the use of **transition-state analogs**, which are molecules with a shape that mimics the transition state of the substrate. Proline racemase catalyzes a reaction that converts L-proline to D-proline. In the progress of the reaction, the  $\alpha$ -carbon must change from a tetrahedral arrangement to a planar form, and then back to tetrahedral, but with the orientation of two bonds reversed (Figure 7.16). An inhibitor of the reaction is pyrrole-2-carboxylate, a chemical that is structurally similar to what proline would look like at its transition state because it is always planar at the equivalent carbon. This inhibitor binds to proline racemase 160 times more strongly than proline does. Transition-state

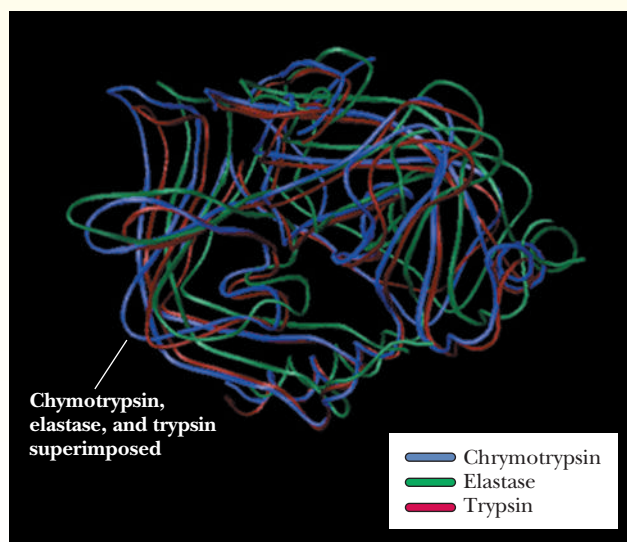
## Biochemical Connections ALLIED HEALTH

### Families of Enzymes: Proteases

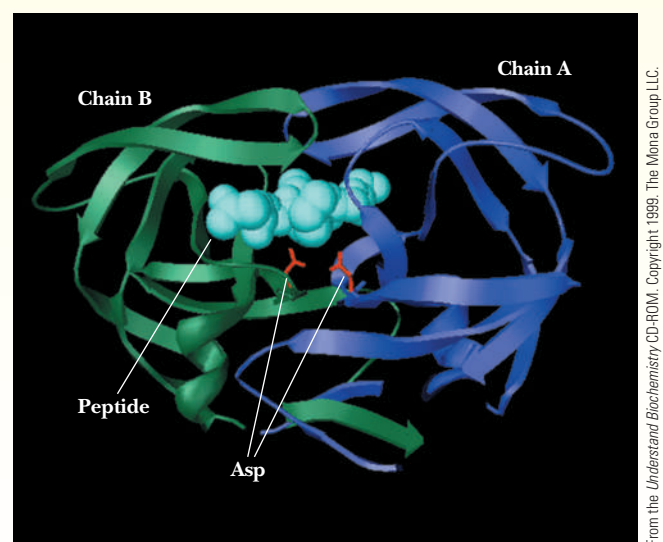
Large numbers of enzymes catalyze similar functions. Many oxidation–reduction reactions take place, each catalyzed by a specific enzyme. We have already seen that kinases transfer phosphate groups. Still other enzymes catalyze hydrolytic reactions. Enzymes that have similar functions may have widely varying structures. The important feature that they have in common is an active site that can catalyze the reaction in question. A number of different enzymes catalyze the hydrolysis of proteins. Chymotrypsin is one example of the class of serine proteases, but many others are known, including elastase, which catalyzes the degradation of the connective tissue protein elastin and the digestive enzyme trypsin. (Recall that we first saw trypsin in its role in protein sequencing.) All these enzymes are similar in structure. Other proteases employ other essential amino acid residues as the nucleophile in the active site. Papain, the basis of commercial meat tenderizers, is a proteolytic enzyme derived from papayas. However, it has a cysteine rather than a serine as the nucleophile in its active site. Aspartyl proteases differ still more widely in structure from the common serine proteases. A pair of aspartate side chains, sometimes on different subunits, participates in the reaction mechanism. A number of aspartyl proteases, such as the digestive enzyme pepsin, are known. However, the most notorious aspartyl protease is the one necessary for the maturation of the human immunodeficiency virus, HIV-1 protease.



■ Papain is a cysteine protease. A critical cysteine residue is involved in the nucleophilic attack on the peptide bonds it hydrolyzes.

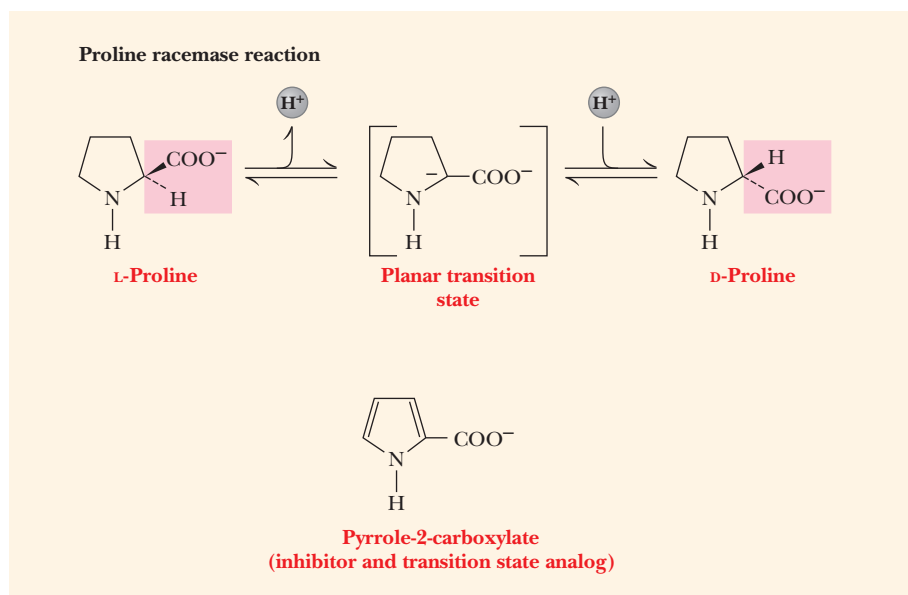


■ Chymotrypsin, elastin, and trypsin are serine proteases and have similar structures.

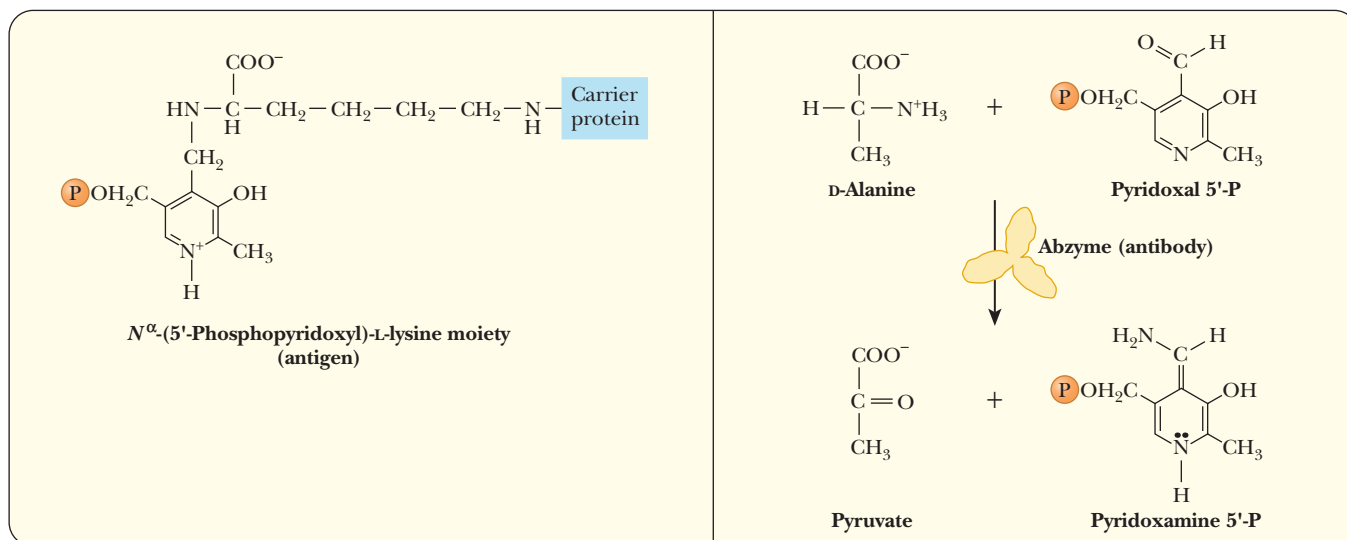


■ HIV-1 protease is a member of the class of enzymes called the aspartic proteases. Two aspartates are involved in the reaction.





■ **FIGURE 7.16** The proline racemase reaction. Pyrrole-2-carboxylate and  $\Delta^1$ -pyrroline-2-carboxylate mimic the planar transition state of the reaction.



**A**  $N^{\alpha}$ -(5'-phosphopyridoxyl)-L-lysine moiety is a transition-state analog for the reaction of an amino acid with pyridoxal 5'-phosphate. When this moiety is attached to a protein and injected into a host, it acts like an antigen, and the host then produces antibodies that have catalytic activity (abzymes).

**B** The abzyme is then used to catalyze the reaction.

■ **FIGURE 7.17** Abzymes.

analogues have been used with many enzymes to help verify a suspected mechanism and structure of the transition state as well as to inhibit an enzyme selectively. Back in 1969, William Jencks proposed that an immunogen (a molecule that elicits an antibody response) would elicit antibodies with catalytic activity if the immunogen mimicked the transition state of the reaction. Richard Lerner and Peter Schultz, who created the first catalytic antibodies, verified this hypothesis in 1986. Because an antibody is a protein designed to bind to specific molecules on the immunogen, the antibody is, in essence, a fake active site. For example, the reaction of pyridoxal phosphate and an amino acid to form the corresponding  $\alpha$ -keto acid and pyridoxamine phosphate is a very important reaction in amino acid metabolism. The molecule,  $N^{\alpha}$ -(5'-phosphopyridoxyl)-L-lysine serves as a transition-state analog for this reaction. When this antigen molecule was used to elicit antibodies, these antibodies, or **abzymes**, had catalytic activity (Figure 7.17). Thus, in addition to

helping to verify the nature of the transition state or making an inhibitor, transition-state analogs now offer the possibility of making designer enzymes to catalyze a wide variety of reactions.

## 7.8 Coenzymes

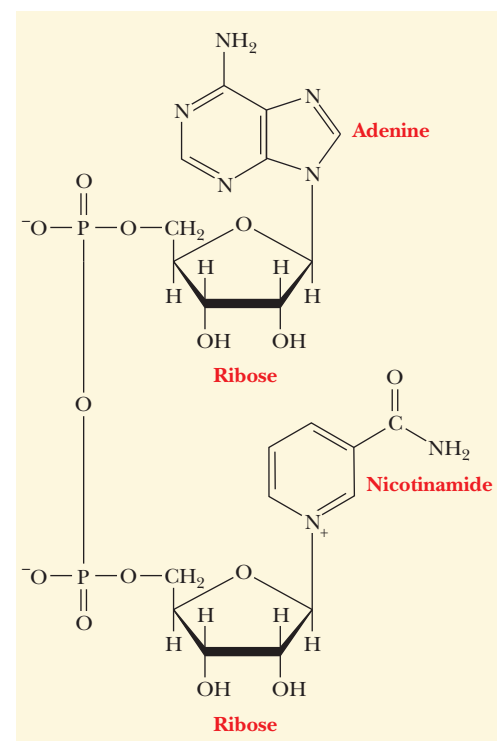
Cofactors are nonprotein substances that take part in enzymatic reactions and are regenerated for further reaction. Metal ions frequently play such a role, and they make up one of two important classes of cofactors. The other important class (**coenzymes**) is a mixed bag of organic compounds; many of them are vitamins or are metabolically related to vitamins.

Because metal ions are Lewis acids (electron-pair acceptors), they can act as Lewis acid–base catalysts. They can also form coordination compounds by behaving as Lewis acids, while the groups to which they bind act as Lewis bases. Coordination compounds are an important part of the chemistry of metal ions in biological systems, as shown by Zn(II) in carboxypeptidase and by Fe(II) in hemoglobin. The coordination compounds formed by metal ions tend to have quite specific geometries, which aid in positioning the groups involved in a reaction for optimum catalysis.

Some of the most important organic coenzymes are vitamins and their derivatives, especially B vitamins. Many of these coenzymes are involved in oxidation–reduction reactions, which provide energy for the organism. Others serve as group-transfer agents in metabolic processes (Table 7.1). We shall see these coenzymes again when we discuss the reactions in which they are involved. For the present, we shall investigate one particularly important oxidation–reduction coenzyme and one group-transfer coenzyme.

Nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ) is a coenzyme in many oxidation–reduction reactions. Its structure (Figure 7.18) has three parts—a nicotinamide ring, an adenine ring, and two sugar–phosphate groups linked together. The nicotinamide ring contains the site at which oxidation and reduction reactions occur (Figure 7.19). Nicotinic acid is another name for the vitamin niacin. The adenine–sugar–phosphate portion of the molecule is structurally related to nucleotides.

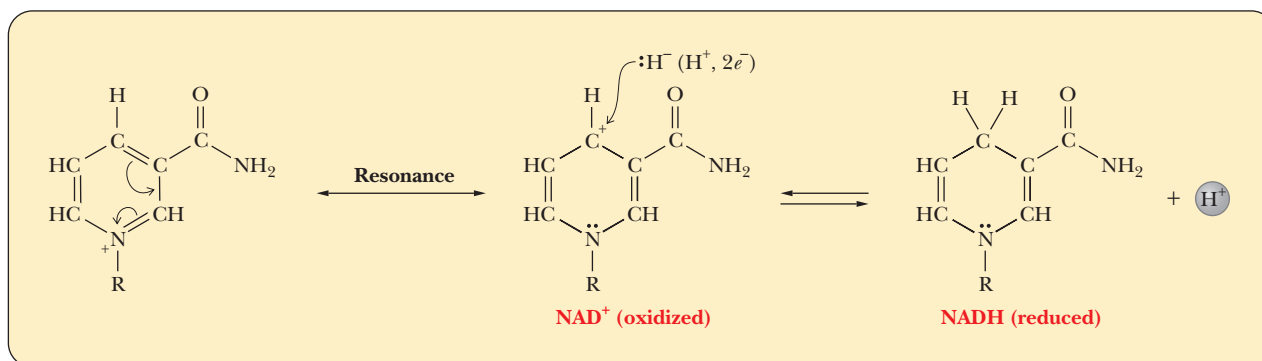
The  $\text{B}_6$  vitamins (pyridoxal, pyridoxamine, and pyridoxine and their phosphorylated forms, which are the coenzymes) are involved in the transfer of amino groups from one molecule to another, an important step in the biosynthesis of amino acids (Figure 7.20). In the reaction, the amino group is transferred from the donor to the coenzyme and then from the coenzyme to the ultimate acceptor (Figure 7.21).



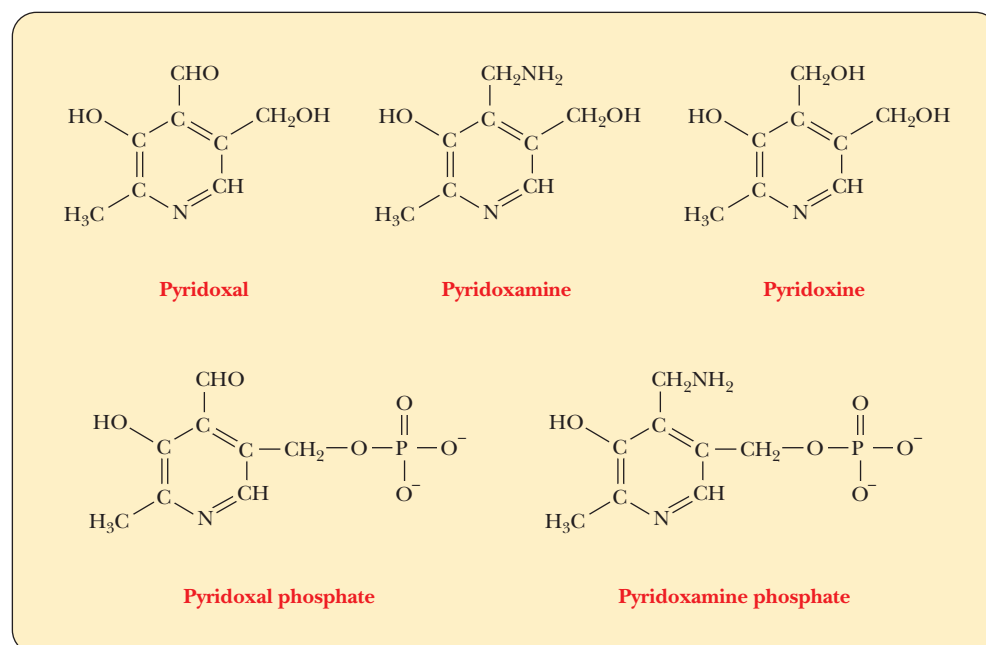
■ **FIGURE 7.18** The structure of nicotinamide adenine dinucleotide ( $\text{NAD}^+$ ).

**TABLE 7.1**

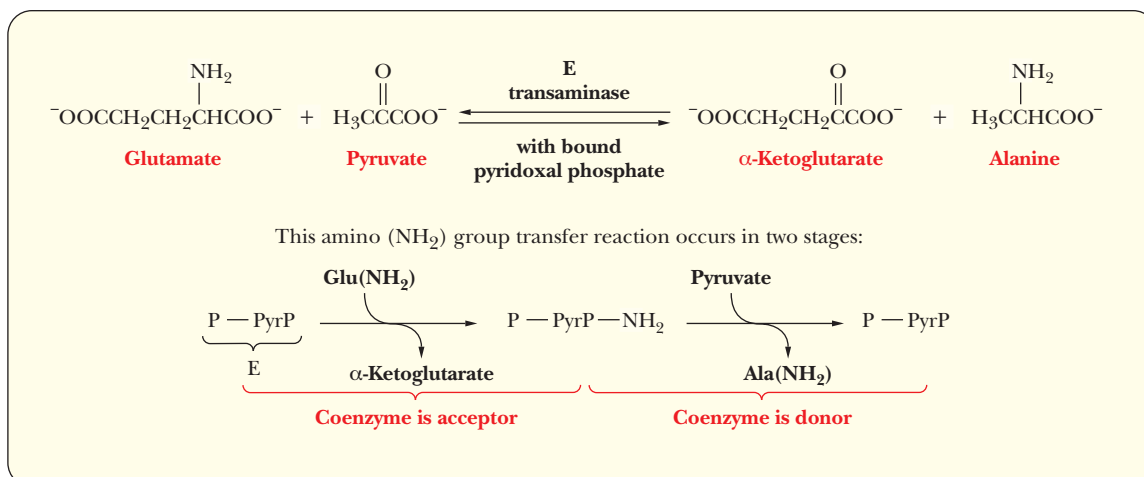
Coenzymes, Their Reactions, and Their Vitamin Precursors			
Coenzyme	Reaction Type	Vitamin Precursors	See Section
Biotin	Carboxylation	Biotin	18.2, 21.6
Coenzyme A	Acyl transfer	Pantothenic acid	15.9, 19.3, 21.6
Flavin coenzymes	Oxidation–reduction	Riboflavin ( $\text{B}_2$ )	15.9, 19.3
Lipoic acid	Acyl transfer	—	19.3
Nicotinamide adenine coenzymes	Oxidation–reduction	Niacin	15.9, 17.3, 19.3
Pyridoxal phosphate	Transamination	Pyridoxine ( $\text{B}_6$ )	23.4
Tetrahydrofolic acid	Transfer of one-carbon units	Folic acid	23.4
Thiamine pyrophosphate	Aldehyde transfer	Thiamin ( $\text{B}_1$ )	17.4, 18.4



■ **FIGURE 7.19** The role of the nicotinamide ring in oxidation–reduction reactions. R is the rest of the molecule. In reactions of this sort, an H<sup>+</sup> is transferred along with the two electrons.



■ **FIGURE 7.20** Forms of vitamin B<sub>6</sub>. The first three structures are vitamin B<sub>6</sub> itself, and the last two structures show the modifications that give rise to the metabolically active coenzyme.



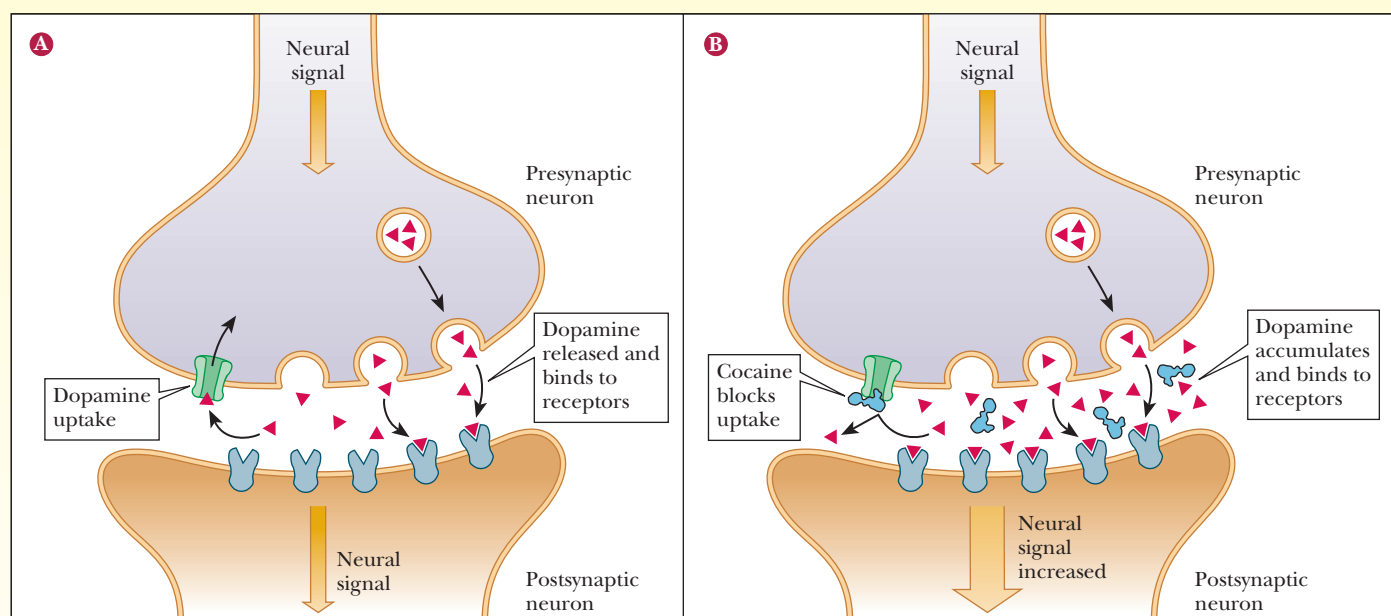
■ **FIGURE 7.21** The role of pyridoxal phosphate as a coenzyme in a transamination reaction. PyrP is pyridoxal phosphate, P is the apoenzyme (the polypeptide chain alone), and E is the active holoenzyme (polypeptide plus coenzyme).

# Biochemical Connections ALLIED HEALTH

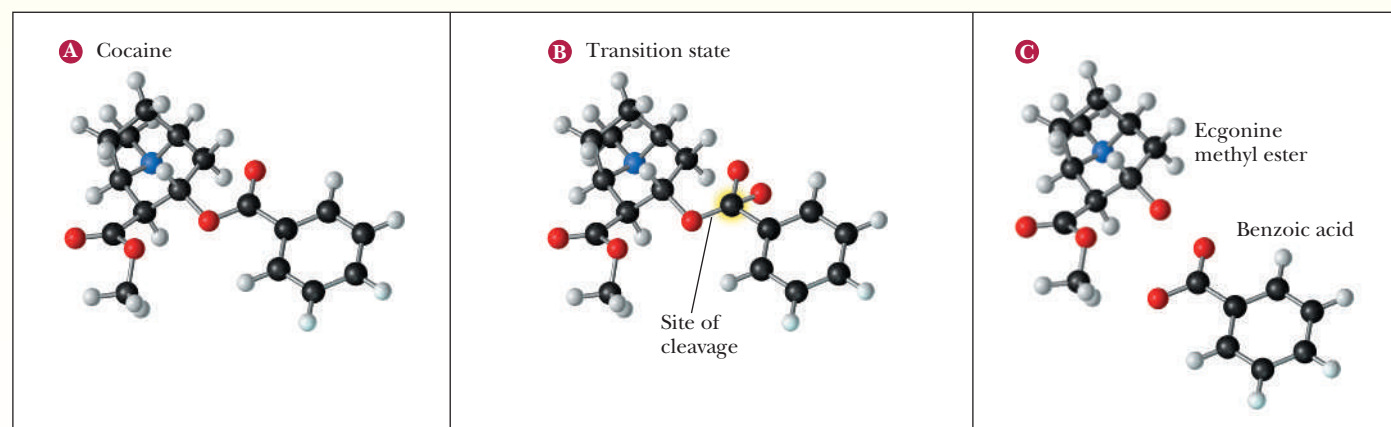
## Catalytic Antibodies against Cocaine

Many addictive drugs, such as heroin, operate by binding to a particular receptor in the neurons, mimicking the action of a neurotransmitter. When a person is addicted to such a drug, a common way to attempt to treat the addiction is to use a compound to block the receptor, thereby denying the drug's access to it. Cocaine addiction has always been difficult to treat, primarily because of its unique modus operandi. As shown, cocaine blocks the reuptake of the neurotransmitter dopamine. Thus, dopamine stays in the system longer, overstimulating the neurons and leading to the reward signals in the brain that lead to addiction. Using a drug to block a receptor would be of no use with cocaine addiction and would probably just make removal

of dopamine even more unlikely. Cocaine can be degraded by a specific esterase, an enzyme that hydrolyzes an ester bond that is part of cocaine's structure. In the process of this hydrolysis, the cocaine must pass through a transition state that changes its shape. Catalytic antibodies to the transition state of the hydrolysis of cocaine were created. When administered to patients suffering from cocaine addiction, the antibodies successfully hydrolyzed cocaine to two harmless degradation products—benzoic acid and ecgonine methyl ester. When degraded, the cocaine cannot block dopamine reuptake. No prolongation of the neuronal stimulus occurs, and the addictive effects of the drug vanish over time.



■ The mechanism of action of cocaine. (a) Dopamine acts as a neurotransmitter. It is released from the presynaptic neuron, travels across the synapse, and bonds to dopamine receptors on the postsynaptic neuron. It is later released and taken up into vesicles in the presynaptic neuron. (b) Cocaine increases the amount of time that dopamine is available to the dopamine receptors by blocking its uptake. (From *Scientific American*, Vol. 276(2), pp. 42–45. Reprinted by permission of Tomoyuki Narashima.)



■ Degradation of cocaine by esterases or catalytic antibodies. Cocaine (a) passes through a transition state (b) on its way to being hydrolyzed to benzoic acid and ecgonine methyl ester (c). Transition-state analogs are used to generate catalytic antibodies for this reaction. (From *Scientific American*, Vol. 276(2), pp. 42–45. Reprinted by permission of Tomoyuki Narashima.)

## Biochemical Connections

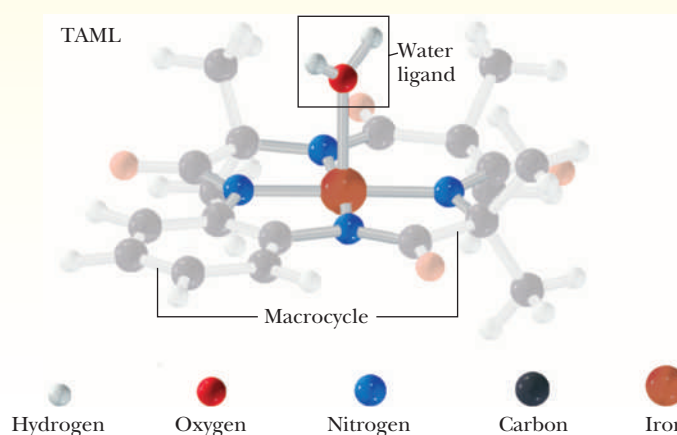
### ENVIRONMENTAL TOXICOLOGY

#### Catalysts for Green Chemistry

Billions of gallons of toxic wastes are dumped into the environment every year. Between the damage caused by our industrial lifestyles and the rampant growth in world population, many scientists have predicted that the Earth is headed for a global environmental collapse. In response, science and industry are both working on ways to reduce and limit the toxicity of compounds produced by industrial syntheses. This has led to the new field of **green chemistry**, in which alternative, less toxic compounds are slowly replacing their more toxic predecessors.

Nature has some detoxification systems of its own, often including hydrogen peroxide and oxygen. These two substances working together are capable of purifying water and cleansing industrial waste. However, in nature such reactions require an enzyme, such as peroxidase, to increase the rate of reaction to a significant level. Current research has come up with some synthetic molecules that possess an enzyme's ability to catalyze a needed reaction. One important set of these molecules is called **TAMLs** (tetra-amido macrocyclic ligands). The heart of the molecule is an iron atom linked to four nitrogen atoms, as shown in the figure, and the two remaining coordination sites are bound to water ligands. Attached to this central unit are carbon rings called the macrocycle. Just as the iron in hemoglobin is reactive and can bind to oxygen, the TAML takes advantage of similar properties. In this case, it reacts with  $\text{H}_2\text{O}_2$  to displace a water ligand. The  $\text{H}_2\text{O}_2$  then expels another water molecule, leaving a very reactive species with a large charge separation between the iron center and the anionic oxygen at the ligand site. This final molecule is

powerful enough to react with and destroy many chemical toxins. By adjusting the components of the TAML, researchers can tailor them for specific toxins, including versions that were able to deactivate more than 99% of spores of *Bacillus atrophaeus*, a bacterial species similar to anthrax. They have also been used to decolorize waste from pulp mills. The researchers working with TAMLs hope to design them to attack other infectious diseases and environmental pollutants.



## SUMMARY

**How are allosteric enzymes controlled?** Allosteric enzymes can be controlled by many different mechanisms, including inhibition and activation by reversibly binding molecules. Feedback inhibition is a common way to regulate an allosteric enzyme that is part of a complicated pathway.

**What is the concerted model for allosteric behavior?** In the concerted model for allosteric behavior, the binding of substrate, inhibitor, or activator to one subunit shifts the equilibrium between an active form of the enzyme, which binds substrate strongly, and an inactive form, which does not bind substrate strongly. The conformational change takes place in all subunits at the same time.

**What is the sequential model for allosteric behavior?** In the sequential model, the binding of substrate induces the conformational change in one subunit, and the change is subsequently passed along to other subunits.

**Does phosphorylation always increase enzyme activity?** Some enzymes are activated or inactivated depending on the presence or absence of phosphate groups. This kind of covalent

modification can be combined with allosteric interactions to allow for a high degree of control over enzymatic pathways.

**How do we determine the essential amino acid residues?** Several questions arise about the events that occur at the active site of an enzyme in the course of a reaction. Some of the most important of these questions address the nature of the critical amino acid residues, their spatial arrangement, and the mechanism of the reaction. The use of labeling reagents and X-ray crystallography allows us to determine the amino acids that are located in the active site and critical to the catalytic mechanism.

**How does the architecture of the active site affect catalysis?** Chymotrypsin is a good example of an enzyme for which most of the questions about its mechanism of action have been answered. Its critical amino acid residues have been determined to be serine 195 and histidine 57. The complete three-dimensional structure of chymotrypsin, including the architecture of the active site, has been determined by X-ray crystallography.



**How do the critical amino acids catalyze the chymotrypsin reaction?** Nucleophilic attack by serine is the main feature of the mechanism, with histidine hydrogen-bonded to serine in the course of the reaction. The reaction takes place in two phases. In the first phase, serine is the nucleophile, and there is an acyl-enzyme intermediate. In the second phase, water acts as the nucleophile and the acyl-enzyme intermediate is hydrolyzed.

**What are the most common types of reactions?** Common organic reaction mechanisms, such as nucleophilic substitution

and general acid–base catalysis, are known to play roles in enzymatic catalysis.

**How do we determine the nature of the transition state?**

Understanding of the nature of catalysis has been aided by the use of transition-state analogs, molecules that mimic the transition state. The compounds usually bind to the enzyme better than the natural substrate and help verify the mechanism. They can also be used to develop potent inhibitors or to create antibodies with catalytic activity, called abzymes.

## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

### 7.1 The Behavior of Allosteric Enzymes

- Recall** What features distinguish enzymes that undergo allosteric control from those that obey the Michaelis–Menten equation?
- Recall** What is the metabolic role of aspartate transcarbamoylase?
- Recall** What molecule acts as a positive effector (activator) of ATCase? What molecule acts as an inhibitor?
- Recall** Is the term  $K_M$  used with allosteric enzymes? What about competitive and noncompetitive inhibition? Explain.
- Recall** What is a K system?
- Recall** What is a V system?
- Recall** What is a homotropic effect? What is a heterotropic effect?
- Recall** What is the structure of ATCase?
- Recall** How is the cooperative behavior of allosteric enzymes reflected in a plot of reaction rate against substrate concentration?
- Recall** Does the behavior of allosteric enzymes become more or less cooperative in the presence of inhibitors?
- Recall** Does the behavior of allosteric enzymes become more or less cooperative in the presence of activators?
- Recall** Explain what is meant by  $K_{0.5}$ .
- Reflect and Apply** Explain the experiment used to determine the structure of ATCase. What happens to the activity and regulatory activities when the subunits are separated?

### 7.2 The Concerted and Sequential Models for Allosteric Enzymes

- Recall** Distinguish between the concerted and sequential models for the behavior of allosteric enzymes.
- Recall** Which allosteric model can explain negative cooperativity?
- Recall** With the concerted model, what conditions favor greater cooperativity?
- Recall** With respect to the concerted model, what is the L value? What is the  $\epsilon$  value?
- Reflect and Apply** Is it possible to envision models for the behavior of allosteric enzymes other than the ones that we have seen in this chapter?

### 7.3 Control of Enzyme Activity by Phosphorylation

- Biochemical Connections** What has been the historical method used in drug design?
- Biochemical Connections** What is a main reason for side effects with traditional drugs that bind to the active site of a receptor?

- Biochemical Connections** What are three advantages of using allosteric drugs as opposed to orthosteric ones?
- Biochemical Connections** How does Valium work?
- Biochemical Connections** Why is taking too much Valium not as dangerous as taking too much Phenobarbital?
- Biochemical Connections** What are two recent allosteric drugs that are currently on the market? What do they do?
- Recall** What is the function of a protein kinase?
- Recall** What amino acids are often phosphorylated by kinases?
- Reflect and Apply** What are some possible advantages to the cell in combining phosphorylation with allosteric control?
- Reflect and Apply** Explain how phosphorylation is involved in the function of the sodium–potassium ATPase.
- Reflect and Apply** Explain how glycogen phosphorylase is controlled allosterically and by covalent modification.

### 7.4 Zymogens

- Recall** Name three proteins that are subject to the control mechanism of zymogen activation.
- Biochemical Connections** List three proteases and their substrates.
- Recall** What are caspases?
- Reflect and Apply** Explain why cleavage of the bond between arginine 15 and isoleucine 16 of chymotrypsinogen activates the zymogen.
- Reflect and Apply** Why is it necessary or advantageous for the body to make zymogens?
- Reflect and Apply** Why is it necessary or advantageous for the body to make inactive hormone precursors?

### 7.5 The Nature of the Active Site

- Recall** What is apoptosis?
- Recall** What diseases are linked to apoptosis?
- Recall** What are the two essential amino acids in the active site of chymotrypsin?
- Recall** Why does the enzyme reaction for chymotrypsin proceed in two phases?
- Reflect and Apply** Briefly describe the role of nucleophilic catalysis in the mechanism of the chymotrypsin reaction.
- Reflect and Apply** Explain the function of histidine 57 in the mechanism of chymotrypsin.
- Reflect and Apply** Explain why the second phase of the chymotrypsin mechanism is slower than the first phase.

43. **Reflect and Apply** Explain how the  $pK_a$  for histidine 57 is important to its role in the mechanism of chymotrypsin action.
44. **Reflect and Apply** An inhibitor that specifically labels chymotrypsin at histidine 57 is N-tosylamido-L-phenylethyl chloromethyl ketone. How would you modify the structure of this inhibitor to label the active site of trypsin?

### 7.6 Chemical Reactions Involved in Enzyme Mechanisms

45. **Reflect and Apply** What properties of metal ions make them useful cofactors?
46. **Recall** In biochemistry mechanisms, what group is often attacked by a nucleophile?
47. **Reflect and Apply** What is meant by general acid catalysis with respect to enzyme mechanisms?
48. **Reflect and Apply** Explain the difference between an  $S_N1$  reaction mechanism and an  $S_N2$  reaction mechanism.
49. **Reflect and Apply** Which of the two reaction mechanisms in Question 48 is likely to cause the loss of stereospecificity? Why?
50. **Reflect and Apply** An experiment is performed to test a suggested mechanism for an enzyme-catalyzed reaction. The results fit the model exactly (to within experimental error). Do the results prove that the mechanism is correct? Why or why not?

### 7.7 The Active Site and Transition States

51. **Reflect and Apply** What would be the characteristics of a transition-state analog for the chymotrypsin reaction?
52. **Reflect and Apply** What is the relationship between a transition-state analog and the induced-fit model of enzyme kinetics?
53. **Reflect and Apply** Explain how a researcher makes an abzyme. What is the purpose of an abzyme?
54. **Biochemical Connections** Why can cocaine addiction not be treated with a drug that blocks the cocaine receptor?
55. **Biochemical Connections** Explain how abzymes can be used to treat cocaine addiction.

### 7.8 Coenzymes

56. **Recall** List three coenzymes and their functions.
57. **Recall** How are coenzymes related to vitamins?
58. **Recall** What type of reaction uses vitamin B<sub>6</sub>?
59. **Reflect and Apply** Suggest a role for coenzymes based on reaction mechanisms.
60. **Reflect and Apply** An enzyme uses  $NAD^+$  as a coenzyme. Using Figure 7.19, predict whether a radiolabeled  $H^-$  ion would tend to appear preferentially on one side of the nicotinamide ring as opposed to the other side.

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).

# Lipids and Proteins Are Associated in Biological Membranes

# 8

© David W. Phillips/The Population Council/  
Photo Researchers, Inc.

Electron micrograph of a fat cell. Much of the cell volume is taken up by lipid droplets.

## 8.1 The Definition of a Lipid

### What are lipids?

Lipids are compounds that occur frequently in nature. They are found in places as diverse as egg yolks and the human nervous system and are an important component of plant, animal, and microbial membranes. The definition of a lipid is based on solubility. Lipids are marginally soluble (at best) in water but readily soluble in organic solvents, such as chloroform or acetone.

Fats and oils are typical lipids in terms of their solubility, but that fact does not really define their chemical nature. In terms of chemistry, lipids are a mixed bag of compounds that share some properties based on structural similarities, mainly a preponderance of nonpolar groups.

Classified according to their chemical nature, lipids fall into two main groups. One group, which consists of open-chain compounds with polar head groups and long nonpolar tails, includes *fatty acids*, *triacylglycerols*, *sphingolipids*, *phosphoacylglycerols*, and *glycolipids*. The second major group consists of fused-ring compounds, the *steroids*; an important representative of this group is cholesterol.

## 8.2 The Chemical Natures of the Lipid Types

### What are fatty acids?

A fatty acid has a carboxyl group at the polar end and a hydrocarbon chain at the nonpolar tail. Fatty acids are **amphipathic** compounds because the carboxyl group is hydrophilic and the hydrocarbon tail is hydrophobic. The carboxyl group can ionize under the proper conditions.

A fatty acid that occurs in a living system normally contains an even number of carbon atoms, and the hydrocarbon chain is usually unbranched (Figure 8.1). If there are carbon-carbon double bonds in the chain, the fatty acid is *unsaturated*; if there are only single bonds, the fatty acid is *saturated*. Tables 8.1 and 8.2 list a few examples of the two classes. In unsaturated fatty acids, the stereochemistry at the double bond is usually *cis* rather than *trans*. The difference between *cis* and *trans* fatty acids is very important to their overall shape. A *cis* double bond puts a kink in the long-chain hydrocarbon tail, whereas the shape of a *trans* fatty acid is like that of a saturated fatty acid in its fully extended conformation. Note that the double bonds are isolated from one another by several singly bonded carbons; fatty acids do not normally have conjugated double-bond systems. The notation used for fatty acids indicates the number of carbon atoms and the number of double bonds. In this system, 18:0 denotes an 18-carbon saturated fatty acid with no double bonds, and 18:1 denotes an 18-carbon fatty acid with one double bond. Note that in the unsaturated fatty acids in Table 8.2 (except arachidonic acid), there is a double bond at the ninth carbon atom from the carboxyl end. The position of the double bond results

## Chapter Outline

### 8.1 The Definition of a Lipid

- What are lipids?

### 8.2 The Chemical Natures of the Lipid Types

- What are fatty acids?
- What are triacylglycerols?
- What are phosphoacylglycerols?
- What are waxes and sphingolipids?
- What are glycolipids?
- What are steroids?

### 8.3 Biological Membranes

- What is the structure of lipid bilayers?
- How does the composition of the bilayer affect its properties?

### 8.4 The Kinds of Membrane Proteins

- How are proteins associated with the bilayer in membranes?

### 8.5 The Fluid-Mosaic Model of Membrane Structure

- How do proteins and the lipid bilayer interact with each other in membranes?

### 8.6 The Functions of Membranes

- How does transport through membranes take place?
- How do membrane receptors work?

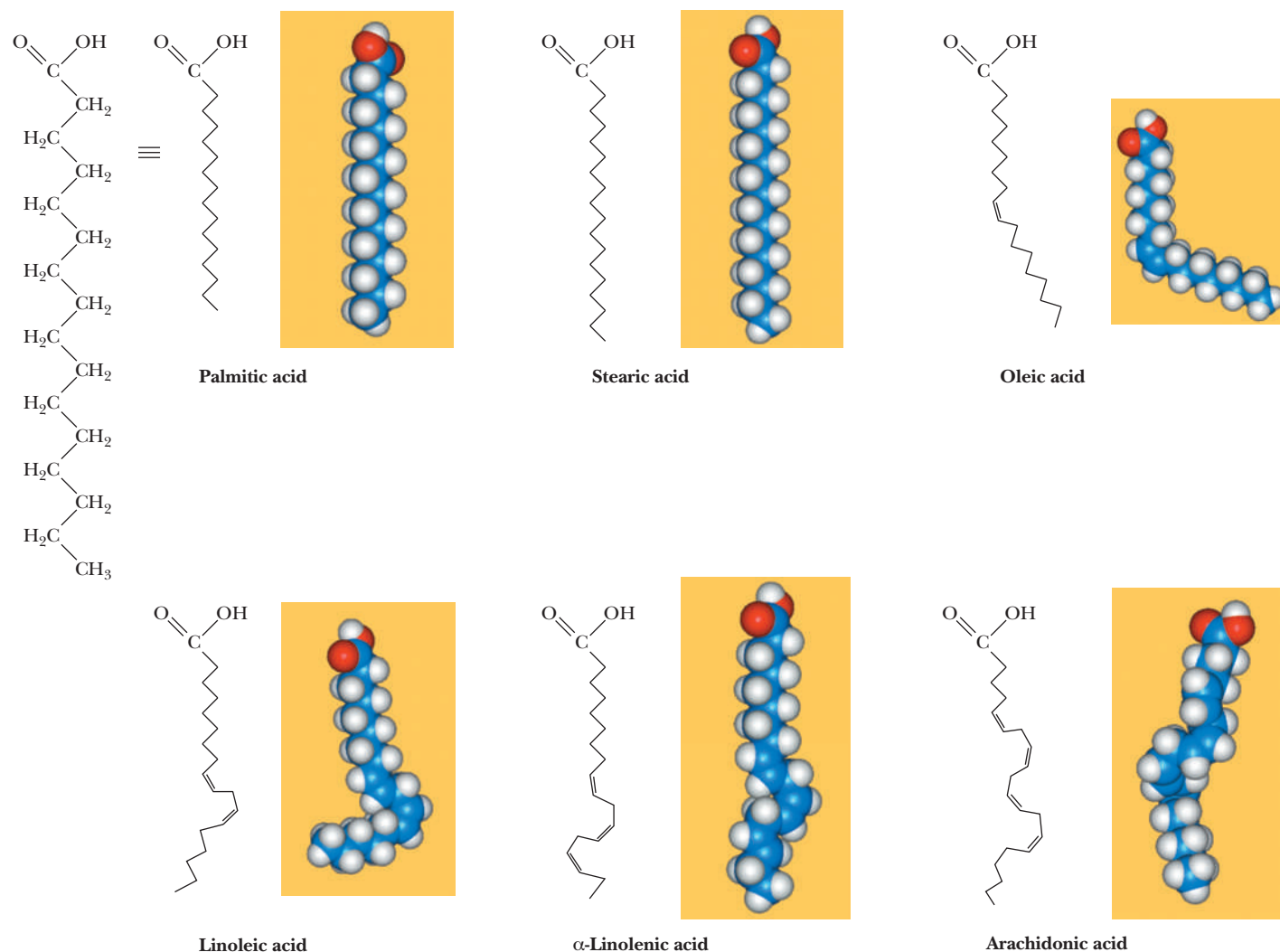
### 8.7 Lipid-Soluble Vitamins and Their Functions

- What is the role of lipid-soluble vitamins in the body?

### 8.8 Prostaglandins and Leukotrienes

- What do prostaglandins and leukotrienes have to do with lipids?

**OWL** Online homework for this chapter may be assigned in OWL.



■ **FIGURE 8.1** The structures of some typical fatty acids. Note that most naturally occurring fatty acids contain even numbers of carbon atoms and that the double bonds are nearly always *cis* and rarely conjugated.

from the way unsaturated fatty acids are synthesized in organisms (Section 21.6). Unsaturated fatty acids have lower melting points than saturated ones. Plant oils are liquid at room temperature because they have higher proportions of unsaturated fatty acids than do animal fats, which tend to be solids. Conversion of oils to fats is a commercially important process. It involves hydrogenation, the process of adding hydrogen across the double bond of unsaturated fatty acids to produce the saturated counterpart. Oleomargarine, in particular, uses partially hydrogenated vegetable oils, which tend to include *trans* fatty acids (see the Biochemical Connections box on page 205).

Fatty acids are rarely found free in nature, but they form parts of many commonly occurring lipids.

### What are triacylglycerols?

**Glycerol** is a simple compound that contains three hydroxyl groups (Figure 8.2). When all three of the alcohol groups form ester linkages with fatty acids, the resulting compound is a **triacylglycerol**; an older name for this type of compound is *triglyceride*. Note that the three ester groups are the polar part of the molecule, whereas the tails of the fatty acids are nonpolar. It is usual for three

### TABLE 8.1

Typical Naturally Occurring Saturated Fatty Acids			
Acid	Number of Carbon Atoms	Formula	Melting Point (°C)
Lauric	12	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>10</sub> CO <sub>2</sub> H	44
Myristic	14	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>12</sub> CO <sub>2</sub> H	58
Palmitic	16	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>14</sub> CO <sub>2</sub> H	63
Stearic	18	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>16</sub> CO <sub>2</sub> H	71
Arachidic	20	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>18</sub> CO <sub>2</sub> H	77

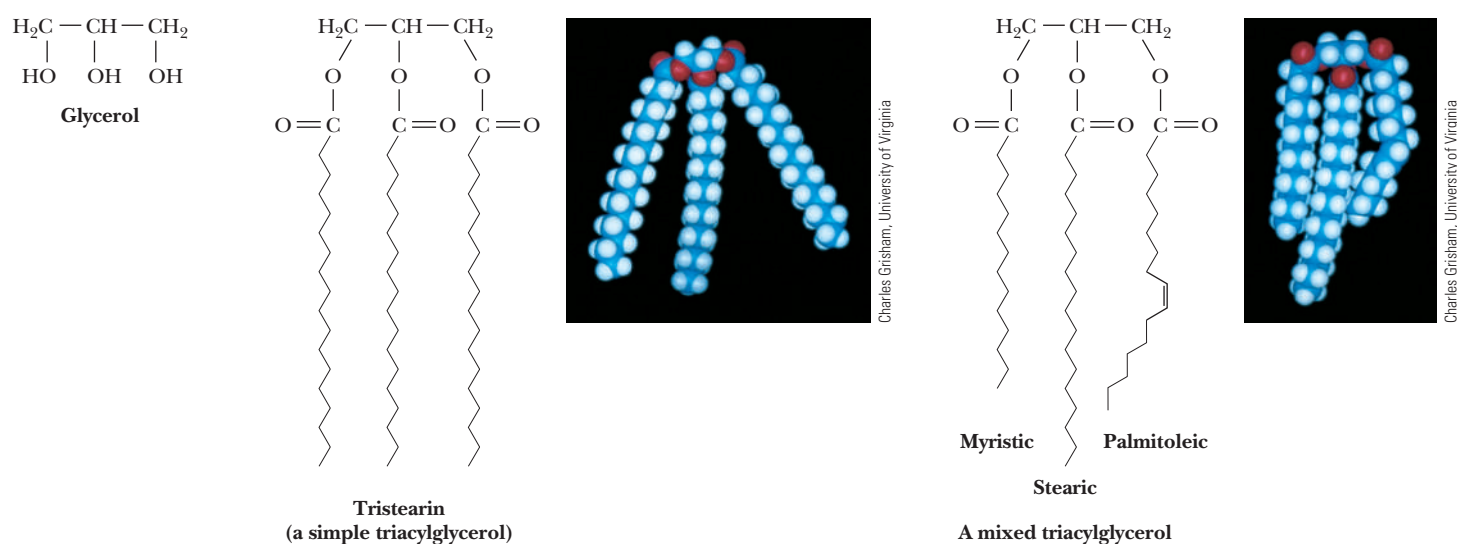
### TABLE 8.2

Typical Naturally Occurring Unsaturated Fatty Acids				
Acid	Number of Carbon Atoms	Degree of Unsaturation*	Formula	Melting Point (°C)
Palmitoleic	16	16:1—Δ <sup>9</sup>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>5</sub> CH=CH(CH <sub>2</sub> ) <sub>7</sub> CO <sub>2</sub> H	−0.5
Oleic	18	18:1—Δ <sup>9</sup>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>7</sub> CH=CH(CH <sub>2</sub> ) <sub>7</sub> CO <sub>2</sub> H	16
Linoleic	18	18:2—Δ <sup>9,12</sup>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>4</sub> CH=CH(CH <sub>2</sub> )CH=CH(CH <sub>2</sub> ) <sub>7</sub> CO <sub>2</sub> H	−5
Linolenic	18	18:3—Δ <sup>9,12,15</sup>	CH <sub>3</sub> (CH <sub>2</sub> CH=CH) <sub>3</sub> (CH <sub>2</sub> ) <sub>7</sub> CO <sub>2</sub> H	−11
Arachidonic	20	20:4—Δ <sup>5,8,11,14</sup>	CH <sub>3</sub> (CH <sub>2</sub> ) <sub>4</sub> CH=CH(CH <sub>2</sub> ) <sub>4</sub> (CH <sub>2</sub> ) <sub>2</sub> CO <sub>2</sub> H	−50

\*Degree of unsaturation refers to the number of double bonds. The superscript indicates the position of double bonds. For example,  $\Delta^9$  refers to a double bond at the ninth carbon atom from the carboxyl end of the molecule.

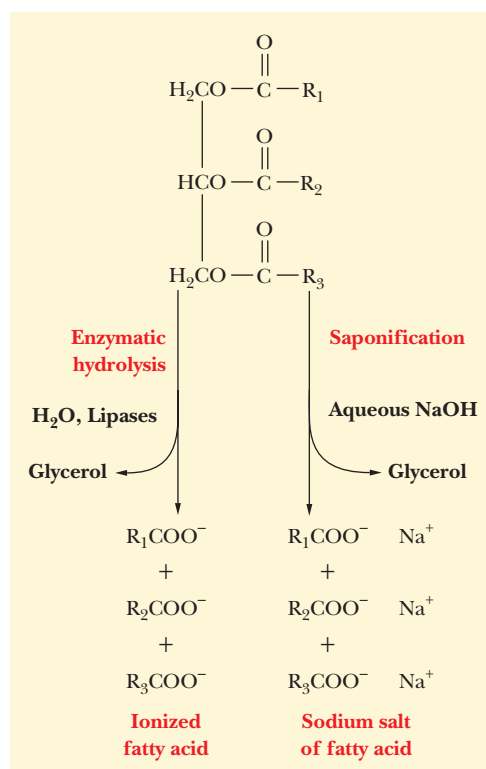
different fatty acids to be esterified to the alcohol groups of the same glycerol molecule. Triacylglycerols do not occur as components of membranes (as do other types of lipids), but they accumulate in adipose tissue (primarily fat cells) and provide a means of storing fatty acids, particularly in animals. They serve as concentrated stores of metabolic energy. Complete oxidation of fats yields about  $9 \text{ kcal g}^{-1}$ , in contrast with  $4 \text{ kcal g}^{-1}$  for carbohydrates and proteins (see Sections 21.3 and 24.2).

When an organism uses fatty acids, the ester linkages of triacylglycerols are hydrolyzed by enzymes called **lipases**. The same hydrolysis reaction can take



■ **FIGURE 8.2** Triacylglycerols are formed from glycerol and fatty acids.





■ **FIGURE 8.3** Hydrolysis of triacylglycerols. The term *saponification* refers to the reactions of glyceryl ester with sodium or potassium hydroxide to produce a soap, which is the corresponding salt of the long-chain fatty acid.

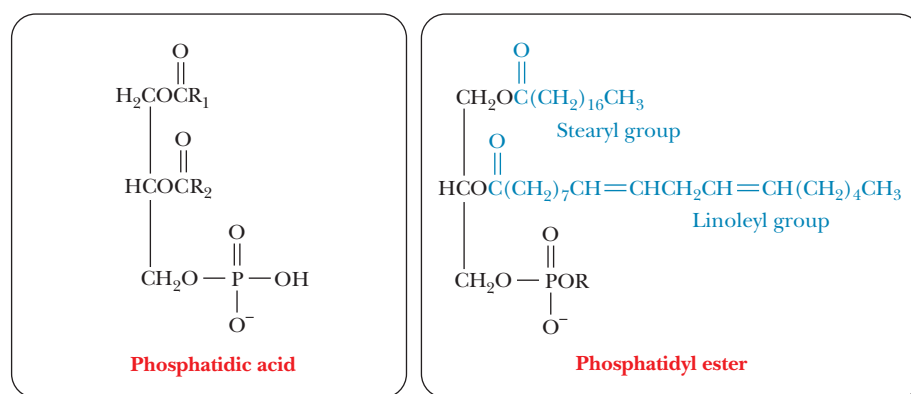
place outside organisms, with acids or bases as catalysts. When a base such as sodium hydroxide or potassium hydroxide is used, the products of the reaction, which is called *saponification* (Figure 8.3), are glycerol and the sodium or potassium salts of the fatty acids. These salts are soaps. When soaps are used with hard water, the calcium and magnesium ions in the water react with the fatty acids to form a precipitate—the characteristic scum left on the insides of sinks and bathtubs. The other product of saponification, glycerol, is used in creams and lotions as well as in the manufacture of nitroglycerin.

### What are phosphoacylglycerols?

It is possible for one of the alcohol groups of glycerol to be esterified by a phosphoric acid molecule rather than by a carboxylic acid. In such lipid molecules, two fatty acids are also esterified to the glycerol molecule. The resulting compound is called a **phosphatidic acid** (Figure 8.4a). Fatty acids are usually monoprotic acids with only one carboxyl group able to form an ester bond, but phosphoric acid is triprotic and thus can form more than one ester linkage. One molecule of phosphoric acid can form ester bonds both to glycerol and to some other alcohol, creating a *phosphatidyl ester* (Figure 8.4b).

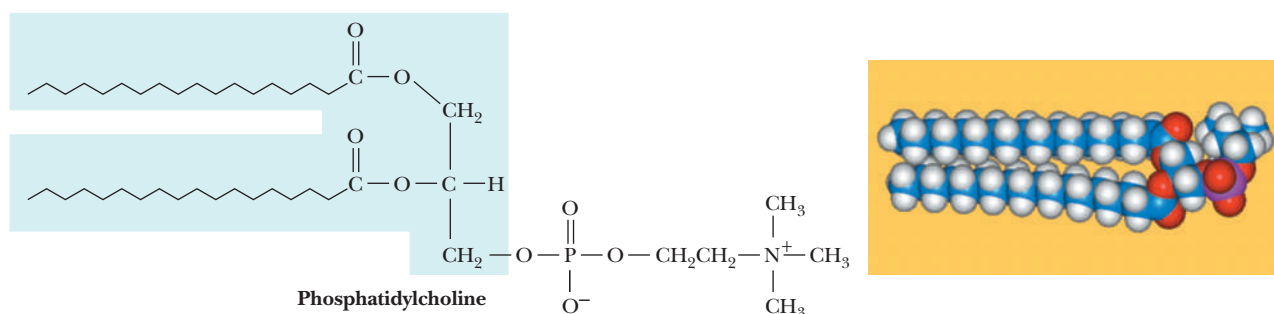
Phosphatidyl esters are classed as **phosphoacylglycerols**. The natures of the fatty acids vary widely, as they do in triacylglycerols. As a result, the names of the types of lipids (such as triacylglycerols and phosphoacylglycerols) that contain fatty acids must be considered generic names.

The classification of a phosphatidyl ester depends on the nature of the second alcohol esterified to the phosphoric acid. Some of the most important lipids in this class are *phosphatidyl ethanolamine* (cephalin), *phosphatidyl serine*, *phosphatidyl choline* (lecithin), *phosphatidyl inositol*, *phosphatidyl glycerol*, and *diphosphatidyl glycerol* (cardiolipin) (Figure 8.5). In each of these types of compounds, the nature of the fatty acids in the molecule can vary widely. All these compounds have long, nonpolar, hydrophobic tails and polar, highly hydrophilic head groups and thus are markedly amphipathic. (We have already seen this characteristic in fatty acids.) In a phosphoacylglycerol, the polar head group is charged, because the phosphate group is ionized at neutral pH. A positively charged amino group is also frequently contributed by an amino alcohol esterified to the phosphoric acid. Phosphoacylglycerols are important components of biological membranes.

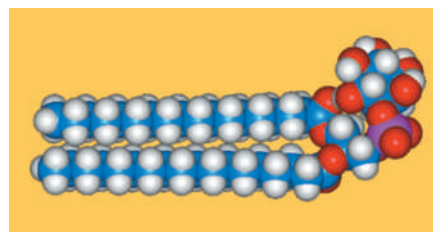
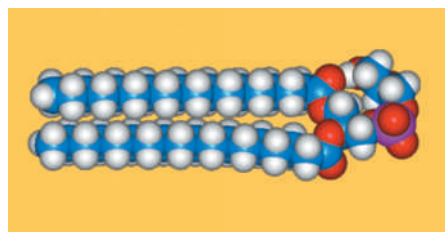
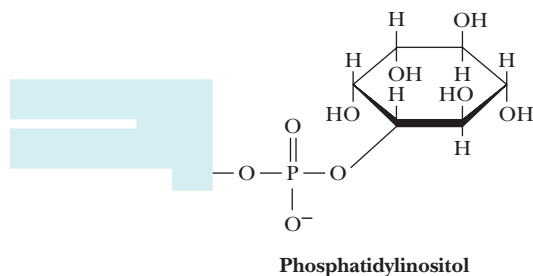
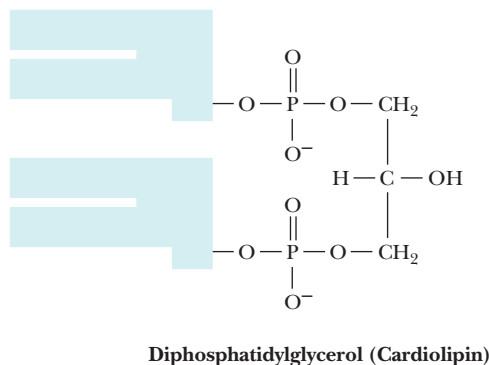
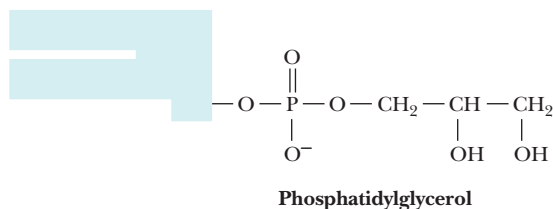
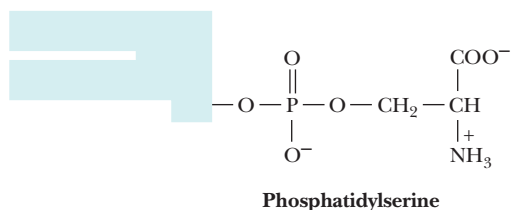
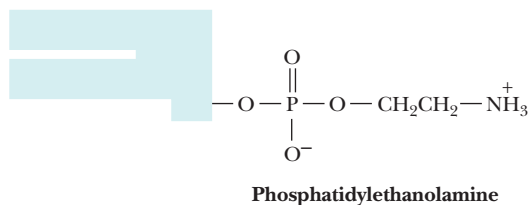


- A** A phosphatidic acid, in which glycerol is esterified to phosphoric acid and to two different carboxylic acids.  $\text{R}_1$  and  $\text{R}_2$  represent the hydrocarbon chains of the two carboxylic acids.
- B** A phosphatidyl ester (phosphoacylglycerol). Glycerol is esterified to two carboxylic acids, stearic acid and linoleic acid, as well as to phosphoric acid. Phosphoric acid, in turn, is esterified to a second alcohol, ROH.

■ **FIGURE 8.4** The molecular architecture of phosphoacylglycerols.



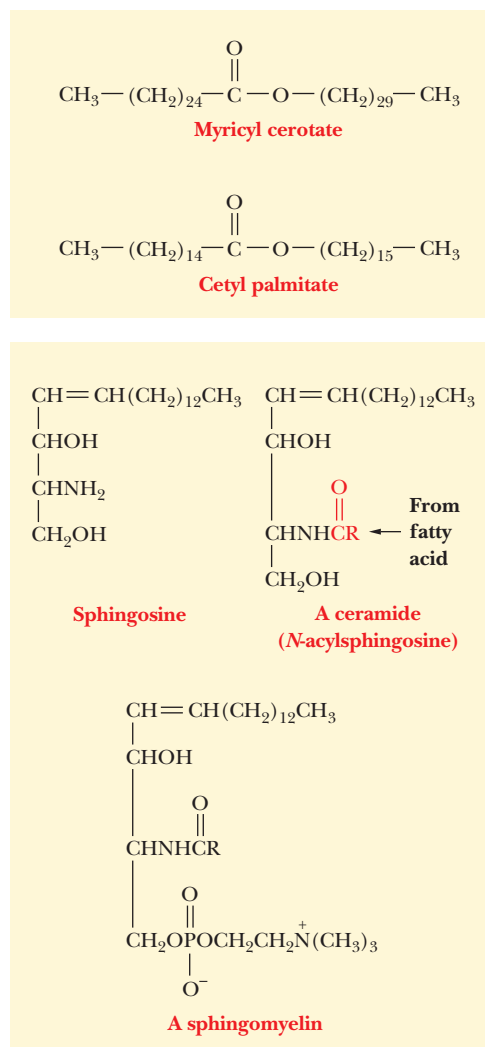
## GLYCEROLIPIDS WITH OTHER HEAD GROUPS:



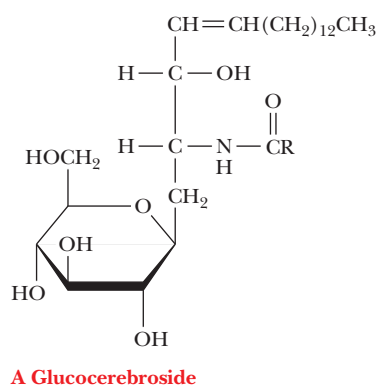
■ **FIGURE 8.5** Structures of some phosphoacylglycerols and space-filling models of phosphatidylcholine, phosphatidylglycerol, and phosphatidylinositol.

### What are waxes and sphingolipids?

**Waxes** are complex mixtures of esters of long-chain carboxylic acids and long-chain alcohols. They frequently serve as protective coatings for both plants and animals. In plants, they coat stems, leaves, and fruit; in animals, they are found on fur, feathers, and skin. Myricyl cerotate (Figure 8.6), the principal component of carnauba wax, is produced by the Brazilian wax palm. Carnauba wax is extensively used in floor wax and automobile wax. The principal component of spermaceti, a wax produced by whales, is cetyl palmitate (Figure 8.6).



■ **FIGURE 8.6** Structures of some waxes and sphingolipids.



■ **FIGURE 8.7** Structure of a glucocerebroside.

The use of spermaceti as a component of cosmetics made it one of the most highly prized products of 19th-century whaling efforts.

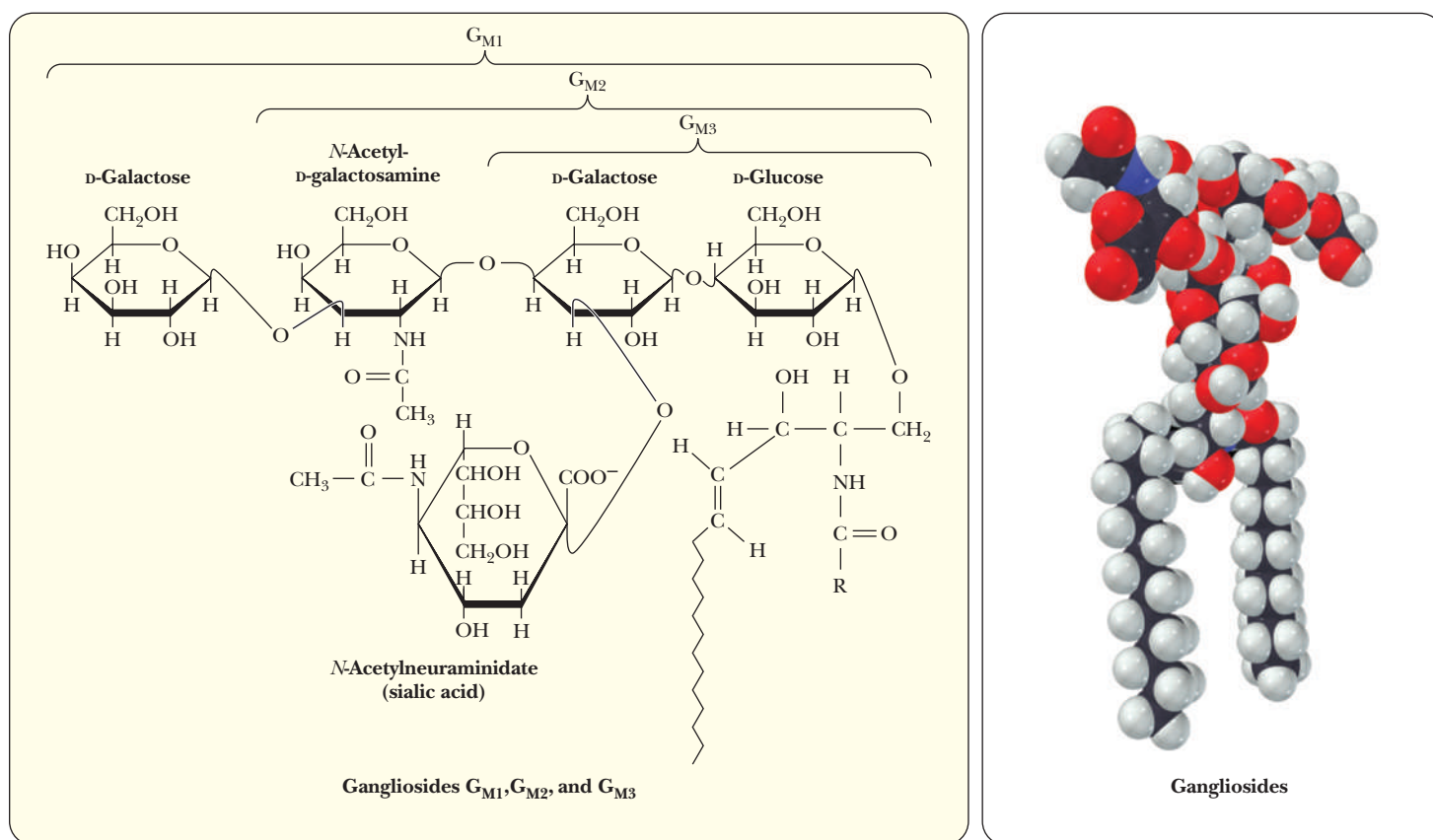
**Sphingolipids** do not contain glycerol, but they do contain the long-chain amino alcohol sphingosine, from which this class of compounds takes its name (Figure 8.6). Sphingolipids are found in both plants and animals; they are particularly abundant in the nervous system. The simplest compounds of this class are the ceramides, which consist of one fatty acid linked to the amino group of sphingosine by an amide bond (Figure 8.6). In **sphingomyelins**, the primary alcohol group of sphingosine is esterified to phosphoric acid, which, in turn, is esterified to another amino alcohol, choline (Figure 8.6). Note the structural similarities between sphingomyelin and other phospholipids. Two long hydrocarbon chains are attached to a backbone that contains alcohol groups. One of the alcohol groups of the backbone is esterified to phosphoric acid. A second alcohol—choline, in this case—is also esterified to the phosphoric acid. We have already seen that choline occurs in phosphoacylglycerols. Sphingomyelins are amphipathic; they occur in cell membranes in the nervous system.

### What are glycolipids?

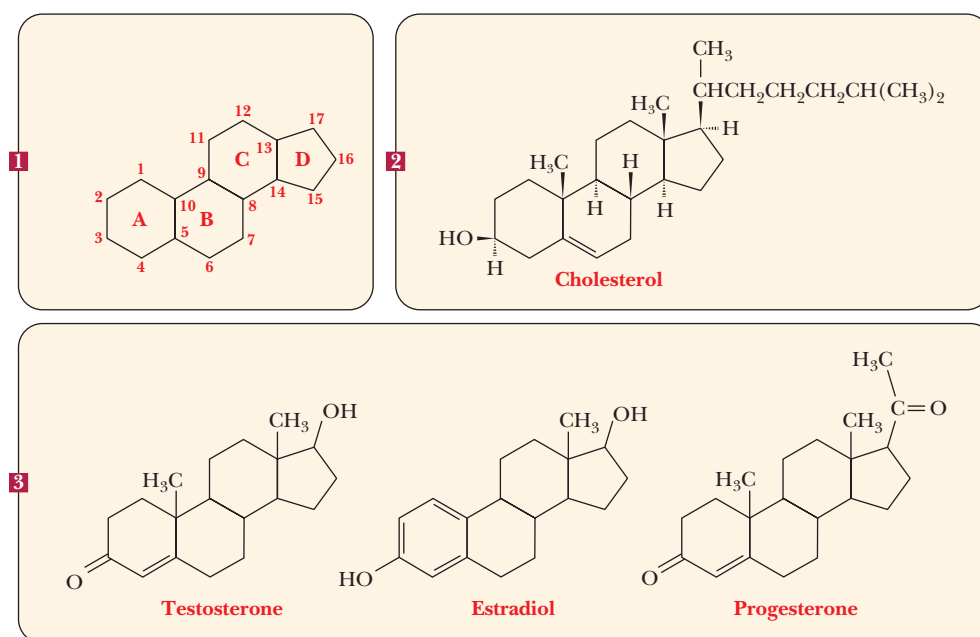
If a carbohydrate is bound to an alcohol group of a lipid by a glycosidic linkage (see Section 16.3 for a discussion of glycosidic linkages), the resulting compound is a **glycolipid**. Quite frequently, **ceramides** (see Figure 8.6) are the parent compounds for glycolipids, and the glycosidic bond is formed between the primary alcohol group of the ceramide and a sugar residue. The resulting compound is called a **cerebroside**. In most cases, the sugar is glucose or galactose; for example, a glucocerebroside is a cerebroside that contains glucose (Figure 8.7). As the name indicates, cerebrosides are found in nerve and brain cells, primarily in cell membranes. The carbohydrate portion of these compounds can be very complex. Gangliosides are examples of glycolipids with a complex carbohydrate moiety that contains more than three sugars. One of them is always a sialic acid (Figure 8.8). These compounds are also referred to as acidic glycosphingolipids because of their net negative charge at neutral pH. Glycolipids are often found as markers on cell membranes and play a large role in tissue and organ specificity. Gangliosides are also present in large quantities in nerve tissues. Their biosynthesis and breakdown are discussed in Section 21.7.

### What are steroids?

Many compounds of widely differing functions are classified as **steroids** because they have the same general structure: a fused-ring system consisting of three six-membered rings (the A, B, and C rings) and one five-membered ring (the D ring). There are many important steroids, including sex hormones. (See Section 24.3 for more steroids of biological importance.) The steroid that is of most interest in our discussion of membranes is **cholesterol** (Figure 8.9). The only hydrophilic group in the cholesterol structure is the single hydroxyl group. As a result, the molecule is highly hydrophobic. Cholesterol is widespread in biological membranes, especially in animals, but it does not occur in prokaryotic cell membranes. The presence of cholesterol in membranes can modify the role of membrane-bound proteins. Cholesterol has a number of important biological functions, including its role as a precursor of other steroids and of vitamin D<sub>3</sub>. We will see a five-carbon structural motif (the isoprene unit) that is common to steroids and to fat-soluble vitamins, which is an indication of their biosynthetic relationship (Sections 8.7 and 21.8). However, cholesterol is best known for its harmful effects on health when it is present in excess in the blood. It plays a role in the development of *atherosclerosis*, a condition in which lipid deposits block the blood vessels and lead to heart disease (see Section 21.8).



■ **FIGURE 8.8** The structures of several important gangliosides. Also shown is a space-filling model of ganglioside  $G_{M1}$ .



■ **FIGURE 8.9** Structures of some steroids. (1) The fused-ring structure of steroids. (2) Cholesterol. (3) Some steroid sex hormones.

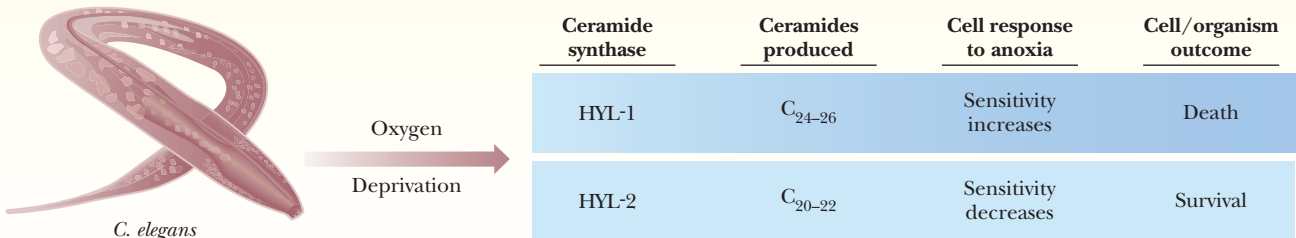
# Biochemical Connections ALLIED HEALTH

## Ceramides, Oxygen, Cancer, and Strokes

The connection between carbon chain length in ceramides and the response to oxygen deprivation (hypoxia) may have implications for serious issues in human health, including cancer and strokes. Cancer cells are resistant to hypoxia, and oxygen deprivation in cells is key to the damage caused by heart attack and strokes. The connection in question came to light in studies on ceramide synthases in the roundworm *Caenorhabditis elegans*, specifically in their response to oxygen deprivation. Ceramides with different chain lengths are produced in response to low oxygen levels, having different effects on survival. The question immediately arises as to whether human ceramide synthases play a role in response to the oxygen deprivation in cells that accompanies a heart attack or a stroke.

As shown in the figure, even-numbered carbon chain lengths of 20 to 26 are crucial. Production of the longer chain-length ceramide, catalyzed by one synthase, leads to cell death. Production of shorter chain-length ceramides, catalyzed by the other synthase, leads to cell survival.

In general, ceramide production leads to cell death, so this survival raises a number of questions. The complete genome of *C. elegans* is known, making it possible to analyze mutations involved in the exceptions that lead to survival. This information can have implications in human health. It will make it possible to design inhibitors for the specific ceramide synthases involved in cell death in heart attacks and strokes. Beyond that, it is possible to envision ways of controlling hypoxia so as to kill cancer cells.



(From Ceramides—Friend or Foe in Hypoxia? by C. Michael Crowder (17 April 2009) Science 324 (5925), 343. Reprinted with permission from AAAS.)

## 8.3 Biological Membranes

Every cell has a cell membrane (also called a plasma membrane); eukaryotic cells also have membrane-enclosed organelles, such as nuclei and mitochondria. The molecular basis of the membrane’s structure lies in its lipid and protein components. Now it is time to see how the interaction between the lipid bilayer and membrane proteins determines membrane function. Membranes not only separate cells from the external environment but also play important roles in transport of specific substances into and out of cells. In addition, a number of important enzymes are found in membranes and depend on this environment for their function.

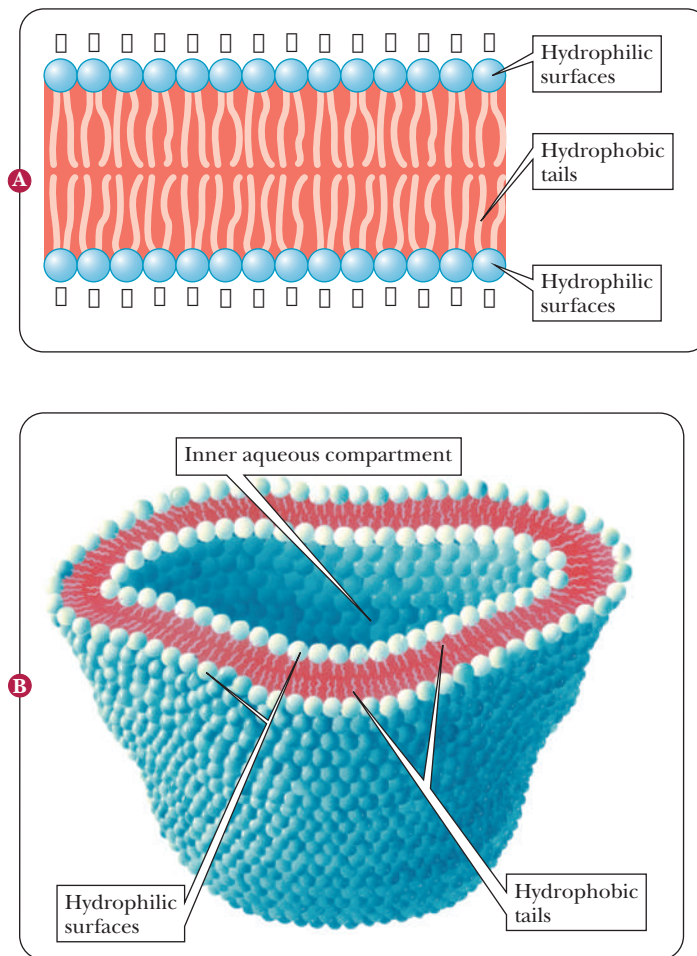
*Phosphoglycerides* are prime examples of amphipathic molecules, and they are the principal lipid components of membranes. The existence of *lipid bilayers* depends on hydrophobic interactions, as described in Section 4.6. These bilayers are frequently used as models for biological membranes because they have many features in common, such as a hydrophobic interior and an ability to control the transport of small molecules and ions, but they are simpler and easier to work with in the laboratory than biological membranes.

The most important difference between lipid bilayers and cell membranes is that the latter contain proteins as well as lipids. The protein component of a membrane can make up from 20% to 80% of its total weight. An understanding of membrane structure requires knowledge of how the protein and lipid components contribute to the properties of the membrane.

### What is the structure of lipid bilayers?

Biological membranes contain, in addition to phosphoglycerides, glycolipids as part of the lipid component. Steroids are present in eukaryotes—cholesterol in animal membranes and similar compounds, called phytosterols, in plants.





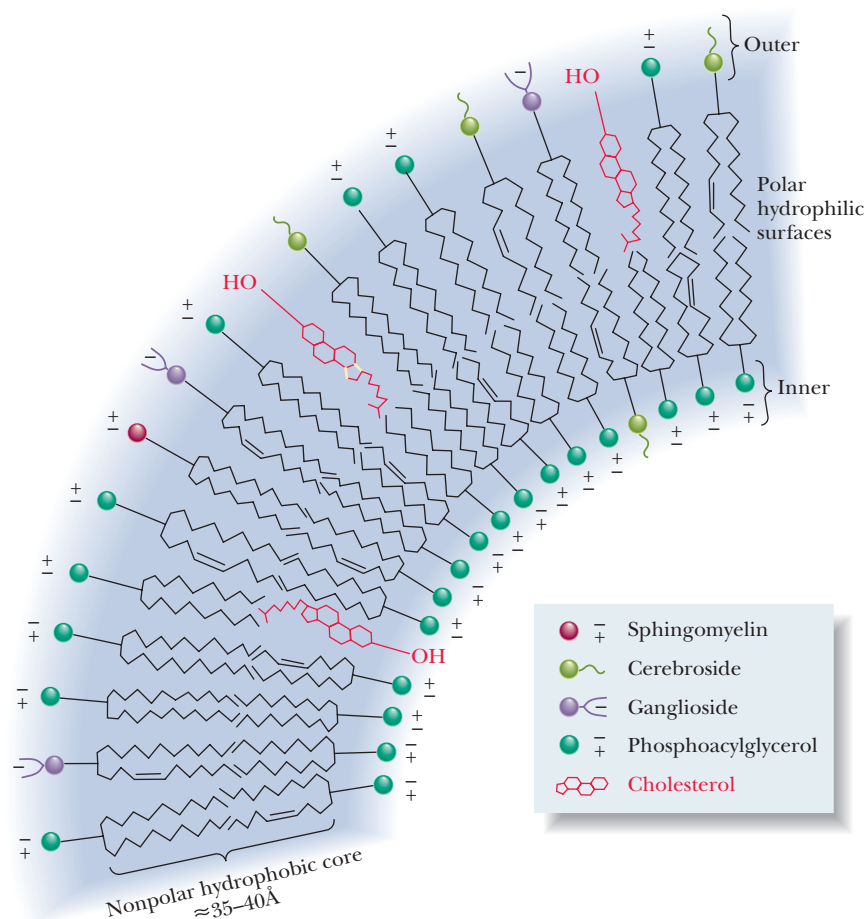
■ **FIGURE 8.10 Lipid bilayers.** (a) Schematic drawing of a portion of a bilayer consisting of phospholipids. The polar surface of the bilayer contains charged groups. The hydrocarbon “tails” lie in the interior of the bilayer. (b) Cutaway view of a lipid bilayer vesicle. Note the aqueous inner compartment and the fact that the inner layer is more tightly packed than the outer layer. (From Bretscher, M. S. *The Molecules of the Cell Membrane*. Scientific American, October 1985, p. 103. Art by Dana Burns-Pizer.)

In the **lipid bilayer** part of the membrane (Figure 8.10), the polar head groups are in contact with water, and the nonpolar tails lie in the interior of the membrane. The whole bilayer arrangement is held together by noncovalent interactions, such as van der Waals and hydrophobic interactions (Section 2.1). The surface of the bilayer is polar and contains charged groups. The nonpolar hydrocarbon interior of the bilayer consists of the saturated and unsaturated chains of fatty acids and the fused-ring system of cholesterol.

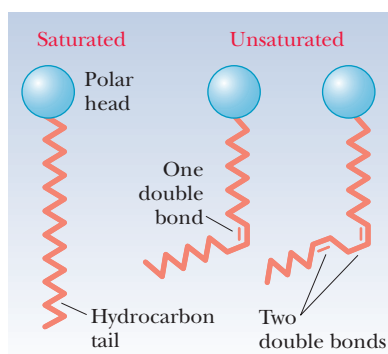
Both the inner and outer layers of the bilayer contain mixtures of lipids, but their compositions differ and can be used to distinguish the inner and outer layers from each other (Figure 8.11). Bulkier molecules tend to occur in the outer layer, and smaller molecules tend to occur in the inner layer.

### How does the composition of the bilayer affect its properties?

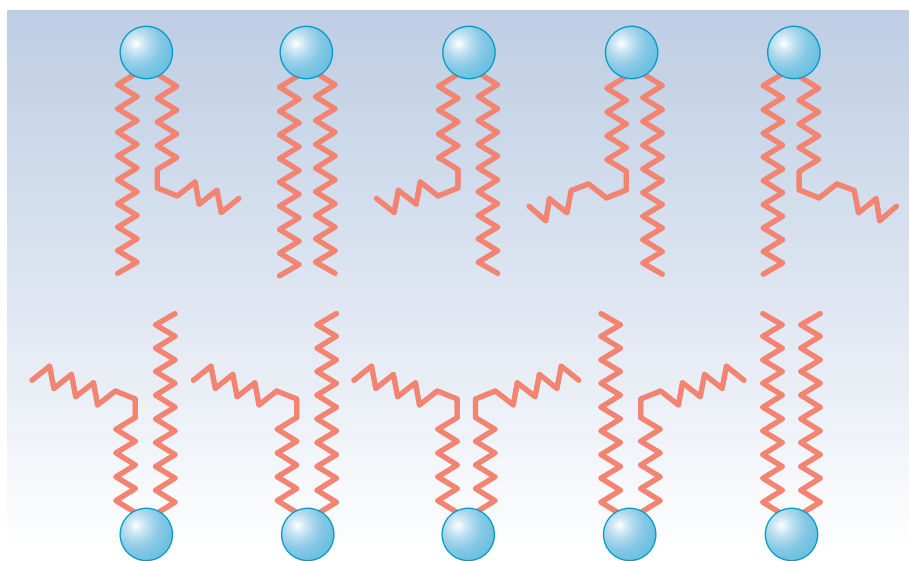
The arrangement of the hydrocarbon interior of the bilayer can be ordered and rigid or disordered and fluid. The bilayer’s fluidity depends on its composition. In saturated fatty acids, a linear arrangement of the hydrocarbon chains leads to close packing of the molecules in the bilayer, and thus to rigidity. Unsaturated fatty acids have a kink in the hydrocarbon chain that does not exist in saturated fatty acids (Figure 8.12). The kinks cause disorder in the packing of the chains, which makes for a more open structure than would be possible for straight saturated chains (Figure 8.13). In turn, the disordered structure caused by the presence of unsaturated fatty acids with *cis* double bonds (and therefore kinks) in their hydrocarbon chains causes greater fluidity in the bilayer. The lipid components of a bilayer are always in motion, to a greater extent in more fluid bilayers and to a lesser extent in more rigid ones.



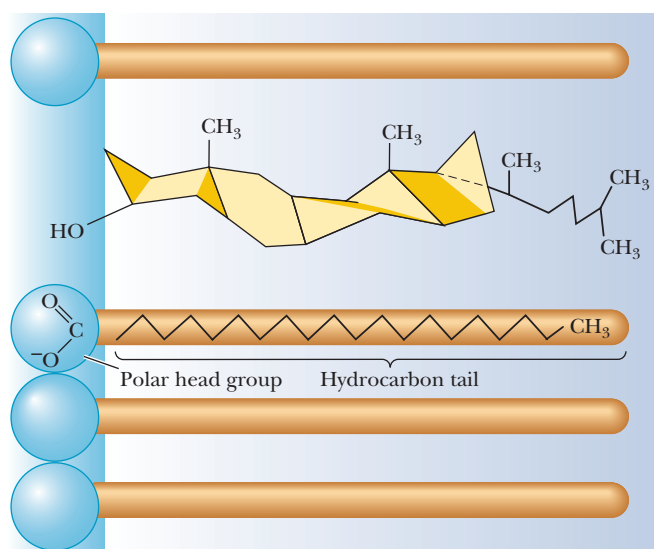
■ **FIGURE 8.11 Lipid bilayer asymmetry.** The compositions of the outer and inner layers differ; the concentration of bulky molecules is higher in the outer layer, which has more room.



■ **FIGURE 8.12 The effect of double bonds on the conformations of the hydrocarbon tails of fatty acids.** Unsaturated fatty acids have kinks in their tails.



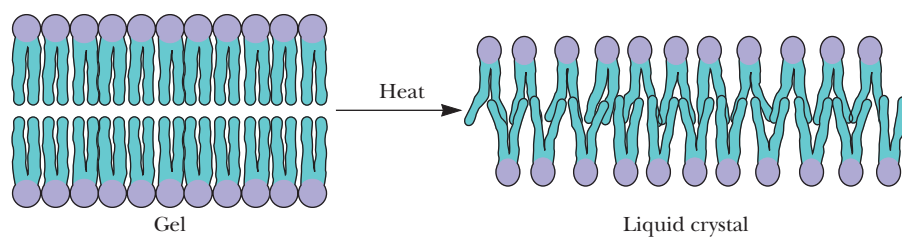
■ **FIGURE 8.13 Schematic drawing of a portion of a highly fluid phospholipid bilayer.** The kinks in the unsaturated side chains prevent close packing of the hydrocarbon portions of the phospholipids.



■ **FIGURE 8.14** Stiffening of the lipid bilayer by cholesterol. The presence of cholesterol in a membrane reduces fluidity by stabilizing extended chain conformations of the hydrocarbon tails of fatty acids, as a result of van der Waals interactions.

The presence of cholesterol may also enhance order and rigidity. The fused-ring structure of cholesterol is itself quite rigid, and the presence of cholesterol stabilizes the extended straight-chain arrangement of saturated fatty acids by van der Waals interactions (Figure 8.14). The lipid portion of a plant membrane has a higher percentage of unsaturated fatty acids, especially polyunsaturated (containing two or more double bonds) fatty acids, than does the lipid portion of an animal membrane. Furthermore, the presence of cholesterol is characteristic of animal, rather than plant, membranes. As a result, animal membranes are less fluid (more rigid) than plant membranes, and the membranes of prokaryotes, which contain no appreciable amounts of steroids, are the most fluid of all. Research suggests that plant sterols can act as natural cholesterol blockers, interfering with the uptake of dietary cholesterol.

With heat, ordered bilayers become less ordered; bilayers that are comparatively disordered become even more disordered. This cooperative transition takes place at a characteristic temperature, like the melting of a crystal, which is also a cooperative transition (Figure 8.15). The transition temperature is higher for more rigid and ordered membranes than it is for relatively fluid and disordered membranes. A powerful method called differential scanning calorimetry (DSC) makes it possible to obtain information about phase transitions in lipid bilayers. A DSC instrument has a sample cell for the bilayer and a reference cell containing a standard that will not undergo a phase transition. The two cells are maintained at a given temperature that is increased in controlled fashion by passing an electric current through the cells. When a phase transition takes place, a different amount of power is needed to maintain the temperature in the two cells. This amount of power can be measured and converted to a graph that gives information about both the transition temperature and the amount of energy needed to bring about the phase transition. The following Biochemical Connections box looks at some connections between



■ **FIGURE 8.15** An illustration of the gel-to-liquid crystalline phase transition, which occurs when a membrane is warmed through the transition temperature,  $T_m$ . Notice that the surface area must increase and the thickness must decrease as the membrane goes through a phase transition. The mobility of the lipid chains increases dramatically.

# Biochemical Connections NUTRITION

## Butter versus Margarine—Which Is Healthier?

We use the terms animal “fats” and plant “oils” because of the solid and fluid nature of these two groups of lipids. The major difference between fats and oils is the percentage of unsaturated fatty acids in the triglycerides and the phosphoglycerides of membranes. This difference is far more important than the fact that the length of the fatty acid chain can affect the melting points. Butter is an exception; it has a high proportion of short-chain fatty acids and thus can “melt in your mouth.” Membranes must maintain a certain degree of fluidity to be functional. Consequently, unsaturated fats are distributed in varying proportions in different parts of the body. The membranes of internal organs of warm-blooded mammals have a higher percentage of saturated fats than do the membranes of skin tissues, which helps keep the membrane more solid at the higher temperature of the internal organ. An extreme example of this is found in the legs and the body of reindeer, where marked differences exist in the percentages of saturated fatty acids.

When bacteria are grown at different temperatures, the fatty acid composition of the membranes changes to reflect more unsaturated fatty acids at lower temperatures and more saturated fatty acids at higher temperatures. The same type of difference can be seen in eukaryotic cells grown in tissue culture.

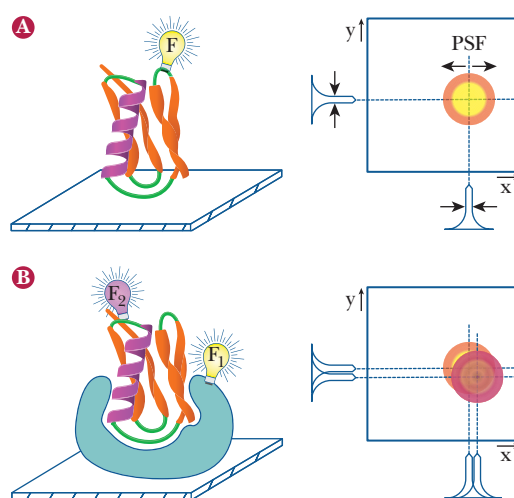
Even if we look at plant oils alone, we find different proportions of saturated fats in different oils. The following table gives the distribution for a tablespoon (14 g) of different oils.

Because cardiovascular disease is correlated with diets high in saturated fats, a diet of more unsaturated fats may reduce the risk of heart attacks and strokes. Canola oil is an attractive dietary choice because it has a high ratio of unsaturated fatty acids to saturated fatty acids. Since the 1960s, we have known that foods higher in polyunsaturated fats were healthier. Unfortunately, even though olive oil is popular in cooking Italian food and canola oil is trendy for other cooking, pouring oil on bread or toast is not appealing. Thus companies began to market butter substitutes that were based on unsaturated fatty acids but that would also have the physical characteristics of butter, such as being solid at room temperature. They accomplished this task by partially hydrogenating the double bonds in the unsaturated fatty acids making up the oils. The irony here is that, to avoid eating the saturated fatty acids in butter, butter substitutes were created from polyunsaturated oils by removing some of the double bonds, thus making them more saturated. In addition, many of the soft spreads that are marketed as being healthy (safflower oil spread and canola oil spread) may indeed pose new health risks. In the hydrogenation process, some double bonds are converted to the *trans* form. Studies now show that *trans* fatty acids raise the ratio of LDL (low-density lipoprotein) cholesterol compared to HDL (high-density lipoprotein) cholesterol, a positive correlator of heart disease. Thus the effects of *trans* fatty acids are similar to those of saturated fatty acids. In the last few years, however, new butter substitutes have been marketed that advertise “no *trans* fatty acids.”

Types of Oil or Fat	Example	Saturated (g)	Monounsaturated (g)	Polyunsaturated (g)
Tropical oils	Coconut oil	13	0.7	0.3
Semitropical oils	Peanut oil	2.4	6.5	4.5
	Olive oil		10.3	1.3
Temperate oils	Canola oil	1	8.2	4.1
	Safflower oil	1.3	1.7	10.4
Animal fat	Lard	5.1	5.9	1.5
	Butter	9.2	4.2	0.6

the fatty acid composition of bilayers and membranes and how they behave at different temperatures.

Recall that the distribution of lipids is not the same in the inner and outer portions of the bilayer. Because the bilayer is curved, the molecules of the inner layer are more tightly packed (refer to Figure 8.11). Bulkier molecules, such as cerebrosides (see Section 8.2), tend to be located in the outer layer. There is very little tendency for “flip-flop” migration of lipid molecules from one layer of the bilayer to another, but it does occur occasionally. Lateral motion of lipid molecules within one of the two layers frequently takes place, however, especially in more fluid bilayers. Several methods exist for monitoring the motions of molecules within a lipid bilayer. One of the most powerful ways uses fluorescence spectroscopy. This method makes use of the fact that some molecules absorb light of a given wavelength and then re-emit light of another, longer wavelength. Lipid molecules are not themselves fluorescent, but they can be “tagged” with groups that are. Fluorescence can be detected even at very low levels. This fact makes it possible to use the technique as the basis of fluorescence microscopy, which will detect the tagged moieties in bilayers. There are many variations in detection techniques, but in all cases they are based on the re-emitted fluorescent light. The use of fluorescence can be expanded in



■ **FIGURE 8.16** Single-molecule fluorescence spectroscopy can be used to determine (a) the position of a single fluorescent center or (b) the distance between two fluorescent centers. (See text below.) (From *Fluorescence Spectroscopy of Single Biomolecules*. Shimon Weiss (12 March 1999) *Science* 283 (5408), p. 1676. Reprinted with permission from AAAS.)

studies on actual membranes rather than bilayers. Membranes contain proteins in addition to the lipid bilayer. The side chains of tryptophan and tyrosine have intrinsic fluorescence, and this property can be used to obtain information about the protein portion of the membrane.

Fluorescence spectroscopy is so sensitive that it can be used to detect information about single molecules. As shown in Figure 8.16a, a single macromolecule can be labeled with a fluorescent moiety (the fluorophore, marked F in the figure). The fluorescent signal is monitored in two dimensions on the oriented sample. It becomes possible to localize the fluorophore in the molecule. “PSF” stands for point-spread function, the error of estimate. In Figure 8.16b, two independent fluorophores are present. The distance between them can be determined by subtracting the distance between the centers of their PSFs.

## 8.4 The Kinds of Membrane Proteins

### *How are proteins associated with the bilayer in membranes?*

Proteins in a biological membrane can be associated with the lipid bilayer in either of two ways—as **peripheral proteins** on the surface of the membrane or as **integral proteins** within the lipid bilayer (Figure 8.17). Notice that the integral protein rhodopsin (shown in purple) consists mostly of helical portions that span the membrane. The peripheral G protein is a trimer. The three different subunits are shown in red, yellow, and blue. Peripheral proteins are usually bound to the charged head groups of the lipid bilayer by polar interactions, electrostatic interactions, or both. They can be removed by such mild treatment as raising the ionic strength of the medium. The relatively numerous charged particles present in a medium of higher ionic strength undergo more electrostatic interactions with the lipid and with the protein, “swamping out” the comparatively fewer electrostatic interactions between the protein and the lipid.

Removing integral proteins from membranes is much more difficult. Harsh conditions, such as treatment with detergents or extensive sonication (exposure to ultrasonic vibrations), are usually required. Such measures frequently denature the protein, which often remains bound to lipids in spite of all efforts to obtain it in pure form. The denatured protein is of course inactive, whether or not it remains bound to lipids. Fortunately, nuclear magnetic resonance techniques enable researchers to study proteins of this





component. **Transport proteins** help move substances in and out of the cell, and **receptor proteins** are important in the transfer of extracellular signals, such as those carried by hormones or neurotransmitters, into the cell. In addition, some enzymes are tightly bound to membranes; examples include many of the enzymes responsible for aerobic oxidation reactions, which are found in specific parts of mitochondrial membranes. Some of these enzymes are on the inner surface of the membrane, and some are on the outer surface. There is an uneven distribution of proteins of all types on the inner and outer layers of all cell membranes, just as there is an asymmetric distribution of lipids.

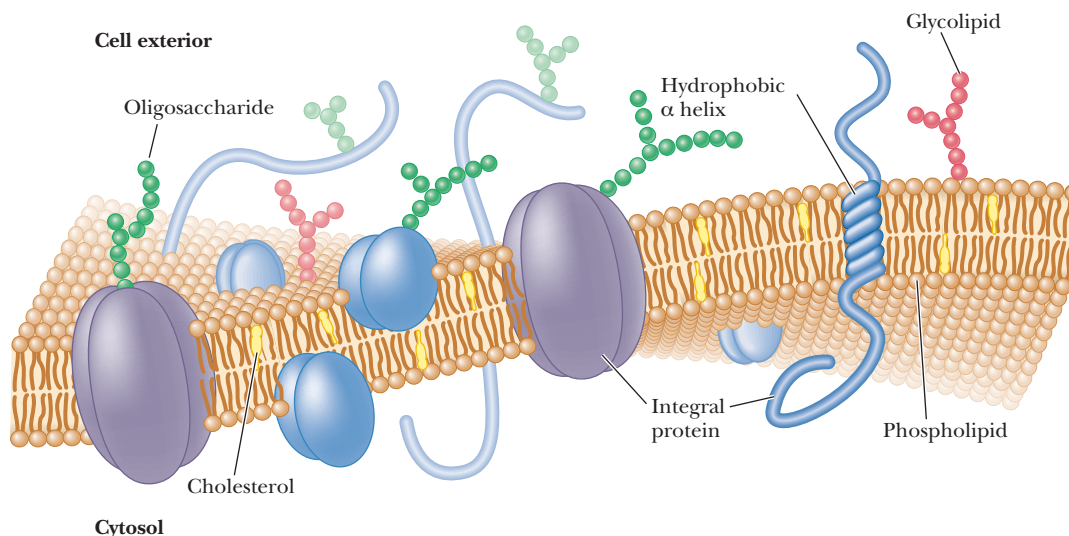
## 8.5 The Fluid-Mosaic Model of Membrane Structure

### *How do proteins and the lipid bilayer interact with each other in membranes?*

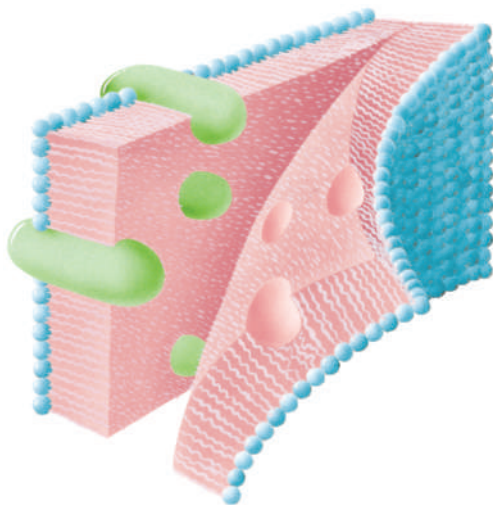
We have seen that biological membranes have both lipid and protein components. How do these two parts combine to produce a biological membrane? Currently, the **fluid-mosaic model** is the most widely accepted description of biological membranes. The term *mosaic* implies that the two components exist side by side without forming some other substance of intermediate nature. The basic structure of biological membranes is that of the lipid bilayer, with the proteins embedded in the bilayer structure (Figure 8.19). As time goes on, it is becoming apparent that preferential association can take place among sphingolipids, sterols, and membrane proteins. Lipids are sorted into assemblages known as rafts, which become the fundamental building blocks on which membrane specificity is based.

Membrane proteins tend to have a specific orientation in the membrane. The term *fluid mosaic* implies that the same sort of lateral motion that we have already seen in lipid bilayers also occurs in membranes. The proteins “float” in the lipid bilayer and can move along the plane of the membrane.

Electron micrographs can be made of membranes that have been frozen and then fractured along the interface between the two layers. The outer layer is removed, exposing the interior of the membrane. The interior has a



■ **FIGURE 8.19** Fluid-mosaic model of membrane structure. Membrane proteins can be seen embedded in the lipid bilayer. (From Singer, S. J., in G. Weissman and R. Claiborne, Eds., *Cell Membranes: Biochemistry, Cell Biology, and Pathology*, New York: HP Pub., 1975, p. 37.)



■ **FIGURE 8.20** Replica of a freeze-fractured membrane. In the freeze-fracture technique, the lipid bilayer is split parallel to the surface of the membrane. The hydrocarbon tails of the two layers are separated from each other, and the proteins can be seen as “hills” in the replica shown. In the other layer, seen edge on, there are “valleys” where the proteins were. (From Singer, S. J., in *G. Weissman and R. Claiborne, Eds., Cell Membranes: Biochemistry, Cell Biology, and Pathology*, New York: HP Pub., 1975, p. 37.)



■ **FIGURE 8.21** Electron micrograph of a freeze-fractured thylakoid membrane of a pea (magnified 110,000X). The grains protruding from the surface are integral membrane proteins.

granular appearance because of the presence of the integral membrane proteins (Figures 8.20 and 8.21). In addition to electron microscopy, atomic force microscopy can provide useful and informative images of membranes. The two methods differ in the physical principle on which the imaging process is based. The usual electron microscopy depends on scattering of a beam of electrons from the surface of the sample. In atomic force microscopy, the sample surface is scanned using a cantilever with a sharp tip. Electrical measurements determine the force generated between the tip and the sample surface, which generates the image.

## Biochemical Connections BIOTECHNOLOGY

### Membranes in Drug Delivery

Because the driving force behind the formation of lipid bilayers is the exclusion of water from the hydrophobic region of lipids, and not some enzymatic process, artificial membranes can be created in the lab. **Liposomes** are stable structures based on a lipid bilayer that form a spherical vesicle. These vesicles can be prepared with therapeutic agents on the inside and then used to deliver the agent to a target tissue.

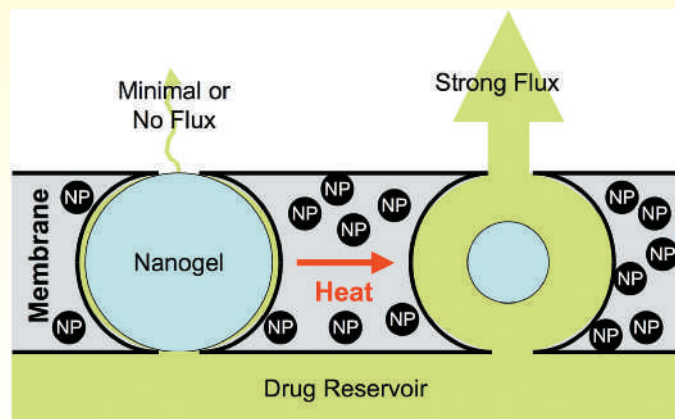
To carry the process even further, vesicles can be prepared with embedded artificial substances that can be used to control the process of release of the therapeutic agent. Some of the embedded substances can be used to create an “on-off” switch for drug delivery. A recent report from researchers at Harvard Medical School describes one such delivery process.

Bilayers were prepared that were embedded with nanogels of a synthetic polymer. This polymer, poly(*N*-isopropylacrylamide) (PNIPAM), forms a hydrogel that is swollen in its native state but collapses on heating. The size of the nanogels in their native state exactly matched the width of the membrane. In addition, nanoparticles of magnetite (iron oxide) were embedded in the membrane matrix.

When a magnetic field is applied, the magnetite particles heat up, leading to a rise in temperature of a few degrees in the PNIPAM. The hydrogel contracts, but the surrounding membrane does not. The result is that channels are formed that allow passage of a drug from one side of the membrane to the other.

When the magnetic field is removed, the particles cool down, the gel expands, and the channels are resealed.

One of the first suggested uses for this technology is for patient-controlled administration of painkillers, but many more applications can be expected as time goes on.



■ The application of a magnetic field leads to heating, which in turn creates a channel for the drug to migrate from the reservoir through the membrane. NP indicates the iron oxide nanoparticles.

© Dr. Andrew Staehlin, University of Colorado

Reprinted with permission from *Nano Letters* (September 2009).  
Hoare, T., Santamaría, J., Goya, G. F., Justa, S., Lin, D., Lau, S., Padara, R.,  
Langer, R., Kihane, D. S. A magnetically triggered composite membrane  
for on-demand drug delivery. Copyright © 2009 American Chemical  
Society.



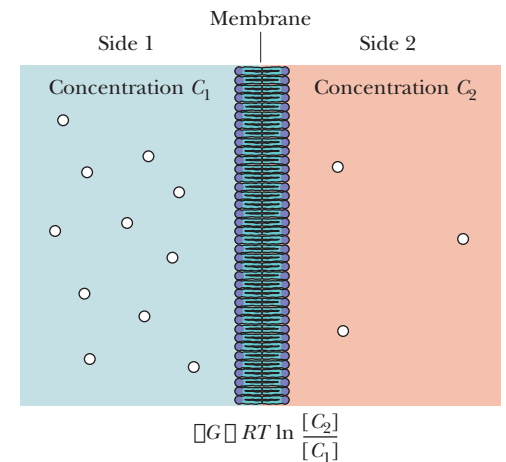
## 8.6 The Functions of Membranes

As already mentioned, three important functions take place in or on membranes (in addition to the structural role of membranes as the boundaries and containers of all cells and of the organelles within eukaryotic cells). The first of these functions is *transport*. Membranes are semipermeable barriers to the flow of substances into and out of cells and organelles. Transport through the membrane can involve the lipid bilayer as well as the membrane proteins. The other two important functions primarily involve the membrane proteins. One of these functions is *catalysis*. As we have seen, enzymes can be bound—in some cases very tightly—to membranes, and the enzymatic reaction takes place on the membrane. The third significant function is the *receptor property*, in which proteins bind specific biologically important substances that trigger biochemical responses in the cell. We shall discuss enzymes bound to membranes in subsequent chapters (especially in our treatment of aerobic oxidation reactions in Chapters 19 and 20). The other two functions we now consider in turn.

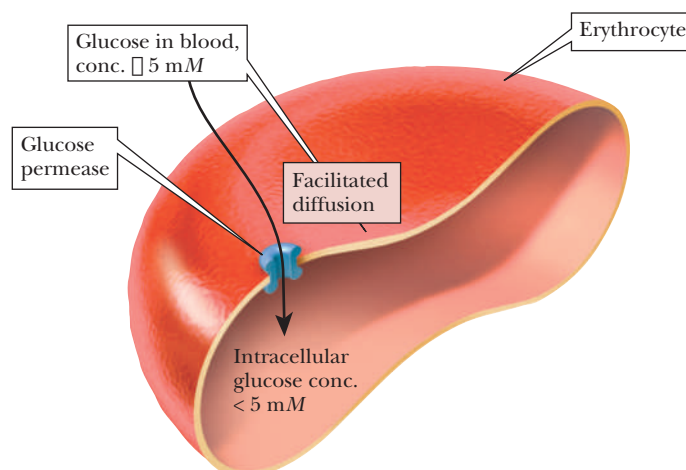
### How does transport through membranes take place?

The most important question about transport of substances across biological membranes is whether the process requires the cell to expend energy. In **passive transport**, a substance moves from a region of higher concentration to one of lower concentration. In other words, the movement of the substance is in the same direction as a *concentration gradient*, and the cell does not expend energy. In **active transport**, a substance moves from a region of lower concentration to one of higher concentration (against a concentration gradient), and this process requires the cell to expend energy.

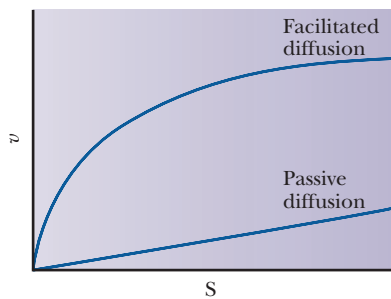
The process of passive transport can be subdivided into two categories—simple diffusion and facilitated diffusion. In **simple diffusion**, a molecule moves directly through the membrane without interacting with another molecule. Small, uncharged molecules, such as  $O_2$ ,  $N_2$ , and  $CO_2$ , can pass through membranes via simple diffusion. The rate of movement through the membrane is controlled solely by the concentration difference across the membrane (Figure 8.22). Larger molecules (especially polar ones) and ions cannot pass through a membrane by simple diffusion. The process of moving a molecule passively through a membrane using a carrier protein, to which molecules bind, is called **facilitated diffusion**. A good example is the movement of glucose into erythrocytes. The concentration of glucose in the blood is about 5 mM. The glucose concentration in the erythrocyte is less than 5 mM. Glucose passes through a carrier protein called glucose permease (Figure 8.23).



**FIGURE 8.22** Passive diffusion. Passive diffusion of an uncharged species across a membrane depends only on the concentrations ( $C_1$  and  $C_2$ ) on the two sides of the membrane.



**FIGURE 8.23** Facilitated diffusion. Glucose passes into an erythrocyte via glucose permease by facilitated diffusion. Glucose flows using its concentration gradient via passive transport. (Adapted from Lehninger, Principles of Biochemistry, Third Edition, by David L. Nelson and Michael M. Cox. © 1982, 1992, 2000 by Worth Publishers. Used with permission of W. H. Freeman and Company.)



■ **FIGURE 8.24** Passive diffusion and facilitated diffusion may be distinguished graphically. The plots for facilitated diffusion are similar to plots of enzyme-catalyzed reactions (Chapter 6), and they display saturation behavior. The value  $v$  stands for velocity of transport.  $S$  is the concentration of the substrate being transported.

This process is labeled as facilitated diffusion because no energy is expended and a protein carrier is used. In addition, facilitated diffusion is identified by the fact that the rate of transport, when plotted against the concentration of the molecule being transported, gives a hyperbolic curve similar to that seen in Michaelis–Menten enzyme kinetics (Figure 8.24). In a carrier protein, a pore is created by folding the backbone and side chains. Many of these proteins have several  $\alpha$ -helical portions that span the membrane; in others, a  $\beta$ -barrel forms the pore. In one example, the helical portion of the protein spans the membrane. The exterior, which is in contact with the lipid bilayer, is hydrophobic, whereas the interior, through which ions pass, is hydrophilic. Note that this orientation is the inverse of that observed in water-soluble globular proteins.

**Active transport** requires moving substances against a concentration gradient. It is identified by the presence of a carrier protein and the need for an energy source to move solutes against a gradient. In **primary active transport**, the movement of molecules against a gradient is directly linked to the hydrolysis of a high-energy molecule, such as ATP. The situation is so markedly similar to pumping water uphill that one of the most extensively studied examples of active transport, moving potassium ions into a cell and simultaneously moving sodium ions out of the cell, is referred to as the **sodium–potassium ion pump** (or  $\text{Na}^+/\text{K}^+$  pump).

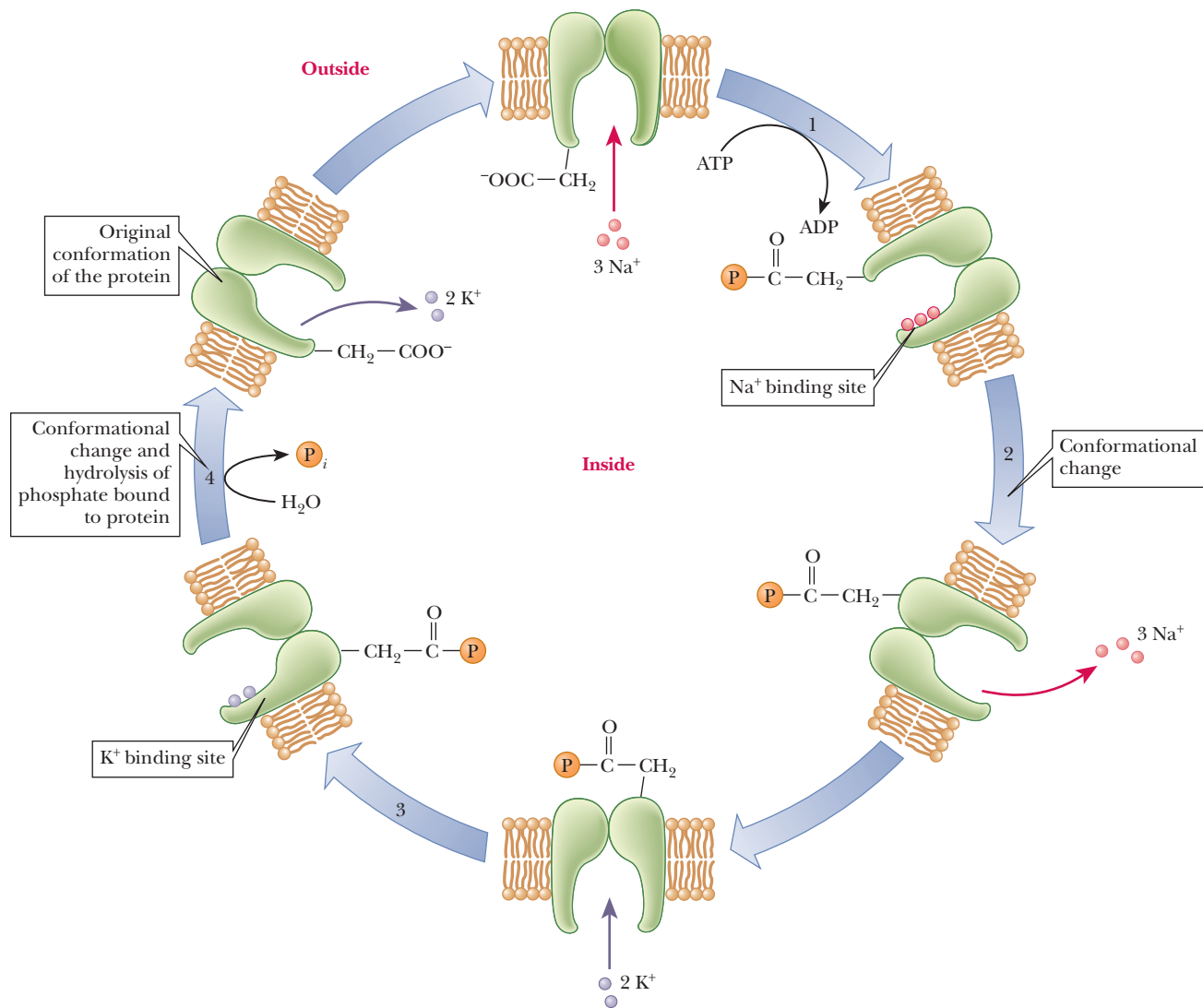
Under normal circumstances, the concentration of  $\text{K}^+$  is higher inside a cell than in extracellular fluids ( $[\text{K}^+]_{\text{inside}} > [\text{K}^+]_{\text{outside}}$ ), but the concentration of  $\text{Na}^+$  is lower inside the cell than out ( $[\text{Na}^+]_{\text{inside}} < [\text{Na}^+]_{\text{outside}}$ ). The energy required to move these ions against their gradients comes from an exergonic (energy-releasing) reaction, the hydrolysis of ATP to ADP and  $\text{P}_i$  (phosphate ion). There can be no transport of ions without hydrolysis of ATP. The same protein appears to serve both as the enzyme that hydrolyzes the ATP (the ATPase) and as the transport protein; it consists of several subunits. The reactants and products of this hydrolysis reaction—ATP, ADP, and  $\text{P}_i$ —remain within the cell, and the phosphate becomes covalently bonded to the transport protein for part of the process.

The  $\text{Na}^+/\text{K}^+$  pump operates in several steps (Figure 8.25). One subunit of the protein hydrolyzes the ATP and transfers the phosphate group to an aspartate side chain on another subunit (step 1). (The bond formed here is a mixed anhydride; see Section 1.2.) Simultaneously, binding of three  $\text{Na}^+$  ions from the interior of the cell takes place. The phosphorylation of one subunit causes a conformational change in the protein, which opens a channel or pore through which the three  $\text{Na}^+$  ions can be released to the extracellular fluid (step 2). Outside the cell, two  $\text{K}^+$  ions bind to the pump enzyme, which is still phosphorylated (step 3). Another conformational change occurs when the bond between the enzyme and the phosphate group is hydrolyzed. This second conformational change regenerates the original form of the enzyme and allows the two  $\text{K}^+$  ions to enter the cell (step 4). The pumping process transports three  $\text{Na}^+$  ions out of the cell for every two  $\text{K}^+$  ions transported into the cell (Figure 8.26).

The operation of the pump can be reversed when there is no  $\text{K}^+$  and a high concentration of  $\text{Na}^+$  in the extracellular medium; in this case, ATP is produced by the phosphorylation of ADP. The actual operation of the  $\text{Na}^+/\text{K}^+$  pump is not completely understood and probably is even more complicated than we now know. There is also a calcium ion ( $\text{Ca}^{2+}$ ) pump, which is a subject of equally active investigation. Unanswered questions about the detailed mechanism of active transport provide opportunities for future research.

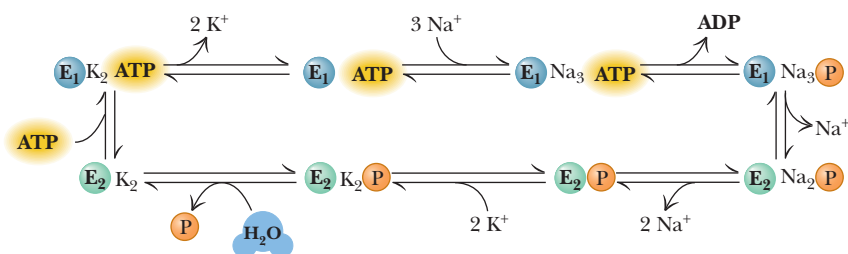
Another type of transport is called **secondary active transport**. An example is the galactoside permease in bacteria (Figure 8.27). The lactose concentration inside the bacterial cell is higher than the concentration outside, so moving lactose into the cell requires energy. The galactoside permease



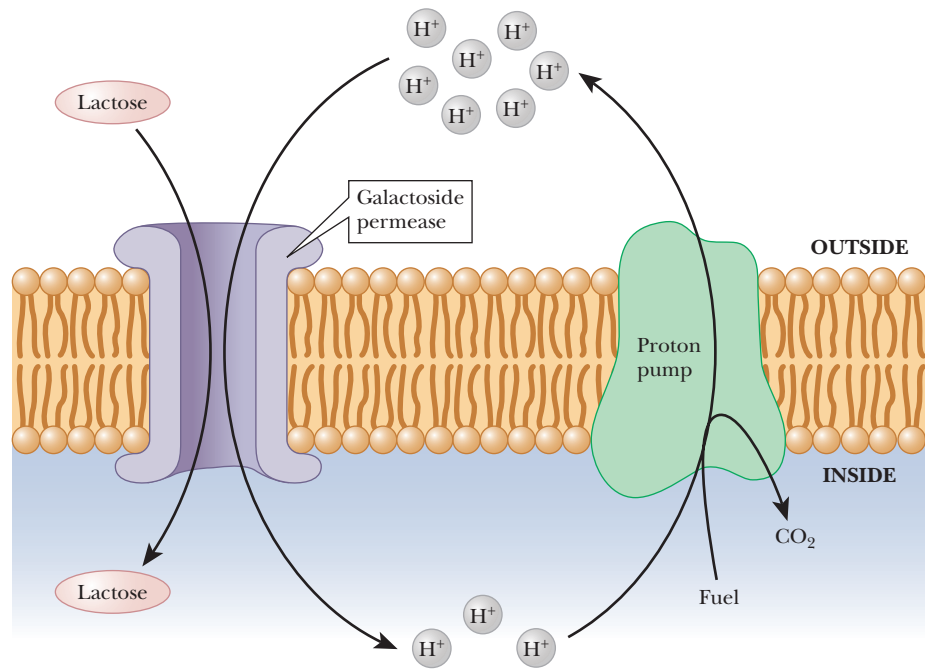


■ **FIGURE 8.25** The sodium–potassium ion pump (see text for details).

does not directly hydrolyze ATP, however. Instead, it harnesses the energy by letting hydrogen ions flow through the permease into the cell with their concentration gradient. As long as more energy is available allowing the hydrogen ions to flow ( $-\Delta G$ ) than is required to concentrate the lactose ( $+\Delta G$ ), the process is possible. However, to arrive at a situation in which there is a higher concentration of hydrogen ions on the outside than on the inside, some other primary active transporter must establish the hydrogen ion gradient. Active transporters that create hydrogen ion gradients are called **proton pumps**.



■ **FIGURE 8.26** A mechanism for  $\text{Na}^+/\text{K}^+$  ATPase (the sodium–potassium ion pump). The model assumes two principal conformations,  $\text{E}_1$  and  $\text{E}_2$ . Binding of  $\text{Na}^+$  ions to  $\text{E}_1$  is followed by phosphorylation and release of ADP.  $\text{Na}^+$  ions are transported and released, and  $\text{K}^+$  ions are bound before dephosphorylation of the enzyme. Transport and release of  $\text{K}^+$  ions complete the cycle.

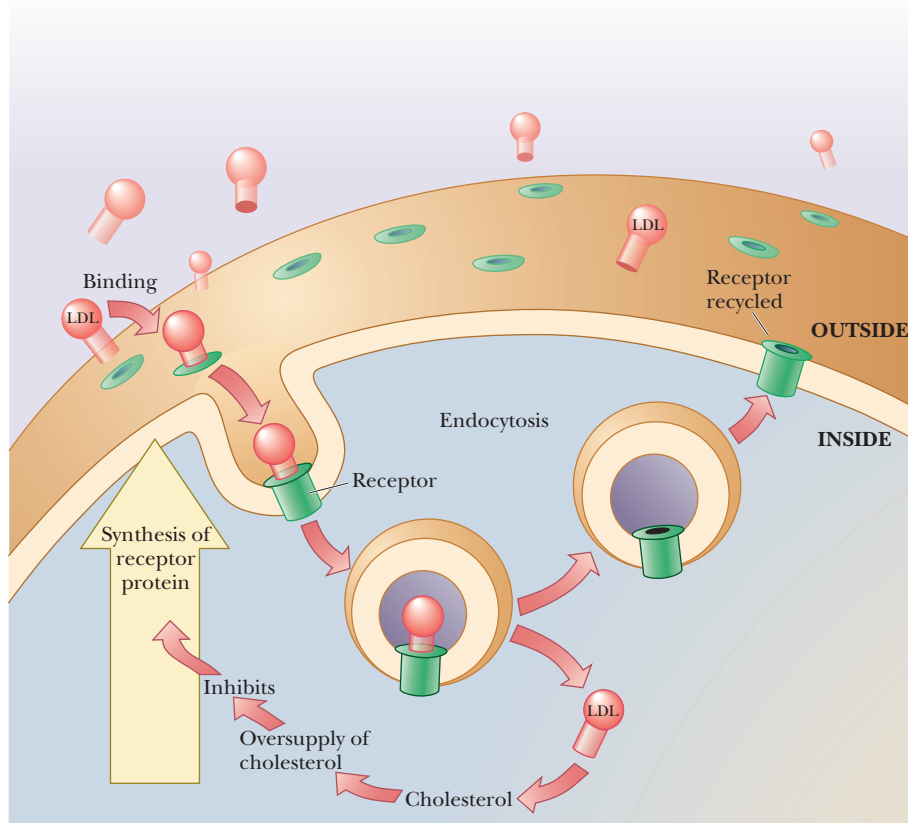


■ **FIGURE 8.27** An example of secondary active transport. Galactoside permease uses the higher concentration of  $H^+$  outside the cell to drive the concentration of lactose inside the cell. (Adapted from Lehninger, Principles of Biochemistry, Third Edition, by David L. Nelson and Michael M. Cox. © 1982, 1992, 2000 by Worth Publishers. Used with permission of W. H. Freeman and Company.)

### How do membrane receptors work?

The first step in producing the effects of some biologically active substances is binding the substance to a protein receptor site on the exterior of the cell. The interaction between receptor proteins and the active substances that bind to them has features in common with enzyme–substrate recognition. There is a requirement for essential functional groups that have the correct three-dimensional conformation with respect to each other. The binding site, whether on a receptor or an enzyme, must provide a good fit for the substrate. In receptor binding, as in enzyme behavior, inhibition of the action of the protein by some sort of “poison” or inhibitor is possible. The study of receptor proteins is less advanced than the study of enzymes because many receptors are tightly bound integral proteins, and their activity depends on the membrane environment. Receptors are often large oligomeric proteins (ones with several subunits), with molecular weights on the order of hundreds of thousands. Also, quite frequently, the receptor has very few molecules in each cell, adding to the difficulties of isolating and studying this type of protein.

An important type of receptor is that for low-density lipoprotein (LDL), the principal carrier of cholesterol in the bloodstream. LDL is a particle that consists of various lipids—in particular, cholesterol and phosphoglycerides—as well as a protein. The protein portion of the LDL particle binds to the LDL receptor of a cell. The complex formed between the LDL and the receptor is pinched off into the cell in a process called *endocytosis*. The receptor protein is then recycled back to the surface of the cell (Figure 8.28). The cholesterol portion of the LDL is used in the cell, but an oversupply of cholesterol causes problems. Excess of cholesterol inhibits the synthesis of LDL receptor. If there are too few receptors for LDL, the level of cholesterol in the bloodstream increases. Eventually, the excess cholesterol is deposited in the arteries, blocking them severely. This blocking of arteries, called *atherosclerosis*, can eventually lead to heart attacks and strokes. In many industrialized countries, typical blood cholesterol levels are high, and the incidence of heart attacks and strokes is correspondingly high. (We will say more about this subject after we have seen the pathway by which cholesterol is synthesized in the body in Section 21.8.)



■ **FIGURE 8.28** The mode of action of the LDL receptor. A portion of the membrane with LDL receptor and bound LDL is taken into the cell as a vesicle. The receptor protein releases LDL and is returned to the cell surface when the vesicle fuses to the membrane. LDL releases cholesterol in the cell. An oversupply of cholesterol inhibits synthesis of the LDL receptor protein. An insufficient number of receptors leads to elevated levels of LDL and cholesterol in the bloodstream. This situation increases the risk of heart attack.

## Biochemical Connections PHYSIOLOGY

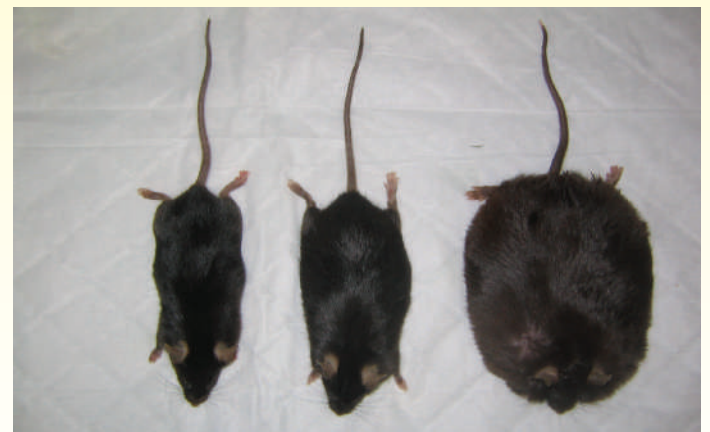
### Lipid Droplets Are Not Just Great Balls of Fat

The chapter-opening photo shows an electron micrograph of a fat cell. Very visible are the large lipid droplets. For decades these structures have been thought of as great balls of fat, a convenient way to store triacylglycerols for consumption. However, these droplets are surrounded by a thin phospholipid membrane that contain many membrane proteins with widely varying activities. On the negative side, they may also be involved in several lipid diseases, cardiovascular diseases, and diabetes. These lipid droplets are now being thought of as a subcellular organelle in their own right.

One of the first clues that lipid droplets were more than a simple pool of fat came in the early 1990s from research by Constantine Londos. He and his colleagues identified a protein called **perilipin** on the membrane of the lipid droplets in fat cells. They discovered that when cells are stimulated to metabolize the fatty acids in the lipid droplets, this protein is phosphorylated. This suggests a more complicated mechanism for controlling lipid digestion in fat cells than previously imagined. More than half a dozen proteins have been identified on the membrane of lipid droplets.

It is now believed that perilipin guards the fat resources of the lipid droplet. When not phosphorylated, the protein does not allow fat-digesting enzymes access to the triacylglycerols. When phosphorylated, the protein shifts conformation and allows access. Studies with mutant mice lacking perilipin have shown that such mice eat much more than their wild-type counterparts, yet burn off two-thirds of the extra calories consumed. The figure shows three different strains of mice. The mouse on the left is normal. The mouse on the right is an obese mouse from a strain that lacks the ability to respond to an appetite-suppressing hormone called **leptin**. The mouse in the middle is a double

mutant that cannot respond to leptin but also lacks perilipin. In this case the extra fat burning that takes place almost makes up for the increased appetite. Further research on such proteins in the membranes of lipid droplets could lead to useful antiobesity therapies.



■ Mouse on the right: an obese mouse that does not produce leptin because of a genetic mutation in the leptin gene. Mouse in the middle: a mouse that does not produce leptin but maintains relatively normal body weight because of its enhanced fat metabolism caused by the lack of perilipin. Mouse on the left: a wild-type, normal mouse that produces leptin as well as perilipin.

Courtesy of Drs. Pradip Saha & Lawrence Chan/Baylor College of Medicine

## 8.7 Lipid-Soluble Vitamins and Their Functions

### What is the role of lipid-soluble vitamins in the body?

Some vitamins, having a variety of functions, are of interest in this chapter because they are soluble in lipids. These lipid-soluble vitamins are hydrophobic, which accounts for their solubility (Table 8.3).

#### Vitamin A

The extensively unsaturated hydrocarbon  **$\beta$ -carotene** is the precursor of **vitamin A**, which is also known as **retinol**. As the name suggests,  $\beta$ -carotene is abundant in carrots, but it also occurs in other vegetables, particularly the yellow ones. When an organism requires vitamin A,  $\beta$ -carotene is converted to the vitamin (Figure 8.29).

A derivative of vitamin A plays a crucial role in vision when it is bound to a protein called *opsin*. The cone cells in the retina of the eye contain several types of opsin and are responsible for vision in bright light and for color vision. The rod cells in the retina contain only one type of opsin; they are responsible for vision in dim light. The chemistry of vision has been more extensively studied in rod cells than in cone cells, and we shall discuss events that take place in rod cells.

Vitamin A has an alcohol group that is enzymatically oxidized to an aldehyde group, forming **retinal** (Figure 8.29b). Two isomeric forms of retinal, involving *cis-trans* isomerization around one of the double bonds, are important in the behavior of this compound in vivo. The aldehyde group of retinal forms an imine (also called a Schiff base) with the side-chain amino group of a lysine residue in rod-cell opsin (Figure 8.30).

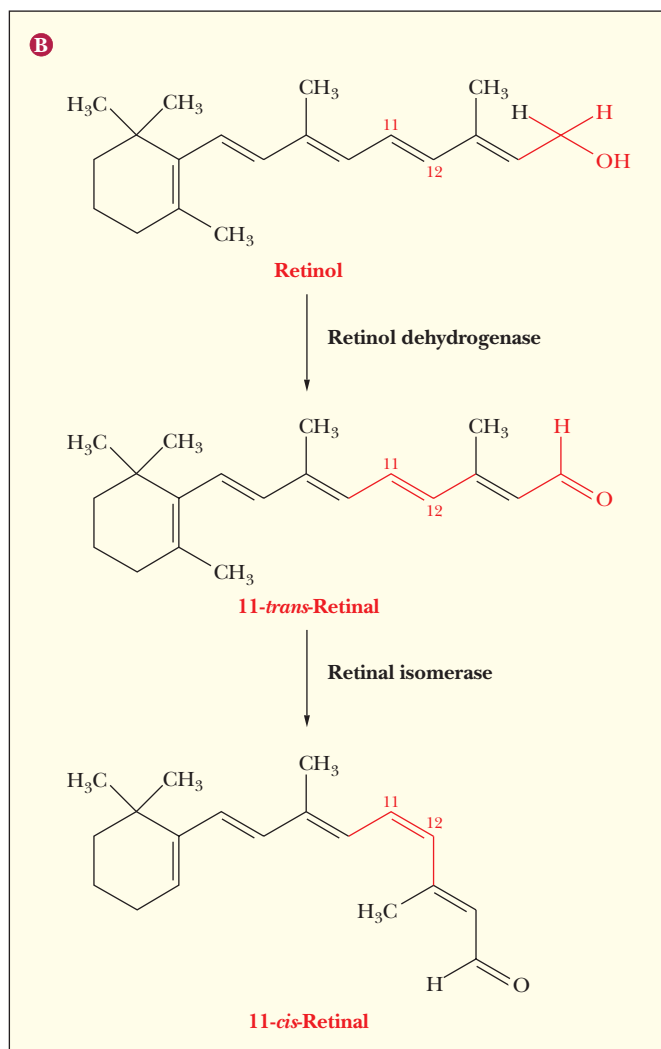
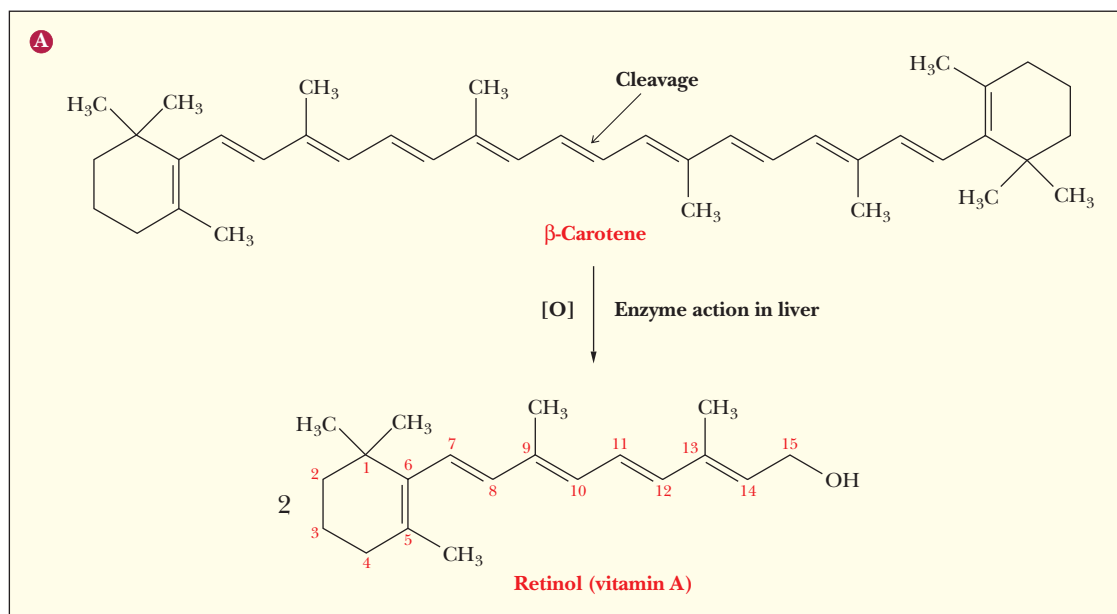
The product of the reaction between retinal and opsin is **rhodopsin**. The outer segment of rod cells contains flat membrane-enclosed discs, the membrane consisting of about 60% rhodopsin and 40% lipid. (For more details about rhodopsin, see the following Biochemical Connections box.)

#### Vitamin D

The several forms of **vitamin D** play a major role in the regulation of calcium and phosphorus metabolism. One of the most important of these compounds, vitamin D<sub>3</sub> (cholecalciferol), is formed from cholesterol by the action of ultraviolet radiation from the Sun. Vitamin D<sub>3</sub> is further processed in the body to form hydroxylated derivatives, which are the metabolically active form of this vitamin (Figure 8.31). The presence of vitamin D<sub>3</sub> leads to increased synthesis

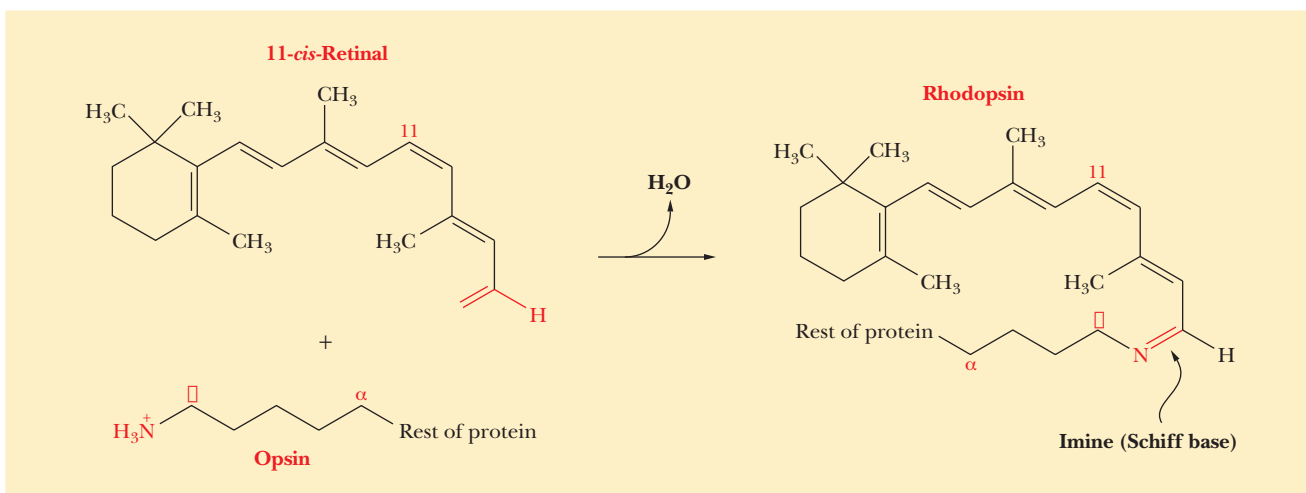
**TABLE 8.3**

Lipid-Soluble Vitamins and Their Functions	
Vitamin	Function
Vitamin A	Serves as the site of the primary photochemical reaction in vision
Vitamin D	Regulates calcium (and phosphorus) metabolism
Vitamin E	Serves as an antioxidant; necessary for reproduction in rats and may be necessary for reproduction in humans
Vitamin K	Has a regulatory function in blood clotting

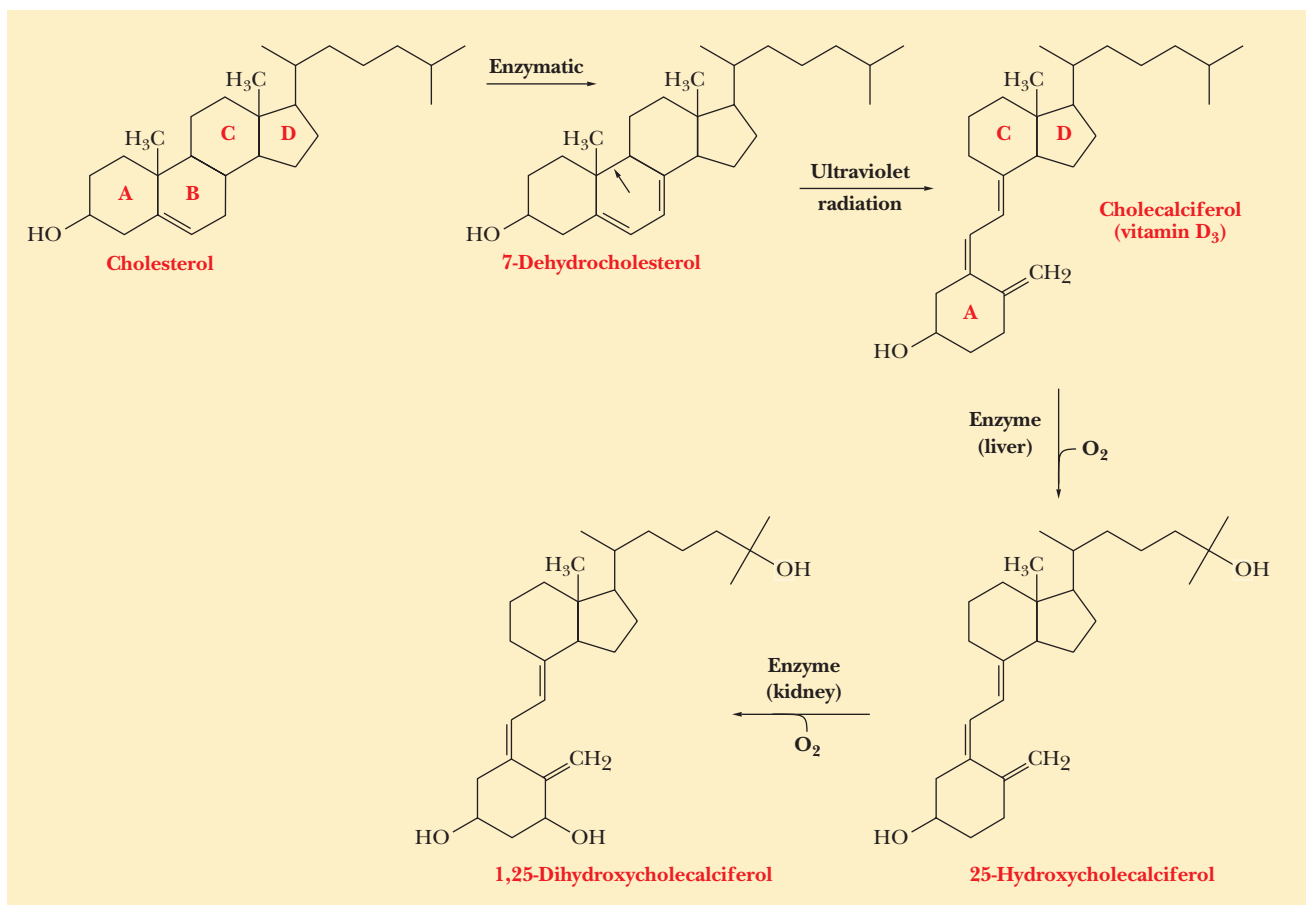


**FIGURE 8.29 Reactions of vitamin A.** (a) The conversion of  $\beta$ -carotene to vitamin A. (b) The conversion of vitamin A to 11-*cis*-retinal.





■ **FIGURE 8.30** The formation of rhodopsin from 11-*cis*-retinal and opsin.



■ **FIGURE 8.31** Reactions of vitamin D. The photochemical cleavage occurs at the bond shown by the arrow; electron rearrangements after the cleavage produce vitamin D<sub>3</sub>. The final product, 1,25-dihydroxycholecalciferol, is the form of the vitamin that is most active in stimulating the intestinal absorption of calcium and phosphate and in mobilizing calcium for bone development.

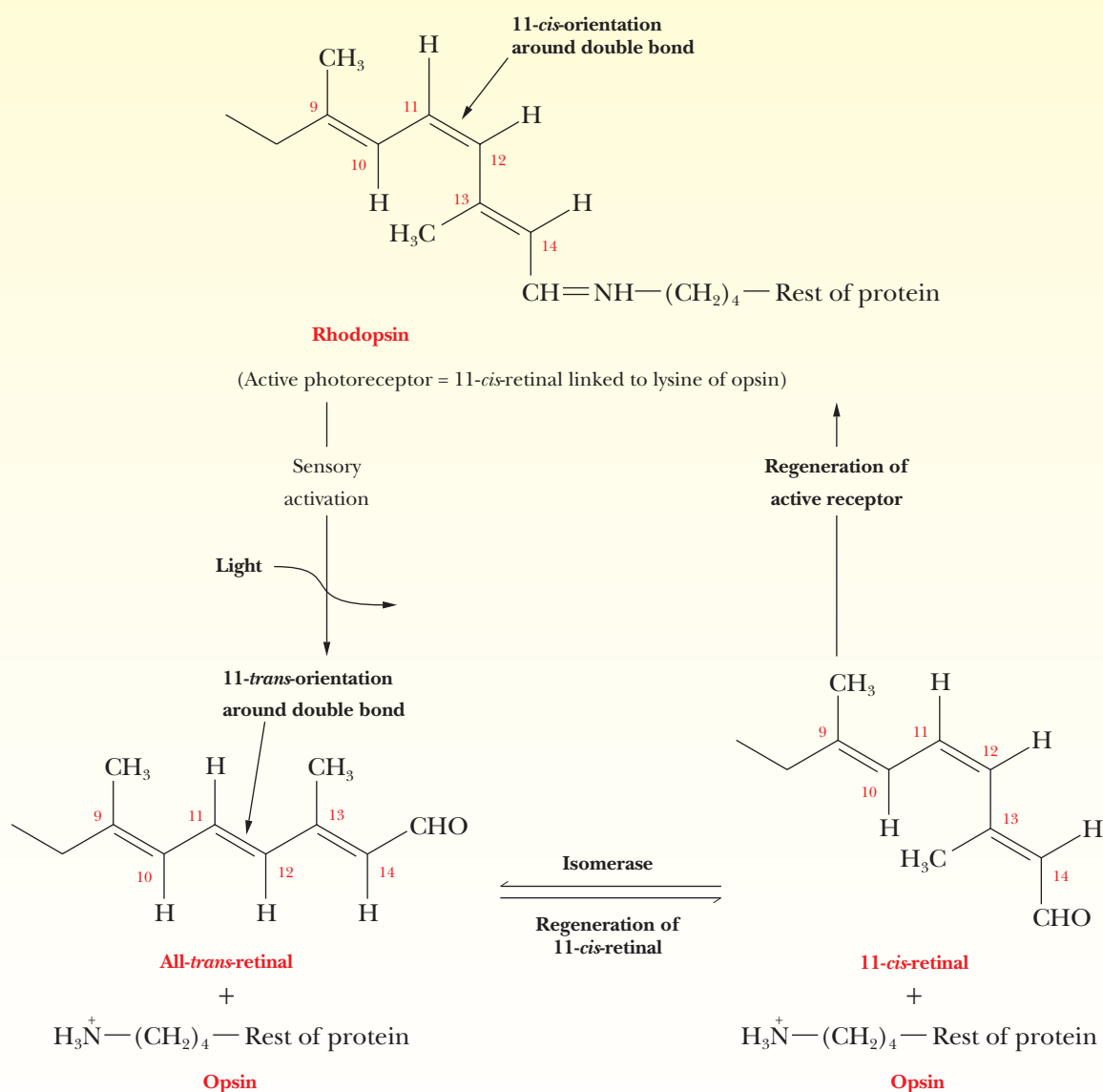
# Biochemical Connections NEUROSCIENCE

## Vision Has Great Chemistry

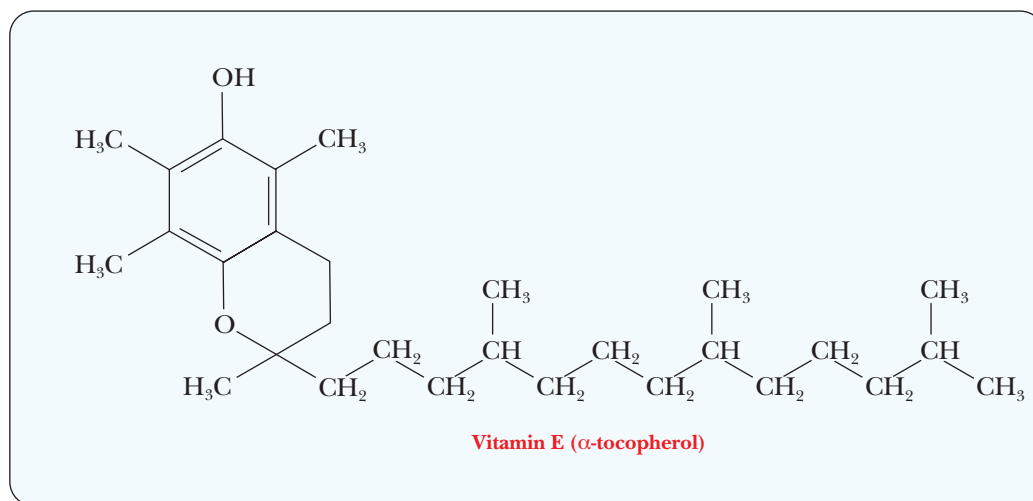
The primary chemical reaction in vision, the one responsible for generating an impulse in the optic nerve, involves *cis-trans* isomerization around one of the double bonds in the retinal portion of rhodopsin. When rhodopsin is active (that is, when it can respond to visible light), the double bond between carbon atoms 11 and 12 of the retinal (11-*cis*-retinal) has the *cis* orientation. Under the influence of light, an isomerization reaction occurs at this double bond, producing all-*trans*-retinal. Because the all-*trans* form of retinal cannot bind to opsin, all-*trans*-retinal and free opsin are released. As a result of this reaction, an electrical impulse is generated in the optic nerve and transmitted to the

brain to be processed as a visual event. The active form of rhodopsin is regenerated by enzymatic isomerization of the all-*trans*-retinal back to the 11-*cis* form and subsequent re-formation of the rhodopsin.

Vitamin A deficiency can have drastic consequences, as would be predicted from its importance in vision. Night blindness—and even total blindness—can result, especially in children. On the other hand, an excess of vitamin A can have harmful effects, such as bone fragility. Lipid-soluble compounds are not excreted as readily as water-soluble substances, and excessive amounts of lipid-soluble vitamins can accumulate in adipose tissue.



■ The primary chemical reaction of vision.



■ **FIGURE 8.32** The most active form of vitamin E is  $\alpha$ -tocopherol.

of a  $\text{Ca}^{2+}$ -binding protein, which increases the absorption of dietary calcium in the intestines. This process results in calcium uptake by the bones.

A deficiency of vitamin D can lead to *rickets*, a condition in which the bones of growing children become soft, resulting in skeletal deformities. Children, especially infants, have higher requirements for vitamin D than do adults. Milk with vitamin D supplements is available to most children. Adults who are exposed to normal amounts of sunlight do not usually require vitamin D supplements.

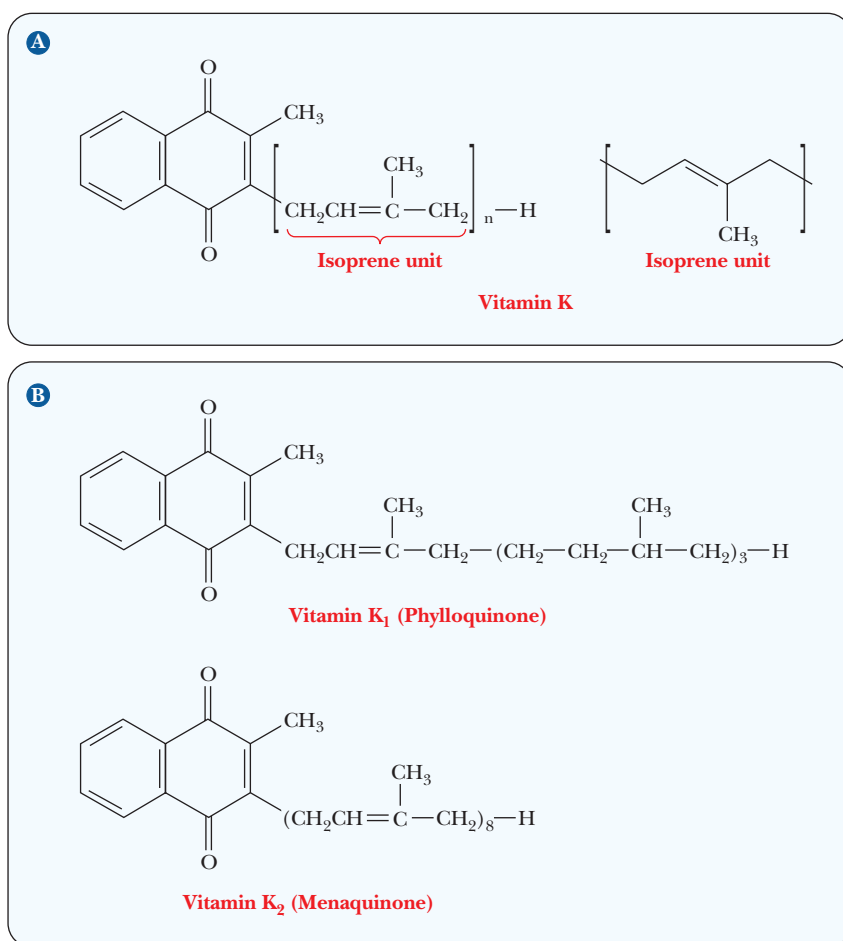
### Vitamin E

The most active form of **vitamin E** is  **$\alpha$ -tocopherol** (Figure 8.32). In rats, vitamin E is required for reproduction and for prevention of the disease *muscular dystrophy*. It is not known whether this requirement exists in humans. A well-established chemical property of vitamin E is that it is an **antioxidant**—that is, a good reducing agent—so it reacts with oxidizing agents before they can attack other biomolecules. The antioxidant action of vitamin E has been shown to protect important compounds, including vitamin A, from degradation in the laboratory; it probably also serves this function in organisms.

Recent research has shown that the interaction of vitamin E with membranes enhances its effectiveness as an antioxidant. Another function of antioxidants such as vitamin E is to react with, and thus to remove, the very reactive and highly dangerous substances known as **free radicals**. A free radical has at least one unpaired electron, which accounts for its high degree of reactivity. Free radicals may play a part in the development of cancer and in the aging process.

### Vitamin K

The name of **vitamin K** comes from the Danish *Koagulation* because this vitamin is an important factor in the blood-clotting process. The bicyclic ring system contains two carbonyl groups, the only polar groups on the molecule (Figure 8.33). A long unsaturated hydrocarbon side chain consists of repeating *isoprene* units, the number of which determines the exact form of vitamin K. Several forms of this vitamin can be found in a single organism, but the reason for this variation is not well understood. Vitamin K is not the first vitamin we have encountered that contains isoprene units, but it is the first one in which the number of isoprene units and their degree of saturation make a



■ **FIGURE 8.33 Vitamin K.** (a) The general structure of vitamin K, which is required for blood clotting. The value of  $n$  is variable, but it is usually  $<10$ . (b) Vitamin K<sub>1</sub> has one unsaturated isoprene unit; the rest are saturated. Vitamin K<sub>2</sub> has eight unsaturated isoprene units.

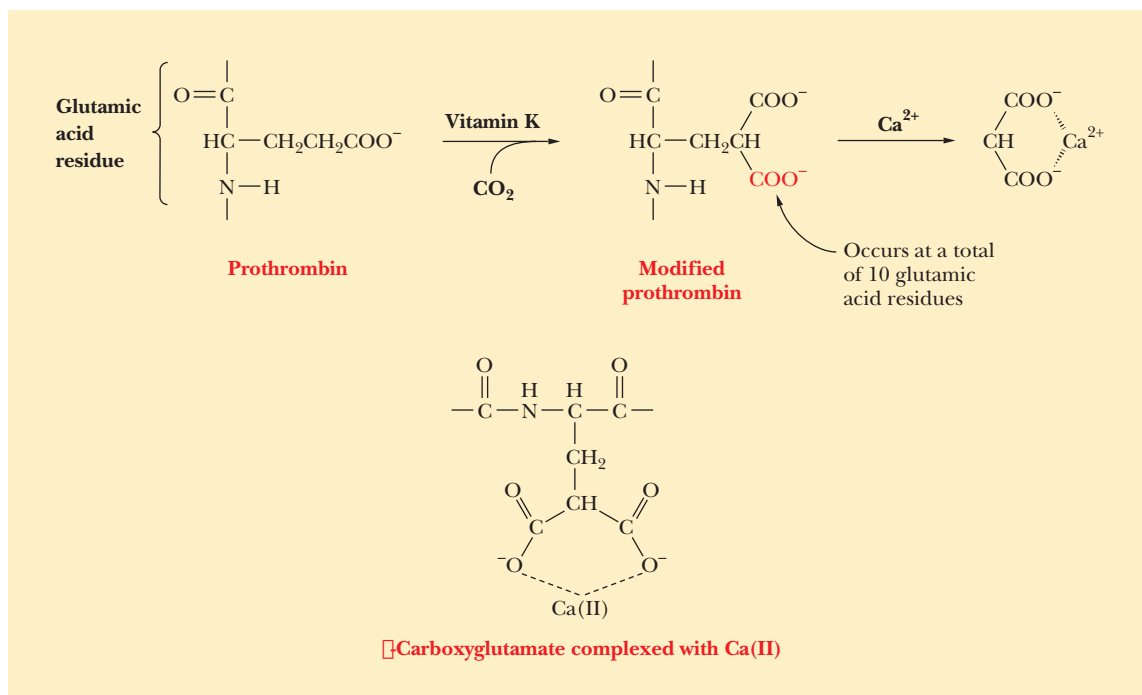
difference. (Can you pick out the isoprene-derived portions of the structures of vitamins A and E?) It is also known that the steroids are biosynthetically derived from isoprene units, but the structural relationship is not immediately obvious (Section 21.8).

The presence of vitamin K is required in the complex process of blood clotting, which involves many steps and many proteins and has stimulated numerous unanswered questions. It is known definitely that vitamin K is required to modify prothrombin and other proteins involved in the clotting process. Specifically, with prothrombin, the addition of another carboxyl group alters the side chains of several glutamate residues of prothrombin. This modification of glutamate produces  $\gamma$ -carboxyglutamate residues (Figure 8.34). The two carboxyl groups in proximity form a *bidentate* (“two teeth”) *ligand*, which can bind calcium ion ( $\text{Ca}^{2+}$ ). If prothrombin is not modified in this way, it does not bind  $\text{Ca}^{2+}$ . Even though there is a lot more to be learned about blood clotting and the role of vitamin K in the process, this point, at least, is well established, because  $\text{Ca}^{2+}$  is required for blood clotting. (Two well-known anticoagulants, dicumarol and warfarin [a rat poison], are vitamin K antagonists.)

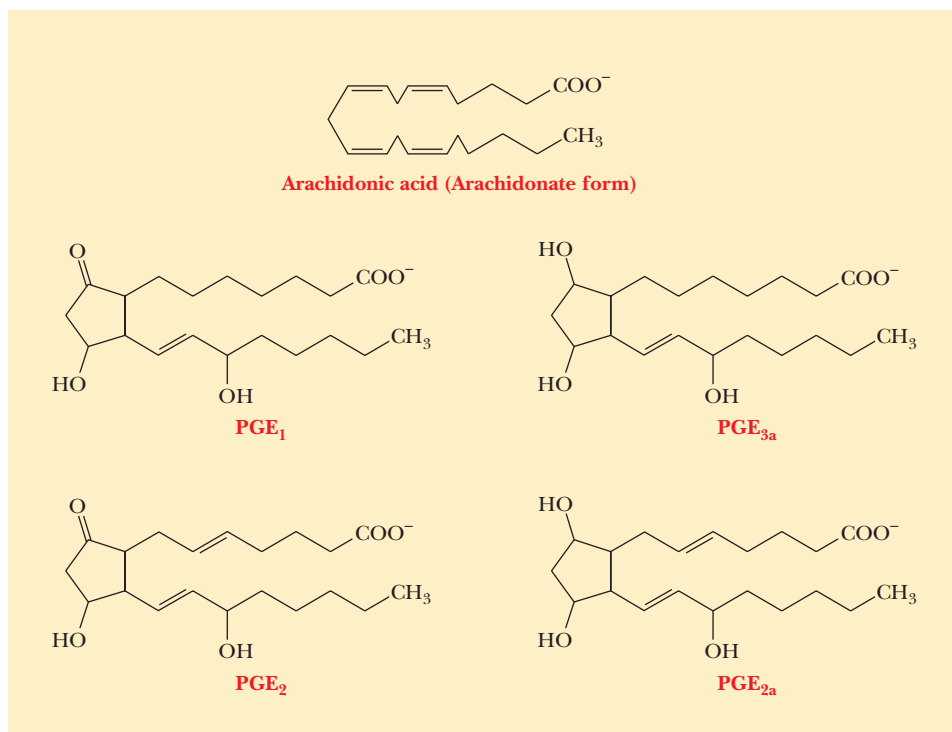
## 8.8 Prostaglandins and Leukotrienes

### What do prostaglandins and leukotrienes have to do with lipids?

A group of compounds derived from fatty acids has a wide range of physiological activities; they are called **prostaglandins** because they were first detected in seminal fluid, which is produced by the prostate gland. It has since been



■ **FIGURE 8.34** The role of vitamin K in the modification of prothrombin. The detailed structure of the  $\gamma$ -carboxyglutamate at the calcium complexation site is shown at the bottom.



■ **FIGURE 8.35** Arachidonic acid and some prostaglandins.

shown that they are widely distributed in a variety of tissues. The metabolic precursor of all prostaglandins is **arachidonic acid**, a fatty acid that contains 20 carbon atoms and four double bonds. The double bonds are not conjugated. The production of the prostaglandins from arachidonic acid takes place in several steps, which are catalyzed by enzymes. The prostaglandins themselves



each have a five-membered ring; they differ from one another in the numbers and positions of double bonds and oxygen-containing functional groups (Figure 8.35).

The structures of prostaglandins and their laboratory syntheses have been topics of great interest to organic chemists, largely because of the many physiological effects of these compounds and their possible usefulness in the pharmaceutical industry. Some of the functions of prostaglandins are control of blood pressure, stimulation of smooth-muscle contraction, and induction of inflammation. Aspirin inhibits the synthesis of prostaglandins, particularly in blood platelets, a property that accounts for its anti-inflammatory and fever-reducing properties. Cortisone and other steroids also have anti-inflammatory effects because of their inhibition of prostaglandin synthesis.

Prostaglandins are known to inhibit the aggregation of platelets. They may thus be of therapeutic value by preventing the formation of blood clots, which can cut off the blood supply to the brain or the heart and cause certain types of strokes and heart attacks. Even if this behavior were the only useful property of prostaglandins, it would justify considerable research effort. Heart attacks and strokes are two of the leading causes of death in industrialized countries. More recently, the study of prostaglandins has been a topic of great interest because of their possible antitumor and antiviral activity.

**Leukotrienes** are compounds that, like prostaglandins, are derived from arachidonic acid. They are found in leukocytes (white blood cells) and have three conjugated double bonds; these two facts account for the name. (Fatty acids and their derivatives do not normally contain conjugated double bonds.) Leukotriene C (Figure 8.36) is a typical member of this group; note the 20 carbon atoms in the carboxylic acid backbone, a feature that relates this compound structurally to arachidonic acid. (The 20-carbon prostaglandins and leukotrienes are also called eicosinoids.) An important property of leukotrienes is their constriction of smooth muscle, especially in the lungs. Asthma attacks may result from this constricting action because the synthesis of leukotriene C appears to be facilitated by allergic reactions, such as a reaction to pollen. Drugs that inhibit the synthesis of leukotriene C are now being used in the treatment of asthma, as are other drugs designed to block leukotriene receptors. In the United States, the incidence of asthma has increased drastically since 1980, providing considerable incentive to find new treatments. The Centers for Disease Control and Prevention have made information available on the Internet at <http://www.cdc.gov/asthma>.) Leukotrienes may also have inflammatory properties and may be involved in rheumatoid arthritis.

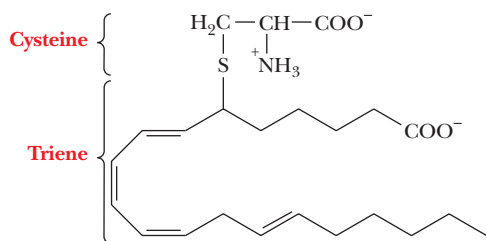
Thromboxanes are a third class of derivatives of arachidonic acid. They contain cyclic ethers as part of their structures. The most widely studied member of the group, thromboxane  $A_2$  ( $TxA_2$ ) (Figure 8.37), is known to induce platelet aggregation and smooth-muscle contraction.

The following Biochemical Connections box explores some connections among topics we have discussed in this chapter.

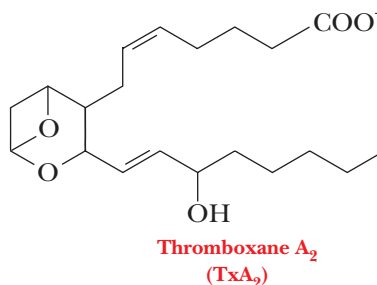


Image copyright Jason Stitt, 2010. Used under license from Shutterstock.com

■ Research on leukotrienes may provide new treatments for asthma, perhaps eliminating the need for inhalers, such as the one shown here.



■ **FIGURE 8.36** Leukotriene C.



■ **FIGURE 8.37** Thromboxane  $A_2$ .

# Biochemical Connections NUTRITION

## Why Should We Eat More Salmon?

Platelets are elements in the blood that initiate blood clotting and tissue repair by releasing clotting factors and platelet-derived growth factor (PDGF). Turbulence in the bloodstream may cause platelets to rupture. Fat deposits and bifurcations of arteries lead to such turbulence, so platelets and PDGF are implicated in blood clotting and growth of atherosclerotic plaque. Furthermore, the anaerobic conditions that exist under a large plaque deposit may lead to weakness and dead cells in the arterial wall, aggravating the problem.

Among peoples who depend on fish as a major food source, including some Eskimo tribes, very little heart disease is diagnosed, even though people in these groups eat high-fat diets and have high levels of blood cholesterol. Analysis of their diet led to the discovery that certain highly unsaturated fatty acids are found in the oils of fish and diving mammals. One class of these fatty acids is called omega-3 ( $\omega_3$ ), an example of which is eicosapentenoic acid (EPA).



Eicosapentenoic acid (EPA)

Note the presence of a double bond at the third carbon atom from the end of the hydrocarbon tail. The omega system of nomenclature is based on numbering the double bonds from the last carbon in the fatty acid instead of the carbonyl group (the delta [ $\Delta$ ] system). Omega is the last letter in the Greek alphabet.

The omega-3 fatty acids inhibit the formation of certain prostaglandins and thromboxane A, which is similar in structure to prostaglandins. Thromboxane released by ruptured arteries causes other platelets to clump in the immediate area and to

increase the size of the blood clot. Any disruption in thromboxane synthesis results in a lower tendency to form blood clots and, thus, in a lower potential for artery damage.

Aspirin also inhibits prostaglandin synthesis, although it is less potent than EPA. Aspirin inhibits the synthesis of the prostaglandins responsible for inflammation and the perception of pain. Aspirin has been implicated in reducing the incidence of heart disease, probably by a mechanism similar to that of EPA. However, people who are being treated with blood thinners or who are prone to easy bleeding should not take aspirin.



© Teutner/Stock Food Creative/Getty

## SUMMARY

**What are lipids?** Lipids are compounds that are insoluble in water but soluble in nonpolar organic solvents. One group of lipids consists of open-chain compounds, each with a polar head group and a long nonpolar tail; this group includes fatty acids, triacylglycerols, phosphoacylglycerols, sphingolipids, and glycolipids. Glycerol, fatty acids, and phosphoric acid are frequently obtained as degradation products of lipids. A second major group consists of fused-ring compounds, the steroids.

**What are fatty acids?** Fatty acids are carboxylic acids that may or may not have double bonds in their hydrocarbon portion.

**What are triacylglycerols?** Triacylglycerols are the storage forms of fatty acids, in which the acid part is esterified to glycerol.

**What are phosphoacylglycerols?** Phosphoacylglycerols differ from triacylglycerols in having a phosphorus-containing portion esterified to glycerol. These compounds are important components of biological membranes.

**What are waxes and sphingolipids?** Waxes are esters of long-chain fatty acids and long-chain alcohols. Sphingolipids do not contain glycerol but have a long-chain alcohol called sphingosine as part of their structure instead.

**What are glycolipids?** In glycolipids, a carbohydrate portion is covalently bonded to the lipid.

**What are steroids?** Steroids have a characteristic fused-ring structure. Other lipids are open-chain compounds.

**What is the structure of lipid bilayers?** A biological membrane consists of a lipid part and a protein part. The lipid part is a bilayer, with the polar head groups in contact with the aqueous interior and exterior of the cell, and the nonpolar portions of the lipid in the interior of the membrane.

**How does the composition of the bilayer affect its properties?** The presence of unsaturated fatty acids in membrane components leads to greater fluidity than with a preponderance of saturated fatty acids. Conversely, the presence of saturated

fatty acids and cholesterol tends to stiffen the bilayer. The packing of molecules in the bilayer can undergo a reversible transition from order to disorder. Lateral motion of lipid molecules within one layer of a membrane occurs frequently.

#### **How are proteins associated with the bilayer in membranes?**

The proteins that occur in membranes can be peripheral proteins, which are found on the surface of the membrane, or integral proteins, which lie within the lipid bilayer. Various structural motifs, such as bundles of seven  $\alpha$ -helices, occur in proteins that span membranes. Peripheral proteins are loosely attached to one surface of the membrane by hydrogen bonds or electrostatic attractions, whereas integral proteins are embedded solidly in the membrane.

**How do proteins and the lipid bilayer interact with each other in membranes?** The fluid-mosaic model describes the interaction of lipids and proteins in biological membranes. The proteins “float” in the lipid bilayer.

**How does transport through membranes take place?** Three important functions take place in or on membranes. The first, transport across the membrane, can involve the lipid bilayer as well as the membrane proteins. (The second, catalysis, is carried out by enzymes bound to the membrane.) The most important question about transport of substances across biological membranes is whether the process requires expenditure of energy by the cell. In passive transport, a substance moves from a region of higher concentration to one of lower concentration, requiring no expenditure of energy by the cell. Active transport requires moving substances against a concentration gradient, a situation similar to pumping water up a hill. Energy, as well as a carrier protein, is required for active transport. The sodium–potassium ion pump is an example of active transport.

**How do membrane receptors work?** Receptor proteins in the membrane bind biologically important substances that trigger a biochemical response in the cell. The first step in the effects of some biologically active substances is binding to a protein receptor site on the exterior of the cell. The interaction between receptor proteins and the active substances to which they bind is very similar to enzyme–substrate recognition. The action of a receptor frequently depends on a conformational change in the receptor protein. Receptors can be ligand-gated channel proteins, in which the binding of ligand transiently opens a channel protein through which substances such as ions can flow in the direction of a concentration gradient.

**What is the role of lipid-soluble vitamins in the body?** Lipid-soluble vitamins are hydrophobic, accounting for their solubility properties. Their structures are ultimately derived from five-carbon isoprene units. A derivative of vitamin A plays a crucial role in vision. Vitamin D controls calcium and phosphorus metabolism, affecting the structural integrity of bones. Vitamin E is known to be an antioxidant; its other metabolic functions are not definitely established. The presence of vitamin K is required in the blood-clotting process.

**What do prostaglandins and leukotrienes have to do with lipids?** The unsaturated fatty acid arachidonic acid is the precursor of prostaglandins and leukotrienes, compounds that have a wide range of physiological activities. Stimulation of smooth-muscle contraction and induction of inflammation are common to both classes of compounds. Prostaglandins are also involved in control of blood pressure and inhibition of blood-platelet aggregation.

## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

### 8.1 The Definition of a Lipid

1. **Recall** Proteins, nucleic acids, and carbohydrates are grouped by common structural features found within their group. What is the basis for grouping substances as lipids?

### 8.2 The Chemical Natures of the Lipid Types

2. **Recall** What structural features do a triacylglycerol and a phosphatidyl ethanolamine have in common? How do the structures of these two types of lipids differ?
3. **Recall** Draw the structure of a phosphoacylglycerol that contains glycerol, oleic acid, stearic acid, and choline.
4. **Recall** What structural features do a sphingomyelin and a phosphatidyl choline have in common? How do the structures of these two types of lipids differ?
5. **Recall** You have just isolated a pure lipid that contains only sphingosine and a fatty acid. To what class of lipids does it belong?
6. **Recall** What structural features does a sphingolipid have in common with proteins? Are there functional similarities?
7. **Recall** Write the structural formula for a triacylglycerol, and name the component parts.
8. **Recall** How does the structure of steroids differ from that of the other lipids discussed in this chapter?
9. **Recall** What are the structural features of waxes? What are some common uses of compounds of this type?
10. **Reflect and Apply** Which is more hydrophilic, cholesterol or phospholipids? Defend your answer.
11. **Reflect and Apply** Write an equation, with structural formulas, for the saponification of the triacylglycerol in Question 7.
12. **Reflect and Apply** Succulent plants from arid regions generally have waxy surface coatings. Suggest why such a coating is valuable for the survival of the plant.

13. **Reflect and Apply** In the produce department of supermarkets, vegetables and fruits (cucumbers are an example) have been coated with wax for shipping and storage. Suggest a reason why this is done.
14. **Reflect and Apply** Egg yolks contain a high amount of cholesterol, but they also contain a high amount of lecithin. From a diet and health standpoint, how do these two molecules complement each other?
15. **Reflect and Apply** In the preparation of sauces that involve mixing water and melted butter, egg yolks are added to prevent separation. How do the egg yolks prevent separation? *Hint:* Egg yolks are rich in phosphatidylcholine (lecithin).
16. **Reflect and Apply** When water birds have had their feathers fouled with crude oil after an oil spill, they are cleaned by rescuers to remove the spilled oil. Why are they not released immediately after they are cleaned?

### 8.3 Biological Membranes

17. **Recall** Which of the following lipids are *not* found in animal membranes?
  - (a) Phosphoglycerides
  - (b) Cholesterol
  - (c) Triacylglycerols
  - (d) Glycolipids
  - (e) Sphingolipids
18. **Recall** Which of the following statements is (are) consistent with what is known about membranes?
  - (a) A membrane consists of a layer of proteins sandwiched between two layers of lipids.
  - (b) The compositions of the inner and outer lipid layers are the same in any individual membrane.
  - (c) Membranes contain glycolipids and glycoproteins.
  - (d) Lipid bilayers are an important component of membranes.
  - (e) Covalent bonding takes place between lipids and proteins in most membranes.
19. **Reflect and Apply** Why might some food companies find it economically advantageous to advertise their product (for example, triacylglycerols) as being composed of polyunsaturated fatty acids with *trans*-double bonds?
20. **Reflect and Apply** Suggest a reason why partially hydrogenated vegetable oils are used so extensively in packaged foods.
21. **Biochemical Connections** Crisco is made from vegetable oils, which are usually liquid. Why is Crisco a solid? *Hint:* Read the label.
22. **Biochemical Connections** Why does the American Heart Association recommend the use of canola oil or olive oil rather than coconut oil in cooking?
23. **Reflect and Apply** In lipid bilayers, there is an order–disorder transition similar to the melting of a crystal. In a lipid bilayer in which most of the fatty acids are unsaturated, would you expect this transition to occur at a higher temperature, a lower temperature, or the same temperature as it would in a lipid bilayer in which most of the fatty acids are saturated? Why?
24. **Biochemical Connections** Briefly discuss the structure of myelin and its role in the nervous system.
25. **Reflect and Apply** Suggest a reason why the cell membranes of bacteria grown at 20°C tend to have a higher proportion of unsaturated fatty acids than the membranes of bacteria of the same species grown at 37°C. In other words, the bacteria grown at 37°C have a higher proportion of saturated fatty acids in their cell membranes.
26. **Reflect and Apply** Suggest a reason why animals that live in cold climates tend to have higher proportions of polyunsaturated fatty acid residues in their lipids than do animals that live in warm climates.

27. **Reflect and Apply** What is the energetic driving force for the formation of phospholipid bilayers?

### 8.4 The Kinds of Membrane Proteins

28. **Recall** Define *glycoprotein* and *glycolipid*.
29. **Recall** Do all proteins associated with membranes span the membrane from one side to another?
30. **Reflect and Apply** A membrane consists of 50% protein by weight and 50% phosphoglycerides by weight. The average molecular weight of the lipids is 800 Da, and the average molecular weight of the proteins is 50,000 Da. Calculate the molar ratio of lipid to protein.
31. **Reflect and Apply** Suggest a reason why the same protein system moves both sodium and potassium ions into and out of the cell.
32. **Reflect and Apply** Suppose that you are studying a protein involved in transporting ions in and out of cells. Would you expect to find the nonpolar residues in the interior or the exterior? Why? Would you expect to find the polar residues in the interior or the exterior? Why?

### 8.5 The Fluid-Mosaic Model of Membrane Structure

33. **Reflect and Apply** Which statements are consistent with the fluid-mosaic model of membranes?
  - (a) All membrane proteins are bound to the interior of the membrane.
  - (b) Both proteins and lipids undergo transverse (flip-flop) diffusion from the inside to the outside of the membrane.
  - (c) Some proteins and lipids undergo lateral diffusion along the inner or outer surface of the membrane.
  - (d) Carbohydrates are covalently bonded to the outside of the membrane.
  - (e) The term *mosaic* refers to the arrangement of the lipids alone.

### 8.6 The Functions of Membranes

34. **Reflect and Apply** Suggest a reason why inorganic ions, such as  $K^+$ ,  $Na^+$ ,  $Ca^{2+}$ , and  $Mg^{2+}$ , do not cross biological membranes by simple diffusion.
35. **Reflect and Apply** Which statements are consistent with the known facts about membrane transport?
  - (a) Active transport moves a substance from a region in which its concentration is lower to one in which its concentration is higher.
  - (b) Transport does not involve any pores or channels in membranes.
  - (c) Transport proteins may be involved in bringing substances into cells.

### 8.7 Lipid-Soluble Vitamins and Their Functions

36. **Recall** What is the structural relationship between vitamin  $D_3$  and cholesterol?
37. **Recall** List an important chemical property of vitamin E.
38. **Recall** What are isoprene units? What do they have to do with the material of this chapter?
39. **Recall** List the fat-soluble vitamins, and give a physiological role for each.
40. **Biochemical Connections** What is the role in vision of the *cis–trans* isomerization of retinal?
41. **Reflect and Apply** Why is it possible to argue that vitamin D is not a vitamin?
42. **Reflect and Apply** Give a reason for the toxicity that can be caused by overdoses of lipid-soluble vitamins.
43. **Reflect and Apply** Why can some vitamin-K antagonists act as anticoagulants?



44. **Reflect and Apply** Why are many vitamin supplements sold as antioxidants? How does this relate to material in this chapter?
45. **Reflect and Apply** A health-conscious friend asks whether eating carrots is better for the eyesight or for preventing cancer. What do you tell your friend? Explain.
- 8.8 Prostaglandins and Leukotrienes**
46. **Biochemical Connections** Define omega-3 fatty acid.
47. **Recall** What are the main structural features of leukotrienes?
48. **Recall** What are the main structural features of prostaglandins?
49. **Reflect and Apply** List two classes of compounds derived from arachidonic acid. Suggest some reasons for the amount of biomedical research devoted to these compounds.
50. **Biochemical Connections** Outline a possible connection between the material in this chapter and the integrity of blood platelets.

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





# Nucleic Acids: How Structure Conveys Information

# 9

Determination of the double-helical structure of DNA has illuminated molecular biology for more than half a century.

## 9.1 Levels of Structure in Nucleic Acids

In Chapter 4, we identified four levels of structure—primary, secondary, tertiary, and quaternary—in proteins. Nucleic acids can be viewed in the same way. The *primary structure* of nucleic acids is the order of bases in the polynucleotide sequence, and the *secondary structure* is the three-dimensional conformation of the backbone. The *tertiary structure* is specifically the supercoiling of the molecule.

There are two principal types of nucleic acids, DNA (deoxyribonucleic acid) and RNA (ribonucleic acid).

### How do DNA and RNA differ?

Important differences between DNA and RNA appear in their secondary and tertiary structures, and so we shall describe these structural features separately for DNA and for RNA. Even though nothing in nucleic acid structure is directly analogous to the quaternary structure of proteins, the interaction of nucleic acids with other classes of macromolecules (for example, proteins) to form complexes is similar to the interactions of the subunits in an oligomeric protein. One well-known example is the association of RNA and proteins in **ribosomes** (the polypeptide-generating machinery of the cell); another is the self-assembly of tobacco mosaic virus, in which the nucleic acid strand winds through a cylinder of coat-protein subunits.

## 9.2 The Covalent Structure of Polynucleotides

Polymers can always be broken down into smaller and smaller units until we are left with the smallest single unit of the polymer, called a monomer. The monomers of nucleic acids are **nucleotides**. An individual nucleotide consists of three parts—a nitrogenous base, a sugar, and a phosphoric acid residue—all of which are covalently bonded together.

The order of bases in the nucleic acids of DNA contains the information necessary to produce the correct amino acid sequence in the cell's proteins.

### What are the structures and components of the nucleotides?

The **nucleic acid bases** (also called **nucleobases**) are of two types—*pyrimidines* and *purines* (Figure 9.1). In this case, the word *base* does not refer to an alkaline compound, such as NaOH; rather, it refers to a one- or two-ring nitrogenous aromatic compound. Three **pyrimidine bases** (single-ring aromatic compounds)—*cytosine*, *thymine*, and *uracil*—commonly occur. Cytosine is found both in RNA and in DNA. Uracil occurs only in RNA. In DNA, thymine is substituted for uracil; thymine is also found to a small extent in some forms

## Chapter Outline

### 9.1 Levels of Structure in Nucleic Acids

- How do DNA and RNA differ?

### 9.2 The Covalent Structure of Polynucleotides

- What are the structures and components of the nucleotides?
- How do nucleotides combine to give nucleic acids?

### 9.3 The Structure of DNA

- What is the nature of the DNA double helix?
- Are there other possible conformations of the double helix?
- How does prokaryotic DNA supercoil into its tertiary structure?
- How does supercoiling take place in eukaryotic DNA?

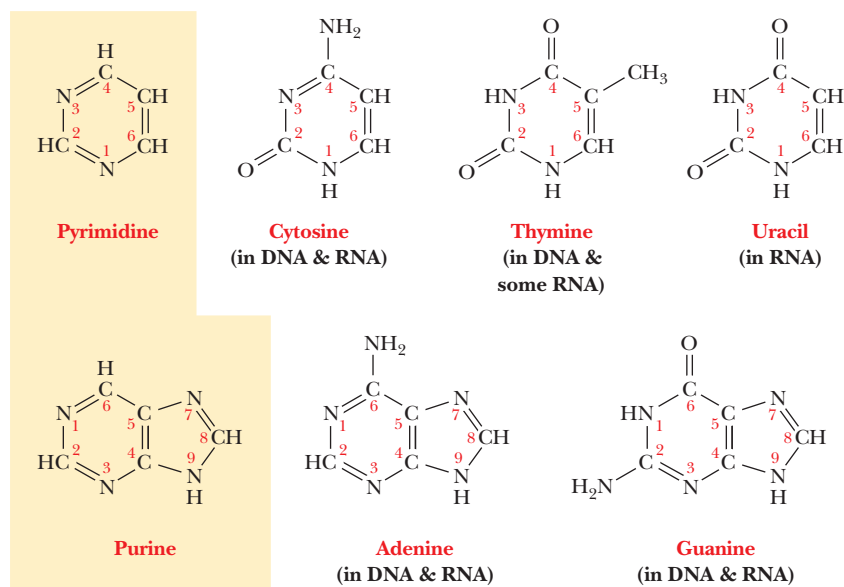
### 9.4 Denaturation of DNA

- How can we monitor DNA denaturation?

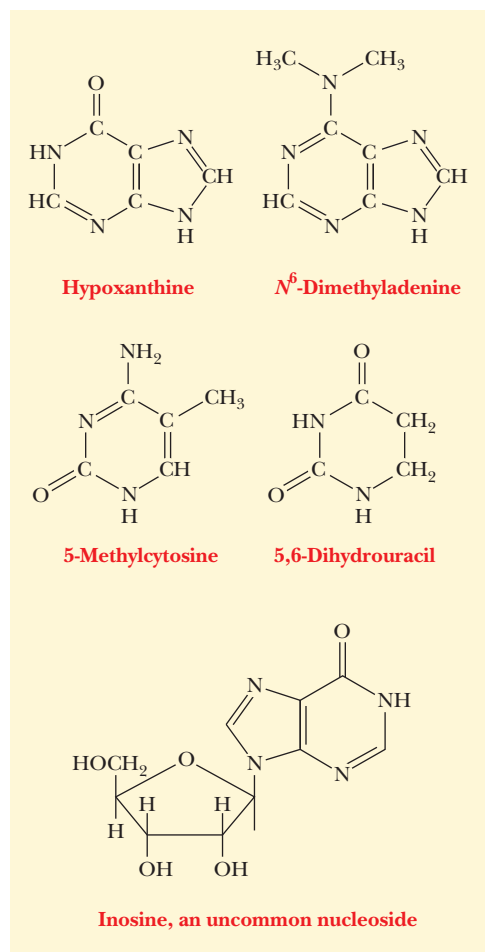
### 9.5 The Principal Kinds of RNA and Their Structures

- What kinds of RNA play a role in life processes?
- What is the role of transfer RNA in protein synthesis?
- How does ribosomal RNA combine with proteins to form the site of protein synthesis?
- How does messenger RNA direct protein synthesis?
- How does small nuclear RNA help with the processing of RNA?
- What is RNA interference, and why is it important?

**OWL** Online homework for this chapter may be assigned in OWL.



■ **FIGURE 9.1** Structures of the common nucleobases. The structures of pyrimidine and purine are shown for comparison.



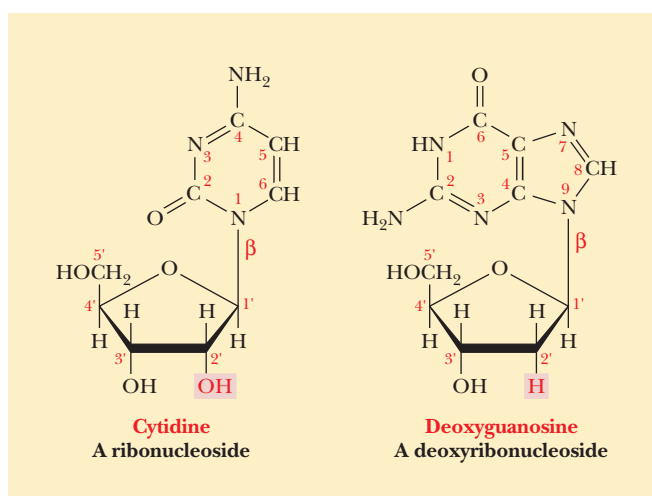
■ **FIGURE 9.2** Structures of some of the less common nucleobases. When hypoxanthine is bonded to a sugar, the corresponding compound is called inosine.

of RNA. The common **purine bases** (double-ring aromatic compounds) are *adenine* and *guanine*, both of which are found in RNA and in DNA (Figure 9.1). In addition to these five commonly occurring bases, there are “unusual” bases, with slightly different structures, that are found principally, but not exclusively, in transfer RNA (Figure 9.2). In many cases, the base is modified by methylation.

A **nucleoside** is a compound that consists of a base and a sugar covalently linked together. It differs from a nucleotide by lacking a phosphate group in its structure. In a nucleoside, a base forms a glycosidic linkage with the sugar. Glycosidic linkages and the stereochemistry of sugars are discussed in detail in Section 16.2. If you wish to look now at the material on the structure of sugars, you will find that it does not depend on material in the intervening chapters. For now, it is sufficient to say that a *glycosidic bond* links a sugar and some other moiety. When the sugar is  $\beta$ -D-ribose, the resulting compound is a **ribonucleoside**; when the sugar is  $\beta$ -D-deoxyribose, the resulting compound is a **deoxyribonucleoside** (Figure 9.3). The glycosidic linkage is from the C-1' carbon of the sugar to the N-1 nitrogen of pyrimidines or to the N-9 nitrogen of purines. The ring atoms of the base and the carbon atoms of the sugar are both numbered, with the numbers of the sugar atoms primed to prevent confusion. Note that the sugar is linked to a nitrogen in both cases (an *N*-glycosidic bond).

When phosphoric acid is esterified to one of the hydroxyl groups of the sugar portion of a nucleoside, a nucleotide is formed (Figure 9.4). A nucleotide is named for the parent nucleoside, with the suffix *-monophosphate* added; the position of the phosphate ester is specified by the number of the carbon atom at the hydroxyl group to which it is esterified—for instance, adenosine 3'-monophosphate or deoxycytidine 5'-monophosphate.

The 5' nucleotides are most commonly encountered in nature. If additional phosphate groups form anhydride linkages to the first phosphate, the corresponding nucleoside diphosphates and triphosphates are formed. Recall this point from Section 2.2. These compounds are also nucleotides.



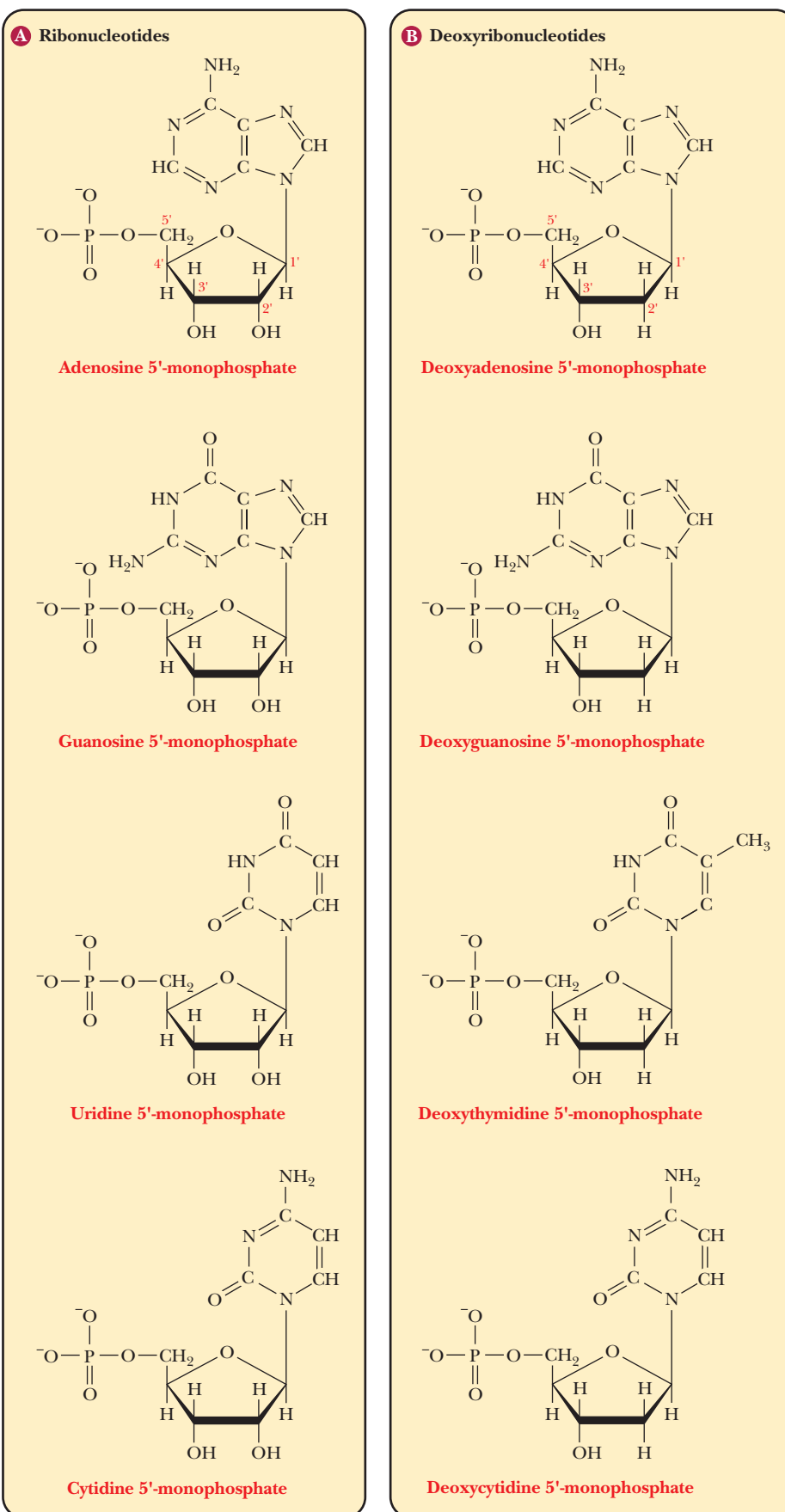
■ **FIGURE 9.3** A comparison of the structures of a ribonucleoside and a deoxyribonucleoside. (A nucleoside does not have a phosphate group in its structure.)

### How do nucleotides combine to give nucleic acids?

The polymerization of nucleotides gives rise to nucleic acids. The linkage between monomers in nucleic acids involves formation of two ester bonds by phosphoric acid. The hydroxyl groups to which the phosphoric acid is esterified are those bonded to the 3' and 5' carbons on adjacent residues. The resulting repeated linkage is a **3', 5'-phosphodiester bond**. The nucleotide residues of nucleic acids are numbered from the 5' end, which normally carries a phosphate group, to the 3' end, which normally has a free hydroxyl group.

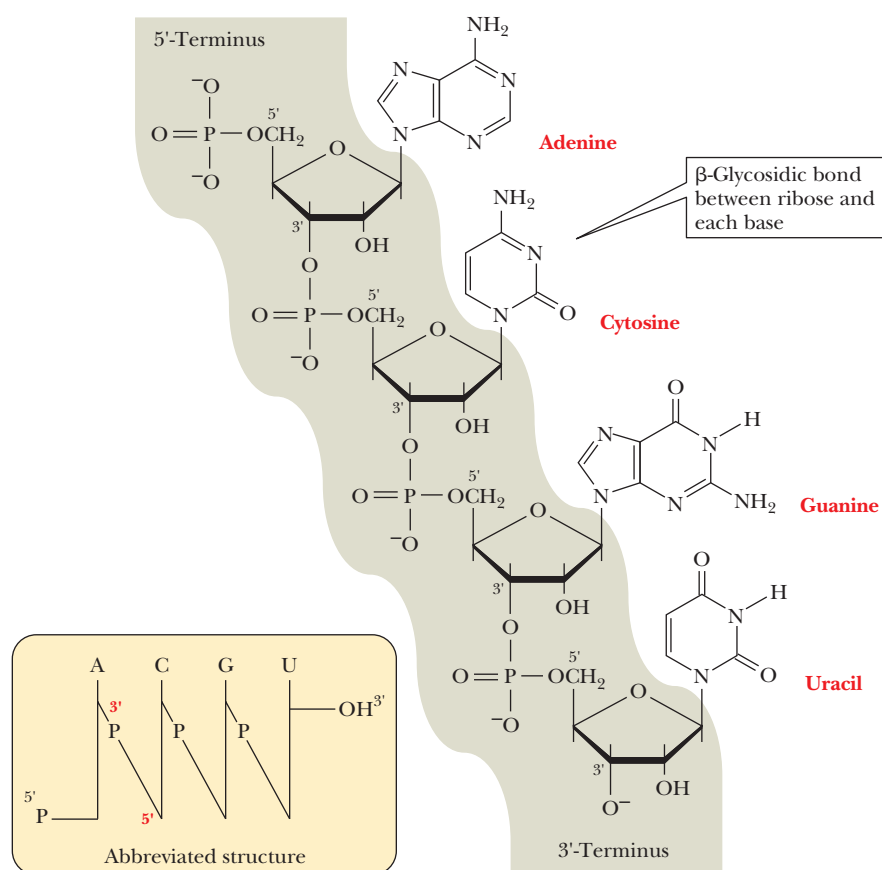
Figure 9.5 shows the structure of a fragment of an RNA chain. The *sugar-phosphate backbone* repeats itself down the length of the chain. The most important features of the structure of nucleic acids are the identities of the bases. Abbreviated forms of the structure can be written to convey this essential information. In one system of notation, single letters, such as A, G, C, U, and T, represent the individual bases. Vertical lines show the positions of the sugar moieties to which the individual bases are attached, and a diagonal line through the letter P represents a phosphodiester bond (Figure 9.5). However, an even more common system of notation uses only the single letters to show the order of the bases. When it is necessary to indicate the position on the sugar to which the phosphate group is bonded, the letter p is written to the left of the single-letter code for the base to represent a 5' nucleotide and to the right to represent a 3' nucleotide. For example, pA signifies 5'-adenosine monophosphate (5'-AMP), and Ap signifies 3'-AMP. The sequence of an oligonucleotide can be represented as pGpApCpApU or, even more simply, as GACAU, with the phosphates understood.

A portion of a DNA chain differs from the RNA chain just described only in the fact that the sugar is 2'-deoxyribose rather than ribose (Figure 9.6). In abbreviated notation, the deoxyribonucleotide is specified in the usual manner. Sometimes a letter d is added to indicate a deoxyribonucleotide residue; for example, dG is substituted for G, and the deoxy analogue of the ribooligonucleotide in the preceding paragraph would be d(GACAT). However, given that the sequence must refer to DNA because of the presence of thymine, the sequence GACAT is not ambiguous and would also be a suitable abbreviation.

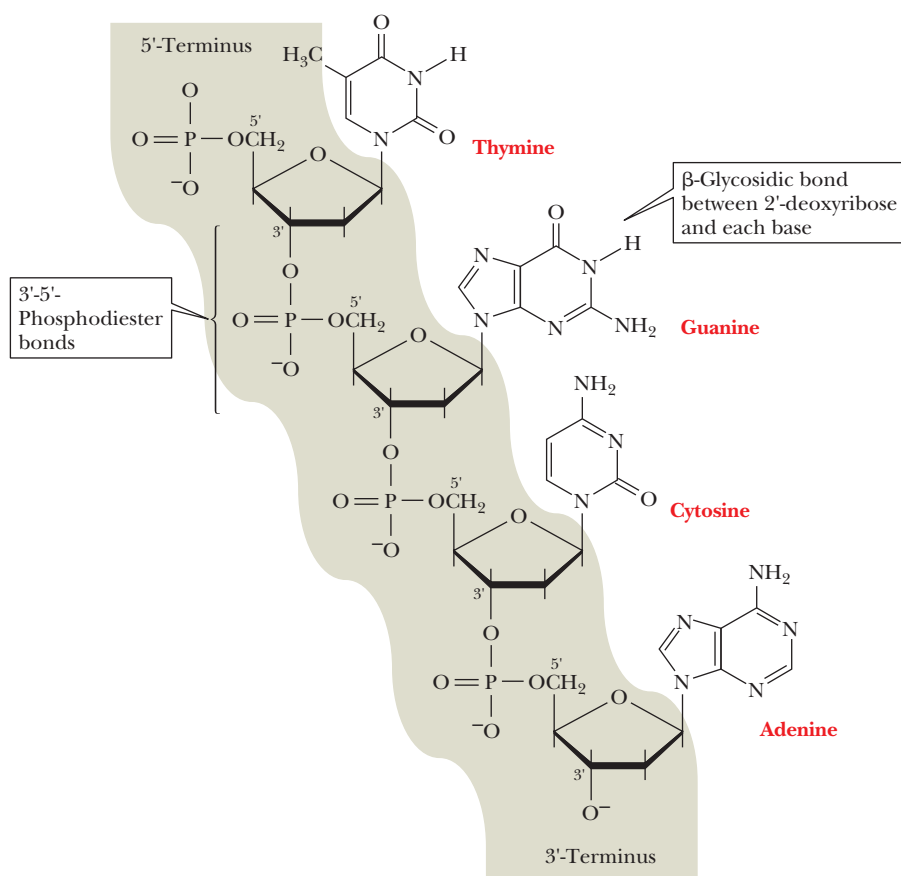


■ **FIGURE 9.4** The structures and names of the commonly occurring nucleotides. Each nucleotide has a phosphate group in its structure. All structures are shown in the forms that exist at pH 7.





■ **FIGURE 9.5** A fragment of an RNA chain.



■ **FIGURE 9.6** A portion of a DNA chain.

# Biochemical Connections GENETICS

## The DNA Family Tree

Because it is easy to determine the sequence of DNA, even using automated and robotic systems that require little supervision, the amount of DNA sequence data available has virtually exploded. Many scientific journals no longer report full sequences; the information is just incorporated into so-called gene banks, large computer systems that store the data.

The sequence information, for proteins as well as for DNA, is readily available to anyone with a web search program. See <http://www.jcvi.org> (the J. Craig Venter Institute) for genomic databases and <http://expasy.hcuge.ch> (the ExPASy molecular biology server maintained by Geneva University Hospital and the University of Geneva). The ExPASy site is a repository for information about protein sequences as well as DNA sequences. A particularly useful site is <http://www.genome.gov>, maintained by the National Human Genome Research Institute of the National Institutes of Health. Another government-sponsored site is <http://genomics.energy.gov>, one of the most important repositories of results of the Human Genome Project. So much information is entered into the databanks that it has become necessary to develop new and more efficient computer technology to search and to compare such sequences. We are just beginning to appreciate the usefulness of so much information (Section 13.9). Many new applications, not even thought of at this time, will undoubtedly be developed. Here are two applications that give molecular information about evolution, the “family tree” of all living things.

1. *Molecular taxonomy.* In ways never before possible, we can compare the sequences not just from existing organisms but also, when DNA is available from fossil specimens, from extinct

ancestors of living organisms. Within given genetic families of limited size, this information has enabled very detailed evolutionary trees to be developed. It has been possible to show that, in some areas, all plants are clones of one another. The largest living organism is a soil fungus that spreads over several acres. Redwood trees grow as clones from a central root system after forest fires. Sadly, many endangered species have such small remaining numbers that all living specimens are closely related to each other. This is true for all nene geese, which are native to Hawaii; for all California condors; and even for some whale species. The lack of genetic diversity in these endangered species may mean that the species are doomed to extinction, in spite of human attempts to ensure their survival.

2. *Ancient DNA.* DNA has been isolated from human fossils, such as mummies, bog people, and the frozen man found in the Alps, allowing comparisons of modern humans to recent relatives. Mitochondrial sequencing has shown that all humans now alive radiated out from one region in Africa some 100,000 to 200,000 years ago. More ancient DNA sequences from insect specimens preserved in amber have been compared to their modern counterparts. The film *Jurassic Park* is based on the suggestion that dinosaurs might be cloned from the DNA in their blood, which survived in the gut of an insect preserved in amber—certainly a far-fetched possibility, although entertaining (and profitable to filmmakers). The acceptance of DNA sequence data from ancient DNA is still controversial because of the likelihood of DNA degradation over time, contamination with modern DNA, and damage due to the initial chemical treatment of the samples.



Charlie Heidecker/Visuals Unlimited

■ Nene geese, native to Hawaii, are an endangered species. Even those in European zoos are related to the ones left in Hawaii.



J. Koivula/Science Source/Photo Researchers, Inc.

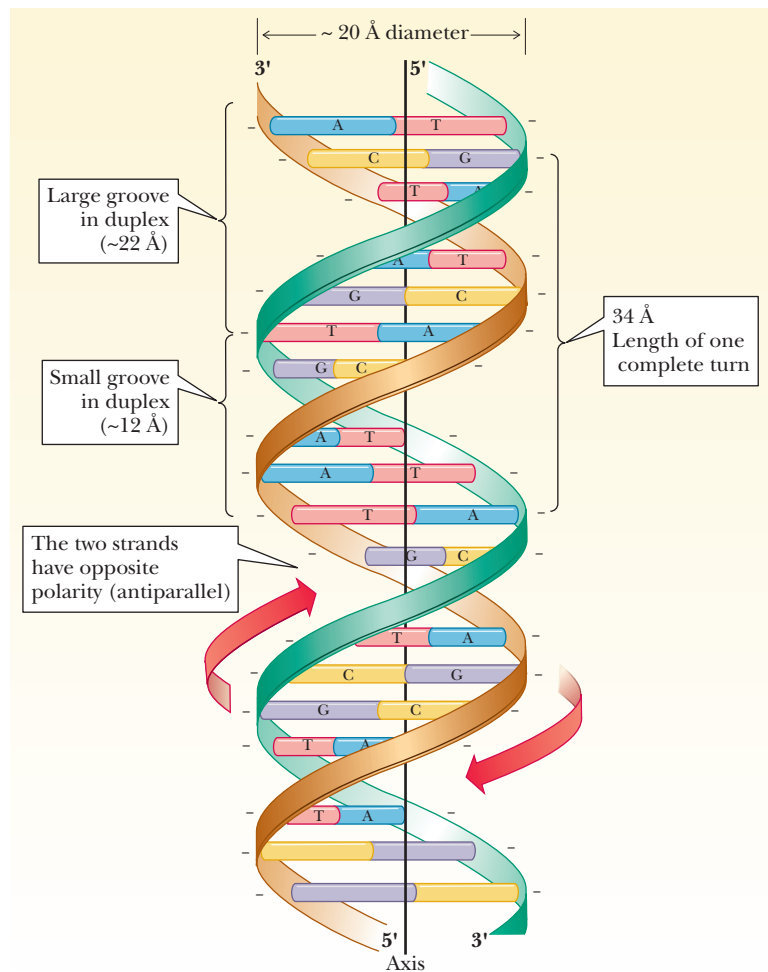
■ Insects preserved in amber.

## 9.3 The Structure of DNA

Representations of the double-helical structure of DNA have become common in the popular press as well as in the scientific literature. When the **double helix** was proposed by James Watson and Francis Crick in 1953, it touched off a flood of research activity, leading to great advances in molecular biology.

### What is the nature of the DNA double helix?

The determination of the double-helical structure was based primarily on model building and X-ray diffraction patterns. Information from X-ray patterns was added to information from chemical analyses that showed that the amount

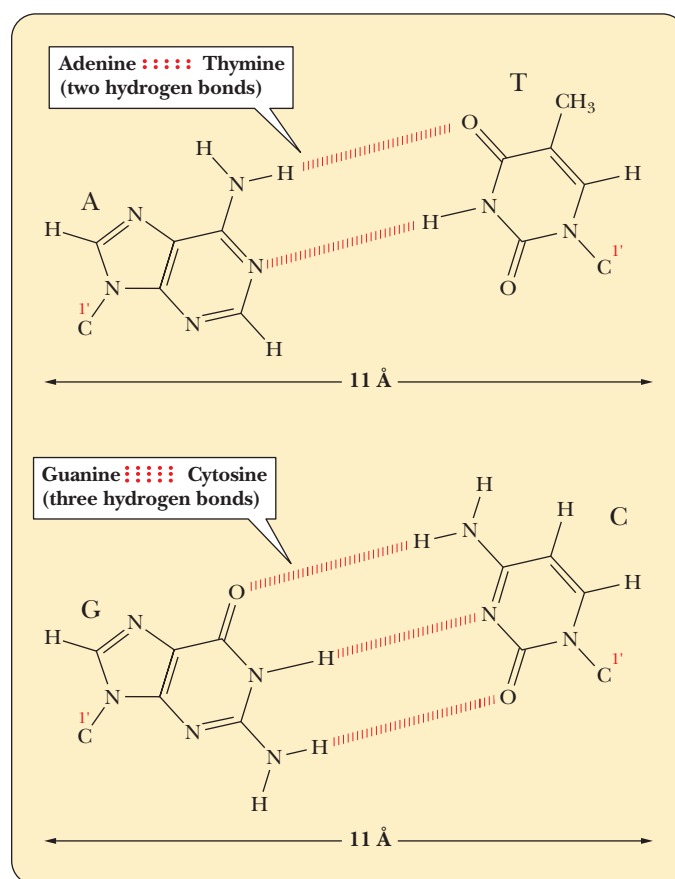


■ **FIGURE 9.7 The double helix.** A complete turn of the helix spans ten base pairs, covering a distance of 34 Å (3.4 nm). The individual base pairs are spaced 3.4 Å (0.34 nm) apart. The places where the strands cross hide base pairs that extend perpendicular to the viewer. The inside diameter is 11 Å (1.1 nm), and the outside diameter is 20 Å (2.0 nm). Within the cylindrical outline of the double helix are two grooves, a small one and a large one. Both are large enough to accommodate polypeptide chains. The minus signs alongside the strands represent the many negatively charged phosphate groups along the entire length of each strand.

of A was always the same as the amount of T, and that the amount of G always equaled the amount of C. Both of these lines of evidence were used to conclude that DNA consists of two polynucleotide chains wrapped around each other to form a helix. Hydrogen bonds between bases on opposite chains determine the alignment of the helix, with the paired bases lying in planes perpendicular to the helix axis. The sugar-phosphate backbone is the outer part of the helix (Figure 9.7). The chains run in antiparallel directions, one 3' to 5' and the other 5' to 3'.

The X-ray diffraction pattern of DNA demonstrated the helical structure and the diameter. The combination of evidence from X-ray diffraction and chemical analysis led to the conclusion that the base pairing is *complementary*, meaning that adenine pairs with thymine and that guanine pairs with cytosine. Because complementary base pairing occurs along the entire double helix, the two chains are also referred to as *complementary strands*. By 1953, studies of the base composition of DNA from many species had already shown that, to within experimental error, the mole percentages of adenine and thymine (moles of these substances as percentages of the total) were equal; the same was found to be the case with guanine and cytosine. An adenine–thymine (A–T) base pair has two hydrogen bonds between the bases; a guanine–cytosine (G–C) base pair has three (Figure 9.8).

The inside diameter of the sugar-phosphate backbone of the double helix is about 11 Å (1.1 nm). The distance between the points of attachment of the bases to the two strands of the sugar-phosphate backbone is the same for the two base pairs (A–T and G–C), about 11 Å (1.1 nm), which allows for a double helix with a smooth backbone and no overt bulges. Base pairs other than A–T and G–C are possible, but they do not have the correct hydrogen



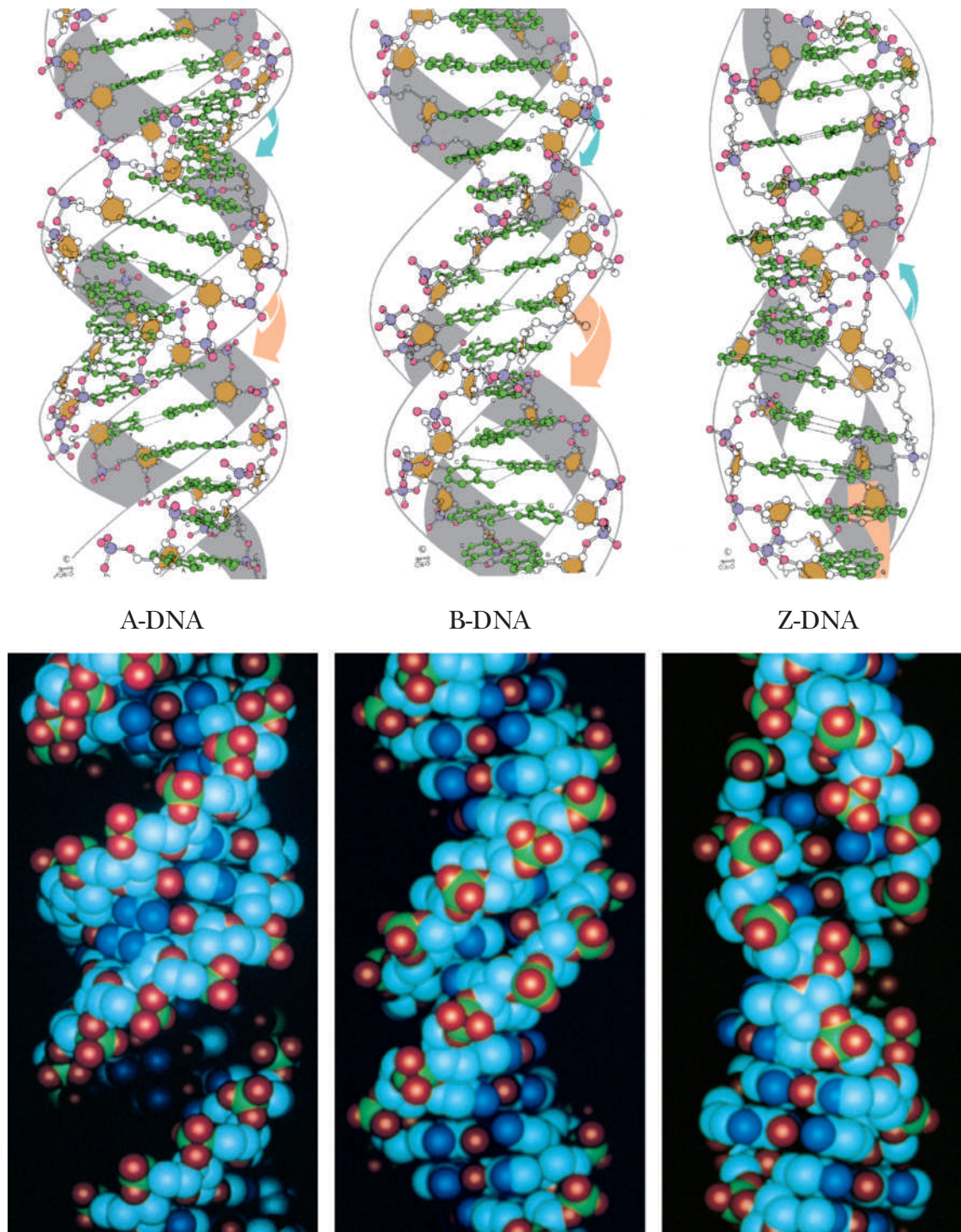
■ **FIGURE 9.8 Base pairing.** The adenine–thymine (A–T) base pair has two hydrogen bonds, whereas the guanine–cytosine (G–C) base pair has three hydrogen bonds.

bonding pattern (A–C or G–T pairs) or the right dimensions (purine–purine or pyrimidine–pyrimidine pairs) to allow for a smooth double helix (Figure 9.8). The outside diameter of the helix is 20 Å (2 nm). The length of one complete turn of the helix along its axis is 34 Å (3.4 nm) and contains 10 base pairs. The atoms that make up the two polynucleotide chains of the double helix do not completely fill an imaginary cylinder around the double helix; they leave empty spaces known as grooves. There is a large **major groove** and a smaller **minor groove** in the double helix; both can be sites at which drugs or polypeptides bind to DNA (see Figure 9.7). At neutral, physiological pH, each phosphate group of the backbone carries a negative charge. Positively charged ions, such as  $\text{Na}^+$  or  $\text{Mg}^{2+}$ , and polypeptides with positively charged side chains must be associated with DNA in order to neutralize the negative charges. Eukaryotic DNA, for example, is complexed with histones, which are positively charged proteins, in the cell nucleus.

### Are there other possible conformations of the double helix?

The form of DNA that we have been discussing so far is called **B-DNA**. It is thought to be the principal form that occurs in nature. However, other secondary structures can occur, depending on conditions such as the nature of the positive ion associated with the DNA and the specific sequence of bases. One of those other forms is **A-DNA**, which has 11 base pairs for each turn of the helix. Its base pairs are not perpendicular to the helix axis but lie at an angle of about  $20^\circ$  to the perpendicular (Figure 9.9). An important shared feature of A-DNA and B-DNA is that both are right-handed helices; that is, the helix winds upward in the direction in which the fingers of the right hand curl when the thumb is pointing upward (Figure 9.10). The A form of DNA was originally found in dehydrated DNA samples, and many researchers believed that the A form was an artifact of DNA preparation. DNA:RNA hybrids can adopt an A





■ **FIGURE 9.9** Comparison of the A, B, and Z forms of DNA. In the A form, the base pairs have a marked propeller-twist with respect to the helix axis. In the B form, the base pairs lie in a plane that is close to perpendicular to the helix axis. Z-DNA is a left-handed helix and in this respect differs from A-DNA and B-DNA, both of which are right-handed helices. (Robert Stodala, Fox Chase Cancer Research Center. Illustration, Irving Geis. Rights owned by Howard Hughes Medical Institute. Not to be reproduced without permission.)

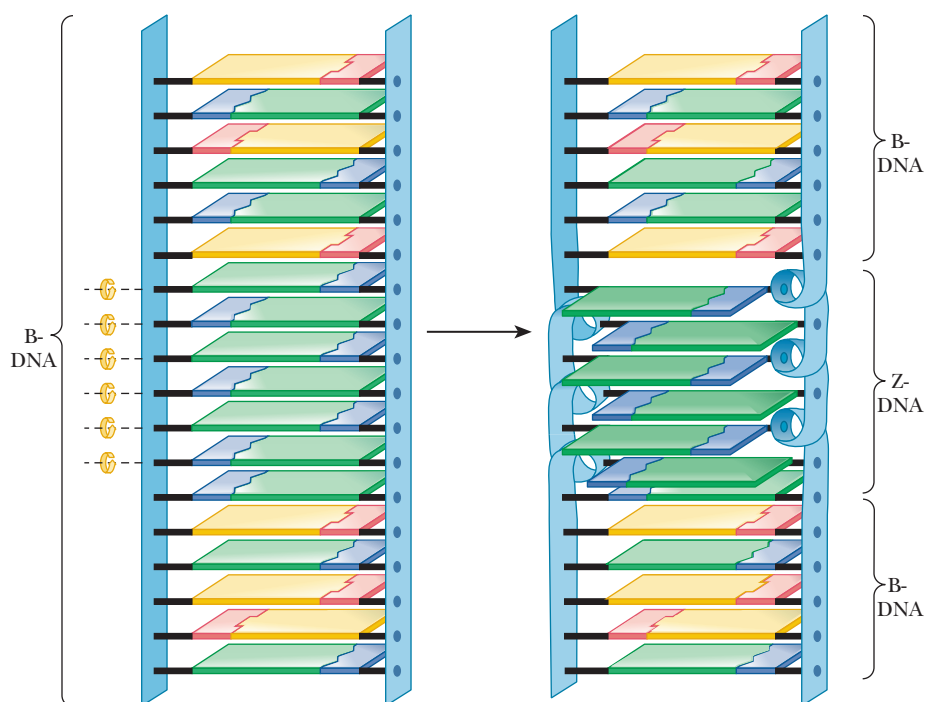
formation because the 2'-hydroxyl on the ribose prevents an RNA helix from adopting the B form; RNA:RNA hybrids may also be found in the A form.

Another variant form of the double helix, **Z-DNA**, is left-handed; it winds in the direction of the fingers of the left hand (Figure 9.10). Z-DNA is known to occur in nature, most often when there is a sequence of alternating

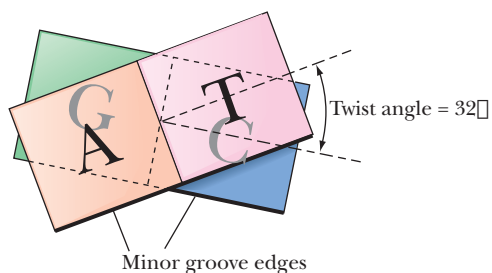




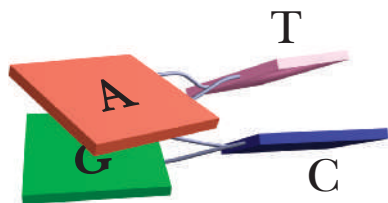
■ **FIGURE 9.10** Right- and left-handed helices are related to each other in the same way as right and left hands.



■ **FIGURE 9.11** Formation of Z-DNA. A Z-DNA section can form in the middle of a section of B-DNA by rotation of the base pairs, as indicated by the curved arrows.



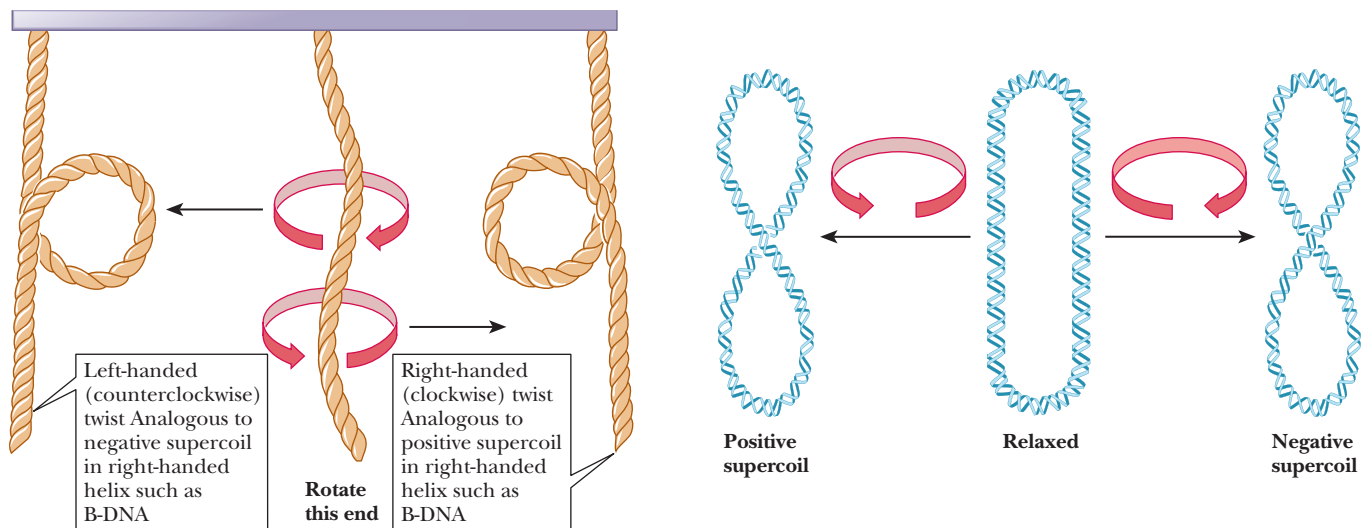
■ **FIGURE 9.12** Helical twists. Two base pairs with  $32^\circ$  of right-handed helical twist; the minor-groove edges are drawn with heavy shading.



■ **FIGURE 9.13** Propeller-twisted base pairs. Note how the hydrogen bonds between bases are distorted by this motion, yet remain intact. The minor-groove edges of the bases are shaded.

purine–pyrimidine, such as dCpGpCpGpCpG. Sequences with cytosine methylated at the number 5 position of the pyrimidine ring can also be found in the Z form. It may play a role in the regulation of gene expression. The Z form of DNA is also a subject of active research among biochemists. The Z form of DNA can be considered a derivative of the B form of DNA, produced by flipping one side of the backbone  $180^\circ$  without having to break either the backbone or the hydrogen bonding of the complementary bases. Figure 9.11 shows how this might occur. The Z form of DNA gets its name from the zigzag look of the phosphodiester backbone when viewed from the side.

The B form of DNA has long been considered the normal, physiological DNA form. It was predicted from the nature of the hydrogen bonds between purines and pyrimidines and later found experimentally. Although it is easy to focus completely on the base pairing and the order of bases in DNA, other features of DNA structure are just as important. The ring portions of the DNA bases are very hydrophobic and interact with each other via hydrophobic bonding of their pi-cloud electrons. This process is usually referred to as **base stacking**, and even single-stranded DNA tends to form structures in which the bases can stack. In standard B-DNA, each base pair is rotated  $32^\circ$  with respect to the preceding one (Figure 9.12). This form is perfect for maximal base pairing, but it is not optimal for maximal overlap of the bases. In addition, the edges of the bases that are exposed to the minor groove must come in contact with water in this form. Many of the bases twist in a characteristic way, called *propeller-twist* (Figure 9.13). In this form, the base-pairing distances are less optimal, but the base stacking is more optimal, and water is eliminated from the minor-groove contacts with the bases. Besides twisting, bases also slide sideways, allowing them to interact better with the bases above and below them. The twist and slide depends on which bases are present, and researchers have identified that a basic unit for studying DNA structure is actually a dinucleotide with its complementary pairs. This is called a *step* in the nomenclature of DNA structure. For example, in Figure 9.13, we see an AG/CT step, which tends to adopt a different structure from a GC/GC step. As more and more is learned about DNA structure, it is evident that the standard B-DNA structure, while a good model,



■ **FIGURE 9.14 Supercoiled DNA topology.** The DNA double helix can be approximated as a two-stranded, right-handed coiled rope. If one end of the rope is rotated counterclockwise, the strands begin to separate (negative supercoiling). If the rope is twisted clockwise (in a right-handed fashion), the rope becomes overwound (positive supercoiling). Get a piece of right-handed multistrand rope, and carry out these operations to convince yourself. (From *Science* Vol. 312, No. 5779, June 9, 2006, p. 1467–68. Used by permission of AAAS.)

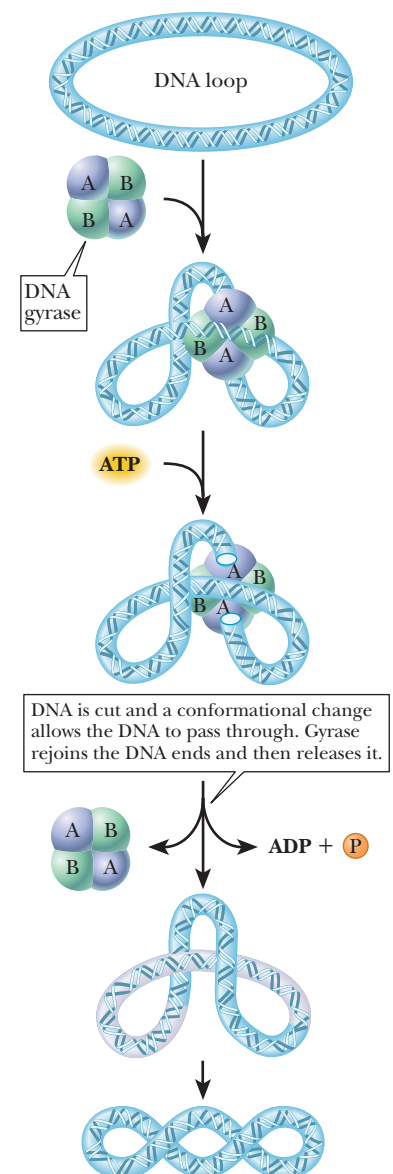
does not really describe local regions of DNA very well. Many DNA-binding proteins recognize the overall structure of a sequence of DNA, which depends upon the sequence but is not the DNA sequence itself.

The DNA molecule has a length considerably greater than its diameter; it is not completely stiff and can fold back on itself in a manner similar to that of proteins as they fold into their tertiary structures. The double helix we have discussed so far is relaxed, which means that it has no twists in it, other than the helical twists themselves. Pictures of relaxed DNA make it easier to understand the basics of DNA structure, but in reality, most DNA does have extra twists in it, which causes the DNA to form **supercoils**.

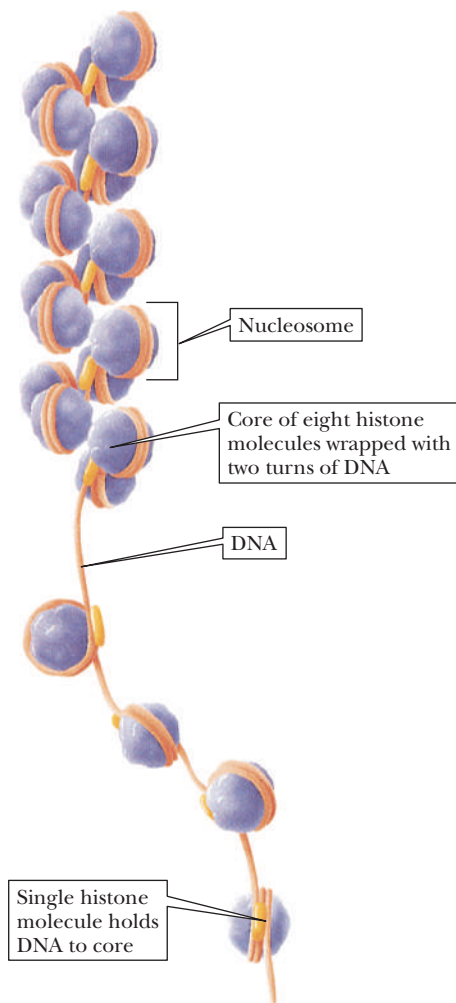
### How does prokaryotic DNA supercoil into its tertiary structure?

Prokaryotic DNA is circular, and this DNA forms supercoils. If the strands are underwound they form **negative supercoils**. If they are overwound they form **positive supercoils** (Figure 9.14). Underwound duplex DNA has fewer than the normal number of turns, whereas overwound DNA has more. DNA supercoiling is analogous twisting or untwisting a rope so that it is torsionally stressed. Negative supercoiling introduces a torsional stress that favors unwinding of the right-handed B-DNA double helix, whereas positive supercoiling overwinds such a helix. Both forms of supercoiling compact the DNA, like a rubber band wound up into a ball.

Enzymes that affect the supercoiling of DNA have been isolated from a variety of organisms. Naturally occurring circular DNA is negatively supercoiled except during replication, when it becomes positively supercoiled. Cellular regulation of this process is critical. Enzymes that are involved in changing the supercoiled state of DNA are called **topoisomerases**, and they fall into two classes. Class I topoisomerases cut the phosphodiester backbone of one strand of DNA, pass the other end through, and then reseal the backbone. Class II topoisomerases cut both strands of DNA, pass some of the remaining DNA helix between the cut ends, and then reseal. In either case, supercoils can be added or removed. As we shall see in upcoming chapters, these enzymes play an important role in replication and transcription, where separation of the helix strands causes supercoiling. **DNA gyrase** is a bacterial topoisomerase that introduces negative supercoils into DNA. The mechanism is shown in Figure 9.15. The enzyme is a tetramer. It cuts both strands of DNA, so it is a class II topoisomerase.



■ **FIGURE 9.15 A model for the action of bacterial DNA gyrase (topoisomerase II).**



■ **FIGURE 9.16** The structure of chromatin. DNA is associated with histones in an arrangement that gives the appearance of beads on a string. The “string” is DNA, and each of the “beads” (nucleosomes) consists of DNA wrapped around a protein core of eight histone molecules. Further coiling of the DNA spacer regions produces the compact form of chromatin found in the cell.

Supercoiling has been observed experimentally in naturally occurring DNA. Particularly strong evidence has come from electron micrographs that clearly show coiled structures in circular DNA from a number of different sources, including bacteria, viruses, mitochondria, and chloroplasts. Ultracentrifugation can be used to detect supercoiled DNA because it sediments more rapidly than the relaxed form. (See Section 9.5 for a discussion of ultracentrifugation.)

Scientists have known for some time that prokaryotic DNA is normally circular, but supercoiling is a relatively recent subject of research. Computer modeling has helped scientists visualize many aspects of the twisting and knotting of supercoiled DNA by obtaining “stop-action” images of very fast changes.

### How does supercoiling take place in eukaryotic DNA?

The supercoiling of the nuclear DNA of eukaryotes (such as plants and animals) is more complicated than the supercoiling of the circular DNA from prokaryotes. Eukaryotic DNA is complexed with a number of proteins, especially with basic proteins that have abundant positively charged side chains at physiological (neutral) pH. Electrostatic attraction between the negatively charged phosphate groups on the DNA and the positively charged groups on the proteins favors the formation of complexes of this sort. The resulting material is called **chromatin**. Thus, topological changes induced by supercoiling must be accommodated by the histone-protein component of chromatin.

The principal proteins in chromatin are the **histones**, of which there are five main types, called H1, H2A, H2B, H3, and H4. All these proteins contain large numbers of basic amino acid residues, such as lysine and arginine. In the chromatin structure, the DNA is tightly bound to all the types of histone except H1. The H1 protein is comparatively easy to remove from chromatin, but dissociating the other histones from the complex is more difficult. Proteins other than histones are also complexed with the DNA of eukaryotes, but they are neither as abundant nor as well studied as histones.

In electron micrographs, chromatin resembles beads on a string (Figure 9.16). This appearance reflects the molecular composition of the protein–DNA complex. Each “bead” is a **nucleosome**, consisting of DNA wrapped around a histone core. This protein core is an octamer, which includes two molecules of each type of histone but H1; the composition of the octamer is  $(\text{H2A})_2(\text{H2B})_2(\text{H3})_2(\text{H4})_2$ . The “string” portions are called *spacer regions*; they consist of DNA complexed to some H1 histone and nonhistone proteins. As the DNA coils around the histones in the nucleosome, about 150 base pairs are in contact with the proteins; the spacer region is about 30 to 50 base pairs long. Histones can be modified by acetylation, methylation, phosphorylation, and ubiquitinylation. Ubiquitin is a protein involved in the degradation of other proteins. It will be studied further in Chapter 12. Modifying histones changes their DNA and protein-binding characteristics, and how these changes affect transcription and replication is a subject of active research (Chapter 11).

The actual structure and spacing of the nucleosomes appears to be important in its own right. While most scientists focus on the actual sequence of DNA, there is evidence that the chromatin structure is important for other reasons too. Recently scientists studied the rate of genetic variation in the DNA of the Japanese killifish, *Oryzias latipes*. They found that the variation was cyclical and corresponded to a chromatin structure. However, what surprised the researchers was that the cyclic nature was different depending on what type of mutation rate they looked at. In the linker region, the rate of insertion or deletion mutations peaked. These are mutations where an extra nucleotide is added or deleted. In this same region, the rate of mutation by substitution, where one nucleotide is substituted for another, was at its low point. How these facts relate



# Biochemical Connections LAW

## Who Owns Your Genes?

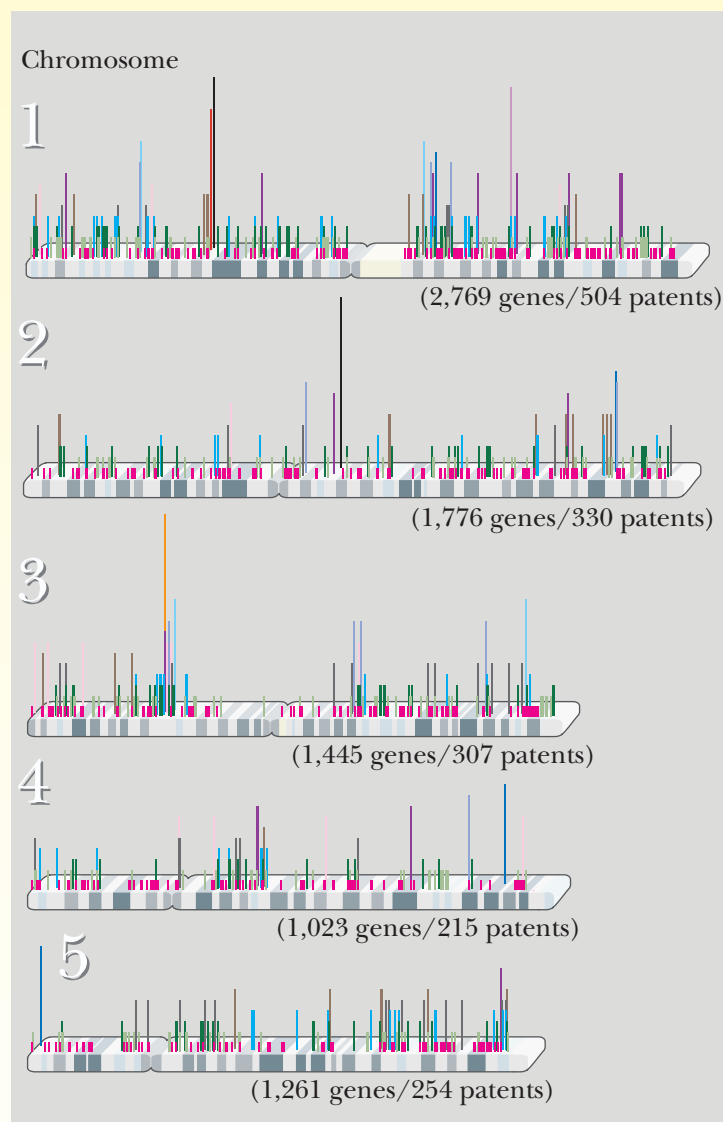
“There is a gene in your body’s cells that plays a key role in early spinal cord development. It belongs to Harvard University. Incyte Corporation, based in Wilmington, Del., has patented the gene for a receptor for histamine, the compound released by cells during the hay fever season. About half of all the genes known to be involved in cancer are patented.”\* Following the explosion in information that came from the Human Genome Project (see Biochemical Connection on page 240), commercial firms, universities, and even government agencies began to look for patents on genes, which began a long philosophical and legal battle that continues to this day. Human cells have about 24,000 genes, which are the blueprint for the 100 trillion cells in our body. About 20 percent of the human genome has been patented. As of 2006, Incyte Corporation owned about 10 percent of all known human genes.

So the question that comes to mind is, “how can a company patent a biological entity?” Well, clearly they cannot actually patent you or your genes, at least not the ones you carry around. What can be patented is purified DNA containing the sequence of the gene and techniques that allow the study of the genes. The idea of patenting information began with a landmark case in 1972 when Ananda M. Chakrabarty, a General Electric engineer, filed for a patent on a strain of *Pseudomonas* bacteria that could break down oil slicks more efficiently. He experimented with the bacteria, getting them to take up DNA from plasmids (rings of DNA; see chapter 13) that conferred the clean-up ability. The patent office rejected the patent on the grounds that products of nature and live organisms cannot be patented. However, the battle was not over, and in 1980 the Supreme Court heard the appeal in the same year that the techniques of molecular biology and recombinant DNA technology really began to take off. Chief Justice Warren Burger declared arguments against patenting life irrelevant by stating, “anything under the sun that is made by man” could be patented. The ruling was close, only 5–4 in favor of Chakrabarty, and the ramifications continue to this day. Patents have been issued for gene sequences, whole organisms such as specific bacteria, and cell types like stem cells. A patent on a clone gene or the protein it produces gives the owner exclusivity in marketing the protein, such as insulin or erythropoietin. As of 2005, the largest holder of scientific patents was the University of California, with more than 1000 patents. The U.S. government was second with 926, and the first commercial enterprise on the list, Sanofi Aventis, came in third at 587.

There are many issues stirring the controversy. Proponents for the patent system point out that it takes money to drive research. Companies will not want to invest hundreds of thousands to millions of dollars in research if they cannot get a tangible gain. Allowing them to patent a product means they can eventually recover their investment. Opponents believe a patent on what amounts to information stifles more research and even prevents the advancement of medicine. If a company holds the patent to a gene known to be involved in a disease, then others cannot study it effectively and perhaps come up with better or cheaper treatments. The latter point of view has come under intense scrutiny recently because patents on diagnostic genes inhibit both research and clinical medicine. At the heart of the conflict are patents for two genes related to breast cancer, *BRCA 1* and *BRCA 2*, both owned by Myriad Genetics, Inc., of Salt Lake City. In 2009 a group of patients, doctors, and research professionals brought a suit to invalidate those patents. They argued that the two genes

are “products of nature” and should never have been patented in the first place. The long-term effects of such a suit are important enough that the American Civil Liberties Union has joined the plaintiffs.

Opponents of gene patents claimed a big victory in March of 2010 when Federal Court Judge Robert Sweet ruled against Myriad in the suit of *BRCA 1* and *BRCA 2* stating the human genes cannot be patented. So who owns your genes? At the moment, you do.



■ This map of the chromosomes offers an indication of how often genes have been patented in the U.S. Each colored bar represents the number of patents in a given segment of a chromosome, which can contain several genes. Patents can claim multiple genes, and one gene may receive multiple patents. As a result, the number of patents indicated for each chromosome does not necessarily match the sum of the values represented by the colored bars. (Laurie Grace)

\* Stix, G. Owning the Stuff of Life. *Scientific American* (February 2006), p. 78.

# Biochemical Connections GENETICS

## The Human Genome Project: Treasure or Pandora's Box?

The Human Genome Project (HGP) was a massive attempt to sequence the entire human genome, some 3.3 billion base pairs spread over 23 pairs of chromosomes. This project, started formally in 1990, was a worldwide effort driven forward by two groups. One is a private company called Celera Genomics, and its preliminary results were published in *Science* in February 2001. The other is a publicly funded group of researchers called the International Human Genome Sequencing Consortium. Their preliminary results were published in *Nature* in February 2001. Researchers were surprised to find only about 30,000 genes in the human genome. This figure has since dwindled to 25,000. This is similar to many other eukaryotes, including some as simple as the roundworm *Caenorhabditis elegans*.

What does one do with the information? From this information, we will eventually be able to identify all human genes and to determine which sets of genes are likely to be involved in all human genetic traits, including diseases that have a genetic basis. There is an elaborate interplay of genes, so it may never be possible to say that a defect in a given gene will ensure that the individual will develop a particular disease. Nevertheless, some forms of genetic screening will certainly become a routine part of medical testing in the future. It would be beneficial, for example, if someone more susceptible to heart disease than the average person were to have this information at an early age. This person could then decide on some minor adjustments in lifestyle and diet that might make heart disease much less likely to develop.

Many people are concerned that the availability of genetic information could lead to genetic discrimination. For that reason, HGP is a rare example of a scientific project in which definite percentages of financial support and research effort have been devoted to the ethical, legal, and social implications (ELSI) of the research. The question is often posed in this form: Who has a right to know your genetic information? You? Your doctor? Your potential spouse or employer? An insurance company? These questions are not trivial, but they have not yet been answered definitively. The 1997 movie *Gattaca* depicted a society in which one's social and economic classes are established at birth based on one's genome. Many citizens have expressed concern that genetic screening would lead to a new type of prejudice and bigotry aimed against "genetically challenged" people. Many people have suggested that there is no point in screening for potentially disastrous genes if there is no meaningful therapy for the disease they may "cause." However, couples often want to know in advance if they are likely to pass on a potentially lethal disease to their children.

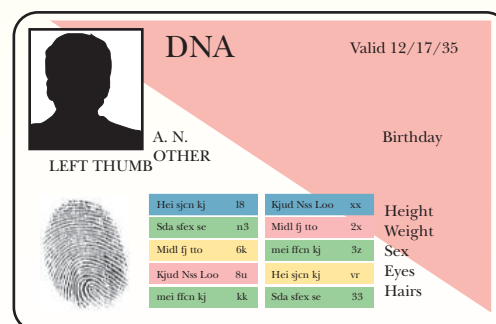
Two specific examples are pertinent here:

1. There is no advantage in testing for the breast-cancer gene if a woman is *not* in a family at high risk for the disease. The presence of a "normal" gene in such a low-risk individual tells nothing about whether a mutation might occur in the future. The risk of breast cancer is not changed if a low-risk person has the normal gene, so mammograms and monthly self-examination are in order.
2. The presence of a gene has not always predicted the development of the disease. Some individuals who have been shown to be carriers of the gene for Huntington's disease have lived to old age without developing the disease. Some males who are functionally sterile have been found to have cystic fibrosis, which carries a side effect of sterility due to the improper chloride-channel function that is a feature of that disease (see Section 13.8). They learn this when they go to a clinic to assess the nature of their fertility problem, even though they

may never have shown true symptoms of the disease as a child, other than perhaps a high occurrence of respiratory ailments.

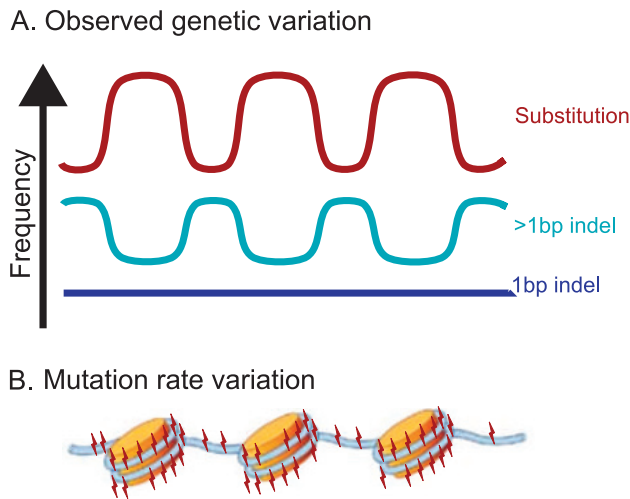
The nature of commercial enterprises coming up with genetic tests has become very controversial in the last couple of years. For example, deCODE Genetics, a company in Iceland, released a breast cancer screen that checked for seven different base changes (called single-nucleotide polymorphisms, or SNPs). The price tag is \$1625. Many scientists balked at the price, and many questioned whether the test is worth it. So much is still unknown about breast cancer that it is not clear whether knowing that a person has one or even all seven of the SNPs would be useful information, as it would indicate only a small percentage change in the chance of getting the disease.

Since the human genome project was finished a decade ago, scientists have been eager to sequence other genomes to compare with human DNA. Such studies could show conserved sequences and shed a light on those genes that are key to the survival of similar biological species as well as those that indicate changes that have occurred during the evolutionary process. Therefore, five years ago the National Human Genome Research Institute (NHGRI) assembled a list of 32 mammals and 24 other vertebrates whose DNA they would like to analyze. At an international symposium of geneticists, they determined that freezers around the world already contained DNA samples from over 16,000 different species. Shortly thereafter they launched the Genome 10K Project, which proposes to sequence 10,000 genomes in the next five years. Many feel that is wishful thinking, given that it means analyzing a genome per day. Also, to make it happen the technology has to keep improving and getting cheaper. However, those involved are confident that the genomes can all be analyzed, although when is still in question.



■ Your genome could appear on an ID card in the foreseeable future. (Fingerprint image: © Powered by Light Royalty-Free/Alamy.)





■ **FIGURE 9.17** The rate of mutation surrounding nucleosomes is not constant. Substitution mutations peak in the nucleosomes themselves, and are at a low point in the linker regions. Conversely, insertion and deletion mutations (indel in the figure) are the opposite. Researchers are attempting to establish why this is. (From Semple, C. A., and Taylor, M. S. (2009). *Molecular biology: The structure of change*. Science 323, 347–348.)

to mutation rates and natural selection is a mystery, but clearly the structure of chromatin and the spacing of the nucleosomes must be very important to both since the rate of mutation is not random and is affected by those structures.

## 9.4 Denaturation of DNA

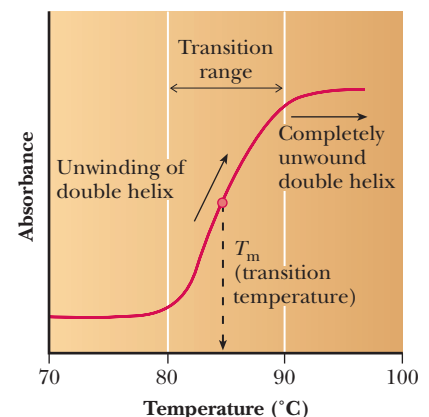
We have already seen that the hydrogen bonds between base pairs are an important factor in holding the double helix together. The amount of stabilizing energy associated with the hydrogen bonds is not great, but the hydrogen bonds hold the two polynucleotide chains in the proper alignment. However, the stacking of the bases in the native conformation of DNA contributes the largest part of the stabilization energy. Energy must be added to a sample of DNA to break the hydrogen bonds and to disrupt the stacking interactions. This is usually carried out by heating the DNA in solution.

### How can we monitor DNA denaturation?

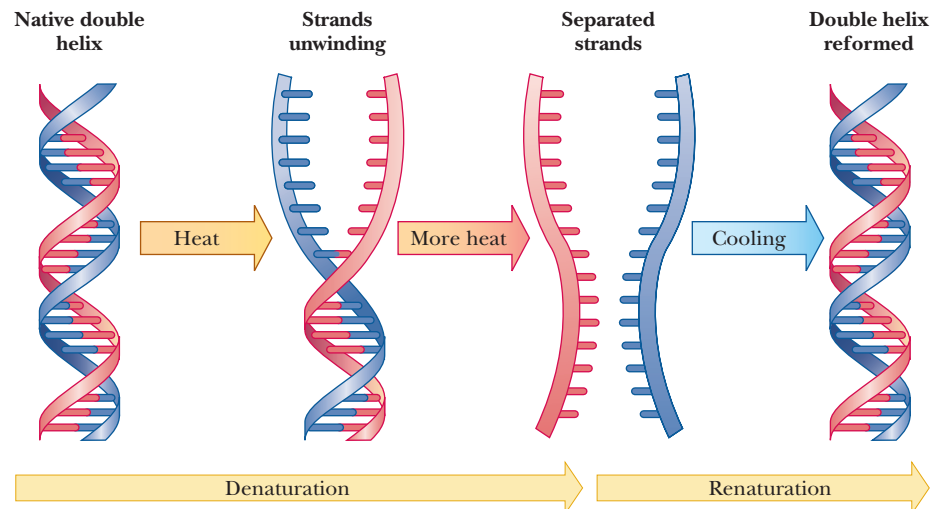
The heat denaturation of DNA, also called *melting*, can be monitored experimentally by observing the absorption of ultraviolet light. The bases absorb light in the 260-nm-wavelength region. As the DNA is heated and the strands separate, the wavelength of absorption does not change, but the amount of light absorbed increases (Figure 9.18). This effect is called *hyperchromicity*. It is based on the fact that the bases, which are stacked on top of one another in native DNA, become unstacked as the DNA is denatured.

Because the bases interact differently in the stacked and unstacked orientations, their absorbance changes. Heat denaturation is a way to obtain single-stranded DNA (Figure 9.19), which has many uses. Some of these uses are discussed in Chapter 14. When DNA is replicated, it first becomes single-stranded so that the complementary bases can be aligned. This same principle is seen during a chemical reaction used to determine the DNA sequence (Chapter 14). A most ambitious example of this reaction is described in the Biochemical Connections box about the Human Genome Project on page 240.

Under a given set of conditions, there is a characteristic midpoint of the melting curve (the transition temperature, or melting temperature, written  $T_m$ ) for DNA from each distinct source. The underlying reason for this property is that each type of DNA has a given, well-defined base composition. A G—C



■ **FIGURE 9.18** The experimental determination of DNA denaturation. This is a typical melting-curve profile of DNA, depicting the hyperchromic effect observed on heating. The transition (melting) temperature,  $T_m$ , increases as the guanine and cytosine (the G—C content) increase. The entire curve would be shifted to the right for a DNA with higher G—C content and to the left for a DNA with lower G—C content.



■ **FIGURE 9.19** Helix unwinding in DNA denaturation. The double helix unwinds when DNA is denatured, with eventual separation of the strands. The double helix is re-formed on renaturation with slow cooling and annealing.

base pair has three hydrogen bonds, and an A—T base pair has only two. The higher the percentage of G—C base pairs, the higher the melting temperature of a DNA molecule. In addition to the effect of the base pairs, G—C pairs are more hydrophobic than A—T pairs, so they stack better, which also affects the melting curve.

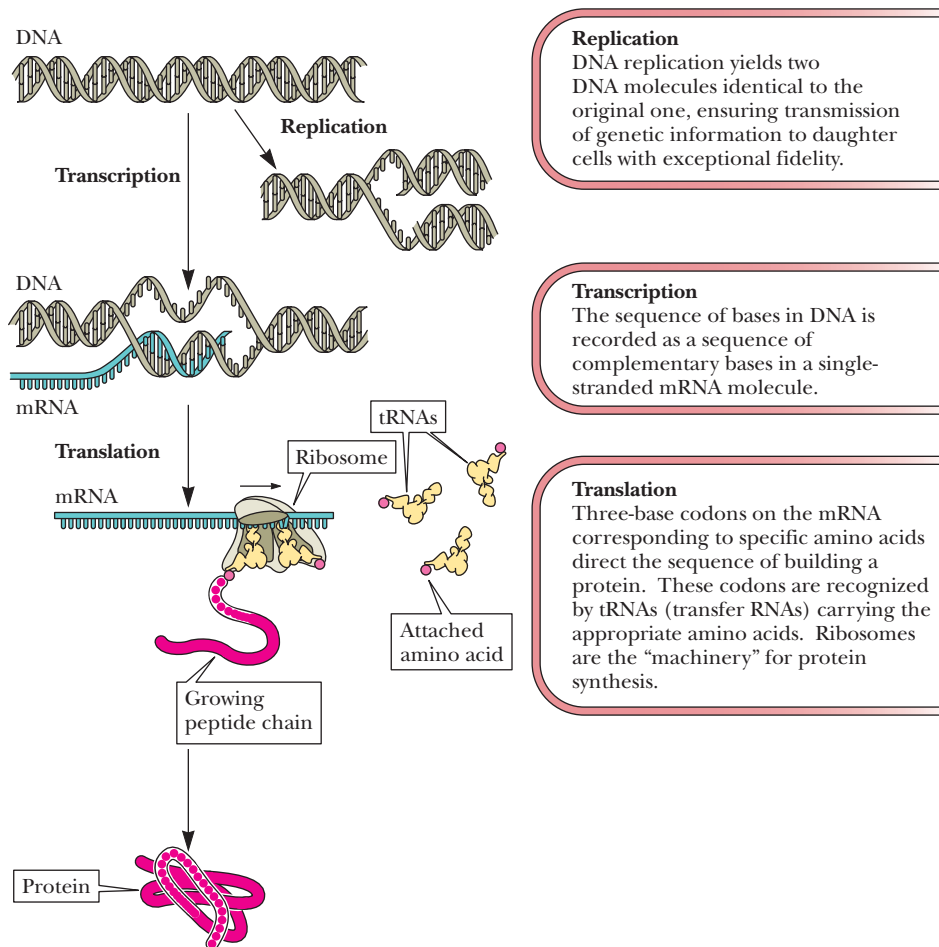
Renaturation of denatured DNA is possible on slow cooling (Figure 9.18). The separated strands can recombine and form the same base pairs responsible for maintaining the double helix.

## 9.5 The Principal Kinds of RNA and Their Structures

### *What kinds of RNA play a role in life processes?*

Six kinds of RNA—transfer RNA (tRNA), ribosomal RNA (rRNA), messenger RNA (mRNA), small nuclear RNA (snRNA), micro RNA (miRNA), and small interfering RNA (siRNA)—play an important role in the life processes of cells. Figure 9.20 shows the process of information transfer. The various kinds of RNA participate in the synthesis of proteins in a series of reactions ultimately directed by the base sequence of the cell's DNA. *The base sequences of all types of RNA are determined by that of DNA.* The process by which the order of bases is passed from DNA to RNA is called transcription (Chapter 11).

Ribosomes, in which rRNA is associated with proteins, are the sites for assembly of the growing polypeptide chain in protein synthesis. Amino acids are brought to the assembly site covalently bonded to tRNA, as aminoacyltRNAs. The order of bases in mRNA specifies the order of amino acids in the growing protein; this process is called **translation** of the genetic message. A sequence of three bases in mRNA directs the incorporation of a particular amino acid into the growing protein chain. (We shall discuss the details of protein synthesis in Chapter 12.) We are going to see that the details of the process differ in prokaryotes and in eukaryotes (Figure 9.21). In prokaryotes, there is no nuclear membrane, so mRNA can direct the synthesis of proteins while it is still being transcribed. Eukaryotic mRNA, on the other hand, undergoes considerable processing. One of the most important parts of the process is splicing out intervening sequences (introns), so that the parts of the mRNA that will be expressed (exons) are contiguous to each other.



■ **FIGURE 9.20** The fundamental process of information transfer in cells. (1) Information encoded in the nucleotide sequence of DNA is transcribed through synthesis of an RNA molecule whose sequence is dictated by the DNA sequence. (2) As the sequence of this RNA is read (as groups of three consecutive nucleotides) by the protein synthesis machinery, it is translated into the sequence of amino acids in a protein. This information transfer system is encapsulated in the dogma DNA → RNA → protein.

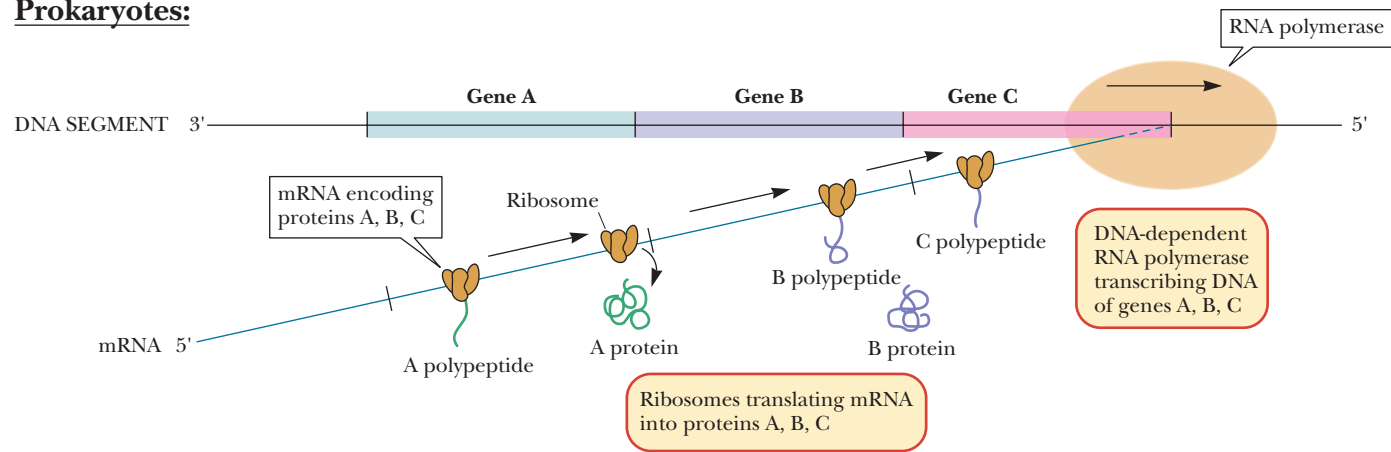
Small nuclear RNAs are found only in the nucleus of eukaryotic cells, and they are distinct from the other RNA types. They are involved in processing of initial mRNA transcription products to a mature form suitable for export from the nucleus to the cytoplasm for translation. Micro RNAs and small interfering RNAs are the most recent discoveries. SiRNAs are the main players in **RNA interference (RNAi)**, a process that was first discovered in plants and later in mammals, including humans. Micro RNAs have been found to be part of one of the oldest evolutionary relationships, that between bacteria and bacteriophages. Bacteria produce these small RNAs, which then bind to sequences of the phage DNA, preventing their infection. They have also been found to be important to repairing nerve damage in muscles (Chapter 11). RNAi is also being used extensively by scientists who wish to eliminate the effect of a gene to help discover its function (see Chapter 13). Table 9.1 summarizes the types of RNA.

### What is the role of transfer RNA in protein synthesis?

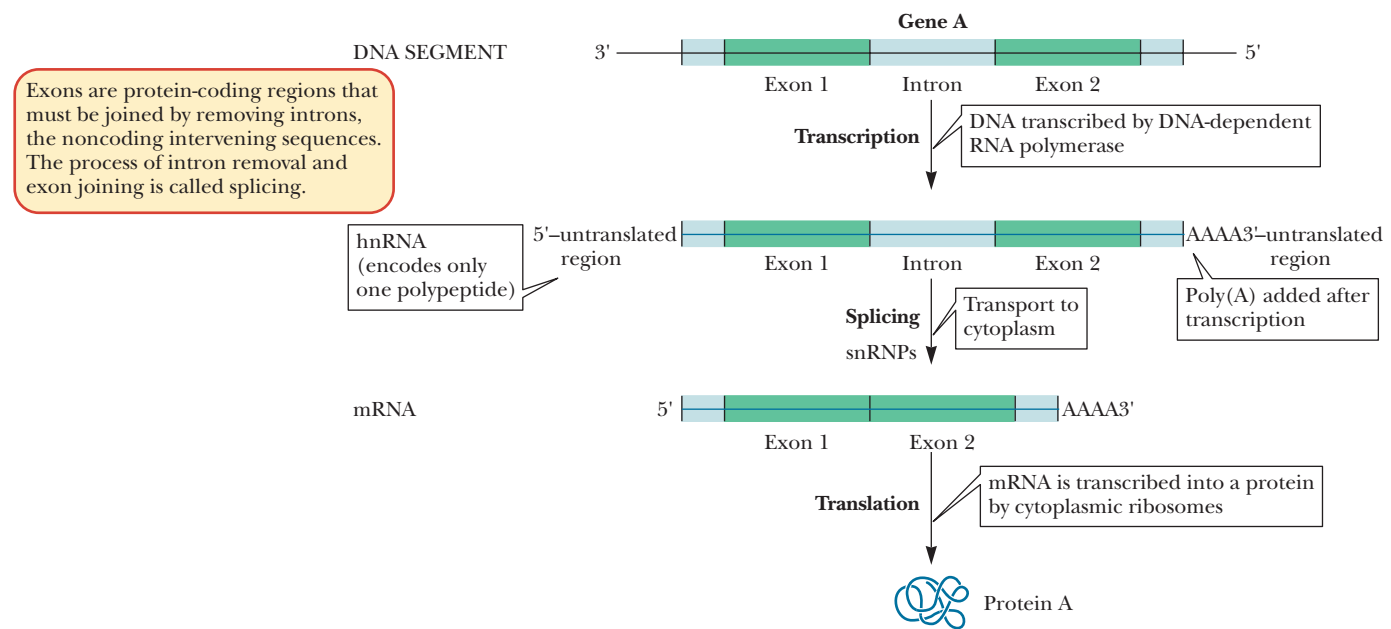
The smallest of the three important kinds of RNA is tRNA. Different types of tRNA molecules can be found in every living cell because at least one tRNA bonds specifically to each of the amino acids that commonly occur in proteins.

Frequently there are several tRNA molecules for each amino acid. A tRNA is a single-stranded polynucleotide chain, between 73 and 94 nucleotide residues

Prokaryotes:



Eukaryotes:



■ **FIGURE 9.21** The role of mRNA in transcription. The properties of mRNA molecules in prokaryotic versus eukaryotic cells during transcription and translation.

**TABLE 9.1**

The Roles of Different Kinds of RNA		
RNA Type	Size	Function
Transfer RNA	Small	Transports amino acids to site of protein synthesis
Ribosomal RNA	Several kinds—variable in size	Combines with proteins to form ribosomes, the site of protein synthesis
Messenger RNA	Variable	Directs amino acid sequence of proteins
Small nuclear RNA	Small	Processes initial mRNA to its mature form in eukaryotes
Small interfering RNA	Small	Affects gene expression; used by scientists to knock out a gene being studied
Micro RNA	Small	Affects gene expression; important in growth and development

long, that generally has a molecular mass of about 25,000 Da. (Note that biochemists tend to call the unit of atomic mass the *dalton*, abbreviated Da.)

Intrachain hydrogen bonding occurs in tRNA, forming A—U and G—C base pairs similar to those that occur in DNA except for the substitution of uracil for thymine. The duplexes thus formed have the A-helical form, rather than the B-helical form, which is the predominant form in DNA (Section 9.3). The molecule can be drawn as a *cloverleaf structure*, which can be considered the secondary structure of tRNA because it shows the hydrogen bonding between certain bases (Figure 9.22). The hydrogen-bonded portions of the molecule are called *stems*, and the non-hydrogen-bonded portions are *loops*. Some of these loops contain modified bases (Figure 9.23). During protein synthesis, both tRNA and mRNA are bound to the ribosome in a definite spatial arrangement that ultimately ensures the correct order of the amino acids in the growing polypeptide chain.

A particular tertiary structure is necessary for tRNA to interact with the enzyme that covalently attaches the amino acid to the 2' or 3' end. To produce this tertiary structure, the tRNA folds into an L-shaped conformation that has been determined by X-ray diffraction (Figure 9.24).

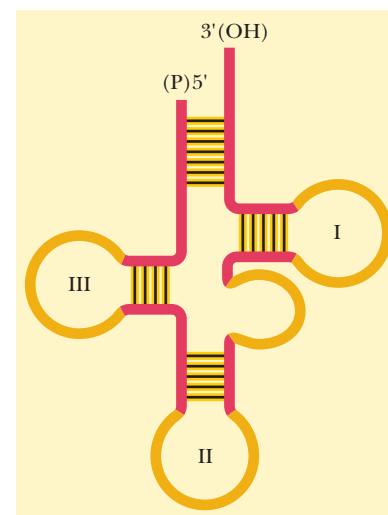
### How does ribosomal RNA combine with proteins to form the site of protein synthesis?

In contrast with tRNA, rRNA molecules tend to be quite large, and only a few types of rRNA exist in a cell. Because of the intimate association between rRNA and proteins, a useful approach to understanding the structure of rRNA is to investigate ribosomes themselves.

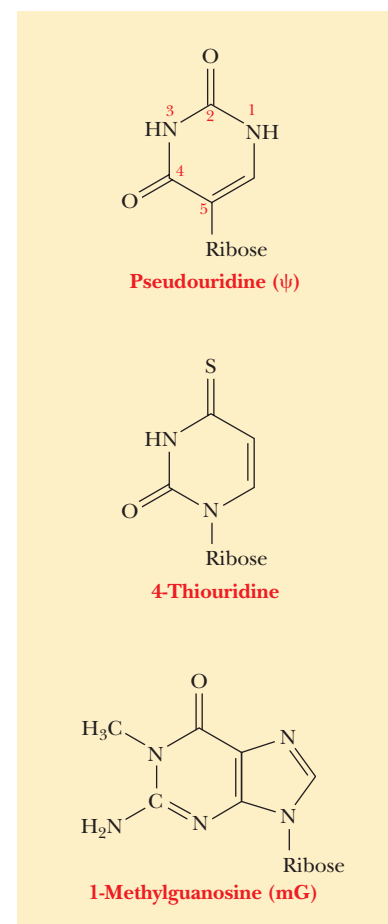
The RNA portion of a ribosome accounts for 60%–65% of the total weight, and the protein portion constitutes the remaining 35%–40% of the weight. Dissociation of ribosomes into their components has proved to be a useful way of studying their structure and properties. A particularly important endeavor has been to determine both the number and the kind of RNA and protein molecules that make up ribosomes. This approach has helped elucidate the role of ribosomes in protein synthesis. In both prokaryotes and eukaryotes, a ribosome consists of two subunits, one larger than the other. In turn, the smaller subunit consists of one large RNA molecule and about 20 different proteins; the larger subunit consists of two RNA molecules in prokaryotes (three in eukaryotes) and about 35 different proteins in prokaryotes (about 50 in eukaryotes). The subunits are easily dissociated from one another in the laboratory by lowering the  $\text{Mg}^{2+}$  concentration of the medium. Raising the  $\text{Mg}^{2+}$  concentration to its original level reverses the process, and active ribosomes can be reconstituted by this method.

A technique called *analytical ultracentrifugation* has proved very useful for monitoring the dissociation and reassociation of ribosomes. Figure 9.25 shows an analytical ultracentrifuge. We need not consider all the details of this technique, as long as it is clear that its basic aim is the observation of the motion of ribosomes, RNA, or protein in a centrifuge. The motion of the particle is characterized by a *sedimentation coefficient*, expressed in *Svedberg units* (S), which are named after Theodor Svedberg, the Swedish scientist who invented the ultracentrifuge. The S value increases with the molecular weight of the sedimenting particle, but it is not directly proportional to it because the particle's shape also affects its sedimentation rate.

Ribosomes and ribosomal RNA have been studied extensively via sedimentation coefficients. Most research on prokaryotic systems has been done with the bacterium *Escherichia coli*, which we shall use as an example here. An *E. coli* ribosome typically has a sedimentation coefficient of 70S. When an intact 70S bacterial ribosome dissociates, it produces a light 30S subunit and a heavy 50S subunit. Note that the values of sedimentation coefficients are not additive,

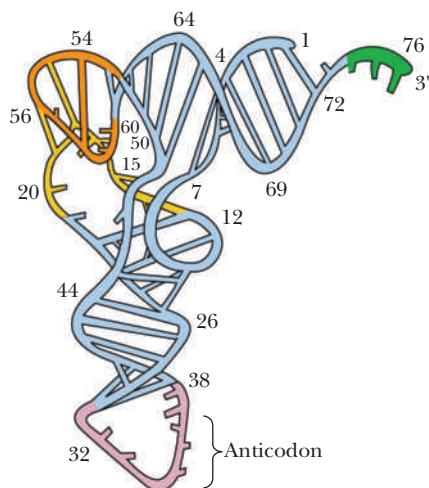


■ **FIGURE 9.22** The cloverleaf depiction of transfer RNA. Double-stranded regions (shown in red) are formed by folding the molecule and stabilized by hydrogen bonds between complementary base pairs. Peripheral loops are shown in yellow. There are three major loops (numbered) and one minor loop of variable size (not numbered).

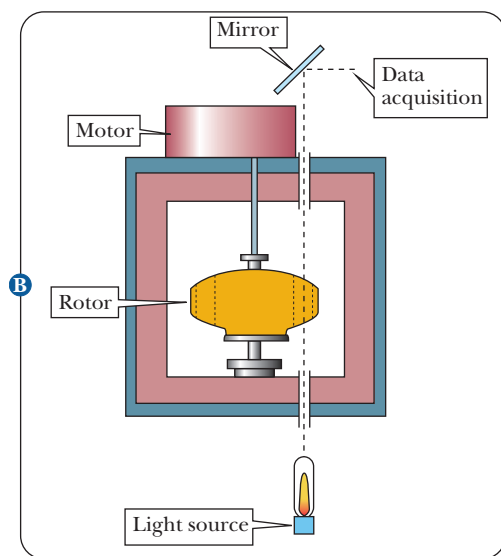
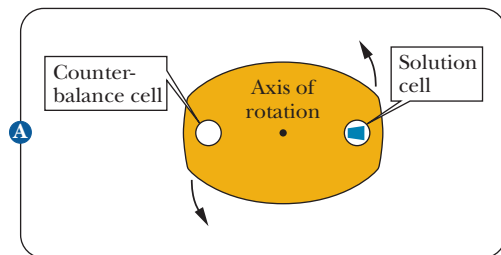


■ **FIGURE 9.23** Structures of some modified bases found in transfer RNA. Note that the pyrimidine in pseudouridine is linked to ribose at C-5 rather than at the usual N-1.





■ **FIGURE 9.24** The three-dimensional structure of yeast phenylalanine tRNA as deduced from X-ray diffraction studies of its crystals. The tertiary folding is illustrated, and the ribose-phosphate backbone is presented as a continuous ribbon; H bonds are indicated by crossbars. Unpaired bases are shown as short, unconnected rods. The anticodon loop is at the bottom and the —CCA 3'—OH acceptor end is at the top right.



■ **FIGURE 9.25** The analytical ultracentrifuge. (a) Top view of an ultracentrifuge rotor. The solution cell has optical windows; the cell passes through a light path once each revolution. (b) Side view of an ultracentrifuge rotor. The optical measurement taken as the solution cell passes through the light path makes it possible to monitor the motion of sedimenting particles.

showing the dependence of the S value on the shape of the particle. The 30S subunit contains a 16S rRNA and 21 different proteins. The 50S subunit contains a 5S rRNA, a 23S rRNA, and 34 different proteins (Figure 9.26). For comparison, eukaryotic ribosomes have a sedimentation coefficient of 80S, and the small and large subunits are 40S and 60S, respectively. The small subunit of eukaryotes contains an 18S rRNA, and the large subunit contains three types of rRNA molecules: 5S, 5.8S, and 28S.

The 5S rRNA has been isolated from many different types of bacteria, and the nucleotide sequences have been determined. A typical 5S rRNA is about 120 nucleotide residues long and has a molecular mass of about 40,000 Da. Some sequences have also been determined for the 16S and 23S rRNA molecules. These larger molecules are about 1500 and 2500 nucleotide residues long, respectively. The molecular mass of 16S rRNA is about 500,000 Da, and that of 23S rRNA is about one million Da. The degrees of secondary and tertiary structure in the larger RNA molecules appear to be substantial.

A secondary structure has been proposed for 16S rRNA (Figure 9.27), and suggestions have been made about the way in which the proteins associate with the RNA to form the 30S subunit.

The *self-assembly of ribosomes* takes place in the living cell, but the process can be duplicated in the laboratory. Elucidation of ribosomal structure is an active field of research. The binding of antibiotics to bacterial ribosomal subunits so as to prevent self-assembly of the ribosome is one focus of the investigation. The structure of ribosomes is also one of the points used to compare and contrast eukaryotes, eubacteria, and archaebacteria (Chapter 1). The study of RNA became much more exciting in 1986, when Thomas Cech showed that certain RNA molecules exhibited catalytic activity (Section 11.8). Equally exciting was the recent discovery that the ribosomal RNA, and not protein, is the part of a ribosome that catalyzes the formation of peptide bonds in bacteria (Chapter 12).

### How does messenger RNA direct protein synthesis?

The least abundant of the main types of RNA is mRNA. In most cells, it constitutes no more than 5%–10% of the total cellular RNA. The sequences of bases in mRNA specify the order of the amino acids in proteins. In rapidly growing cells, many different proteins are needed within a short time interval. Fast turnover in protein synthesis becomes essential. Consequently, it is logical that mRNA is formed when it is needed, directs the synthesis of proteins, and then is degraded so that the nucleotides can be recycled. Of the main types of RNA, mRNA is the one that usually turns over most rapidly in the cell. Both tRNA and rRNA (as well as ribosomes themselves) can be recycled intact for many rounds of protein synthesis.

The sequence of mRNA bases that directs the synthesis of a protein reflects the sequence of DNA bases in the gene that codes for that protein, although this mRNA sequence is often altered after it is produced from the DNA. Messenger RNA molecules are heterogeneous in size, as are the proteins whose sequences they specify. Less is known about possible intrachain folding in mRNA, with the exception of folding that occurs during termination of transcription (Chapter 11). It is also likely that several ribosomes are associated with a single mRNA molecule at some time during the course of protein synthesis. In eukaryotes, mRNA is initially formed as a larger precursor molecule called **heterogeneous nuclear RNA (hnRNA)**. These contain lengthy portions of intervening sequences called **introns** that do not encode a protein. These introns are removed by posttranscriptional splicing. In addition, protective units called **5'-caps** and **3' poly(A) tails** are added before the mRNA is complete (Section 11.7).

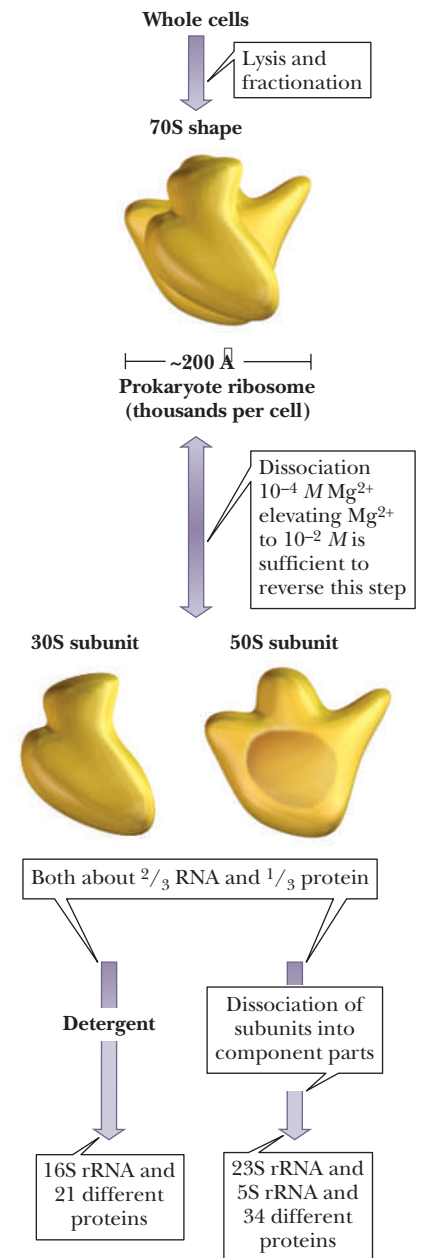
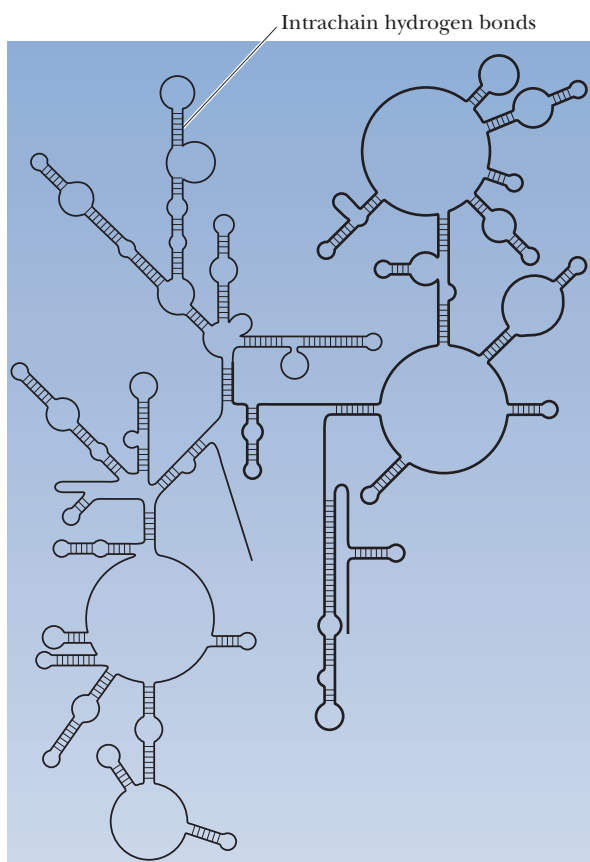
### How does small nuclear RNA help with the processing of RNA?

A recently discovered RNA molecule is the small nuclear RNA (snRNA), which is found, as the name implies, in the nucleus of eukaryotic cells. This type of RNA is small, about 100 to 200 nucleotides long, but it is not a tRNA molecule nor a small subunit of rRNA. In the cell, it is complexed with proteins forming **small nuclear ribonucleoprotein particles**, usually abbreviated **snRNPs** (pronounced “snurps”). These particles have a sedimentation coefficient of 10S. Their function is to help with the processing of the initial mRNA transcribed from DNA into a mature form that is ready for export out of the nucleus. In eukaryotes, transcription occurs in the nucleus, but because most protein synthesis occurs in the cytosol, the mRNA must first be exported. Many researchers are working on the processes of RNA splicing, which will be described further in Section 11.8.

### What is RNA interference, and why is it important?

The process called RNA interference was heralded as the breakthrough of the year in 2002 in *Science* magazine. Short stretches of RNA (20–30 nucleotides long) have been found to have an enormous control over gene expression.

This process has been found to be a protection mechanism in many species, with the siRNAs being used to eliminate expression of an undesirable gene, such as one that is causing uncontrolled cell growth or a gene that came from a virus. These small RNAs are also being used by scientists who wish to study gene expression. In what has become an explosion of new biotechnology, many



■ **FIGURE 9.26** The structure of a typical prokaryotic ribosome. The individual components can be mixed, producing functional subunits. Reassociation of subunits gives rise to an intact ribosome.

■ **FIGURE 9.27** A schematic drawing of a proposed secondary structure for 16S rRNA. The intrachain folding pattern includes loops and double-stranded regions. Note the extensive intrachain hydrogen bonding.

# Biochemical Connections GENETICS

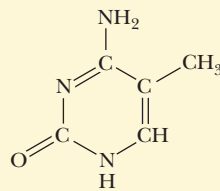
## Why Identical Twins Are Not Identical

A lot has been learned about the differences between nature and nurture by studying twins. Often twins separated at birth are later studied to see how different they have become. Their differences and similarities give us insight into how much of our physiology and our behavior is governed by genetics. Sometimes, however, twins raised together under seemingly identical circumstances can turn out very different. While the actual DNA sequence of identical twins is the same, two twins can be very different in other ways. The study of **epigenetics** is an active area of research. Epigenetics refers to changes in DNA that are not reflected in the actual base sequence. Epigenetic modifications of DNA act like switches that turn on or off certain genes. If these modifications are not the same in each of two twins, then the twins will no longer be identical.

The best known example of an epigenetic mechanism is DNA methylation, where a cytosine is tagged with a methyl group, as shown in the figure. This is usually associated with shutting off the expression of the gene. Another epigenetic mechanism is

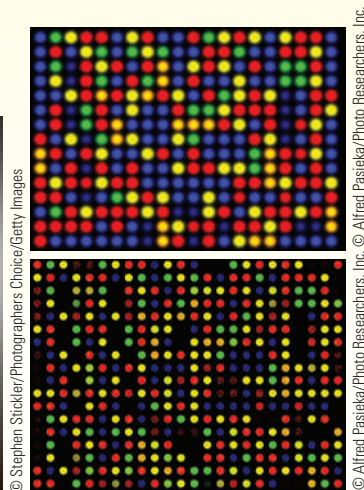
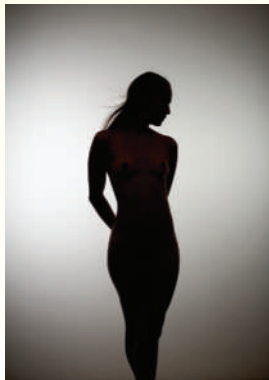
chromatin remodeling. The histone proteins discussed in Section 9.3 can be modified by the addition of methyl, acetyl, or phosphate groups. This, in turn, influences the activity of adjacent genes. Acetylation generally switches on the expression of the genes, while methylation usually silences the expression, as shown in the figure.

Because certain disease states can be linked to epigenetic states, it is possible for one individual to develop a disease while his or her identical twin does not. Susceptibility to diseases is often a family trait, but the actual mechanism of getting the disease may require epigenetic changes in the DNA of a cell. Epigenetic changes are known to be important in the field of cancer research, but only recently have scientists begun to study the relationship between epigenetic state and other diseases, such as schizophrenia, immune deficiencies, obesity, diabetes, and heart disease.

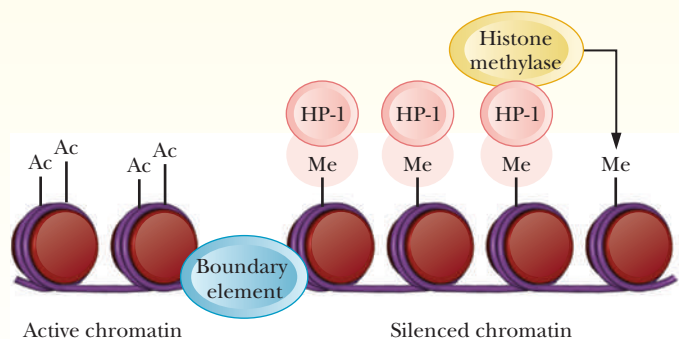


5-Methylcytosine

■ Structure of 5-methylcytosine.



■ Sophisticated microarrays coded to show epigenetic differences could show why identical twins are not the same. (Reprinted with permission from Nature.)



■ Methylation and acetylation of histones regulates expression of genes. From Bannister, A., Zegerman, P., Partridge, J.F., Miska E., Thomas, J.O., Allshire, R.C., Kouzarides, T.: Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410**, 120–124 (1 March 2001).

companies have been created to produce and to market designer siRNA to knock out hundreds of known genes. This technology also has medical applications: siRNA has been used to protect mouse liver from hepatitis and to help clear infected liver cells of the disease. The biotech applications of RNA interference will be discussed more in Chapter 13. For the moment, we can say that many new biotechnology companies have sprung up in recent years to exploit possible applications of RNA interference.



# Biochemical Connections GENOMICS

## Synthetic Genome Created

For those researchers interested in studying the nature of life and its relationship to DNA chemistry, the Holy Grail of experiments is to demonstrate that synthetic DNA can lead to life. In May of 2010, science took a giant step towards that goal. The laboratory of Human Genome Project pioneer, J. Craig Venter, successfully designed synthetic DNA and used it to drive the reproduction of a bacterial species.

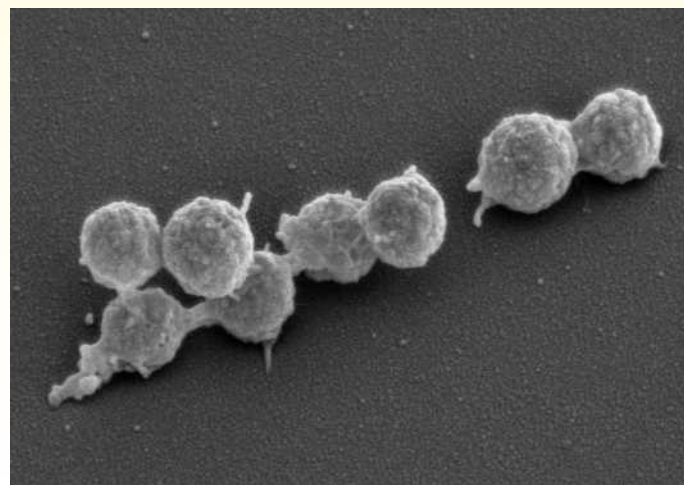
The experiment was completed in several stages, eventually taking 40 million dollars and a team of over 20 researchers a decade to complete. It began when Ventner and two colleagues, Clyde Hutchinson and Hamilton Smith, demonstrated they could transplant the DNA from one bacterial species to another. In 2008 they created an artificial chromosome of the bacterium *Mycoplasma genitalium*, which they chose as it has the smallest genome of a free-living organism, with only 600,000 bases. Besides the natural DNA sequence, their synthetic version contained constructed “watermark” sequences that allowed them to tell the synthetic version from the natural one. Unfortunately, the *M. genitalium* bacteria reproduced too slowly to be studied efficiently, so they switched species to the faster growing *Mycoplasma mycoides*, which contains 1 million bases.

In 2009, they demonstrated that they could transplant natural *M. mycoides* DNA into a close cousin, *M. capricolum*. They culminated the experiment in 2010 by taking the synthetic version of the *M. mycoides* DNA and transplanting it into *M. capricolum* cells that had had their DNA removed. The watermark sequences allow for the growth of blue bacterial colonies to show that any colonies growing have the synthetic instead of the natural DNA (see figure).

Some scientists have called their experiment, “life recreated” because it did not quite demonstrate creation of life from chemistry because they transplanted DNA into cells that were previ-

ously alive. However, this experiment is a keystone demonstration of the importance of DNA to all life processes, since the chemically synthesized DNA was able to take over the evacuated *M. capricolum* cells and begin to grow colonies of *M. mycoides*.

While it will be years or decades before scientists can begin to create designer organisms, the potential to create microbes that can synthesize pharmaceuticals or fuels has molecular biologists excited to see what organisms come out of Ventner’s and other researchers’ labs in the future.



T. Deernick and M. Ellisman/NCMIR, University of California, San Diego

■ Life re-created. SEM of bacterial group.

## SUMMARY

**How do DNA and RNA differ?** DNA (deoxyribonucleic acid) and RNA (ribonucleic acid) are the two kinds of nucleic acids. DNA contains the sugar deoxyribose, but RNA has ribose in the same position. The difference in the sugars gives rise to differences in their secondary and tertiary structures. The primary structure of nucleic acids is the order of bases in the polynucleotide sequence, and the secondary structure is the three-dimensional conformation of the backbone. The tertiary structure is specifically the supercoiling of the molecule.

**What are the structures and components of the nucleotides?** The monomers of nucleic acids are nucleotides. An individual nucleotide consists of three parts—a nitrogenous base, a sugar, and a phosphoric acid residue—all of which are covalently bonded together. The bases are bonded to the sugars, forming nucleosides.

**How do nucleotides combine to give nucleic acids?** Nucleosides are linked by ester bonds to phosphoric acid to form the phosphodiester backbone.

**What is the nature of the DNA double helix?** The double helix originally proposed by Watson and Crick is the most striking feature of DNA structure. The two coiled strands run in antiparallel directions with hydrogen bonds between complementary bases. Adenine pairs with thymine, and guanine pairs with cytosine.

**Are there other possible conformations of the double helix?** Some variations on the usual representation of the double helix (B-DNA) are known to exist. In A-DNA, the base pairs lie at an angle to the helix axis, and in Z-DNA the helix is left handed, rather than the more usual right-handed form of B-DNA. These variant forms are known to have physiological roles.

**How does prokaryotic DNA supercoil into its tertiary structure?** Supercoiling is a feature of DNA structure both in prokaryotes and in eukaryotes. Prokaryotic DNA is normally circular and is twisted into a supercoiled form before the circle is closed. The form of supercoiling plays a role in DNA replication.

**How does supercoiling take place in eukaryotic DNA?** Eukaryotic DNA is complexed with histones and other basic proteins, but less is known about proteins bound to prokaryotic DNA.

**How can we monitor DNA denaturation?** When DNA is denatured, the double-helical structure breaks down; the progress of this phenomenon can be followed by monitoring the absorption of ultraviolet light. The temperature at which DNA becomes denatured by heat depends on its base composition; higher temperatures are needed to denature DNA rich in G—C base pairs.

**What kinds of RNA play a role in life processes?** The six kinds of RNA—transfer RNA (tRNA), ribosomal RNA (rRNA), messenger RNA (mRNA), small nuclear RNA (snRNA), micro RNA (miRNA), and small interfering RNA (siRNA)—differ in structure and function.

**What is the role of transfer RNA in protein synthesis?** Transfer RNA is relatively small, about 80 nucleotides long. It exhibits

extensive intrachain hydrogen bonding, represented in two dimensions by a cloverleaf structure. Amino acids are brought to the site of protein synthesis bonded to transfer RNAs.

**How does ribosomal RNA combine with proteins to form the site of protein synthesis?** Ribosomal RNA molecules tend to be quite large and are complexed with proteins to form ribosomal subunits. Ribosomal RNA also exhibits extensive internal hydrogen bonding.

**How does messenger RNA direct protein synthesis?** The sequence of bases in a given mRNA determines the sequence of amino acids in a specified protein. The size of mRNA molecules varies with the size of the protein.

**How does small nuclear RNA help with the processing of RNA?** Eukaryotic mRNA is processed in the nucleus by a fourth type of RNA, small nuclear RNA, which is complexed with proteins to give small nuclear ribonuclear protein particles (snRNPs). Eukaryotic mRNA is initially produced in an immature form that must be processed by removing introns and adding protective units at the 5' and 3' ends.

**What is RNA interference, and why is it important?** Micro RNA and small interfering RNA are both very small, about 20–30 bases long. They function in the control of gene expression and were the most recent discoveries in RNA research.

## REVIEW EXERCISES

**OWL** Interactive versions of these problems are assignable in OWL

### 9.1 Levels of Structure in Nucleic Acids

- 1. Reflect and Apply** Consider the following in light of the concept of levels of structure (primary, secondary, tertiary, quaternary) as defined for proteins.
  - (a) What level is shown by double-stranded DNA?
  - (b) What level is shown by tRNA?
  - (c) What level is shown by mRNA?

### 9.2 The Covalent Structure of Polynucleotides

- 2. Recall** What is the structural difference between thymine and uracil?
- 3. Recall** What is the structural difference between adenine and hypoxanthine?
- 4. Recall** Give the name of the base, the ribonucleoside or deoxyribonucleoside, and the ribonucleoside triphosphate for A, G, C, T, and U.
- 5. Recall** What is the difference between ATP and dATP?
- 6. Recall** Give the sequence on the opposite strand for ACGTAT, AGATCT, and ATGGTA (all read 5' → 3').
- 7. Recall** Are the sequences shown in Question 6 those of RNA or DNA? How can you tell?
- 8. Reflect and Apply** (a) Is it biologically advantageous that DNA is stable? Why or why not? (b) Is it biologically advantageous that RNA is unstable? Why or why not?

- 9. Reflect and Apply** A friend tells you that only four different kinds of bases are found in RNA. What would you say in reply?
- 10. Reflect and Apply** In the early days of molecular biology, some researchers speculated that RNA, but not DNA, might have a branched rather than linear covalent structure. Why might this speculation have come about?
- 11. Reflect and Apply** Why is RNA more vulnerable to alkaline hydrolysis than DNA?

### 9.3 The Structure of DNA

- 12. Recall** In what naturally occurring nucleic acids would you expect to find A-form helices, B-form helices, Z-form helices, nucleosomes, and circular DNA?
- 13. Recall** Draw a G—C base pair. Draw an A—T base pair.
- 14. Recall** Which of the following statements is (are) true?
  - (a) Bacterial ribosomes consist of 40S and 60S subunits.
  - (b) Prokaryotic DNA is normally complexed with histones.
  - (c) Prokaryotic DNA normally exists as a closed circle.
  - (d) Circular DNA is supercoiled.
- 15. Biochemical Connections** What are the two principal opposing views regarding the patenting of genes?
- 16. Biochemical Connections** Describe the landmark case that set the stage for the biotech patent battles of today.



17. **Biochemical Connections** What two genes are at the heart of a current lawsuit and what is their importance?
18. **Recall** How do the major and minor grooves in B-DNA compare to those in A-DNA?
19. **Recall** Which of the following statements is (are) true?
  - (a) The two strands of DNA run parallel from their 5' to their 3' ends.
  - (b) An adenine–thymine base pair contains three hydrogen bonds.
  - (c) Positively charged counterions are associated with DNA.
  - (d) DNA base pairs are always perpendicular to the helix axis.
20. **Recall** Define *supercoiling*, *positive supercoil*, *topoisomerase*, and *negative supercoil*.
21. **Recall** What is propeller-twist?
22. **Recall** What is an AG/CT step?
23. **Recall** Why does propeller-twist occur?
24. **Recall** What is the difference between B-DNA and Z-DNA?
25. **Recall** If circular B-DNA is positively supercoiled, will these supercoils be left- or right-handed?
26. **Recall** Briefly describe the structure of chromatin.
27. **Biochemical Connections** What is the motive behind Genome Project 10K?
28. **Reflect and Apply** List three mechanisms that relax the twisting stress in helical DNA molecules.
29. **Reflect and Apply** Explain how DNA gyrase works.
30. **Reflect and Apply** Explain, and draw a diagram to show, how acetylation or phosphorylation could change the binding affinity between DNA and histones.
31. **Reflect and Apply** Would you expect to find adenine–guanine or cytosine–thymine base pairs in DNA? Why?
32. **Reflect and Apply** One of the original structures proposed for DNA had all the phosphate groups positioned at the center of a long fiber. Give a reason why this proposal was rejected.
33. **Reflect and Apply** What is the complete base composition of a double-stranded eukaryotic DNA that contains 22% guanine?
34. **Reflect and Apply** Why was it necessary to specify that the DNA in Question 33 is double-stranded?
35. **Reflect and Apply** What would be the most obvious characteristic of the base distribution of a single-stranded DNA molecule?
36. **Biochemical Connections** What is the purpose of the Human Genome Project? Why do researchers want to know the details of the human genome?
37. **Biochemical Connections** Explain the legal and ethical considerations involved in human gene therapy.
38. **Biochemical Connections** A recent commercial for a biomedical company talked about a future in which every individual would have a card that told his or her complete genotype. What would be some advantages and disadvantages of this?
39. **Reflect and Apply** A technology called PCR is used for replicating large quantities of DNA in forensic science (Chapter 13). With this technique, DNA is separated by heating with an automated system. Why is information about the DNA sequence needed to use this technique?

#### 9.4 Denaturation of DNA

40. **Reflect and Apply** Why does DNA with a high A—T content have a lower transition temperature,  $T_m$ , than DNA with a high G—C content?

#### 9.5 The Principal Kinds of RNA and Their Structures

41. **Recall** Sketch a typical cloverleaf structure for transfer RNA. Point out any similarities between the cloverleaf pattern and the proposed structures of ribosomal RNA.
42. **Recall** What is the purpose of small nuclear RNA? What is an snRNP?
43. **Recall** Which type of RNA is the biggest? Which is the smallest?
44. **Recall** Which type of RNA has the least amount of secondary structure?
45. **Recall** Why does the absorbance increase when a DNA sample unwinds?
46. **Recall** What is RNA interference?
47. **Reflect and Apply** Would you expect tRNA or mRNA to be more extensively hydrogen bonded? Why?
48. **Reflect and Apply** The structures of tRNAs contain several unusual bases in addition to the typical four. Suggest a function for the unusual bases.
49. **Reflect and Apply** Would you expect mRNA or rRNA to be degraded more quickly in the cell? Why?
50. **Reflect and Apply** Which would be more harmful to a cell, a mutation in DNA or a transcription mistake that leads to an incorrect mRNA? Why?
51. **Reflect and Apply** Explain briefly what happens to eukaryotic mRNA before it can be translated to protein.
52. **Reflect and Apply** Explain why a 50S ribosomal subunit and a 30S ribosomal subunit combine to form a 70S subunit, instead of an 80S subunit.

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# Biosynthesis of Nucleic Acids: Replication

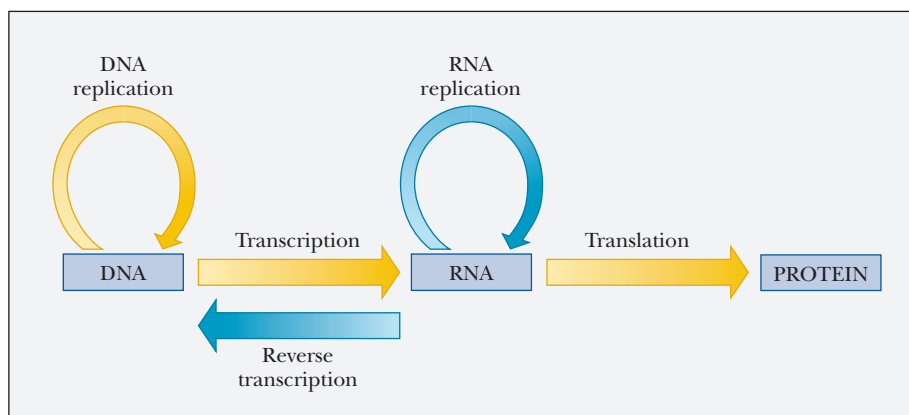
# 10

CNR/Photo Researchers, Inc.

Prokaryotic cells divide by pinching in two.

## 10.1 The Flow of Genetic Information in the Cell

The sequence of bases in DNA encodes genetic information. The duplication of DNA, giving rise to a new DNA molecule with the same base sequence as the original, is necessary whenever a cell divides to produce daughter cells. This duplication process is called **replication**. The actual formation of gene products requires RNA; the production of RNA on a DNA template is called **transcription**, which will be studied in Chapter 11. The base sequence of DNA is reflected in the base sequence of RNA. Three kinds of RNA are involved in the biosynthesis of proteins. Of the three, messenger RNA (mRNA) is particularly important. A sequence of three bases in mRNA specifies the identity of one amino acid in a manner directed by the genetic code. The process by which the base sequence directs the amino acid sequence is called **translation**, which will be studied in Chapter 12. In nearly all organisms, the flow of genetic information is DNA → RNA → protein. The only major exceptions are some viruses (called retroviruses) in which RNA, rather than DNA, is the genetic material. In those viruses, RNA can direct its own synthesis as well as that of DNA. The enzyme *reverse transcriptase* catalyzes this process. (Not all viruses in which RNA is the genetic material are retroviruses, but all retroviruses have a reverse transcriptase. In fact, that is the origin of the term *retrovirus*, referring to the reverse of the usual situation with transcription. In cases of infection by retroviruses, such as HIV, reverse transcriptase is a target for drug design. Figure 10.1 shows ways in which information is transferred in the cell. This scheme has been called the “Central Dogma” of molecular biology.



■ **FIGURE 10.1** Mechanisms for transfer of information in the cell. The yellow arrows represent general cases, and the blue arrows represent special cases (mostly in RNA viruses).

## Chapter Outline

### 10.1 The Flow of Genetic Information in the Cell

#### 10.2 Replication of DNA

- How did scientists figure out that replication is semiconservative?
- In which direction does replication go?

#### 10.3 DNA Polymerase

- How can replication proceed along the DNA if the two strands are going in opposite directions?

#### 10.4 Proteins Required for DNA Replication

- How does replication work with supercoiled DNA?
- How is single-stranded DNA protected long enough for replication?
- Where does the primer come from?

#### 10.5 Proofreading and Repair

- How does proofreading improve replication fidelity?

#### 10.6 DNA Recombination

#### 10.7 Eukaryotic DNA Replication

- How is replication tied to cell division?
- How are eukaryotic polymerases similar to prokaryotic ones?

**OWL** Online homework for this chapter may be assigned in OWL.

## 10.2 Replication of DNA

Naturally occurring DNA exists in many forms. Single- and double-stranded DNAs are known, and both can exist in linear and circular forms. As a result, it is difficult to generalize about all possible cases of DNA replication. Since many DNAs are double-stranded, we can present some general features of the replication of double-stranded DNA, features that apply both to linear and to circular DNA. Most of the details of the process that we shall discuss here were first investigated in prokaryotes, particularly in the bacterium *Escherichia coli*. We shall use information obtained by experiments on this organism for most of our discussion of the topic. Section 10.7 will discuss differences between prokaryotic and eukaryotic replication.

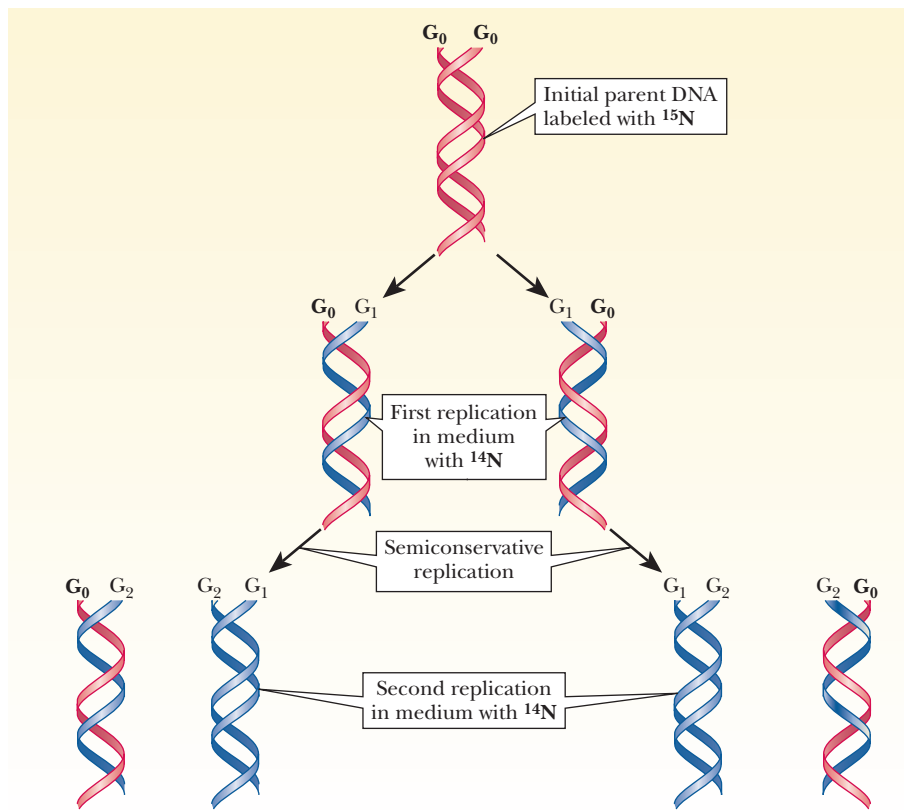
The process by which one double-helical DNA molecule is duplicated to produce two such double-stranded molecules is complex. The very complexity allows for a high degree of fine-tuning which, in turn, ensures considerable fidelity in replication. The cell faces three important challenges in carrying out the necessary steps. The first challenge is *separating the two DNA strands*. The two strands of DNA are wound around each other in such a way that they must be unwound if they are to be separated. In addition to achieving continuous unwinding of the double helix, the cell also must protect the unwound portions of DNA from the action of **nucleases** that preferentially attack single-stranded DNA. The second task involves *synthesizing of DNA from the 5' to the 3' end*. Two antiparallel strands must be synthesized in the same direction on antiparallel templates. In other words, the template has one 5' → 3' strand and one 3' → 5' strand, as does the newly synthesized DNA. The third task is *guarding against errors in replication*, ensuring that the correct base is added to the growing polynucleotide chain. Finding the answers to these challenges requires an understanding of the material in this section and the three following sections.

### Semiconservative Replication

DNA replication involves separation of the two original strands and production of two new strands with the original strands as templates. Each new DNA molecule contains one strand from the original DNA and one newly synthesized strand. This situation is called **semiconservative replication** (Figure 10.2). The details of the process differ in prokaryotes and eukaryotes, but the semiconservative nature of replication is observed in all organisms.

### How did scientists figure out that replication is semiconservative?

Semiconservative replication of DNA was established unequivocally in the late 1950s by experiments performed by Matthew Meselson and Franklin Stahl. *E. coli* bacteria were grown with  $^{15}\text{NH}_4\text{Cl}$  as the sole nitrogen source,  $^{15}\text{N}$  being a heavy isotope of nitrogen. (The usual isotope of nitrogen is  $^{14}\text{N}$ .) In such a medium, all newly formed nitrogen compounds, including purine and pyrimidine nucleobases, become labeled with  $^{15}\text{N}$ . The  $^{15}\text{N}$ -labeled DNA has a higher density than unlabeled DNA, which contains the usual isotope,  $^{14}\text{N}$ . In this experiment, the  $^{15}\text{N}$ -labeled cells were then transferred to a medium that contained only  $^{14}\text{N}$ . The cells continued to grow in the new medium. With every new generation of growth, a sample of DNA was extracted and analyzed by the technique of **density-gradient centrifugation** (Figure 10.3). This technique depends on the fact that heavy  $^{15}\text{N}$  DNA (DNA that contains  $^{15}\text{N}$  alone) forms a band at the bottom of the tube; light  $^{14}\text{N}$  DNA (containing  $^{14}\text{N}$  alone) appears at the top of the tube. DNA containing a 50–50 mixture of  $^{14}\text{N}$  and  $^{15}\text{N}$  appears at a position halfway between the two bands. In the actual experiment, this



■ **FIGURE 10.2** The labeling pattern of  $^{15}\text{N}$  strands in semiconservative replication. (G<sub>0</sub> indicates original strands; G<sub>1</sub> indicates new strands after the first generation; G<sub>2</sub> indicates new strands after the second generation.)

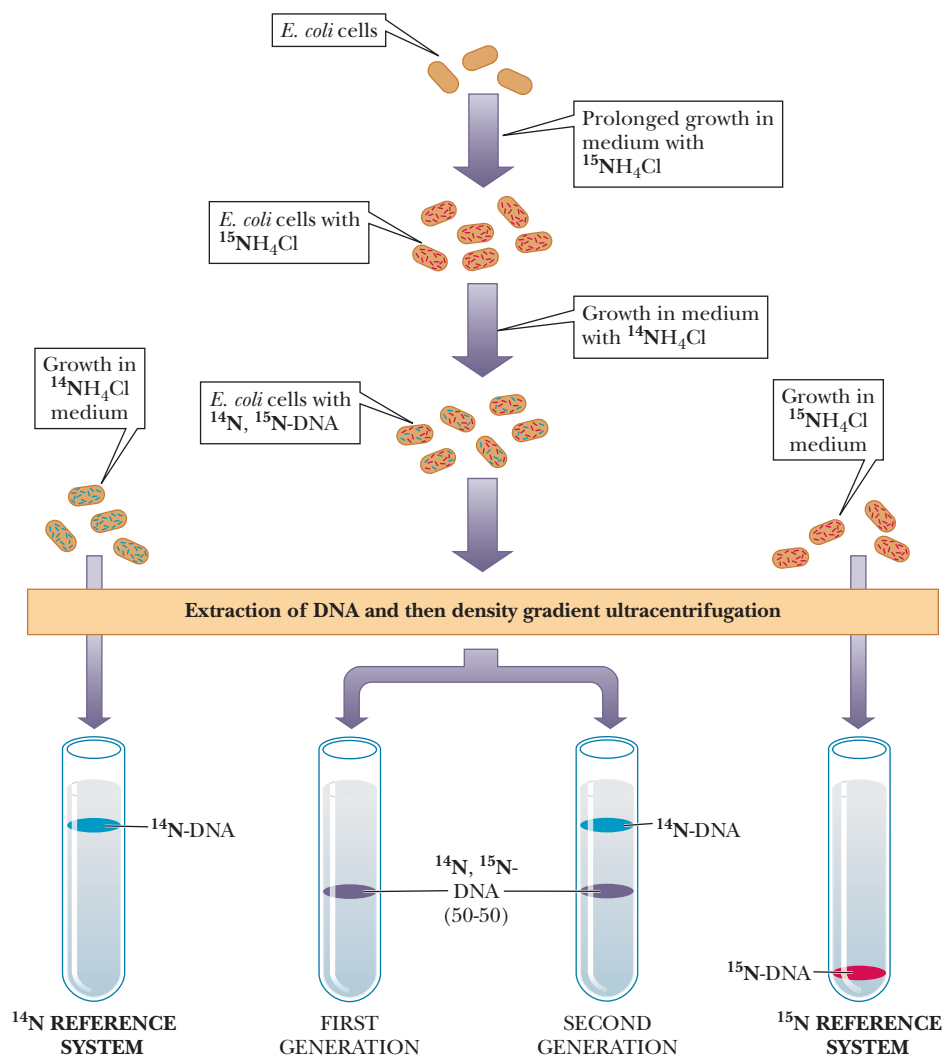
50–50 hybrid DNA was observed after one generation, a result to be expected with semiconservative replication. After two generations in the lighter medium, half of the DNA in the cells should be the 50–50 hybrid and half should be the lighter  $^{14}\text{N}$  DNA. This prediction of the kind and amount of DNA that should be observed was confirmed by the experiment.

### In which direction does replication go?

During replication, the DNA double helix unwinds at a specific point called the **origin of replication** (OriC in *E. coli*). New polynucleotide chains are synthesized using each of the exposed strands as a template. Two possibilities exist for the growth of the new strands: synthesis can take place in both directions from the origin of replication, or in one direction only. It has been established that DNA synthesis is bidirectional in most organisms, with the exception of a few viruses and plasmids. (Plasmids are rings of DNA that are found in bacteria and that replicate independently from the regular bacterial genome. They are discussed in Section 13.3.) For each origin of replication, there are two points (**replication forks**) at which new polynucleotide chains are formed. A “bubble” (also called an “eye”) of newly synthesized DNA between regions of the original DNA is a manifestation of the advance of the two replication forks in opposite directions. This feature is also called a  $\theta$  structure because of its resemblance to the lowercase Greek letter theta.

One such bubble (and one origin of replication) exists in the circular DNA of prokaryotes (Figure 10.4a). In eukaryotes, several origins of replication, and thus several bubbles, exist (Figure 10.4b). The bubbles grow larger and eventually merge, giving rise to two complete daughter DNAs. This bidirectional growth of both new polynucleotide chains represents *net chain growth*. Both new polynucleotide chains are synthesized in the 5' to 3' direction.



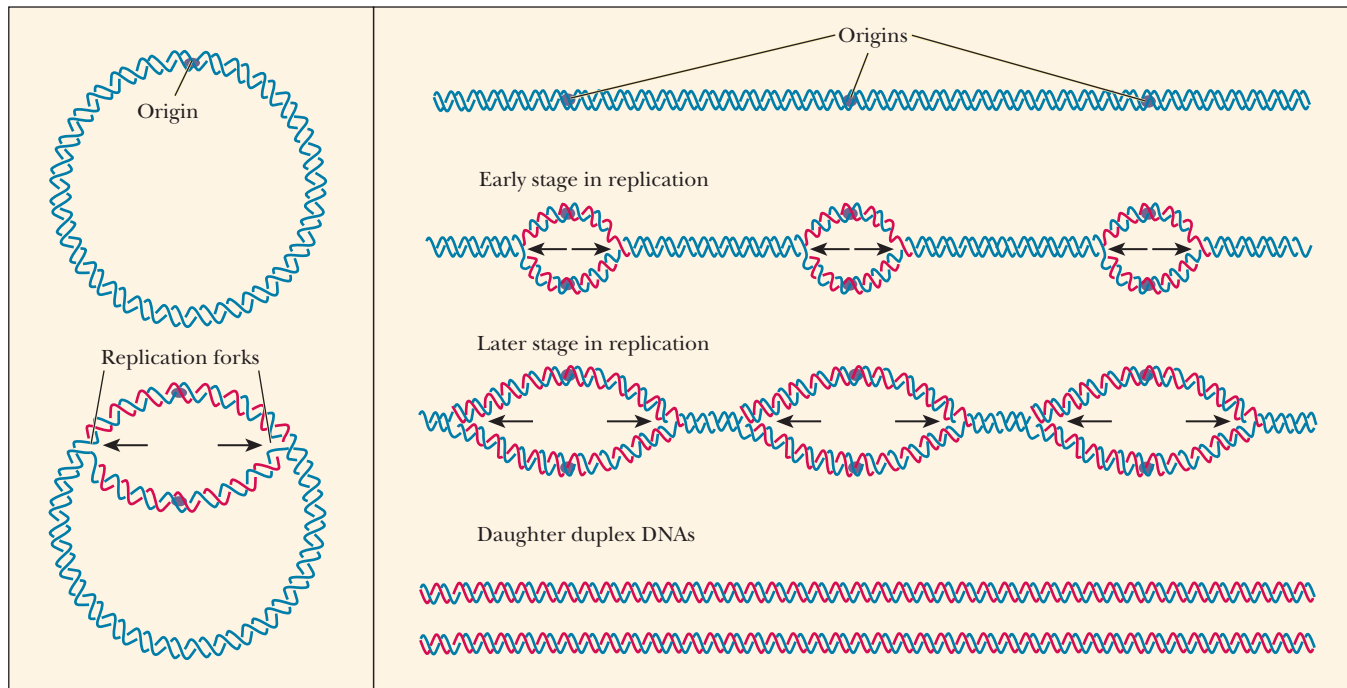


■ **FIGURE 10.3** The experimental evidence for semiconservative replication. Heavy DNA labeled with  $^{15}\text{N}$  forms a band at the bottom of the tube, and light DNA with  $^{14}\text{N}$  forms a band at the top. DNA that forms a band at an intermediate position has one heavy strand and one light strand.

## 10.3 DNA Polymerase

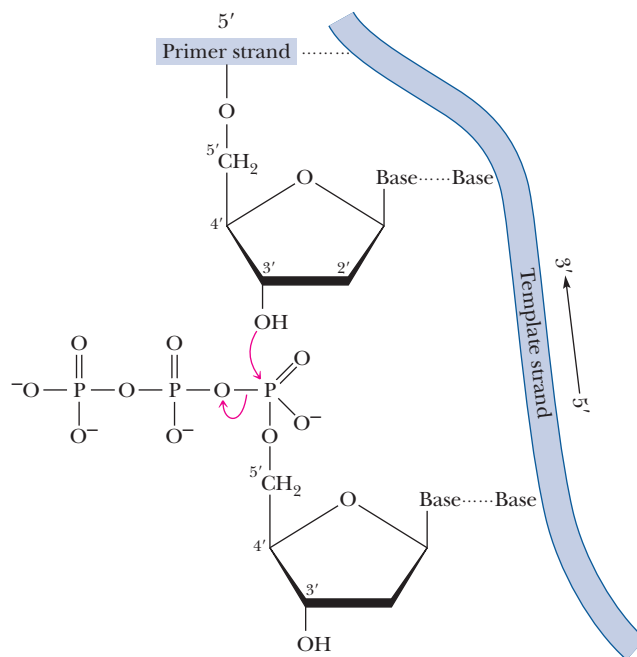
### Semidiscontinuous DNA Replication

All synthesis of nucleotide chains occurs in the  $5' \rightarrow 3'$  direction from the perspective of the chain being synthesized. This is due to the nature of the reaction of DNA synthesis. The last nucleotide added to a growing chain has a 3'-hydroxyl on the sugar. The incoming nucleotide has a 5'-triphosphate on its sugar. The 3'-hydroxyl group at the end of the growing chain is a nucleophile. It attacks the phosphorus adjacent to the sugar in the nucleotide to be added to the growing chain, leading to the elimination of the pyrophosphate and the formation of a new phosphodiester bond (Figure 10.5). We discussed nucleophilic attack by a hydroxyl group at length in the case of serine proteases (Section 7.5); here we see another instance of this kind of mechanism. It is helpful to always keep this mechanism in mind. The further in depth we study DNA, the more the directionality of  $5' \rightarrow 3'$  can lead to confusion over which strand of DNA we are discussing. If you always remember that all synthesis of nucleotides occurs in the  $5' \rightarrow 3'$  direction from the perspective of the growing chain, it will be much easier to understand the processes to come.



- A** Replication of the chromosome of *E. coli*, a typical prokaryote. There is one origin of replication, and there are two replication forks.
- B** Replication of a eukaryotic chromosome. There are several origins of replication, and there are two replication forks for each origin. The "bubbles" that arise from each origin eventually coalesce.

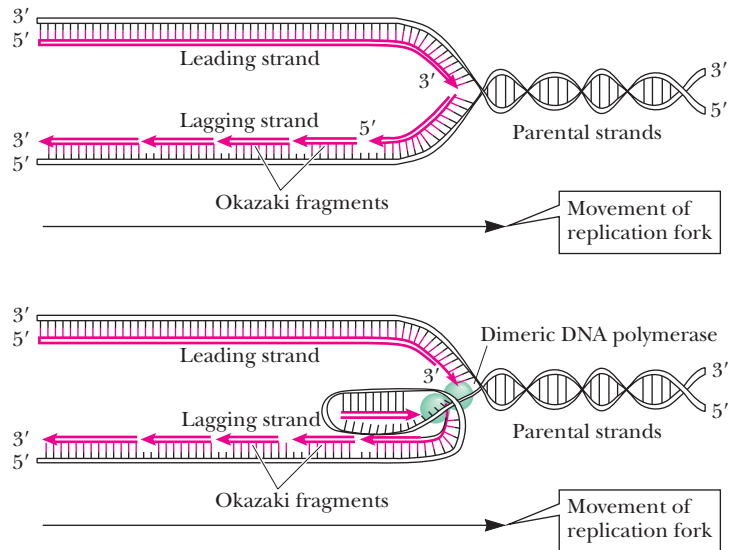
■ **FIGURE 10.4 Bidirectional replication.** Bidirectional replication of DNA is shown for prokaryotes (one origin of replication) and eukaryotes (several origins). Bidirectional replication refers to overall synthesis (compare this with Figure 10.6).



■ **FIGURE 10.5 The addition of a nucleotide to a growing DNA chain.** The 3'-hydroxyl group at the end of the growing DNA chain is a nucleophile. It attacks at the phosphorus adjacent to the sugar in the nucleotide, which is added to the growing chain. Pyrophosphate is eliminated, and a new phosphodiester bond is formed.

**A** As the helix unwinds, the other parental strand (the 5' → 3' strand) is copied in a discontinuous fashion through synthesis of a series of fragments 1000 to 2000 nucleotides in length, called the Okazaki fragments; the strand constructed from Okazaki fragments is called the lagging strand.

**B** Because both strands are synthesized in concert by a dimeric DNA polymerase situated at the replication fork, the 5' → 3' parental strand must wrap around in trombone fashion so that the unit of the dimeric DNA polymerase replicating it can move along it in the 3' → 5' direction. This parental strand is copied in a discontinuous fashion because the DNA polymerase must occasionally dissociate from this strand and rejoin it further along. The Okazaki fragments are then covalently joined by DNA ligase to form an uninterrupted DNA strand.



**FIGURE 10.6 The semidiscontinuous model for DNA replication.** Newly synthesized DNA is shown in red. Because DNA polymerases only polymerize nucleotides 5' → 3', both strands must be synthesized in the 5' → 3' direction. Thus, the copy of the parental 3' → 5' strand is synthesized continuously; this newly made strand is designated the leading strand.

This universal nature of synthesis presents a problem for the cell because as the DNA synthesis proceeds along a replication fork, the two strands are going in opposite directions.

### How can replication proceed along the DNA if the two strands are going in opposite directions?

The problem is solved by different modes of polymerization for the two growing strands. One newly formed strand (the leading strand) is formed continuously from its 5' end to its 3' end at the replication fork on the exposed 3' to 5' template strand. The other strand (the lagging strand) is formed semidiscontinuously in small fragments (typically 1000 to 2000 nucleotides long), sometimes called Okazaki fragments after the scientist who first studied them (Figure 10.6). The 5' end of each of these fragments is closer to the replication fork than the 3' end. The fragments of the lagging strand are then linked together by an enzyme called **DNA ligase**.

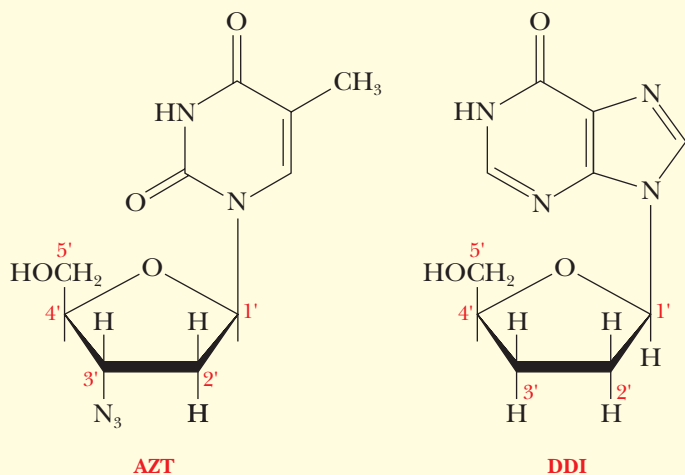
### DNA Polymerase from *E. coli*

The first DNA polymerase discovered was found in *E. coli*. **DNA polymerase** catalyzes the successive addition of each new nucleotide to the growing chain.

#### Apply Your Knowledge

##### DNA Structure

A nucleoside derivative that has been very much in the news is 3'-azido-3'-deoxythymidine (AZT). This compound has been widely used in the treatment of AIDS (acquired immune deficiency syndrome), as has 2'-3'-dideoxyinosine (DDI). Propose a reason for the effectiveness of these two compounds. *Hint:* How might these two compounds fit into a DNA chain?



### Solution

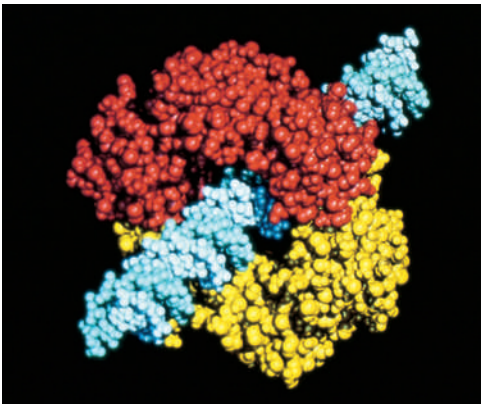
Both compounds lack a hydroxyl group at the 3'-position of the sugar moiety. They cannot form the phosphodiester linkages found in nucleic acids. Thus, they interfere with the replication of the AIDS virus by preventing nucleic acid synthesis.

At least five DNA polymerases are present in *E. coli*. Three of them have been studied more extensively, and some of their properties are listed in Table 10.1. DNA polymerase I (Pol I) was discovered first, with the subsequent discovery of polymerases II (Pol II) and polymerase III (Pol III). Polymerase I consists of a single polypeptide chain, but polymerases II and III are multisubunit proteins that share some common subunits. Polymerase II is not required for replication; rather, it is strictly a repair enzyme. Recently, two more polymerases, Pol IV and Pol V, were discovered. They, too, are repair enzymes, and both are involved in a unique repair mechanism called the SOS response (see the Biochemical Connections box on page 271.) Two important considerations regarding the polymerases are the speed of the synthetic reaction (turnover number) and the **processivity**, which is the number of nucleotides joined before the enzyme dissociates from the template (Table 10.1).

**TABLE 10.1**

Properties of DNA Polymerases of <i>E. coli</i>			
Property	Pol I	Pol II	Pol III
Mass (kDa)	103	90	830
Turnover number ( $\text{min}^{-1}$ )	600	30	1200
Processivity	200	1500	$\geq 500,000$
Number of subunits	1	$\geq 4$	$\geq 10$
Structural gene	<i>polA</i>	<i>polB</i> *	<i>polC</i> *
Polymerization 5' $\rightarrow$ 3'	Yes	Yes	Yes
Exonuclease 5' $\rightarrow$ 3'	Yes	No	No
Exonuclease 3' $\rightarrow$ 5'	Yes	Yes	Yes

\* Polymerization subunit only. These enzymes have multiple subunits, and some of them are shared between both enzymes.



Courtesy of John Kuriyan/University of California, Berkeley

■ **FIGURE 10.7** The dimer of  $\beta$ -subunits of DNA polymerase III bound to DNA. One monomer is shown in yellow, the other in red. Note that the dimer forms a closed loop around the DNA (shown in blue). The rest of the polymerase III holoenzyme is not shown. The remainder of the holoenzyme consists of the core enzyme responsible for the polymerization and the 3' exonuclease activity ( $\alpha$ -,  $\epsilon$ -, and  $\theta$ -subunits) and the  $\gamma$ -complex (consisting of  $\gamma$ -,  $\delta$ -,  $\delta'$ -,  $\chi$ -, and  $\psi$ -subunits), which allows the  $\beta$ -subunits to form a clamp that surrounds the DNA and slides along it as polymerization proceeds. [Adapted from Kong, X. P., et al. *Three-Dimensional Structure of the Subunit of E. Coli DNA Polymerase Holoenzyme: A Sliding DNA Clamp*. *Cell* **69**, 425–437 (1992).]

Polymerase III consists of a core enzyme responsible for the polymerization and 3' exonuclease activity—consisting of  $\alpha$ -,  $\epsilon$ -, and  $\theta$ -subunits—and a number of other subunits, including a dimer of  $\alpha$ -subunits responsible for DNA binding, and the  $\gamma$ -complex—consisting of  $\gamma$ -,  $\delta$ -,  $\delta'$ -,  $\chi$ -, and  $\psi$ -subunits—which allows the  $\beta$ -subunits to form a clamp that surrounds the DNA and slides along it as polymerization proceeds (Figure 10.7). Table 10.2 gives the subunit composition of the DNA polymerase III complex. All of these polymerases add nucleotides to a growing polynucleotide chain but have different roles in the overall replication process. As can be seen in Table 10.1, DNA polymerase III has the highest turnover number and a huge processivity compared to polymerases I and II.

If DNA polymerases are added to a single-stranded DNA template with all the deoxynucleotide triphosphates necessary to make a strand of DNA, no reaction occurs. It was discovered that DNA polymerases cannot catalyze de novo synthesis. All three enzymes require the presence of a **primer**, a short oligonucleotide strand to which the growing polynucleotide chain is covalently attached in the early stages of replication. In essence, DNA polymerases must have a nucleotide with a free 3'-hydroxyl already in place so that they can add the first nucleotide as part of the growing chain. In natural replication, this primer is RNA.

DNA polymerase reaction requires all four deoxyribonucleoside triphosphates—dTTP, dATP, dGTP, and dCTP.  $Mg^{2+}$  and a DNA template are also necessary. Because of the requirement for an RNA primer, all four ribonucleoside triphosphates—ATP, UTP, GTP, and CTP—are needed as well; they are incorporated into the primer. The primer (RNA) is hydrogen-bonded to the template (DNA); the primer provides a stable framework on which the nascent chain can start to grow. The newly synthesized DNA strand begins to grow by forming a covalent linkage to the free 3'-hydroxyl group of the primer.

It is now known that DNA polymerase I has a specialized function in replication—repairing and “patching” DNA—and that DNA polymerase III is the enzyme primarily responsible for the polymerization of the newly formed DNA strand. The major function of DNA polymerases II, IV, and V is as repair enzymes. The exonuclease activities listed in Table 10.1 are part of the proofreading-and-repair functions of DNA polymerases, a process by which incorrect nucleotides are removed from the polynucleotide so that the correct nucleotides can be incorporated. The 3'  $\rightarrow$  5' exonuclease activity, which all three polymerases possess, is part of the **proofreading** function; incorrect

TABLE 10.2

The Subunits of <i>E. coli</i> DNA Polymerase III Holoenzyme			
Subunit	Mass (kDa)	Structural Gene	Function
$\alpha$	130.5	<i>polC</i> ( <i>dnaE</i> )	Polymerase
$\epsilon$	27.5	<i>dnaQ</i>	3'-exonuclease
$\theta$	8.6	<i>holE</i>	$\alpha$ , $\epsilon$ assembly?
$\tau$	71	<i>dnaX</i>	Assembly of holoenzyme on DNA
$\beta$	41	<i>dnaN</i>	Sliding clamp, processivity
$\gamma$	47.5	<i>dnaX(Z)</i>	Part of the $\gamma$ complex*
$\delta$	39	<i>holA</i>	Part of the $\gamma$ complex*
$\delta'$	37	<i>holB</i>	Part of the $\gamma$ complex*
$\chi$	17	<i>holC</i>	Part of the $\gamma$ complex*
$\psi$	15	<i>holD</i>	Part of the $\gamma$ complex*

\* Subunits  $\gamma$ ,  $\delta$ ,  $\delta'$ ,  $\chi$ , and  $\psi$  form the so-called  $\gamma$  complex, which is responsible for the placement of the  $\beta$ -subunits (the sliding clamp) on the DNA. The  $\gamma$  complex is referred to as the clamp loader. The  $\delta$  and  $\tau$  subunits are encoded by the same gene.



nucleotides are removed in the course of replication and are replaced by the correct ones. Proofreading is done one nucleotide at a time. The 5' → 3' exonuclease activity clears away short stretches of nucleotides during **repair**, usually involving several nucleotides at a time. This is also how the RNA primers are removed. The proofreading-and-repair function is less effective in some DNA polymerases.

## 10.4 Proteins Required for DNA Replication

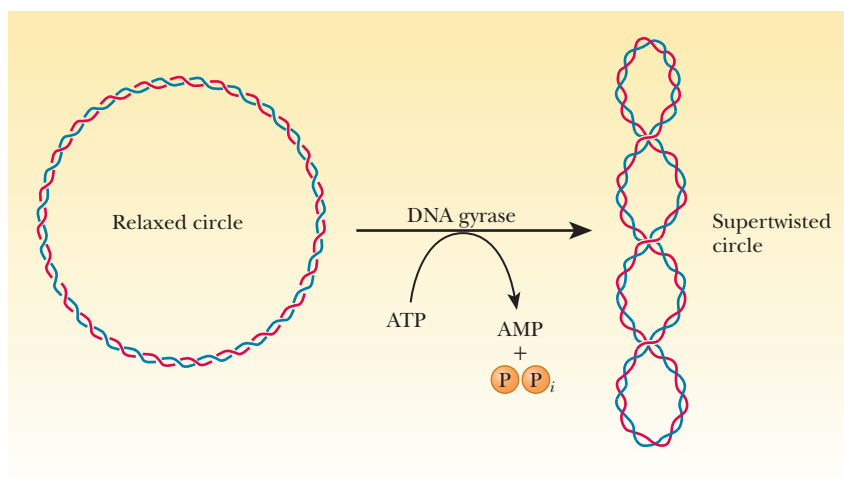
Two questions arise in separating the two strands of the original DNA so that it can be replicated. The first is how to achieve continuous unwinding of the double helix. This question is complicated by the fact that prokaryotic DNA exists in a supercoiled, closed-circular form (see “Tertiary Structure of DNA: Supercoiling” in Section 9.3). The second related question is how to protect single-stranded stretches of DNA that are exposed to intracellular nucleases as a result of the unwinding.

### Supercoiling and Replication

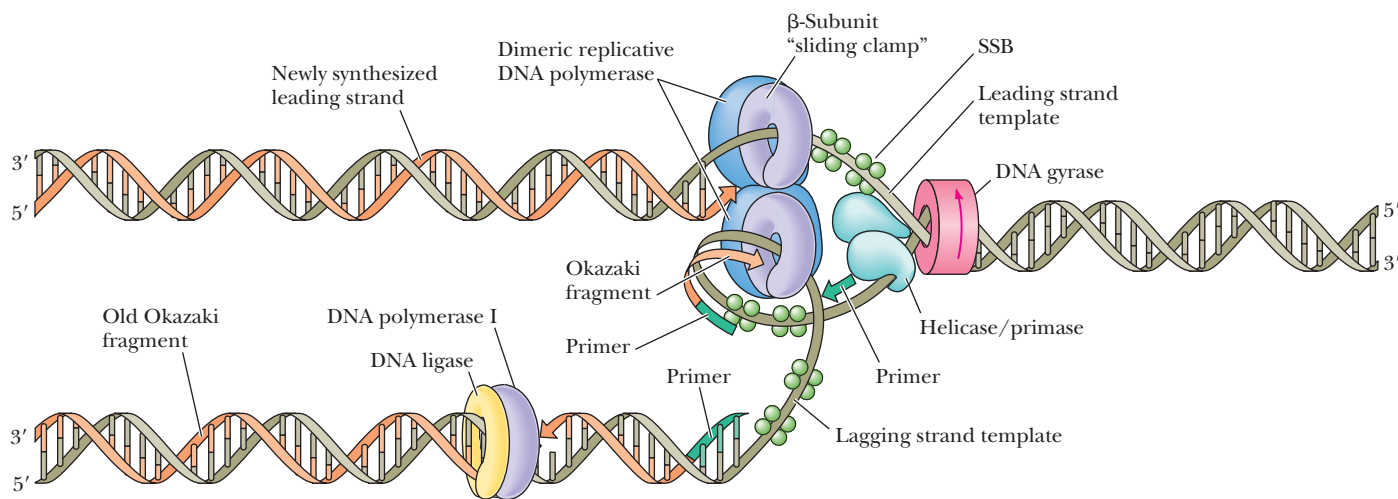
An enzyme called **DNA gyrase** (class II topoisomerase) catalyzes the conversion of relaxed, circular DNA with a nick in one strand to the supercoiled form with the nick sealed that is found in normal prokaryotic DNA (Figure 10.8). A slight unwinding of the helix before the nick is sealed introduces the supercoiling. The energy required for the process is supplied by the hydrolysis of ATP. Some evidence exists that DNA gyrase causes a double-strand break in DNA in the process of converting the relaxed, circular form to the supercoiled form.

### How does replication work with supercoiled DNA?

In replication, the role of the gyrase is somewhat different. The prokaryotic DNA is negatively supercoiled in its natural state; however, opening the helix during replication would introduce positive supercoils ahead of the replication fork. To see this phenomenon for yourself, look for an old phone cord and try straightening out a section of the cord. You will be able to see the result in the coils ahead. If the replication fork continued to move, the torsional strain of the positive supercoils would eventually make further replication impossible. DNA gyrase fights these positive supercoils by putting negative supercoils ahead of the replication fork (Figure 10.9). A helix-destabilizing protein, called a **helicase**, promotes unwinding by binding at the



■ **FIGURE 10.8** DNA gyrase introduces supertwisting in circular DNA.



■ **FIGURE 10.9** General features of a replication fork. The DNA duplex is unwound by the action of DNA gyrase and helicase, and the single strands are coated with SSB (ssDNA-binding protein). Primase periodically primes synthesis on the lagging strand. Each half of the dimeric replicative polymerase is a holoenzyme bound to its template strand by a  $\beta$ -subunit sliding clamp. DNA polymerase I and DNA ligase act downstream on the lagging strand to remove RNA primers, replace them with DNA, and ligate the Okazaki fragments.

replication fork. A number of helicases are known, including the *DnaB* protein and the *rep* protein.

### How is single-stranded DNA protected long enough for replication?

Single-stranded regions of DNA are very susceptible to degradation by nucleases. If left unchecked, this would make it very difficult to complete replication before DNA damage would occur. Another protein, called the **single-strand binding protein (SSB)**, stabilizes the single-stranded regions by binding tightly to these portions of the molecule. The presence of this DNA-binding protein protects the single-stranded regions from hydrolysis by the nucleases.

### The Primase Reaction

One of the great surprises in studies of DNA replication was the discovery that *RNA serves as a primer in DNA replication*. In retrospect, it is not surprising at all, because RNA can be formed *de novo* without a primer, even though DNA synthesis requires a primer. This finding lends support to theories of the origin of life in which RNA, rather than DNA, was the original genetic material. The fact that RNA has been shown to have catalytic ability in several cases has added support to that theory (Chapter 11). A primer in DNA replication must have a free 3'-hydroxyl to which the growing chain can attach, and both RNA and DNA can provide this group. The primer activity of RNA was first observed *in vivo*. In some of the original *in vitro* experiments, DNA was used as a primer because a primer consisting of DNA was expected. Living organisms are, of course, far more complex than isolated molecular systems and, as a result, can be full of surprises for researchers.

### Where does the primer come from?

It has subsequently been found that a separate enzyme, called **primase**, is responsible for copying a short stretch of the DNA template strand to produce the RNA primer sequence. The first primase was discovered in *E. coli*. The enzyme consists of a single polypeptide chain, with a molecular weight of about 60,000. There are 50 to 100 molecules of primase in a typical *E. coli* cell.

TABLE 10.3

## A Summary of DNA Replication in Prokaryotes

1. DNA synthesis is bidirectional. Two replication forks advance in opposite directions from an origin of replication.
2. The direction of DNA synthesis is from the 5' end to the 3' end of the newly formed strand. One strand (the leading strand) is formed continuously, while the other strand (the lagging strand) is formed discontinuously. On the lagging strand, small fragments of DNA (Okazaki fragments) are subsequently linked.
3. Five DNA polymerases have been found in *E. coli*. Polymerase III is primarily responsible for the synthesis of new strands. The first polymerase enzyme discovered, polymerase I, is involved in synthesis, proofreading, and repair. Polymerases II, IV, and V function as repair enzymes under unique conditions.
4. DNA gyrase introduces a swivel point in advance of the movement of the replication fork. A helix-destabilizing protein, a helicase, binds at the replication fork and promotes unwinding. The exposed single-stranded regions of the template DNA are stabilized by a DNA-binding protein.
5. Primase catalyzes the synthesis of an RNA primer.
6. The synthesis of new strands is catalyzed by Pol III. The primer is removed by Pol I, which also replaces the primer with deoxynucleotides. DNA ligase seals the remaining nicks.

The primer and the protein molecules at the replication fork constitute the **primosome**. The general features of DNA replication, including the use of an RNA primer, appear to be common to all prokaryotes (Figure 10.9).

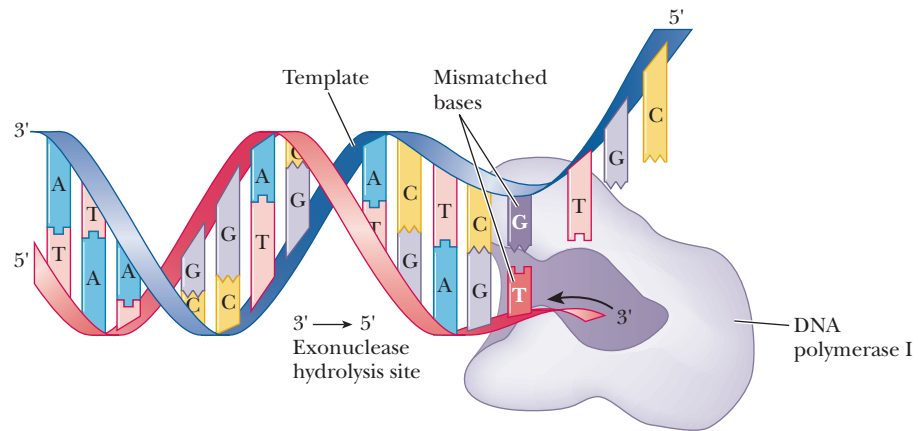
### Synthesis and Linking of New DNA Strands

The synthesis of two new strands of DNA is begun by DNA polymerase III. The newly formed DNA is linked to the 3'-hydroxyl of the RNA primer, and synthesis proceeds from the 5' end to the 3' end on both the leading and the lagging strands. Two molecules of Pol III, one for the leading strand and one for the lagging strand, are physically linked to the *primosome*. The resulting multiprotein complex is called the **replisome**. As the replication fork moves, the RNA primer is removed by polymerase I, using its exonuclease activity. The primer is replaced by deoxynucleotides, also by DNA polymerase I, using its polymerase activity. (The removal of the RNA primer and its replacement with the missing portions of the newly formed DNA strand by polymerase I are the repair function we mentioned earlier.) None of the DNA polymerases can seal the nicks that remain; DNA ligase is the enzyme responsible for the final linking of the new strand. Table 10.3 summarizes the main points of DNA replication in prokaryotes.

## 10.5 Proofreading and Repair

DNA replication takes place only once each generation in each cell, unlike other processes, such as RNA and protein synthesis, which occur many times. It is essential that the fidelity of the replication process be as high as possible to prevent **mutations**, which are errors in replication. Mutations are frequently harmful, even lethal, to organisms. Nature has devised several ways to ensure that the base sequence of DNA is copied faithfully.

Errors in replication occur spontaneously only once in every  $10^9$  to  $10^{10}$  base pairs. *Proofreading* refers to the removal of incorrect nucleotides immediately after they are added to the growing DNA during the replication process. DNA polymerase I has three active sites, as demonstrated by Hans Klenow. Pol I can be cleaved into two major fragments. One of them (the Klenow fragment) contains



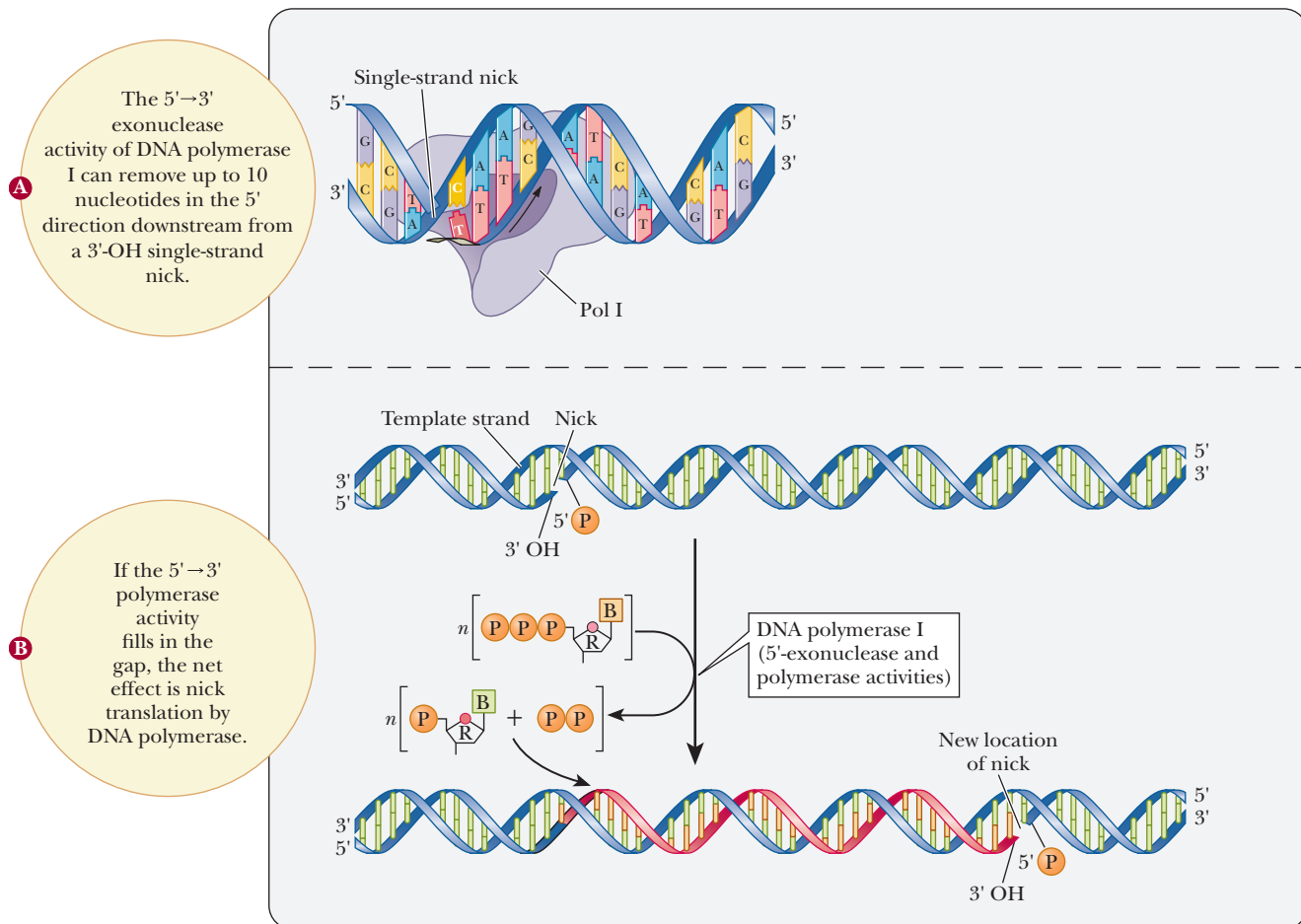
■ **FIGURE 10.10** DNA polymerase proofreading. The 3' → 5' exonuclease activity of DNA polymerase I removes nucleotides from the 3' end of the growing DNA chain.

the polymerase activity and the proofreading activity. The other contains the 5' → 3' repair activity. Figure 10.10 shows the proofreading activity of Pol I. Errors in hydrogen bonding lead to the incorporation of an incorrect nucleotide into a growing DNA chain once in every  $10^4$  to  $10^5$  base pairs. DNA polymerase I uses its 3' exonuclease activity to remove the incorrect nucleotide. Replication resumes when the correct nucleotide is added, also by DNA polymerase I. Although the specificity of hydrogen-bonded base pairing accounts for one error in every  $10^4$  to  $10^5$  base pairs, the proofreading function of DNA polymerase improves the fidelity of replication to one error in every  $10^9$  to  $10^{10}$  base pairs.

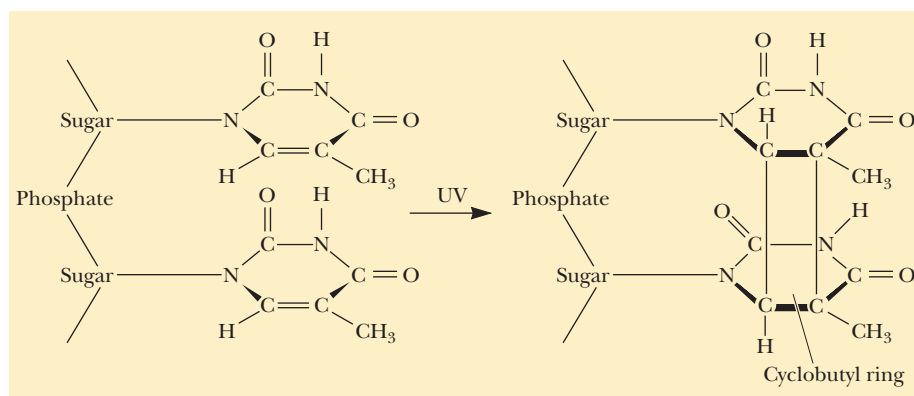
### How does proofreading improve replication fidelity?

During replication, a *cut-and-patch process* catalyzed by polymerase I takes place. The cutting is the removal of the RNA primer by the 5' exonuclease function of the polymerase, and the patching is the incorporation of the required deoxynucleotides by the polymerase function of the same enzyme. Note that this part of the process takes place after polymerase III has produced the new polynucleotide chain. Existing DNA can also be repaired by polymerase I, using the cut-and-patch method, if one or more bases have been damaged by an external agent, or if a mismatch was missed by the proofreading activity. DNA polymerase I is able to use its 5' → 3' exonuclease activity to remove RNA primers or DNA mistakes as it moves along the DNA. It then fills in behind it with its polymerase activity. This process is called **nick translation** (Figure 10.11). In addition to experiencing those spontaneous mutations caused by misreading the genetic code, organisms are frequently exposed to **mutagens**, agents that produce mutations. Common mutagens include ultraviolet light, ionizing radiation (radioactivity), and various chemical agents, all of which lead to changes in DNA over and above those produced by spontaneous mutation. The most common effect of ultraviolet light is the creation of pyrimidine dimers (Figure 10.12). The  $\pi$  electrons from two carbons on each of two pyrimidines form a cyclobutyl ring, which distorts the normal shape of the DNA and interferes with replication and transcription. Chemical damage, which is often caused by free radicals (Figure 10.13), can lead to a break in the phosphodiester backbone of the DNA strand. This is one of the primary reasons that antioxidants are so popular as dietary supplements these days.

When damage has managed to escape the normal exonuclease activities of DNA polymerases I and III, prokaryotes have a variety of other repair mechanisms at their disposal. In **mismatch repair**, enzymes recognize that two bases are incorrectly paired. The area with the mismatch is removed, and DNA polymerases replicate the area again. If there is a mismatch, the challenge for the repair system is to know which of the two strands is the correct one. This is possible only because prokaryotes alter their DNA at certain locations (Chapter 13)



■ **FIGURE 10.11** DNA polymerase repair.

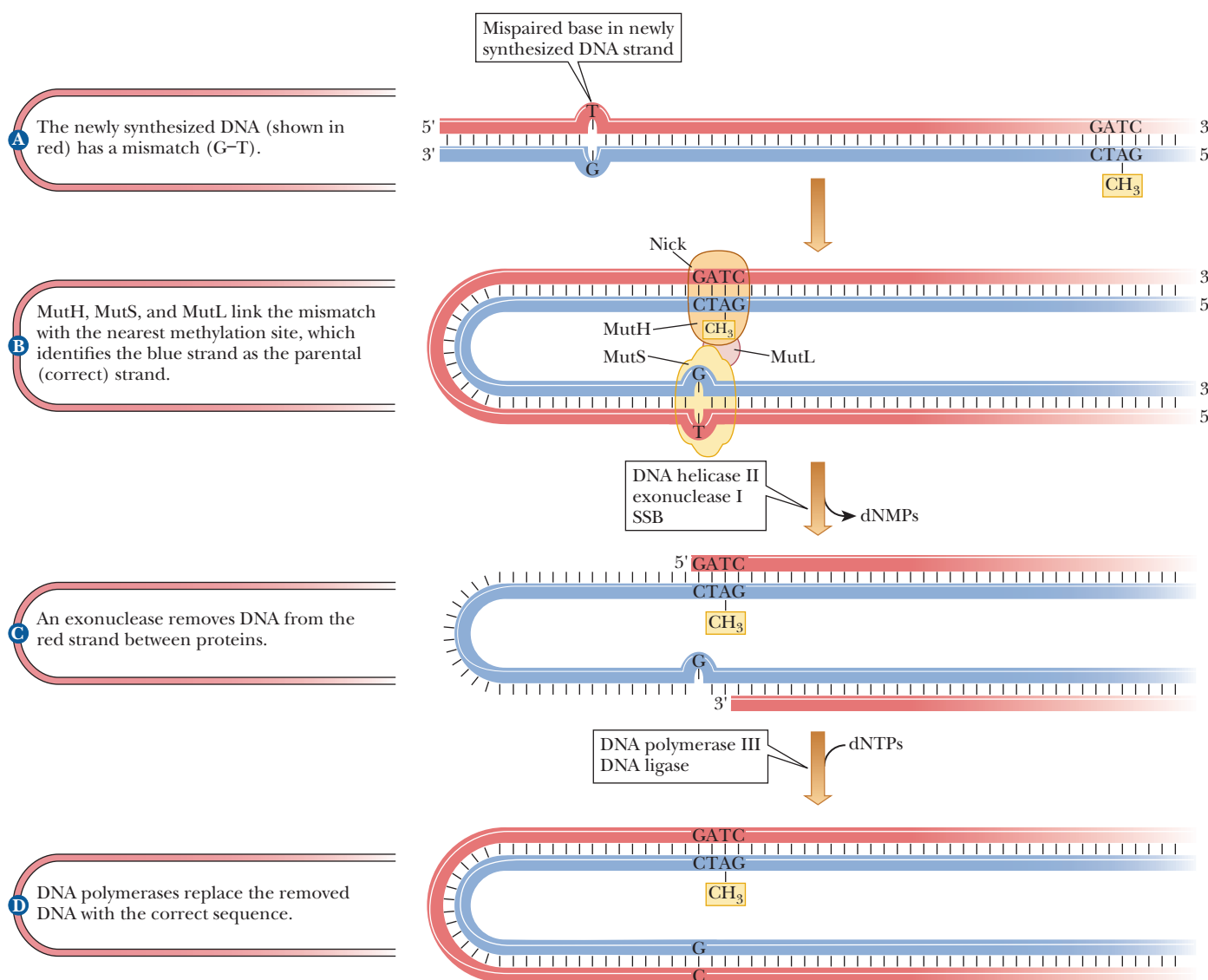
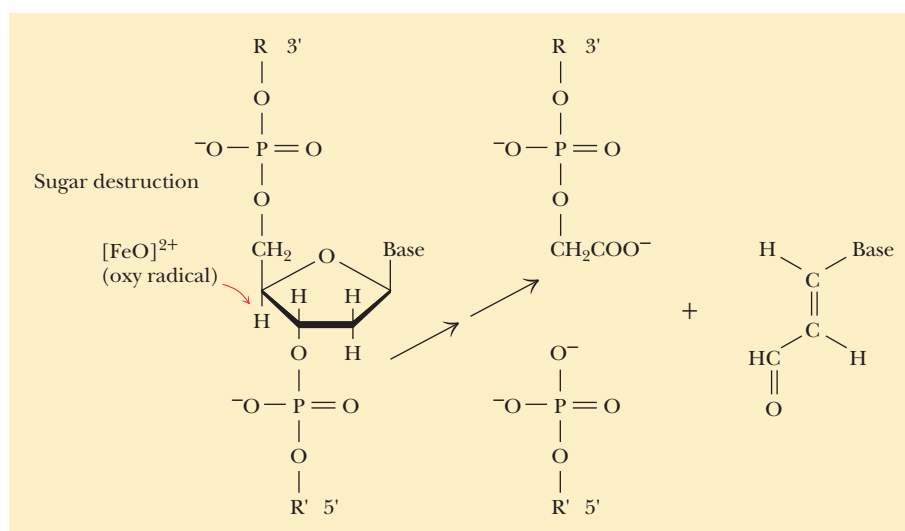


■ **FIGURE 10.12** UV irradiation causes dimerization of adjacent thymine bases. A cyclobutyl ring is formed between carbons 5 and 6 of the pyrimidine rings. Normal base pairing is disrupted by the presence of such dimers.

by modifying bases with added methyl groups. This **methylation** occurs shortly after replication. Thus, immediately after replication, there is a window of opportunity for the mismatch-repair system. Figure 10.14 shows how this works. Assume that a bacterial species methylates adenines that are part of a unique sequence. Originally, both parental strands are methylated. When the DNA is replicated, a mistake is made, and a T is placed opposite a G (Figure 10.14a). Because the parental strand contained methylated adenines, the enzymes can distinguish the parental strand from the newly synthesized daughter strand without the modified bases. Thus, the T is the mistake and not the G. Several



■ **FIGURE 10.13** **Oxidation damage.** Oxygen radicals, in the presence of metal ions such as  $\text{Fe}^{2+}$ , can destroy sugar rings in DNA, breaking the strand.



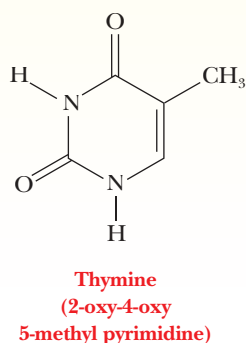
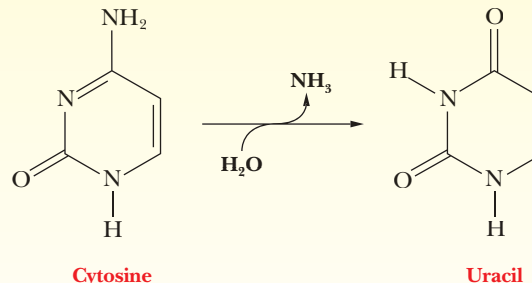
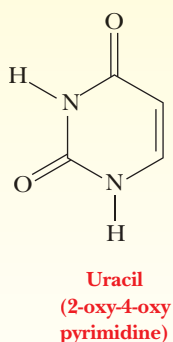
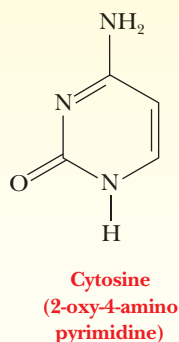
■ **FIGURE 10.14** **Mismatch repair in *E. coli*.** (Adapted from *Lehninger, Principles of Biochemistry, Third Edition*, by David L. Nelson and Michael M. Cox. © 1982, 1992, 2000 by Worth Publishers. Used with permission of W. H. Freeman and Company.)

# Biochemical Connections GENETICS

## Why Does DNA Contain Thymine and Not Uracil?

Given that both uracil and thymine base-pair with adenine, why does RNA contain uracil and DNA contain thymine? Scientists now believe that RNA was the original hereditary molecule, and that DNA developed later. If we compare the structure of uracil and thymine, the only difference is the presence of a methyl group at C-5 of thymine. This group is not on the side of the molecule involved in base pairing. Because carbon sources and

energy are required to methylate a molecule, there must be a reason for DNA developing with a base that does the same thing as uracil but that requires more energy to produce. The answer is that thymine helps guarantee replication fidelity. One of the most common spontaneous mutations of bases is the natural deamination of cytosine.

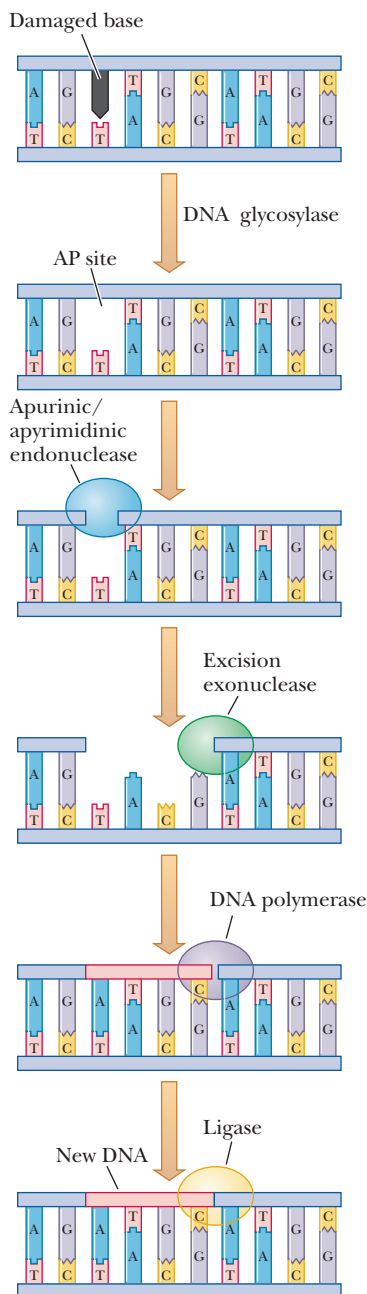


At any moment, a small but finite number of cytosines lose their amino groups to become uracil. Imagine that during replication, a C–G base pair separates. If at that moment the C deaminates to U, it would tend to base-pair to A instead of to G. If U were a natural base in DNA, the DNA polymerases would just line up an adenine across from the uracil, and there would be no way to know that the uracil was a mistake. This would lead to a much higher level of mutation during replication. Because uracil is an unnatural base in DNA, DNA polymerases can recognize it as a mistake and can replace it. Thus, the incorporation of thymine into DNA, though energetically more costly, helps ensure that the DNA is replicated faithfully.

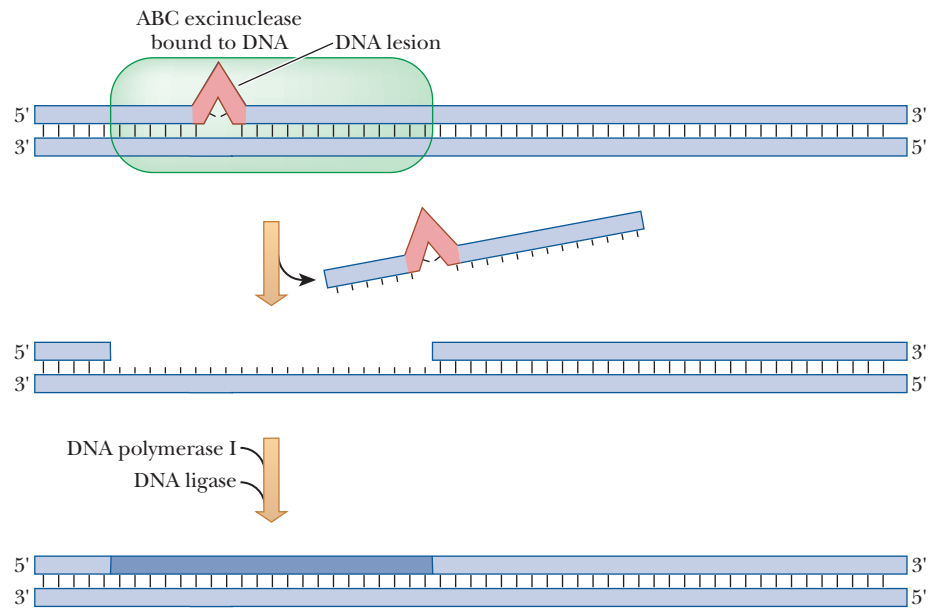
proteins and enzymes are then involved in the repair process. *MutH*, *MutS*, and *MutL* form a loop between the mistake and a methylation site. DNA helicase II helps unwind the DNA. *Exonuclease I* removes the section of DNA containing the mistake (Figure 10.14b). Single-stranded binding proteins protect the template (blue) strand from degradation. DNA polymerase III then fills in the missing piece (Figure 10.14c).

Another repair system is called **base-excision repair** (Figure 10.15). A base that has been damaged by oxidation or chemical modification is removed by *DNA glycosylase*, leaving an *AP site*, so called because it is apurinic or apyrimidinic (without purine or pyrimidine). An *AP endonuclease* then removes the sugar and phosphate from the nucleotide. An *excision exonuclease* then removes several more bases. Finally, DNA polymerase I fills in the gap, and DNA ligase seals the phosphodiester backbone.

**Nucleotide-excision repair** is common for DNA lesions caused by ultraviolet or chemical means, which often lead to deformed DNA structures. Figure 10.16 demonstrates how a large section of DNA containing the lesion is removed by *ABC excinuclease*. DNA polymerase I and DNA ligase then work to fill in the gap. This type of repair is also the most common repair for ultraviolet damage in mammals. Defects in DNA repair mechanisms can have drastic consequences.



■ **FIGURE 10.15 Base-excision repair.** A damaged base (●) is excised from the sugar–phosphate backbone by DNA glycosylase, creating an AP site. Then an apurinic/apyrimidinic endonuclease severs the DNA strand, and an excision nuclease removes the AP site and several nucleotides. DNA polymerase I and DNA ligase then repair the gap.



■ **FIGURE 10.16 Nucleotide-excision repair.** When a serious lesion, such as a pyrimidine dimer, is detected, ABC excinuclease binds to the region and cuts out a large piece of DNA, including the lesion. DNA polymerase I and DNA ligase then resynthesize and seal the DNA. (Adapted from *Lehninger, Principles of Biochemistry, Third Edition*, by David L. Nelson and Michael M. Cox. © 1982, 1992, 2000 by Worth Publishers. Used with permission of W. H. Freeman and Company.)

One of the most remarkable examples is the disease *xeroderma pigmentosum*. Affected individuals develop numerous skin cancers at an early age because they do not have the repair system to correct damage caused by ultraviolet light. The endonuclease that nicks the damaged portion of the DNA is probably the missing enzyme. The repair enzyme that recognizes the lesion has been named XPA protein after the disease. The cancerous lesions eventually spread throughout the body, causing death.

## 10.6 DNA Recombination

**Genetic recombination** is a natural process in which genetic information is rearranged to form new associations. For example, compared to their parents, progeny may have new combinations of traits because of recombination. At the molecular level, genetic recombination is the exchange of one DNA sequence with another or the incorporation of a DNA sequence into another. If the recombination involves a reaction between homologous sequences, then the process is called **homologous recombination**. When very different nucleotide sequences recombine, it is **nonhomologous recombination**.

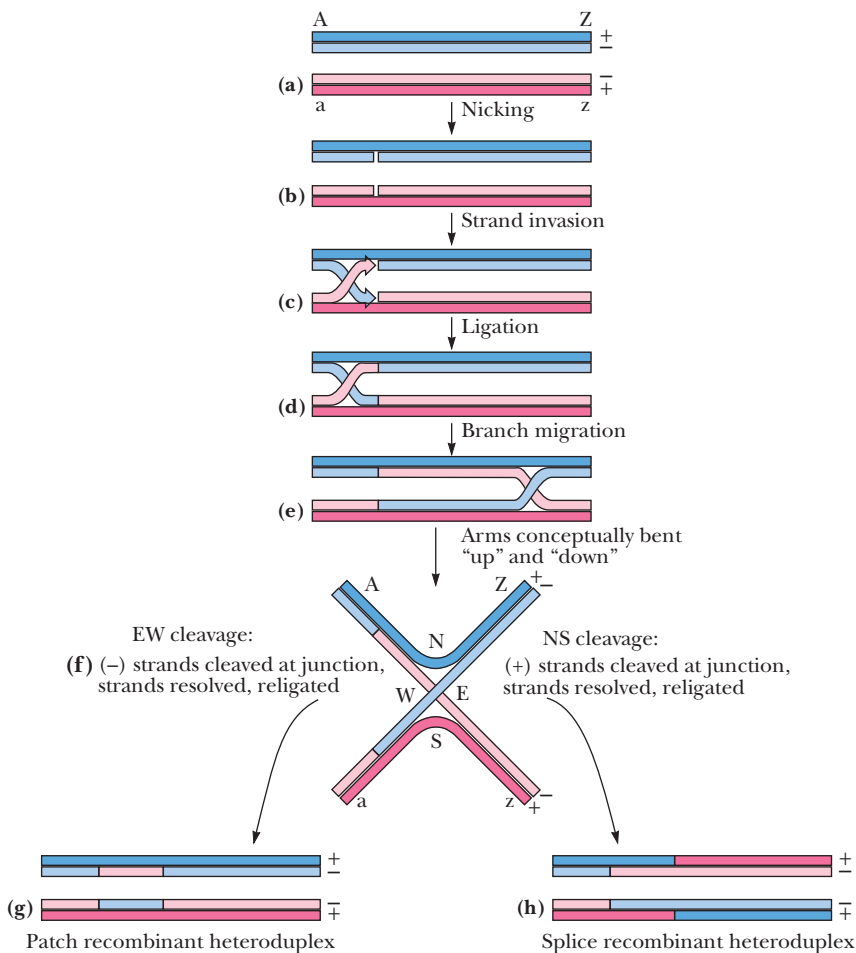
The process involved in homologous recombination is also termed general recombination because the enzymes that mediate the exchange can use essentially any pair of homologous DNA sequences. It occurs in all organisms and is prevalent during the production of gametes in diploid organisms during meiosis. In higher animals, it also occurs in somatic cells and is responsible for the rearrangements within immune cells that lead to the tremendous diversity of immunoglobulins that vertebrates possess (see Chapter 14). Recombination does not occur randomly around a chromosome. There are some areas of a chromosome much more likely to show recombination. These zones are called **hot spots**.

Recombination was first shown by Meselson and Weigle, who used two different phages to infect bacteria (see Figure 10.17). One of the phages had light DNA and one had heavy DNA. Without recombination, the light DNA would always package into light virus particles, and the heavy would package into heavy particles. This would lead to only two populations of phages after infection. Meselson and Weigle's

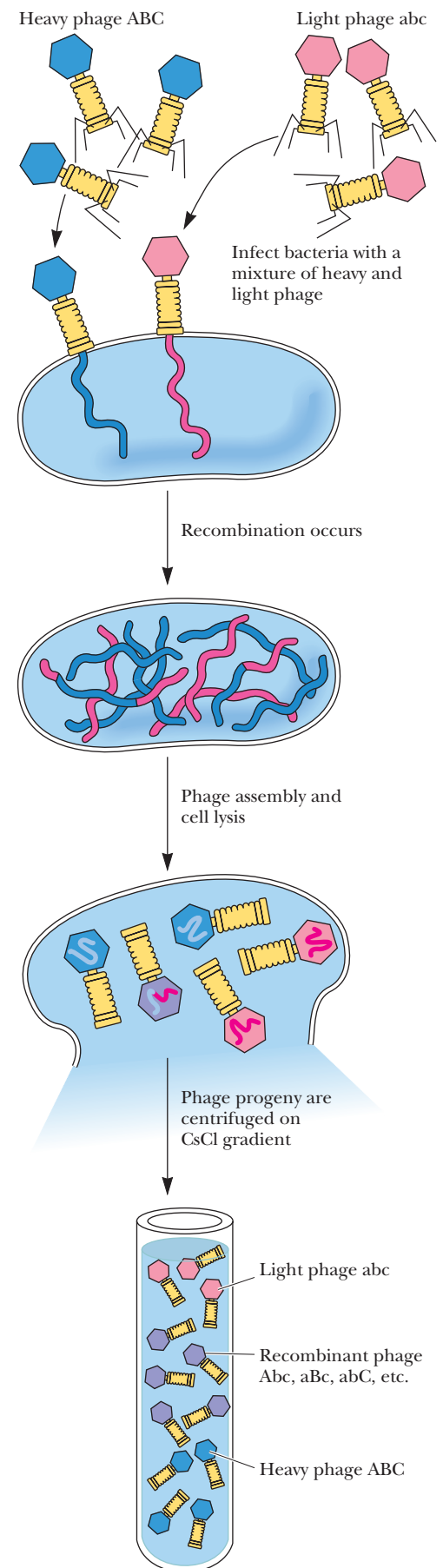
■ **FIGURE 10.17** Meselson and Weigle's experiment demonstrated that a physical exchange of chromosome parts actually occurs during recombination. Density labeled "heavy" phage, symbolized as ABC in this diagram, was used to coinfect bacteria along with "light" phage, the abc phage. The progeny from the infection were collected and subjected to CsCl density gradient centrifugation. Parental-type ABC and abc phage were well-separated in the gradient, but recombinant phage (ABc, Abc, etc.) were distributed diffusely between the two parental strands because they contained chromosomes constituted from fragments of both "heavy" and "light" DNA. These recombinant chromosomes formed by breakage and reunion of parental "heavy" and "light" chromosomes. (From *Biochemistry* by Reginald H. Garrett and Charles M. Grisham, © 2010 by Brooks/Cole, Cengage Learning.)

results showed, however, that there were intermediate combinations that had DNA of different weights. This demonstrated that the phage DNA was recombining.

Recombination occurs by the breakage and reunion of DNA strands so that physical exchange of DNA parts takes place. The mechanism was deduced in 1964 by Robin Holliday and is referred to as the **Holliday Model** (see Figure 10.18). First the two homologous DNA segments align. In eukaryotes, this is called **chromosome pairing**. A nick occurs at the same place on two homologous strands, shown as the (–) strands in Figure 10.18b. The DNAs on the two strands then swap places, or cross over, at the nick by the process of **strand invasion**, shown in Figure 10.18c. The crossing over can then proceed down each strand of DNA, like opening a zipper and reclosing on a different one. The branch migration leads to strand exchange between the two homologous DNA pieces. This leads to exchange of genes and traits caused by them.



■ **FIGURE 10.18** The Holliday model for homologous recombination. The + signs and – signs label strands of like polarity. For example, assume that the two strands running 5' 3' as read left to right are labeled +, and the two strands running 3' 5' as read left to right are labeled –. Only strands of like polarity exchange DNA during recombination (see text for detailed description). (From *Biochemistry* by Reginald H. Garrett and Charles M. Grisham, © 2010 by Brooks/Cole, Cengage Learning.)



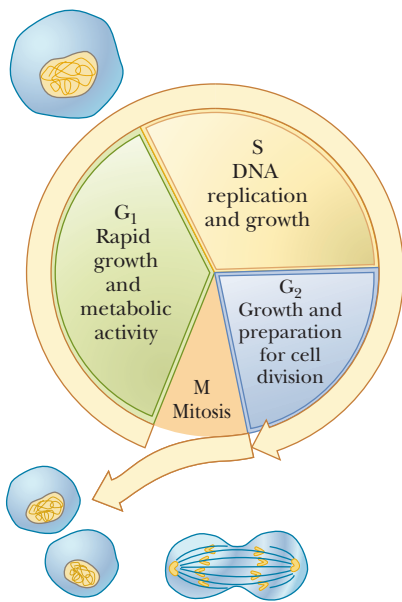
In *E. coli*, the principal molecules involved are the **RecBCD** enzyme complex, which initiates recombination; the **RecA** protein, which binds single-stranded DNA; and the **RuvA**, **RuvB**, and **RuvC** proteins, which drive branch migration. Eukaryotic homologs of these prokaryotic proteins have been found, indicating that the fundamental process of general recombination is conserved.

Recombination is a critical process during meiosis. Surprisingly, the segregation of chromosomes during formation of gametes is quite inaccurate, with estimates indicating that abnormal chromosome numbers in gametes, called **aneuploidy**, occur in 10%–25% of all conceptions. This is the leading cause of pregnancy loss and birth defects. The process of recombination is critical to the correct segregation of the chromosomes and is an active area of research for that reason. Recently scientists discovered several genes involved in controlling recombination. One gene discovered, *PRDM9*, was found to be very important in maintaining the recombination in known hot spots. The protein product of *PRDM9* is a zinc-finger protein (Section 11.7) that acts as a histone methyltransferase. Several variations of the sequence for *PRDM9* have been discovered, and researchers are studying the effects of changes in this gene on recombination and correct meiosis.

## 10.7 Eukaryotic DNA Replication

Our understanding of replication in eukaryotes is not as extensive as that in prokaryotes, owing to the higher level of complexity in eukaryotes and the consequent difficulty in studying the processes. Even though many of the principles are the same, eukaryotic replication is more complicated in three basic ways: there are multiple origins of replication, the timing must be controlled to that of cell divisions, and more proteins and enzymes are involved.

In a human cell, a few billion base pairs of DNA must be replicated once, and only once, per cell cycle. Cell growth and division are divided into phases—M, G<sub>1</sub>, S, and G<sub>2</sub> (Figure 10.19). DNA replication takes place during a few hours in the S phase, and pathways exist to make sure that the DNA is replicated only once per cycle. Eukaryotic chromosomes accomplish this DNA synthesis by having replication begin at multiple origins of replication, also called **replicators**. These are specific DNA sequences that are usually between gene sequences. An average human chromosome may have several hundred replicators. The zones where replication is proceeding are called **replicons**, and the size of these varies with the species. In higher mammals, replicons may span 500 to 50,000 base pairs.



■ **FIGURE 10.19** The eukaryotic cell cycle. The stages of mitosis and cell division define the M phase ("M" for *mitosis*). G<sub>1</sub> ("G" for *gap*, not *growth*) is typically the longest part of the cell cycle; G<sub>1</sub> is characterized by rapid growth and metabolic activity. Cells that are quiescent—that is, not growing and dividing (such as neurons)—are said to be in G<sub>0</sub>. The S phase is the time of DNA synthesis. S is followed by G<sub>2</sub>, a relatively short period of growth in which the cell prepares for division. Cell cycle times vary from less than 24 hours (rapidly dividing cells, such as the epithelial cells lining the mouth and gut) to hundreds of days.

### How is replication tied to cell division?

The best-understood model for control of eukaryotic replication is from yeast cells (Figure 10.20). Only chromosomes from cells that have reached the G<sub>1</sub> phase are competent to initiate DNA replication. Many proteins are involved in the control of replication and its link to the cell cycle. As usual, these proteins are usually given an abbreviation that makes them easier to say, but more difficult for the uninitiated to comprehend at first glance. The first proteins involved are seen during a window of opportunity that occurs between the early and late G<sub>1</sub> phase (see Figure 10.20, *top*). Replication is initiated by a multi-subunit protein called the **origin recognition complex (ORC)**, which binds to the origin of replication. This protein complex appears to be bound to the DNA throughout the cell cycle, but it serves as an attachment site for several proteins that help control replication. The next protein to bind is an activation factor called the **replication activator protein (RAP)**. After the activator



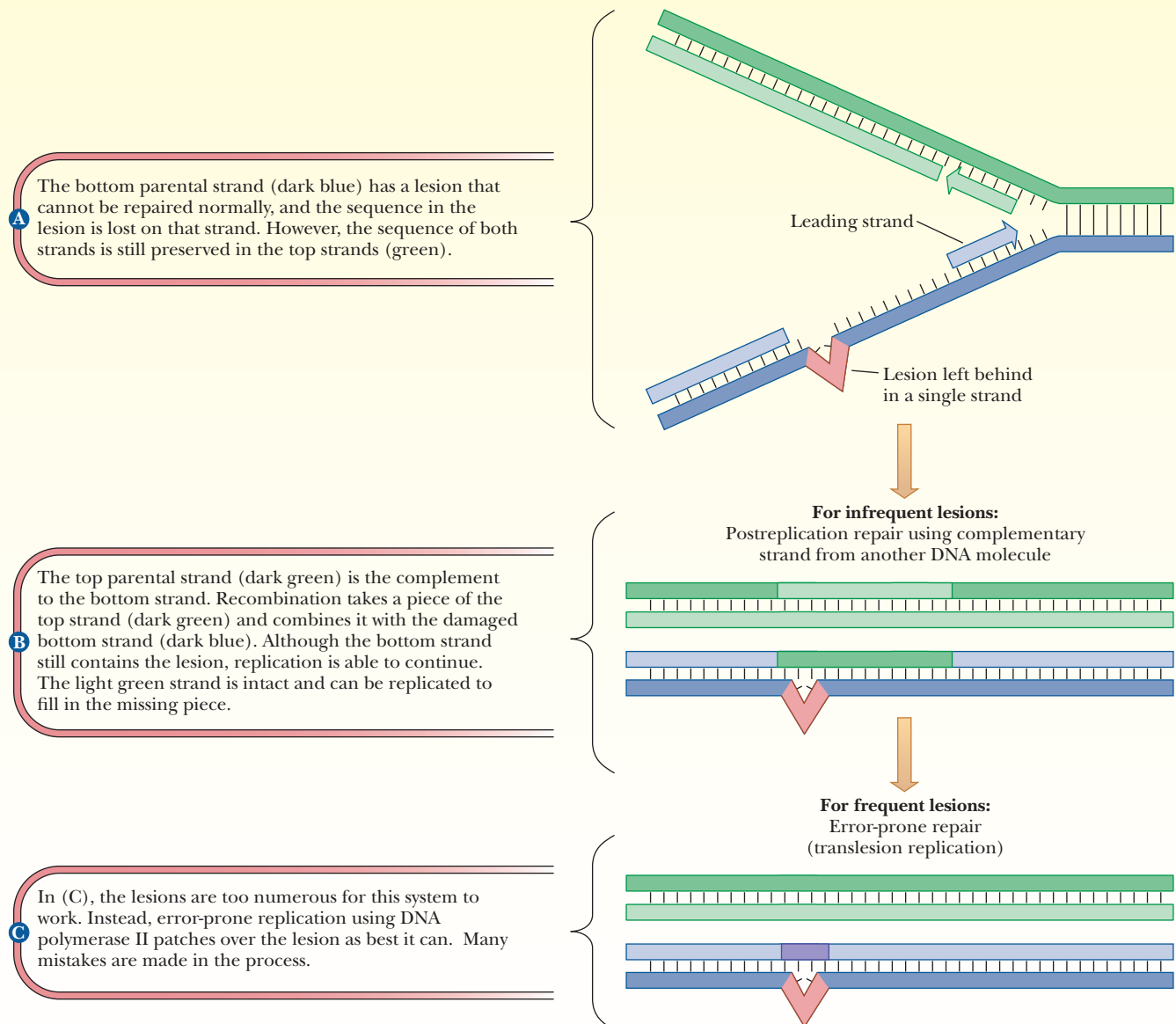
# Biochemical Connections MICROBIOLOGY

## The SOS Response in *E. coli*

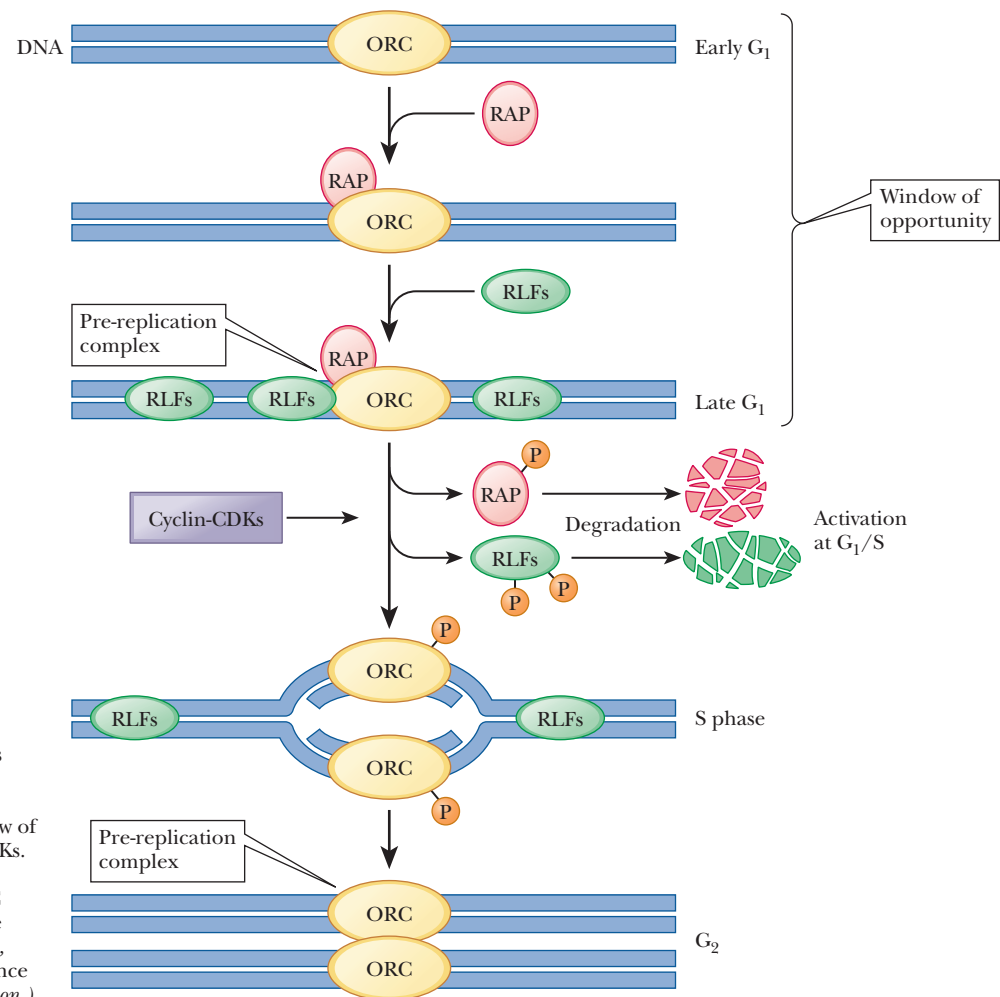
When bacteria are subjected to extreme conditions and a great deal of DNA damage occurs, the normal repair mechanisms are not up to the task of repairing it. However, bacteria have one last card to play, which is called, appropriately, the SOS response. At least 15 proteins are activated as part of this response, including the mysterious DNA polymerase II. Another important protein is called *recA*. It gets its name from the fact that it is involved in a recombination event. Homologous DNA can recombine by a variety of mechanisms (see Section 10.6).

The figure below demonstrates the mechanics of prokaryotic DNA repair under two different circumstances. The starting situation is shown in (a). The parental strands are shown in dark green and dark blue. The bottom parental strand has a serious lesion

(pink wedge) that cannot be fixed by the other repair mechanisms we have seen in this section. In the bottom strand, the sequence at the lesion could be lost. If the lesions are infrequent, recombination can be used to repair the bottom strand. In (b), a piece of the top parental strand (dark green) is used to recombine with the bottom parental strand (dark blue). In this way replication is able to continue. The light green strand can be replicated normally to fill in the missing piece. This type of recombination is seen when the lesions are infrequent. If the lesions are too frequent, then recombination will not work. In that case, error-prone replication occurs. The DNA polymerase just “has a good guess” over the lesion and continues. Many errors occur, and the mutation rate is high, but some of the bacteria will survive.



■ **Recombination can be used to repair infrequent lesions.** (Adapted from *Lehninger, Principles of Biochemistry, Third Edition*, by David L. Nelson and Michael M. Cox. © 1982, 1992, 2000 by Worth Publishers. Used with permission of W. H. Freeman and Company.)



■ **FIGURE 10.20** Model for initiation of the DNA replication cycle in eukaryotes. ORC is present at the replicators throughout the cell cycle. The prereplication complex (pre-RC) is assembled through sequential addition of the RAP (replication activator protein) and RLFs (replication licensing factors) during a window of opportunity defined by the state of cyclin-CDKs. Phosphorylation of the RAP, ORC, and RLFs triggers replication. After initiation, a post-RC state is established, and the RAP and RLFs are degraded. (Adapted from Figure 2 in Stillman, B., 1996. *Cell Cycle Control of DNA Replication*. Science 274: 1659–1663. © 1996 AAAS. Used by permission.)

protein is bound, **replication licensing factors (RLFs)** can bind. Yeast contains at least six different RLFs. They get their name from the fact that replication cannot proceed until they are bound. One of the keys to linking replication to cell division is that some of the RLF proteins have been found to be cytosolic. Thus, they have access to the chromosome only when the nuclear membrane dissolves during mitosis. Until they are bound, replication cannot occur. After RLFs bind, the DNA is then competent for replication. The combination of the DNA, ORC, RAP, and RLFs constitutes what researchers call the **pre-replication complex (pre-RC)**.

The next step involves other proteins and protein kinases. In Chapter 7, we learned that many processes are controlled by kinases phosphorylating target proteins. One of the great discoveries in this field was the existence of **cyclins**, which are proteins that are produced in one part of a cell cycle and degraded in another. Cyclins are able to combine with specific protein kinases, called **cyclin-dependent protein kinases (CDKs)**. When these cyclins combine with CDKs, they can activate DNA replication and also block reassembly of a pre-RC after initiation. The state of activity of the CDKs and the cyclins determines the window of opportunity for DNA synthesis. Cyclin-CDK complexes phosphorylate sites on the RAP, the RLFs, and the ORC itself. Once phosphorylated, the RAP dissociates from the pre-RC, as do the RLFs. Once phosphorylated and released, the RAP and the RLFs are degraded (Figure 10.20, *middle*). Thus, the activation of cyclin-CDKs serves both to initiate DNA replication and to

prevent formation of another pre-RC. In the G<sub>2</sub> phase, the DNA has been replicated. During mitosis, the DNA is separated into the daughter cells. At the same time, the dissolved nuclear membrane allows entrance of the licensing factors that are produced in the cytosol so that each daughter cell can initiate a new round of replication.

## Eukaryotic DNA Polymerases

At least 19 different polymerases are present in eukaryotes, of which 5 have been studied more extensively (Table 10.4). The use of animals rather than plants for study avoids the complication of any DNA synthesis in chloroplasts. The five best-studied polymerases are called  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ . The  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\epsilon$  enzymes are found in the nucleus, and the  $\gamma$  form occurs in mitochondria.

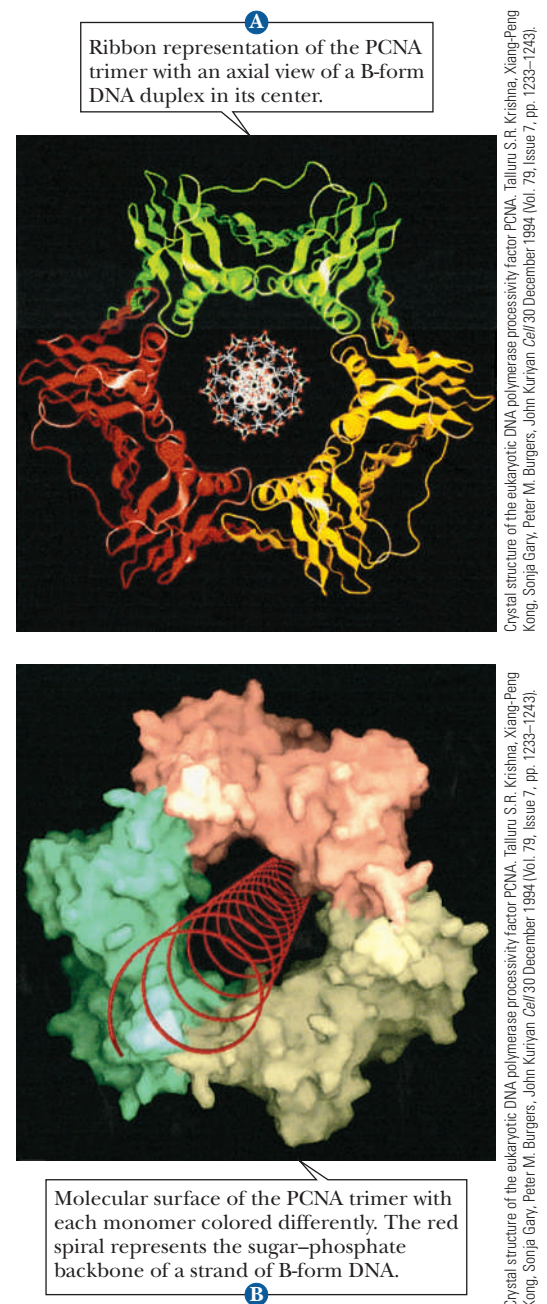
### How are eukaryotic polymerases related to prokaryotic ones?

Polymerase  $\alpha$  was the first discovered, and it has the most subunits. It also has the ability to make primers, but it lacks a 3'  $\rightarrow$  5' proofreading activity and has low processivity. After making the RNA primer, Pol  $\alpha$  adds about 20 nucleotides and is then replaced by Pol  $\delta$  and  $\epsilon$ . Polymerase  $\delta$  is the principal DNA polymerase in eukaryotes. It interacts with a special protein called *PCNA* (for *proliferating cell nuclear antigen*). PCNA is the eukaryotic equivalent of the part of Pol III that functions as a sliding clamp ( $\beta$ ). It is a trimer of three identical proteins that surround the DNA (Figure 10.21). The role of DNA polymerase  $\epsilon$  is less clear, but recent evidence suggests it is involved in leading strand replication. It may replace polymerase  $\delta$  in lagging strand synthesis. DNA polymerase  $\beta$  appears to be a repair enzyme. DNA polymerase  $\gamma$  carries out DNA replication in mitochondria. Several of the DNA polymerases isolated from animals lack exonuclease activity (the  $\alpha$  and  $\beta$  enzymes). In this regard, the animal enzymes differ from prokaryotic DNA polymerases. Separate exonucleolytic enzymes exist in animal cells.

**TABLE 10.4**

The Biochemical Properties of Eukaryotic DNA Polymerases					
	$\alpha$	$\delta$	$\epsilon$	$\beta$	$\gamma$
Mass (kDa)					
Native	>250	170	256	36-38	160-300
Catalytic core	165-180	125	215	36-38	125
Other subunits	70, 50, 60	48	55	None	35, 47
Location	Nucleus	Nucleus	Nucleus	Nucleus	Mitochondria
Associated functions					
3' $\rightarrow$ 5' exonuclease	No	Yes	Yes	No	Yes
Primase	Yes	No	No	No	No
Properties					
Processivity	Low	High	High	Low	High
Fidelity	High	High	High	Low	High
Replication	Yes	Yes	Yes	No	Yes
Repair	No	?	Yes	Yes	No

Source: Adapted from Kornberg, A., and Baker, T. A., 1992. *DNA Replication*, 2nd ed. New York: W. H. Freeman and Co.



**FIGURE 10.21 Structure of the PCNA homotrimer.** Note that the trimeric PCNA ring of eukaryotes is remarkably similar to its prokaryotic counterpart, the dimeric  $\beta$  sliding clamp (Figure 10.7). (Adapted from Figure 3 in Krishna, T. S., et al., 1994. *Crystal Structure of the Eukaryotic DNA Polymerase Processivity Factor PCNA*. *Cell* 79: 1233–1243.)

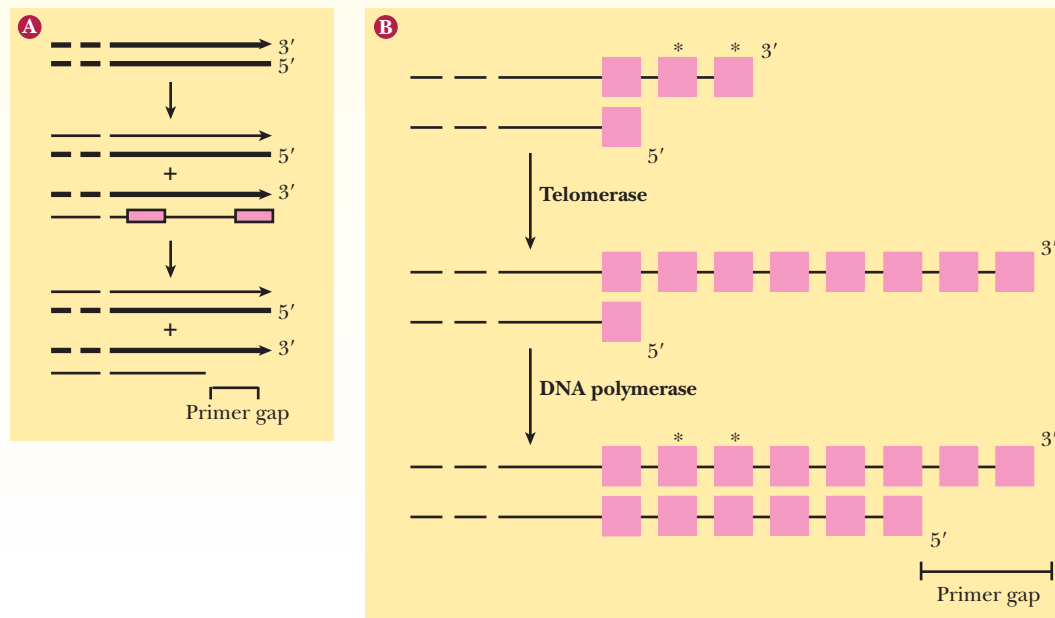
# Biochemical Connections ALLIED HEALTH

## Telomerase and Cancer

Replication of linear DNA molecules poses particular problems at the ends of the molecules. Remember that at the 5' end of a strand of DNA being synthesized, there is initially a short RNA primer, which must later be removed and replaced by DNA. This is never a problem with a circular template because the DNA polymerase I that is coming from the 5' side of the primer (the previous Okazaki fragment) can then patch over the RNA with DNA. However, with a linear chromosome, this is not possible. At each end is a 3' and a 5' DNA chain. The 5'-end template strand is not a problem because a DNA polymerase copying it is moving from 5' to 3' and can proceed to the end of the chromosome from the last RNA primer. The 3'-end template strand poses a problem, however—see part (a) of the figure. The RNA primer at the 5' end of the new strand (shown in green on the opposite page) does not have any way of being replaced. Remember that all DNA polymerases require a primer and, because nothing is upstream (to the 5' side), the RNA primer cannot be replaced with DNA. RNA is unstable, and the RNA primer is degraded in time. In effect, unless some special mechanism is created, the linear molecule gets shorter each time it is replicated.

The ends of eukaryotic chromosomes have a special structure called a telomere, which is a series of repeated DNA sequences. In human sperm-cell and egg-cell DNA, the sequence is 5'TTAGGG3', and this sequence is repeated more than 1000 times at the end of the chromosomes. This repetitive DNA is non-coding and acts as a buffer against degradation of the DNA sequence at the ends, which would occur with each replication as the RNA primers are degraded. Some evidence shows a relationship between longevity and telomere length, and some researchers have suggested that the loss of the telomere DNA with age is part of the natural aging process. Eventually, the DNA becomes nonviable and the cell dies.

However, even with long telomeres, cells eventually die when their DNA gets shorter with each replication unless there is some compensatory mechanism. The creative solution is an enzyme called telomerase, which provides a mechanism for synthesis of the telomeres—see part (b) of the figure. The enzyme telomerase is a ribonuclear protein, containing a section of RNA that is the complement of the telomere. In humans, this sequence is 5'CCCUAA3'. Telomerase binds to the 5' strand at



■ **Telomere replication.** (a) In replication of the lagging strand, short RNA primers are added (pink) and extended by DNA polymerase. When the RNA primer at the 5' end of each strand is removed, there is no nucleotide sequence to read in the next round of DNA replication. The result is a gap (primer gap) at the 5' end of each strand (only one end of a chromosome is shown in this figure). (b) Asterisks indicate sequences at the 3' end that cannot be copied by conventional DNA replication. Synthesis of telomeric DNA by telomerase extends the 5' ends of DNA strands, allowing the strands to be copied by normal DNA replication.

## The Eukaryotic Replication Fork

The general features of DNA replication in eukaryotes are similar to those in prokaryotes. Table 10.5 summarizes the differences. As with prokaryotes, DNA replication in eukaryotes is semiconservative. There is a leading strand with continuous synthesis in the 5' → 3' direction and a lagging strand with discontinuous synthesis in the 5' → 3' direction. An RNA primer is formed by a specific enzyme in eukaryotic DNA replication, as is the case with prokaryotes, but in this case the primase activity is associated with Pol  $\alpha$ . The structures involved at the eukaryotic replication fork are shown in Figure 10.22. The formation of

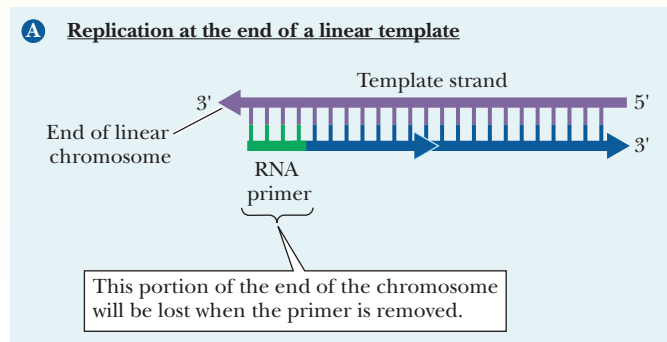


## Biochemical Connections (CONTINUED)

the chromosome end and uses a **reverse transcriptase** activity to synthesize DNA (shown in red) on the 3' strand, using its own RNA as the template. This allows the template strand (shown in purple) to be elongated, effectively lengthening the telomere.

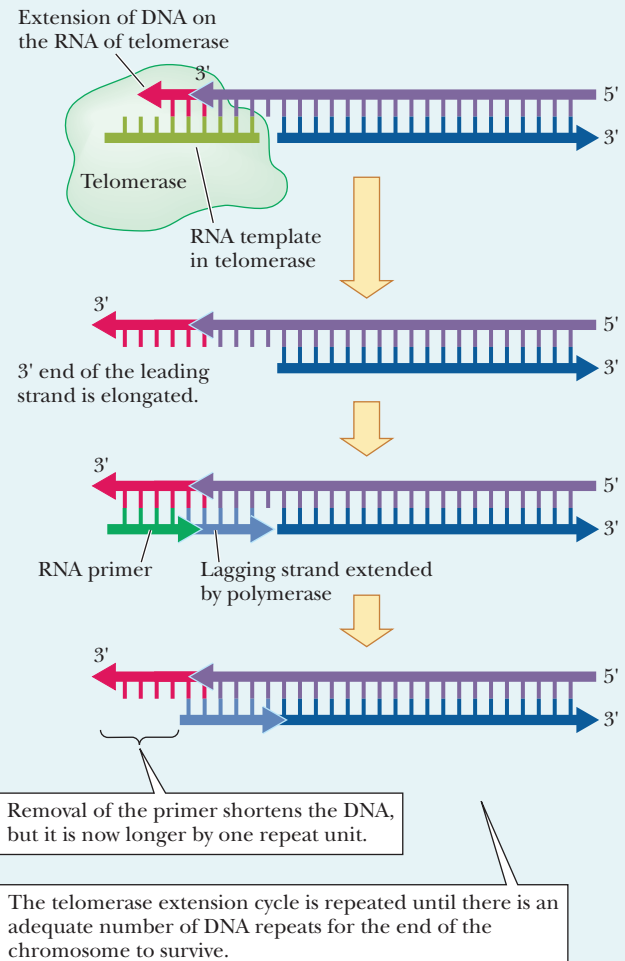
When the nature of telomerase was discovered, researchers originally believed that it was a “fountain of youth” and that, if we could figure out how to keep it going, cells (and perhaps individuals) would never die. Very recent work has shown that even though the enzyme telomerase *does* remain active in rapidly growing tissues such as blood cells, the intestinal lumen, skin, and others, it is *not* active in most adult tissues. When the cells of most adult tissues divide, for replacement or for repair, they do not preserve the chromosome ends. Eventually, enough DNA is gone, a vital gene is lost, and the cell dies. This may be a part of the normal aging and death process.

The big surprise was the discovery that telomerase is reactivated in cancer cells, explaining in part their immortality and their ability to keep dividing rapidly. This observation has opened a new possibility for cancer therapy: if we can prevent the reactivation of telomerase in cancerous tissues, the cancer might die of natural causes. The study of telomerase is just the tip of the iceberg. Other mechanisms must exist to protect the integrity of chromosomes besides telomerase. Using techniques described in Chapter 13, researchers have genetically engineered mice to lack telomerase. These mice showed continued shortening of their telomeres with successive replication and generations, but eventually the chromosome shortening stopped, indicating that some other process also conserved the length of the chromosomes.



Currently, the relationship between telomeres, recombination, and DNA repair is being studied.

**B A mechanism by which telomerase may work. (In this case, RNA of the telomerase acts as a template for reverse transcription)**



Okazaki fragments (typically 150 to 200 nucleotides long in eukaryotes) is initiated by Pol  $\alpha$ . After the RNA primer is made and a few nucleotides are added by Pol  $\alpha$ , the polymerase dissociates and is replaced by Pol  $\delta$  and its attached PCNA protein. Another protein, called *RFC* (replication factor C), is involved in attaching PCNA to Pol  $\delta$ . The RNA primer is eventually degraded, but, in the case of eukaryotes, the polymerases do not have the 5'  $\rightarrow$  3' exonuclease activity to do it. Instead, separate enzymes, FEN-1 and RNase H1, degrade the RNA. Continued movement of Pol  $\delta$  fills in the gaps made by primer removal. As with prokaryotic replication, topoisomerases relieve the torsional strain

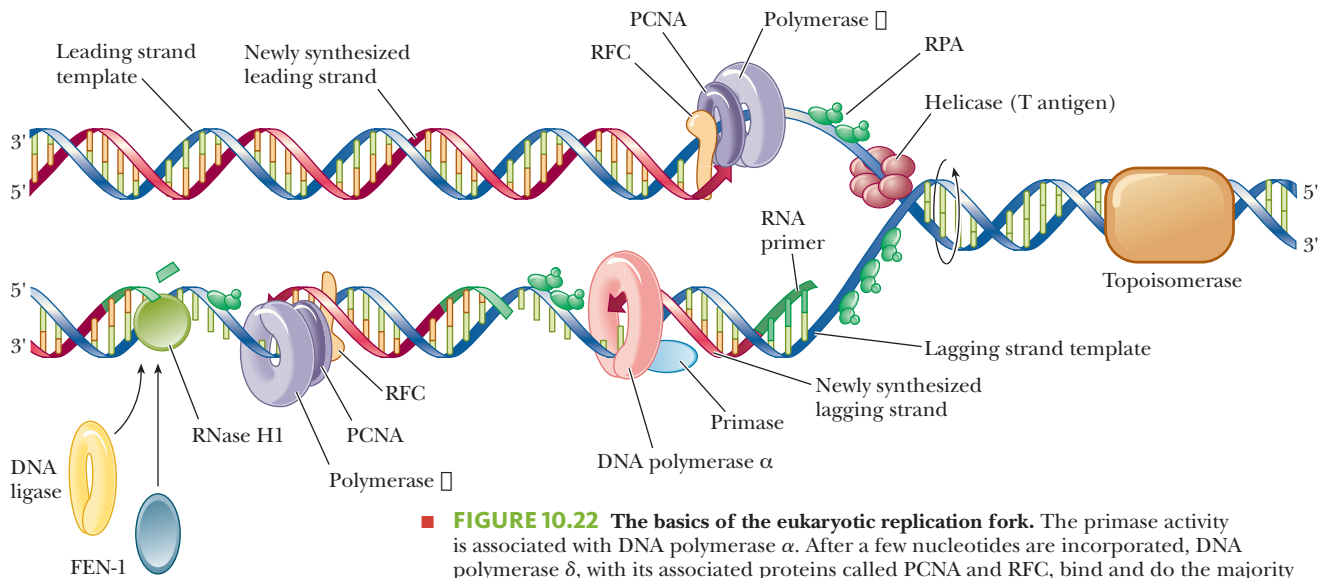


**TABLE 10.5**

Differences in DNA Replication in Prokaryotes and Eukaryotes	
Prokaryotes	Eukaryotes
Five polymerases (I, II, III, IV, V)	Five polymerases ( $\alpha$ , $\beta$ , $\gamma$ , $\delta$ , $\epsilon$ )
Functions of polymerase: I is involved in synthesis, proof-reading, repair, and removal of RNA primers II is also a repair enzyme III is the main polymerizing enzyme  IV, V are repair enzymes under unusual conditions	Functions of polymerases: $\alpha$ is a polymerizing enzyme  $\beta$ is a repair enzyme $\gamma$ is involved in mitochondrial DNA synthesis  $\delta$ is the main polymerizing enzyme $\epsilon$ is the leading strand replication enzyme
Polymerases are also exonucleases	Not all polymerases are exonucleases
One origin of replication	Several origins of replication
Okazaki fragments 1000–2000 residues long	Okazaki fragments 150–200 residues long
No proteins complexed to DNA	Histones complexed to DNA

from unwinding the helix, and a single-stranded binding protein, called RPA, protects the DNA from degradation. Finally, DNA ligase seals the nicks that separate the fragments.

Another important difference between DNA replication in prokaryotes and in eukaryotes is that prokaryotic DNA is not complexed to histones, as is eukaryotic DNA. Histone biosynthesis occurs at the same time and at the same rate as DNA biosynthesis. In eukaryotic replication, histones are associated with DNA as it is formed. An important aspect of DNA replication in eukaryotes, specifically affecting humans, is described in the Biochemical Connections box on pages 274 and 275.



**FIGURE 10.22** The basics of the eukaryotic replication fork. The primase activity is associated with DNA polymerase  $\alpha$ . After a few nucleotides are incorporated, DNA polymerase  $\delta$ , with its associated proteins called PCNA and RFC, bind and do the majority of the synthesis. The enzymes FEN-1 and RNase H1 degrade the RNA primers in eukaryotic replication. (From Cellular and Molecular Biology by Karp, Figure 13–22. Used by permission of John Wiley & Sons, Inc.)

# Biochemical Connections EVOLUTIONARY BIOLOGY

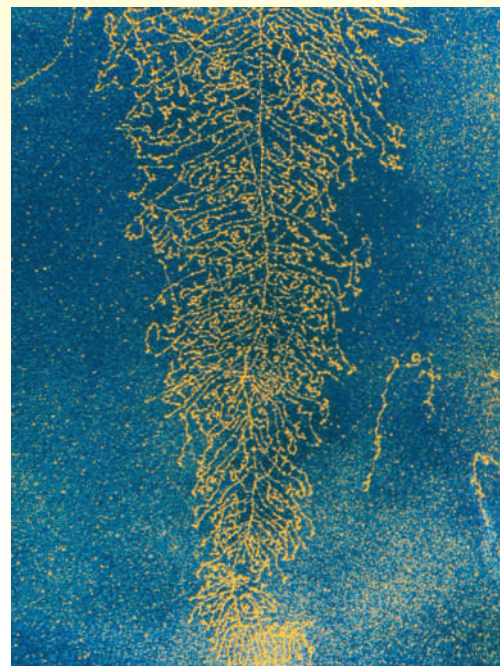
## Self-Replicating RNAs

While this chapter has dealt with DNA replication, RNA replication is also a hot topic in biochemistry. Proponents of the “RNA world” hypothesis believe that RNA was the original molecule of heredity, the first that took simple compounds and turned them into larger molecules having a function. To demonstrate how evolution could have started, scientists have been looking for a way to bridge the gap between the simple molecules that were available before life on earth began and those molecules that could reproduce, metabolize, form cells, etc. In essence, scientists are looking for ways to create life in a test tube. If scientists can create artificial life, then it becomes much easier to imagine how it could have happened naturally with billions of years of trial and error.

In January of last year, Gerald Joyce and Tracey Lincoln published work showing that they had created a series of small RNA molecules that could reproduce, change, and compete for limited resources, exactly the traits that Charles Darwin envisioned at the species level when he proposed the *survival of the fittest*. In their study they started with 24 RNA variants, all of which reproduced, but at different rates. Sometimes the reproduction was not faithful and mutations were made. Some of the mutations survived better than the originals and replaced them. They let the system run for 100 hours, during which time they saw an amplification of the replicating molecules by a factor of  $10^{23}$ . The original molecules died out and were replaced by other molecules, essentially demonstrating chemical evolution *in vitro*.

The studies are very exciting to evolutionary biologists, but do not, by themselves, prove evolution. The definition of life includes many aspects, and reproductive success is only one of them. It is a big step from finding that RNA molecules can reproduce and change to finding that they can develop metabolism and form

cells. The next step will be to demonstrate that as the RNA molecules evolve, they can develop new abilities and functions. This is an area of active research, with many scientists trying to find the missing pieces.



Science Photo Lib./CMSP

## SUMMARY

**How did scientists figure out that replication is semiconservative?** *E. coli* bacteria were grown with  $^{15}\text{NH}_4\text{Cl}$  as the sole nitrogen source. In such a medium, all newly formed nitrogen compounds, including purine and pyrimidine nucleobases, become labeled with  $^{15}\text{N}$ . The  $^{15}\text{N}$ -labeled cells were then transferred to a medium that contained only  $^{14}\text{N}$ . With every new generation of growth, a sample of DNA was extracted and analyzed by density-gradient. DNA containing a 50–50 mixture of  $^{14}\text{N}$  and  $^{15}\text{N}$  appeared at a position halfway between the two bands after one generation, a result to be expected with semiconservative replication.

**In which direction does replication go?** Replication goes in both directions from the origin of replication. That is to say, each strand of DNA is replicated, and from each origin, two replication forks move in opposite directions.

**How can replication proceed along the DNA if the two strands are going in opposite directions?** Different modes of polymerization exist for the two growing strands. One newly formed strand (the leading strand) is formed continuously from its 5' end to its 3' end at the replication fork on the exposed

3'-to-5' template strand. The other strand (the lagging strand) is formed semidiscontinuously in small fragments, called Okazaki fragments. The fragments of the lagging strand are linked together by DNA ligase.

**How does replication work with supercoiled DNA?** DNA gyrase is used to induce negative supercoils in the DNA to compensate for the positive supercoils that form in front of the replication fork where the strands are separated. The action of DNA gyrase reduces the torsional strain on the DNA.

**How is single-stranded DNA protected long enough for replication?** Specific proteins called single-strand binding proteins bind to the single-stranded regions and protect them from nucleases.

**Where does the primer come from?** The enzyme primase uses a DNA strand as a template and creates a complementary RNA strand that is the primer for DNA synthesis.

**How does proofreading improve replication fidelity?** Spontaneous mutation of bases and insertion of the wrong nucleotide would normally lead to an error every  $10^4$  to  $10^5$  bases.

However, the proofreading capability of the DNA polymerases allows mismatched nucleotides to be removed, reducing the errors to one in  $10^9$  to  $10^{10}$ .

**How is replication tied to cell division?** Replication is tied to cell division by several proteins including the origin recognition complex, replication activator protein, and replication licensing factors. The process is controlled by cyclins, proteins produced during the  $G_1$  and S phases that bind to cyclin-dependent kinases and activate replication.

## REVIEW EXERCISES

**OWL** Interactive versions of these problems are assignable in OWL

### 10.1 The Flow of Genetic Information in the Cell

1. **Recall** Define *replication*, *transcription*, and *translation*.
2. **Reflect and Apply** Is the following statement true or false? Why? “The flow of genetic information in the cell is always DNA  $\rightarrow$  RNA  $\rightarrow$  protein.”
3. **Reflect and Apply** Why is it more important for DNA to be replicated accurately than transcribed accurately?

### 10.2 Replication of DNA

4. **Recall** Why is the replication of DNA referred to as a semiconservative process? What is the experimental evidence for the semiconservative nature of the process? What experimental results would you expect if replication of DNA were a conservative process?
5. **Recall** What is a replication fork? Why is it important in replication?
6. **Recall** Describe the structural features of an origin of replication.
7. **Recall** Why is it necessary to unwind the DNA helix in the replication process?
8. **Reflect and Apply** In the Meselson–Stahl experiment that established the semiconservative nature of DNA replication, the extraction method produced short fragments of DNA. What sort of results might have been obtained with longer pieces of DNA?
9. **Reflect and Apply** Suggest a reason why it would be unlikely for replication to take place without unwinding the DNA helix.

### 10.3 DNA Polymerase

10. **Recall** Do DNA-polymerase enzymes also function as exonucleases?
11. **Recall** Compare and contrast the properties of the enzymes DNA polymerase I and polymerase III from *E. coli*.
12. **Reflect and Apply** Define *processivity*, and indicate the importance of this concept in DNA replication.
13. **Reflect and Apply** Comment on the dual role of the monomeric reactants in replication.
14. **Reflect and Apply** What is the importance of pyrophosphatase in the synthesis of nucleic acids?
15. **Reflect and Apply** DNA synthesis always takes place from the 5' to the 3' end. The template strands have opposite directions. How does nature deal with this situation?
16. **Reflect and Apply** What would happen to the replication process if the growing DNA chain did not have a free 3' end?
17. **Reflect and Apply** Suggest a reason for the rather large energy “overkill” in inserting a deoxyribonucleotide into a growing DNA molecule. (About  $15 \text{ kcal mol}^{-1}$  is used in forming a phosphate ester bond that actually requires only about a third as much energy.)
18. **Reflect and Apply** Why is it not surprising that the addition of nucleotides to a growing DNA chain takes place by nucleophilic substitution?

### How are eukaryotic polymerases similar to prokaryotic ones?

Five polymerases are present in eukaryotes, compared to three in prokaryotes. They are labeled  $\alpha$  through  $\epsilon$ . The main polymerizer is the  $\delta$  version, which is similar to Pol III in prokaryotes. The various polymerases vary in size, complexity, processivity, and their levels of exonuclease and repair activities. The  $\gamma$  version is found only in mitochondria, but the others are found in the nucleus.

19. **Reflect and Apply** Is it unusual that the  $\beta$ -subunits of DNA polymerase III that form a sliding clamp along the DNA do not contain the active site for the polymerization reaction? Explain your answer.

### 10.4 Proteins Required for DNA Replication

20. **Recall** List the substances required for replication of DNA catalyzed by DNA polymerase.
21. **Recall** Describe the discontinuous synthesis of the lagging strand in DNA replication.
22. **Recall** What are the functions of the gyrase, primase, and ligase enzymes in DNA replication?
23. **Recall** Single-stranded regions of DNA are attacked by nucleases in the cell, yet portions of DNA are in a single-stranded form during the replication process. Explain.
24. **Recall** Describe the role of DNA ligase in the replication process.
25. **Recall** What is the primer in DNA replication?
26. **Reflect and Apply** How does the replication process take place on a supercoiled DNA molecule?
27. **Reflect and Apply** Why is a short RNA primer needed for replication?

### 10.5 Proofreading and Repair

28. **Recall** How does proofreading take place in the process of DNA replication?
29. **Recall** Does proofreading always take place by the same process in replication?
30. **Recall** Describe the excision repair process in DNA, using the excision of thymine dimers as an example.
31. **Biochemical Connections** Of what benefit is it for DNA to have thymine rather than uracil?
32. **Reflect and Apply** Your book contains about 2 million characters (letters, spaces, and punctuation marks). If you could type with the accuracy with which the prokaryote *E. coli* incorporates, proofreads, and repairs bases in replication (about one uncorrected error in  $10^9$  to  $10^{10}$  bases), how many such books would you have to type before an uncorrected error is “permitted”? (Assume that the error rate is one in  $10^{10}$  bases.)
33. **Reflect and Apply** *E. coli* incorporates deoxyribonucleotides into DNA at a rate of 250 to 1000 bases per second. Using the higher value, translate this into typing speed in words per minute. (Assume five characters per word, using the typing analogy from Question 32.)
34. **Reflect and Apply** Given the typing speed from Question 33, how long must you type, nonstop, at the fidelity shown by *E. coli* (see Question 32) before an uncorrected error would be permitted?
35. **Reflect and Apply** Can methylation of nucleotides play a role in DNA replication? If so, what sort of role?

- 36. **Reflect and Apply** How can breakdown in DNA repair play a role in the development of human cancers?
- 37. **Biochemical Connections** Can prokaryotes deal with drastic DNA damage in ways that are not available to eukaryotes?

### 10.6 DNA Recombination

- 38. **Recall** What is homologous recombination?
- 39. **Recall** How did Messelson and Weigle demonstrate recombination?
- 40. **Reflect and Apply** How has the use of DNA labeled with heavy isotopes been instrumental in our understanding of replication?
- 41. **Recall** What is the Holliday Model?

### 10.7 Eukaryotic DNA Replication

- 42. **Recall** Do eukaryotes have fewer origins of replication than prokaryotes, or more origins, or the same number?
- 43. **Recall** How does DNA replication in eukaryotes differ from the process in prokaryotes?
- 44. **Recall** What role do histones play in DNA replication?
- 45. **Reflect and Apply** (a) Eukaryotic DNA replication is more complex than prokaryotic replication. Give one reason why this should be so. (b) Why might eukaryotic cells need more kinds of DNA polymerases than bacteria?

- 46. **Reflect and Apply** How do the DNA polymerases of eukaryotes differ from those of prokaryotes?
- 47. **Reflect and Apply** What is the relationship between control of DNA synthesis in eukaryotes and the stages of the cell cycle?
- 48. **Biochemical Connections** What would be the effect on DNA synthesis if the telomerase enzyme were inactivated?
- 49. **Reflect and Apply** Would it be advantageous to a eukaryotic cell to have histone synthesis take place at a faster rate than DNA synthesis?
- 50. **Reflect and Apply** What are replication licensing factors? How did they get their name?
- 51. **Reflect and Apply** Is DNA synthesis likely to be faster in prokaryotes or in eukaryotes?
- 52. **Reflect and Apply** Outline a series of steps by which reverse transcriptase produces DNA on an RNA template.
- 53. **Biochemical Connections** Name an important difference in the replication of circular DNA versus linear double-stranded DNA.
- 54. **Reflect and Apply** Why is it reasonable that eukaryotes have a DNA polymerase (Pol  $\gamma$ ) that operates only in mitochondria?
- 55. **Biochemical Connections** What is meant by the “RNA world”?
- 56. **Biochemical Connections** Why are scientists excited to find that RNA molecules can be made that are self-replicating?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





# Transcription of the Genetic Code: The Biosynthesis of RNA

# 11

In transcription, the template strand of DNA is used to produce a complementary strand of RNA.

## 11.1 Overview of Transcription

As we saw in Chapter 10, the central dogma of molecular biology is that DNA makes RNA, and RNA makes proteins. The process of making RNA from DNA is called **transcription**, and it is the major control point in the expression of genes and the production of proteins.

In the use of genetic information, one of the strands of the double-stranded DNA molecule is transcribed into a complementary sequence of RNA. The RNA sequence differs from DNA in one respect: The DNA base thymine (T) is replaced by the RNA base uracil (U). Of all the DNA in a cell, only some is transcribed. Transcription produces all the types of RNA—mRNA, tRNA, rRNA, snRNA, miRNA, and siRNA. New types of RNA and new functions for it are found every year.

The details of RNA transcription differ somewhat in prokaryotes and eukaryotes. For example, the process is much more complicated in eukaryotes, involving a large number of transcription factors. Most of the research on the subject has been done in prokaryotes, especially *E. coli*, but some general features are found in all organisms except in the case of cells infected by RNA viruses.

### What are the basics common to all transcription?

Although many differences exist between transcription in prokaryotes and eukaryotes, and even differences between transcription of different types of RNA in eukaryotes, some aspects are constant. Table 11.1 summarizes the main features of the process.

TABLE 11.1

#### General Features of RNA Synthesis

1. RNA is initially synthesized using a DNA template in the process called transcription; the enzyme that catalyzes the process is **DNA-dependent RNA polymerase**.
2. All four ribonucleoside triphosphates (ATP, GTP, CTP, and UTP) are required, as is  $Mg^{2+}$ .
3. A primer is not needed in RNA synthesis, but a DNA template is required.
4. As is the case with DNA biosynthesis, the RNA chain grows from the 5' to the 3' end. The nucleotide at the 5' end of the chain retains its triphosphate group (abbreviated ppp).
5. The enzyme uses one strand of the DNA as the template for RNA synthesis. The base sequence of the DNA contains signals for initiation and termination of RNA synthesis. The enzyme binds to the template strand and moves along it in the 3'-to-5' direction.
6. The template is unchanged.

## Chapter Outline

### 11.1 Overview of Transcription

- What are the basics common to all transcription?

### 11.2 Transcription in Prokaryotes

- What do the subunits of RNA polymerase do?
- Which of the DNA strands is used in transcription?
- How does RNA polymerase know where to begin transcription?

### 11.3 Transcription Regulation in Prokaryotes

- How is transcription controlled in prokaryotes?
- What is the difference between an enhancer and a promoter?
- How does repression work in the *lac* operon?
- How are RNA secondary structures involved in transcription attenuation?

### 11.4 Transcription in Eukaryotes

- How does Pol II recognize the correct DNA to transcribe?
- What do eukaryotic transcription factors do?

### 11.5 Transcription Regulation in Eukaryotes

- How do response elements work?

### 11.6 Non-Coding RNAs

- What are Micro RNAs?
- What are Small Interfering RNAs?
- What is RNA Silencing?

### 11.7 Structural Motifs in DNA-Binding Proteins

- What are DNA-binding domains?
- What are transcription-activation domains?

### 11.8 Posttranscriptional RNA Modification

- Why is mRNA modified after initial transcription?
- How are introns spliced out to make mature RNA?

### 11.9 Ribozymes

- What are the characteristics of a ribozyme?

**OWL** Online homework for this chapter may be assigned in OWL.

## 11.2 Transcription in Prokaryotes

### RNA Polymerase in *Escherichia coli*

The enzyme that synthesizes RNA is called **RNA polymerase**, and the most extensively studied one was isolated from *E. coli*. The molecular weight of this enzyme is about 470,000 Da, and it has a multisubunit structure. Five different types of subunits, designated  $\alpha$ ,  $\omega$ ,  $\beta$ ,  $\beta'$ , and  $\sigma$ , have been identified. The actual composition of the enzyme is  $\alpha_2\omega\beta\beta'\sigma$ . The  $\sigma$ -subunit is rather loosely bound to the rest of the enzyme (the  $\alpha_2\omega\beta\beta'$  portion), which is called the **core enzyme**. The **holoenzyme** consists of all the subunits, including the  $\sigma$ -subunit.

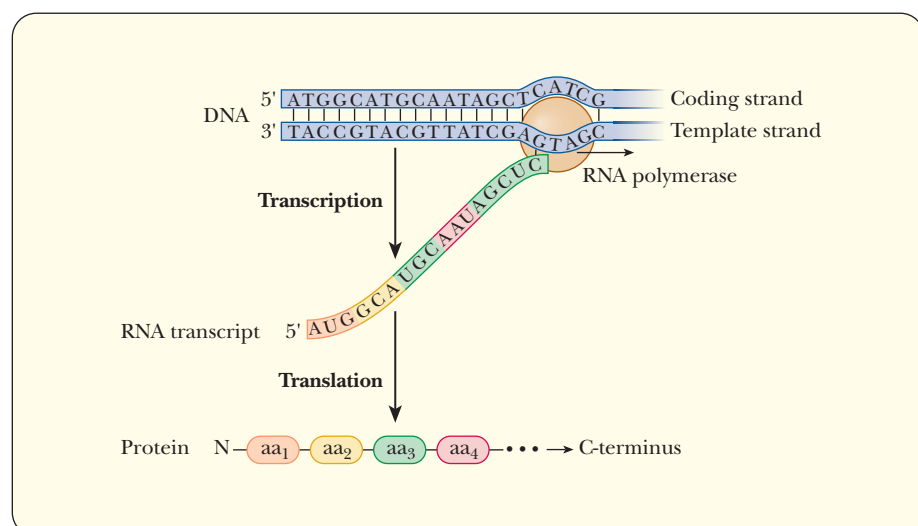
### What do the subunits of RNA polymerase do?

The  $\sigma$ -subunit is involved in the recognition of specific promoters, whereas the  $\beta$ -,  $\beta'$ -,  $\alpha$ -, and  $\omega$ -subunits combine to make the active site for polymerization. The mechanism of the polymerization reaction is an active area of research.

### Which of the DNA strands is used in transcription?

Figure 11.1 shows the basics of information transfer from DNA to protein. Of the two strands of DNA, one of them is the template for RNA synthesis of a particular RNA product. RNA polymerase reads it from 3' to 5'. This strand has several names. The most common is the **template strand**, because it is the strand that directs the synthesis of the RNA. It is also called the **antisense strand**, because its code is the complement of the RNA that is produced. It is sometimes called the **(-) strand** by convention. The other strand is called the **coding strand** because its sequence of DNA will be the same as the RNA sequence that is produced (with the exception of U replacing T). It is also called the **sense strand**, because the RNA sequence is the sequence that we use to determine what amino acids are produced in the case of mRNA. It is also called the **(+) strand** by convention, or even the **nontemplate strand**. For our purposes, we will use the terms *template strand* and *coding strand* throughout. Because the DNA in the coding strand has the same sequence as the RNA that is produced, it is used when discussing the sequence of genes for proteins or for promoters and controlling elements on the DNA.

The core enzyme of RNA polymerase is catalytically active but lacks specificity. The core enzyme alone would transcribe both strands of DNA when only one strand contains the information in the gene. The holoenzyme of RNA polymerase binds to specific DNA sequences and transcribes only the correct



■ **FIGURE 11.1** The basics of transcription. RNA polymerase uses the template strand of DNA to make an RNA transcript that has the same sequence as the nontemplate DNA strand, with the exception that T is replaced by U. If this RNA is mRNA, it can later be translated to protein.

strand. The essential role of the  $\sigma$ -subunit is recognition of the **promoter locus** (a DNA sequence that signals the start of RNA transcription; see Section 11.3). The loosely bound  $\sigma$ -subunit is released after transcription begins and about 10 nucleotides have been added to the RNA chain. Prokaryotes can have more than one type of  $\sigma$ -subunit. The nature of the  $\sigma$ -subunit can direct RNA polymerases to different promoters and cause the transcription of various genes to reflect different metabolic conditions.

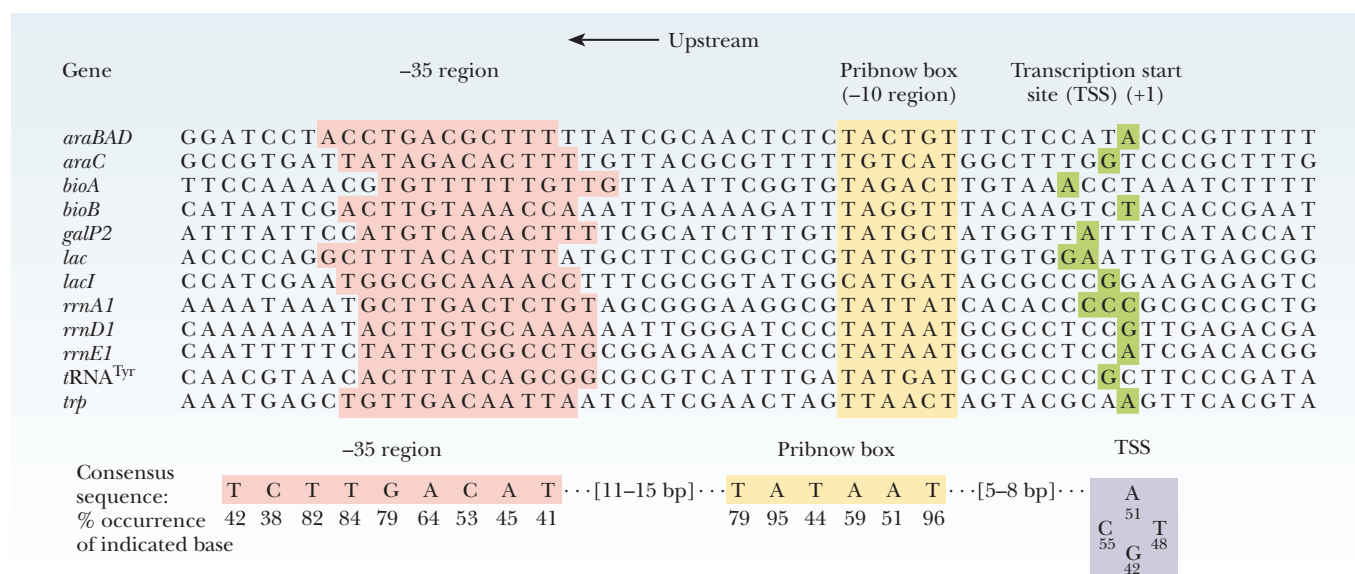
## Promoter Structure

Even the simplest organisms contain a great deal of DNA that is not transcribed. RNA polymerase must have a way of knowing which of the two strands is the template strand, which part of the strand is to be transcribed, and where the first nucleotide of the gene to be transcribed is located.

### How does RNA polymerase know where to begin transcription?

Promoters are DNA sequences that provide direction for RNA polymerase. The promoter region to which RNA polymerase binds is closer to the 3' end of the template strand than is the actual gene for the RNA to be synthesized. The RNA is formed from the 5' end to the 3' end, so the polymerase moves along the template strand from the 3' end to the 5' end. However, by convention, all control sequences are given for the coding strand, which is 5' to 3'. The binding site for the polymerase is said to lie *upstream* of the start of transcription, which is farther to the 5' side of the coding strand. This is often a source of confusion to students, and it is important to remember the correct orientation. The promoter sequence will be given based on the coding strand, even though the RNA polymerase is actually binding to the template strand. Promoters are upstream, which means to the 5' side of the coding strand and to the 3' side of the template strand.

Most bacterial promoters have at least three components. Figure 11.2 shows some typical promoter sequences for *E. coli* genes. The component closest to the first nucleotide to be incorporated is about 10 bases upstream. Also by convention, the first base to be incorporated into the RNA chain is said to be at



**FIGURE 11.2** Sequences of representative promoters from *E. coli*. By convention, these are given as the sequence that would be found on the coding strand going from left to right as the 5' to 3' direction. The numbers below the consensus sequences indicate the percentage of the time that a certain position is occupied by the indicated nucleotide.

position +1 and is called the **transcription start site (TSS)**. All the nucleotides upstream from this start site are given negative numbers.

Because the first promoter element is about 10 bases upstream, it is called the  $-10$  region, but is also called the **Pribnow box** after its discoverer. After the Pribnow box, there are 16 to 18 bases that are completely variable. The next promoter element is about 35 bases upstream of the TSS and is simply called the  **$-35$  region** or  **$-35$  element**. An element is a general term for a DNA sequence that is somehow important in controlling transcription. The area from the  $-35$  element to the TSS is called the **core promoter**. Upstream of the core promoter can be an **UP element**, which enhances the binding of RNA polymerase. UP elements usually extend from  $-40$  to  $-60$ . The region from the end of the UP element to the transcription start site is known as the **extended promoter**.

The base sequence of promoter regions has been determined for a number of prokaryotic genes, and a striking feature is that they contain many bases in common. These are called **consensus sequences**. Promoter regions are A–T rich, with two hydrogen bonds per base pair; consequently, they are more easily unwound than G–C-rich regions, which have three hydrogen bonds per base pair. Figure 11.2 shows the consensus sequences for the  $-10$  and  $-35$  regions.

Even though the  $-10$  and  $-35$  regions of many genes are similar, there are also some significant variations that are important to the metabolism of the organism. Besides directing the RNA polymerase to the correct gene, the promoter base sequence controls the frequency with which the gene is transcribed. Some promoters are strong, and others are weak. A strong promoter binds RNA polymerase tightly, and the gene therefore is transcribed more often. In general, as a promoter sequence varies from the consensus sequence, the binding of RNA polymerase becomes weaker.

### Chain Initiation

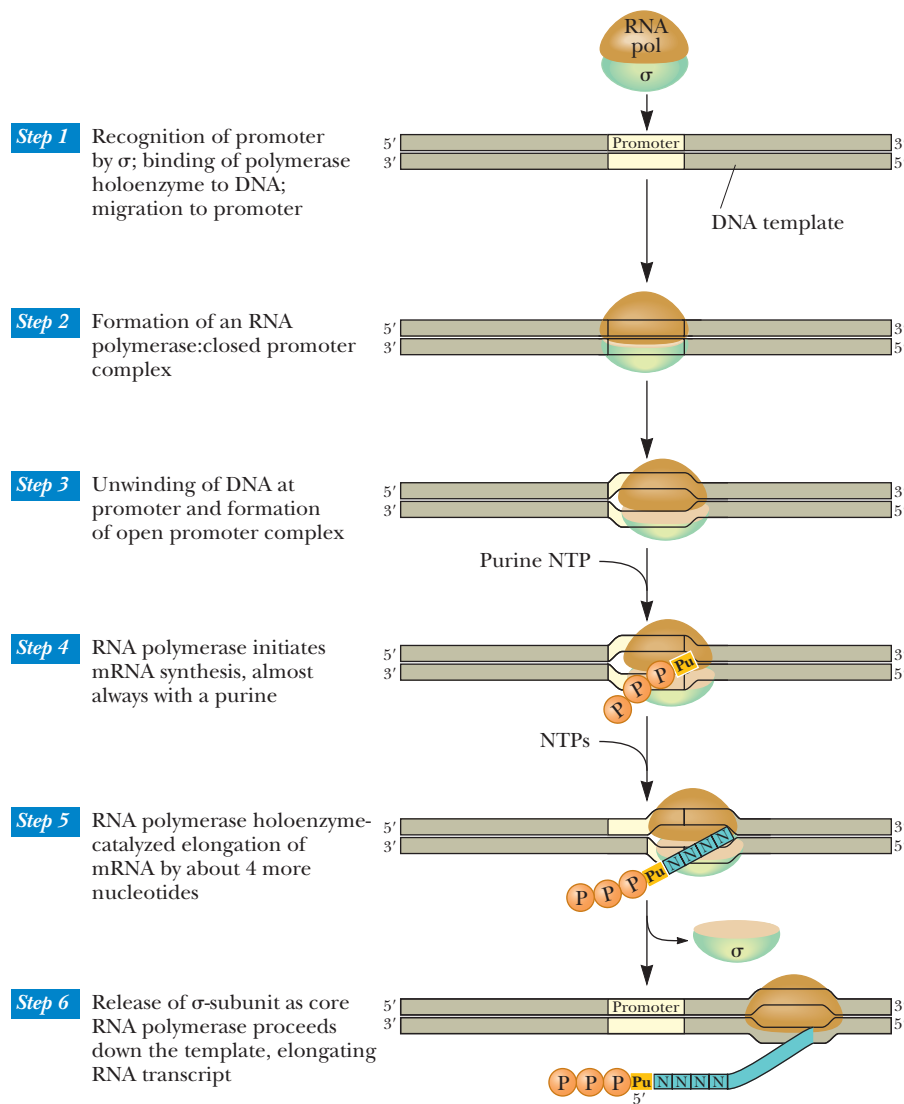
The process of transcription (and translation as well, as we will see in Chapter 12) is usually broken down into phases for easier studying. The first phase of transcription is called **chain initiation**, and it is the part of transcription that has been studied the most. It is also the part that is the most controlled.

Chain initiation begins when RNA polymerase (RNA pol) binds to the promoter and forms what is called the **closed complex** (Figure 11.3). The  $\sigma$ -subunit directs the polymerase to the promoter. It bridges the  $-10$  and  $-35$  regions of the promoter to the RNA polymerase core via a flexible “flap” in the  $\sigma$ -subunit. Core enzymes lacking the  $\sigma$ -subunit bind to areas of DNA that lack promoters. The holoenzyme may bind to “promoterless” DNA, but it dissociates without transcribing.

Chain initiation requires formation of the **open complex**. Recent studies show that a portion of the  $\beta'$  and the  $\sigma$ -subunits initiate strand separation, melting about 14 base pairs surrounding the transcription start site. A purine ribonucleoside triphosphate is the first base in RNA, and it binds to its complementary DNA base at position +1. Of the purines, A tends to occur more often than G. This first residue retains its 5'-triphosphate group (indicated by ppp in Figure 11.3).

### Chain Elongation

After the strands have separated, a transcription bubble of about 17 base pairs moves down the DNA sequence to be transcribed (Figure 11.3), and RNA polymerase catalyzes the formation of the phosphodiester bonds between the incorporated ribonucleotides. When about 10 nucleotides have been incorporated, the  $\sigma$ -subunit dissociates and is later recycled to bind to another RNA polymerase core enzyme.



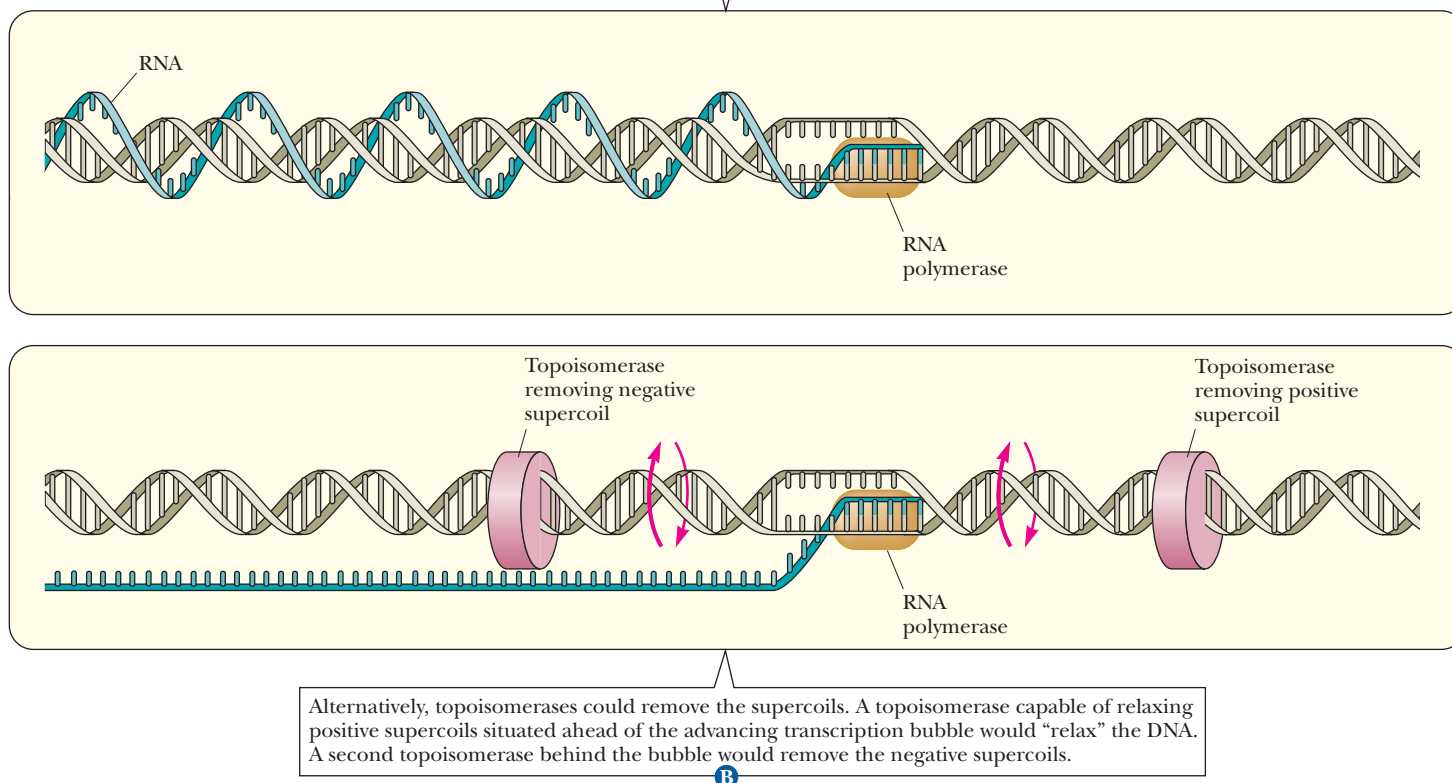
**FIGURE 11.3** Sequence of events in the initiation and elongation phases of transcription as it occurs in prokaryotes. Nucleotides in this region are numbered with reference to the base at the transcription start site, which is designated +1.

The transcription process supercoils DNA, with negative supercoiling upstream of the transcription bubble and positive supercoiling downstream, as shown in Figure 11.4. Topoisomerases relax the supercoils in front of and behind the advancing transcription bubble. The rate of chain elongation is not constant. The RNA polymerase moves quickly through some DNA regions and slowly through others. It may pause for as long as one minute before continuing.

Instead of finishing every RNA chain, RNA polymerase actually releases most chains near the beginning of the process, after about 5 to 10 nucleotides have been assembled, in a process called *abortive transcription*. The cause of abortive transcription is the failure of RNA polymerase to break its own bonds to the promoter via the  $\sigma$ -subunit. Studying this process has led to the current model of the mechanism of transcription. In order for chain elongation to occur, the RNA polymerase must be able to launch itself off the promoter. Given the tight binding between the  $\sigma$ -subunit and the promoter, this requires substantial energy. Figure 11.5 shows the current model of the open complex that allows progression into chain elongation. The RNA polymerase is bound tightly to the DNA promoter. It “scrunches” the DNA into itself, causing torsional strain of the separated DNA strands. Like a bow loading up with potential energy as the bowstring is pulled, this provides the energy to allow the polymerase to break free.

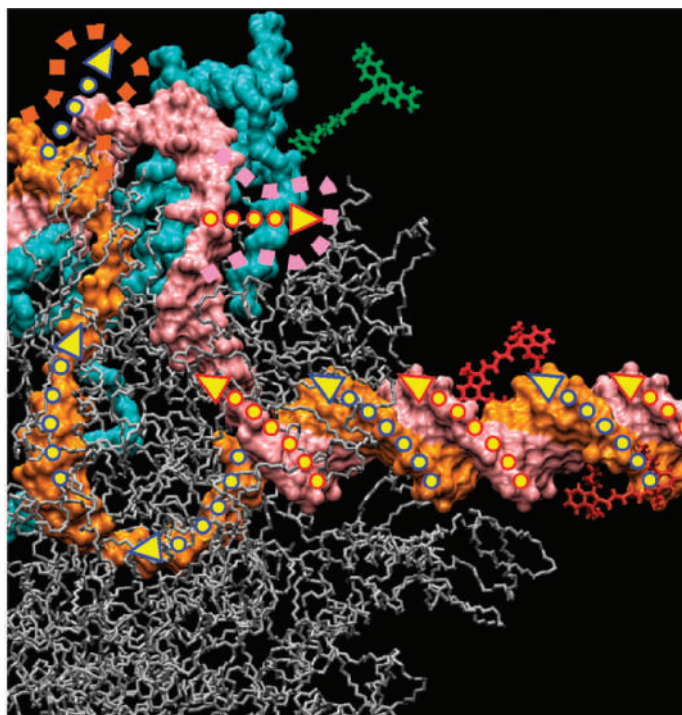


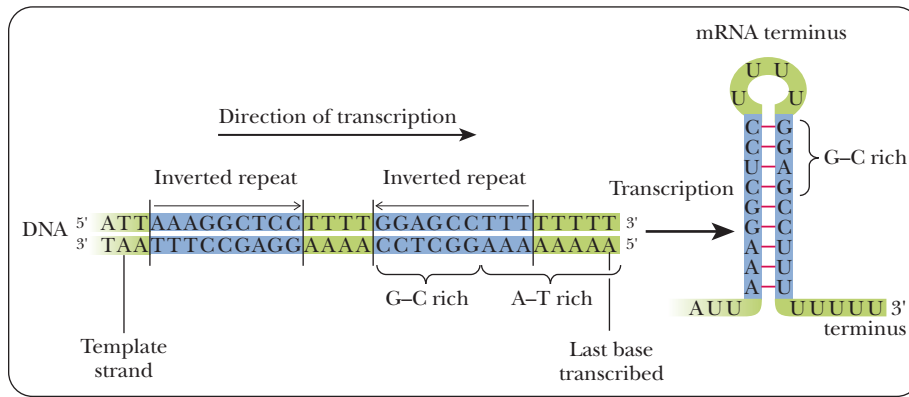
**A**  
If the RNA polymerase followed the template strand around the axis of the DNA duplex, there would be no strain, and no supercoiling of the DNA would occur, but the RNA chain would be wrapped around the double helix once every 10 base pairs. This possibility seems unlikely because it would be difficult to disentangle the transcript from the DNA complex.



**FIGURE 11.4** Two models for transcription elongation. (Adapted from Futcher, B., 1988. *Supercoiling and Transcription, or Vice Versa?* Trends in Genetics 4, 271–272. Used by permission of Elsevier Science.)

- FIGURE 11.5** Ready to scrunch. The model shows an open promoter complex of bacterial RNA polymerase (RNAP) poised to begin RNA synthesis, with arrows designating motions of DNA segments that occur as the first nucleotides are combined to form an RNA chain. Downstream DNA (on the right) rotates inward and separates in the active site channel of the polymerase. Template (orange) and coding (pink) DNA strands follow the indicated paths, moving as the RNA chain (not shown) is polymerized on the template DNA strand. Sites of extrusion of single-stranded DNA are shown by outlined arrows at the top.  $\sigma^{70}$  (blue); RNAP core chains (gray). Fluorescent tags shown in red and green were used to monitor the movement that demonstrated the "scrunching" phenomenon. (Courtesy of Achillefs Kapandis [University of Oxford], Shimon Weiss [University of California, Los Angeles], and Richard H. Ebright [Howard Hughes Medical Institute, Rutgers University].)





■ **FIGURE 11.6 Inverted repeats terminate transcription.** Inverted repeats in the DNA sequence being transcribed can lead to an mRNA molecule that forms a hairpin loop. This is often used to terminate transcription.

### Chain Termination

Termination of RNA transcription also involves specific sequences *downstream* of the actual gene for the RNA to be transcribed. There are two types of termination mechanisms. The first is called **intrinsic termination**, and it is controlled by specific sequences called **termination sites**. The termination sites are characterized by two inverted repeats spaced by a few other bases (Figure 11.6). Inverted repeats are sequences of bases that are complementary, such that they can loop back on themselves. The DNA then encodes a series of uracils. When the RNA is created, the inverted repeats form a hairpin loop. This tends to stall the advancement of RNA polymerase. At the same time, the presence of the uracils causes a series of A–U base pairs between the template strand and the RNA. A–U pairs are weakly hydrogen-bonded compared with G–C pairs, and the RNA dissociates from the transcription bubble, ending transcription.

The other type of termination involves a special protein called *rho* ( $\rho$ ). Rho-dependent termination sequences also cause a hairpin loop to form. In this case, the  $\rho$  protein binds to the RNA and chases the polymerase, as shown in Figure 11.7. When the polymerase transcribes the RNA that forms a hairpin loop (not shown in the figure), it stalls, giving the  $\rho$  protein a chance to catch up. When the  $\rho$  protein reaches the termination site, it facilitates the dissociation of the transcription machinery. The movement of the  $\rho$  protein and the dissociation require ATP.

## 11.3 Transcription Regulation in Prokaryotes

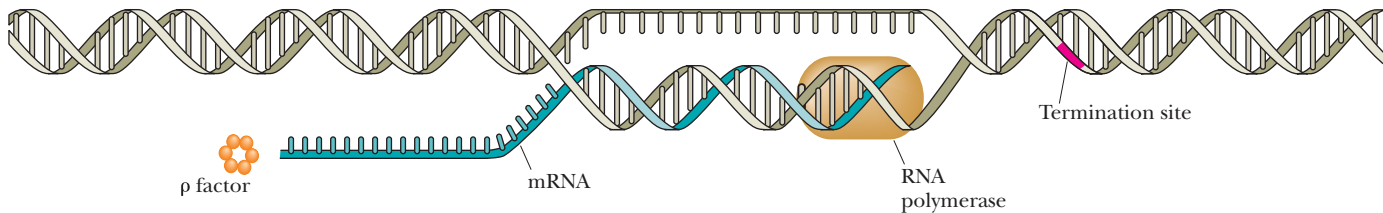
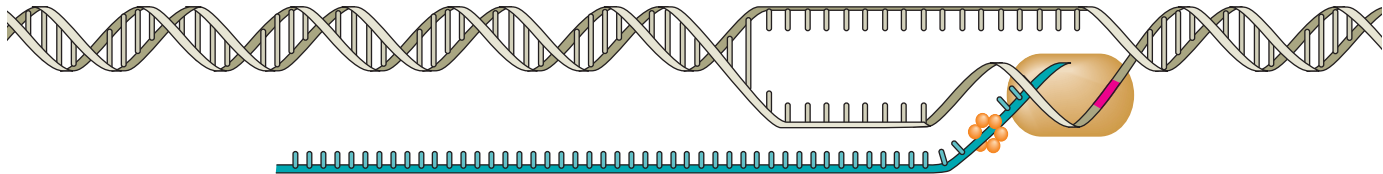
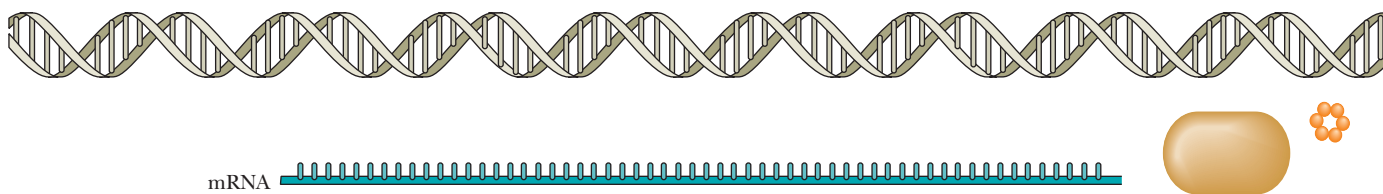
Although many RNAs and proteins are produced in even a simple prokaryotic cell, not all of them are produced at the same time or in the same quantities. In prokaryotes, the control of transcription is largely responsible for controlling the level of protein production. In fact, many equate transcription control with gene expression.

### How is transcription controlled in prokaryotes?

In prokaryotes, transcription is controlled in four principal ways—alternative  $\sigma$  factors, enhancers, operons, and transcription attenuation. They will be discussed in turn.

### Alternative $\sigma$ Factors

Viruses and bacteria can exert some control over which genes are expressed by producing different  $\sigma$ -subunits that direct the RNA polymerase to different genes. A classic example of how this works is the action of phage SPO1, a virus that infects the bacteria *Bacillus subtilis*. The virus has a set of genes

**A** The rho-factor mechanism of transcription termination.**B** Rho factor attaches to a recognition site on mRNA and moves it along behind RNA polymerase.**C** When RNA polymerase pauses at the termination site, rho factor unwinds the DNA:RNA hybrid in the transcription bubble...**D** ...releasing the nascent mRNA.

**FIGURE 11.7** The rho-factor mechanism of transcription termination. Rho ( $\rho$ ) factor (a) attaches to a recognition site on mRNA and (b) moves along it behind RNA polymerase. (c) When RNA polymerase pauses at the termination site, rho factor unwinds the DNA:RNA hybrid in the transcription bubble, releasing the nascent mRNA (d).

called the *early genes*, which are transcribed by the host's RNA polymerase, using its regular  $\sigma$ -subunit (Figure 11.8). One of the viral early genes codes for a protein called *gp28*. This protein is actually another  $\sigma$ -subunit, which directs the RNA polymerase to transcribe preferentially more of the viral genes during the *middle phase*. Products of the middle phase transcription are *gp33* and *gp34*, which together make up another  $\sigma$  factor that directs the transcription of the *late genes*. Remember that  $\sigma$  factors are recycled. Initially, the *B. subtilis* uses the standard  $\sigma$  factor. As more and more of the *gp28* is produced, it competes for binding with standard  $\sigma$  for the RNA polymerase, eventually subverting the transcription machinery for the virus instead of the bacterium.

Another example of alternative  $\sigma$  factors is seen in the response of *E. coli* to heat shock. The normal  $\sigma$ -subunit in this species is called  $\sigma^{70}$  because it has a molecular weight of 70,000 Da. When *E. coli* are grown at higher temperatures than their optimum, they produce another set of proteins in response. Another  $\sigma$  factor, called  $\sigma^{32}$ , is produced. It directs the RNA polymerase to bind to different promoters that are not normally recognized by  $\sigma^{70}$ .

## Enhancers

Certain *E. coli* genes include sequences upstream of the extended promoter region. The genes for ribosomal RNA production have three upstream sites, called *Fis* sites because they are binding sites for the protein called Fis (Figure 11.9). These sites extend from the end of the UP element at  $-60$  to  $-150$ , and are examples of a class of DNA sequences called **enhancers**. Enhancers are sequences that can be bound by proteins called **transcription factors**, a class of molecule we will see a lot of in Sections 11.4 and 11.5.

### What is the difference between an enhancer and a promoter?

When a DNA sequence is labeled a promoter, it implies that the RNA polymerase binds to that region of DNA. An enhancer, on the other hand, is a DNA sequence that is usually upstream of the promoter. The polymerase does not bind to enhancers. When enhancers allow a response to changing metabolic conditions, such as temperature shock, they are usually referred to as **response elements**. When binding the transcription factor increases the level of transcription, the element is said to be an enhancer. When binding the transcription factor decreases transcription, the element is said to be a **silencer**. The position and orientation of enhancers is less important than for sequences that are part of the promoter. Molecular biologists can study the nature of control elements by making changes to them. When enhancer sequences are moved from one place on the DNA to another or have their sequences reversed, they still function as enhancers. The study of the number and nature of transcription factors is the most common research in molecular biology these days.

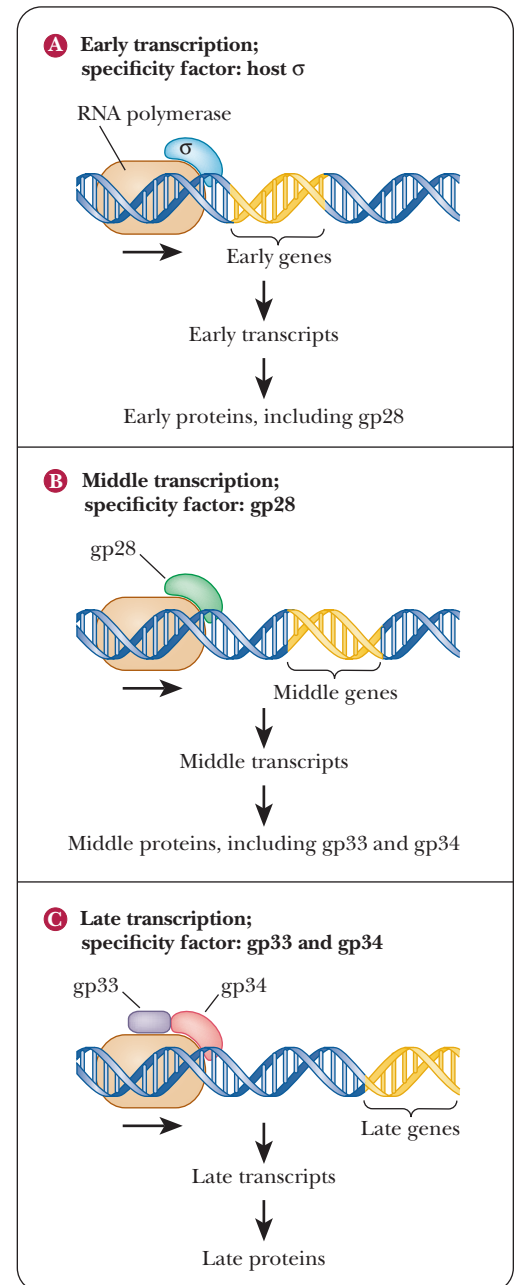
## Operons

In prokaryotes, genes that encode enzymes of certain metabolic pathways are often controlled as a group, with the genes encoding the proteins of the pathway being close together and under the control of a common promoter. Such a group of genes is called an **operon**. Usually these genes are not transcribed all the time. Rather, the production of these proteins can be triggered by the presence of a suitable substance called an **inducer**. This phenomenon is called **induction**. A particularly well-studied example of an inducible protein is the enzyme  $\beta$ -galactosidase in *E. coli*.

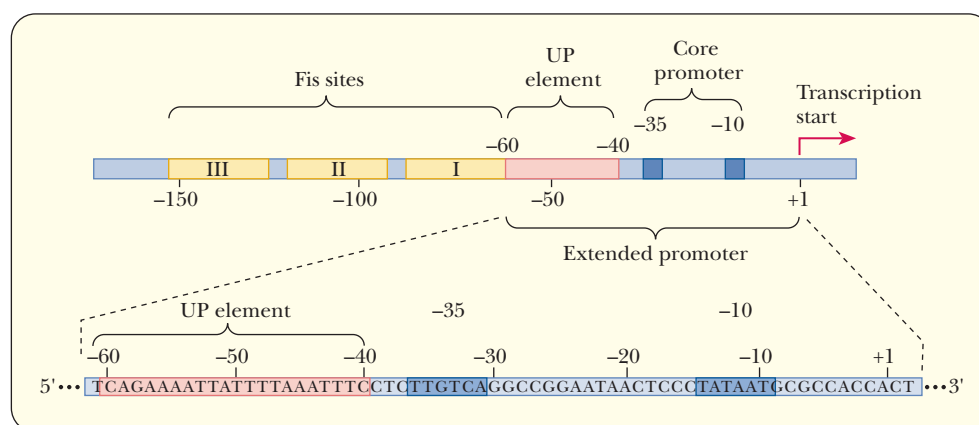
The disaccharide *lactose* (a  $\beta$ -galactoside; Section 16.3) is the substrate of  $\beta$ -galactosidase. The enzyme hydrolyzes the glycosidic linkage between galactose and glucose, the monosaccharides that are the component parts of lactose. *E. coli* can survive with lactose as its sole carbon source. To do so, the bacterium needs  $\beta$ -galactosidase to catalyze the first step in lactose degradation.

The production of  $\beta$ -galactosidase takes place only in the presence of lactose, not in the presence of other carbon sources, such as glucose. A metabolite of lactose, allolactose, is the actual inducer, and  $\beta$ -galactosidase is an *inducible enzyme*.

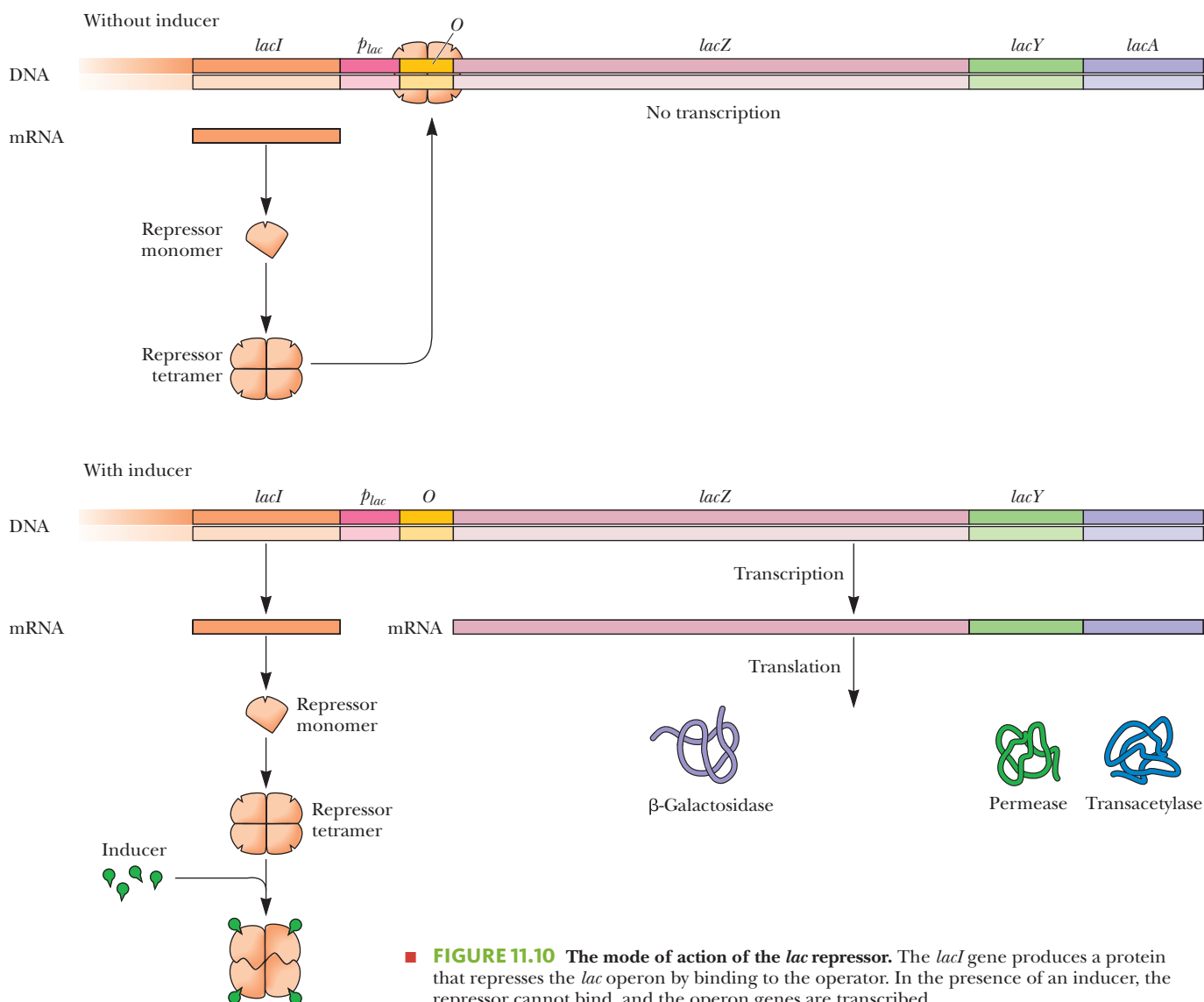
The production of  $\beta$ -galactosidase takes place only in the presence of lactose, not in the presence of other carbon sources, such as glucose. A metabolite of lactose, allolactose, is the actual inducer, and  $\beta$ -galactosidase is an *inducible enzyme*.  $\beta$ -Galactosidase is coded for by a **structural gene** (*lacZ*) (Figure 11.10). Structural genes encode the gene products that are involved in the biochemical pathway of the operon. Two other structural genes are part of the operon. One is *lacY*, which encodes the enzyme lactose permease, which allows lactose to enter the cell. The other is *lacA*, which encodes an enzyme called transacetylase. The function of this last enzyme is not known, but some hypothesize that its role is to inactivate certain antibiotics that may enter the cell through the lactose permease. The expression of these structural genes is in turn under control of a **regulatory gene** (*lacI*), and the mode of operation of the regulatory gene is the most important part of the *lac* operon mechanism. The regulatory gene is responsible for the production of a protein, the **repressor**. As the name indicates, the repressor inhibits the expression of the structural genes. In the presence of the inducer, this inhibition is removed. This is an example



**FIGURE 11.8 Control of transcription via different  $\sigma$  subunits.** (a) When the phage SPO1 infects *B. subtilis*, the host RNA polymerase (tan) and  $\sigma$ -subunit (blue) transcribe the early genes of the infecting viral DNA. One of the early gene products is gp28 (green), an alternative  $\sigma$ -subunit. (b) The gp28 directs the RNA polymerase to transcribe the middle genes, which produces gp33 (purple) and gp34 (red). (c) The gp33 and gp34 direct the host's RNA polymerase to transcribe the late genes. (Adapted by permission from Molecular Biology, by R. F. Weaver, McGraw-Hill, 1999.)



■ **FIGURE 11.9 Elements of a bacterial promoter.** The core promoter includes the -10 and -35 regions. The extended promoter includes the UP element. Upstream of the UP element, there may be enhancers, such as the Fis sites seen in the promoters for genes that code for ribosomal RNA in *E. coli*. The protein Fis is a transcription factor. (Adapted by permission from Molecular Biology, by R. F. Weaver, McGraw-Hill, 1999.)



■ **FIGURE 11.10 The mode of action of the *lac* repressor.** The *lacI* gene produces a protein that represses the *lac* operon by binding to the operator. In the presence of an inducer, the repressor cannot bind, and the operon genes are transcribed.



of **negative regulation** because the *lac* operon is turned on unless something is present to turn it off, which is the repressor in this case.

### How does repression work in the *lac* operon?

The repressor protein that is made by the *lacI* gene forms a tetramer when it is translated. It then binds to a portion of the operon called the **operator** (*O*) (Figure 11.10). When the repressor is bound to the operator, RNA polymerase cannot bind to the adjacent promoter region (*plac*), which facilitates the expression of the structural genes. The operator and promoter together constitute the **control sites**.

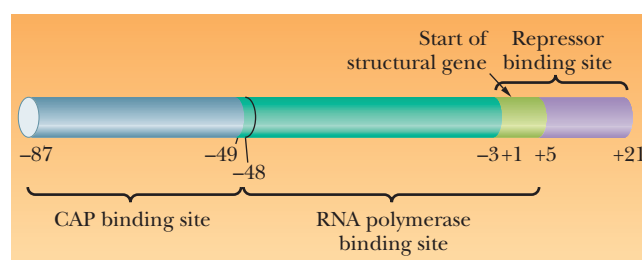
In induction, the inducer binds to the repressor, producing an inactive repressor that cannot bind to the operator (Figure 11.10). Because the repressor is no longer bound to the operator, RNA polymerase can now bind to the promoter, and transcription and translation of the structural genes can take place. The *lacI* gene is adjacent to the structural genes in the *lac* operon, but this need not be the case. Many operons are known in which the regulatory gene is far removed from the structural genes.

The *lac* operon is induced when *E. coli* has lactose, and no glucose, available to it as a carbon source. When both glucose and lactose are present, the cell does not make the *lac* proteins. The repression of the synthesis of the *lac* proteins by glucose is called **catabolite repression**. The mechanism by which *E. coli* recognizes the presence of glucose involves the promoter. The promoter has two regions. One is the binding site for RNA polymerase, and the other is the binding site for another regulatory protein, the **catabolite activator protein (CAP)** (Figure 11.11). The binding site for RNA polymerase also overlaps the binding site for the repressor in the operator region.

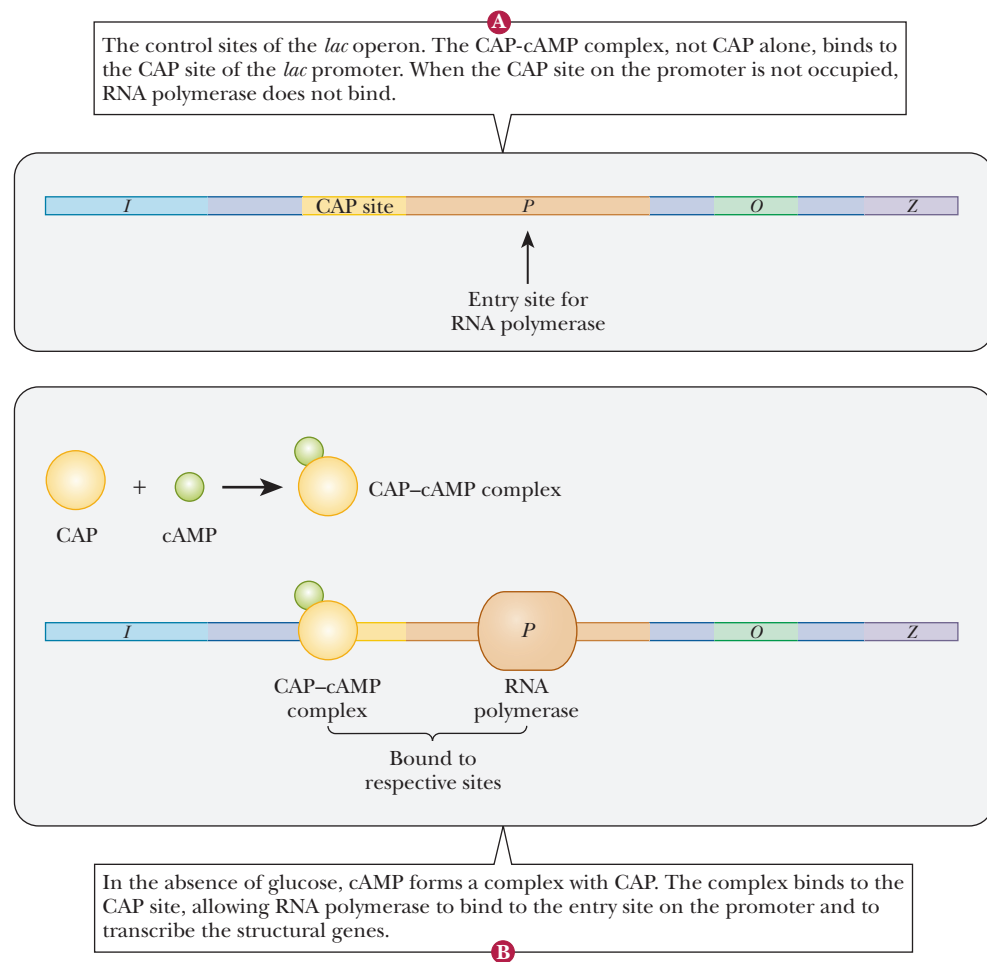
The binding of CAP to the promoter depends on the presence or absence of 3',5'-cyclic AMP (cAMP). When glucose is not present, cAMP is formed, serving as a “hunger signal” for the cell. CAP forms a complex with cAMP. The complex binds to the CAP site in the promoter region. When the complex is bound to the CAP site on the promoter, the RNA polymerase can bind at the binding site available to it and proceed with transcription (Figure 11.12). The *lac* promoter is particularly weak, and RNA polymerase binding is minimal in the absence of the CAP—cAMP complex bound to the CAP site. The CAP site is an example of an enhancer element, and the CAP—cAMP complex is a transcription factor. The modulation of transcription by CAP is a type of **positive regulation**.

When the cell has an adequate supply of glucose, the level of cAMP is low. CAP binds to the promoter only when it is complexed to cAMP. The combination of positive and negative regulation with the *lac* operon means that the presence of lactose is necessary, but not sufficient, for transcription of the operon structural genes. It takes the presence of lactose *and* the absence of glucose for the operon to be active. As we shall see later, many transcription factors and response elements involve the use of cAMP, a common messenger in the cell.

Operons can be controlled by positive or negative regulation mechanisms. They are also classified as **inducible**, **repressible**, or both, depending on how



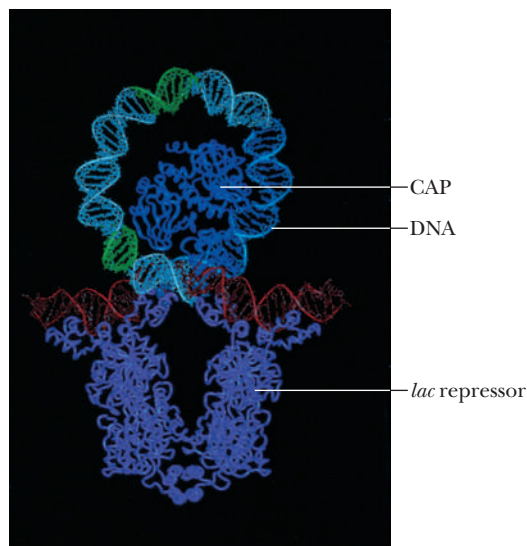
■ **FIGURE 11.11** Binding sites in the *lac* operon. Numbering refers to base pairs. Negative numbers are assigned to base pairs in the regulatory sites. Positive numbers indicate the structural gene, starting with base pair +1. The CAP binding site is seen next to the RNA polymerase binding site.



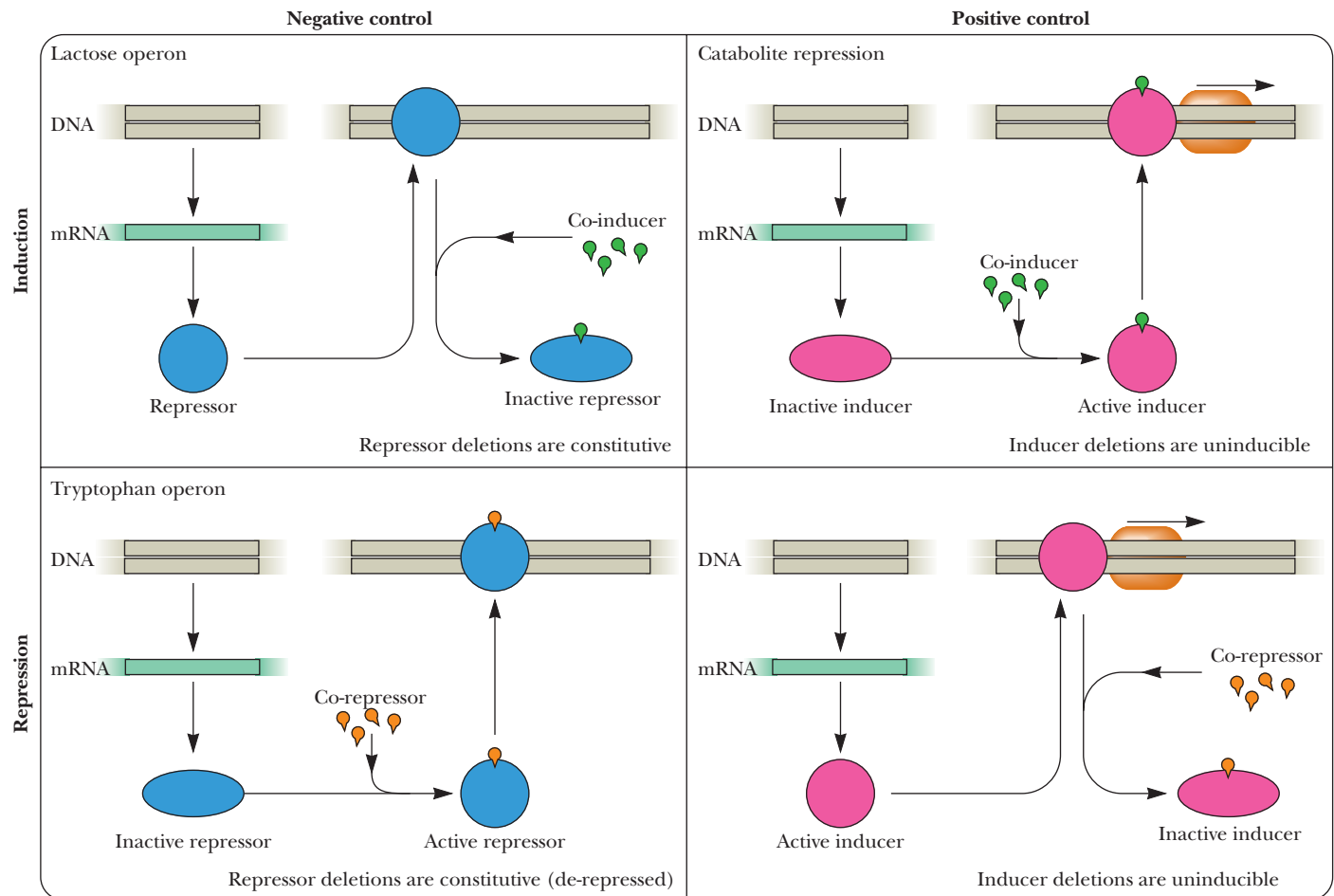
■ **FIGURE 11.12** Catabolite repression.

they respond to the molecules that control their expression. There are four general possibilities, as shown in Figure 11.13. The top left figure shows a negative control system with induction. It is negative control because a repressor protein stops transcription when it binds to the promoter. It is an inducible system because the presence of the inducer or **co-inducer**, as it is often called, releases the repression, as we saw with the *lac* operon. Negative control systems can be identified by the fact that, if the gene for the repressor is mutated in some way that stops the expression of the repressor, the operon is always expressed. Genes that are always expressed are called **constitutive**. The top right figure shows a positively controlled inducible system. The controlling protein is an inducer that binds to the promoter, stimulating transcription, but it works only when bound to its co-inducer. This is what is seen with the catabolite activator protein with the *lac* operon. Such positively controlled systems can be identified by the fact that, if the gene for the inducer is mutated, it cannot be expressed—that is, it is **uninducible**. The bottom left figure shows a negatively controlled repressible system. A repressor stops transcription, but this repressor functions only in the presence of a **co-repressor**. The bottom right figure shows a positively controlled repressible system. An inducer protein binds to the promoter, stimulating transcription; but, in the presence of the co-repressor, the inducer is inactivated.

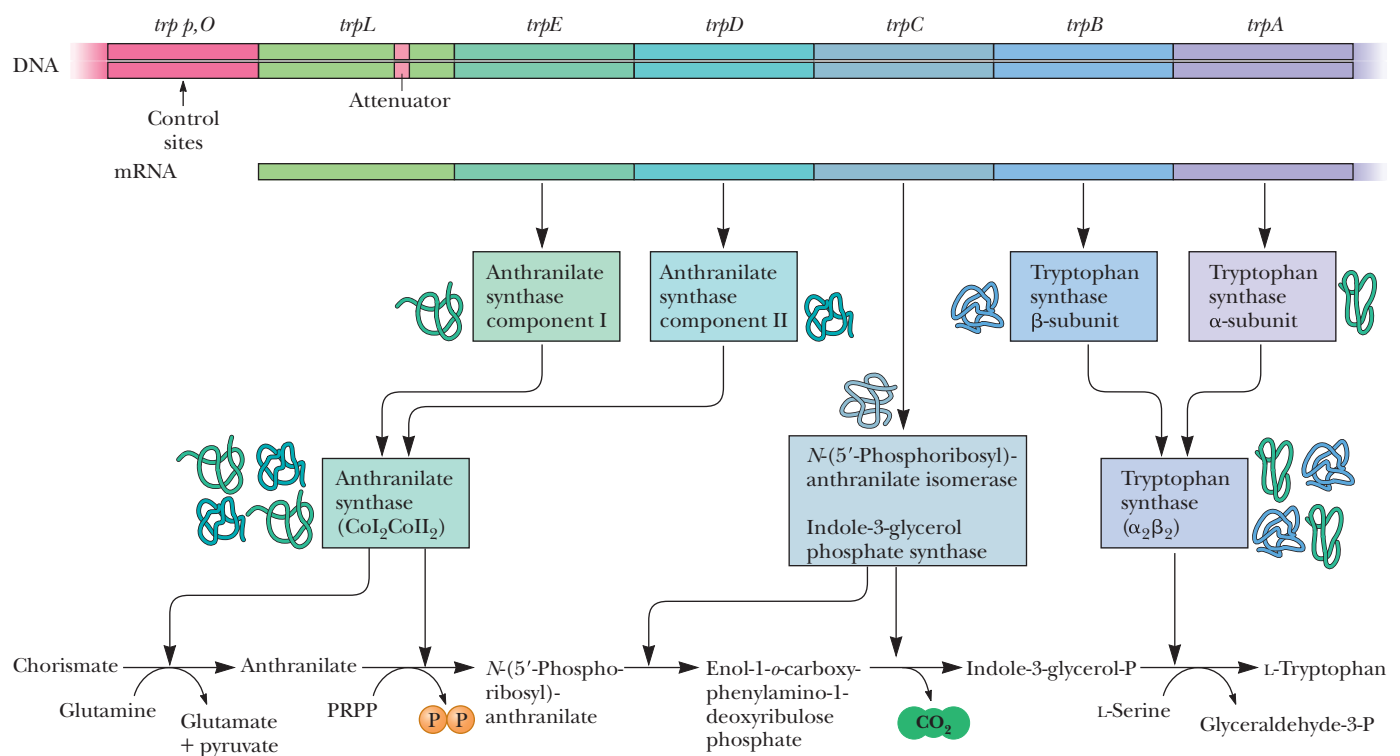
The *trp* operon of *E. coli* codes for a leader sequence (*trpL*) and five polypeptides, *trpE* through *trpA*, as shown in Figure 11.14. The five proteins make up four different enzymes (shown in the three boxes near the bottom of the figure). These enzymes catalyze the multistep process that converts chorismate to tryptophan. Control of the operon is via a repressor protein that binds to



■ **The *lac* repressor and CAP bound to DNA.** (From Mitchell Lewis, et al., (1 March 1996) Science 271 (5253), 1247. Used with permission of AAAS.)



**FIGURE 11.13** Basic control mechanisms seen in the control of genes. They may be inducible or repressible, and they may be positively or negatively controlled.



**FIGURE 11.14** The *trp* operon of *E. coli*.

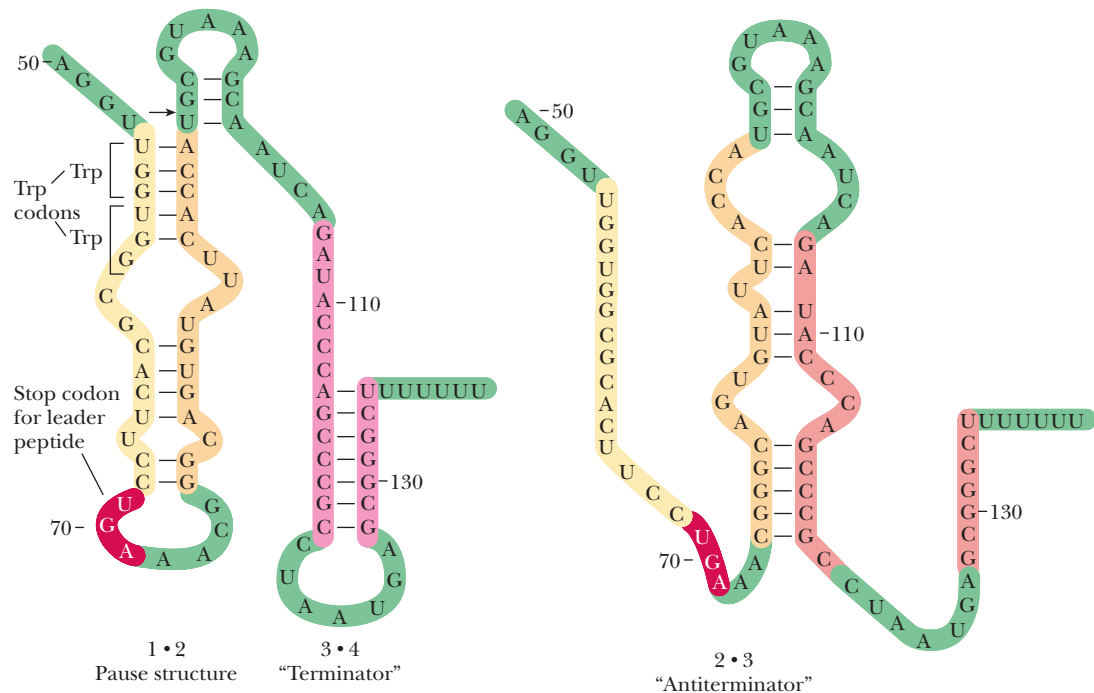
two molecules of tryptophan. When tryptophan is plentiful, this repressor–tryptophan complex binds to the *trp* operator that is next to the *trp* promoter. This binding prevents the binding of RNA polymerase, so the operon is not transcribed. When tryptophan levels are reduced, the repression is lifted because the repressor will not bind to the operator in the absence of the co-repressor, tryptophan. This is an example of a system that is repressible and under negative regulation, as shown in Figure 11.13. The *trp* repressor protein is itself produced by the *trpR* operon and also represses that operon. It is an example of **autoregulation**, because the product of the *trpR* operon regulates its own production.

### Transcription Attenuation

In addition to repression, the *trp* operon is regulated by **transcription attenuation**. This control mechanism works by altering transcription *after* it has begun via transcription termination or pausing. Prokaryotes have no separation of transcription and translation as eukaryotes do, so the ribosomes are attached to the mRNA while it is being transcribed. The *trp* operon's first gene is the *trpL* sequence that codes for a leader peptide. This leader peptide has two key tryptophan residues in it. Translation of the mRNA leader sequence depends on having an adequate supply of tryptophan-charged tRNA (Chapters 9 and 12). When tryptophan is scarce, the operon is translated normally. When it is plentiful, transcription is terminated prematurely after only 140 nucleotides of the leader sequence have been transcribed. Secondary structures formed in the mRNA of the leader sequence are responsible for this effect (Figure 11.15).

#### How are RNA secondary structures involved in transcription attenuation?

Three possible hairpin loops can form in this RNA—the **1•2 pause structure**, the **3•4 terminator**, or the **2•3 antiterminator**. Transcription begins normally



■ **FIGURE 11.15** Alternative secondary structures can form in the leader sequence of mRNA for the *trp* operon. Binding between regions 1 and 2 (yellow and tan) is called a pause structure. Regions 3 and 4 (purple) then form a terminator hairpin loop. Alternative binding between regions 2 and 3 forms an antiterminator structure.

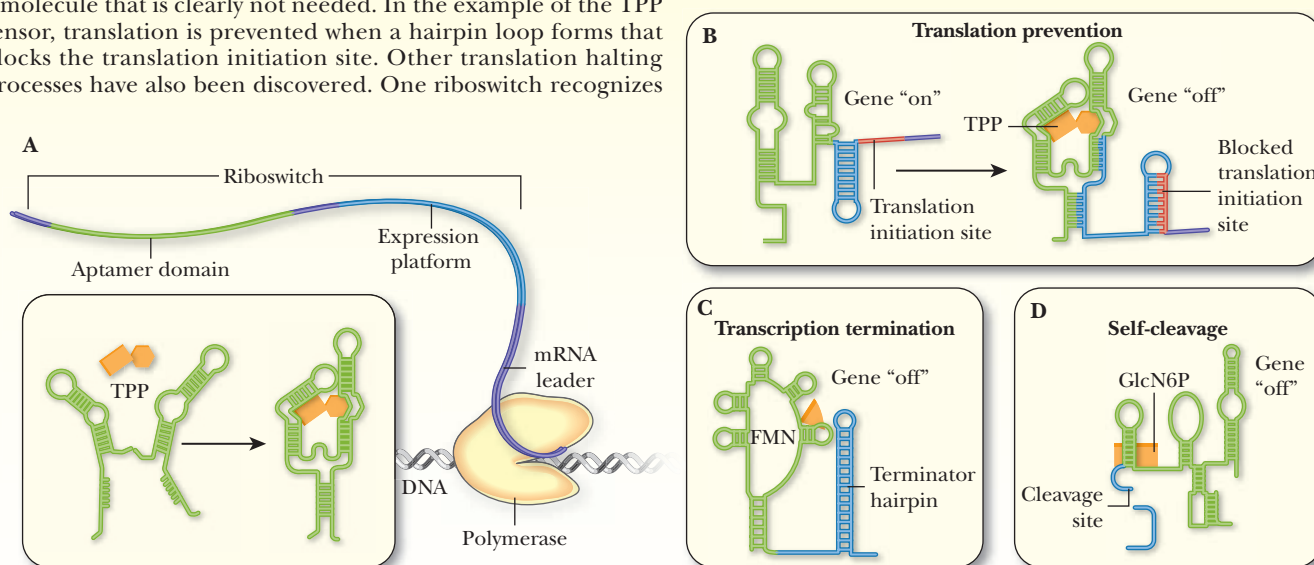
# Biochemical Connections BACTERIOLOGY

## Riboswitches Provide Another Weapon Against Pathogens

Besides the transcription controls discussed in the previous section, another recent discovery is that prokaryotes also have control mechanisms based on the mRNA transcript itself. While studying the metabolism of certain vitamins and cofactors, scientists discovered that the mRNA of pathways involved in making these vitamins could bind the vitamins themselves. The mRNA has two functions—sensing and decision making. Together, these functions are called **riboswitches**. A riboswitch is made up of two parts. At the 5' end is a sensing domain called an **aptamer**. One of the known riboswitches binds to thiamine pyrophosphate (TPP), a vitamin we will see during our study of metabolism (see the figure). The aptamer portion of the mRNA is able to bind to TPP and therefore acts as a sensor for this vitamin. If the riboswitch senses that the vitamin is present, then it can respond in a way that prevents the translation of the mRNA, thereby avoiding a pathway that would produce more of a molecule that is clearly not needed. In the example of the TPP sensor, translation is prevented when a hairpin loop forms that blocks the translation initiation site. Other translation halting processes have also been discovered. One riboswitch recognizes

another vitamin metabolite, flavin mononucleotide (FMN). When FMN is bound to the aptamer, a terminator hairpin is created similar to the one that forms during transcription attenuation. In another example, a riboswitch that recognizes the sugar, glucosamine-6-phosphate (GlcN6P), self-destructs in its presence.

More than a dozen human pathogens have riboswitches as part of their metabolism. Researchers are hoping to find molecules that can act like a competitive inhibitor and fool the riboswitch into acting as though the natural substrate were present. If the riboswitch controls a vital process, then shutting off the riboswitch would kill the pathogen. Riboswitch targets have been identified in many different bacteria that can be harmful to man, including *Bacillus anthracis*, *Hemophilus influenza*, *Helicobacter pylori*, *Salmonella enterica*, and *Streptococcus pneumoniae*.

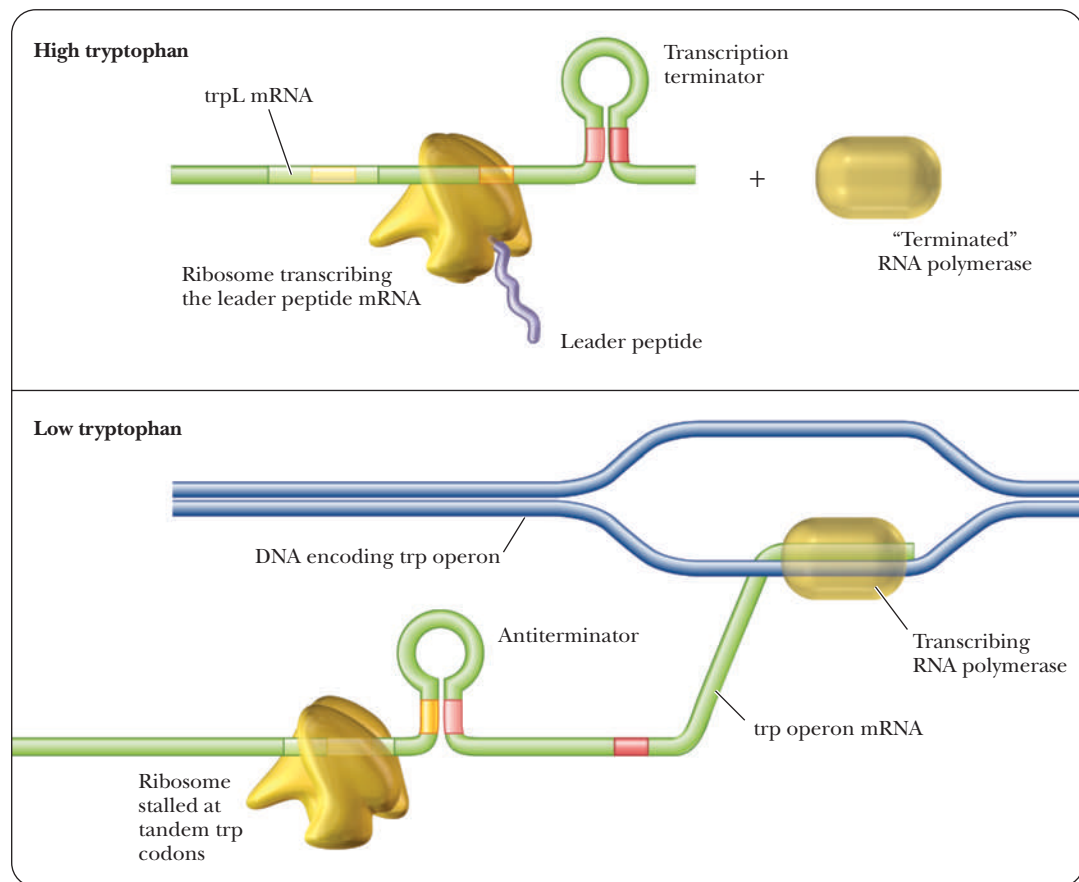


■ **Metabolite Sensing** (a): An aptamer for the coenzyme thiamine pyrophosphate (TPP) assumes a defined shape (left) as it exits the polymerase. When TPP is present, the aptamer binds to it, grasping the molecule tightly (right). **Riboswitch Responses:** Riboswitches employ a variety of strategies to control protein manufacture. When TPP is absent, for example, the expression platform can leave a translation initiation site open to ribosomes, allowing expression of the gene's instructions to remain "on" (b, left). When TPP is bound by the aptamer, the expression platform can form a hairpin that blocks translation, turning the gene "off" (b, right). A riboswitch sensing the coenzyme flavin mononucleotide (FMN) forms a terminator hairpin that halts transcription of its message by polymerase (c). An unusual ribozyme triggered by glucosamine-6-phosphate (GlcN6P) self-destructs by cleaving itself (d). (Reprinted from Scientific American, *The Power of Riboswitches* by Jeffrey E. Barrick and Ronald R. Breaker, January 2007, page 54.)

and proceeds until position 92, at which point the 1·2 pause structure can form. This causes RNA polymerase to pause in its RNA synthesis. A ribosome begins to translate the leader sequence, which releases the RNA polymerase from its pause and allows transcription to resume. The ribosome follows closely behind the RNA polymerase shown in Figure 11.16. The ribosome stops over the UGA stop codon of the mRNA, which prevents the 2·3 antiterminator hairpin from forming and allows instead the 3·4 terminator hairpin to form. This hairpin has the series of uracils characteristic of rho-independent termination. The RNA polymerase ceases transcription when this terminator structure forms.

If tryptophan is limiting, the ribosome stalls out over the tryptophan codons on the mRNA of the leader sequence. This leaves the mRNA free to





■ **FIGURE 11.16** The attenuation mechanism in the *trp* operon. The pause structure forms when the ribosome passes over the Trp codons quickly when tryptophan levels are high. This causes premature abortion of the transcript as the terminator loop is allowed to form. When tryptophan is low, the ribosome stalls at the Trp codons, allowing the antiterminator loop to form, and transcription continues.

form the 2·3 antiterminator hairpin, which stops the 3·4 terminator sequence from forming, so that the RNA polymerase continues to transcribe the rest of the operon. Transcription is attenuated in several other operons dealing with amino acid synthesis. In these cases, there are always codons for the amino acid, which is the product of the pathway that acts in the same way as the tryptophan codons in this example.

## 11.4 Transcription in Eukaryotes

We have seen that prokaryotes have a single RNA polymerase that is responsible for the synthesis of all three kinds of prokaryotic RNA—mRNA, tRNA, and rRNA. The polymerase can switch  $\sigma$  factors to interact with different promoters, but the core polymerase stays the same. The transcription process is predictably more complex in eukaryotes than in prokaryotes. Three primary RNA polymerases with different activities are known to exist. Each one transcribes a different set of genes and recognizes a different set of promoters:

1. RNA polymerase I is found in the nucleolus and synthesizes precursors of most, but not all, ribosomal RNAs.
2. RNA polymerase II is found in the nucleoplasm and synthesizes mRNA precursors.

3. RNA polymerase III is found in the nucleoplasm and synthesizes the tRNAs, precursors of 5S ribosomal RNA, and a variety of other small RNA molecules involved in mRNA processing and protein transport.

All three of these eukaryotic RNA polymerases are large (500–700 kDa), complex proteins consisting of 10 or more subunits. Their overall structures differ, but they all have a few subunits in common. They all have two larger subunits that share sequence homology with the  $\beta$ - and  $\beta'$ -subunits of prokaryotic RNA polymerase that make up the catalytic unit. There are no  $\sigma$ -subunits to direct polymerases to promoters. The detection of a gene to be transcribed is accomplished in a different way in eukaryotes, and the presence of transcription factors, of which there are hundreds, plays a larger role. Plants are known to have two other RNA Polymerases, Pol IV and Pol V, that have poorly understood functions, although they are involved in producing small RNA molecules that are part of gene silencing (see Section 11.6). We shall restrict our discussion to transcription by Pol II.

### Structure of RNA Polymerase II

Of the three RNA polymerases, RNA polymerase II is the most extensively studied, and the yeast *Saccharomyces cerevisiae* is the most common model system. Yeast RNA polymerase II consists of 12 subunits, as shown in Table 11.2. The subunits are called RPB1 through RPB12. **RPB** stands for **RNA polymerase B** because another nomenclature system refers to the polymerases as A, B, and C, instead of I, II, and III.

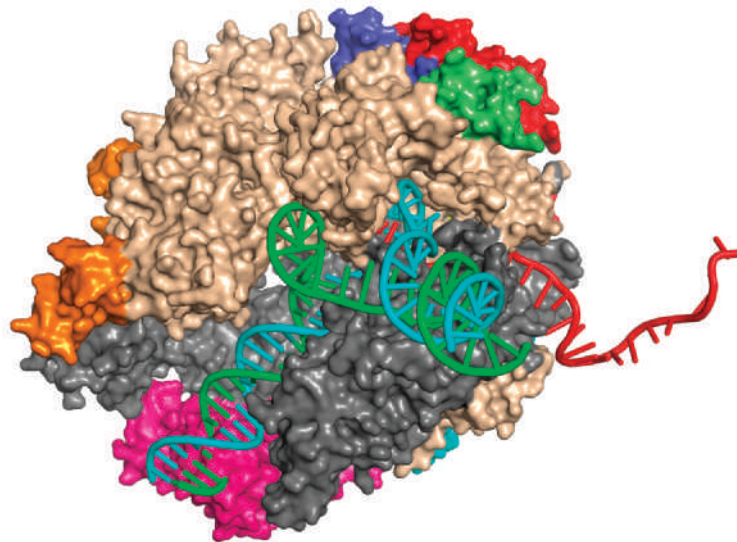
The function of many of the subunits is not known. The core subunits, RPB1 through RPB3, seem to play a role similar to their homologues in prokaryotic RNA polymerase. Five of them are present in all three RNA polymerases. RPB1 has a repeated sequence of PTSPSYS in the **C-terminal domain (CTD)**, which, as the name applies, is found at the C-terminal region of the protein. Threonine, serine, and tyrosine are all substrates for phosphorylation, which is important in the control of transcription initiation.

X-ray crystallography has been used to determine the structure of RNA polymerase II. Notable features include a pair of jaws formed by subunits RPB1, RPB5, and RPB9, which appear to grip the DNA downstream of the active site. A clamp near the active site is formed by RPB1, RPB2, and RPB6, which may be involved in locking the DNA:RNA hybrid to the polymerase, increasing the stability of the transcription unit. Figure 11.17 shows a diagram of the structure of RNA polymerase II.

**TABLE 11.2**

Yeast RNA Polymerase II Subunits			
Subunit	Size (kDa)	Features	<i>E. coli</i> Homologue
RPB1	191.6	Phosphorylation site	$\beta'$
RPB2	138.8	NTP binding site	$\beta$
RPB3	35.3	Core assembly	$\alpha$
RPB4	25.4	Promoter recognition	$\sigma$
RPB5	25.1	In Pol I, II, and III	
RPB6	17.9	In Pol I, II, and III	
RPB7	19.1	Unique to Pol II	
RPB8	16.5	In Pol I, II, and III	
RPB9	14.3		
RPB10	8.3	In Pol I, II, and III	
RPB11	13.6		
RPB12	7.7	In Pol I, II, and III	

■ **FIGURE 11.17** Architecture of yeast RNA polymerase II. Transcription of DNA (helical structure) into RNA (red) is shown. The template strand of DNA is shown in blue and the coding strand in green. Transcription takes place in the clamp region of the active site, shown at center right. The jaws that hold DNA downstream of the active site are shown at lower left. (Courtesy of Roger Kornberg/Stanford University)



Recent structural work on RNA polymerases from prokaryotes and eukaryotes has led to some exciting conclusions regarding their evolution. Extensive homology exists between the core regions of RNA polymerases from bacteria, yeast, and humans, leading researchers to speculate that RNA polymerase evolved eons ago, at a time when only prokaryotes existed. As more complex organisms developed, layers of other subunits were added to the core polymerase to reflect the more complicated metabolism and compartmentalization of eukaryotes.

### How does Pol II recognize the correct DNA to transcribe?

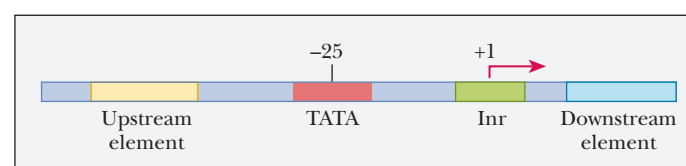
#### Pol II Promoters

Pol II promoters have four elements (Figure 11.18). The first includes a variety of **upstream elements**, which act as **enhancers** and **silencers**. Specific binding proteins either activate transcription above basal levels, in the case of enhancers, or suppress it, in the case of silencers. Two common elements that are close to the core promoter are the GC box (−40), which has a **consensus sequence** of GGGCGG, and the CAAT box (extending to −110), which has a consensus sequence of GGCCAATCT.

The second element, found at position −25, is the **TATA box**, which has a consensus sequence of TATAA(T/A).

The third element includes the transcription start site at position +1, but, in the case of eukaryotes, it is surrounded by a sequence called the **initiator element** (*Inr*). This sequence is not well conserved. For instance, the sequence for a particular gene type may be  $_{-3}\text{YYCAYYYY}_{+6}$ , in which Y indicates either pyrimidine, and A is the purine at the transcription start site (TSS).

The fourth element is a possible downstream regulator, although these are more rare than upstream regulators. Many natural promoters lack at least one of the four elements. The initiator plus the TATA box make up the core promoter and are the two most consistent parts across different species and genes. Some genes do not have TATA boxes; they are called “TATA-less” promoters. In some genes, the TATA box is necessary for transcription, and deletion of the TATA



■ **FIGURE 11.18** Four elements of Pol II promoters.

box causes a loss of transcription. In others, the TATA box orients the RNA polymerase correctly. Elimination of the TATA box in these genes causes transcription at random starting points. Whether a particular regulatory element is considered part of the promoter or not is often a judgment call. Those that are considered part of the promoter are close to the TSS (50–200 bp) and show specificity with regard to distance and orientation of the sequence. Regulatory sequences that are not considered part of the promoter can be far removed from the TSS, and their orientation is irrelevant. Experiments have shown that when such sequences are reversed, they still work, and when they are moved several thousand base pairs upstream, they still work.

## Initiation of Transcription

The biggest difference between transcription in prokaryotes and eukaryotes is the sheer number of proteins associated with the eukaryotic version of the process. Any protein that regulates transcription but is not itself a subunit of RNA polymerase is a **transcription factor**. There are many transcription factors for eukaryotic transcription, as we shall see. The molecular mass of the entire complex of Pol II and all of the associated factors exceeds 2.5 million Da.

Transcription initiation begins by the formation of a **preinitiation complex**, and most of the control of transcription occurs at this step. This complex normally contains RNA polymerase II and six **general transcription factors (GTFs)**—**TFIIA**, **TFIIB**, **TFIID**, **TFIIE**, **TFIIF**, and **TFIIH**.

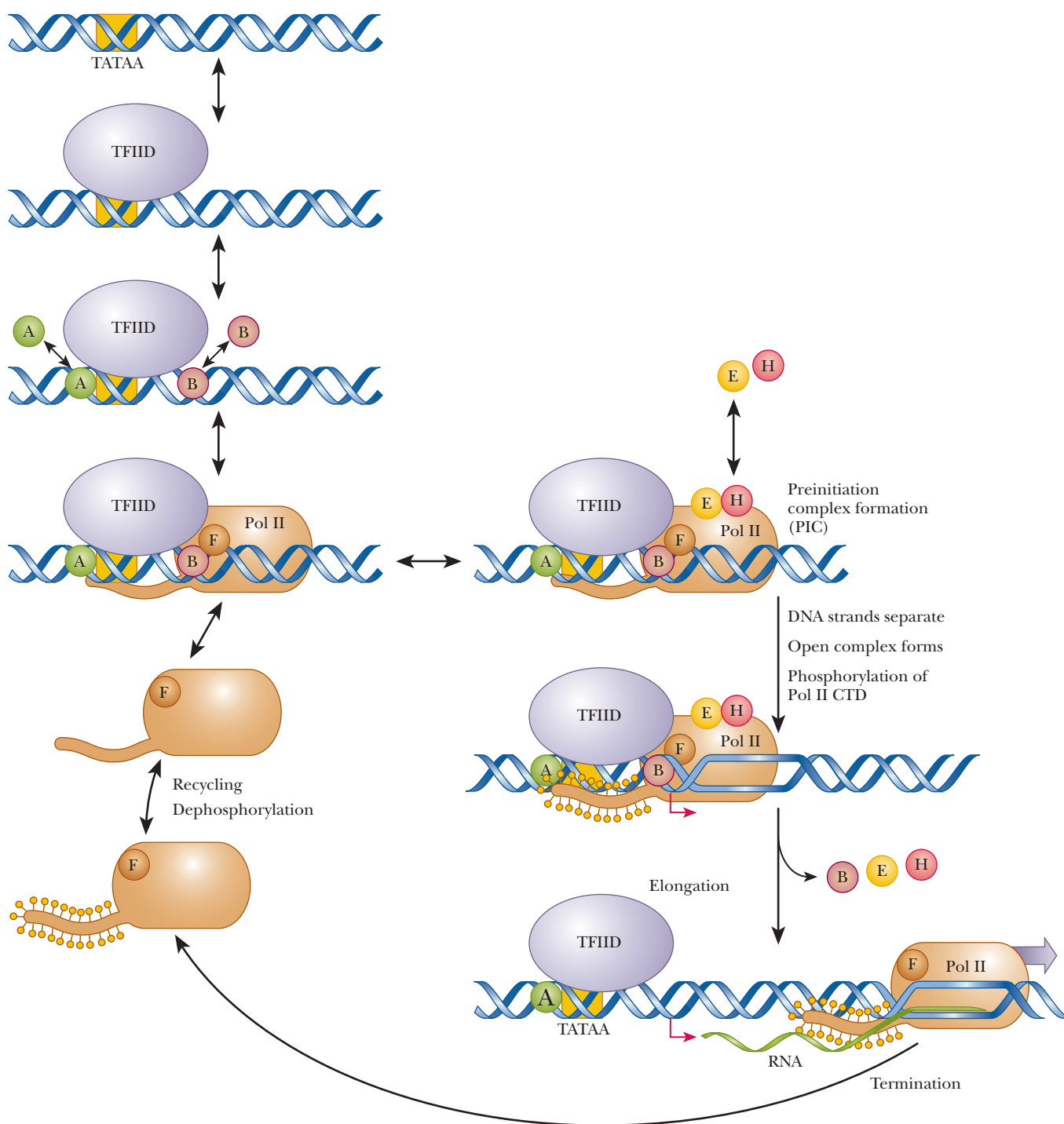
### What do eukaryotic transcription factors do?

The general transcription factors are required for all promoters. Much work is still going on to determine the structure and function of each of the parts of the preinitiation complex. Each GTF has a specific function, and each is added to the complex in a defined order. Table 11.3 is a summary of the components of the preinitiation complex.

Figure 11.19 shows the sequence of events in Pol II transcription. The first step in the formation of the preinitiation complex is the recognition of the TATA box by **TFIID**. This transcription factor is actually a combination of several proteins. The primary protein is called **TATA-binding protein (TBP)**. Associated with TBP are many **TBP-associated factors (TAFIIIs)**. Because TBP is also

**TABLE 11.3**

General Transcription Initiation Factors			
Factor	Subunits	Size (kDa)	Function
TFIID-TBP	1	27	TATA box recognition, positioning of TATA box DNA around TFIIB and Pol II
TFIID-TAF <sub>II</sub> s	14	15–250	Core promoter recognition (non-TATA elements), positive and negative regulation
TFIIA	3	12, 19, 35	Stabilization of TBP binding; stabilization of TAF–DNA binding
TFIIB	1	38	Recruitment of Pol II and TFIIF; start-site recognition for Pol II
TFIIF	3	156 total	Promoter targeting of Pol II
TFIIE	2	92 total	TFIIH recruitment; modulation of TFIIH helicase ATPase, and kinase activities; promoter melting
TFIIH	9	525 total	Promoter melting; promoter clearance via phosphorylation of CTD



■ **FIGURE 11.19** A schematic representation of the order of events of transcription. TFIIID (which contains the TATA-box binding protein, TBP) binds to the TATA box. TFIIA and TFIIB then bind, followed by recruitment of RNA polymerase II and TFIIF. TFIIE and TFIIH then bind to form the preinitiation complex (PIC). Kinases phosphorylate the C-terminal domain of Pol II, leading to the open complex in which the DNA strands are separated. RNA is produced during elongation as Pol II and TFIIF leave the promoter and the other general transcription factors behind. Pol II dissociates during the termination phase, and the CTD is dephosphorylated. Pol II/TFIIF is then recycled to bind to another promoter.



present and required for Pol I and Pol III, it is a universal transcription factor. TBP is highly conserved. From species as different as yeast, plants, fruit flies, and humans, the TBPs have more than 80% identical amino acids. The TBP protein binds to the minor groove of the DNA at the TATA box via the last 180 amino acids of its C-terminal domain. As shown in Figure 11.20, the TBP sits on the TATA box like a saddle. The minor groove of the DNA is opened, and the DNA is bent to an 80° angle.

As shown in Figure 11.19, once TFIID is bound, TFIIA binds, and TFIIA also interacts with both the DNA and TFIID. TFIIB also binds to TFIID, bridging the TBP and Pol II. TFIIA and TFIIB can bind in either order, and they do not interact with each other. TFIIB is critical for the assembly of the initiation complex and for the location of the correct transcription start site. TFIIF then binds tightly to Pol II and suppresses nonspecific binding. Pol II and TFIIF then bind stably to the promoter. TFIIF interacts with Pol II, TBP, TFIIB, and the TAFIIs. It also regulates the activity of the CTD phosphatase.

The last two factors to be added are TFIIE and TFIIH. TFIIE interacts with unphosphorylated Pol II. These two factors have been implicated in the phosphorylation of polymerase II. TFIIH also has helicase activity. After all these GTFs have bound to unphosphorylated Pol II, the preinitiation complex is complete. TFIIH has been found to have other functions as well, such as DNA repair (see the Biochemical Connections box on page 303).

Before transcription can begin, the preinitiation complex must form the *open complex*. In the open complex, the Pol II CTD is phosphorylated, and the DNA strands are separated (Figure 11.19).

## Elongation and Termination

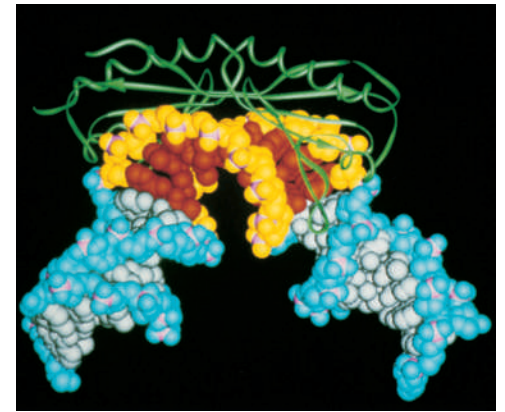
Less is known about elongation and termination in eukaryotes than in prokaryotes. Most of the research efforts have focused on the preinitiation complex and on the regulation by enhancers and silencers. As shown in Figure 11.19, the phosphorylated Pol II synthesizes RNA and leaves the promoter region behind. At the same time, the GTFs either are left at the promoter or dissociate from Pol II.

Pol II does not elongate efficiently when alone in vitro. Under those circumstances, it can synthesize only 100–300 nucleotides per minute, whereas the in vivo rates are between 1500 and 2000 nucleotides per minute. The difference is due to elongation factors. One is TFIIF, which, in addition to its role in the formation of the preinitiation complex, also has a separate stimulatory effect on elongation. A second elongation factor, which was named *TFIIS*, was more recently discovered.

Elongation is controlled in several ways. There are sequences called *pause sites*, where the RNA polymerase hesitates. This is very similar to the transcription attenuation we saw with prokaryotes. Elongation can also be aborted, leading to premature termination. In fact, more and more research is showing that abortive termination occurs more often than correct elongation, usually after just a few nucleotides have been linked. Finally, elongation can proceed past the normal termination point. This is called *antitermination*. The TFIIF class of elongation factors promotes a rapid read-through of pause sites, perhaps locking the Pol II into an elongation-competent form that does not pause and dissociate.

The TFIIS class of elongation factors are called *arrest release factors*. They help the RNA polymerase move again after it has paused. A third class of elongation factors consists of the *P-TEF* and *N-TEF* proteins (*positive-transcription elongation factor* and *negative-transcription elongation factor*). They increase the productive form of transcription and decrease the abortive form, or vice versa. At some point during either elongation or termination, TFIIF dissociates from Pol II.

Termination begins by stopping the RNA polymerase. There is a eukaryotic consensus sequence for termination, which is AAUAAA. This sequence may



■ **FIGURE 11.20** Model of yeast TATA-binding protein (TBP) binding to DNA. The DNA backbone of the TATA box is shown in yellow, and the TATA bases are shown in red. Adjacent DNA sequences are shown in turquoise. The TBP, which is shown in green, sits on the minor groove of the DNA like a saddle. (Reprinted by permission from *Crystal Structure of a Yeast TBP/TATA-Box Complex* by Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B. *Nature* 365, p. 512 [1993].)

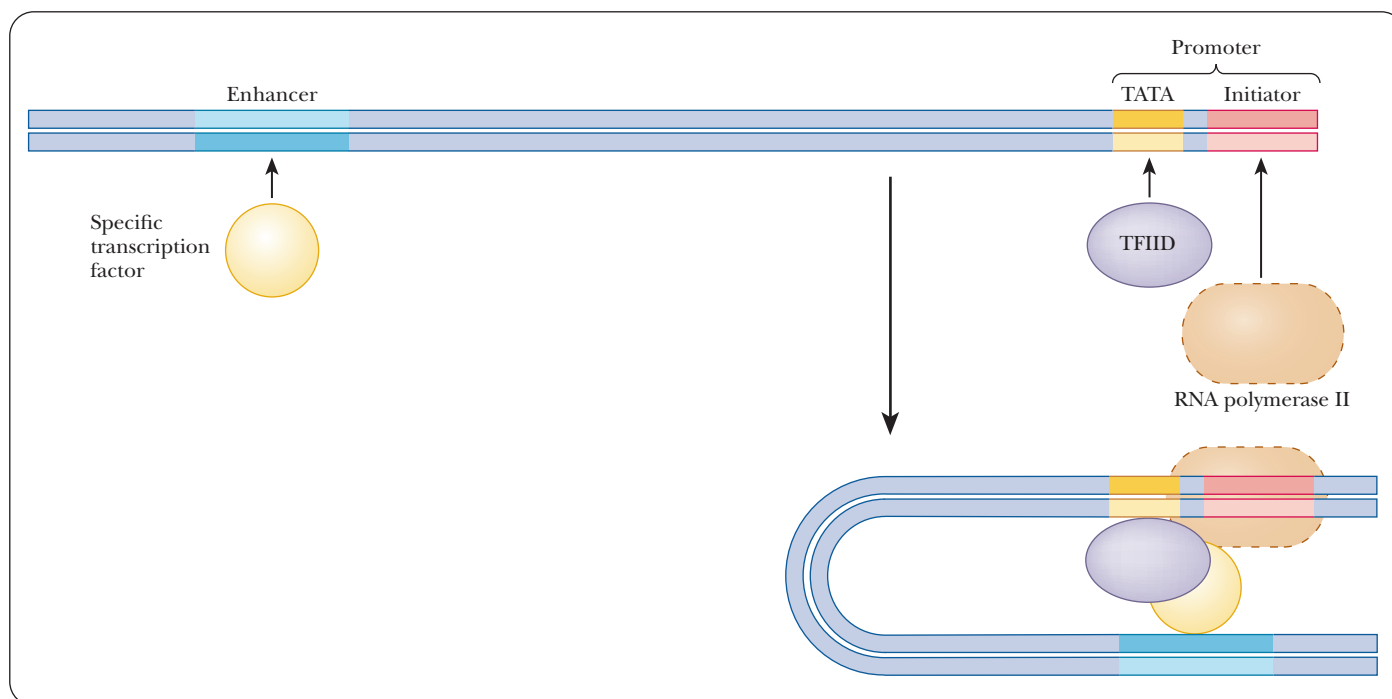
be 100–1000 bases away from the actual end of the mRNA. After termination occurs, the transcript is released, and the Pol II open form (phosphorylated) is released from the DNA. The phosphates are removed by phosphatases, and the Pol II/TFIIF complex is recycled for another round of transcription (Figure 11.19).

## 11.5 Transcription Regulation in Eukaryotes

In the last section, we saw how the general transcription machinery, consisting of the RNA polymerase and general transcription factors, functions to initiate transcription. This is the general case that is consistent for all transcription of mRNA. However, this machinery alone produces only a low level of transcription called the **basal level**. The actual transcription level of some genes may be many times the basal level. The difference is gene-specific transcription factors, otherwise known as **activators**. Recall that eukaryotic DNA is complexed to histone proteins in chromatin. The DNA is wound tightly around the histone proteins, and many of the promoters and other regulatory DNA sequences may be inaccessible much of the time.

### Enhancers and Silencers

As seen in prokaryotic transcription, enhancers and silencers are regulatory sequences that augment or diminish transcription, respectively. They can be upstream or downstream from the transcription initiator, and their orientation doesn't matter. They act through the intermediary of a gene-specific transcription-factor protein. As shown in Figure 11.21, the DNA must loop back so that the enhancer element and its associated transcription factor can contact the preinitiation complex. How this looping enhances transcription is still unknown.



■ **FIGURE 11.21** DNA looping brings enhancers in contact with transcription factors and RNA polymerase.

# Biochemical Connections

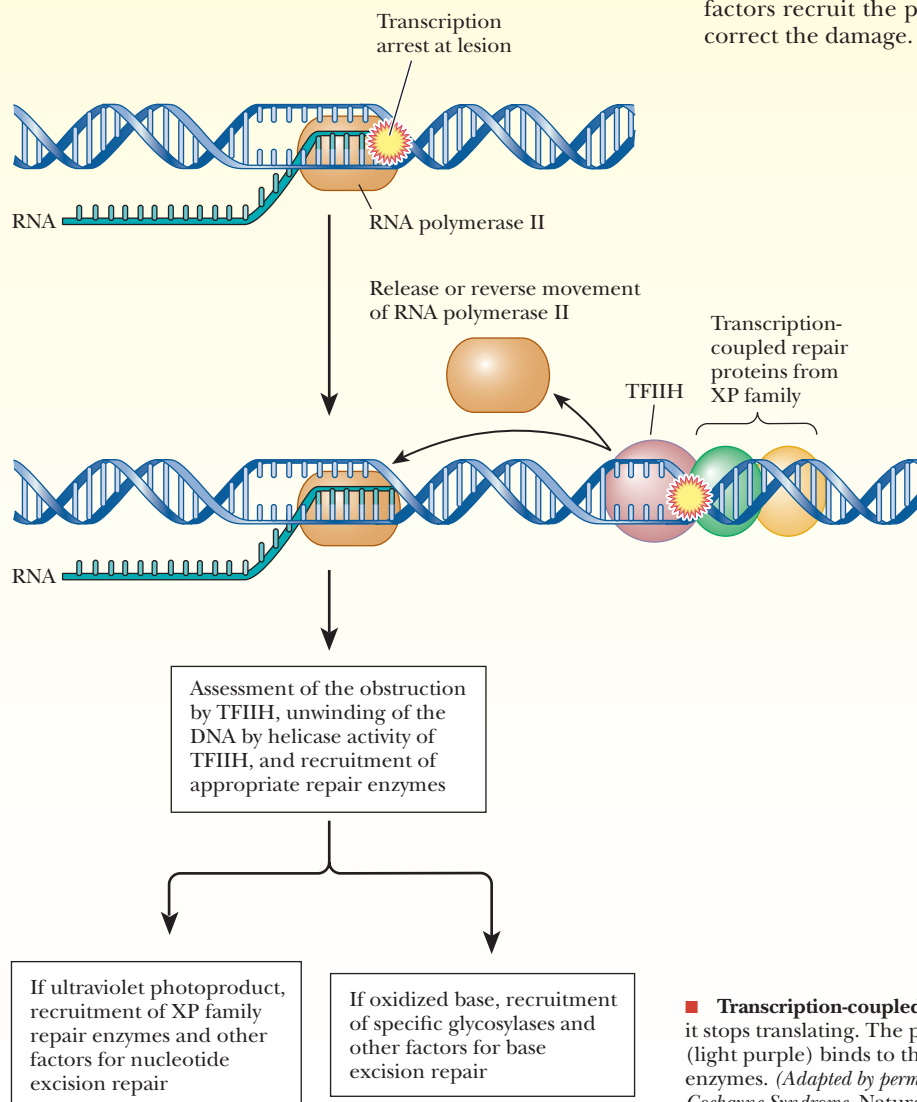
## GENETICS AND ALLIED HEALTH

### TFIIH—Making the Most Out of the Genome

The dogma for decades had always been that humans were more complex than other species, and this complexity was supposedly due to our having a larger amount of DNA and a greater number of genes. With the preliminary data from the Human Genome Project just in, it is now clear that we are not that much more complicated in terms of gene number. How, then, can very different structures and metabolisms between humans and nematodes, for example, be explained? Scientists must now look both at the effects of the proteins produced and at the control of their production, rather than simply counting the number of genes that encode proteins. A complex organism must get a lot of bang for the buck out of its gene products. This is seen clearly in the field of transcription. Eukaryotes have three RNA polymerases, but they all share some common subunits. Each polymerase has a unique organization of subunits and transcription factors, but many of these are shared among the multiple polymerases. Transcription factor TFIIH is particularly versatile. Besides its role in

initiation of transcription of Pol II, it also has a cyclin-dependent kinase activity. Cyclins are proteins that are involved in the control of the cell cycle. Thus TFIIH is involved not only in tying transcription and cell division together but also in repairing DNA, as seen in Chapter 10.

Two human genetic diseases, xeroderma pigmentosum (XP) and Cockayne syndrome, are characterized by extreme skin sensitivity to sunlight. Several genes are involved in the former disease, and most of the mutations lead to missing or defective DNA polymerases that act as repair enzymes. However, in a couple of the XP mutations and in Cockayne syndrome, there is a defect in the TFIIH protein. Besides its role in general transcription, it has been implicated in a DNA repair mechanism called *transcription-coupled repair* (TCR). The figure shows the model of the function of TFIIH. When RNA polymerase is attempting to transcribe DNA and it encounters a lesion, it cannot continue. The polymerase is released. TFIIH and one of the protein products of the XP family, XPG, bind to the DNA. It is believed that these factors recruit the particular repair enzymes that are needed to correct the damage.



■ **Transcription-coupled repair.** When RNA Pol II encounters a lesion, it stops translating. The polymerase must release or back up, while TFIIH (light purple) binds to the lesion and helps recruit the correct DNA repair enzymes. (Adapted by permission from Hanawalt, P. C. *DNA Repair: The Bases for Cockayne Syndrome*. *Nature* **405**, 415 [2000].)

Response Elements

Some transcription control mechanisms can be categorized based on a common response to certain metabolic factors. Enhancers that respond to these factors are called **response elements**. Examples include the **heat-shock element (HSE)**, the **glucocorticoid-response element (GRE)**, the **metal-response element (MRE)**, and the **cyclic-AMP-response element (CRE)**.

How do response elements work?

These response elements all bind proteins (transcription factors) that are produced under certain cell conditions, and several related genes are activated. This is not the same as an operon because the genes are not linked in sequence and are not controlled by a single promoter. Several different genes, all with unique promoters, may all be affected by the same transcription factor binding the response element.

In the case of HSE, elevated temperatures lead to the production of specific heat-shock transcription factors that activate the associated genes. Glucocorticoid hormones bind to a steroid receptor. Once bound, this becomes the transcription factor that binds to the GRE. Table 11.4 summarizes some of the best-understood response elements.

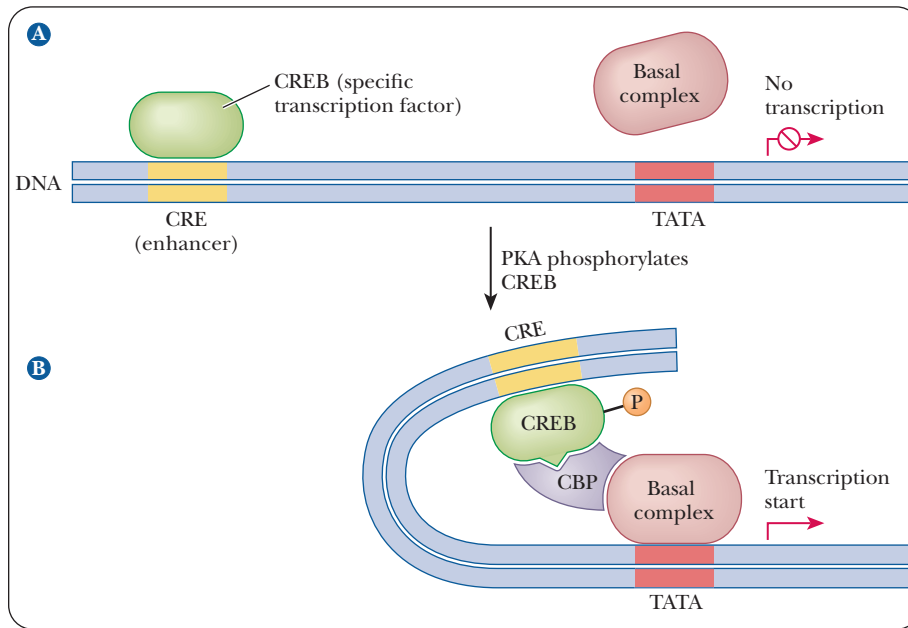
We will look more closely at the cyclic-AMP-response element as an example of eukaryotic control of transcription. Hundreds of research papers deal with this topic as more and more genes are found to have this response element as part of their control. Remember that cAMP was also involved in the control of prokaryotic operons via the CAP protein.

Cyclic AMP is produced as a second messenger from several hormones, such as epinephrine and glucagon (see Chapter 24). When the levels of cAMP rise, the activity of **cAMP-dependent protein kinase** (protein kinase A) is stimulated. This enzyme phosphorylates many other proteins and enzymes inside the cell and is usually associated with switching the cell to a catabolic mode, in which macromolecules are broken down for energy. Protein kinase A phosphorylates a protein called **cyclic-AMP-response-element binding protein (CREB)**, which binds to the cyclic-AMP-response element and activates the associated genes (see the Biochemical Connections box on page 309). The CREB does not directly contact the basal transcription machinery (RNA polymerase and GTFs), however, and the activation requires another protein. **CREB-binding protein (CBP)** binds to CREB after it has been phosphorylated and bridges the response element and the promoter region, as shown in Figure 11.22. After this bridge is made, transcription is activated above basal levels. CBP is called a

TABLE 11.4

Response Elements and Their Characteristics				
Response Element	Physiological Signal	Consensus Sequence	Transcription Factor	Size (kDa)
CRE	cAMP-dependent activation of protein kinase A	TGACGTCA	CREB, CREM, ATF1	43
GRE	Presence of glucocorticoids	TGGTACAAA TGTTC	Glucocorticoid receptor	94
HSE	Heat shock	CNNGAANNT CCNG*	HSTF	93
MRE	Presence of cadmium	CGNCCCCGN CNC*	?	?

\* N stands for any nucleotide.



**FIGURE 11.22** Activation of transcription via CREB and CBP. (a) Unphosphorylated CREB does not bind to CREB binding protein, and no transcription occurs. (b) Phosphorylation of CREB causes binding of CREB to CBP, which forms a complex with the basal complex (RNA polymerase and GTFs), thereby activating transcription. (Adapted by permission from Molecular Biology, by R. F. Weaver, McGraw-Hill, 1999.)

**TABLE 11.5**

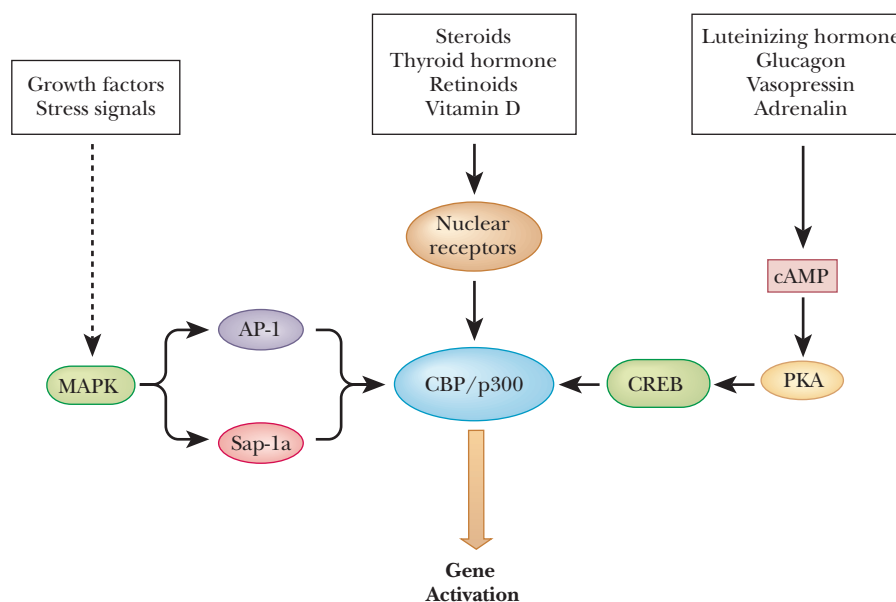
Abbreviations Used in Transcription			
bZIP	Basic-region leucine zipper	NTD	N-terminal domain
CAP	Catabolite activator protein	N-TEF	Negative transcription elongation factor
CBP	CREB-binding protein	Pol II	RNA polymerase II
CRE	Cyclic-AMP-response element	P-TEF	Positive transcription elongation factor
CREB	Cyclic-AMP-response-element binding protein	RPB	RNA polymerase B (Pol II)
CREM	Cyclic-AMP-response-element modulating protein	RNP	Ribonucleoprotein particle
CTD	C-terminal domain	snRNP	Small nuclear ribonucleoprotein particle ("snurp")
GRE	Glucocorticoid-response element	TAF	TBP-associated factor
GTF	General transcription factor	TATA	Consensus promoter element in eukaryotes
HSE	Heat-shock-response element	TBP	TATA-box binding protein
HTH	Helix–turn–helix	TCR	Transcription-coupled repair
Inr	Initiator element	TF	Transcription factor
MRE	Metal-response element	TSS	Transcription start site
MAPK	Mitogen-activated protein kinase	XP	Xeroderma pigmentosum

*mediator* or *coactivator*. Many abbreviations are used in the language of transcription, and Table 11.5 summarizes the more important ones.

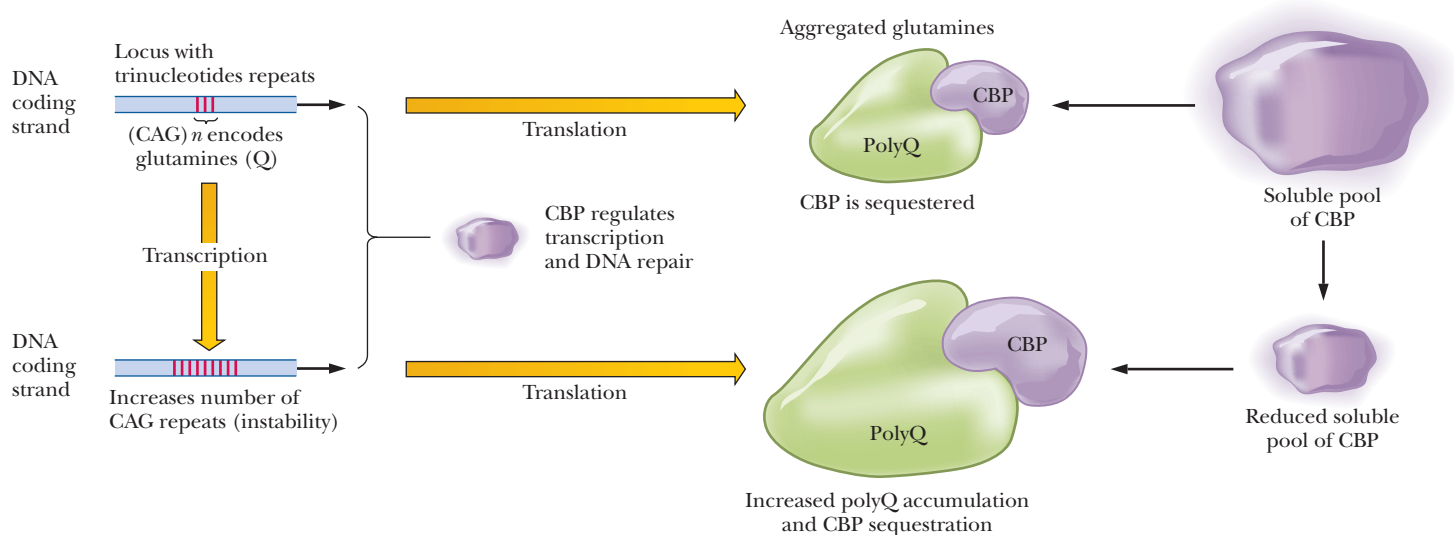
The CBP protein and a similar one called p300 are a major bridge to several different hormone signals, as can be seen in Figure 11.23. Several hormones that act through cAMP cause the phosphorylation and binding of CREB to CBP. Steroid and thyroid hormones and some others act on receptors in the nucleus to bind to CBP/p300. Growth factors and stress signals cause *mitogen-activated protein kinase* (MAPK) to phosphorylate transcription factors *AP-1* (*activating protein 1*) and *Sap-1a*, both of which bind to CBP.



■ **FIGURE 11.23** Multiple ways in which CREB binding protein (CBP) and p300 are involved in gene expression. MAPK is mitogen-activated protein kinase. It acts on two other transcription factors, AP-1 and Sap-1a, which bind to CBP. Steroid hormones affect nuclear receptors, which then bind to CBP. Other hormones activate a cAMP cascade, leading to phosphorylation of CREB, which then binds to CBP. (Adapted by permission from *Molecular Biology*, by R. F. Weaver, McGraw-Hill, 1999.)



Researchers have now linked several known human diseases to reduction of active levels of CBP, including Huntington's disease, spinocerebellar ataxia, and fragile X syndrome, all of which are characterized by mutations that increase levels of a trinucleotide repeat of CAG. This repeat produces polyglutamine when transcribed and translated. The polyglutamine product sequesters CBP, making less of it available for molecular processes, such as transcription and DNA repair. The loss of the DNA repair leads to a propagation of the CAG repeats and leads to the disease becoming worse in successive generations, a situation known as *genetic anticipation*. Figure 11.24 diagrams this syndrome.



■ **FIGURE 11.24** A toxic cycle accelerated. Polyglutamine (polyQ) proteins interfere with CBP, which influences transcription and DNA repair. In a vicious cycle, this may lead to more genetic instability of the trinucleotide repeat CAG, increased polyQ production, more interference with CBP function, and recurring instability of expanded CAG repeats. (Reprinted by permission of *Science magazine* from *Anticipating Trouble from Gene Transcription* by Mark E. Fortini, *Science* **315**, 1800–1801 [2007].)

## 11.6 Non-Coding RNAs

While so far we have focused on traditional transcription of DNA that leads to protein producing mRNA, the reality is that most of the RNA transcribed in our genome does not make protein. Scientists believe that most of the transcription of RNA leads to gene regulation and explains much of the complexity of higher organisms, as well as the many differences between higher organisms that have very similar DNA. Some estimates suggest that as much as 98% of the transcriptional output of our genomes is comprised of non-coding RNAs (ncRNA). NcRNAs have been linked to many processes, including regular transcription, gene silencing, replication, RNA processing, RNA modification, translation, protein stabilization, and protein translocation. While the number and type of RNAs seem to grow daily, the main focus of research is on two types, **Micro RNAs (miRNA)** and **Small Interfering RNA (siRNA)**.

### What are Micro RNAs?

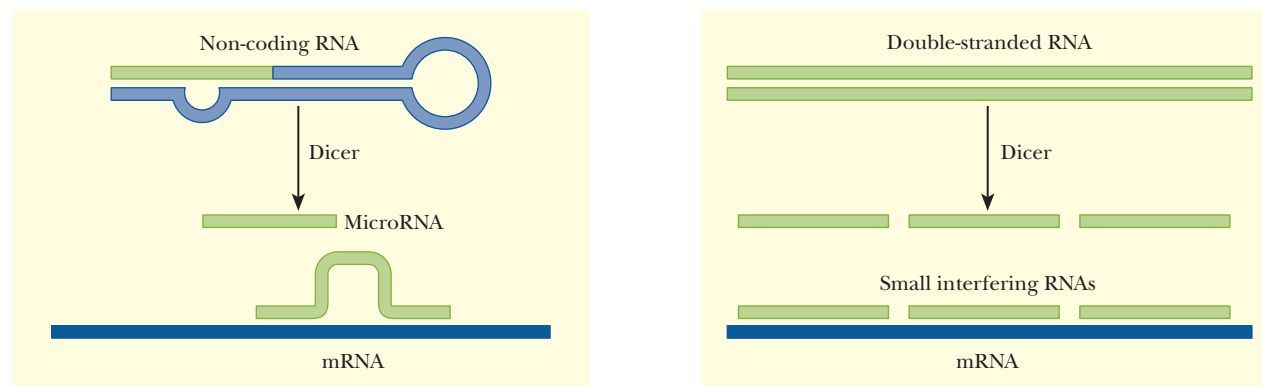
Micro RNAs are about 22 nucleotides long, and are cut from a longer, hairpin-shaped RNA by the enzyme Dicer (Figure 11.25a) The first miRNA genes, called *lin-4* and *let-7*, were found in the roundworm, *C. elegans* in 1993 and 2000, respectively. Although they were initially thought to be peculiar to the roundworm, it was soon discovered that *let-7* was present in many species, including humans. Now there are hundreds of miRNAs known. These miRNAs bind imperfectly to specific mRNAs and block their transcription.

### What are Small Interfering RNAs?

SiRNAs are formed in a way similar to miRNA, by the enzyme Dicer (Figure 11.25b) When a cell detects specific double-stranded RNA molecules, Dicer cuts them into small pieces of 21–25 nucleotides. These then bind to mRNA molecules in the process known as **RNA interference (RNAi)**, targeting them for destruction. The difference between the effect of miRNA and siRNA is subtle, but important. Binding of miRNA prevents the translation of the mRNA. Binding of siRNA causes the destruction of the mRNA. RNAi is the molecular process that is the backbone of a protective mechanism called **RNA silencing**.

### What is RNA Silencing?

Genomes of higher organisms are targets for invasion by viruses and transposable elements. Estimates are that nearly 50% of the human genome is a remnant of previous invasions by viruses and transposons over evolutionary time. In essence, our DNA has to fight off foreign invaders, just as our antibodies have to fight off foreign proteins and organisms such as parasites and bacteria.



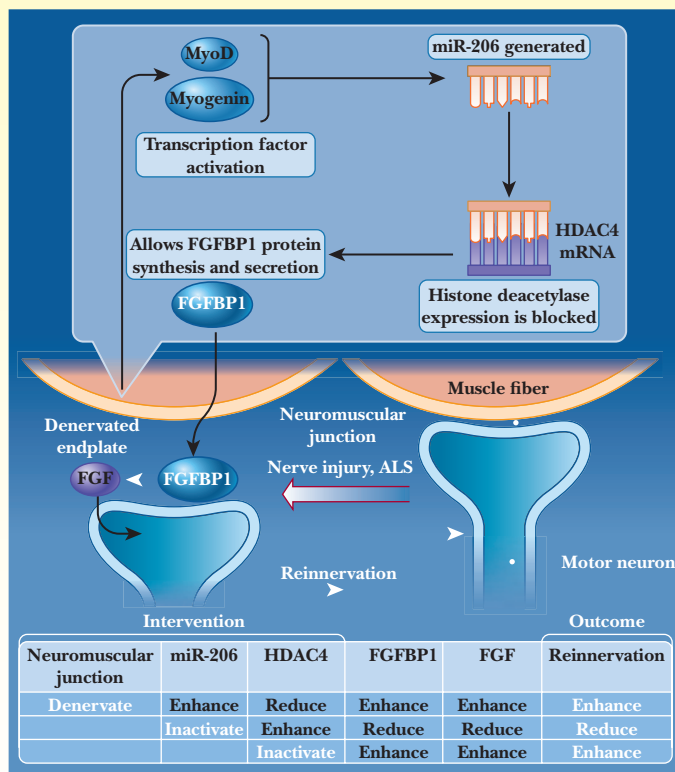
■ **FIGURE 11.25** The enzyme Dicer generates micro RNAs that bind imperfectly to target mRNAs, blocking protein production (left); in RNAi, the RNAs made by Dicer bind tightly to mRNAs, targeting them for destruction. (Reprinted by permission from *Gene regulation: The brave new world of RNA* by Carina Davis, *Nature* 418, 122–124 [2002].)

## Biochemical Connections MEDICINE

### A MicroRNA Helps Regenerate Nerve Synapses After Injury

Over the past 15 years, many new roles have been found for the hundreds of known miRNA molecules, including ties to various disorders such as cancer and heart disease. Recently a new role for an miRNA was discovered. MiRNA-206 was found to be involved in reinnervating the neuromuscular junction after injury. It was also found to improve survival in a mouse model of the disease amyotrophic lateral sclerosis (ALS), otherwise known as Lou Gehrig's disease. It has been known for some time that miRNAs play a role in muscle response to stress, and researchers used this as a basis for their hypothesis that diseases of muscle tissue might be due to loss of particular miRNAs. In normal mice, sciatic nerve injury results in loss of nerve function in the muscle and leads to an increase in miRNA-206. Using a line of mice that had ALS and inactivated miRNA-206, the time from onset of the disease was shortened, indicating this miRNA has a protective effect on nerves in the muscle. Further research indicated that the direct effect of the micro RNA is to repress the expression of histone deacetylase 4 (HDAC4) in the muscle cells. HDAC4 normally inhibits reinnervation by blocking expression of fibroblast growth factor binding protein 1 (BGFBP1), which acts on other fibroblast growth factors (FGF).

As can be seen in the figure, the miRNA-206 binds to the mRNA that expresses the histone deacetylase, blocking its translation. This allows the production of the FGFBP1 in the muscle fiber, which then passes into the neuromuscular junction and stimulates FGF. This growth factor moves into the motor neuron and promotes reinnervation. This pathway was shown to work both with damaged nerves and the diseased nerves from the ALS mice. This finding supports a growing body of evidence for the importance of miRNA to neurological function and susceptibility to disease. Micro RNA networks have been implicated in Parkinson's disease, Huntington's disease, and Alzheimer's disease.



■ **From injury to repair.** A damaged adult neuromuscular junction triggers a signaling cascade in the muscle fiber that involves the micro RNA miR-206. This increases the effect of the growth factor FGF on the neuron, and promotes reinnervation of the muscle. Reinnervation can be enhanced or diminished by altering upstream events. (Reprinted by permission from *A Reinnervating MicroRNA* by Robert H. Brown, *Science* 326, 1494–1495 [2009]. Used with permission of AAAS.)

RNA silencing is now believed to be an evolutionarily conserved process that is analogous to the protection of our genomes by an immune system. Researchers have used a variety of techniques to establish the importance of RNA silencing to the health of the organism, including creating strains of mice that lacked the proteins that make miRNA. The results were a variety of health problems for the mice, including heart disease and cancer.

Studies in humans have also showed the importance of miRNA-based processes. For example, loss of a particular miRNA, called miRNA-101, leads to overexpression of a particular histone methyltransferase that helps the progression of prostate cancer. Studies with HIV-infected cells have also shown that miRNA is important to controlling the spread of the disease and that HIV actively suppresses the production of the cell's miRNAs that would otherwise be able to fight it. Micro RNAs have been implicated in various heart conditions through their influence on production of alpha and beta myosin heavy chain, the predominant contractile proteins in the heart.

In late 2006, a California-based research team claimed to have found that small RNA molecules can also be found to work as activators. Researchers were attempting to use RNAi to block transcription of the human tumor suppressor gene, E-cadherin. When they added synthetic RNAs designed to target the gene's DNA, they found instead that production of the tumor suppressor went up instead of down. This process is tentatively being referred to as RNA

# Biochemical Connections

## GENETICS AND ENDOCRINOLOGY

### CREB—The Most Important Protein You Have Never Heard Of?

Hundreds of genes are controlled by the cyclic-AMP-response element. CREs are bound by a family of transcription factors that include CREB, cyclic-AMP-response-element modulating protein (CREM), and activating transcription factor 1 (ATF-1). All of these proteins share a high degree of homology, and all belong to the basic-region leucine zipper class of transcription factors (see Section 11.7). CREB itself is a 43-kDa protein with a critical serine at position 133 that can be phosphorylated. Transcription is activated only when CREB is phosphorylated at this site. CREB can be phosphorylated by a variety of mechanisms. The classical mechanism is via protein kinase A, which is stimulated by cAMP release. Protein kinase C, which is stimulated by  $\text{Ca}^{2+}$  release, and MAPK also phosphorylate CREB. The ultimate signals for these processes can be peptide hormones, growth or stress factors, or neuronal activity. Phosphorylated CREB does not act alone to stimulate transcription of its target genes. It works in concert with a 265-kDa protein, the CREB-binding protein (CBP), which connects CREB and the basal transcription machinery. More than 100 known transcription factors also bind to CBP. To add to the diversity of transcriptional control, CREB and CREM are both synthesized in

alternate forms because of different posttranscriptional splicing mechanisms (see Section 11.8). In the case of CREM, some of the isoforms are stimulatory while others are inhibitory.

Although the research is still ongoing, CREB-mediated transcription has been implicated in a tremendous variety of physiological processes, such as *cell proliferation*, *cell differentiation*, and *spermatogenesis*. It controls *release of somatostatin*, a hormone that inhibits growth hormone secretion. It has been shown to be critical for *development of mature T-lymphocytes* (immune-system cells), and has been shown to confer *protection to nerve cells* in the brain under hypoxic conditions. It is involved in *metabolism of the pineal gland* and *control of circadian rhythms*. CREB levels have been shown to be elevated during the body's adaptation to strenuous physical exercise. It is involved in the *regulation of gluconeogenesis* by the peptide hormones, glucagon and insulin, and it directly affects *transcription of metabolic enzymes*, such as phosphoenolpyruvate carboxykinase (PEPCK) and lactate dehydrogenase. Most interestingly, CREB has been shown to be *critical in learning and storage in long-term memory*, and low levels of CREB have been found in brain tissue of those with Alzheimer's disease.

activation, or RNAa. It is not clear at the moment whether this process was a positive activation of the gene in question or an interference with some other gene that led to the indirect activation of the E-cadherin gene. This process might be another powerful tool in the scientist's arsenal, as it would allow expanded ways to manipulate genes and approach the fight against genetic diseases.

## 11.7 Structural Motifs in DNA-Binding Proteins

Proteins that bind to DNA during the course of transcription do so by the same types of interactions that we have seen in protein structures and enzymes—hydrogen bonding, electrostatic attractions, and hydrophobic interactions. Most proteins that activate or inhibit transcription by RNA polymerase II have two functional domains. One of them is the **DNA-binding domain**, and the other is the **transcription-activation domain**.

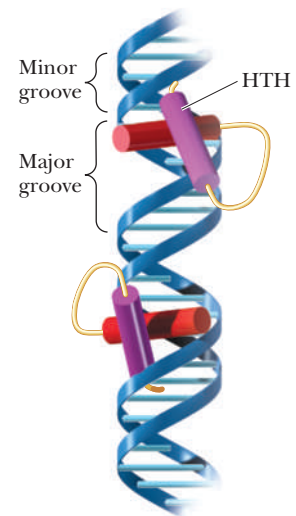
### What are DNA-binding domains?

#### DNA-Binding Domains

Most DNA-binding proteins have domains that fall into one of three categories: **helix–turn–helix (HTH)**, **zinc fingers**, and **basic-region leucine zipper (bZIP)**. These domains interact with DNA in either the major or minor groove, with the major groove being more common.

#### Helix–Turn–Helix Motifs

A common feature seen in proteins that bind to DNA is the presence of a segment of  $\alpha$ -helix that fits into the major groove. The width of the major groove and the  $\alpha$ -helix are similar, so the protein helix can fit snugly. This is the most common motif because the standard form of DNA, B-DNA, has the major groove of the correct size, and no alterations in its topology are necessary. Such binding proteins are often dimers with two regions of HTH, as shown in Figure 11.26.



■ **FIGURE 11.26** The helix–turn–helix motif. Proteins containing the HTH motif bind to DNA via the major groove. (With permission, from the Annual Review of Biochemistry, Volume 58 © 1989 by Annual Reviews. [www.annualreviews.org](http://www.annualreviews.org).)

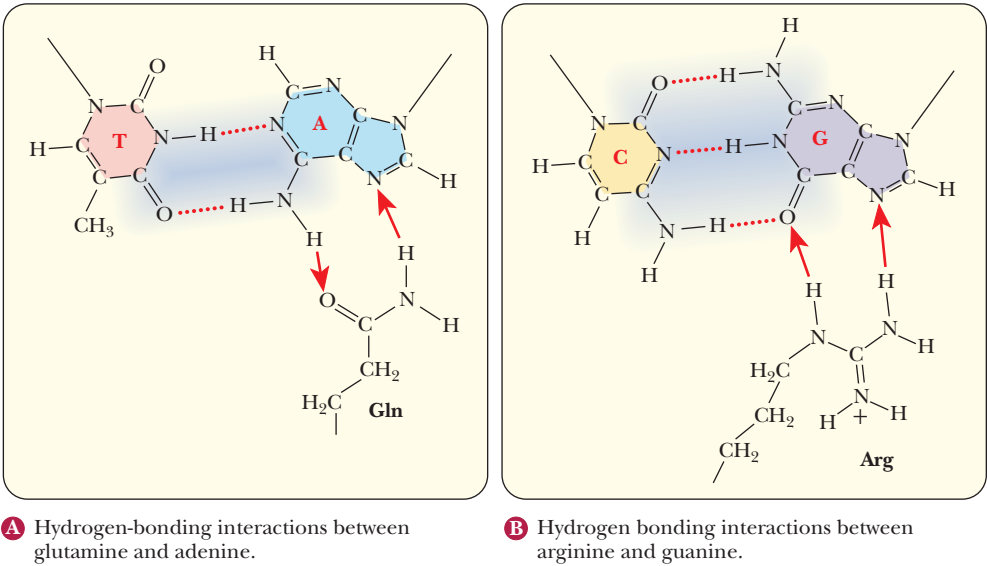
TABLE 11.6

Amino Acid Sequences in the HTH Regions of Selected Transcription Regulatory Proteins																				
	Helix								Turn				Helix							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
434 Rep	17-Gln	Ala	Glu	Leu	Ala	Gln	Lys	Val	Gly	Thr	Thr	Gln	Gln	Ser	Ile	Glu	Gln	Leu	Glu	Asn-36
434 Cro	17-Gln	Thr	Glu	Leu	Ala	Thr	Lys	Ala	Gly	Val	Lys	Gln	Gln	Ser	Ile	Gln	Leu	Ile	Glu	Ala-36
Lam Rep	33-Gln	Glu	Ser	Val	Ala	Asp	Lys	Met	Gly	Met	Gly	Gln	Ser	Gly	Val	Gly	Ala	Leu	Phe	Asn-52
Lam Cro	16-Gln	Thr	Lys	Thr	Ala	Lys	Asp	Leu	Gly	Val	Tyr	Gln	Ser	Ala	Ile	Asn	Lys	Ala	Ile	His-35
CAP	169-Arg	Gln	Glu	Ile	Gly	Glu	Ile	Val	Gly	Cys	Ser	Arg	Glu	Thr	Val	Gly	Arg	Ile	Leu	Lys-18
Trp Rep	68-Gln	Arg	Glu	Leu	Lys	Asn	Glu	Leu	Gly	Ala	Gly	Ile	Ala	Thr	Ile	Thr	Arg	Gly	Ser	Asn-87
Lac Rep	6-Leu	Tyr	Asp	Val	Ala	Arg	Leu	Ala	Gly	Val	Ser	Tyr	Gln	Thr	Val	Ser	Arg	Val	Val	Asn-25
Antp	31-Arg	Ile	Glu	Ile	Ala	His	Ala	Leu	Cys	Leu	Thr	Glu	Arg	Gln	Ile	Lys	Ile	Trp	Phe	Gln-50

Source: Adapted from Harrison, S. C., and Aggarwal, A. K., 1990, DNA recognition by proteins with the helix–turn–helix motif. *Annual Review of Biochemistry* 59, 933–969.

The HTH motif is a sequence of 20 amino acids that is relatively conserved in many different DNA-binding proteins. Table 11.6 shows the sequence for the HTH region of several transcription factors. The first helical region is composed of the first eight residues of the region. A sequence of three or four amino acids separates it from the second helical region. Position 9 is a glycine involved in a  $\beta$ -turn (Chapter 4).

Proteins that recognize DNA with specific base sequences are more likely to bind to the major groove. The orientation of the bases in the standard base pairings puts more of the unique structure into the major groove. Figure 11.27 shows how glutamine and arginine can interact favorably with adenine and guanine, respectively. Some interactions, however, including many in the minor groove, only read the DNA indirectly. As discussed in Chapter 9, the B form of DNA is not as constant as was once thought. Local variations in helix



■ FIGURE 11.27 DNA–amino acid interactions.



structure occur based on the actual sequence, especially when there are A–T-rich areas. The bases undergo extensive propeller-twist. Many proteins bind to the edges of the bases that protrude into the minor groove. Studies have shown that artificial molecules that mimic the base protrusions into the minor groove are equally capable of binding to many transcription factors.

Thus, although the base pairing is clearly important to DNA, sometimes the overall shape of the bases is important for other reasons. A particular binding protein might be recognizing not the part of the base that is involved in hydrogen bonds, but rather a part that is protruding into the grooves.

## Zinc Fingers

In 1985, it was discovered that a transcription factor of RNA polymerase III, TFIIIA, had nine repeating structures of 30 amino acids each. Each repeat contained two closely spaced cysteines and two closely spaced histidines 12 amino acids later. It was also found that this factor had enough associated zinc ions to bind to each of the repeats. This led to the discovery of the zinc-finger domain in DNA-binding proteins, which is represented in Figure 11.28.

The motif gets its name from the shape adopted by the 12 amino acids that are looped out from the intersection of the two cysteines and two histidines with the zinc ion. When TFIIIA binds to DNA, the repeated zinc fingers follow the major groove around the DNA, as shown in Figure 11.29.

## Basic-Region Leucine Zipper Motif

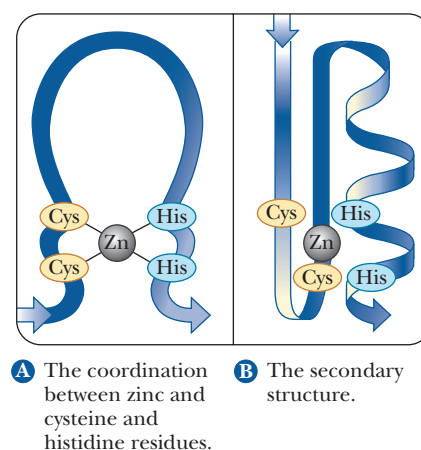
The third major class of sequence-dependent DNA-binding proteins is called the basic-region leucine zipper motif. Many transcription factors are known to contain this motif, including CREB (see the Biochemical Connections box on page 309). Figure 11.30 shows the sequence homology of several such transcription factors. Half of the protein is composed of the basic region, with many conserved residues of lysine, arginine, and histidine. The second half contains a series of leucines every seven residues. The significance of the spacing of the leucines is clear. It takes 3.6 amino acids to make a turn of an  $\alpha$ -helix. With a seven-residue spacing, the leucines all line up on one side of an  $\alpha$ -helix, as shown in Figure 11.31. The motif gets the name *zipper* from the fact that the line of hydrophobic residues interacts with a second analogous protein fragment via hydrophobic bonds, interweaving themselves like a zipper. DNA-binding proteins with leucine zippers bind the DNA in the major groove via the strong electrostatic interactions between the basic region and the sugar phosphates. Protein dimers form, and the leucine half interacts with the other subunit, while the basic part interacts with the DNA, as shown in Figure 11.32.

## What are transcription-activation domains?

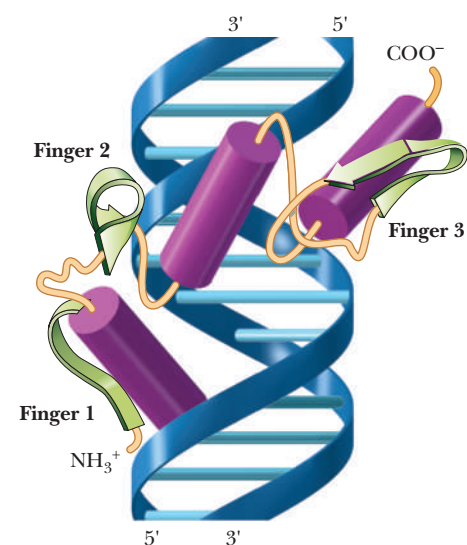
### Transcription-Activation Domains

The three motifs mentioned above are involved in the binding of transcription factors to DNA. Not all transcription factors bind directly to DNA, however. Some bind to other transcription factors and never contact the DNA. An example is CBP, which bridges CREB and the RNA polymerase II transcription-initiation complex. The motifs whereby transcription factors recognize other proteins can be broken down into three categories:

1. *Acidic domains* are regions rich in acidic amino acids. *Gal4* is a transcription factor in yeast that activates the genes for metabolizing galactose. It has a domain of 49 amino acids, 11 of which are acidic.
2. *Glutamine-rich domains* are seen in several transcription factors. *Sp1* is an upstream transcription factor that activates transcription in the presence of an additional promoter element called a *GC box*. It has two glutamine-rich



**FIGURE 11.28** Cys<sub>2</sub>His<sub>2</sub> zinc-finger motifs. (Adapted from Evans, R. M., and Hollenberg, S. M., 1988. *Cell* 52, 1, Figure 1.)



**FIGURE 11.29** Zinc-finger proteins follow the major groove of DNA. (Adapted with permission from Pavletich, N., and Pabo, C. O., 1991, *Science* 252, 809, Figure 2. Copyright © 1991 AAAS.)

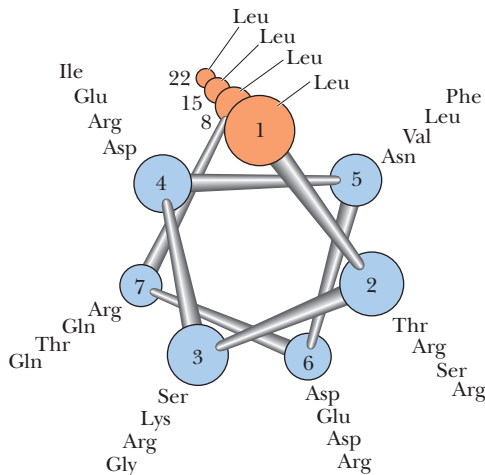
Protein		Basic region A	Basic region B	Leucine zipper	
C/EBP	278-DKNSNEYRVRRERNNIAVRKSHDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELDTLRG-341				
Jun	257-SQERIKAEKRKMRNRI AASKCHKRKLER IARLEEKVKT LKAQNSELASTANMLTEQVAQLKO-320				
Fos	233-EERRRIRRI RRRERNKMAAAKCRNRRRELTDTLQ AETDQLEDKKSALQTEIANLLKEKEKLEF-296				
GCN4	221-PESSDPAALKRARNT EAARRSRARKLQRMKOLEDKVEELL SKNYHLENEVARLKKLVGER-COOH				
YAP1	60-DLDPETKQKRTAQNRAAQRAFHERKERKMKEL EKKVQSLES IQQQNEVEATFLRDQLITLVN-123				
CREB	279-EEAARKREVRLMKNREAAARECRRKKKEYVKCL ENRVAVLENQNKT LIEELKALKDLYCHKSD-342				
Cys-3	95-ASRLAAEEDKRKRNTAASARFR I KKKQREQALEKSAKEMSEKVTQLEGR I QALETENKYLKG-148				
CPC1	211-EDPSDVVAMKRARNTLAARKSBERKAQRLEELEAKIEEL I AERDRYKNLALAHGASTE-COOH				
HBPI	176-WDERELKKQKRLSNRESARRSRLRKQAECEELGQRAEAL KSENSSLRIELDR I KKEYEELLS-239				
TGA1	68-SKPVEKVLRRLLAQRNEAARKSRLRKKA YVQQL ENSKLKLIQLEQELERARKQGMCVGGGVDA-131				
Opaque2	223-MPTEERVVRKRKESNRESARRSRYRKA AHLKELEDQVAQLKA ENSCLLRRI AALNQKYNDANV-286				

■ **FIGURE 11.30** Amino acid sequences of DNA-binding proteins with basic-region leucine zippers. (Adapted with permission from Vinson, C. R., Sigler, P. B., and McKnight, S. L. 1989. *Science* **246**, 912, Figure 1. Copyright © 1989 AAAS.)

domains, one of which contains 39 glutamines in 143 amino acids. CREB and CREM (see the Biochemical Connections box on page 309) also have this domain.

3. A *proline-rich domain* is seen in the activator *CTF-1*. It has a domain of 84 amino acids, 19 of which are prolines. CTF-1 is a member of a class of transcription factors that bind to an extended promoter element called a CCAAT box. The N-terminal domain has been shown to regulate transcription of certain genes. The C-terminal end is a transcription regulator and is known to bind to histone proteins via the proline repeats. An active area of study is how transcription is linked to the acetylation of histones. The coactivator CBP, which was discussed in the previous section, is also a histone acetyl transferase.

Despite the seemingly overwhelming complexity of transcription factors, their elucidation has been made more manageable by the similarities in the motifs described in this section. For example, if a new protein is discovered or a new DNA sequence is elucidated, evidence of its role as a transcription factor can be determined by locating the DNA-binding protein motifs discussed in this section.



■ **FIGURE 11.31** The helical wheel structure of a basic-region leucine zipper of a typical DNA-binding protein. The amino acids listed show the progression down the helix. Note that the leucines line up along one side, forming a hydrophobic spine. (Adapted with permission from Landschultz, W. H., Johnson, P. F., and McKnight, S. L., 1988, *Science* **240**, 1759–1764, Figure 1. Copyright © 1988 AAAS.)

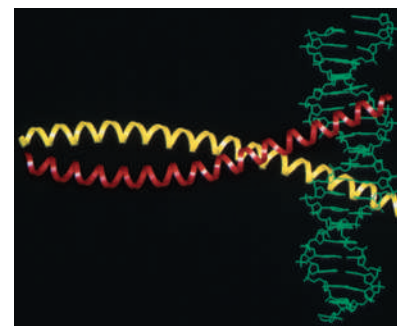
## 11.8 Posttranscriptional RNA Modification

The three principal kinds of RNA—tRNA, rRNA, and mRNA—are all modified enzymatically after transcription to give rise to the functional form of the RNA in question. The type of processing in prokaryotes can differ greatly from that in eukaryotes, especially in the case of mRNA. The initial size of the RNA transcripts is greater than the final size because of leader sequences at the 5' end and trailer sequences at the 3' end. The leader and trailer sequences must be removed, and other forms of *trimming* are also possible. *Terminal sequences* can be added after transcription, and *base modification* is frequently observed, especially in tRNA.

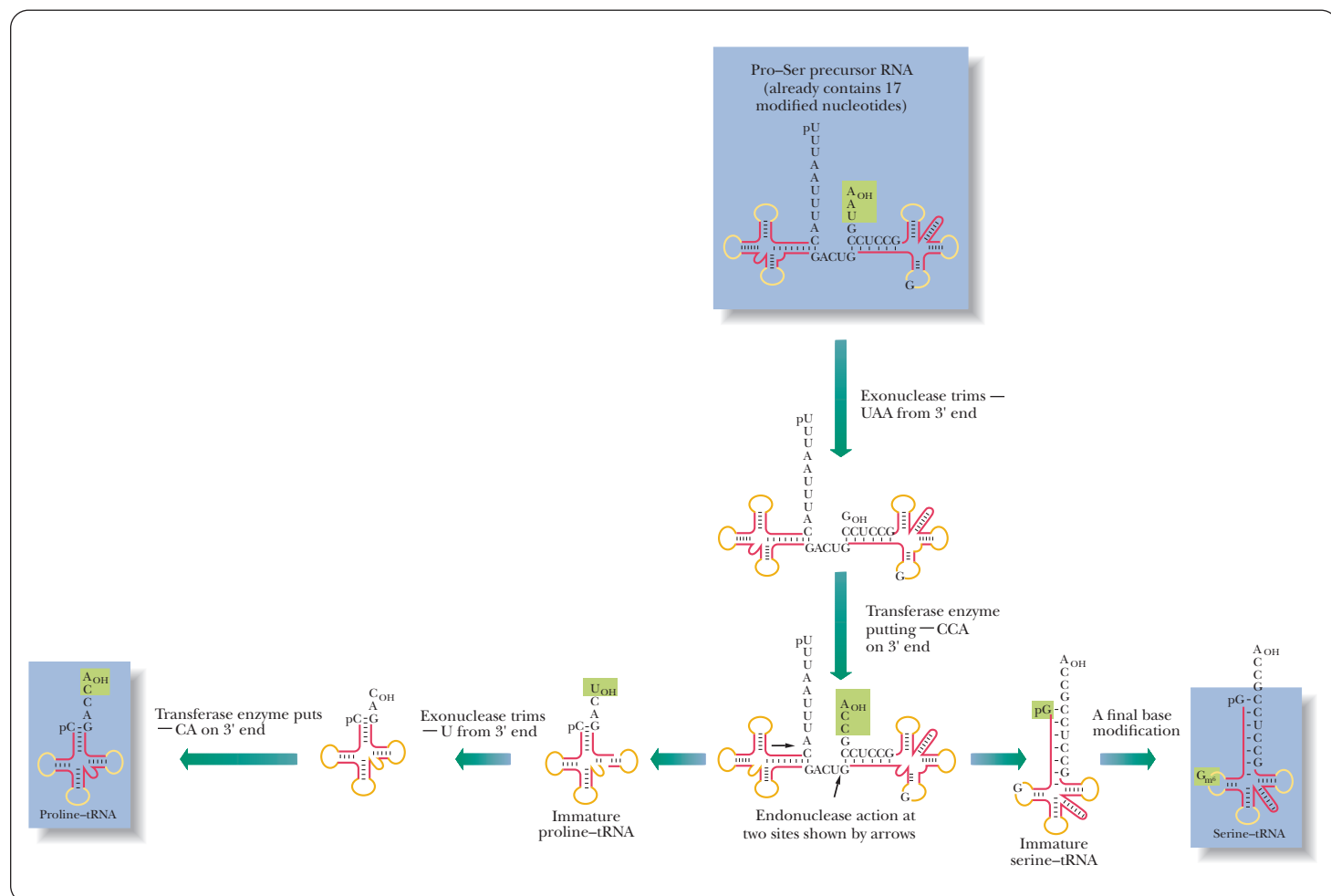
## Transfer RNA and Ribosomal RNA

The precursor of several tRNA molecules is frequently transcribed in one long polynucleotide sequence. All three types of modification—trimming, addition of terminal sequences, and base modification—take place in the transformation of the initial transcript to the mature tRNAs (Figure 11.33). (The enzyme responsible for generating the 5' ends of all *E. coli* tRNAs, *RNase P*, consists of both RNA and protein.) The RNA moiety is responsible for the catalytic activity. This was one of the first examples of catalytic RNA (Section 11.9). Some base modifications take place before trimming, and some occur after. Methylation and substitution of sulfur for oxygen are two of the more usual types of base modification. (See Section 9.2 and the discussion of transfer RNA in Section 9.5 for the structures of some of the modified bases.) One type of methylated nucleotide found only in eukaryotes contains a 2'-*O*-methylribosyl group (Figure 11.34).

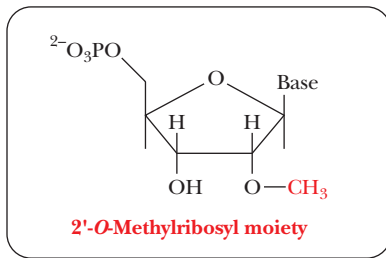
The trimming and addition of terminal nucleotides produce tRNAs with the proper size and base sequence. Every tRNA contains a CCA sequence at the 3' end. The presence of this portion of the molecule is of great importance in protein synthesis because the 3' end is the acceptor for amino acids to be added to a growing protein chain (Chapter 12). Trimming of large precursors of eukaryotic tRNAs takes place in the nucleus, but most methylating enzymes occur in the cytosol.



**FIGURE 11.32 Structure of the bZIP transcription factor.** Crystal structure of the bZIP transcription factor c-Fos:c-Jun bound to a DNA oligomer containing the AP-1 consensus target sequence TGACTCA. The basic region binds to the DNA while the leucine regions of the two helices bind via hydrophobic interactions. (Reprinted by permission from *Crystal structure of the heterodimeric bZIP transcription factor c-Fos:c-Jun bound to DNA* by Glover, J. N. M., and Harrison, S. C., *Nature* 373, 257–261 [1995].)



**FIGURE 11.33 Posttranscriptional modification of a tRNA precursor.** Dashes represent hydrogen-bonded base pairs. The symbols  $G_{OH}$ ,  $C_{OH}$ ,  $A_{OH}$ , and  $U_{OH}$  refer to a free 3' end without a phosphate group;  $G_m^2$  is a methylated guanine.



■ **FIGURE 11.34** The structure of a nucleotide containing a 2'-O-methylribosyl group.

The processing of rRNAs is primarily a matter of methylation and of trimming to the proper size. In prokaryotes, there are three rRNAs in an intact ribosome, which has a sedimentation coefficient of 70S. (Sedimentation coefficients and some aspects of ribosomal structure are reviewed in the discussion of ribosomal RNA in Section 9.5.) In the smaller subunit, which has a sedimentation coefficient of 30S, one RNA molecule has a sedimentation coefficient of 16S. The 50S subunit contains two kinds of RNA, with sedimentation coefficients of 5S and 23S. The ribosomes of eukaryotes have a sedimentation coefficient of 80S, with 40S and 60S subunits. The 40S subunit contains an 18S RNA, and the 60S subunit contains a 5S RNA, a 5.8S RNA, and a 28S RNA. Base modifications in both prokaryotic and eukaryotic rRNA are accomplished primarily by methylation.

## Messenger RNA

Extensive processing takes place in eukaryotic mRNA. Modifications include **capping** of the 5' end, **polyadenylating** (adding a poly-A sequence to) the 3' end, and **splicing** of coding sequences. Such processing is not a feature of the synthesis of prokaryotic mRNA.

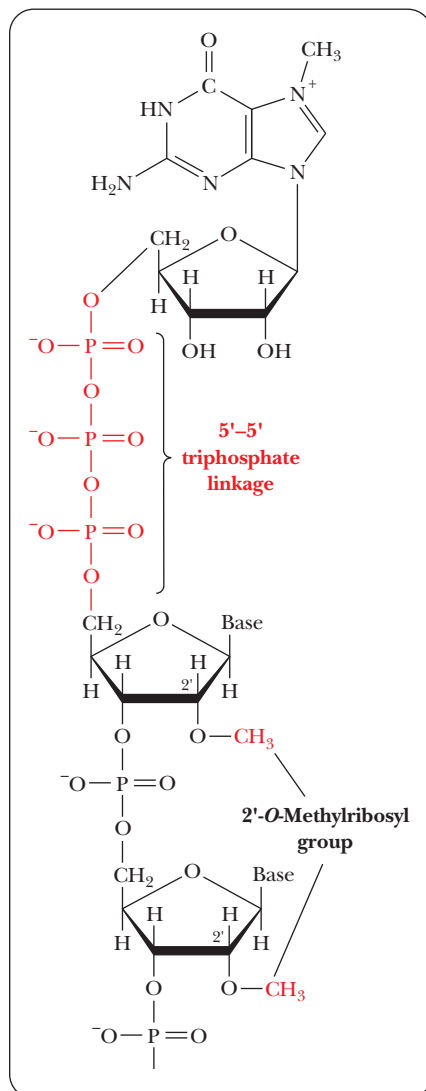
### Why is mRNA modified after initial transcription?

The cap at the 5' end of eukaryotic mRNA is a guanylate residue that is methylated at the N-7 position. This modified guanylate residue is attached to the neighboring residue by a 5'-5' triphosphate linkage (Figure 11.35). The 2'-hydroxyl group of the ribosyl portion of the neighboring residue is frequently methylated, and sometimes that of the next nearest neighbor is as well. The **polyadenylate tail** (abbreviated *poly-A* or *poly[r(A)<sub>n</sub>]*) at the 3' end of a message (typically 100 to 200 nucleotides long) is added before the mRNA leaves the nucleus. It is thought that the presence of the tail protects the mRNA from nucleases and phosphatases, which would degrade it. According to this point of view, the adenylate residues would be cleaved off before the portion of the molecule that contains the actual message is attacked. The presence of the 5' cap also protects the mRNA from exonuclease degradation.

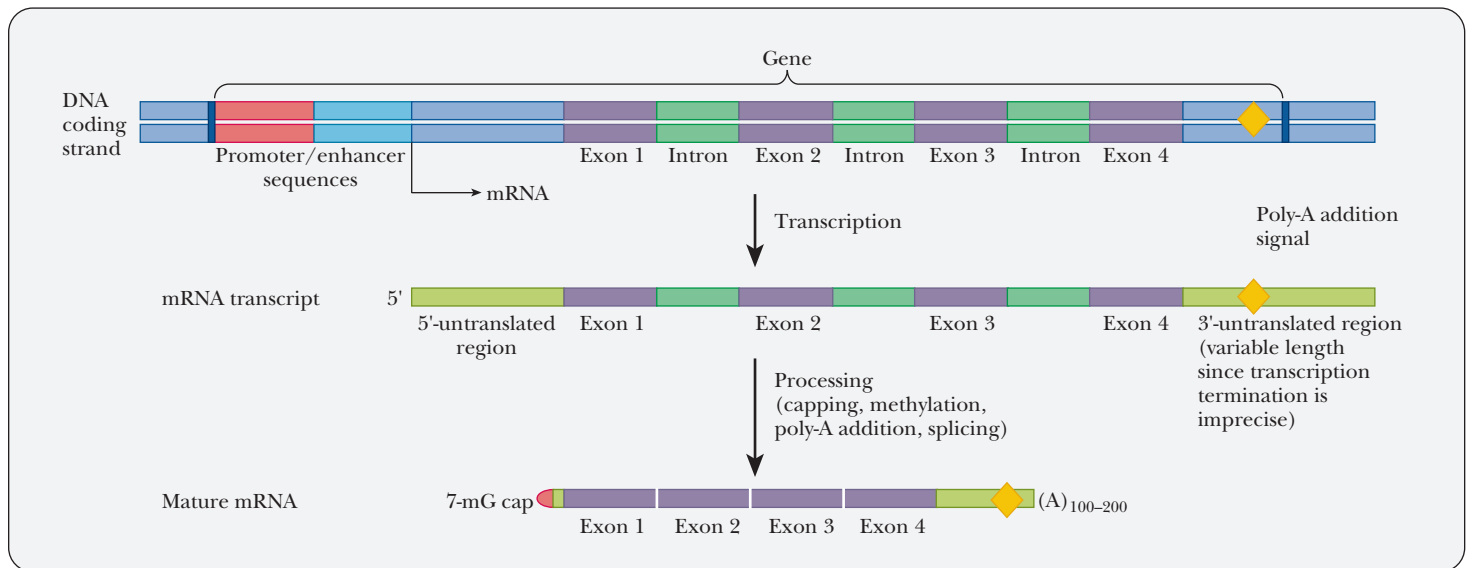
The presence of the poly-A tail has been very fortuitous for researchers. By designing an affinity chromatography column (Chapters 5 and 13) with a **poly-T tail** (or poly[d(T)] tail), the isolation of mRNA from a cell lysate can be quickly accomplished. This enables the study of transcription by looking at which genes are being transcribed at a particular time under various cell conditions.

The genes of prokaryotes are continuous; every base pair in a continuous prokaryotic gene is reflected in the base sequence of mRNA. The genes of eukaryotes are not necessarily continuous; eukaryotic genes frequently contain intervening sequences that do not appear in the final base sequence of the mRNA for that gene product. The DNA sequences that are expressed (the ones actually retained in the final mRNA product) are called **exons**. The intervening sequences, which are not expressed, are called **introns**. Such genes are often referred to as **split genes**. The expression of a eukaryotic gene involves not just its transcription but also the processing of the primary transcript into its final form. Figure 11.36 shows how a split gene might be processed. When the gene is transcribed, the mRNA transcript contains regions at the 5' and 3' ends that are not translated and several introns shown in green. The introns are removed, linking the exons together. The 3' end is modified by adding a poly-A tail and a 7-mG cap to yield the mature mRNA.

Some genes have very few introns, while others have many. There is one intron in the gene for the muscle protein actin; there are two for both the  $\alpha$ - and  $\beta$ -chains of hemoglobin, three for lysozyme, and so on, up to as many as 50



■ **FIGURE 11.35** The structures of some typical mRNA caps.



■ **FIGURE 11.36** The organization of split genes in eukaryotes.

introns in a single gene. The pro  $\alpha$ -2 collagen gene in chickens is about 40,000 base pairs long, but the actual coding regions amount to only 5000 base pairs spread out over 51 exons. With so much splicing needed, the splicing mechanisms must be very accurate. Splicing is a little easier because the genes have the exons in the correct order, even if they are separated by introns. Also, the primary transcript is usually spliced in the same positions in all tissues of the organism.

A major exception to this is the splicing that occurs with immunoglobulins, in which antibody diversity is maintained by having multiple ways of splicing mRNA. In the last few years, more eukaryotic proteins that are the products of alternative splicing have been discovered. The need for this was also demonstrated by the data from the Human Genome Project. Differential splicing was necessary to explain the fact that the known number of proteins exceeds the number of human genes found.

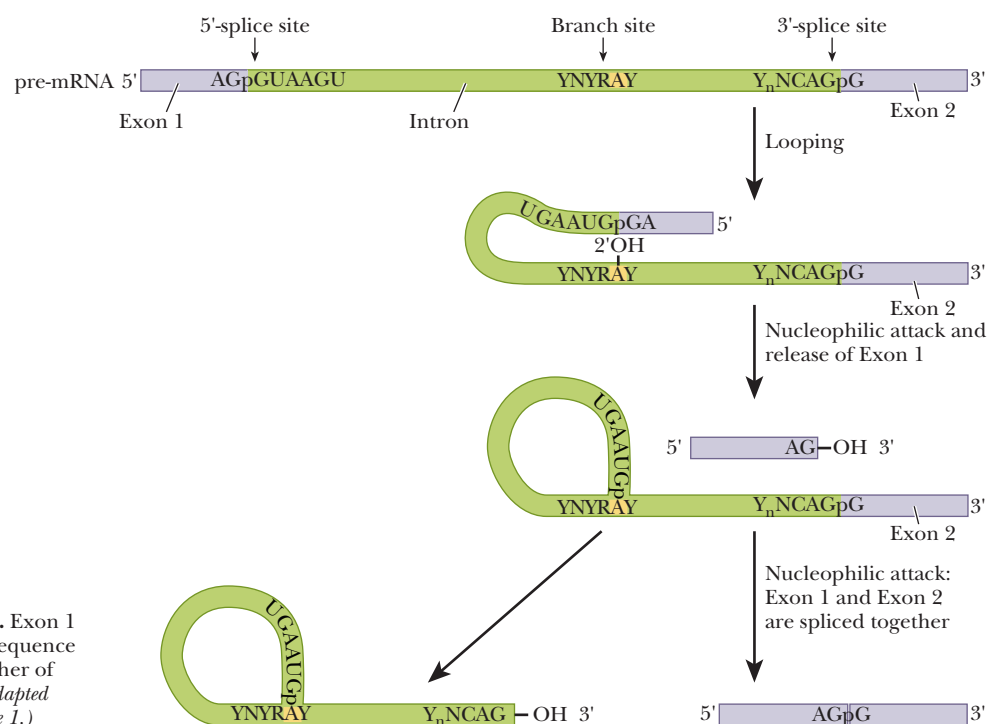
### How are introns spliced out to make mature RNA?

#### The Splicing Reaction: Lariats and Snurps

The removal of intervening sequences takes place in the nucleus, where RNA forms **ribonucleoprotein particles (RNPs)** through association with a set of nuclear proteins. These proteins interact with RNA as it is formed, keeping it in a form that can be accessed by other proteins and enzymes. The substrate for splicing is the capped, polyadenylated pre-mRNA. Splicing requires cleavage at the 5' and 3' end of introns and the joining of the two ends. This process must be done with great precision to avoid shifting the sequence of the mRNA product. Specific sequences make up the *splice sites* for the process, with GU at the 5' end and AG at the 3' end of the introns in higher eukaryotes. A *branch site* within the intron also has a conserved sequence. This site is found 18 to 40 bases upstream from the 3' splice site. The branch site sequence in higher eukaryotes is PyNPuAPy, where Py represents any pyrimidine and Pu any purine. N can be any nucleotide. The A is invariant.

Figure 11.37 shows how splicing occurs. The G that is always present on the 5' end of the intron loops back in close contact with the invariant A from the branch point. The 2' hydroxyl of the A performs a nucleophilic attack on





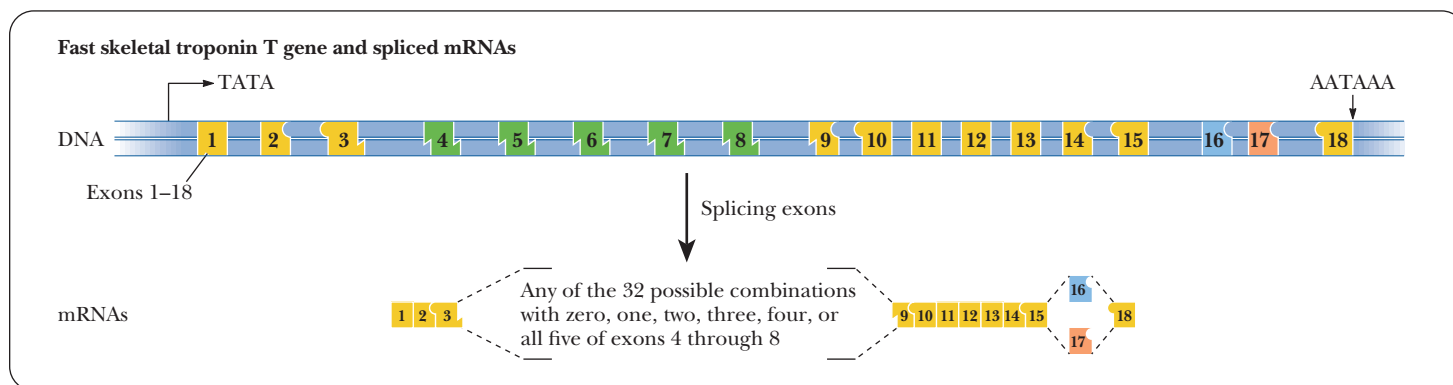
■ **FIGURE 11.37 Splicing of mRNA precursors.** Exon 1 and exon 2 are separated by the intervening sequence (intron) shown in green. In the splicing together of the two exons, a lariat forms in the intron. (Adapted from Sharp, P. A., 1987, *Science* 235, 766, Figure 1.)

the phosphodiester backbone at the 5' splice site, forming a *lariat* structure and releasing exon 1. The AG at the 3' end of the exon then does the same to the G at the 3' splice site, fusing the two exons. These lariat structures can be seen with an electron microscope, although the structure is inherently unstable and soon is linearized.

The splicing also depends on small nuclear ribonucleoproteins, or snRNPs (pronounced “snurps”), to mediate the process. This snRNP is another basic type of RNA, separate from mRNA, tRNA, and rRNA. The snRNPs, as their name implies, contain both RNA and proteins. The RNA portion is between 100 and 200 nucleotides in higher eukaryotes, and there are 10 or more proteins. With more than 100,000 copies of some snRNPs in eukaryotic cells, snRNPs are one of the most abundant gene products. They are enriched in uridine residues and are therefore often given names like U1 and U2. snRNPs also have an internal consensus sequence of AUUUUUG. The snRNPs bind to the RNAs being spliced via complementary regions between the snRNP and the branch and splice sites. The actual splicing involves a 50S to 60S particle called the **spliceosome**, which is a large multisubunit particle similar in size to a ribosome. Several different snRNPs are involved, and there is an ordered addition of them to the complex. In addition to their role in splicing, certain snRNPs have been found to stimulate transcription elongation. It is now widely recognized that some RNAs can catalyze their own self-splicing, as will be discussed in Section 11.9. The present process involving ribonucleoproteins may well have evolved from the self-splicing of RNAs. An important similarity between the two processes is that both proceed via a lariat mechanism by which the splice sites are brought together.

### Alternative RNA Splicing

Gene expression can also be controlled at the level of RNA splicing. Many proteins are always spliced in the same way, but many others can be spliced in different ways to give different **isoforms** of the protein to be produced. In humans, 5% of the proteins produced have isoforms based on alternative splicing.



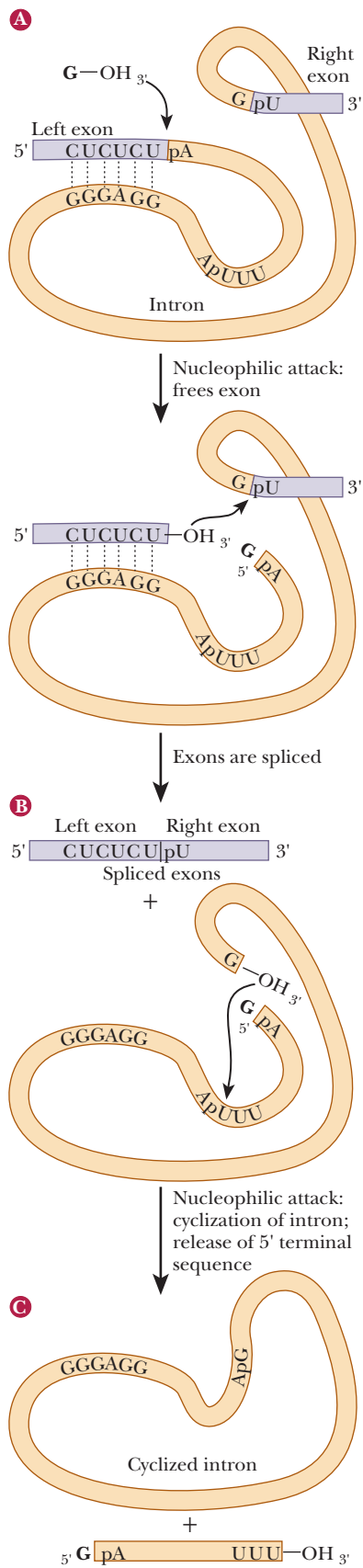
■ **FIGURE 11.38** Organization of the fast skeletal muscle troponin T gene and the 64 possible mRNAs that can be generated from it. Exons shown in orange are constitutive, appearing in all mRNAs produced. Exons in green are combinatorial, giving rise to all possible combinations from zero to all five. The exons in blue and red are mutually exclusive: only one or the other may be used. (With permission, from the Annual Review of Biochemistry, Volume 58, © 1989 by Annual Reviews. [www.annualreviews.org](http://www.annualreviews.org).)

These differences may be seen by having two forms of the mRNA in the same cell, or there might be only one form in one tissue, but a different form in another tissue. Regulatory proteins can affect the recognition of splice sites and direct the alternative splicing.

It has been found that a protein called Tau accumulates in the brain of people with Alzheimer's disease. This protein has six isoforms generated by differential splicing, with the forms appearing during specific developmental stages. The human troponin T gene produces a muscle protein that has many isoforms because of differential splicing. Figure 11.38 shows the complexity of this gene. Eighteen exons can be linked together to make the mature mRNA. Some of them are always present, such as exons 1–3 and 9–15, which are always linked together in their respective orders. However, exons 4–8 can be added in any combination group of 32 possible combinations. On the right side, either exon 16 or 17 is used, but not both. This leads to a total of 64 possible troponin molecules, which highlights the tremendous diversity in protein structure and function that can come from splicing mRNA.

## 11.9 Ribozymes

There was a time when proteins were considered the only biological macromolecules capable of catalysis. The discovery of the catalytic activity of RNA has thus had a profound impact on the way biochemists think. A few enzymes with RNA components had been discovered, such as telomerase (Chapter 10) and RNase P, an enzyme that cleaves extra nucleotides off the 5' ends of tRNA precursors. It was later shown that the RNA portion of RNase P has the catalytic activity. The field of catalytic RNA (ribozymes) was launched in earnest by the discovery of RNA that catalyzes its own self-splicing. It is easy to see a connection between this process and the splicing of mRNA by snRNPs. More recently, it has been shown that RNAs can catalyze reactions involved in protein synthesis, as will be explained further in Chapter 12. The catalytic efficiency of catalytic RNAs is less than that of protein enzymes, and the catalytic efficiency of currently existing RNA systems is greatly enhanced by the presence of protein subunits in addition to the RNA. Recall that many important coenzymes include an adenosine phosphate moiety in their structure (Section 7.8). Compounds of such central importance in metabolism must be of ancient origin,



another piece of evidence in support of the idea of an RNA-based world, where RNA was the original genetic molecule and the original catalytic one as well.

### What are the characteristics of a ribozyme?

Several groups of ribozymes are known to exist. In **Group I ribozymes**, there is a requirement for an external guanosine, which becomes covalently bonded to the splice site in the course of excision. An example is the self-splicing that takes place in pre-rRNA of the ciliate protist *Tetrahymena* (Figure 11.39). The transesterification (of phosphoric acid esters) that takes place here releases one end of the intron. The free 3'-OH end of the exon attacks the 5' end of the other exon, splicing the two exons and releasing the intron. The free 3'-OH end of the intron then attacks a nucleotide 15 residues from the 5' end, cyclizing the intron and releasing a 5' terminal sequence. The precision of this sequence of reactions depends on the folded conformation of the RNA, which remains internally hydrogen bonded throughout the process. In vitro, this catalytic RNA can act many times, being regenerated in the usual way for a true catalyst. In vivo, however, it appears to act only once by splicing itself out. **Group II ribozymes** display a lariat mechanism of operation similar to the mechanism seen in Section 11.8 that was facilitated by snRNPs. There is no requirement for an external nucleotide; the 2'-OH of an internal adenosine attacks the phosphate at the 5' splice site. Clearly, DNA cannot self-splice in this fashion because it does not have a 2'-OH.

The folding of the RNA is crucial to its catalytic activity, as is the case with protein catalysts. A divalent cation ( $\text{Mg}^{2+}$  or  $\text{Mn}^{2+}$ ) is required; it is quite likely that metal ions stabilize the folded structure by neutralizing some of the negative charges on the phosphate groups of the RNA. A divalent cation is essential for the functioning of the smallest ribozymes known, the hammerhead ribozymes, which can be catalytically active with as few as 43 nucleotides. (The name comes from the fact that their structures resemble the head of a hammer when shown in conventional representations of hydrogen-bonded secondary structure.) The folding of RNA is such that large-scale conformational changes can take place with great precision. Similar large-scale changes take place in the ribosome in protein synthesis and in the spliceosome in the processing of mRNA. Note that they remain RNA machines when proteins have taken over much of the catalytic functioning of the cell. The ability of RNA to undergo the requisite large-scale conformational changes may well play a role in the process. A recently proposed clinical application of ribozymes has been suggested. If a ribozyme can be devised that can cleave the RNA genome of HIV, the virus that causes AIDS (Section 14.2), it will be a great step forward in the treatment of this disease. Research on this topic is in progress in several laboratories.

**FIGURE 11.39** The self-splicing of pre-rRNA of the ciliate protist *Tetrahymena*, a Class I ribozyme. (a) A guanine nucleotide attacks at the splice site of the exon on the left, giving a free 3'-OH end. (b) The free 3'-OH end of the exon attacks the 5' end of the exon on the right, splicing the two exons and releasing the intron. (c) The free 3'-OH end of the intron then attacks a nucleotide 15 residues from the 5' end, cyclizing the intron and releasing a 5' terminal sequence.

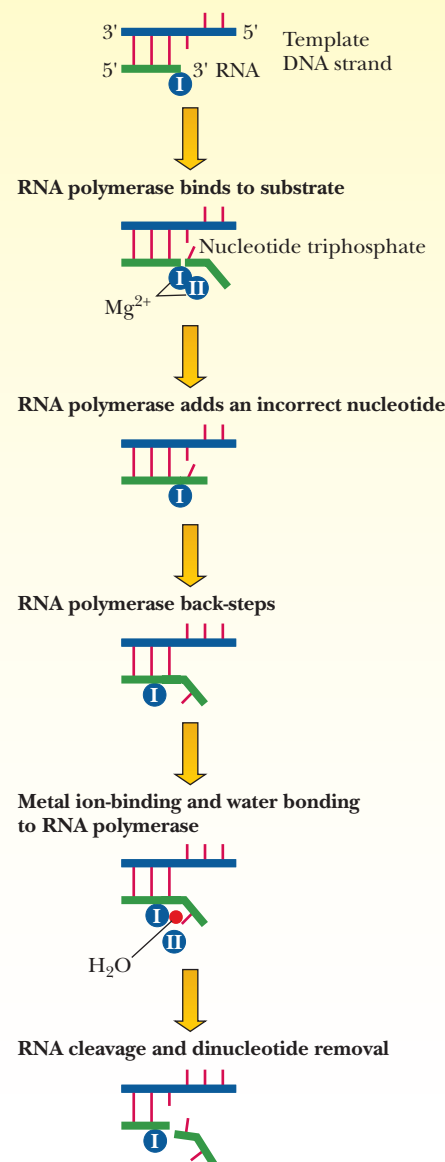
# Biochemical Connections EVOLUTIONARY BIOLOGY

## Proofreading in Transcription? RNA Fills In Another Missing Piece

The process of proofreading in replication has long been understood (Chapter 10), but until recently no such ability had been shown in transcription. To some extent, proofreading was not considered as necessary. An analogy would be the difference between a misprint in a cookbook and misreading the cookbook. In the first case, if the cookbook itself were wrong (DNA), then every attempt to read it would lead to a poor baking attempt. In the latter case, a person who misread the cookbook would make the mistake only once, analogous to making a poor RNA copy one time. Although there could be ramifications to making an incorrect RNA and/or protein, the consequences would be much worse if the DNA itself was wrong.

However, with our current model of the origins of life being based on RNA that had both replicative and catalytic ability, the puzzle was missing a piece. If RNA were the original genetic molecule, then the RNA would have to be replicated faithfully, and some proofreading ability would be necessary. With DNA proofreading, separate proteins do the proofreading, but for the RNA-world hypothesis to hold, RNA would need to have this ability without any protein, which had not evolved yet. Work by Zenkin et al. cited in the bibliography has shown that RNA can indeed catalyze its own proofreading, filling in our missing piece. They demonstrated that a misincorporated nucleotide could bend back on the previous nucleotide and, in the presence of  $Mg^{2+}$  and water, catalyze its own cleavage of the phosphodiester bond, as shown in the figure.

■ **RNA-assisted transcriptional proofreading.** Correction of misincorporation errors at the growing end of the transcribed RNA is stimulated by the misincorporated nucleotide.  $Mg^{2+}$  ions are bound to the catalytic region of RNA polymerase. (Reprinted by permission of Science, from *Self-Correcting Messages* by Patrick Cramer, Science, 313, 447 [2006].)



## SUMMARY

**What are the basics common to all transcription?** RNA synthesis is the transcription of the base sequence of DNA to that of RNA. All RNAs are synthesized on a DNA template. The enzyme that catalyzes the process is DNA-dependent RNA polymerase. All four ribonucleoside triphosphates—ATP, GTP, CTP, and UTP—are required, as is  $Mg^{2+}$ . There is no need for a primer in RNA synthesis. As is the case with DNA biosynthesis, the RNA chain grows from the 5' to the 3' end. The

enzyme uses one strand of the DNA (the antisense strand, or template strand) as the template for RNA synthesis. The RNA product has a sequence that matches the other strand of DNA, the coding strand.

**What do the subunits of RNA polymerase do?** In prokaryotes, the subunits of RNA polymerase,  $\alpha$ ,  $\beta$ ,  $\beta'$ , and  $\omega$ , make up the core of the enzyme and are responsible for the enzymatic

activity that catalyzes nucleotide incorporation. The  $\sigma$ -subunit is used for promoter recognition.

**Which of the DNA strands is used in transcription?** To make any particular RNA product, the RNA polymerase reads one of the DNA strands, called the template strand. It moves along the template strand from 3' to 5' and produces the RNA from 5' to 3'. The other strand of DNA is called the coding strand, and its sequence matches that of the RNA produced. In eukaryotes, the opposite strand is often used to produce small noncoding RNAs; their function in gene expression is being actively studied.

**How does RNA polymerase know where to begin transcription?** In prokaryotic transcription, RNA polymerase is directed to the gene to be transcribed by the interactions between the polymerase's  $\sigma$ -subunit and sequences of DNA near the start site called promoters. Consensus sequences have been established for prokaryotic promoters, and the key elements are sequences at -35 and -10, the latter called the Pribnow box. In eukaryotic transcription, RNA polymerase binds to promoters as well, but there is no  $\sigma$ -subunit, although there is a specific subunit, RBP4, that is involved in promoter recognition.

**How is transcription controlled in prokaryotes?** Frequency of transcription is controlled by the promoter sequence. Additional sequences upstream can also be involved in regulating prokaryotic transcription. These sequences are called enhancers or silencers, and they stimulate or inhibit transcription, respectively. Proteins called transcription factors can bind to these enhancer or silencer elements. Many prokaryotic genes that produce proteins that are part of a pathway are controlled in groups called operons, and expression of some genes is controlled by transcription attenuation.

**What is the difference between an enhancer and a promoter?** A promoter is a DNA sequence near the transcription start site that is bound by RNA polymerase during transcription initiation. Its position and orientation are critical to its function. Enhancers are DNA sequences that are farther away from the start site. Their position and orientation are not as important. They bind proteins called transcription factors and stimulate transcription above basal levels.

**How does repression work in the lac operon?** The regulatory gene of the *lac* operon, *lacI*, produces a protein called the repressor. This protein monomer, once transcribed and translated, combines to form a tetramer. The tetramer is the active repressor and it binds to the operator portion of the *lac* promoter. In the presence of the inducer, lactose, the repressor no longer binds and the repression is released.

**How are RNA secondary structures involved in transcription attenuation?** In prokaryotes, transcription and translation are linked. With attenuation, the RNA being transcribed is also being translated. Depending on the speed of simultaneous

translation, the RNA produced can form different hairpin loop structures. In one orientation the hairpin loop acts as a terminator and aborts the transcription before the actual proteins can be translated. In another orientation the transcription is allowed to proceed.

**How does Pol II recognize the correct DNA to transcribe?** One of the subunits of RNA pol II is used for promoter recognition. The most well-studied promoter region is called the TATA box. Although some genes lack TATA boxes, they are the most consistent part of eukaryotic promoters that RNA polymerase recognizes.

**What do eukaryotic transcription factors do?** There are a large number of transcription factors. Some are called general transcription factors, and they are involved in transcription initiation. They aid in promoter recognition and binding. They have a specific order and location of binding. Other transcription factors bind to enhancers or response elements and increase the rate of transcription above basal levels.

**How do response elements work?** Response elements are DNA sequences similar to enhancers, but they are involved in a bigger picture of metabolic responses. Common response elements are the heat-shock element (HSE) and the cyclic-AMP response element (CRE). Hundreds of processes are linked to transcription involving CRE.

**What are DNA-binding domains?** There are some common and easily recognizable domains in transcription factors, such as the helix-turn-helix, zinc fingers, and the basic-region leucine zipper. These sections of the protein allow easy binding to the DNA.

**What are transcription-activation domains?** In addition to binding DNA, transcription factors often bind other proteins. The sites of protein-protein interactions can also be identified by common motifs, such as acidic domains, glutamine-rich domains, and proline-rich domains.

**Why is mRNA modified after initial transcription?** Messenger RNA is modified in several ways in eukaryotes. Two modifications believed to be protection mechanisms are 5' capping and 3' polyadenylation. The 5' cap uses a unique 5'-5' bond that standard nucleases would not be able to degrade; the 3' the mRNA. The other major modification is the removal of introns. Because eukaryotic DNA is not continuous, the mRNA produced has intervening sequences that are not correct. These introns must be removed by splicing to yield the correct final mRNA.

**How are introns spliced out to make mature RNA?** Introns are removed by specific splicing reactions involving RNA. Sometimes the splicing reaction involves a separate ribonucleoprotein molecule called a snRNP ("snurp"). Other times the RNA being spliced catalyzes the reaction itself. The general mechanism involves a 3' splice point, a 5' splice point, and



a branch point. An intermediate in the process has a lariat shape.

**What are the characteristics of a ribozyme?** Ribozymes are RNA molecules with catalytic ability. Some ribozymes are a combination of RNA and protein and others are just RNA.

## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

### 11.2 Transcription in Prokaryotes

- Recall** What is the difference in the requirement for a primer in RNA transcription compared to DNA replication?
- Recall** List three important properties of RNA polymerase from *E. coli*.
- Recall** What is the subunit composition of *E. coli* RNA polymerase?
- Recall** What is the difference between the core enzyme and the holoenzyme?
- Recall** What are the different terms used to describe the two strands of DNA involved in transcription?
- Recall** Define *promoter region* and list three of its properties.
- Recall** Put the following in linear order: UP element, Pribnow box, TSS,  $-35$  region, Fis site.
- Recall** Distinguish between rho-dependent termination and intrinsic termination.
- Reflect and Apply** Diagram a section of DNA being transcribed. Give the various names for the two strands of DNA.

### 11.3 Transcription Regulation in Prokaryotes

- Recall** Define *inducer* and *repressor*.
- Recall** What is a  $\sigma$  factor? Why is it important in transcription?
- Recall** What is the difference between  $\sigma^{70}$  and  $\sigma^{32}$ ?
- Recall** What is the function of the catabolite activator protein?
- Recall** What is transcription attenuation?
- Reflect and Apply** What role does an operon play in the synthesis of enzymes in prokaryotes?
- Reflect and Apply** Diagram a termination of transcription showing how inverted repeats can be involved in releasing the RNA transcript.
- Reflect and Apply** Give an example of a system in which alternative  $\sigma$  factors can control which genes are transcribed. Explain how this works.
- Reflect and Apply** Explain, with diagrams, how transcription attenuation works in the *trp* operon.

### 11.4 Transcription in Eukaryotes

- Biochemical Connections** What is an aptamer?
- Biochemical Connections** What is a riboswitch?
- Biochemical Connections** What are various ways that a riboswitch shuts off translation when it binds to its target molecule?
- Biochemical Connections** How is the discovery of riboswitches relevant to bacterial pathology?
- Recall** Define *exon* and *intron*.
- Recall** What are some of the main differences between transcription in prokaryotes and in eukaryotes?

To be a ribozyme, it must be the RNA portion that is involved in the catalysis of the reaction. Scientists believe that RNA was the first molecule that could combine the ability to carry genetic information for replication as well as the ability to catalyze reactions. It was recently shown that RNA can also catalyze its own proofreading during transcription.

- Recall** What are the products of the reactions of the three eukaryotic RNA polymerases?
- Recall** List the components of eukaryotic Pol II promoters.
- Recall** List the Pol II general transcription factors.
- Reflect and Apply** What are the functions of TFIIF?

### 11.5 Transcription Regulation in Eukaryotes

- Recall** Describe the function of three eukaryotic response elements.
- Recall** What is the purpose of CREB?
- Reflect and Apply** How does regulation of transcription in eukaryotes differ from regulation of transcription in prokaryotes?
- Reflect and Apply** What is the mechanism of transcription attenuation?
- Reflect and Apply** How do the roles of enhancers and silencers differ from each other?
- Reflect and Apply** How do response elements modulate RNA transcription?
- Reflect and Apply** Diagram a gene that is affected by CRE and CREB, showing which proteins and nucleic acids contact each other.
- Reflect and Apply** Explain the relationship between TFIID, TBP, and TAFs.
- Reflect and Apply** Defend or attack this statement: "All eukaryotic promoters have TATA boxes."
- Reflect and Apply** Explain the different ways in which eukaryotic transcription elongation is controlled.
- Reflect and Apply** Explain the importance of CREB, giving examples of genes activated by it.
- Reflect and Apply** Give examples of structural motifs found in transcription factors that interact with other proteins instead of DNA.

### 11.6 Non-Coding RNAs

- Recall** What are micro RNAs?
- Recall** What are small interfering RNAs?
- Recall** In what processes are non-coding RNAs important?
- Recall** What is RNA interference?
- Reflect and Apply** Why do scientists think that RNA silencing is an evolutionarily conserved process?
- Reflect and Apply** What is a potential link between prostate cancer and miRNA-101?
- Biochemical Connections** How is miRNA-206 beneficial to an organism?
- Biochemical Connections** MicroRNAs play a role in physical nerve damage. What other neurological diseases are they associated with?

**11.7 Structural Motifs in DNA-Binding Proteins**

49. **Recall** List three important structural motifs in DNA-binding proteins.
50. **Reflect and Apply** Give examples of the major structural motifs in DNA-binding proteins, and explain how they bind.

**11.8 Posttranscriptional RNA Modification**

51. **Recall** List several ways in which RNA is processed after transcription.
52. **Recall** What do the proteins Tau and troponin have in common?
53. **Reflect and Apply** Why is a trimming process important in converting precursors of tRNA and rRNA to the active forms?
54. **Reflect and Apply** List three molecular changes that take place in the processing of eukaryotic mRNA.

55. **Reflect and Apply** What are snRNPs? What is their role in the processing of eukaryotic mRNAs?
56. **Reflect and Apply** What roles can RNA play, other than that of transmission of the genetic message?
57. **Reflect and Apply** Diagram the formation of a lariat in RNA processing.
58. **Reflect and Apply** Explain how differential splicing of RNA is thought to be relevant to the information gathered from the Human Genome Project.

**11.9 Ribozymes**

59. **Recall** What is a ribozyme? List some examples of ribozymes.
60. **Reflect and Apply** Outline a mechanism by which RNA can catalyze its own self-splicing.
61. **Reflect and Apply** Why are proteins more effective catalysts than RNA molecules?

**ANNOTATED BIBLIOGRAPHY**

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).

# Protein Synthesis: Translation of the Genetic Message

# 12

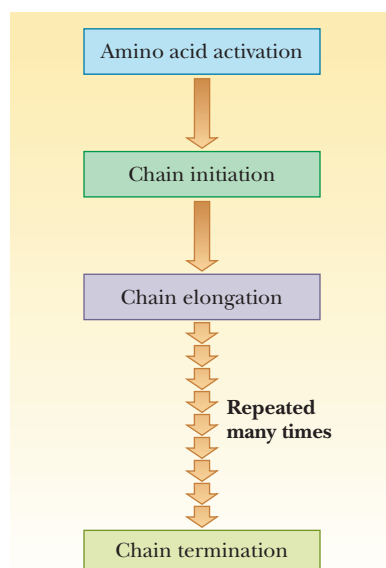
© Alfred Pasieka/Photo Researchers, Inc.

Transfer RNA brings amino acids to the site at which they are incorporated into a growing polypeptide chain.

## 12.1 Translating the Genetic Message

Protein biosynthesis is a complex process requiring ribosomes, messenger RNA (mRNA), transfer RNA (tRNA), and a number of protein factors. The ribosome is the site of protein synthesis. The mRNA and tRNA, which are bound to the ribosome in the course of protein synthesis, are responsible for the correct order of amino acids in the growing protein chain.

Before an amino acid can be incorporated into a growing protein chain, it must first be **activated**, a process involving both tRNA and a specific enzyme of the class known as **aminoacyl-tRNA synthetases**. The amino acid is covalently bonded to the tRNA in the process, forming an aminoacyl-tRNA. The actual formation of the polypeptide chain occurs in three steps. In the first step, **chain initiation**, the first aminoacyl-tRNA is bound to the mRNA at the site that encodes the start of polypeptide synthesis. In this complex, the mRNA and the ribosome are bound to each other. The next aminoacyl-tRNA forms a complex with the ribosome and with mRNA. The binding site for the second aminoacyl-tRNA is close to that for the first aminoacyl-tRNA. A peptide bond is formed between the amino acids in the second step, called **chain elongation**. The chain-elongation process repeats itself until the polypeptide chain is complete. Finally, in the third step, **chain termination** takes place. Each of these steps has many distinguishing features (Figure 12.1), and we shall look at each of them in detail.



■ **FIGURE 12.1** The steps in protein biosynthesis.

## Chapter Outline

### 12.1 Translating the Genetic Message

#### 12.2 The Genetic Code

- How did scientists determine the genetic code?
- If there are 64 codons, how can there be fewer than 64 tRNA molecules?

#### 12.3 Amino Acid Activation

- What is the “second genetic code”?

#### 12.4 Prokaryotic Translation

- How does the ribosome know where to start translating?
- Why is EF-Tu so important in *E. coli*?

#### 12.5 Eukaryotic Translation

- How is translation different in eukaryotes?

#### 12.6 Posttranslational Modification of Proteins

- Once modified, do proteins always have the correct three-dimensional structure?

#### 12.7 Protein Degradation

- How does the cell know which proteins to degrade?

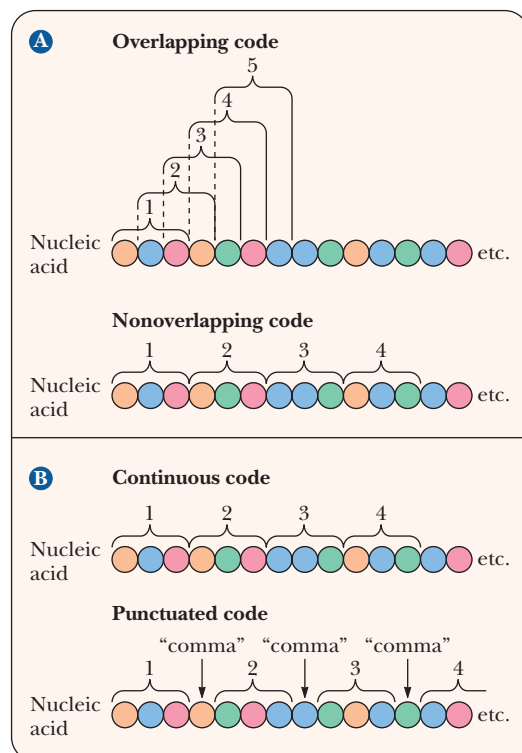
**OWL** Online homework for this chapter may be assigned in OWL.

## 12.2 The Genetic Code

Some of the most important features of the code can be specified by saying that the genetic message is contained in a *triplet*, *nonoverlapping*, *commaless*, *degenerate*, *universal code*. Each of these terms has a definite meaning that describes the way in which the code is translated.

A **triplet** code means that a sequence of three bases (called a **codon**) is needed to specify one amino acid. The genetic code must translate the language of DNA, which contains four bases, into the language of the 20 common amino acids that are found in proteins. If there were a one-to-one relationship between bases and amino acids, then the four bases could encode only four amino acids, and all proteins would have to be combinations of these four. If it took two bases to make a codon, then there would be  $4^2$  possible combinations of two bases for 16 possible amino acids, which is still not enough. Thus, one could have guessed that a codon would have to be at least three bases long. With three bases, there are  $4^3$  possibilities, or 64 possible codons, which is more than enough to encode the 20 amino acids. The term *nonoverlapping* indicates that no bases are shared between consecutive codons; the ribosome moves along the mRNA three bases at a time rather than one or two at a time (Figure 12.2). If the ribosome moved along the mRNA more than three bases at a time, this situation would be referred to as “a punctuated code.” Because no intervening bases exist between codons, the code is *commaless*. In a *degenerate* code, more than one triplet can encode the same amino acid. Sixty-four ( $4 \times 4 \times 4$ ) possible triplets of the four bases occur in RNA, and all are used to encode the 20 amino acids or one of the three stop signals. Note that there is a big difference between a degenerate code and an ambiguous one. Each amino acid may have more than one codon, so the genetic code is a little redundant, but no codon can encode more than one amino acid. If it did, the code would be ambiguous, and the protein-synthesizing machinery would not know which amino acid should be inserted in the sequence. All 64 codons have been assigned meanings, with 61 of them coding for amino acids and the remaining 3 serving as the termination signals (Table 12.1).

Two amino acids, tryptophan and methionine, have only one codon each, but the rest have more than one. A single amino acid can have as many as six codons, as is the case with leucine and arginine. Originally, the genetic code was thought to be a random selection of bases encoding amino acids. More recently, it is becoming clear why the code has withstood billions of years of natural selection. Multiple codons for a single amino acid are not randomly distributed in Table 12.1 but have one or two bases in common. The bases that are common to several codons are usually the first and second bases, with more room for variation in the third base, which is called the “wobble” base. The degeneracy of the code acts as a buffer against deleterious mutations. For example, for eight of the amino acids (L, V, S, P, T, A, G, and R), the third base is completely irrelevant. Thus, any mutation in the third base of these codons would not change the amino acid at that location. A mutation in the DNA that does not lead to a change in the amino acid translated is called a *silent mutation*. In addition, the second base of the codon also appears to be very important for determining the type of amino acid. For example, when the second base is U, all the amino acids generated from the codon possibilities are hydrophobic. Thus, if the first or third base were mutated, the mutation would not be silent, but the damage would not be as great because one hydrophobic amino acid would be replaced with another. Codons sharing the same first letter often code for amino acids that are products of one another or precursors of one another. A paper in *Scientific American* looked at the error rate for other hypothetical genetic codes and calculated that, of 1 million possible genetic codes that could be conceived of, only 100 would have the effect of reducing errors in protein function when compared with the real code. Indeed, it seems that



■ **FIGURE 12.2** Theoretically possible genetic codes. (a) An overlapping versus a nonoverlapping code. (b) A continuous versus a punctuated code.

TABLE 12.1

The Genetic Code									
First Position (5'-end)		Second Position				Third Position (3'-end)	Third-Base Degeneracy Is Color-Coded		
							Third-Base Relationship	Third Bases with Same Meaning	Number of Codons
	U	C	A	G					
U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys	U	<div></div> Third base irrelevant	U, C, A, G	32 (8 families)	
	UUC Phe	UCC Ser	UAC Tyr	UGC Cys	C		<div></div> Purines	A or G	12 (6 pairs)
	UUA Leu	UCA Ser	UAA Stop	UGA Stop	A		<div></div> Pyrimidines	U or C	14 (7 pairs)
	UUG Leu	UCG Ser	UAG Stop	UGG Trp	G		<div></div> Three out of four	U, C, A	3 (AUX = Ile)
C	CUU Leu	CCU Pro	CAU His	CGU Arg	U	<div></div> Unique definitions	G only	2 (AUG = Met) (UGG = Trp)	
	CUC Leu	CCC Pro	CAC His	CGC Arg	C				
	CUA Leu	CCA Pro	CAA Gln	CGA Arg	A				
	CUG Leu	CCG Pro	CAG Gln	CGG Arg	G				
A	AUU Ile	ACU Thr	AAU Asn	AGU Ser	U	<div></div> Unique definition	A only	1 (UGA = Stop)	
	AUC Ile	ACC Thr	AAC Asn	AGC Ser	C				
	AUA Ile	ACA Thr	AAA Lys	AGA Arg	A				
	AUG Met*	ACG Thr	AAG Lys	AGG Arg	G				
G	GUU Val	GCU Ala	GAU Asp	GGU Gly	U				
	GUC Val	GCC Ala	GAC Asp	GGC Gly	C				
	GUA Val	GCA Ala	GAA Glu	GGA Gly	A				
	GUG Val	GCG Ala	GAG Glu	GGG Gly	G				

\* AUG signals translation initiation as well as coding for Met residues.

the genetic code has withstood the test of time because it is one of the best ways to protect an organism from DNA mutations.

However, to make matters more interesting, scientists have recently discovered that some silent mutations are not as silent as they once thought. See the Biochemical Connections on page 348 for more on this topic.

### How did scientists determine the genetic code?

The assignment of triplets in the genetic code was based on several types of experiments. One of the most significant experiments involved the use of synthetic polyribonucleotides as messengers. When homopolynucleotides (polyribonucleotides that contain only one type of base) are used as a *synthetic mRNA* for polypeptide synthesis in laboratory systems, homopolypeptides (polypeptides that contain only one kind of amino acid) are produced. When poly U is the messenger, the product is polyphenylalanine. With poly A as the messenger, polylysine is formed. The product for poly C is polyproline, and the product for poly G is polyglycine. This procedure was used to establish the code for the four possible homopolymers quickly. When an alternating copolymer (a polymer with an alternating sequence of two bases) is the messenger, the product is an alternating polypeptide (a polypeptide with an alternating sequence of two amino acids). For example, when the sequence of the polynucleotide is –ACACACACACACACACAC–, the polypeptide produced has alternating threonines and histidines. There are two types of coding triplets in this polynucleotide, ACA and CAC, but this experiment cannot establish which one codes for threonine and which one codes for histidine. More information is needed



for an unambiguous assignment, but it is interesting that this result proves that the code is a triplet code. If it were a doublet code, the product would be a mixture of two homopolymers, one specified by the codon AC and the other by the codon CA. (The terminology for the different ways of reading this message as a doublet is to say that they have different **reading frames**, /AC/AC/ and /CA/CA/. In a triplet code, only one reading frame is possible, namely, /ACA/CAC/ACA/CAC/, which gives rise to an alternating polypeptide.) Use of other synthetic polynucleotides can yield other coding assignments, but, as in our example here, many questions remain.

Other methods are needed to answer the remaining questions about codon assignment. One of the most useful methods is the **filter-binding assay** (Figure 12.3). In this technique, various tRNA molecules, one of which is radioactively labeled with carbon-14 ( $^{14}\text{C}$ ), are mixed with ribosomes and synthetic trinucleotides that are bound to a filter. The mixture of tRNAs is passed through the filter, and some bind and others pass through. If the radioactive label is detected on the filter, then it is known that the particular tRNA did bind. If the radioactive label is found in a solution that flowed through the filter, then the tRNA did not bind. This technique depends on the fact that aminoacyl-tRNAs bind strongly to ribosomes in the presence of the correct trinucleotide. In this situation, the trinucleotide plays the role of an mRNA codon. The possible trinucleotides are synthesized by chemical methods, and binding assays are repeated with each type of trinucleotide. For example, if the aminoacyl-tRNA for histidine binds to the ribosome in the presence of the trinucleotide CAU, the sequence CAU is established as a codon for histidine. About 50 of the 64 codons were identified by this method.

### Codon–Anticodon Pairing and Wobble

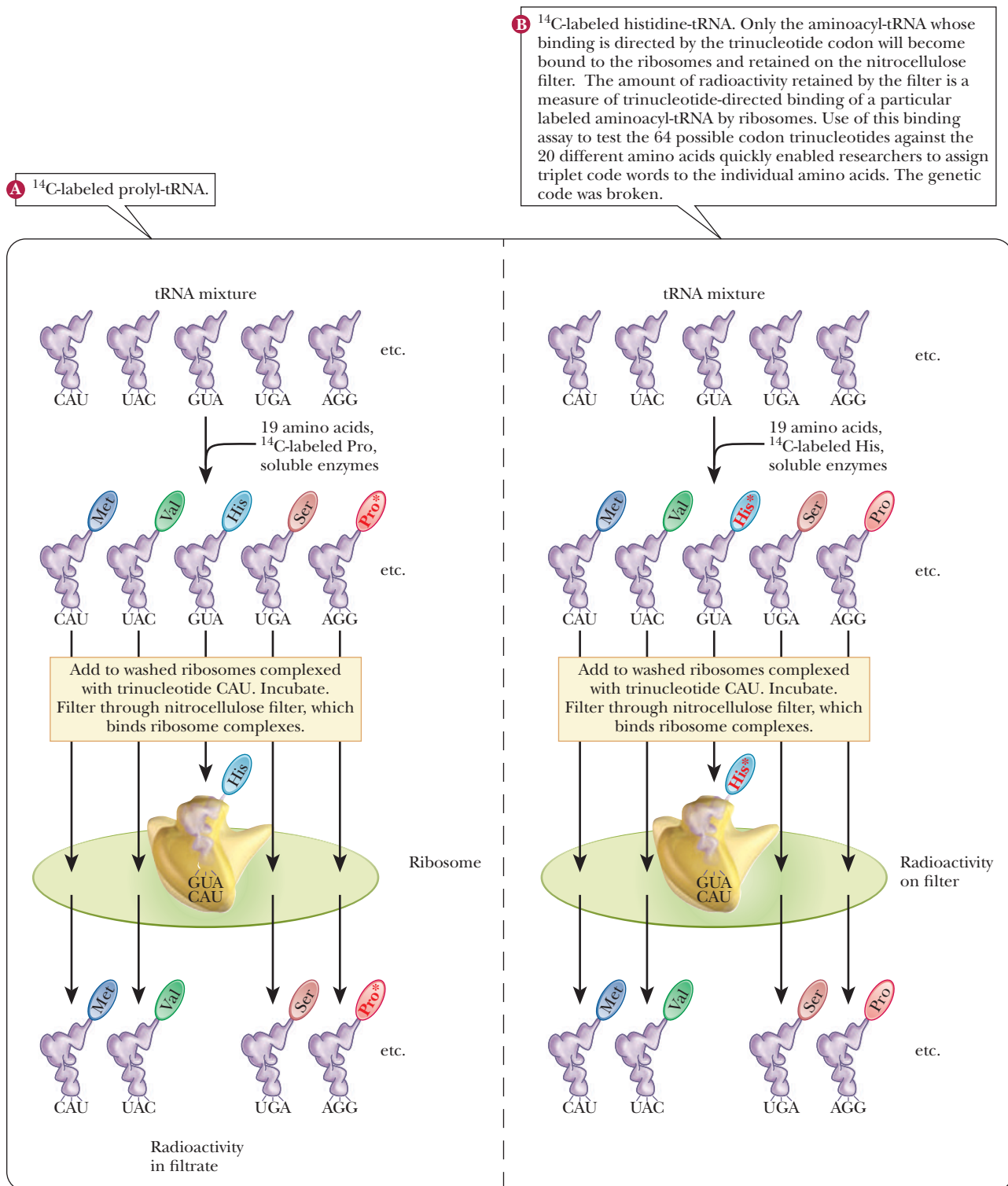
A codon forms base pairs with a complementary **anticodon** of a tRNA when an amino acid is incorporated during protein synthesis. Because there are 64 possible codons, one might expect to find 64 types of tRNA but, in fact, the number is less than 64 in all cells.

#### *If there are 64 codons, how can there be fewer than 64 tRNA molecules?*

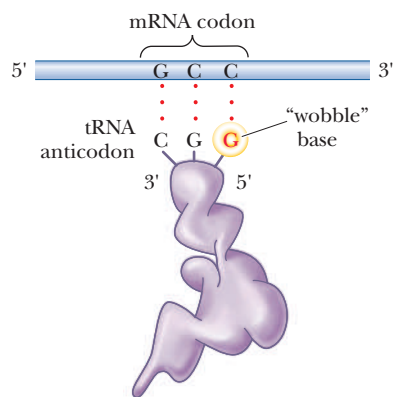
Some tRNAs bond to one codon exclusively, but many of them can recognize more than one codon because of variations in the allowed pattern of hydrogen bonding. This variation is called “**wobble**” (Figure 12.4), and it applies to the first base of an anticodon, the one at the 5' end, but not to the second or the third base. Recall that mRNA is read from the 5' to the 3' end. The first (wobble) base of the anticodon hydrogen-bonds to the third base of the codon, the one at the 3' end. The base in the wobble position of the anticodon can base-pair with several different bases in the codon, not just the base specified by Watson–Crick base pairing (Table 12.2).

When the wobble base of the anticodon is uracil, it can base-pair not only with adenine, as expected, but also with guanine, the other purine base. When the wobble base is guanine, it can base-pair with cytosine, as expected, and also with uracil, the other pyrimidine base. The purine base hypoxanthine frequently occurs in the wobble position in many tRNAs, and it can base-pair with adenine, cytosine, and uracil in the codon (Figure 12.5). Adenine and cytosine do not form any base pairs other than the expected ones with uracil and guanine, respectively (Table 12.2). To summarize, when the wobble position is occupied by I (from inosine, the nucleoside made up of ribose and hypoxanthine), G, or U, variations in hydrogen bonding are allowed; when the wobble position is occupied by A or C, these variations do not occur.

The wobble model provides insight into some aspects of the degeneracy of the code. In many cases, the degenerate codons for a given amino acid differ in



**FIGURE 12.3 The filter-binding assay for elucidation of the genetic code.** A reaction mixture combines washed ribosomes,  $\text{Mg}^{2+}$ , a particular trinucleotide, and all 20 aminoacyl-tRNAs, one of which is radioactively ( $^{14}\text{C}$ ) labeled. (Adapted from Nirenberg, M. W., and Leder, P., 1964. RNA Codewords and Protein Synthesis. *Science* **145**, 1399–1407.)



■ **FIGURE 12.4** “Wobble” base pairing. The wobble base of the anticodon is the one at the 5' end; it forms hydrogen bonds with the last base of the mRNA codon, the one at the 3' end of the codon. (Adapted by permission from Crick, F. H. C., 1966. Codon–anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology* **19**, 548–555.)

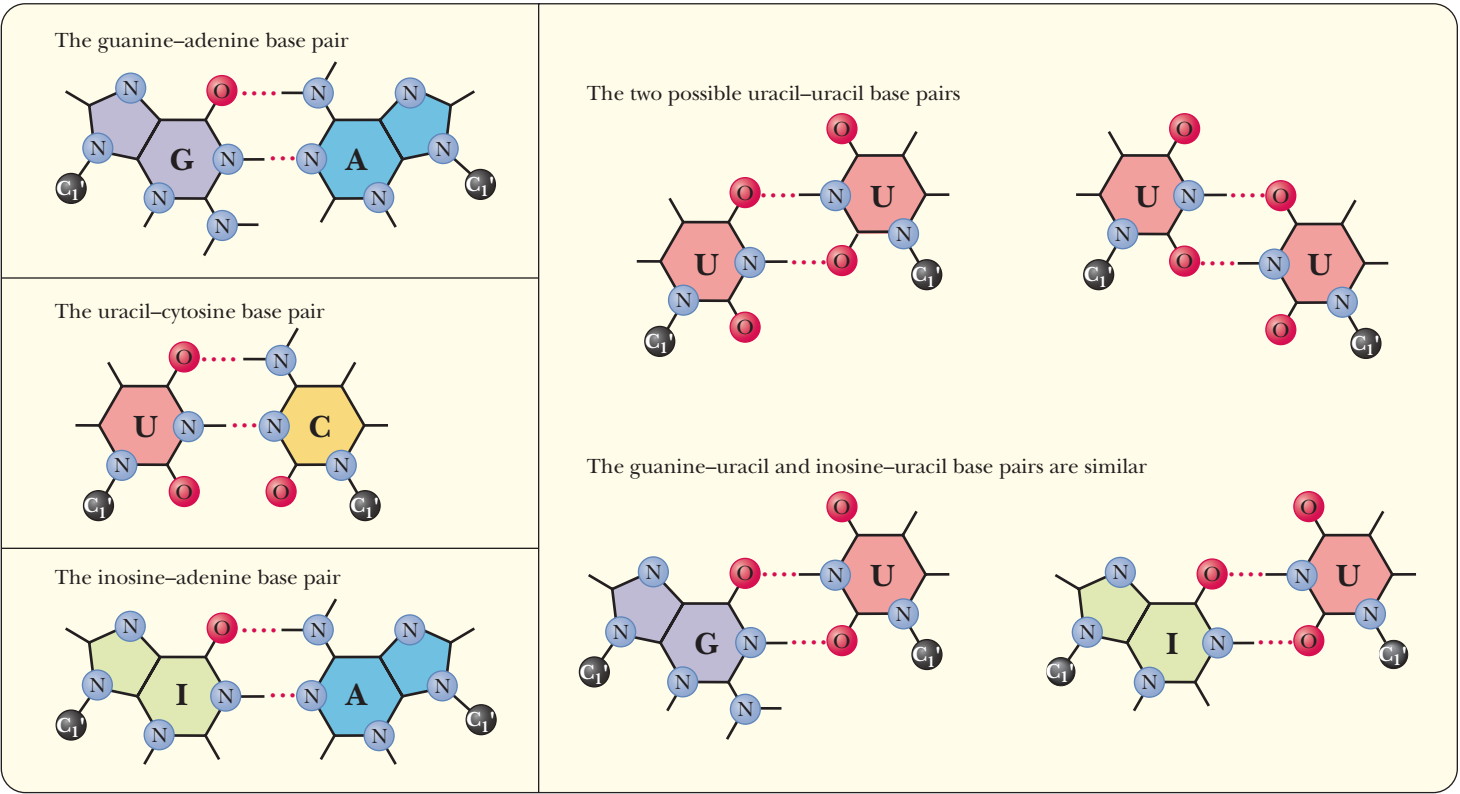
**TABLE 12.2**

Base-Pairing Combinations in the Wobble Scheme	
Base at 5' End of Anticodon	Base at 5' End of Codon
I*	A, C, or U
G	C or U
U	A or G
A	U
C	G

\* I = hypoxanthine.  
Note that there are no variations in base pairing when the wobble position is occupied by A or C.

the third base, the one that pairs with the wobble base of the anticodon. Fewer different tRNAs are needed because a given tRNA can base-pair with several codons. As a result, a cell would have to invest less energy in the synthesis of needed tRNAs. The existence of wobble also minimizes the damage that can be caused by misreading of the code. If, for example, a leucine codon, CUU, were to be misread as CUC, CUA, or CUG during transcription of mRNA, this codon would still be translated as leucine during protein synthesis; no damage to the organism would occur. We saw in earlier chapters that drastic consequences can result from misreading the genetic code in other codon positions, but here we see that such effects are not inevitable.

A *universal* code is one that is the same in all organisms. The universality of the code has been observed in viruses, prokaryotes, and eukaryotes. However,

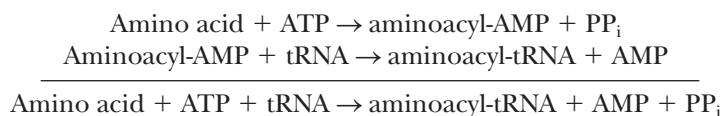


■ **FIGURE 12.5** Various base-pairing alternatives. G:A is unlikely, because the 2-NH<sub>2</sub> of G cannot form one of its H bonds; even water is sterically excluded. U:C may be possible, even though the two C=O are juxtaposed. Two U:U arrangements are feasible. G:U and I:U are both possible and somewhat similar. The purine pair I:A is also possible. (Adapted from Crick, F. H. C., 1966. Codon–anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology* **19**, 548–555.)

there are some exceptions. Some codons seen in mitochondria are different from those seen in the nucleus. There are also at least 16 organisms that have code variations. For example, the marine alga *Acetabularia* translates the standard stop codons, UAG and UAA, as a glycine rather than as a stop. Fungi of the genus *Candida* translate the codon CUG as a serine, where that codon would specify leucine in most organisms. The evolutionary origin of these differences is not known at this writing, but many researchers believe that understanding these code variations is important to understanding evolution.

## 12.3 Amino Acid Activation

The activation of the amino acid and the formation of the aminoacyl-tRNA take place in two separate steps, both of which are catalyzed by the aminoacyl-tRNA synthetase (Figure 12.6). First, the amino acid forms a covalent bond to an adenine nucleotide, producing an aminoacyl-AMP. The free energy of hydrolysis of ATP provides energy for bond formation. The aminoacyl moiety is then transferred to tRNA, forming an aminoacyl-tRNA.

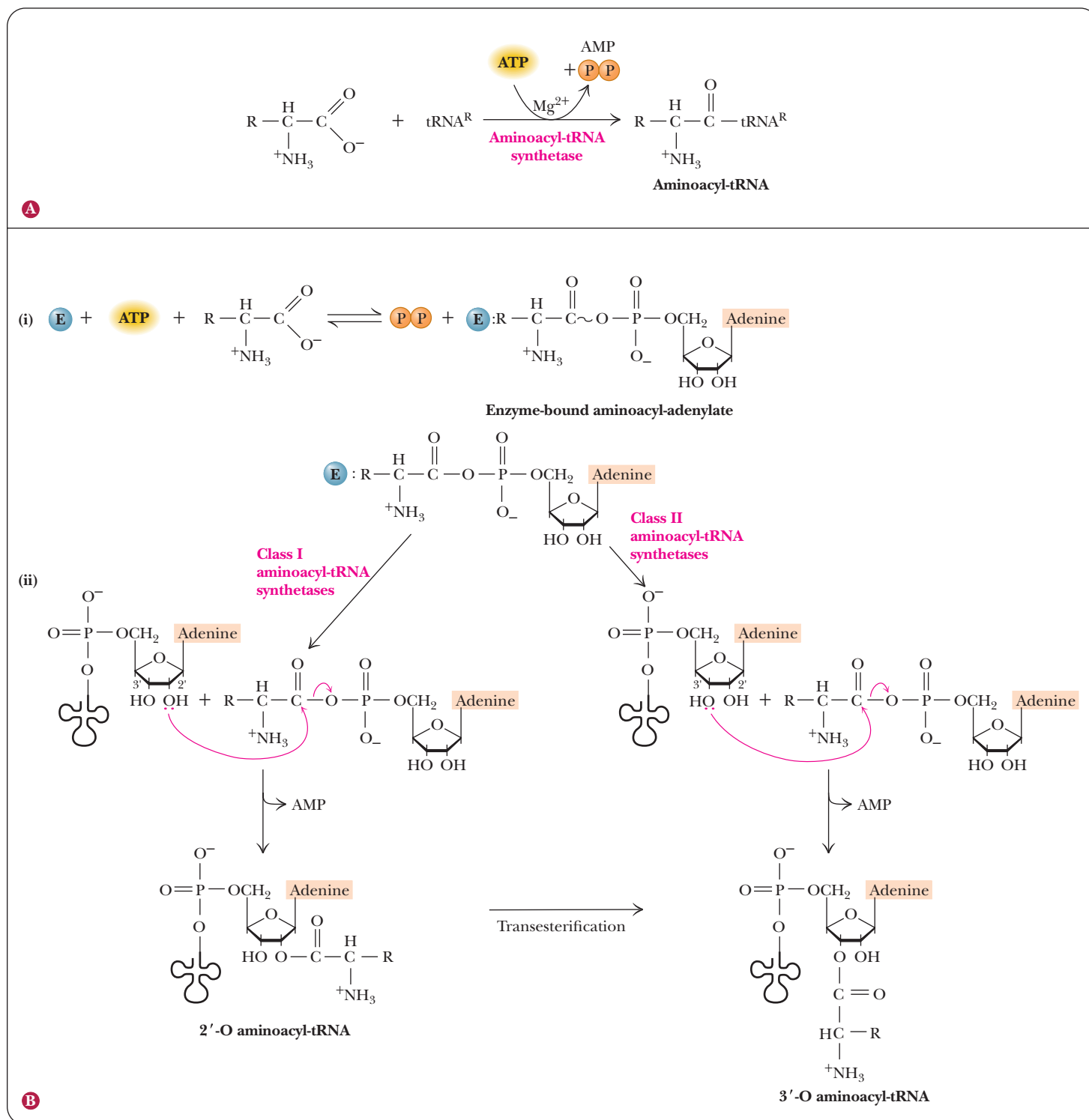


Aminoacyl-AMP is a mixed anhydride of a carboxylic acid and a phosphoric acid. Because anhydrides are reactive compounds, the free-energy change for the hydrolysis of aminoacyl-AMP favors the second step of the overall reaction. Another point that favors the process is the energy released when pyrophosphate ( $\text{PP}_i$ ) is hydrolyzed to orthophosphate ( $\text{P}_i$ ) to replenish the phosphate pool in the cell.

In the second part of the reaction, an ester linkage is formed between the amino acid and either the 3'-hydroxyl or the 2'-hydroxyl of the ribose at the 3' end of the tRNA. There are two classes of aminoacyl-tRNA synthetases. Class I loads the amino acid onto the 2' hydroxyl. Class II uses the 3' hydroxyl. These two classes of enzyme appear to be unrelated and indicate a convergent evolution. Several tRNAs can exist for each amino acid, but a given tRNA does not bond to more than one amino acid. The synthetase enzyme requires  $\text{Mg}^{2+}$  and is highly specific both for the amino acid and for the tRNA. A separate synthetase exists for each amino acid, and this synthetase functions for all the different tRNA molecules for that amino acid. The specificity of the enzyme contributes to the accuracy of the translation process. A student who used an earlier edition of this book compared the mode of action of the aminoacyl-tRNA synthetases to a “dating service” for amino acids and tRNAs. The synthetase assures that the right amino acid pairs up with the right tRNA, and this is its primary function. The synthetase has another level of activity as well. An extra level of proofreading by the synthetase is part of what is sometimes called the “second genetic code.”

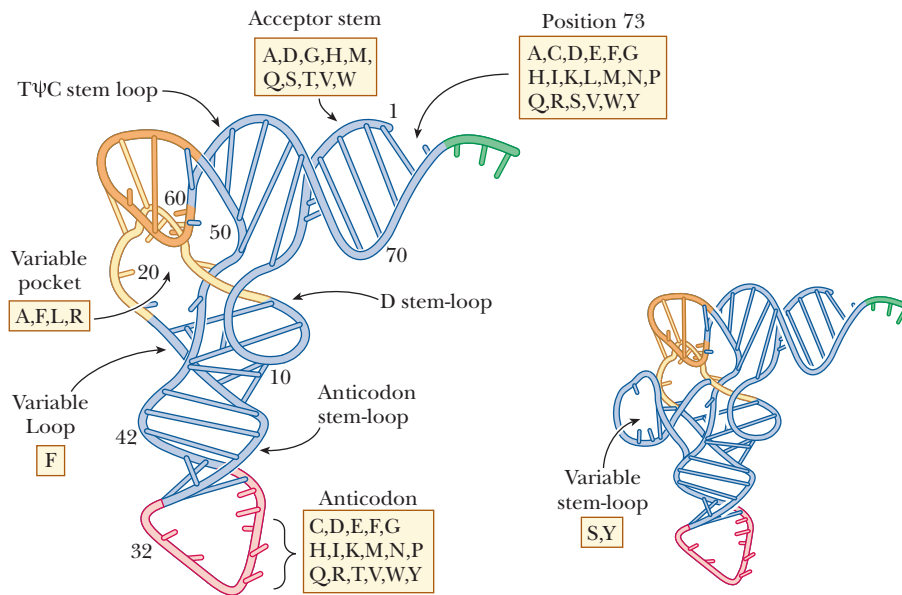
### What is the “second genetic code”?

The two-stage reaction allows for selectivity to operate at two levels: that of the amino acid and that of the tRNA. The specificity of the first stage uses the fact that the aminoacyl-AMP remains bound to the enzyme. For example, isoleucyl-tRNA synthetase can form an aminoacyl-AMP of isoleucine or the structurally similar valine. If the valyl moiety is then transferred to the tRNA for isoleucine, it is detected by an editing site in the tRNA synthetase, which then hydrolyzes



■ **FIGURE 12.6 The aminoacyl-tRNA synthetase reaction.** (a) The overall reaction. Ever-present pyrophosphatases in cells quickly hydrolyze the  $\text{PP}_i$  produced in the aminoacyl-tRNA synthetase reaction, rendering aminoacyl-tRNA synthesis thermodynamically favorable and essentially irreversible. (b) The overall reaction commonly proceeds in two steps: (i) formation of an aminoacyl-adenylate and (ii) transfer of the activated amino acid moiety of the mixed anhydride to either the 2'-OH (class I aminoacyl-tRNA synthetases) or 3'-OH (class II aminoacyl-tRNA synthetases) of the ribose on the terminal adenylic acid at the 3'-OH terminus common to all tRNAs. Those aminoacyl-tRNAs formed as 2'-OH esters undergo a transesterification that moves the aminoacyl group to the 3'-OH of tRNA. Only the 3'-esters are substrates for protein synthesis.





**FIGURE 12.7** Ribbon diagram of the tRNA tertiary structure. Numbers represent the consensus nucleotide sequence. The locations of nucleotides recognized by the various aminoacyl-tRNA synthetases are indicated; shown within the boxes are one-letter designations of the amino acids whose respective aminoacyl-tRNA synthetases interact at the discriminator base (position 73), acceptor stem, variable pocket and/or loop, or anticodon. The inset shows additional recognition sites in those tRNAs having a variable loop that forms a stem-loop structure. (Adapted with permission from Saks, M. E., Sampson, J. R., and Abelson, J. N., 1994. *The transfer RNA problem: A search for rules*. Science 263, 191–197, Figure 2. Copyright © 1994 AAAS.)

the incorrectly acylated aminoacyl-tRNA. The selectivity resides in the tRNA, not in the amino acid.

The second aspect of selectivity depends on the selective recognition of tRNAs by aminoacyl-tRNA synthetases. Specific binding sites on tRNA are recognized by aminoacyl-tRNA synthetases. The exact position of the recognition site varies with different synthetases, and this feature, in and of itself, is a source of greater specificity. Contrary to what one might expect, the anticodon is not always the part of the tRNA that is recognized by the aminoacyl-tRNA synthetase, although it frequently is involved. Figure 12.7 shows the locations of the recognition sites for the tRNAs for various amino acids.

The recognition of the correct tRNA by the synthetase is vital to the fidelity of translation because most of the final proofreading occurs at this step.

## 12.4 Prokaryotic Translation

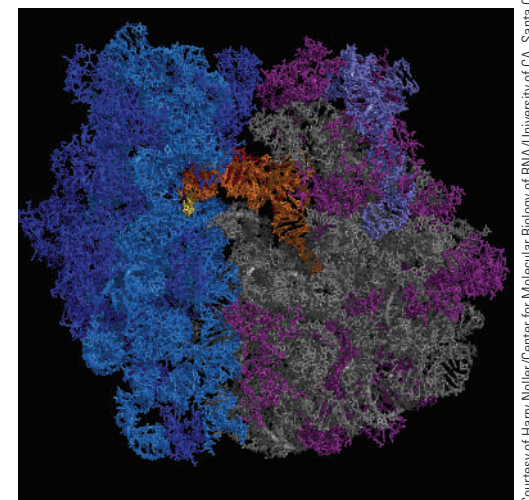
The details of the chain of events in translation differ somewhat in prokaryotes and eukaryotes. Like DNA and RNA synthesis, this process has been more thoroughly studied in prokaryotes. We shall use *Escherichia coli* as our principal example, because all aspects of protein synthesis have been most extensively studied in this bacterium. As was the case with replication and transcription, translation can be divided into stages—chain initiation, chain elongation, and chain termination.

### Ribosomal Architecture

Protein synthesis requires the specific binding of mRNA and aminoacyl-tRNAs to the ribosome. Ribosomes have a specific architecture that facilitates the binding. In Figure 12.8, a tRNA molecule (shown in orange) is base pairing with part of the mRNA (gold) on the left. The tRNA extends into the peptidyl-transferase center on the right. Elucidation of the details of ribosomal structure is a recent triumph of X-ray crystallography.

### Chain Initiation

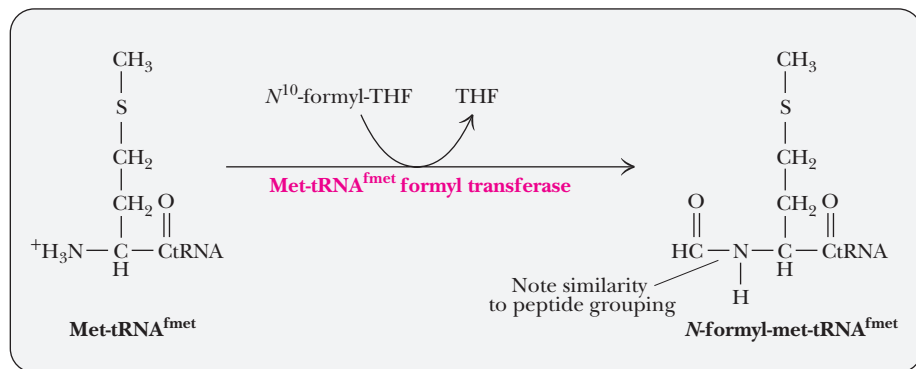
In all organisms, the synthesis of polypeptide chains starts at the N-terminal end; the chain grows from the N-terminal end to the C-terminal end. This



**FIGURE 12.8** Ribosome structure as determined by X-ray crystallography. A tRNA is shown in orange. The portion of the mRNA hydrogen bonded to the tRNA can be seen in the left (gold). The peptidyl transferase is shown in gray on the right.

Courtesy of Harry Noller/Center for Molecular Biology of RNA/University of CA, Santa Cruz

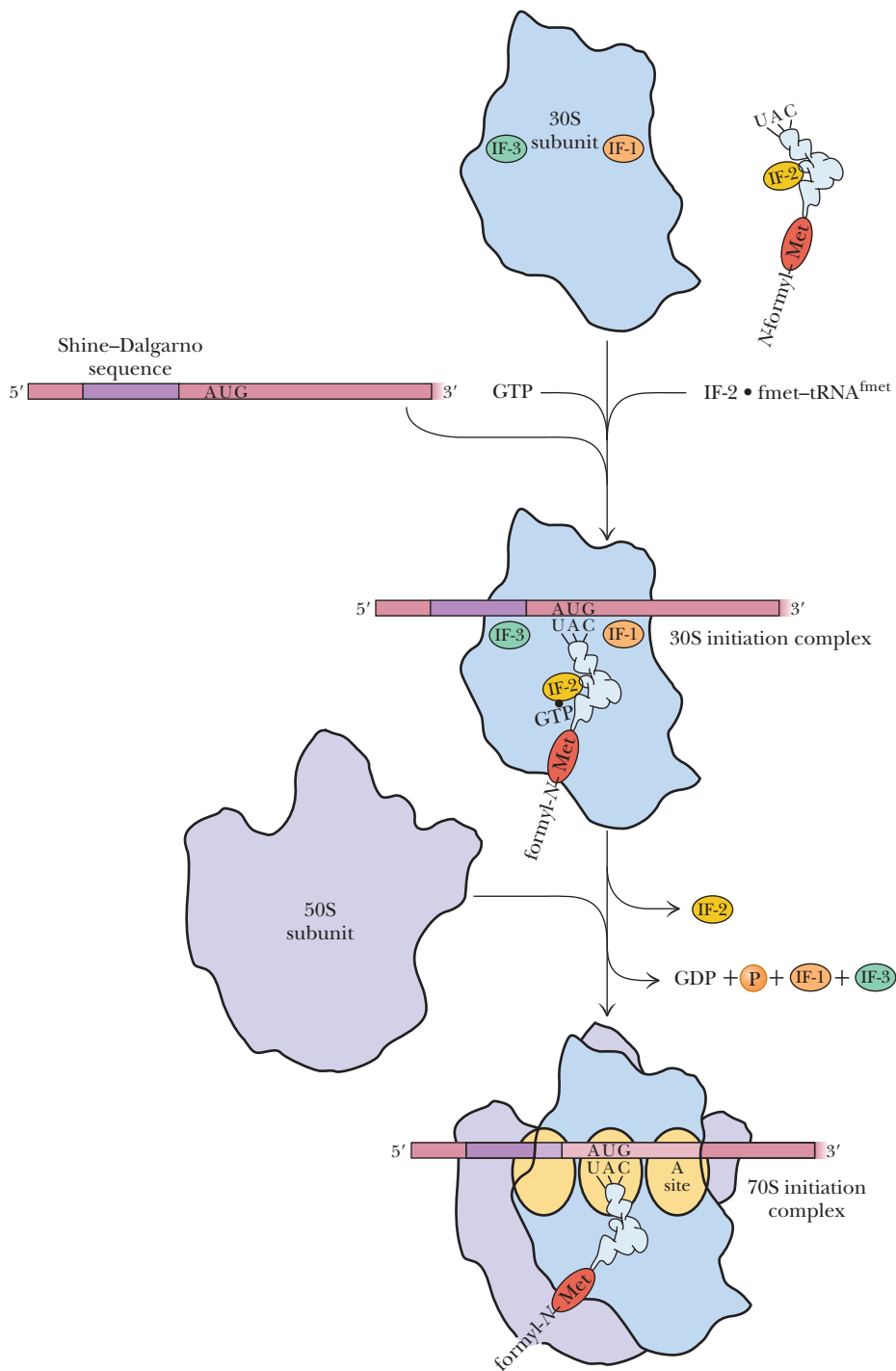
■ **FIGURE 12.9** Formation of the *N*-formylmethionine-tRNA<sup>fmet</sup> (first reaction). Methionine must be bound to tRNA<sup>fmet</sup> to be formylated.



is one of the reasons that scientists chose to record DNA sequences from 5' to 3' and to focus on the coding strand of DNA and the mRNA. The coding strand sequences are read from 5' to 3', the mRNA is read from 5' to 3', and the proteins are built from the N-terminus to the C-terminus. In prokaryotes, the initial N-terminal amino acid of all proteins is *N*-formylmethionine (fmet) (Figure 12.9). However, this residue often is removed by posttranslational processing after the polypeptide chain is synthesized. There are two different tRNAs for methionine in *E. coli*, one for unmodified methionine and one for *N*-formylmethionine. These two tRNAs are called tRNA<sup>met</sup> and tRNA<sup>fmet</sup>, respectively (the superscript identifies the tRNA). The aminoacyl-tRNAs that they form with methionine are called met-tRNA<sup>met</sup> and met-tRNA<sup>fmet</sup>, respectively (the prefix identifies the bound amino acid). In the case of met-tRNA<sup>fmet</sup>, a formylation reaction takes place after methionine is bonded to the tRNA, producing ***N*-formylmethionine-tRNA<sup>fmet</sup>** (fmet-tRNA<sup>fmet</sup>). The source of the formyl group is *N*<sup>10</sup>-formyltetrahydrofolate (see “One-Carbon Transfers and the Serine Family” in Section 23.4). Methionine bound to tRNA<sup>met</sup> is not formylated.

Both tRNAs (tRNA<sup>met</sup> and tRNA<sup>fmet</sup>) contain a specific sequence of three bases (a triplet), 3'-UAC-5', which base-pairs with the sequence 5'-AUG-3' in the mRNA sequence. The tRNA<sup>fmet</sup> triplet in question, 3'-UAC-5', recognizes the AUG triplet, which is the **start signal** when it occurs at the beginning of the mRNA sequence that directs the synthesis of the polypeptide. The same 3'-UAC-5' triplet in tRNA<sup>met</sup> recognizes the AUG triplet when it is found in an internal position in the mRNA sequence.

The start of polypeptide synthesis requires the formation of an **initiation complex** (Figure 12.10). At least eight components enter into the formation of the initiation complex, including mRNA, the 30S ribosomal subunit, fmet-tRNA<sup>fmet</sup>, GTP, and three protein initiation factors, called IF-1, IF-2, and IF-3. The IF-3 protein facilitates the binding of mRNA to the 30S ribosomal subunit. It also appears to prevent premature binding of the 50S subunit, which takes place in a subsequent step of the initiation process. IF-2 binds GTP and aids in the selection of the initiator tRNA (fmet-tRNA<sup>fmet</sup>) from all the other aminoacylated tRNAs available. The function of IF-1 is less clear; it appears to bind to IF-3 and to IF-2, and it facilitates the action of both. It also catalyzes the separation of the 30S and the 50S ribosomal subunits being recycled for another round of translation. The resulting combination of mRNA, the 30S ribosomal subunit, and fmet-tRNA<sup>fmet</sup> is the **30S initiation complex** (Figure 12.10). A 50S ribosomal subunit binds to the 30S initiation complex to produce the **70S initiation complex**. The hydrolysis of GTP to GDP and P<sub>i</sub> favors the process by providing energy; the initiation factors are released at the same time. The correct positioning of the initiator tRNA is maintained as a result of a small difference between it and tRNA for an internal methionine. A single C–A mismatched base pair near the acceptor stem allows the 30S subunit to recognize the initiator tRNA.



■ **FIGURE 12.10** The formation of an initiation complex. The 30S ribosomal subunit binds to mRNA and fmet-tRNA<sup>fmet</sup> in the presence of GTP and the three initiation factors, IF-1, IF-2, and IF-3, forming the 30S initiation complex. The 50S ribosomal subunit is added, forming the 70S initiation complex.

### How does the ribosome know where to start translating?

For the mRNA to be translated correctly, the ribosome must be placed at the correct start location. The start signal is preceded by a purine-rich leader segment of mRNA, called the **Shine-Dalgarno sequence** (5'-GGAGGU-3') (Figure 12.10), which usually lies about 10 nucleotides upstream of the AUG start signal (also known as the initiation codon) and acts as a ribosomal binding site. Figure 12.11 gives some characteristic Shine-Dalgarno sequences. This purine-rich area binds to a pyrimidine-rich sequence on the 16S ribosomal RNA part of the 30S subunit and aligns it for proper translation beginning with the AUG start codon.

■ **FIGURE 12.11** Various Shine–Dalgarno sequences recognized by *E. coli* ribosomes. These sequences lie about 10 nucleotides upstream from their respective AUG initiation codon and are complementary to the UCCU core sequence element of *E. coli* 16S rRNA. G:U as well as canonical G:C and A:U base pairs are involved here.

		Initiation codon
<i>araB</i>	- U U U G G A U G G A G U G A A A C G A U G G C G A U U -	AUG
<i>galE</i>	- A G C C U A A U G G A G C G A A U U A U G A G A G U U -	AUG
<i>lacI</i>	- C A A U U C A G G G U G G U G A U U G U G A A A C C A -	AUG
<i>lacZ</i>	- U U C A C A C A G G A A A C A G C U A U G A C C A U G -	AUG
Q $\beta$ phage replicase	- U A A C U A A G G A U G A A A U G C A U G U C U A A G -	AUG
$\phi$ X174 phage A protein	- A A U C U U G G A G G C U U U U U U A U G G U U C G U -	AUG
R17 phage coat protein	- U C A A C C G G G G U U U G A A G C A U G G C U U C U -	AUG
ribosomal protein S12	- A A A A C C A G G A G C U A U U U A A U G G C A A C A -	AUG
ribosomal protein L10	- C U A C C A G G A G C A A A G C U A A U G G C U U U A -	AUG
<i>trpE</i>	- C A A A A U U A G A G A A U A A C A A U G C A A A C A -	AUG
<i>trpL</i> leader	- G U A A A A A G G G U A U C G A C A A U G A A A G C A -	AUG
3'-end of 16S rRNA	3' HO A U U C C U C C A C U A G - 5'	

## Chain Elongation

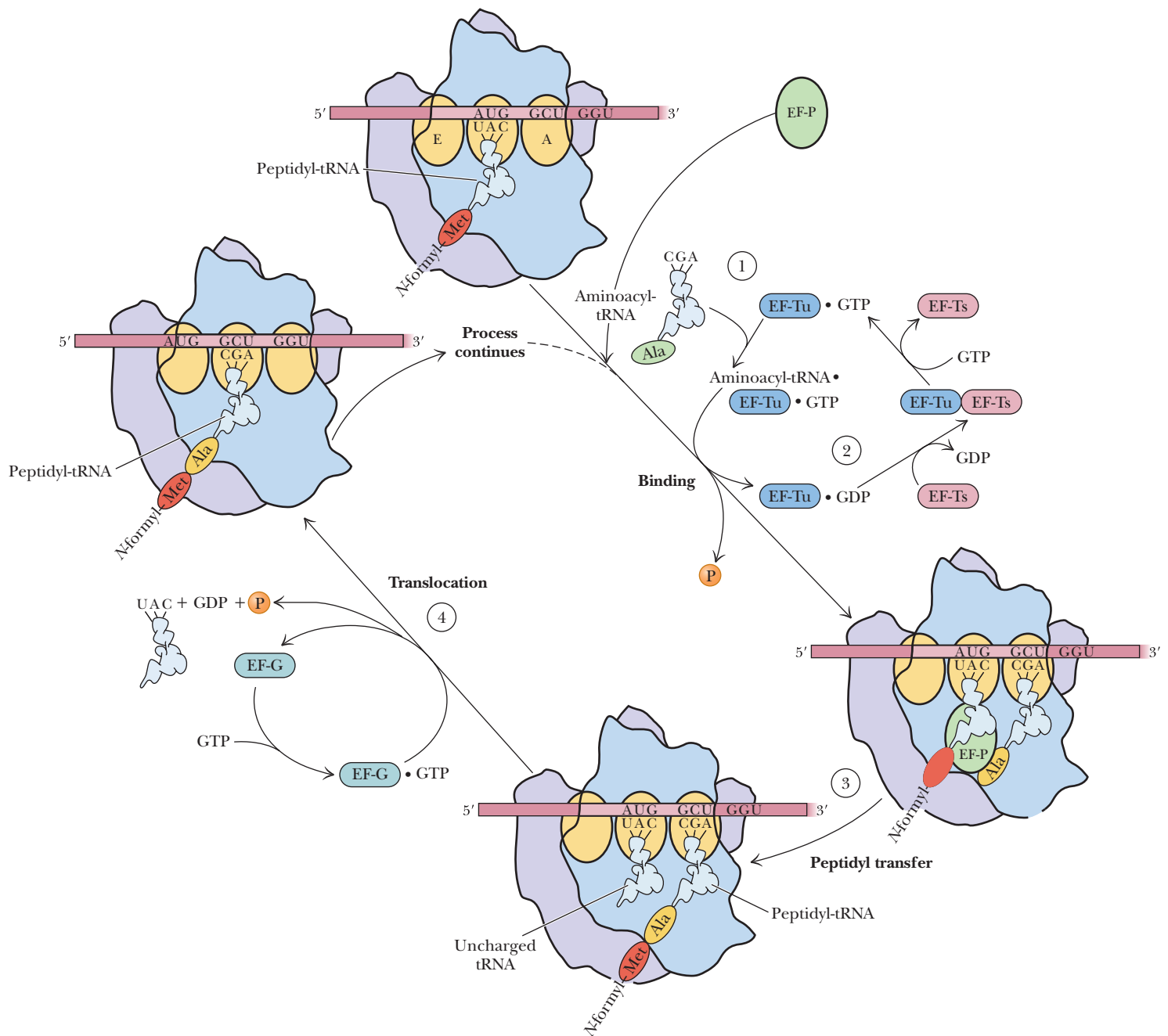
The elongation phase of prokaryotic protein synthesis (Figure 12.12) uses the fact that three binding sites for tRNA are present on the 50S subunit of the 70S ribosome. The three tRNA binding sites are called the **P (peptidyl) site**, the **A (aminoacyl) site**, and the **E (exit) site**. The P site binds a tRNA that carries a peptide chain, and the A site binds an incoming aminoacyl-tRNA. The E site carries an uncharged tRNA that is about to be released from the ribosome. Chain elongation begins with the addition of the second amino acid specified by the mRNA to the 70S initiation complex (Step 1). The P site on the ribosome is the one initially occupied by the fmet-tRNA<sup>fmet</sup> in the 70S initiation complex. The second aminoacyl-tRNA binds at the A site. A triplet of tRNA bases (the anticodon AGC in our example) forms hydrogen bonds with a triplet of mRNA bases (GCU, the codon for alanine, in this example). In addition, GTP and three protein elongation factors, EF-P, EF-Tu and EF-Ts (temperature-unstable and temperature-stable elongation factors, respectively), are required (Step 2). EF-Tu guides the aminoacyl-tRNA into part of the A site and aligns the anticodon with the mRNA codon. Only when the match is found to be correct is the aminoacyl-tRNA inserted completely into the A site. GTP is hydrolyzed and EF-Tu dissociates. EF-P is bound adjacent to the P site and E site and is thought to help catalyze the first peptide bond formed. EF-Ts is involved in regeneration of EF-Tu-GTP. This small EF-Tu protein (43 kDa) is the most abundant protein in *E. coli*, comprising 5% of the dry weight of the cell. The exact mechanism of hydrolysis of GTP and the release of the tRNA by EF-Tu is an active area of research. In late 2009, researchers determined the structure of the ribosome and EF-Tu complex to 3.6-angstrom resolution.

### Why is EF-Tu so important in *E. coli*?

It has recently been shown that EF-Tu is involved in another level of translation fidelity. When the correct amino acid is bound to the correct tRNA, EF-Tu is efficient at delivering the activated tRNA to the ribosome. If the tRNA and amino acid are mismatched, then either the EF-Tu does not bind the activated tRNA very well, in which case it does not deliver it well to the ribosome, or it binds the activated tRNA too well, in which case it does not release it from the ribosome.

A **peptide bond** is then formed in a reaction catalyzed by *peptidyl transferase*, which is a part of the 50S subunit (Step 3). The mechanism for this reaction is shown in Figure 12.13. The  $\alpha$ -amino group of the amino acid in the A site performs a nucleophilic attack on the carbonyl group of the amino acid linked to the tRNA in the P site. There is now a dipeptidyl-tRNA at the A site and a tRNA with no amino acid attached (an “*uncharged tRNA*”) at the P site.

A **translocation** step then takes place before another amino acid can be added to the growing chain (Figure 12.12, Step 4). In the process, the



■ **FIGURE 12.12** A summary of the steps in chain elongation. Step 1: An aminoacyl-tRNA is bound to the A site on the ribosome. Elongation factor EF-Tu (Tu) and GTP are required. The P site on the ribosome is already occupied. Step 2: Elongation factor EF-Tu is released from the ribosome and regenerated in a process requiring elongation factor EF-Ts (Ts) and GTP. Step 3: The peptide bond is formed, leaving an uncharged tRNA at the P site. Step 4: In the translocation step, the uncharged tRNA is released. The peptidyl-tRNA is translocated to the P site, leaving an empty A site. The uncharged tRNA is translocated to the E site and subsequently released. Elongation factor EF-G and GTP are required.

uncharged tRNA moves from the P site to the E site, from which it is subsequently released; the peptidyl-tRNA moves from the A site to the vacated P site. In addition, the mRNA moves with respect to the ribosome. Another elongation factor, EF-G, also a protein, is required at this point, and once again GTP is hydrolyzed to GDP and P<sub>i</sub>.

The three steps of the chain elongation process are aminoacyl-tRNA binding, peptide bond formation, and translocation (Steps 1, 3, and 4 in



■ **FIGURE 12.13** Peptide bond formation in protein synthesis. Nucleophilic attack by the  $\alpha$ -amino group of the A-site aminoacyl-tRNA on the carbonyl-C of the P-site peptidyl-tRNA is facilitated when a purine moiety of rRNA abstracts a proton.

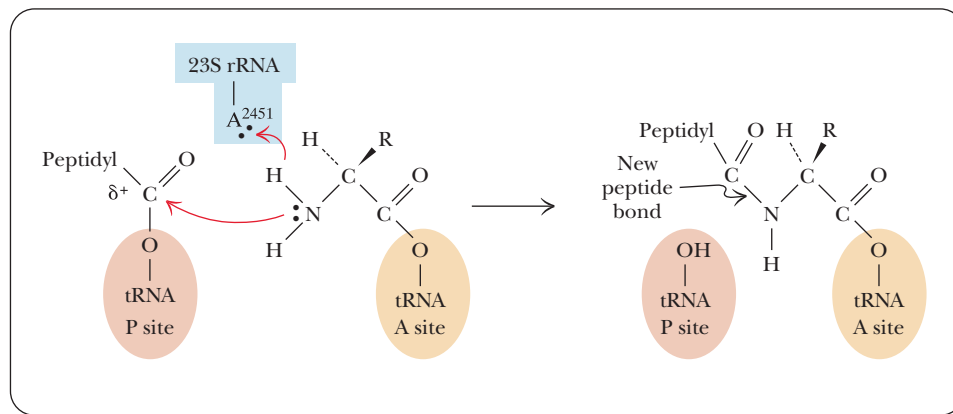


Figure 12.12). They are repeated for each amino acid specified by the genetic message of the mRNA until the stop signal is reached. Step 2 in Figure 12.12 shows the regeneration of aminoacyl-tRNA.

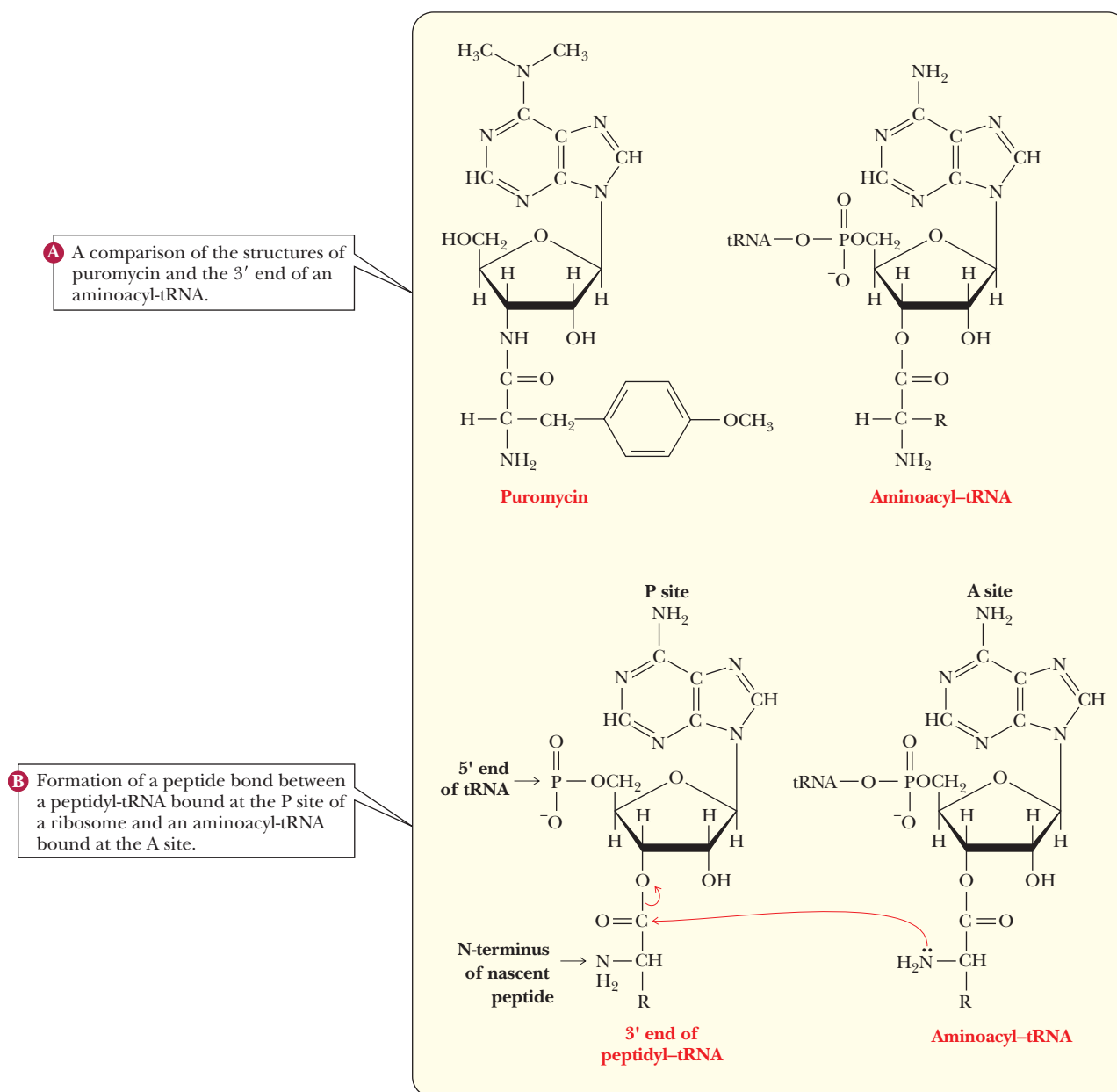
Much of the information about this phase of protein synthesis has been gained from the use of inhibitors. Puromycin is a structural analog for the 3' end of an aminoacyl-tRNA, making it a useful probe to study chain elongation (Figure 12.14). In an experiment of this sort, puromycin binds to the A site, and a peptide bond is formed between the C-terminus of the growing polypeptide and the puromycin. The peptidyl puromycin is weakly bound to the ribosome and dissociates from it easily, resulting in premature termination and a defective protein. Puromycin also binds to the P site and blocks the translocation process, although it does not react with peptidyl-tRNA in this case. The existence of A and P sites was determined by these experiments with puromycin.

### Chain Termination

A stop signal is required for the termination of protein synthesis. The codons UAA, UAG, and UGA are the stop signals. These codons are not recognized by any tRNAs, but they are recognized by proteins called release factors (Figure 12.15). One of two protein release factors (RF-1 or RF-2) is required, as is GTP, which is bound to a third release factor, RF-3. RF-1 binds to UAA and UAG, and RF-2 binds to UAA and UGA. RF-3 does not bind to any codon, but it does facilitate the activity of the other two release factors. Either RF-1 or RF-2 is bound near the A site of the ribosome when one of the termination codons is reached. Recent studies have shown that a conserved sequence of Pro-X-Thr in RF1 and Ser-Pro-Phe in RF2 controls their codon specificities. The release factor not only blocks the binding of a new aminoacyl-tRNA but also affects the activity of the peptidyl transferase so that the bond between the carboxyl end of the peptide and the tRNA is hydrolyzed. A conserved sequence of Gly-Gly-Gln is essential for the hydrolysis reaction of the RF. GTP is hydrolyzed in the process. The whole complex dissociates, setting free the release factors, tRNA, mRNA, and the 30S and 50S ribosomal subunits. All these components can be reused in further protein synthesis. Table 12.3 summarizes the steps in protein synthesis and the components required for each step. The following Biochemical Connections box describes an interesting variation on stop codons.

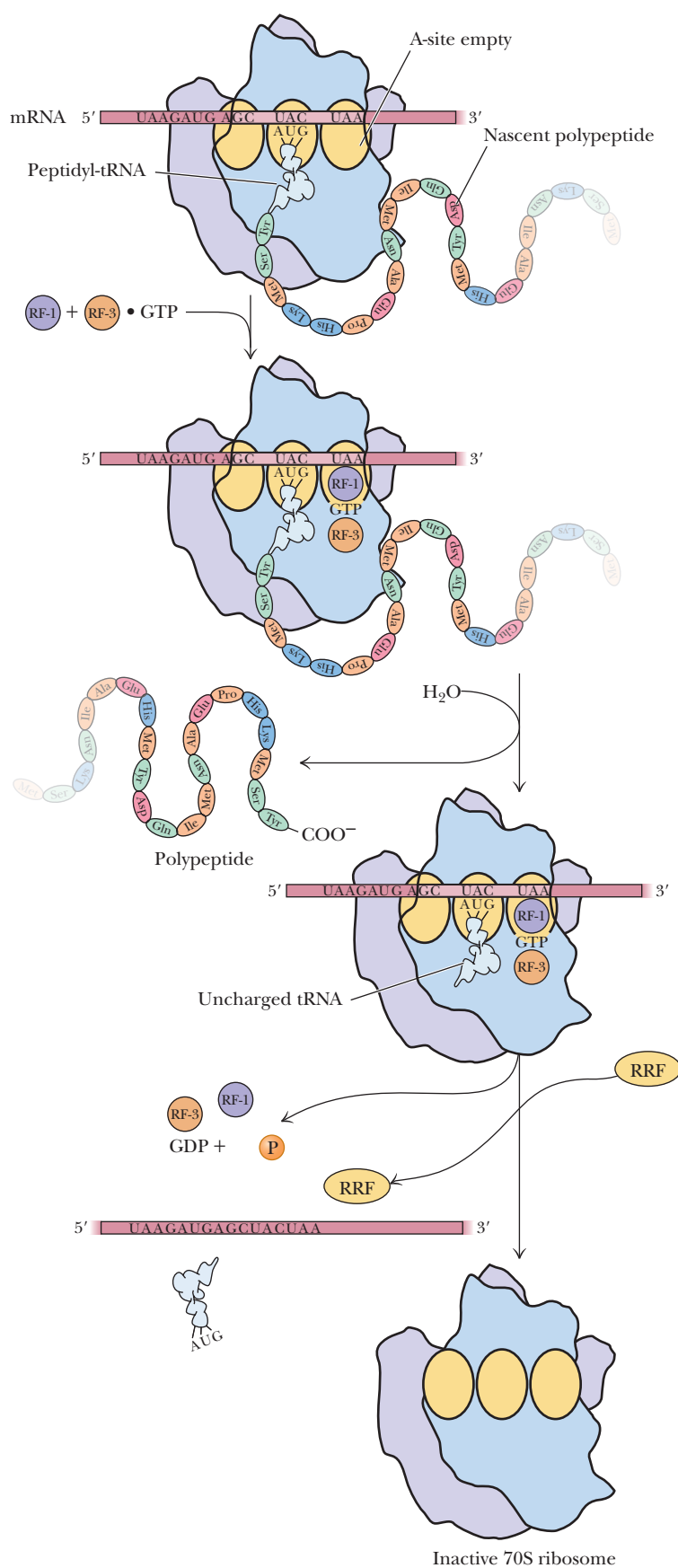
### The Ribosome Is a Ribozyme

Until recently, proteins were thought to be the only molecules with catalytic ability. Then the self-splicing ability of the *Tetrahymena* snRNP showed that RNA can also catalyze reactions. In 2000, the complete structure of the large ribosomal subunit was determined by X-ray crystallography to 2.4 Å (0.24-nm)



■ **FIGURE 12.14** The mode of action of puromycin.

resolution (Figure 12.16). Ribosomes had been studied for 40 years, but the complete structure had been elusive. When the active sites for peptidyl transferase were looked at, it turned out that there is no protein in the vicinity of the new peptide bond, showing once again that RNA has catalytic ability. This is an exciting finding because it answers questions that have been plaguing scientists for decades. It was assumed that RNA was the first genetic material, and RNA can encode proteins that act as catalysts; but, because it takes proteins to do the translation, how could the first proteins have been created? With the discovery of an RNA-based peptidyl transferase, it was suddenly possible to imagine an “RNA world” in which the RNA both carried the message and processed it. This discovery is very intriguing, but it has not yet been accepted by many researchers, and some evidence questions the nature of catalytic RNA. One study showed that mutations of the putative RNA bases involved in the catalytic mechanism do not significantly reduce the efficiency of peptidyl transferase, throwing into question whether the RNA is chemically involved in the catalysis.



■ **FIGURE 12.15** The events in peptide chain termination.

TABLE 12.3

Components Required for Each Step of Protein Synthesis in <i>Escherichia coli</i>	
Step	Components
Amino acid activation	Amino acids tRNAs Aminoacyl-tRNA synthetases ATP, Mg <sup>2+</sup>
Chain initiation	fmet-tRNA <sup>fmet</sup> Initiation codon (AUG) of mRNA 30S ribosomal subunit 50S ribosomal subunit Initiation factors (IF-1, IF-2, and IF-3) GTP, Mg <sup>2+</sup>
Chain elongation	70S ribosome Codons of mRNA Aminoacyl-tRNAs Elongation factors (EF-Tu, EF-Ts, EF-P, and EF-G) GTP, Mg <sup>2+</sup>
Chain termination	70S ribosome Termination codons (UAA, UAG, and UGA) of mRNA Release factors (RF-1, RF-2, and RF-3) GTP, Mg <sup>2+</sup>

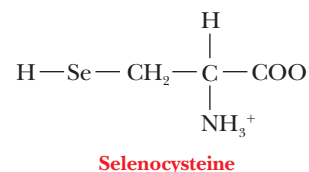
## Biochemical Connections MOLECULAR GENETICS

### The 21st Amino Acid

Many amino acids, such as citrulline and ornithine (which are found in the urea cycle), are not building blocks of proteins. Other nonstandard amino acids, such as hydroxyproline, are formed after translation by posttranslational modification. When discussing amino acids and translation, the magic number was always 20. Only 20 standard amino acids were put onto tRNA molecules for protein synthesis. In the late 1980s, another amino acid was found in proteins from eukaryotes and prokaryotes alike, including humans. It is selenocysteine, a cysteine residue in which the sulfur atom has been replaced by a selenium atom.

Selenocysteine (Sec) is unique because it is the only amino acid that both lacks its own tRNA synthetase and is synthesized while bound to its tRNA. Its importance in life was demonstrated by the fact that strains of mice lacking the ability to make selenocysteine are not viable. It has been found in the active sites of enzymes involved in clearing reactive oxidative molecules during thyroid hormone activation.

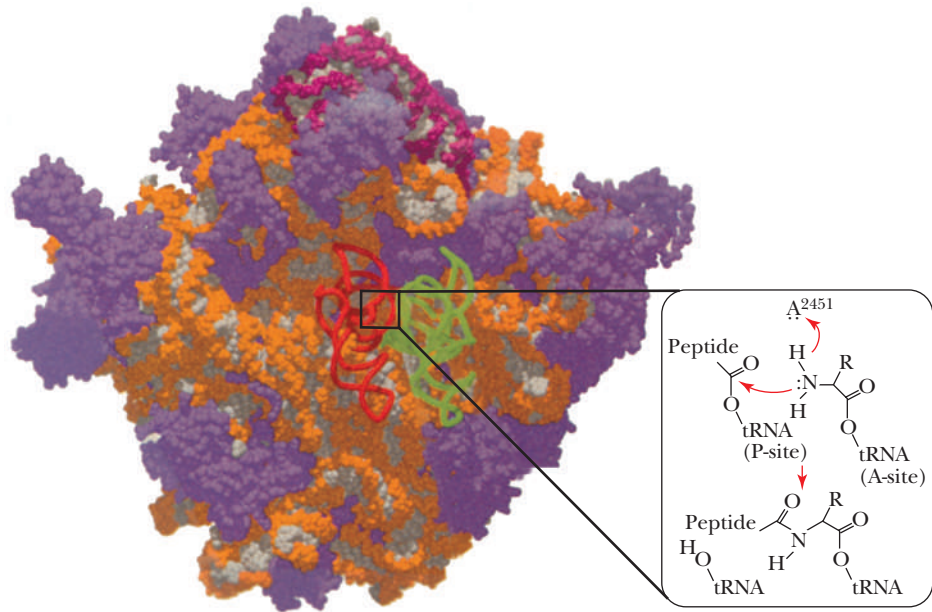
The codon for Sec is UGA, which normally is a translation stop codon. During translation of an mRNA that should lead to a Sec-bearing protein, the UGA is interpreted differently due to interactions with specialized elongation factors, SelB in bacteria and EFsec in humans. The first step in Sec synthesis is the misacylation



■ Selenocysteine.

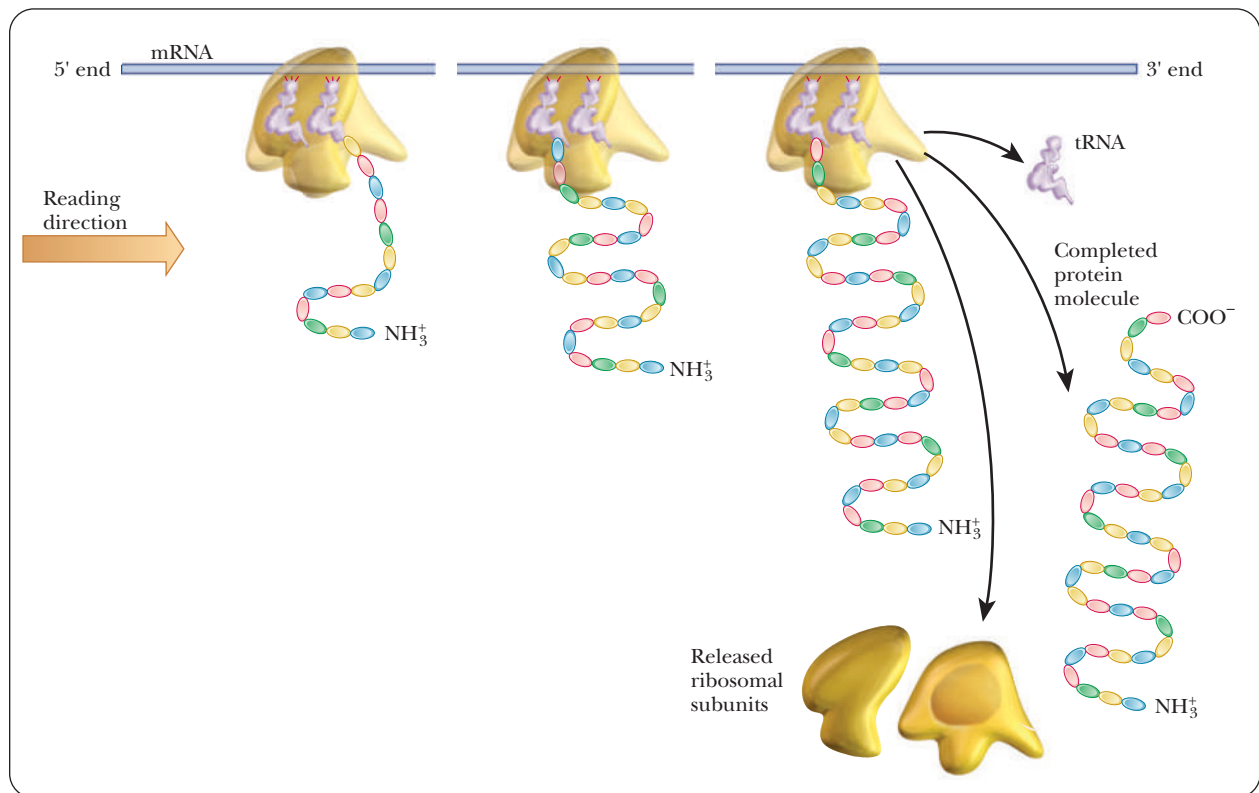
of tRNA<sup>Sec</sup> by the synthetase that normally uses serine, seryl-tRNA synthetase, to yield Ser-tRNA<sup>Sec</sup>. Then the  $\beta$ -hydroxyl group of Ser-tRNA<sup>Sec</sup> is phosphorylated by *O*-phosphoseryl-tRNA kinase to give *O*-phosphoseryl-tRNA<sup>Sec</sup>, which in turn gives the substrate used for the final reaction that catalyzes the conversion of the phosphoseryl group into the selenocysteinyl group.

■ **FIGURE 12.16 Ribosome structure.** The large subunit of the ribosome seen from the viewpoint of the small subunit, with proteins in purple, 23S rRNA in orange and white, 5S rRNA (at the top) in burgundy and white, and A-site tRNA (green) and P-site tRNA (red) docked. In the box, the peptidyl transfer mechanism is catalyzed by RNA. The general base (adenine 2451 in *E. coli* 23S rRNA) is rendered unusually basic by its environment within the folded structure; it could abstract a proton at any of several steps, one of which is shown here. (Reprinted by permission of Thomas Cech. *The ribosome is a ribozyme*. Science **289**, p. 878. Copyright © 2000 AAAS.)



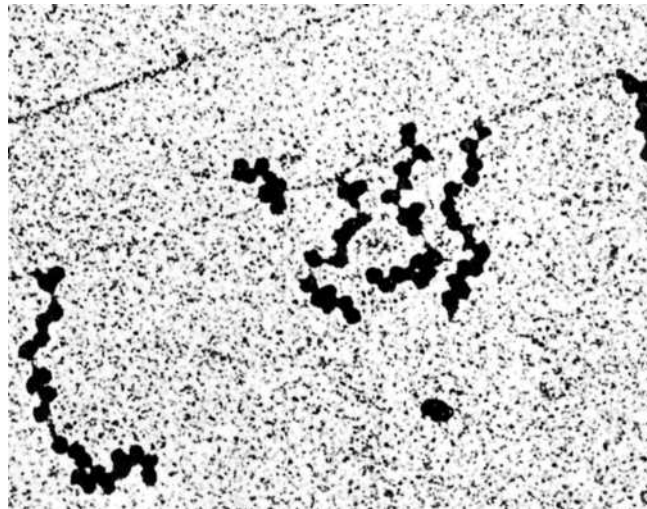
## Polysomes

In our description of protein synthesis, we have considered, up to now, the reactions that take place at one ribosome. It is, however, not only possible but quite usual for several ribosomes to be attached to the same mRNA. Each of these ribosomes bears a polypeptide in one of various stages of completion, depending on the position of the ribosome as it moves along the mRNA (Figure 12.17).



■ **FIGURE 12.17 Simultaneous protein synthesis on polysomes.** A single mRNA molecule is translated by several ribosomes simultaneously. Each ribosome produces one copy of the polypeptide chain specified by the mRNA. When the protein has been completed, the ribosome dissociates into subunits that are used in further rounds of protein synthesis.



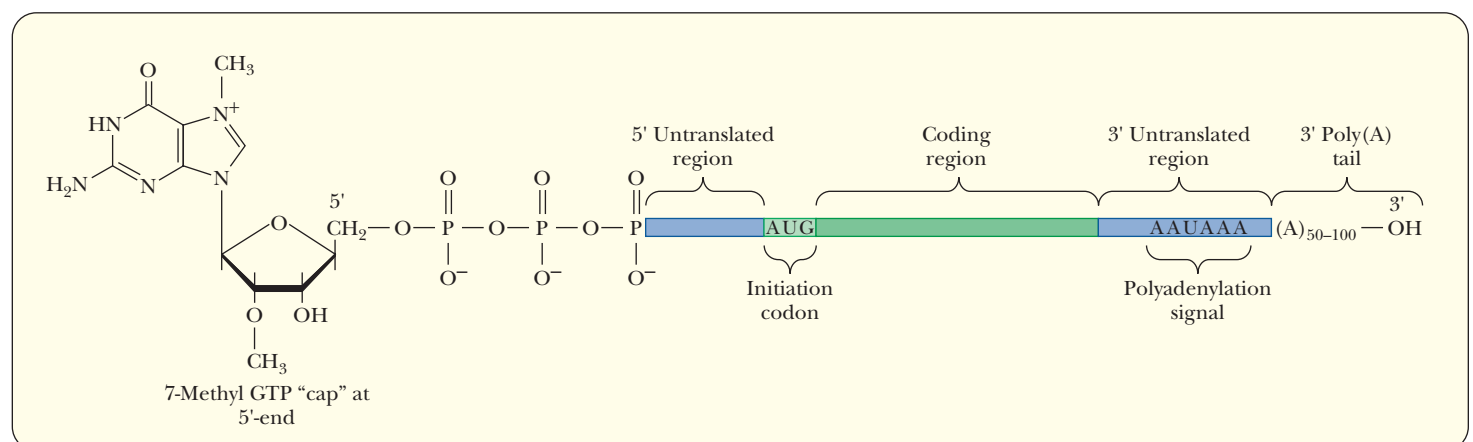


■ **FIGURE 12.18** Electron micrograph showing coupled translation. The dark spots are ribosomes, arranged in clusters on a strand of mRNA. Several mRNAs have been transcribed from one strand of DNA (diagonal line from center left to upper right). (From *Visualization of Bacterial Genes in Action* by O. L. Miller, Jr., B. A. Hamkalo, and C. A. Thomas, Jr. (24 July 1970) Science 169 (3943), 392. Used with permission of AAAS.)

This complex of mRNA with several ribosomes is called a **polysome**; an alternative name is *polyribosome*. In prokaryotes, translation begins very soon after mRNA transcription. It is possible for a molecule of mRNA that is still being transcribed to have a number of ribosomes attached to it that are in various stages of translating that mRNA. It is also possible for DNA to be in various stages of being transcribed. In this situation, several molecules of RNA polymerase are attached to a single gene, giving rise to several mRNA molecules, each of which has a number of ribosomes attached to it. The prokaryotic gene is being simultaneously transcribed and translated. This process, which is called *coupled translation* (Figure 12.18), is possible in prokaryotes because of the lack of cell compartmentalization. In eukaryotes, mRNA is produced in the nucleus, and most protein synthesis takes place in the cytosol.

## 12.5 Eukaryotic Translation

The main features of translation are the same in prokaryotes and eukaryotes, but the details differ. The messenger RNAs of eukaryotes are characterized by two major posttranscriptional modifications. The first is the 5' cap, and the second is the 3' poly A tail (Figure 12.19). Both modifications are essential to eukaryotic translation.



■ **FIGURE 12.19** The characteristic structure of eukaryotic mRNAs. Untranslated regions ranging between 40 and 150 bases in length occur at both the 5' and 3' ends of the mature mRNA. An initiation codon at the 5' end, invariably AUG, signals the translation start site.

How is translation different in eukaryotes?

Chain Initiation

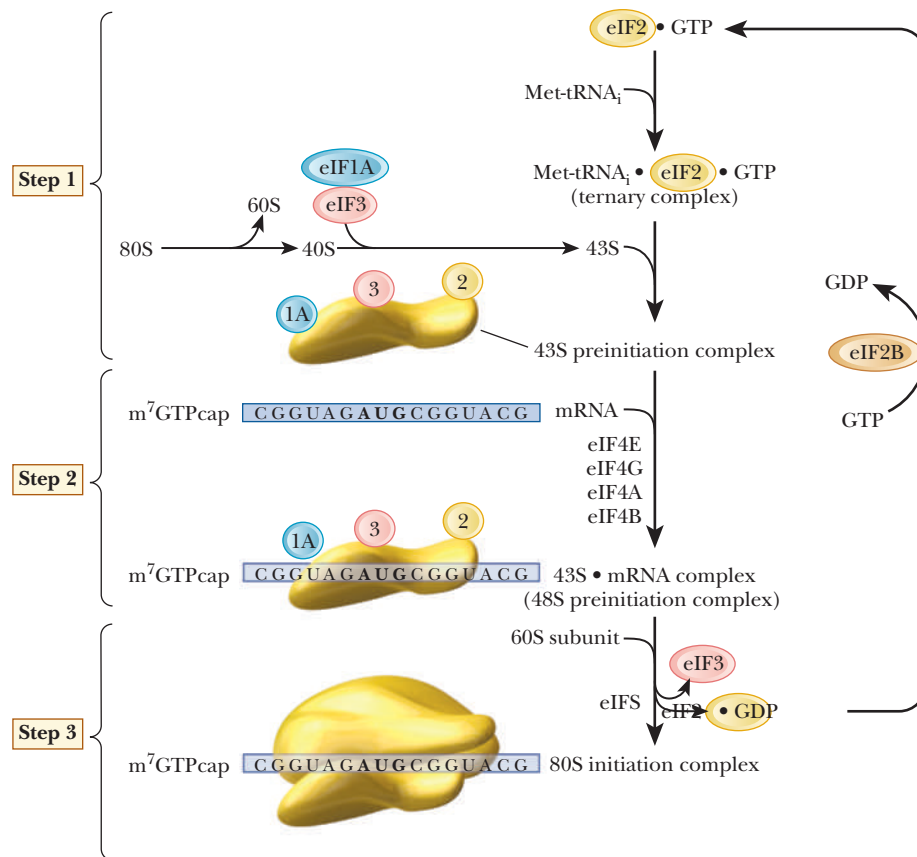
This is the part of eukaryotic translation that is the most different from that in prokaryotes. Thirteen more initiation factors are given the designation **eIF**, for **eukaryotic initiation factor**. Many of them are multisubunit proteins. Table 12.4 summarizes pertinent information about these initiation factors.

Step 1 in chain initiation involves the assembly of a 43S preinitiation complex (Figure 12.20). The initial amino acid is methionine, which is attached to a special tRNA<sub>i</sub> that serves only as the initiator tRNA. There is no fmet in eukaryotes. The met-tRNA<sub>i</sub> is delivered to the 40S ribosomal subunit as a complex with GTP and eIF2. The 40S ribosome is also bound to eIF1A and eIF3. This order of events is different from that in prokaryotes in that the first tRNA binds to the ribosome without the presence of the mRNA. In Step 2, the mRNA is recruited. There is no Shine–Dalgarno sequence for location of the start codon. The **5' cap** orients the ribosome to the correct AUG via what is called a *scanning mechanism*, which is driven by ATP hydrolysis. The eIF4E is also a cap-binding protein, which forms a complex with several other eIFs. A **poly A binding**

TABLE 12.4

Properties of Eukaryotic Translation Initiation Factors			
Factor	Subunit	Size (kDa)	Function
eIF1		15	Enhances initiation complex formation
eIF1A		17	Stabilizes Met-tRNA <sub>i</sub> binding of 40S ribosomes
eIF2		125	GTP-dependent Met-tRNA <sub>i</sub> binding to 40S ribosomes
	α	36	Regulated by phosphorylation
	β	50	Binds Met-tRNA <sub>i</sub>
	γ	55	Binds GTP, Met-tRNA <sub>i</sub>
eIF2B		270	Promotes guanine nucleotide exchange on eIF2
	α	26	Binds GTP
	β	39	Binds ATP
	γ	58	Binds ATP
	δ	67	Regulated by phosphorylation
	ε	82	
eIF2C		94	Stabilizes ternary complex in presence of RNA
eIF3		550	Promotes Met-tRNA <sub>i</sub> and mRNA binding
	p35	35	
	p36	36	
	p40	40	
	p44	44	
	p47	47	
	p66	66	Binds RNA
	p115	115	Major phosphorylated subunit
	p170	170	
eIF4A		46	Binds RNA; ATPase; RNA helicase; promotes mRNA binding to 40S ribosomes
eIF4B		80	Binds mRNA; promotes RNA helicase activity and mRNA binding to 40S ribosomes
eIF4E		25	Binds to mRNA caps
eIF4G		153.4	Binds eIF4A, eIF4E, and eIF3
eIF4F			Complex binds to mRNA caps; RNA helicase activity; promotes mRNA binding to 40S
eIF5		48.9	Promotes GTPase of eIF2, ejection of eIF
eIF6			Dissociates 80S; binds to 60S

Adapted from Clark, B. F. C., et al., eds. 1996. Prokaryotic and eukaryotic translation factors. *Biochimie* **78**, 1119–1122.



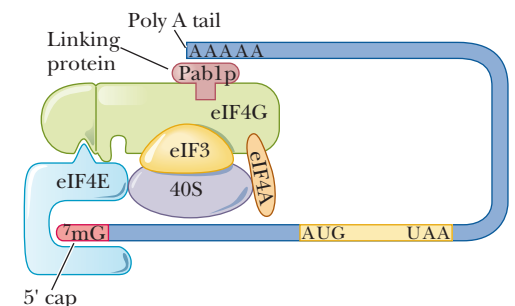
■ **FIGURE 12.20** The three stages in the initiation of translation in eukaryotic cells. See Table 12.4 for a description of the functions of the eukaryotic initiation factors (eIFs).

**protein (Pab1p)** links the **poly A tail** to eIF4G. The eIF-40S complex is initially positioned upstream of the start codon (Figure 12.21). It moves downstream until it encounters the first AUG in the correct context. The context is determined by a few bases surrounding the start codon, called the **Kozak sequence**. It is characterized by the consensus sequence  $_{-3}\text{ACCAUGG}_{+4}$ . The ribosome may skip the first AUG it finds if the next one has the Kozak sequence. Another factor is the presence of mRNA secondary structure. If hairpin loops form downstream of an AUG, an earlier AUG may be chosen. The mRNA and the seven eIFs constitute the 48S preinitiation complex. In Step 3, the 60S ribosome is recruited, forming the 80S initiation complex. GTP is hydrolyzed, and the initiation factors are released. The following Biochemical Connections describes an important link between initiation factors and memory.

## Chain Elongation

Peptide chain elongation in eukaryotes is very similar to that of prokaryotes. The same mechanism of peptidyl transferase and ribosome translocation is seen. The structure of the eukaryotic ribosome is different in that there is no E site, only the A and P sites. There are two eukaryotic elongation factors, eEF1 and eEF2. The eEF1 consists of two subunits, eEF1A and eEF1B. The 1A subunit is the counterpart of EF-Tu in prokaryotes. The 1B subunit is the equivalent of the EF-Ts in prokaryotes. The eEF2 protein is the counterpart of the prokaryotic EF-G, which causes translocation.

Many of the differences between translation in prokaryotes and eukaryotes can be seen in the response to inhibitors of protein synthesis and to toxins. The antibiotic chloramphenicol (a trade name is Chloromycetin) binds to the A site and inhibits peptidyl transferase activity in prokaryotes, but not in eukaryotes. This property has made chloramphenicol useful in treating



■ **FIGURE 12.21** mRNA looping in eukaryotic translation initiation. Initiation factor eIF4G serves as a multipurpose adapter to engage the 7-methyl-G cap:eIF4E complex, the Pab1p:poly(A) tract, and the 40S ribosomal subunit in eukaryotic translation initiation. (Adapted with permission from Heutze, H. W., 1997. *eIF4G: A multipurpose ribosome adapter?* Science 275, 500–501. Copyright © 1997 AAAS.)

## Biochemical Connections NEUROLOGY

### Protein Synthesis Makes Memories

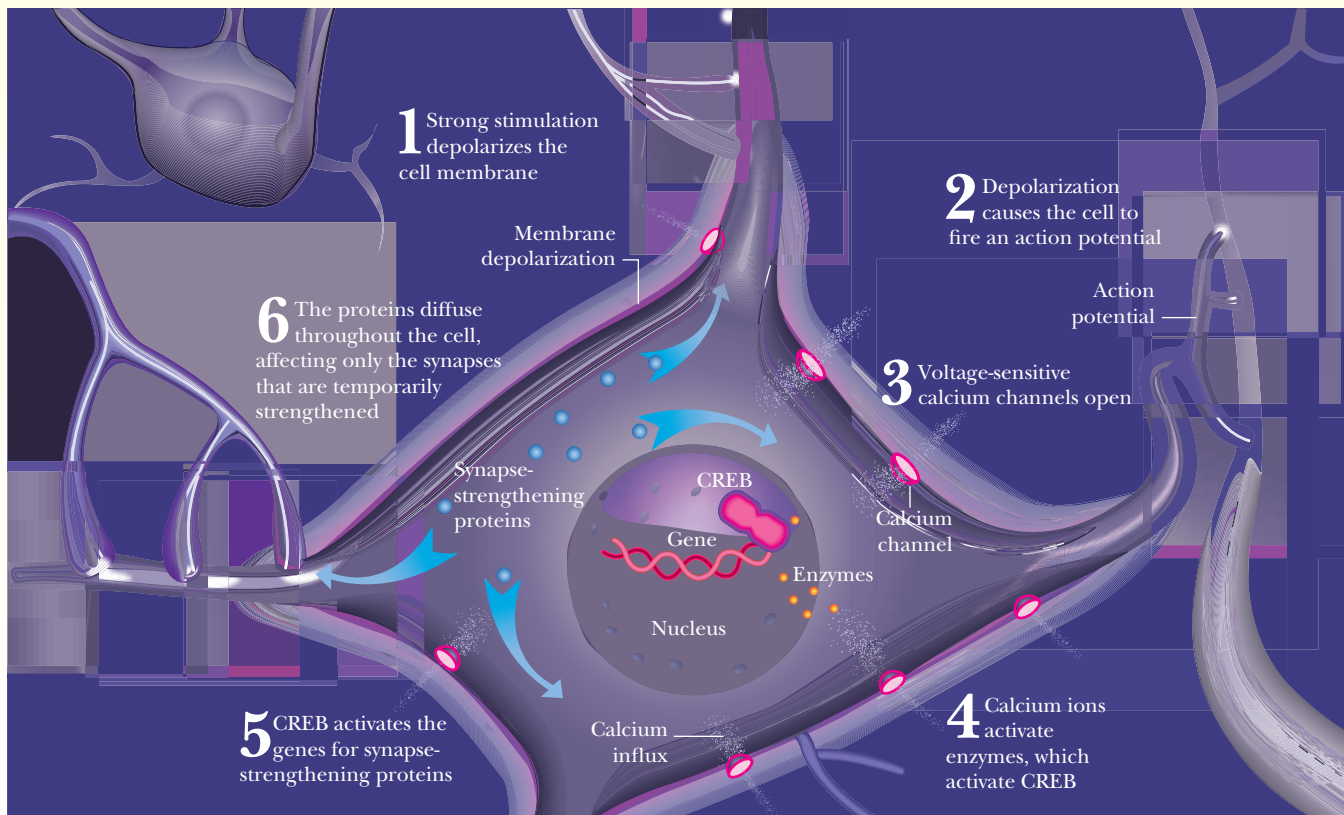
Memories are of two types—short-term and long-term. Short-term memories last from seconds to minutes, while long-term memories last days, months, or even a lifetime. Neuroscientists have long been fascinated by what makes one memory stick while another does not. When you meet someone at a party, and she tells you her name, you might forget it in seconds. However, your best friend's name is converted to long-term memory and might last your whole life. While there are tremendous individual variations in memory capability, one thing that is known for sure is that making long-term memories relies on new protein synthesis. Animals given drugs that block protein synthesis cannot form new long-term memories, yet their ability to make short-term memories is preserved. Perhaps people with good long-term memory are better at making new brain proteins.

Both long- and short-term memories come from connections between neurons, called synapses, where one neuron emits a signal from its axon that is received by the next neuron's dendrite. Memories are made when the synapse is made stronger, or “sensitized” to further signals. When the memory is long-term, this strengthening is permanent. In order for this strengthening to occur, genes in the nucleus of the neuron must be activated and proteins produced. The central paradox for neuroscientists studying memory has always been, “how does a gene activated in a neuron's nucleus ‘know’ when to strengthen the synapse permanently?” In searching for that answer, they set off to find out what proteins were involved in the process. By the mid '90s,

researchers had determined that a critical role is played by the transcription factor CREB (Chapter 11) in turning short-term memories into long-term ones. More recent research is searching for other proteins, especially ones that specifically strengthen the synapse. The entire process can be envisioned as shown in the figure below.

The process begins when strong stimulation depolarizes the cell membrane of the nerve (1). This stimulation could come from multiple firings involving the action potentials of a single synapse, or from simultaneous firings from multiple synapses. Not all impulses received cause a nerve to fire its own impulse, thereby passing on the impulse. Only ones that are strong enough will do that. This may be one reason that minor stimuli are quickly forgotten, or why we have to concentrate to remember something. When the incoming signal is strong enough, the receiving neuron fires (2). The depolarization of the neuron opens calcium channels (3). Calcium enters the neuron and activates key signaling enzymes, which activate CREB (4). CREB activates the genes for the putative synapse-strengthening proteins (5). In the last step, these proteins diffuse throughout the cell but only affect those synapses that were temporarily strengthened.

Memory is a complicated business, and we have just scratched the surface of understanding it. It is clear that a combination of external factors (the stimulus) and internal factors (the chemistry) affect how we remember. Key among the internal factors are several processes involving protein synthesis.



■ **Protein Synthesis makes memories.** (From *Making Memories Stick* by R. Douglas Fields, *Scientific American*, February 2005, page 80.)



bacterial infections. In eukaryotes, diphtheria toxin is a protein that interferes with protein synthesis by decreasing the activity of the eukaryotic elongation factor eEF2.

### Chain Termination

As in prokaryotic termination, the ribosome encounters a stop codon, either UAG, UAA, or UGA, and these are not recognized by a tRNA molecule. In prokaryotes, three different release factors—RF1, RF2, and RF3—were used, with two of them alternating, depending on which stop codon was found. In eukaryotes, only one release factor binds to all three stop codons and catalyzes the hydrolysis of the bond between the C-terminal amino acid and the tRNA.

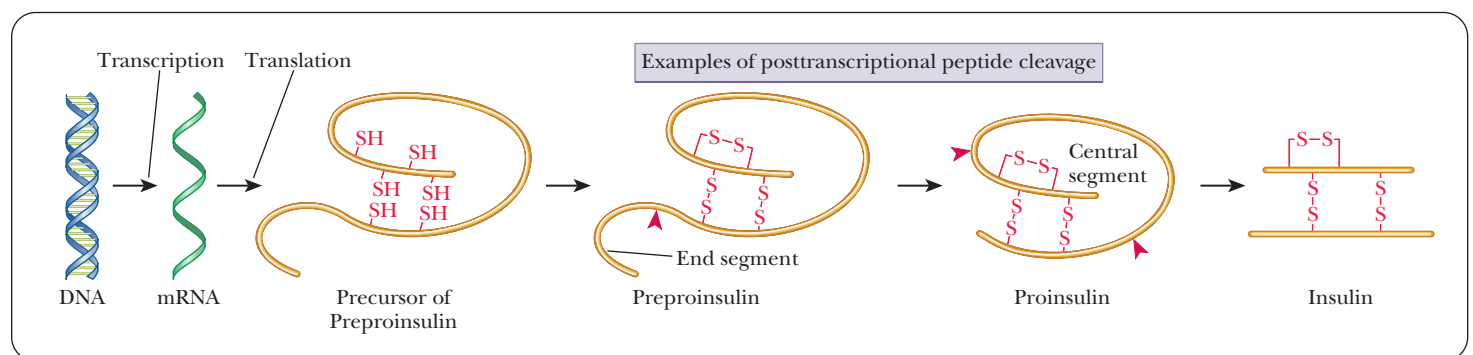
There is a special tRNA called a **suppressor tRNA**, which allows translation to continue through a stop codon (see the Biochemical Connections box on page 339 for a description of a unique tRNA that inserts a selenocysteine residue because the anticodon of the corresponding tRNA binds to the stop codon). Suppressor tRNAs tend to be found in cells in which a mutation has introduced a stop codon.

### Coupled Transcription and Translation in Eukaryotes?

Until recently, the dogma of eukaryotic translation was that it was physically separated from transcription. Transcription occurred in the nucleus, and mRNA was then exported to the cytosol for translation. Although this system is accepted as the normal process, recent evidence has shown that the nucleus has all of the components (mRNA, ribosomes, protein factors) necessary for translation. In addition, evidence shows that, in isolated test systems, proteins are translated in the nucleus. Some researchers have suggested that 10%–15% of the cell's protein synthesis occurs in the nucleus.

## 12.6 Posttranslational Modification of Proteins

Newly synthesized polypeptides are frequently processed before they reach the form in which they have biological activity. We have already mentioned that, in prokaryotes, *N*-formylmethionine is cleaved off. Specific bonds in precursors can be hydrolyzed, as in the cleavage of preproinsulin to proinsulin and of proinsulin to insulin (Figure 12.22). Proteins destined for export to specific parts of the cell or from the cell have leader sequences at their N-terminal ends. These leader sequences, which direct the proteins to their proper destination,



■ **FIGURE 12.22** Some examples of posttranslational modification of proteins. After a precursor of preproinsulin is formed by the transcription–translation process, it is transformed into preproinsulin by formation of three disulfide bonds. Specific cleavage that removes an end segment converts preproinsulin to proinsulin. Finally, two further specific cleavages remove a central segment, with insulin as the end result.



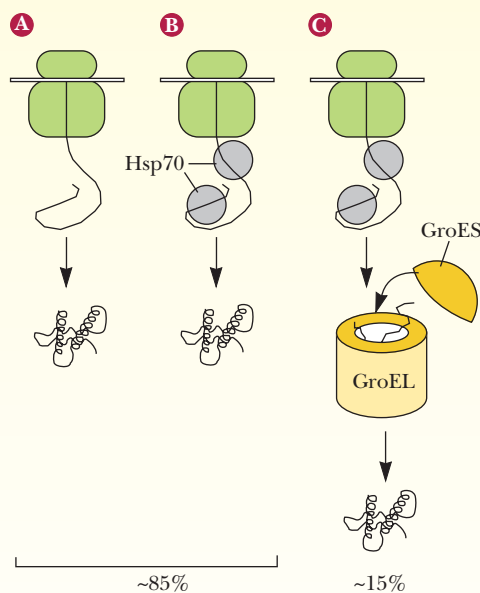
# Biochemical Connections BIOPHYSICAL CHEMISTRY

## Chaperones: Preventing Unsuitable Associations

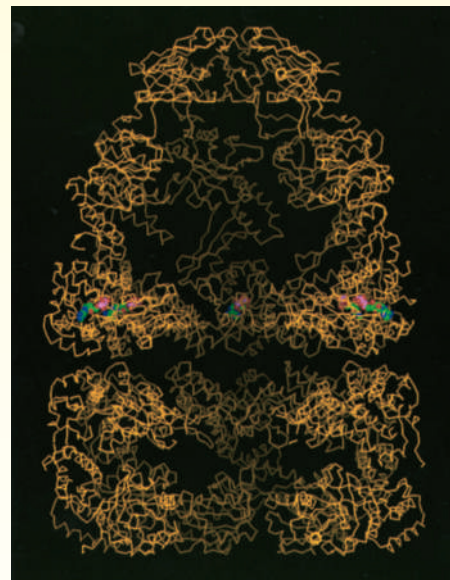
It is sometimes said that a chaperone's task is to prevent unsuitable associations. The class of proteins known as molecular chaperones operate in this way by preventing aggregation of newly formed proteins until they fold into their active forms. The information necessary for protein folding is present in the amino acid sequence, and many proteins fold correctly without any outside help, as shown in part (a) of the figure. However, some proteins may form aggregates with other proteins, or may fold with incorrect secondary and tertiary structures, unless they interact first with a chaperone. Well-known examples include heat-shock proteins, which are produced by cells as a result of heat stress. The prime examples of this are the *Hsp70* class of proteins, named after the 70-kDa heat-shock protein that occurs in mammalian cytosol, shown in part (b) of the figure. The *Hsp70* protein binds to the nascent polypeptide and prevents it from interacting with other proteins or from folding into an unproductive form. Completion of correct folding requires release from the chaperone and is driven by ATP hydrolysis. All proteins in this class, which were first studied as a response to heat stress in cells of all types,

have highly conserved primary structures, in both prokaryotes and eukaryotes.

About 85% of proteins fold as shown in parts (a) and (b) of the figure. Another group, the **chaperonins** (also called the *Hsp60* proteins from their 60-kDa molecular weight) are known to be involved in the folding of the other 15% of proteins. A large multisubunit protein forms a cage of 60-kDa subunits around the nascent protein to protect it during the folding process, shown in part (c) of the figure. **GroEL** and **GroES** are the best-characterized chaperonins from *E. coli*. GroEL is formed by two stacked seven-membered rings of 60-kDa subunits with a central cavity, as shown in the figure below. Protein folding occurs in the central cavity and is dependent on ATP hydrolysis. GroES is a single seven-membered ring of 10-kDa subunits that sits on top of GroEL. During protein folding, the polypeptide chain goes through cycles of binding and unbinding to the surface of the central cavity. In some cases, more than 100 ATP molecules must be hydrolyzed before protein folding is complete.



■ **Protein folding pathways.** (a) Chaperone-independent folding. (b) Folding facilitated by a chaperone (gray)—in this case, the *Hsp70* protein. About 85% of proteins fold by one of the two mechanisms shown in (a) and (b). (c) Folding facilitated by chaperonins—in this case, GroEL and GroES. (Adapted from Netzer, W. J., and Hartl, F. U., 1998. Protein folding in the cytosol: Chaperonin-dependent and -independent mechanisms. *Trends in Biochemical Sciences* 23, 68–73, Figure 2.)



■ **Structure of the GroEL–GroES complex, emphasizing the central cavity.** The protein is shown in orange. Bound ADP molecules are shown in green, and their associated magnesium ions are shown in red. (Adapted by permission from Figure 1 in Xu, Z., Horwich, A. L., and Sigler, P. B., 1997, *Nature* 388, 741–750. Molecular graphics courtesy of Paul B. Sigler, Yale University.)

are recognized and removed by specific proteases associated with the *endoplasmic reticulum*. The finished protein then enters the *Golgi apparatus*, which directs it to its final destination.

In addition to the processing of proteins by breaking bonds, other substances can be linked to the newly formed polypeptide. Various cofactors, such as heme groups, are added, and disulfide bonds are formed (Figure 12.22). Some amino acid residues are also covalently modified, as in the conversion of proline to hydroxyproline. Other covalent modifications can take place, an example being the addition of carbohydrates or lipids to yield an active final form of the protein in question. Proteins can also be methylated, phosphorylated, and ubiquitinated (Section 12.7).

### Once modified, do proteins always have the correct three-dimensional structure?

A highly important question concerns the proper folding of the newly synthesized protein. In principle, the primary structure of the protein conveys enough information to specify its three-dimensional structure. In the cell, the complexity of the process and the number of possible conformations make it less likely that a protein would spontaneously fold into the correct conformation. The Biochemical Connections box on page 346 describes the processes involved in protein folding in vivo.

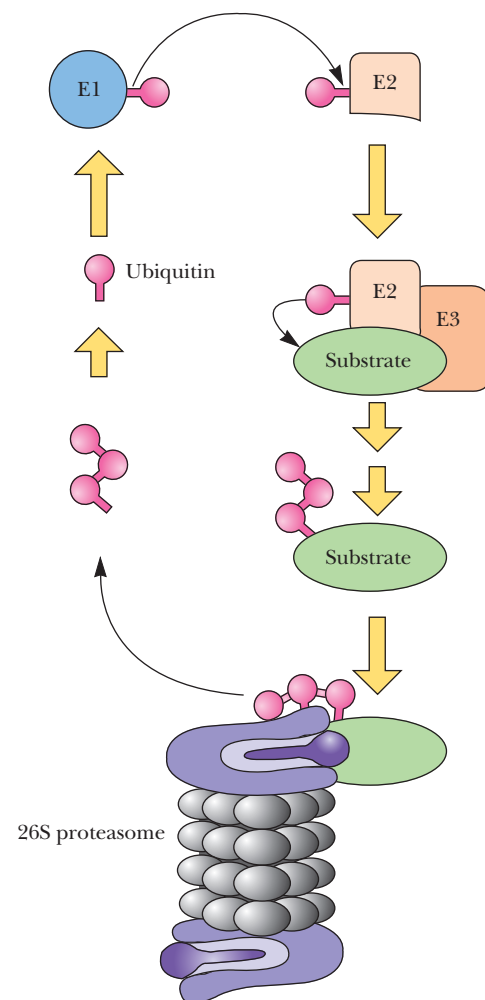
## 12.7 Protein Degradation

One of the most often overlooked controls of gene expression occurs at the level of the degradation of proteins. Proteins are in a dynamic state in which they are turned over often. Athletes are painfully aware of this because it means that they must work very hard to get in shape, but then they get out of shape very quickly. Some classes of proteins experience a 50% turnover every three days. In addition, abnormal proteins that were formed from errors in either transcription or translation are degraded quickly. It is believed that a single break in the peptide backbone of a protein is enough to trigger the rapid degradation of the pieces, because breakdown products from natural proteins are rarely seen in vivo.

### How does the cell know which proteins to degrade?

If protein degradation is so quick, clearly it is a process that must be heavily controlled to avoid destruction of the wrong polypeptides. The degradation pathways are restricted to degradative subcellular organelles, such as lysosomes, or to macromolecular structures called **proteasomes**. Proteins are directed to lysosomes by specific signal sequences, often added in a posttranslational modification step. Once in the lysosome, the destruction is nonspecific. Proteasomes are found in both prokaryotes and eukaryotes, and specific pathways exist to target a protein so that it complexes with a proteasome and is degraded.

In eukaryotes, the most common mechanism for targeting protein for destruction in a proteasome is by **ubiquitylation**. Ubiquitin is a small polypeptide (76 amino acids) that is highly conserved in eukaryotes. There is a high degree of homology between the sequences in species as widespread as yeast and humans. When ubiquitin is linked to a protein, it condemns that protein to destruction in a proteasome. Figure 12.23 shows the mechanism of ubiquitylation. Three enzymes are involved—*ubiquitin-activating enzyme* (E1), *ubiquitin-carrier protein* (E2), and *ubiquitin-protein ligase* (E3). The ligase transfers the ubiquitin to free amino groups on the targeted protein, either the N-terminus or lysine side chains. Proteins must have a free  $\alpha$ -amino group to



■ **FIGURE 12.23** Diagram of the ubiquitin–proteasome degradation pathway. Pink “lollipop” structures symbolize ubiquitin molecules. (Adapted with permission from Hilt, W., and Wolf, D. H., 1996. *Proteasomes: Destruction as a program*. Trends in Biochemical Sciences 21, 96–102, Figure 1.)

## Biochemical Connections GENETICS

### Silent Mutations Are Not Always Silent

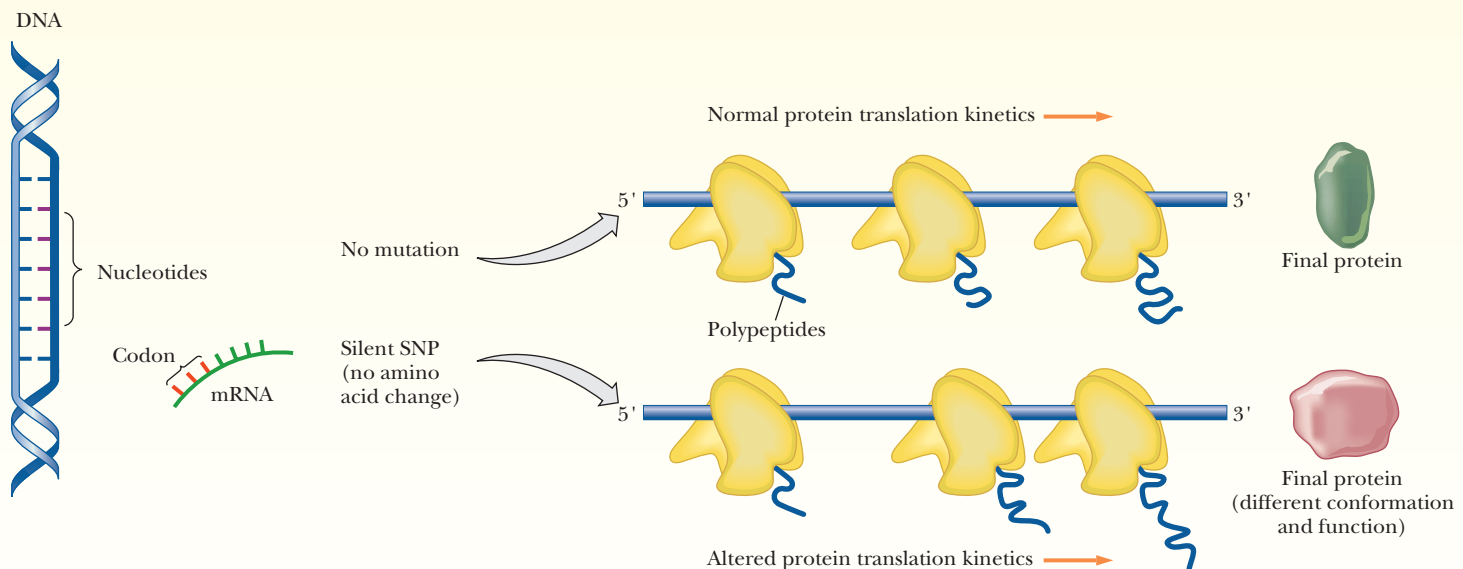
A silent mutation is a mutation that changes DNA but not the amino acid incorporated. For example, if the DNA coding strand has a UUC, it codes for phenylalanine. If a mutation in the DNA changes the sequence to UUU, then the DNA has undergone a silent mutation because both UUU and UUC code for the same amino acid. At least that is what scientists believed for decades. Recent evidence, however, has shown that this is not always true. Researchers at the National Cancer Institute were studying a gene called *MDR1*, which is named for its association with multiple drug resistance in tumor cells. They had sequences of this gene and knew that there were some common silent mutations. Interestingly, they discovered that there was a phenotypic response to silent mutations of this gene that influenced patients' response to certain drugs. This was striking, as a silent mutation should have no effect on the final product.

Apparently, not all codons are translated equally. Different codons may require alternate versions of the tRNA for a particular amino acid. Even though the amino acid incorporated is the same, the pace with which the ribosome is able to incorporate the amino acid differs depending on which codon it is. This is a situation reminiscent of transcription attenuation that we saw in Chapter 11. As shown in the figure, translation kinetics can affect the form of the final protein. If the wild-type codon is used, then translation proceeds normally and produces the

normal conformation of the protein. However, if a silent mutation changes the pace of the movement of the ribosome, then because of folding differences, an abnormal protein conformation is created.

In the last 10 years, many new examples of phenotypic effects due to silent mutations have been discovered. Besides the protein-folding differences shown above, there are other processes based on synonymous codon differences. One of the most frequent is based on differences in processing of the mRNA after transcription (Section 11.8). Research has shown that the mRNA exons not only contain the code that should lead to production of the correct amino acids, but they also contain the information necessary for correct removal of introns. There are specific sequences, called **exonic splicing enhancers (ESE)**, that tell the splicing machinery where to splice out introns. The codons GGA and GGG both encode glycine and both can occur in ESEs. However, GGA is a much more potent splicing enhancer. As shown in the figure below, a silent mutation change in an ESE could lead to incorrect removal of introns and an entire exon being left out of the final mRNA.

Another process involving silent mutations is caused by secondary structures of mRNA. While textbooks often depict mRNA as a linear molecule for simplicity, mRNA can also fold up into many different structures. Changing any nucleotide can



■ **Translation kinetics and protein folding.** Unaffected translation kinetics results in a correctly folded protein. Abnormal kinetics, caused by the ribosome moving faster or slower through certain mRNA regions, can produce a different final protein conformation. Abnormal kinetics may arise from a silent single nucleotide polymorphism (SNP) in a gene that creates a codon synonymous to the wild type codon. However, this synonymous codon substitution may lead to different kinetics of mRNA translation, thus yielding a protein with a different final structure and function. (Reprinted by permission of Science magazine, from SNPs, *Silent but not Invisible* by Anton A. Komar, Science **315**, 466–467 [2007].)

# Biochemical Connections (CONTINUED)

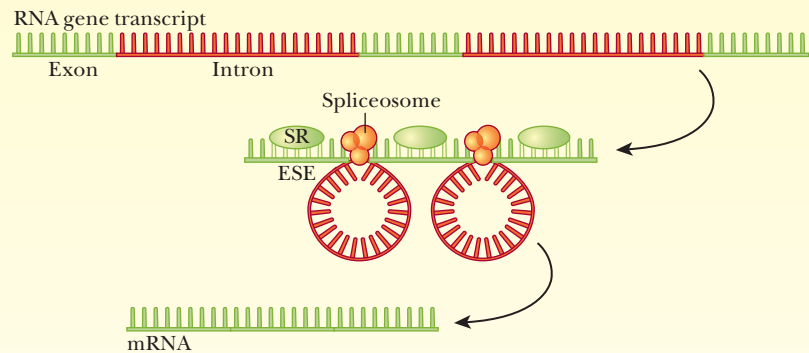
potentially change the way the mRNA folds, and such changes can affect the speed with which the mRNA is translated, if it can be translated at all. Researchers studying a gene involved in pain tolerance discovered that silent mutations in an intron for the gene *COMT* (catechol-O-methyltransferase) are responsible for different levels of an important enzyme that affects how we perceive pain. They showed that the different mutations led to different

folding patterns of the mRNA, which translated to different levels of the enzyme in the cell (see the figure below).

Researchers continue to study this phenomenon because it is now known that more than 50 human diseases are caused by silent mutations, including Marfan syndrome, androgen-insensitivity syndrome, cholesteryl ester storage disease, McArdle disease, and phenylketonurea.

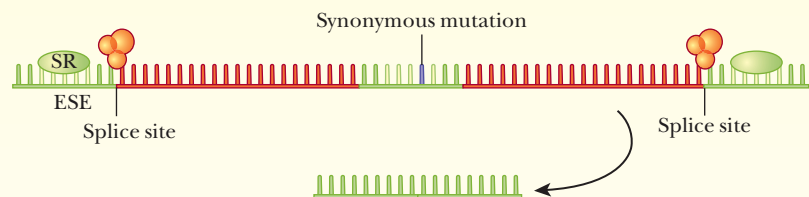
## NORMAL RNA SPLICING

The raw RNA transcript of a gene contains exons, which encode amino acids, and long noncoding intron segments that must be edited out of the final mRNA. Within each exon, short nucleotide sequences act as exonic splicing enhancers (ESE) that flag the boundaries of the exon to cellular editing machinery. The binding of splicing regulatory (SR) proteins to enhancer sites directs “spliceosome” proteins to both ends of an intron, which they excise from the transcript, before joining the exon ends together.



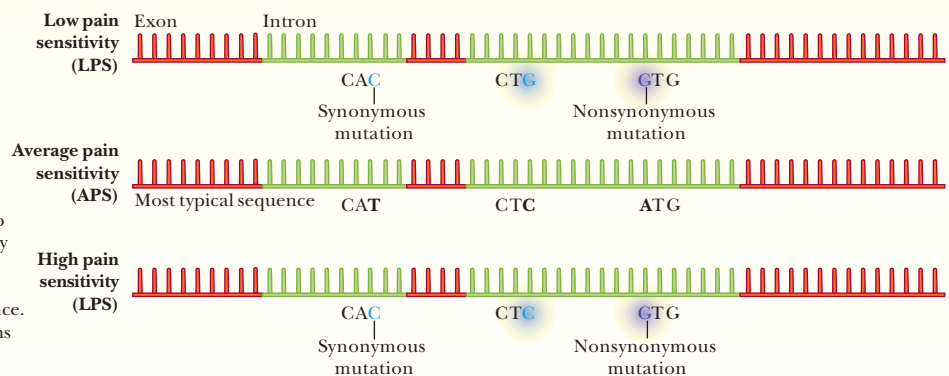
## EXON SKIPPING

Single-nucleotide synonymous changes in an exon can render splicing enhancer sequences invisible to the splicing machinery, causing an entire exon to be left out of the final mRNA.

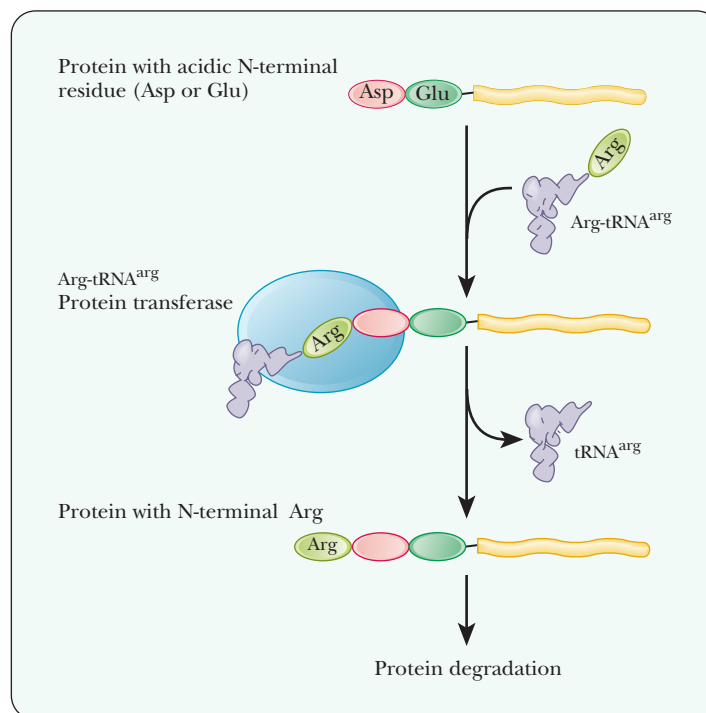


## COMT GENE VARIANTS

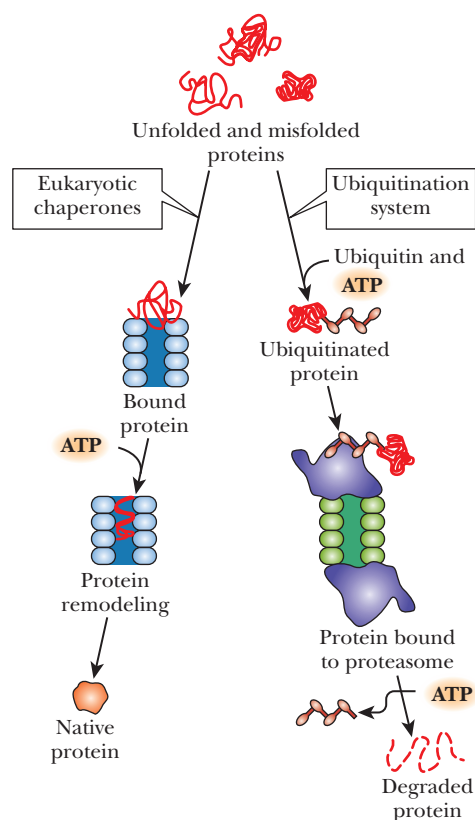
Three common versions of the gene for catechol-O-methyltransferase (*COMT*) are associated with low (LPS), average (APS) or high pain sensitivity (HPS). Differences between the most typical sequence (APS) and the other variants occur at three sites in the gene's exons; however, only one of the changes (purple glow) alters the encoded amino acid. That mutation was once thought to account for differences in pain sensitivity among individuals, but both LPS and HPS subjects have the same G in that position, so it cannot be the sole influence. In fact, one of the synonymous mutations (blue glow) was found to account for 7 percent of the pain sensitivity variation.



■ Different sequences in the introns in a gene can lead to different phenotypes, as shown for a gene for pain tolerance. (Reprinted with permission. Copyright © 2009 Scientific American, a division of Nature America, Inc. All rights reserved.)



■ **FIGURE 12.24** Degradation of proteins with acidic N-termini. Proteins with acidic N-termini show a tRNA requirement for degradation. Arginyl-tRNA<sup>arg</sup>;protein transferase catalyzes the transfer of Arg to the free α-NH<sub>2</sub> of proteins with Asp or Glu N-terminal residues. Arg-tRNA<sup>arg</sup>;protein transferase serves as part of the protein degradation recognition system.



■ **FIGURE 12.25** Misfolded proteins can be repaired and folded correctly by eukaryotic chaperones or they can be targeted for destruction by ubiquitination. (From *Biochemistry*, by R. H. Garrett and C. M. Grisham, Brooks/Cole Cengage Learning, 2010, p. 1004.)

be susceptible, so proteins that are modified at the N-terminus—with an acetyl group, for instance—are protected from ubiquitin-mediated degradation. The nature of the N-terminal amino acid also influences its susceptibility to ubiquitylation. Proteins with Met, Ser, Ala, Thr, Val, Gly, or Cys at the N-terminus are resistant. Those with Arg, Lys, His, Phe, Tyr, Trp, Leu, Asn, Gln, Asp, or Glu at the N-terminus have very short half-lives, between 2 and 30 minutes. Proteins with an acidic residue at the N-terminus have a requirement for tRNA as part of their destruction pathway. The tRNA for arginine, Arg-tRNA<sup>arg</sup>, is used to transfer arginine to the N-terminus, making the protein much more susceptible to the ubiquitin ligase (Figure 12.24). If we put together the chaperone and protein degradation process, we have a good picture of how the cell monitors quality control of proteins. Cells start with proteins that may be unfolded or misfolded. If they are unfolded, they can go through the chaperone process for correct folding (see Figure 12.25). If they are misfolded, then they go through the ubiquitination system and are degraded. The following Biochemical Connections box gives an interesting example of how transcription regulation and protein degradation work together to control the process of acclimation to high altitude.



# Biochemical Connections PHYSIOLOGY

## How Do We Adapt to High Altitude?

Those of us who live at low altitudes are quite aware of the sensations of oxygen deprivation and the associated physiological changes that occur with prolonged exposure to high altitudes. Physiologists have studied this phenomenon for many years, and biochemists have tried to find the mechanism that explains how cells sense the oxygen partial pressure and make the adaptive changes. Two of the major changes are the increase in red blood cells, which is stimulated by the hormone **erythropoietin (EPO)**, and **angiogenesis**, the stimulation of formation of new capillaries, which is stimulated by **vascular endothelial growth factor (VEGF)**. Researchers have learned a great deal about how cells respond to low oxygen pressure, or **hypoxia**, and the results have many applications, including the production of drugs to treat inflammation, heart disease, and cancer.

A family of transcription factors called *hypoxia inducible factors (HIFs)* is the key to these processes (see the figure). Heterodimers composed of HIF $\alpha$  and HIF $\beta$  subunits bind to DNA and up-regulate a variety of genes when the partial pressure of oxygen in the blood is low. Oxygen can be low for many reasons, such as a person being at high altitude or a tissue having to work unusually hard. During a heart attack or stroke, the oxygen partial pressure can drop, and these transcription factors can help reduce the damage.

The genes that are controlled by HIF are responsible for production of EPO and angiogenesis, as well as production of glycolytic enzymes that can provide energy for the cells when aerobic metabolism is compromised. In addition, many types of cancers are found associated with elevated levels of HIF. This may be involved in the tumor's ability to grow, which requires a large oxygen supply. As it turns out, if cell growth is uncontrolled, the dimerization of the two HIF subunits is unchecked, which would lead to constant expression of the adaptive genes. When this happens, overgrowth of endothelial cells occurs, leading to tumors. Any given cell maintains relatively constant levels of the HIF $\beta$  subunit, but the level of the HIF $\alpha$  subunit is regulated. The system is mainly controlled by the degradation of the HIF $\alpha$  subunit (see the right side of the figure). Proline 564 in the HIF $\alpha$  subunit can be hydroxylated by an enzyme called *proline hydroxylase (PH)*. After it is hydroxylated, it binds to a protein called von Hippel–Lindau protein (pVHL), which was first discovered to be a tumor suppressor (see Section 14.8 for more on tumor suppressors). After the pVHL is bound, it stimulates the organization of a complex with *ubiquitin ligase (UL)*, which ubiquitinates the HIF $\alpha$  subunit. Ubiquitin is a 76-residue polypeptide that is very abundant and conserved in eukaryotes. When ubiquitin is put onto a protein, it targets the protein for transport to a proteasome, where the protein is degraded. Searching for how this pathway is related to the body's ability to sense the oxygen partial pressure is currently an area of active research. Evidence exists that the proline hydroxylase has a requirement for

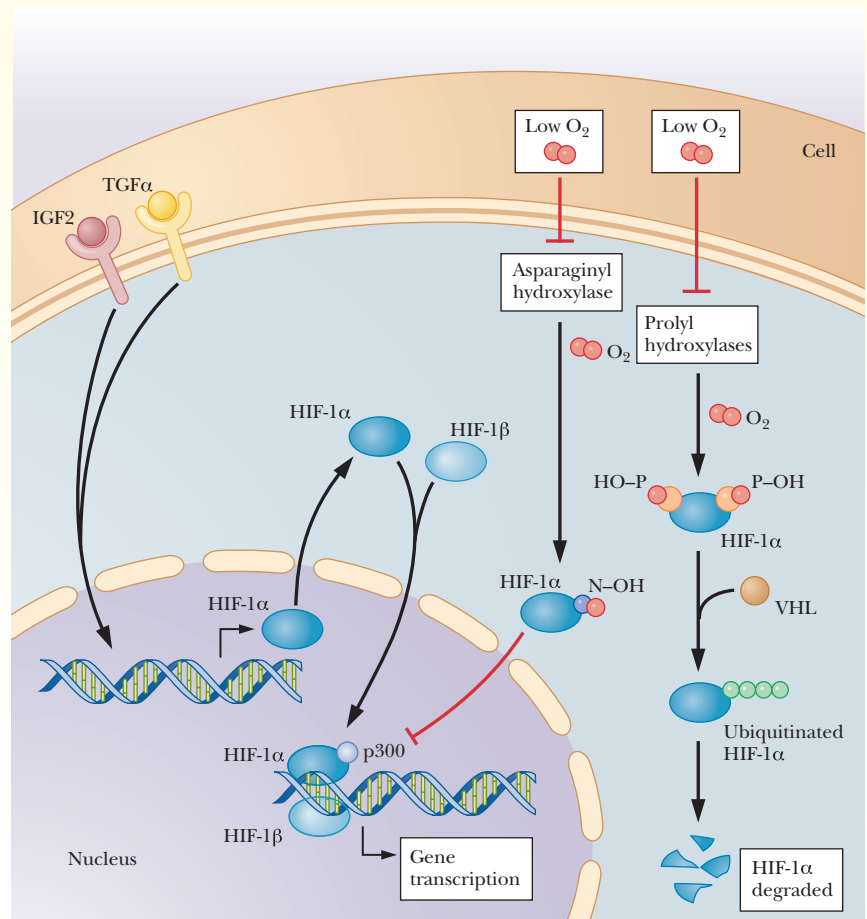
iron and oxygen. It is possible that, in the absence of sufficient oxygen, the HIF proline hydroxylase cannot function. Thus, the HIF $\alpha$  subunit is not targeted for destruction and is available to bind to the HIF $\beta$  subunit, forming the active dimer and stimulating the adaptive effects of reduced oxygen pressure.

A second control point was recently discovered to be mediated by another hydroxylase. This time the target is an asparagine residue on HIF $\alpha$ . When the HIF $\alpha$  subunit is hydroxylated, it cannot bind to the transcription mediator, (see Chapter 11), and therefore cannot induce transcription. Thus, when oxygen is low, two different hydroxylations proceed at reduced levels.

Reduction of the asparagine's hydroxylase reaction allows the HIF $\alpha$  to do its job, and reduction of the prolyl hydroxylase reaction stops degradation of the HIF $\alpha$  by the ubiquitin-linked pathway. Many researchers believe that these two reactions are the mechanism for oxygen sensing, but others believe they are just incidental to the true mechanism, which has yet to be discovered.

HIF is also controlled in a positive way by **insulin growth factor 2 (IGF2)** and by **transforming growth factor  $\alpha$  (TGF $\alpha$ )**. These are common growth factors that are active in several growth and differentiation pathways, and they are not related to oxygen availability.

Researchers are looking at control of HIF as a potential therapy against cancer because so many cancers are found to be associated with high HIF levels. Because tumors need oxygen to grow, stopping the tumor's activation of HIF-mediated transcription could effectively choke out the tumor before it progresses.



■ **The control of the level and function of hypoxia inducible factor  $\alpha$  by the level of oxygen.** (Adapted with permission from Marx, J., 2004. How cells endure low oxygen. *Science* 303, 1454–1456. Illustration by Carin Cain. Copyright © 2004 AAAS.)

## SUMMARY

**How did scientists determine the genetic code?** A variety of techniques were used to determine the genetic code. The initial experiments used synthetic mRNA sequences with only a single nucleotide, such as AAAAAAAAA. These were then translated to see what protein homopolymer was produced. In this way the translation for AAA, UUU, GGG, and CCC was possible to determine. The most comprehensive experiment was the filter-binding experiment of Nirenberg, which used specific trinucleotides bound to a filter to see which tRNA molecules would bind.

**If there are 64 codons, how can there be fewer than 64 tRNA molecules?** There are only about half as many tRNA molecules as there are codons for amino acids. This occurs because of wobble in the third base (5' end) of the anticodon on the tRNA. The wobble base can break Watson–Crick base-pairing rules under certain circumstances. For example, if the wobble base is a U, it can bind to an A or a G of the mRNA codon, and thus fewer tRNA molecules are needed.

**What is the “second genetic code”?** The second genetic code refers to the specificity with which the aminoacyl-tRNA synthetases put the correct amino acid onto the correct tRNA. There is only minimal proofreading past that point, so the correct loading of amino acids is critical. If the wrong amino acid is loaded on the tRNA, it is quickly hydrolyzed by the synthetase to avoid such errors.

**How does the ribosome know where to start translating?** In prokaryotic translation, the correct AUG start codon is identified by its proximity to a consensus sequence called the Shine–Dalgarno sequence. This sequence is complementary to a sequence on the small subunit of the prokaryotic ribosome. The ribosome is initially positioned on the Shine–Dalgarno sequence, which aligns it for correct translation initiation at the start codon.

**Why is EF-Tu so important in *E. coli*?** The elongation factors EF-Tu and EF-Ts add a level of complexity to the elongation

of proteins. This complexity seems at first to be burdensome, but is actually very important as it leads to another level of proofreading. EF-Tu delivers the aminoacyl-tRNA to the A site of the ribosome only when the codon and anticodon match. In addition, if the tRNA is not matched correctly to its amino acid, then EF-Tu is not efficient at delivering the tRNA to the ribosome.

**How is translation different in eukaryotes?** In eukaryotes, the process differs in several details. Eukaryotic mRNA undergoes a lot of processing that is not observed in prokaryotes. Eukaryotic mRNA has a 5' cap and a 3' poly A tail, both of which are involved in formation of the initiation complex. There is no Shine–Dalgarno sequence, but there is a Kozak sequence surrounding the correct AUG start codon. The number of initiation factors and elongation factors is much higher in eukaryotes than in prokaryotes. Stop codons in eukaryotes are sometimes recognized by suppressor tRNAs, which allow insertion of nonstandard amino acids, such as selenocysteine.

**Once modified, do proteins always have the correct three-dimensional structure?** In theory, the primary structure of the protein determines its three-dimensional structure. However, in reality proteins often need the help of a chaperone to arrive at the correct structure. This is due to possible interactions with other proteins before the nascent protein chain is complete and also the possibility that a protein will begin to fold incorrectly in its early stages of translation before it is complete.

**How does the cell know which proteins to degrade?** Proteins are degraded in subcellular organelles, such as lysosomes, or in macromolecular structures called proteasomes. Many proteins are targeted for destruction by being bound to a protein called ubiquitin. The nature of the amino acid sequence at the N-terminus is often very important to control the timing of the destruction of a protein. Damaged proteins are degraded very quickly.

## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

### 12.1 Translating the Genetic Message

1. **Recall** Prepare a flow chart showing the stages of protein synthesis.

### 12.2 The Genetic Code

2. **Recall** A genetic code in which two bases encode a single amino acid is not adequate for protein synthesis. Give a reason why.
3. **Recall** Define *degenerate code*.
4. **Recall** How can the binding assay technique be used to assign coding triplets to the corresponding amino acids?

5. **Recall** Which nucleotides break the rules of Watson–Crick base pairing when they are found at the wobble position of the anticodon? Which ones do not?
6. **Recall** Describe the role of stop codons in the termination of protein synthesis.
7. **Reflect and Apply** Consider a three-base sequence in the template of DNA: 5' ... 123 ... 3', in which 1, 2, and 3 refer to the relative positions of deoxyribonucleotides. Comment on the probable effect on the resulting protein if the following point mutations (one-base substitutions) occurred.

- (a) changing one purine for another in position 1
  - (b) changing one pyrimidine for another in position 2
  - (c) changing a purine to a pyrimidine in position 2
  - (d) changing one purine for another in position 3
8. **Reflect and Apply** It is possible for the codons for a single amino acid to have the first two bases in common and to differ in the third base. Why is this experimental observation consistent with the concept of wobble?
  9. **Reflect and Apply** The nucleoside inosine frequently occurs as the third base in codons. What role does inosine play in wobble base pairing?
  10. **Reflect and Apply** Is it reasonable that codons for the same amino acid have one or two nucleotides in common? Why or why not?
  11. **Reflect and Apply** How would protein synthesis be affected if a single codon could specify the incorporation of more than one amino acid (an ambiguous code)?
  12. **Reflect and Apply** Comment on the evolutionary implications of the differences in the genetic code observed in mitochondria.

### 12.3 Amino Acid Activation

13. **Recall** What is the role of ATP in amino acid activation?
14. **Recall** Outline the proofreading processes in amino acid activation.
15. **Recall** What ensures fidelity in protein synthesis? How does this compare with the fidelity of replication and transcription?
16. **Recall** Can the same enzyme esterify more than one amino acid to its corresponding tRNA?
17. **Reflect and Apply** A friend tells you that she is starting a research project on aminoacyl esters. She asks you to describe the biological role of this class of compounds. What do you tell her?
18. **Reflect and Apply** Suggest a reason why the proofreading step in protein synthesis takes place at the level of amino acid activation rather than that of codon–anticodon recognition.
19. **Reflect and Apply** Is amino acid activation energetically favored? Why or why not?

### 12.4 Prokaryotic Translation

20. **Recall** Identify the following by describing their functions: EF-G, EF-Tu, EF-Ts, EF-P, and peptidyl transferase.
21. **Recall** What are the components of the initiation complex in protein synthesis? How do they interact with one another?
22. **Recall** What is the role of the 50S ribosomal subunit in prokaryotic protein synthesis?
23. **Recall** What are the A site and the P site? How are their roles in protein synthesis similar? How do they differ? What is the E site?
24. **Recall** How does puromycin function as an inhibitor of protein synthesis?
25. **Recall** Describe the role of the stop signals in protein synthesis.
26. **Recall** Does mRNA bind to one or to both ribosomal subunits in the course of protein synthesis?
27. **Recall** What is the Shine–Dalgarno sequence? What role does it play in protein synthesis?
28. **Reflect and Apply** You are studying with a friend who says that the hydrogen-bonded portions of tRNA play no important role in its function. What is your reply?
29. **Reflect and Apply** *E. coli* has two tRNAs for methionine. What is the basis for the distinction between the two?
30. **Reflect and Apply** In prokaryotic protein synthesis, formylmethionine (fmet) is the first amino acid incorporated, whereas (normal) methionine is incorporated in eukaryotes. The same codon (AUG) serves both. What prevents methionine from being inserted into the beginning and formylmethionine in the interior?
31. **Reflect and Apply** Describe the recognition process by which the tRNA for *N*-formylmethionine interacts with the portion of mRNA that specifies the start of transcription.

32. **Reflect and Apply** The fidelity of protein synthesis is assured twice during protein synthesis. How and when?
33. **Reflect and Apply**
  - (a) How many activation cycles are needed for a protein with 150 amino acids?
  - (b) How many initiation cycles are needed for a protein with 150 amino acids?
  - (c) How many elongation cycles are needed for a protein with 150 amino acids?
  - (d) How many termination cycles are needed for a protein with 150 amino acids?
34. **Reflect and Apply** What is the energy cost per amino acid in prokaryotic protein synthesis? Relate this to low entropy.
35. **Reflect and Apply** Would it be possible to calculate the cost of protein synthesis, including the cost of making mRNA and DNA?
36. **Reflect and Apply** Suggest a possible conclusion from the fact that peptidyl transferase is one of the most conserved sequences in all of biology.
37. **Reflect and Apply** In the early days of research on protein synthesis, some scientists observed that their most highly purified ribosome preparations, containing almost exclusively single ribosomes, were less active than preparations that were less highly purified. Suggest an explanation for this observation.
38. **Reflect and Apply** Suggest a scenario for the origin and development of peptidyl transferase as an integral part of the ribosome.
39. **Reflect and Apply** Would you expect electron microscopy to give detailed information about ribosomal structure? *Hint:* Look at Figure 12.18.
40. **Reflect and Apply** How does it improve the efficiency of protein synthesis to have several binding sites for tRNA close to each other on the ribosome?
41. **Reflect and Apply** A virus does not contain ribosomes. How does it manage to ensure the synthesis of its proteins?

### 12.5 Eukaryotic Translation

42. **Recall** What conserved sequences are important to codon recognition in prokaryotic release factors?
43. **Recall** What sequence in the prokaryotic release factors is important to the hydrolysis of the peptide bond?
44. **Biochemical Connections** Why is selenocysteine called the 21st amino acid when there are many more amino acids found than the 20 basic ones coded for by the genetic code?
45. **Biochemical Connections** What is unique about selenocysteine?
46. **Recall** What are two major similarities between protein synthesis in bacteria and in eukaryotes? What are two major differences?
47. **Reflect and Apply** Why do amino acids other than methionine occur in the N-terminal position of proteins from eukaryotes?
48. **Reflect and Apply** Would puromycin be useful for the treatment of a virus infection? Why or why not? Would chloramphenicol be useful?
49. **Reflect and Apply** Protein synthesis takes place much more slowly in eukaryotes than in prokaryotes. Suggest a reason why this is so.
50. **Reflect and Apply** Why is it advantageous to have a mechanism to override the effect of stop codons in protein synthesis?

### 12.6 Posttranslational Modification of Proteins

51. **Biochemical Connections** What process distinguishes a short-term memory from a long-term one?
52. **Biochemical Connections** What transcription factor in the nucleus of a neuron is known to play a role in creating long-term memories?
53. **Biochemical Connections** What is one experiment done to show that protein synthesis is required to make long-term memories?
54. **Biochemical Connections** How does the strength of an incoming nerve impulse affect memory?

55. **Reflect and Apply** The amino acid hydroxyproline is found in collagen. There is no codon for hydroxyproline. Explain the occurrence of this amino acid in a common protein.

### 12.7 Protein Degradation

56. **Recall** What role does ubiquitin play in the degradation of proteins?

57. **Reflect and Apply** Consider protein degradation in the absence of ubiquitinylation. Is the process likely to be more or less efficient?

58. **Reflect and Apply** Is it reasonable to expect that protein degradation can take place at any location in a cell?

59. **Biochemical Connections** What are exonic splicing enhancers and how are they important?

60. **Biochemical Connection** What is a silent mutation? Why is the name “silent mutation” a bit of a misnomer?

61. **Biochemical Connections** How can a silent mutation lead to “exon skipping”?

62. **Biochemical Connections** What process in mRNA folding leads to sensitivity to pain due to an effect on the COMT gene?

63. **Biochemical Connections** What diseases are known to be caused by silent mutations?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# Nucleic Acid Biotechnology Techniques

# 13

Gopal Murti/CNRI/Phototake, NYC

Human chromosomes viewed through a scanning electron microscope.

## 13.1 Purification and Detection of Nucleic Acids

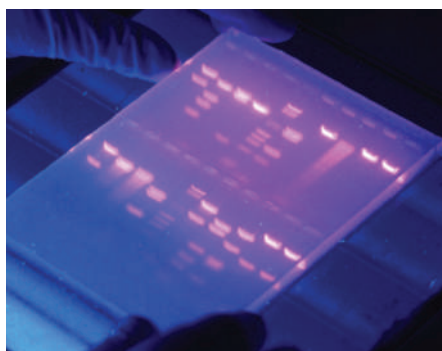
In early 1997, headlines around the world reported the successful cloning of a sheep by Scottish scientists, using a technique called somatic cell nuclear transfer. Since then more than 20 mammalian species have been cloned by similar techniques. Such striking examples of the power of the techniques for manipulating DNA have sparked enormous amounts of discussion. In this chapter, we will focus our attention on some of the most important methods used in biotechnology.

Experiments on nucleic acids frequently involve extremely small quantities of materials of widely varying molecular size. Two of the primary necessities are to separate the components of a mixture and to detect the presence of nucleic acids. Fortunately, powerful methods exist for accomplishing both goals.

### *How are nucleic acids separated?*

#### Separation Techniques

Any separation method depends on the differences between the items to be separated. Charge and size are two properties of molecules that are frequently used for separation. One of the most widely used techniques in molecular biology, **gel electrophoresis**, uses both these properties (see Section 5.3). Electrophoresis is based on the motion of charged particles in an electric field. For our purposes, it is enough to know that the motion of a charged molecule in an electric field depends on the ratio of its charge to its mass. A sample is applied to a supporting medium. With the use of electrodes, an electric current is passed through the medium to achieve the desired separation. Polymeric gels, such as agarose and polyacrylamide, are frequently used as supporting media for electrophoresis (Figure 13.1). They are prepared and cast as a continuous cross-linked matrix. The cross-linking gives rise to pores, and the choice of agarose versus polyacrylamide gels depends on the size of the molecules to be separated—agarose for larger fragments (thousands of oligonucleotides) and polyacrylamide for smaller ones (hundreds of oligonucleotides).



© D. VoItung/LookatSciences/Phototake, NYC

■ **FIGURE 13.1** Separation of oligonucleotides by gel electrophoresis. Each band seen in the gel represents a different oligonucleotide.

## Chapter Outline

### 13.1 Purification and Detection of Nucleic Acids

- How are nucleic acids separated?
- How can we visualize DNA?

### 13.2 Restriction Endonucleases

- Why are sticky ends important?

### 13.3 Cloning

- What is cloning?
- What are plasmids?

### 13.4 Genetic Engineering

- What is the purpose of genetic engineering?
- How can human proteins be made by bacteria?
- What is an expression vector?

### 13.5 DNA Libraries

- How do we find the piece of DNA we want in a library?

### 13.6 The Polymerase Chain Reaction

- What are the advantages of PCR?

### 13.7 DNA Fingerprinting

- How can differences in DNA from individuals be seen?

### 13.8 Sequencing DNA

- How do dideoxy nucleotides allow us to sequence DNA?

### 13.9 Genomics and Proteomics

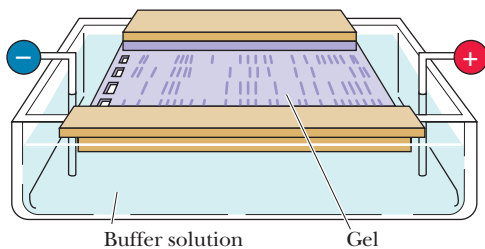
- How do microarrays work?

**OWL** Online homework for this chapter may be assigned in OWL.





■ The late Dolly, the most famous sheep in the world, produced by the cloning technique called somatic cell nuclear transfer.



■ **FIGURE 13.2** The experimental setup for gel electrophoresis. The samples are placed on the left side of the gel. When the current is applied, the negatively charged oligonucleotides migrate toward the positive electrode.

The charge on the molecules to be separated leads them to move through the gel toward an electrode of opposite charge. Nucleic acids and oligonucleotide fragments are negatively charged at neutral pH because of the presence of the phosphate groups. When these negatively charged molecules are placed in an electric field between two electrodes, they all migrate toward the positive electrode. In nucleic acids, each nucleotide residue contributes a negative charge from the phosphate to the overall charge of the fragment, but the mass of the nucleic acid or oligonucleotide increases correspondingly. Thus, the ratio of charge to mass remains approximately the same regardless of the size of the molecule in question. As a result, the separation takes place simply on the basis of size and is due to the sieving action of the gel. In a given amount of time, with a sample consisting of a mixture of oligonucleotides, a smaller oligonucleotide moves farther than a larger one in an electrophoretic separation. The oligonucleotides move in the electric field because of their charges; the distances they move in a given time depend on their sizes.

Most separations are done with an agarose gel in a horizontal position, called a *submarine gel* because it is underneath the buffer in the chamber. However, when DNA sequencing is done (see Section 13.8), a polyacrylamide gel is run in a vertical position. Many different samples can be separated on a single gel. Each sample is loaded at a given place (a distinct well) at the negative-electrode end of the gel, and an electric current flows until the separation is complete (Figure 13.2).

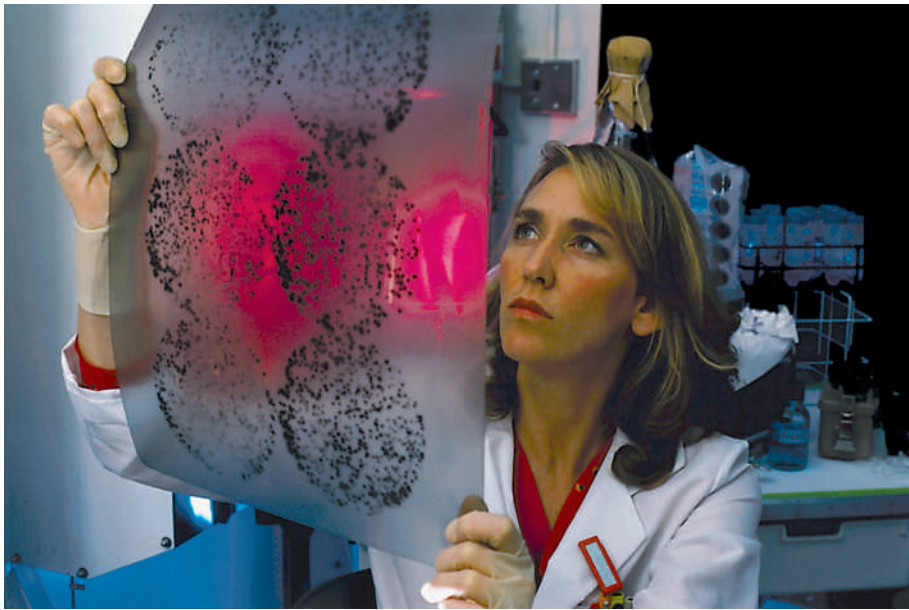
## Detection Methods

After the DNA pieces have been separated, they must be treated in some way that allows them to be seen. Some of these techniques allow all of the DNA to be seen, but others are more specific for certain DNA pieces.

### How can we visualize DNA and RNA?

The original method for detecting the separated products is based on radioactive labeling of the sample. A label, or tag, is an atom or molecule that allows visualization of another molecule. The isotope of phosphorus of mass number 32 ( $^{32}\text{P}$ , spoken as “P-thirty-two”) was widely used in the past for this purpose. More recently,  $^{35}\text{S}$ , or the isotope of sulfur of mass number 35 (spoken as “S-thirty-five”), has been used extensively. The DNA molecules undergo a reaction that incorporates the radioactive isotope into the DNA. When the labeled oligonucleotides have been separated, the gel is placed in contact with a piece of X-ray film. The radioactively labeled oligonucleotides expose the portions of the film with which they are in contact. When the film is developed, the positions of the labeled substances show up as dark bands. This technique is called **autoradiography**, and the resulting film image is an *autoradiograph* (Figure 13.3). While we will principally mention DNA in this section, most of the visualization techniques work with RNA too.

Many examples of autoradiographs can be seen in the scientific literature, but, as time goes on, autoradiography is being replaced by detection methods that do not use radioactive materials with their associated hazards. Many of these methods depend on emission of light (**luminescence**) by a chemical label attached to the fragments, and they can detect amounts of substances measured in picomoles. The way in which the label emits light depends on the application. When the base sequence of DNA is to be determined, the label is a series of four fluorescent compounds, one for each base. The gel with the separated products is irradiated with a laser; the wavelength of the laser light is absorbed by each of the four labels. Each of the four labeled compounds re-emits light at a different, characteristic, longer wavelength. This is called **fluorescence**. Another detection method that uses fluorescence involves the compound ethidium bromide. Its molecular structure includes a planar portion



Hank Morgan/Photo Researchers, Inc.

■ **FIGURE 13.3** An example of an autoradiograph.

that can slip between the bases of DNA, giving ethidium bromide fluorescent properties when it binds to DNA that differ from those observed when it is free in solution. An ethidium bromide solution is used as a stain for DNA in a gel. The solution soaks into the gel, and the DNA fragments in the gel can be seen as orange bands by shining ultraviolet light on the gel.

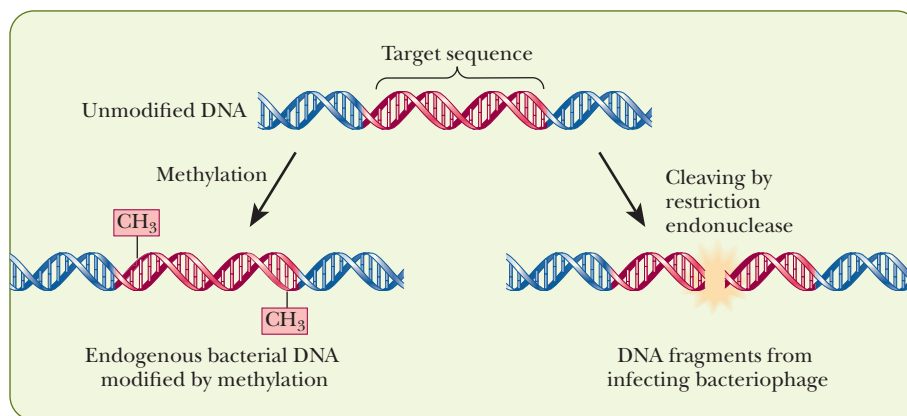
Ethidium bromide is a strong carcinogen, and great care must be taken in its use and disposal. Because of this, newer fluorescent dyes have been developed, such as SyBr Green and SyBr Gold, which are not as harmful and can be disposed of more easily.

## 13.2 Restriction Endonucleases

Many enzymes act on nucleic acids. A group of specific enzymes acts in concert to ensure the faithful replication of DNA, and another group directs the transcription of the base sequence of DNA into that of RNA. (We needed all of Chapters 10 and 11 to describe the manner in which these enzymes operate.) Other enzymes, called **nucleases**, catalyze the hydrolysis of the phosphodiester backbones of nucleic acids. Some nucleases are specific for DNA; others are specific for RNA. Cleavage from the ends of the molecule (by *exonucleases*) is known, as is cleavage in the middle of the chain (by *endonucleases*). Some enzymes are specific for single-stranded nucleic acids, and others cleave double-stranded ones. One group of nucleases, **restriction endonucleases**, has played a crucial role in the development of recombinant DNA technology.

Restriction enzymes were discovered in the course of genetic investigations of bacteria and **bacteriophages** (**phages** for short; from the Greek *phagein*, “to eat”), the viruses that infect bacteria. The researchers noted that bacteriophages that grew well in one strain of the bacterial species they infected frequently grew poorly (had *restricted* growth) in another strain of the same species. Further work showed that this phenomenon arises from a subtle difference between the phage DNA and the DNA of the strain of bacteria in which phage growth is restricted. This difference is the presence of methylated bases at certain sequence-specific sites in the host DNA and not in the viral DNA.

The growth-restricting host cells contain cleavage enzymes, the restriction endonucleases, that produce double-chain breaks at the unmethylated



■ **FIGURE 13.4 Methylation of DNA.** Methylation of endogenous DNA protects it from cleavage by its own restriction endonucleases.

specific sequences in phage DNA; the cells' own corresponding DNA sequences, in which methylated bases occur, are not attacked, as shown in Figure 13.4. These cleavage enzymes consequently degrade DNA from any source *but* the host cell. The most immediate consequence is a slowing of the growth of the phage in that bacterial strain, but the important thing for our discussion is that DNA from any source can be cleaved by such an enzyme if it contains the target sequence. More than 800 restriction endonucleases have been discovered in a variety of bacterial species. More than 100 specific sequences are recognized by one or more of these enzymes. Table 13.1 shows several target sequences.

### Many Restriction Endonucleases Produce “Sticky Ends”

Each restriction endonuclease hydrolyzes only a specific bond of a specific sequence in DNA. The sequences recognized by restriction endonucleases—their sites of action—read the same from left to right as they do from right to left (on the complementary strand). The term for such a sequence is a **palindrome**. (“Able was I ere I saw Elba” and “Madam, I’m Adam” are well-known linguistic palindromes.) A typical restriction endonuclease called *EcoRI* is isolated from *E. coli* (each restriction endonuclease is designated by an abbreviation of the name of the organism in which it occurs, and this abbreviation is the first letter of the name of the genus and the first two letters of the name of the species). The *EcoRI* site in DNA is 5'-GAATTC-3', where the base sequence on the other strand is 3'-CTTAAG-5'. The sequence from left to right on one strand is the same as the sequence from right to left on the other strand. The phosphodiester bond between G and A is the one hydrolyzed. This same break is made on both strands of the DNA. There are four nucleotide residues—two adenines and two thymines in each strand—between the two breaks on opposite strands, leaving **sticky ends**, which can still be joined by hydrogen bonding between the complementary bases. With the ends held in place by the hydrogen bonds, the two breaks can then be resealed covalently by the action of DNA ligases (Figure 13.5). If no ligase is present, the ends can remain separated, and the hydrogen bonding at the sticky ends holds the molecule together until gentle warming or vigorous stirring effects a separation. Some enzymes, such as *HaeIII*, cut in a way that leave a blunt end.

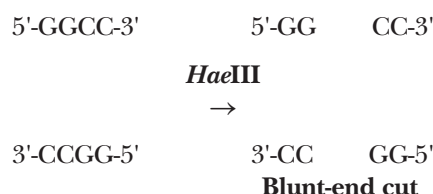


TABLE 13.1

Restriction Endonucleases and Their Cleavage Sites	
Enzyme*	Recognition and Cleavage Site
<i>Bam</i> HI	↓ 5'-GGATCC-3' 3'-CCTAGG-5' ↑
<i>Eco</i> RI	↓ 5'-GAATTC-3' 3'-CTTAAG-5' ↑
<i>Hae</i> III	↓ 5'-GGCC-3' 3'-CCGG-5' ↑
<i>Hind</i> III	↓ 5'-AAGCTT-3' 3'-TTCGAA-5' ↑
<i>Hpa</i> II	↓ 5'-CCGG-3' 3'-GGCC-5' ↑
<i>Not</i> I	↓ 5'-GCGGCCGC-3' 3'-CGCCGGCG-5' ↑
<i>Pst</i>	↓ 5'-CTGCAG-3' 3'-GACGTC-5' ↑

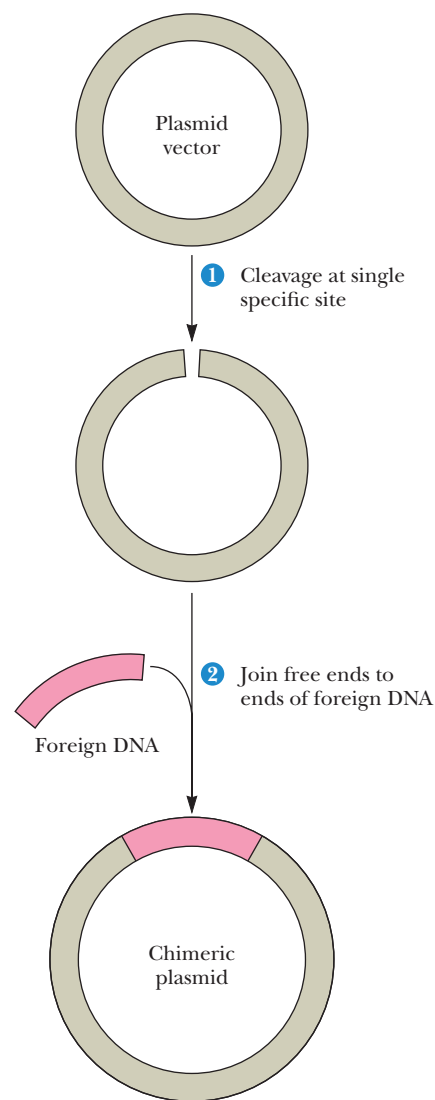
Arrows indicate the phosphodiester bonds cleaved by the restriction endonucleases.  
\* The name of the restriction endonuclease consists of a three-letter abbreviation of the bacterial species from which it is derived—for example *Eco* for *Escherichia coli*.

To make life more challenging for molecular biologists, some enzymes can also cut with less than absolute specificity. This is called *star (\*) activity* and can often be seen if the enzyme concentration is too high or the enzyme is incubated with the DNA too long. Also, different enzymes from different species may have the same specificity for cut sites. Such enzymes are called isoschizomers.

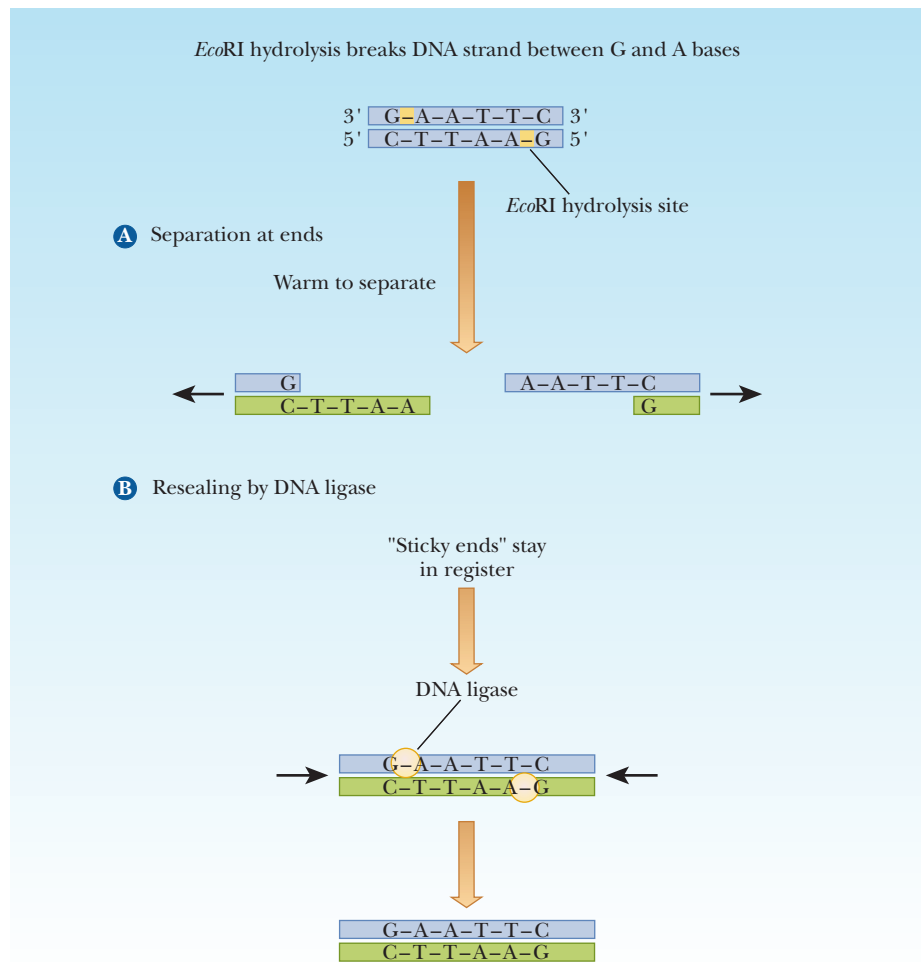
Why are sticky ends important?

As we will see in the next section, the production of sticky ends by a restriction enzyme is very important to the process of creating recombinant DNA. Once DNA is cut with a particular enzyme, the sticky ends fit back together. At that point it does not matter whether the pieces came from the original piece of DNA. In other words, pieces of DNA from different sources can be put together as long as they were both cut with the same restriction enzyme.





■ **FIGURE 13.6** Production of recombinant DNA. (1) Foreign DNA sequences can be inserted into plasmid vectors by opening the circular plasmid with a restriction endonuclease. (2) The ends of the linearized plasmid DNA are then joined with the ends of a foreign sequence, reclosing the circle to create a chimeric plasmid.



■ **FIGURE 13.5** Hydrolysis of DNA by restriction endonucleases. (a) Separation of ends. (b) Resealing of ends by DNA ligase.

### 13.3 Cloning

DNA molecules containing covalently linked segments derived from two or more DNA sources are called **recombinant DNA**. (Another name for recombinant DNA is **chimeric DNA**, named after the chimera, a monster in Greek mythology that had the head of a lion, the body of a goat, and the tail of a serpent.) The production of recombinant DNA was made possible by the isolation of restriction endonucleases.

#### Using “Sticky Ends” to Construct Recombinant DNA

If samples of DNA from two different sources are digested with the same restriction enzyme and then mixed, in some cases the sticky ends that anneal to one another are from different sources. The nicks in the covalent structure can be sealed with **DNA ligases** (Section 10.4), producing recombinant DNA (Figure 13.6).

Unfortunately, when two different kinds of DNA are combined using restriction enzymes and DNA ligase, relatively few product molecules are collected. Further experiments with the DNA will require large amounts to work with, and this is made possible by inserting the DNA into a viral or bacterial source. The virus is usually a bacteriophage; the bacterial DNA typically is derived from a **plasmid**, a small circular DNA molecule that is not part of the main circular



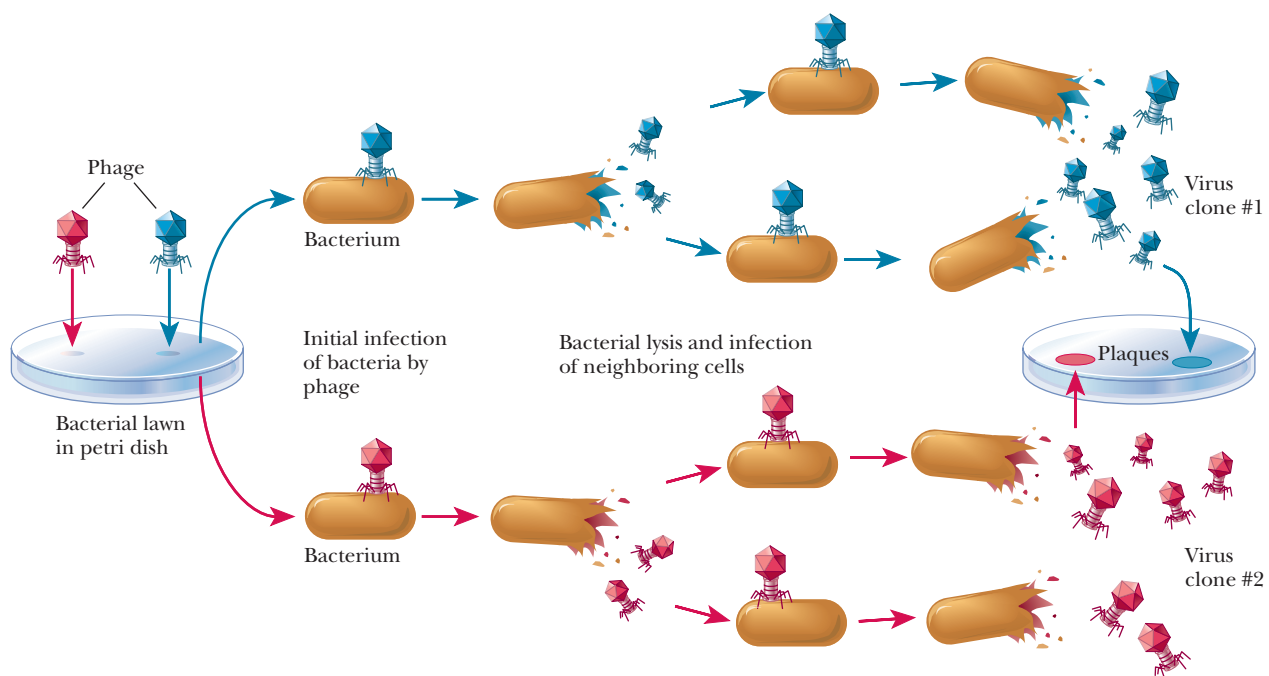
DNA chromosome of the bacterium. Using DNA from a viral or bacterial source as one of the components of a recombinant DNA enables scientists to take advantage of the rapid growth of viruses and bacteria and to obtain greater amounts of the recombinant DNA. This process of making identical copies of DNA is called **cloning**.

### What is cloning?

The term **clone** refers to a genetically identical population, whether of organisms, cells, viruses, or DNA molecules. Every member of the population is derived from a single cell, virus, or DNA molecule. It is particularly easy to see how individual bacteriophages and bacterial cells can produce large numbers of progeny. Bacteria grow rapidly, and large populations can be obtained relatively easily under laboratory conditions. Viruses also grow easily. We shall examine each of these examples in turn.

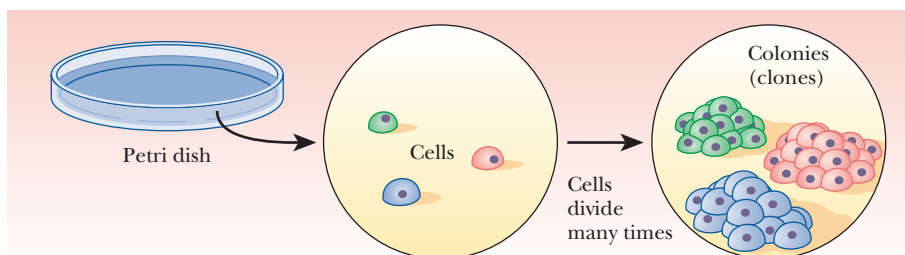
A virus can be considered a genome with a protein coat, usually consisting of many copies of one kind of protein or, at most, a small number of different kinds of proteins. The viral genome can be double-stranded or single-stranded DNA or RNA. For purposes of this discussion, we confine our attention to DNA viruses with double-stranded DNA. In the cloning of bacteriophages, a “lawn” of bacteria covering a petri dish is infected with the phage. Each individual virus infects a bacterial cell and reproduces, as do its progeny when they infect and destroy other bacterial cells. As the virus multiplies, a clear spot, called a *plaque*, appears on the petri dish, marking the area in which the bacterial cells have been killed (Figure 13.7). The plaque consists of the progeny viruses that are clones of the original.

To clone individual cells, whether from a bacterial or a eukaryotic source, a small number of cells is spread thinly over a suitable growth medium in a dish. Spreading the cells thinly ensures that each cell will multiply in isolation from the others. Each colony of cells that appears on the dish will then be a clone



■ **FIGURE 13.7** The cloning of a virus. The progeny of each individual phage (bacterial virus) infects and destroys bacteria on the petri dish, leaving clear spots known as plaques. Each plaque indicates the presence of a clone. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)

■ **FIGURE 13.8** The cloning of cells. Each individual cell divides many times, producing a colony of progeny. Each colony is a clone. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)



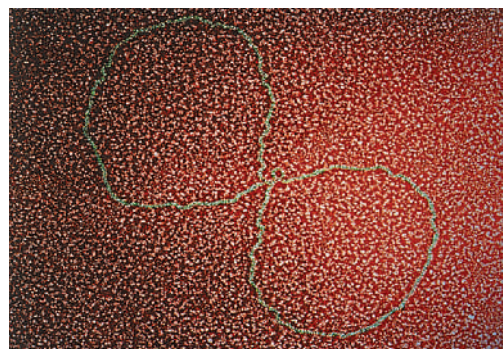
derived from a single cell (Figure 13.8). Because large quantities of bacteria and bacteriophages can be grown in short time intervals under laboratory conditions, it is useful to introduce DNA from a larger, slower-growing organism into bacteria or phages and to produce more of the desired DNA by cloning. If, for example, we want to take a portion of human DNA, which would be hard to acquire, and clone it in a virus, we can treat the human DNA and the virus DNA with the same restriction endonuclease, mix the two, and allow the sticky ends to anneal. If we then treat the mixture with DNA ligase, we have produced recombinant DNA. To clone it, we incorporate the chimeric DNA into virus particles by adding viral coat protein and allowing the virus to assemble itself. The virus particles are spread on a lawn of bacteria, and the cloned segments in each plaque can then be identified (Figure 13.9). The bacteriophage is called a **vector**, the carrier for the gene of interest that was cloned. The gene of interest is called many things, such as the “foreign DNA,” the “insert,” “geneX,” or even “YFG,” for “your favorite gene.”

### What are plasmids?

The other principal vector is a plasmid—bacterial DNA that is not part of the main circular DNA chromosome of the bacterium. This DNA, which usually exists as a closed circle, replicates independently of the main bacterial genome and can be transferred from one strain of a bacterial species to another by cell-to-cell contact. The foreign gene can be inserted into the plasmid by the successive actions of restriction endonucleases and DNA ligase, as was seen in Figure 13.6. When the plasmid is taken up by a bacterium, the DNA insert goes along for the ride (Figure 13.10). The bacteria that contain the DNA insert can then be grown in fermentation tanks under conditions that allow them to divide rapidly, amplifying the inserted gene many thousandfold.

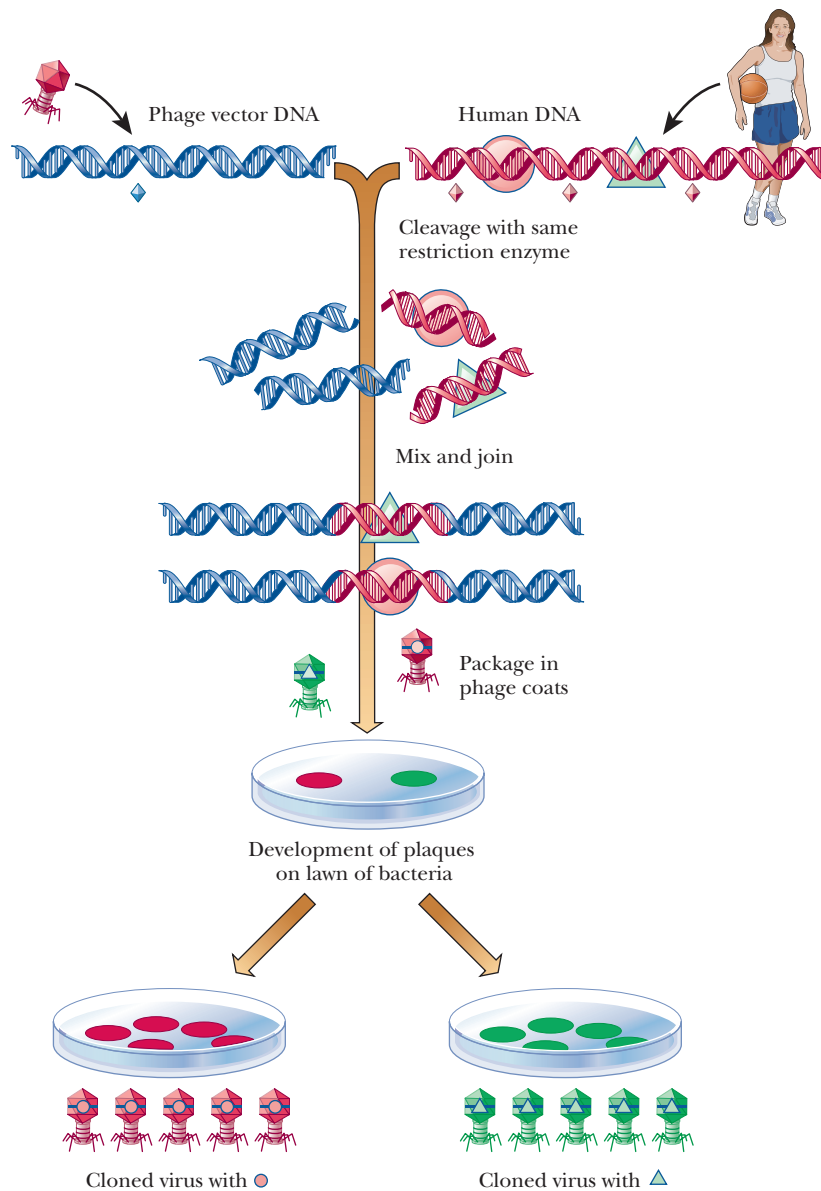
While the theory of cloning DNA into a plasmid is straightforward, there are several considerations for a successful experiment. When bacteria take up a plasmid, we say they have been *transformed*. Transformation is the process whereby new DNA is incorporated into a host. Bacteria are encouraged to take up foreign DNA by a couple of methods. One is to heat-shock the bacteria at 42°C, followed by placing them on ice. Another is to place them in an electric field, a technique called *electroporation*.

How are we to know which of the bacteria have taken up the plasmid? Because bacteria divide quickly, we would not want all the bacteria to grow—rather, only the ones that have the plasmid. This process is called **selection**. Each plasmid chosen for cloning must have some type of **selectable marker** that lets us know that the growing bacterial colonies contain the plasmid. These markers are usually genes that confer resistance to antibiotics. After transformation, the bacteria are plated on a medium containing the antibiotic to which the plasmid carries resistance. In that way, only the bacteria that took up the plasmids will grow (Figure 13.10). One of the first plasmids used for cloning is pBR322 (Figure 13.11). This simple plasmid was created from a naturally occurring one found in *E. coli*. Like all plasmids, it has an origin of replication, so it can

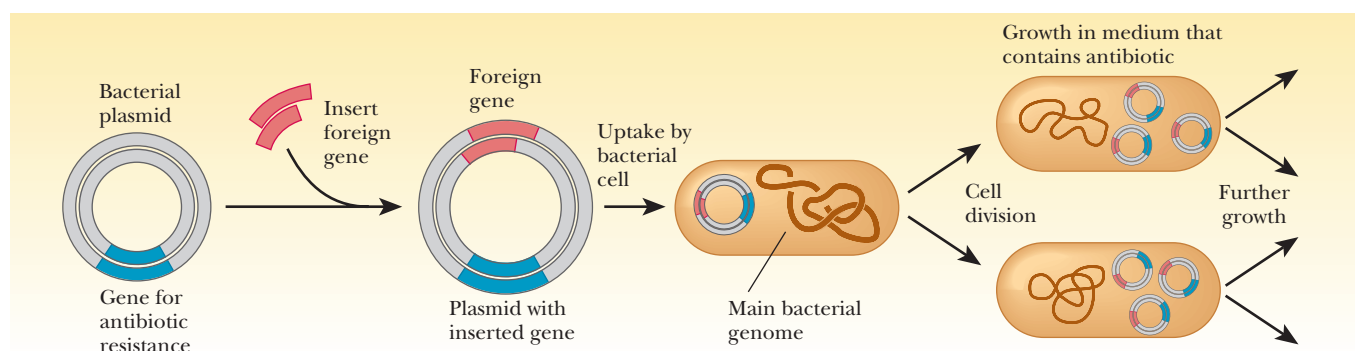


© Dan Electra/Phototake, NYC

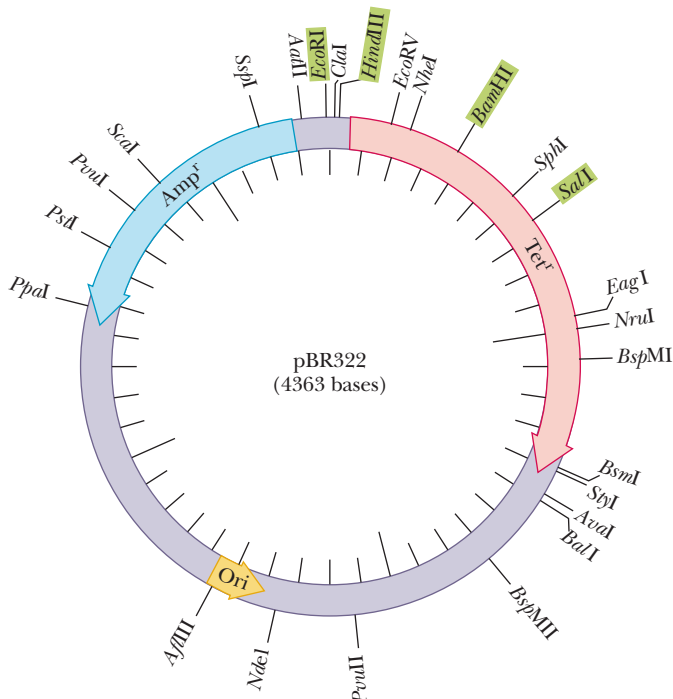
■ **DNA plasmids**—extrachromosomal self-replicating genetic elements of a bacterial cell.



■ **FIGURE 13.9** The cloning of human DNA fragments with a viral vector. Human DNA is inserted into viral DNA and then cloned. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)



■ **FIGURE 13.10** Selecting for recombinant DNA in a bacterial plasmid. The plasmid also contains a gene for antibiotic resistance. When bacteria are grown in a medium that contains the antibiotic, those that have acquired a plasmid grow. Bacteria without a plasmid cannot grow in this medium. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)

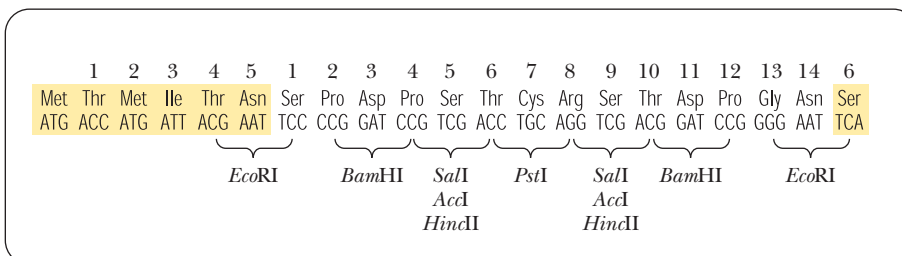


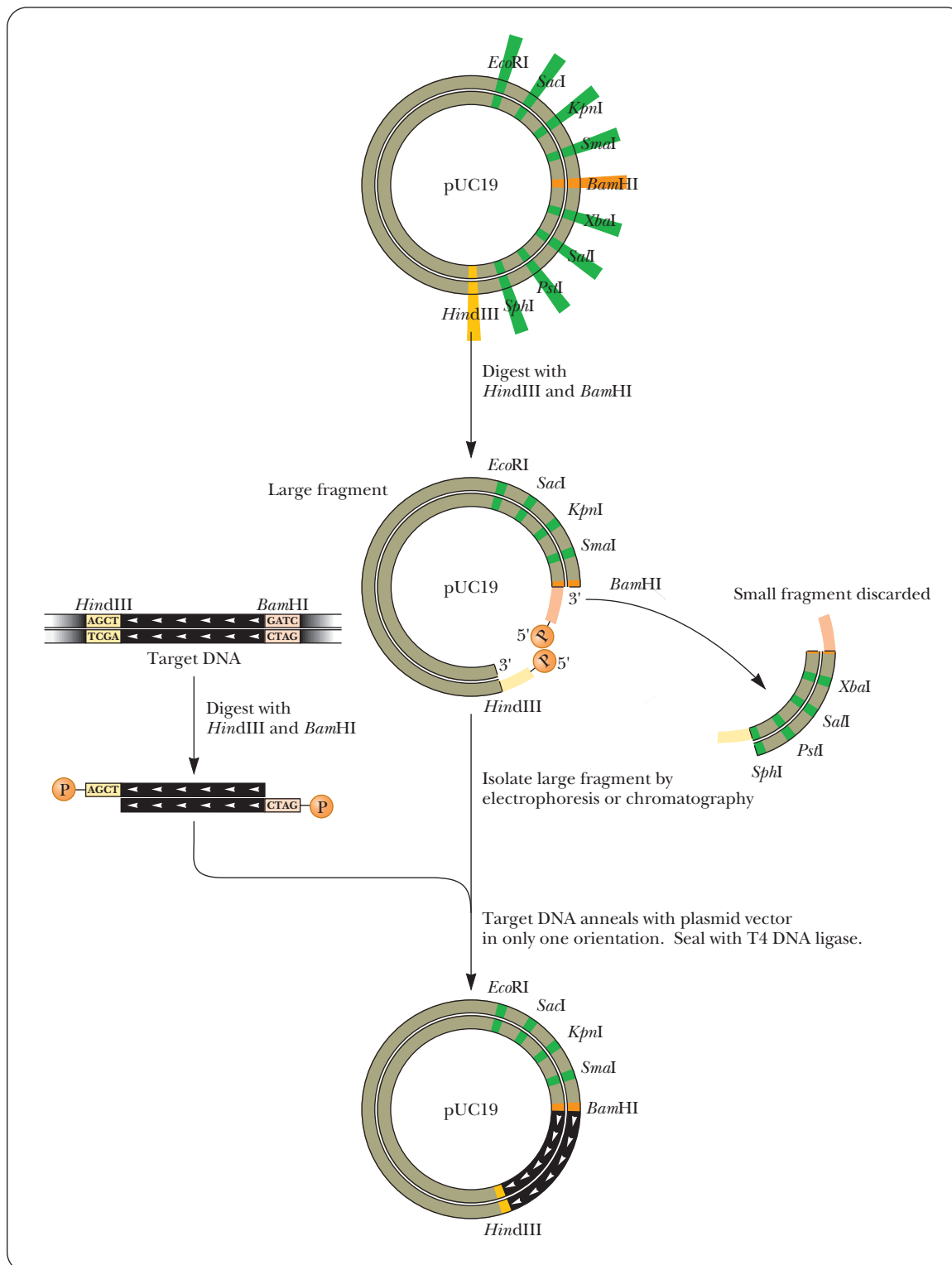
■ **FIGURE 13.11 Plasmid pBR322.** One of the first widely used cloning vectors is the plasmid pBR322. This 4363-base-pair plasmid contains an origin of replication (*ori*) and genes encoding resistance to the drugs ampicillin (*amp<sup>r</sup>*) and tetracycline (*tet<sup>r</sup>*). The locations of restriction endonuclease cleavage sites are indicated.

replicate independently of the rest of the genome. It has genes that confer resistance to two antibiotics, tetracycline and ampicillin. The genes are indicated *tet<sup>r</sup>* and *amp<sup>r</sup>*. The pBR322 plasmid has several restriction enzyme sites. The number and location of restriction sites is very important to a cloning experiment. The foreign DNA must be inserted at unique restriction sites so that the use of restriction enzymes opens up the plasmid at only one point. Also, if the restriction site chosen is inside one of the selection markers, the resistance to the antibiotic is lost upon insertion of the foreign DNA. This was, in fact, the original way selection was done with plasmids. Foreign DNA was inserted using restriction sites in the *tet<sup>r</sup>* gene. Selection was achieved by noting the loss of the ability of bacteria to grow on a medium containing tetracycline.

One of the early stumbling blocks to cloning was finding the right plasmid, one that had restriction sites that matched those enzymes needed to cut out the foreign DNA. As the technology to design plasmids improved, they were created with regions that had many different restriction sites in a small space. This region was called a **multiple cloning site (MCS)** or a **polylinker** (Figure 13.12). A popular cloning vector series is based on the pUC plasmids (Figure 13.13). The acronym *pUC* stands for *universal doning plasmid*. Each of these cloning vectors has an extensive MCS, which helps solve another problem with cloning—the directionality of the inserted DNA. Depending on what is to be done with the cloned DNA, controlling its orientation in the vector may be important. If only one restriction enzyme is used, such as *Bam*H1, then the

■ **FIGURE 13.12** A vector cloning site containing multiple restriction sites. This is called a polylinker or multiple cloning site (MCS). The colored amino acids are from the *lacZ* gene that is part of the plasmid. The MCS does not disrupt the normal reading frame of this sequence, so this plasmid can be used for blue/white screening (see text). (Adapted from Figure 1.14.2, in Ausubel, F. M., et al., 1987, Current Protocols in Molecular Biology. New York: John Wiley and Sons. Used by permission.)



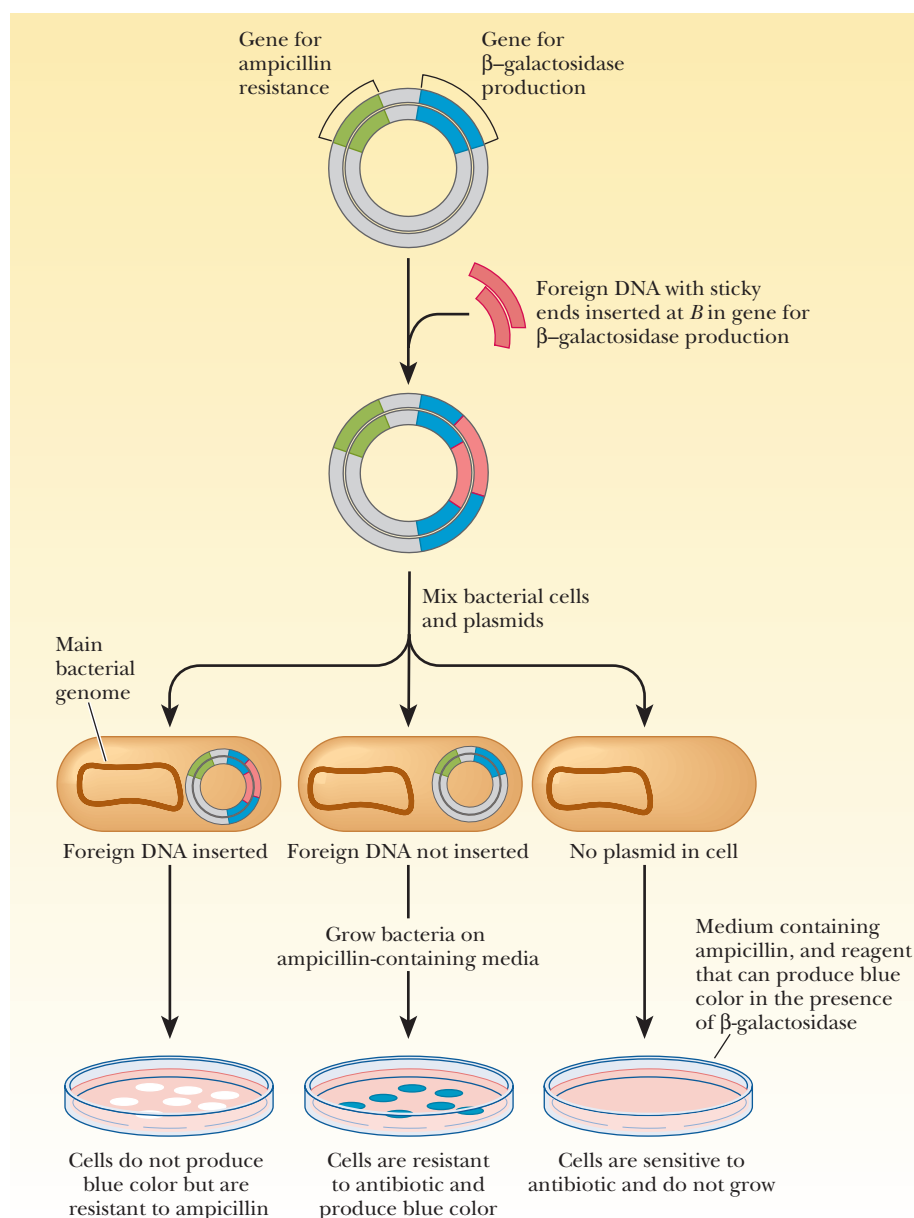


■ **FIGURE 13.13** Cloning with pUC plasmids. The pUC series of plasmids is very popular. They have extensive multiple cloning sites. Here we see an example of directional cloning. Two different restriction enzymes are used to cut open the MCS and to cut out a piece of DNA to be cloned. As a result, the DNA that is to be inserted can be incorporated in only one orientation.



foreign DNA can enter the plasmid in either of two directions. However, if the foreign DNA is cut out of its source at one end with *Bam*HI and at the other end with *Hind*III, and if these same two restriction enzymes are used to open up the plasmid, then the ends match in only one direction (Figure 13.13).

The use of the pUC plasmids also aids in the selection procedure. The older plasmids that were based on pBR322 had the shortcoming that the foreign DNA was inserted into the tetracycline resistance gene. This meant that the only way to spot bacteria that had taken up a plasmid that had also taken up the insert was that the bacteria would *not* grow on a medium containing tetracycline. This lack of growth by the successful clone made it challenging to go back and find the proper bacterial colonies. The pUC plasmids, however, have a characteristic that alleviates this procedure—they contain the *lacZ* gene, which is the basis for a selection technique called **blue/white screening**. The *lacZ* gene codes for the  $\alpha$ -subunit of the enzyme  $\beta$ -galactosidase, which is used to cleave disaccharides, such as lactose (Chapter 16). The MCS is located inside the *lacZ* gene, so that, when foreign DNA is inserted, it inactivates the gene. Figure 13.14 shows how this characteristic is useful. The pUC plasmid is cut open in its MCS by restriction enzymes, and the foreign DNA (red) is cut out of its source with the same



**FIGURE 13.14 Clone selection via blue/white screening.** The pUC plasmid contains a gene for ampicillin resistance and the *lacZ* gene. The latter produces the  $\alpha$ -subunit of  $\beta$ -galactosidase. Transformed cells are plated on an agar medium containing ampicillin and a dye called X-gal. The *lacZ* gene is found inside the multiple cloning site of the plasmid. When the plasmid and the DNA to be cloned are cut with a restriction enzyme and then mixed together, two possibilities result. The DNA insert can be incorporated as shown (red insert seen inside plasmid), or the plasmid can recircularize without the insert. When this mixture is used to transform bacteria, there can be three products. (left) The bacteria take up a plasmid that has the insert. This plasmid confers ampicillin resistance to the cells, but the *lacZ* gene is inactivated by the presence of the insert. These cells grow and are the normal off-white color of bacterial colonies. (middle) The bacteria take up the recircularized plasmid. This plasmid confers ampicillin resistance, so the cells grow. The plasmid makes the  $\alpha$ -subunit of  $\beta$ -galactosidase. The  $\beta$ -galactosidase cleaves the X-gal, causing the dye to turn blue, so these cells grow with a blue color. (right) Bacteria take up no plasmid at all. These cells do not grow, because of their sensitivity to ampicillin. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)

enzymes. These are then combined and joined together with DNA ligase to yield two products in the ligation reaction. The desired product is the plasmid that now contains the foreign DNA. The other is a plasmid that has reclosed upon itself without the inserted DNA. This type is much less common when two different restriction enzymes are used to open the plasmid, but it still occurs infrequently. When this mixture is used to transform the bacteria, there are three possible products: (1) bacteria that took up the plasmid with the insert, (2) bacteria that took up the plasmid without the insert, and (3) bacteria that took up no plasmid at all. The mixture of bacteria from the transformation is plated on a medium containing ampicillin and a dye, such as X-gal.  $\beta$ -Galactosidase hydrolyzes a bond in the X-gal molecule, turning it blue. The bacteria to be transformed are mutants that make a defective version of the  $\beta$ -galactosidase that lacks the  $\alpha$ -subunit. If the bacteria take up no plasmid at all, they lack the ampicillin-resistance gene and do not grow. If the bacteria take up a plasmid lacking the insert, they have a functional *lacZ* gene and produce the  $\alpha$ -subunit of  $\beta$ -galactosidase. These colonies produce active  $\beta$ -galactosidase, which cleaves the dye, X-gal. These colonies grow with a blue color. If the bacteria take up a plasmid that contains the insert, the *lacZ* gene is inactivated. These colonies are the off-white color of normal bacterial colonies on agar.

## 13.4 Genetic Engineering

The previous sections dealt with how DNA of interest could be inserted into a vector and amplified by cloning. One of the most important purposes for doing this is to be able to produce the gene product in larger quantities than could be acquired by other means. When an organism is intentionally changed at the molecular level so that it exhibits different traits, we say it has been *genetically engineered*.

In a sense, **genetic engineering** on an organismal level has been around since humans first started to use selective breeding on plants and animals. This procedure did not deal directly with the molecular nature of genetic material, nor was the appearance of traits under human control. Breeders had to cope with changes that arose spontaneously, and the only choice was whether to breed for a trait or to let it die out. An understanding of the molecular nature of heredity and the ability to manipulate those molecules in the laboratory have, of course, added to our ability to control the appearance of these traits.

### What is the purpose of genetic engineering?

The practice of selective alteration of organisms for both agricultural and medical purposes has profited greatly from recombinant DNA methods. Genetic engineering of crop plants is an active field of research. Genes for increased yields, frost resistance, and resistance to pests are introduced into commercially important plants such as strawberries, tomatoes, and corn. Similarly, animals of commercial importance—mostly mammals, but also including fish—are also genetically altered. Some variations introduced in animals have medical implications. Mice with altered genetic makeup are used in the research laboratory. In another medically related field, researchers working with insect-borne diseases, such as malaria, are trying to engineer strains of insects, such as the mosquito, *Anopheles gambiae*, that can no longer transmit the infection to humans (Figure 13.15). In all cases, the focus of the research is to introduce *traits that can be inherited* by the descendants of the treated organisms. In the treatment of human genetic disease, however, the aim is not to produce heritable changes. Serious ethical questions arise with the manipulation of human genetics; consequently, the focus of research has been on forms of **gene therapy** in which cells of specific tissues in a living person are altered in a way that alleviates the effects of the disease. Examples of diseases that may someday be treated in this



James Gathany and the Centers for Disease Control

■ **FIGURE 13.15** Two adult female *Anopheles gambiae* mosquitoes (ventral view). The one on the left is a mutant. Scientists are attempting to produce strains of these mutant mosquitoes, which are unable to transmit malaria to humans, in hopes that they will replace the malaria carriers.

way include cystic fibrosis, hemophilia, Duchenne muscular dystrophy, and severe combined immune deficiency (SCID). The last of these is also known as “bubble-boy syndrome,” because those who have it must live in isolation (in a large “bubble”) to avoid infection.

### DNA Recombination Occurs in Nature

When recombinant DNA technology was in its early stages in the 1970s, considerable concern arose both about safety and about ethical questions. Some of the ethical questions are still matters of concern. One that has definitely been laid to rest is the question of whether the process of cutting and splicing DNA is an unnatural process. Indeed, DNA recombination is a common part of the crossing over of chromosomes. There are many, varied reasons for *in vivo* recombination of DNA, two of which are the maintenance of genetic diversity and the repair of damaged DNA (Section 10.5).

Until recently, heritable changes in organisms were solely those that arose from mutations. Researchers in the field took advantage of both spontaneous mutations and those produced by exposure of organisms to radioactive materials and other substances known to induce mutations. Selective breeding was then used to increase the population of desired mutants. It was not possible to produce “custom-tailored” changes in genes.

Since the advent of recombinant DNA technology, it is possible (within limits) to change specific genes, and even to change specific DNA sequences within those genes, to alter the inherited characteristics of organisms. Bacteria can be altered to produce large amounts of medically and economically important proteins. Animals can be manipulated to cure, or to alleviate the symptoms of, their genetic diseases, and agriculturally important plants can be made to produce greater crop yields or be given increased resistance to pests. The following Biochemical Connections box gives some examples of agricultural applications of genetic engineering.

### Bacteria as “Protein Factories”

We can use the reproductive power of bacteria to express large quantities of a mammalian protein of interest; however, the process is often more complicated than it might seem because most mammalian proteins are heavily processed after their initial transcription and translation (Section 12.7). Because bacteria have little posttranslational modification of their proteins, they lack the enzymes necessary for this processing.

#### *How can human proteins be made by bacteria?*

An application of genetic engineering that is of considerable practical importance is the production of human insulin by *E. coli*. This was one of the first human proteins produced through genetic engineering, and its production eliminated the problems related to harvesting insulin from large numbers of laboratory animals and giving humans a peptide from another species. The process is far from straightforward, however. A significant problem is that the insulin gene is split. It contains an **intron**, a DNA sequence that codes for RNA that is eventually deleted in the processing of the mRNA that directs the synthesis of the protein (see Section 11.6). Only the RNA transcribed from DNA sequences called **exons** appears in mature mRNA (Figure 13.16). Bacteria do not have the cellular apparatus for splicing introns out of RNA transcripts to give functional mRNA. One might think that the problem could be solved by using cDNA (Section 13.6) obtained from the mRNA for insulin in a reaction catalyzed by reverse transcriptase. The problem here is that the polypeptide encoded by this mRNA contains an end peptide and a central peptide, which is to be removed from it by further processing in insulin-producing cells to yield two polypeptide chains, designated A and B (Figure 13.16).

# Biochemical Connections PLANT SCIENCE

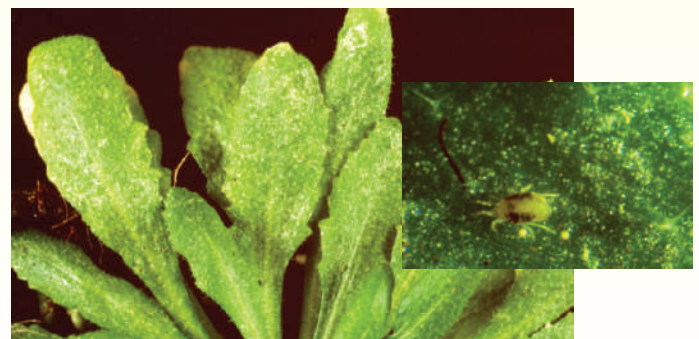
## Genetic Engineering in Agriculture

The idea of genetic engineering in humans often upsets people because they are fearful of “playing God.” Genetic manipulations in plants appear to cause less controversy, although many people are still concerned about the practice. Nevertheless, many types of modifications have been made, and some have been introduced with little fanfare and have shown some signs of success. Several examples are listed here. It is important to realize that many of the modifications that have been made using genetic engineering are just controlled versions of the selective breeding used for centuries to improve crop and animal production.

1. **Disease resistance.** Because most high-production crops involve specialized strains, many are more susceptible to fungal disease and insect damage. As a consequence, many herbicides and insecticides are applied liberally during the growing season. In many cases, other plants have natural resistance to these pests. When the gene that produces the resistance can be isolated, it can be transferred into other plants. There has been limited success in transferring such resistance to crop species. In 2000, the public became more aware of genetically modified (GM) food crops because of news about Bt corn, a crop carrying a bacterial gene that produces a toxin poisonous to certain caterpillars. The Bt gene, from *Bacillus thuringiensis*, has been put into corn and cotton to increase crop yields by killing the caterpillars that would otherwise eat those crops. It also reduces the amount of pesticides needed to grow the crops. On the negative side is the potential harm to other species. For example, environmental groups oppose the planting of Bt corn because of its effect on the monarch butterfly in lab tests.
2. **Nitrogen fixation.** Nitrogen fixation is most easily accomplished by bacteria that grow in nodules on the roots of certain legumes, such as beans, peas, and alfalfa. We now realize that the genes of the nitrogen-fixing bacteria actually become shared with the genes of the host plant. Much research is going on to determine whether these genes could be incorporated into other plants, which would reduce the amount of nitrogen fertilizer needed for maximal plant growth and crop production.
3. **Frost-free plants.** Many marine organisms, such as fish found in the Antarctic, produce a so-called antifreeze protein, a protein characterized by its hydrophobic surface, which prevents the formation of ice crystals at low temperature. In plants exposed to freezing temperatures, ice crystals in their tissues actually cause the frost damage. Insertion of the gene for the antifreeze protein into strawberries and potatoes, for example, has resulted in crops that are stable during late spring frost or in areas with very short growing seasons. Much controversy has ensued over these crops because, in each case, a foreign gene from a nonplant species was introduced into a plant. This type of cross-species gene modification causes widespread concern, even though the taste and texture of the modified product is indistinguishable from those of foods produced by unmodified plants.
4. **Tomatoes with a long shelf life.** The Flavr Savr tomato has a genetic modification, but one in which no new gene has been introduced. Rather, one of the plant's own genes has been deactivated. The gene in question is the one that allows the tomato plant to produce ethylene, a key compound in the ripening process. Because this gene has been deactivated,

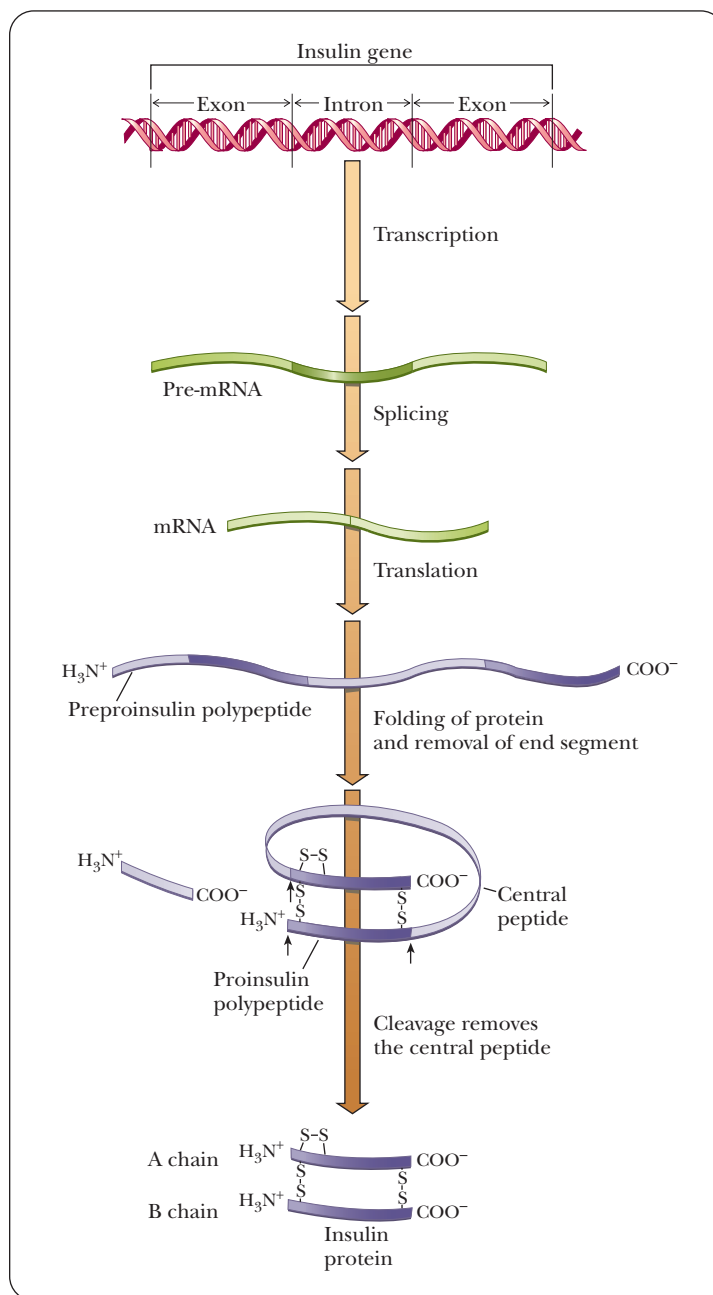
the tomatoes mature on the plant until they just begin to show some pink color—the exact stage at which tomatoes are picked and sent to market. Typically, fields of unmodified tomatoes need to be picked as many as nine times in a season, because the tomatoes all ripen at different rates. When an unmodified tomato ripens, it continues to make ethylene, which leads to overripening, softening, and deterioration. Fields of the modified tomatoes can be picked once or twice, and the harvested fruits can be made to ripen as needed by exposing them to exogenous ethylene gas. Because the Flavr Savr tomato does not make its own ethylene, the shelf life of the ripe, ready-to-eat tomato is increased. The financial saving made possible by less time spent picking tomatoes and the extended shelf life results in a cheaper and better product. Best of all, the taste is indistinguishable from that of unmodified tomatoes. (One of the authors of this book has eaten Flavr Savr tomatoes many times.)

5. **Increased milk production.** Much controversy exists over providing cows with supplemental bovine somatotropin (BST), also called growth hormone, a hormone that increases metabolism and milk production in dairy cows. The controversy centers on human consumption of the hormone. However, BST is a peptide hormone, so it is hydrolyzed in the digestive tract and is not absorbed directly into the human bloodstream (Section 24.2). Furthermore, all milk must contain some of this hormone, because the cow cannot produce milk without it. A more soundly based concern about BST supplementation in dairy cows is that the cows given extra BST often develop mastitis, an inflammation of the udder caused by bacterial infection. Mastitis is frequently treated with high doses of antibiotics, and some of these antibiotics could end up in the milk, causing problems for people with certain food sensitivity.
6. **Good predator attraction.** Researchers in the Netherlands were studying the mustard plant, *Arabidopsis thaliana*, which is very susceptible to attack by herbivorous spider mite predators. They introduced a strawberry gene into the mustard plants that produces a chemical attractant for predatory mites that eat the herbivorous spider mites.



■ **Transgenic mustard plant attracts good predators.** Researchers created a transgenic version of the mustard plant, *Arabidopsis thaliana*, that contained a strawberry gene that produces an attractant for predatory mites that will eat herbivorous spider mites (shown in inset). (Photos courtesy of Marcel Dicke/Wageningen University – The Netherlands – [www.ent.wur.nl](http://www.ent.wur.nl))





### ■ FIGURE 13.16 Synthesis of insulin in humans.

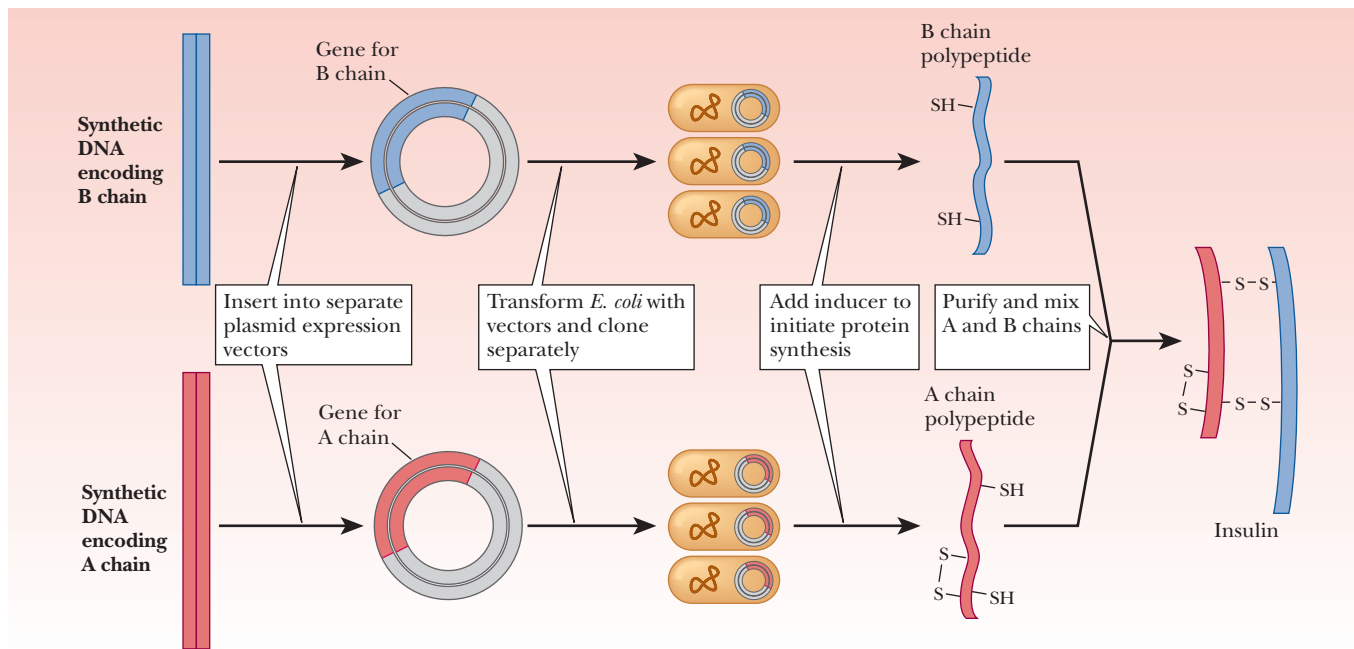
The insulin gene is a split gene. The intervening sequence (intron) encodes an RNA transcript that is spliced out of the mRNA. Only the portions of the gene called exons are reflected in the base sequence of mRNA. Once protein synthesis takes place, the polypeptide is folded, cut, and spliced. The end product, active insulin, has two polypeptide chains as a result. (*Adapted with permission from Dealing with Genes: The Language of Heredity, by Paul Berg and Maxine Singer, © 1992 by University Science Books.*)

The approach to this problem is to use two synthetic DNAs, one encoding the A chain of insulin and the other encoding the B chain. These synthetic DNAs are produced in the laboratory using methods that were developed by synthetic organic chemists. Each DNA is inserted into a separate plasmid vector (Figure 13.17). The vectors are taken up by two different populations of *E. coli*. The two groups are then cloned separately; each group of bacteria produces one of the two polypeptide chains of insulin. The A and B chains are extracted and mixed, finally producing functional human insulin. The Biochemical Connections box on page 373 describes some other successes in the production of human proteins for which the demand greatly exceeds the supply.

### Protein Expression Vectors

The plasmid vectors pBR322 and pUC are referred to as *cloning vectors*. They are used to insert the foreign DNA and to amplify it. However, if the goal is to



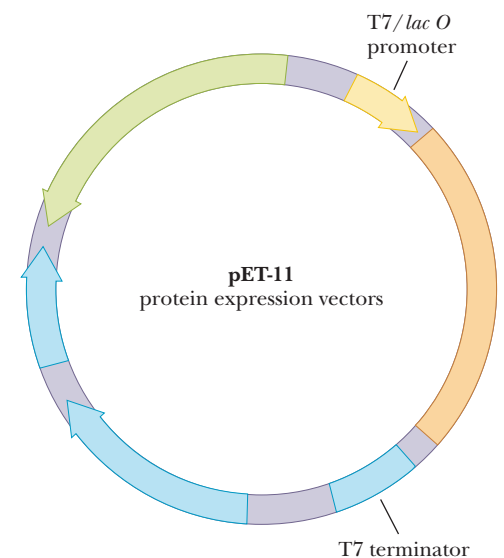


■ **FIGURE 13.17** Production of recombinant human insulin. Active human insulin can be produced in bacteria by the use of two separate batches of *E. coli*. Each batch produces one of the two chains, the A chain or the B chain. The two chains are mixed to produce active insulin. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)

produce the protein product from the foreign DNA, they are not suitable. An **expression vector** is needed.

### What is an expression vector?

An expression vector has many of the same attributes as a cloning vector, such as the origin of replication, a multiple cloning site, and at least one selectable marker. In addition, it must be able to be transcribed by the genetic machinery of the bacteria into which it is transformed. This means that it must have a promoter for RNA polymerase, and the RNA transcribed must have a ribosomal binding site so that it can be translated. It must also have a transcription termination sequence; otherwise, the entire plasmid is transcribed instead of just the inserted gene. Figure 13.18 shows a schematic of an expression vector. Upstream of the site where the foreign DNA is inserted is the transcription promoter. Often this is the promoter for a viral RNA polymerase called **T7 polymerase**. There is also a T7 terminator at the other end of the MCS. After the insert is successfully ligated, the plasmid is transformed into an expression strain of bacteria, such as *E. coli* JM109 DE3. What makes this strain unique is that it has a gene that produces T7 RNA polymerase, but the gene is under the control of the *lac* operon (Chapter 11). Once the bacteria are growing well with the plasmid, the cells are given a lactose analogue, IPTG (isopropylthiogalactoside). This stimulates the *lac* operon in the bacteria, which then produce T7 RNA polymerase, which then binds to the plasmid T7 promoter and transcribes the gene. The bacterial cells then translate the mRNA into protein. This selective control of the expression is important because many foreign proteins are toxic to the cells; expression must be timed carefully. The plasmid shown in Figure 13.18 also has the *lacI* gene, although it is transcribed in the opposite direction. This produces the repressor for the *lac* operon to help make sure that none of the foreign proteins are transcribed unless the system is induced by IPTG. The following Biochemical Connections box gives an example of how protein expression can be linked to a novel purification scheme.



■ **FIGURE 13.18** The pET expression vectors. These plasmids have the usual plasmid components, such as an origin of replication, MCS, and an antibiotic resistance gene (which provides resistance to ampicillin). In addition, their MCS is between the promoter for binding T7 RNA polymerase and a termination site for T7 RNA polymerase. When these vectors carry inserts, the insert DNA can be transcribed in the cell by T7 RNA polymerase. The cell then translates the mRNA into protein. (Courtesy of Stratagene.)

## Genetic Engineering in Eukaryotes

When the target organism for genetic engineering is an animal or a plant, one must consider that these are multicellular organisms with multiple kinds of tissues. In bacteria, altering the genetic makeup of a cell implies a change in the whole single-celled organism. In multicellular organisms, one possibility is to change a gene in a specific tissue, one that contains only one kind of differentiated cell. In other words, the change is *somatic*, affecting only the body tissues of the altered organism. In contrast, changes in germ cells (egg and sperm cells), called *germ-line* changes, are passed on to succeeding generations. If germ cells are to be modified, the change must be made at an early stage in development, before the germ cells are sequestered from the rest of the organism. Attempts to produce such changes have succeeded in comparatively few organisms, such as plants, fruit flies, and some other animals such as mice. Genetic engineering in plants frequently uses a vector based on a bacterial plasmid from the crown gall bacterium, *Agrobacterium tumefaciens*. Cells of this bacterium bind to wounded plant tissue, allowing plasmids to move from the bacterial cells into the plant cells. Some of the plasmid DNA inserts itself into the DNA of the plant cells in the only known natural transfer of genes from a bacterial plasmid to a eukaryotic genome. Expression of plasmid genes in the plant gives rise to a tumor called a *crown gall*. Whole, healthy plants can grow from gall cells, even though they are not germ cells. (This process, of course, does not take place in animals.) The plants that grow from the gall cells can produce fertile seeds, allowing the gene that has been transferred to be continued in a new strain of the plant. Genes from any desired source can be incorporated into the *A. tumefaciens* plasmid and then transferred to a plant. This method was used to genetically engineer tomato plants that resist defoliation by caterpillars (Figure 13.19). A gene that encodes a protein toxic to caterpillars was taken from the bacterium *Bacillus thuringiensis* to bring about this modification. Work is continuing on other useful modifications of food crops. Many observers of this whole line of research have raised questions about both the safety and the ethics of the process. The public became more aware of the extent to which genetically modified (GM) foods were in circulation in 2000, when corn that had been modified with the gene from *Bacillus thuringiensis* (Bt corn) showed up in taco shells. This had been an accident, as the Bt corn had been approved only for animal feed and not for human consumption, pending studies of potential allergenic effects. Environmentalists are also concerned about the effect of GM crops for two reasons. The first concern is the effect on nontargeted insects, such as the monarch butterfly, which may be particularly sensitive to the toxin produced by the Bt gene. Second is the potential



■ **FIGURE 13.19** A transgenic tomato plant. Recombinant DNA methods have produced plants that resist defoliation by caterpillars. Tomatoes with a longer shelf life are another result of this research.

Courtesy Calgene LLC

# Biochemical Connections ALLIED HEALTH

## Human Proteins through Genetic Recombination Techniques

Genetic engineering techniques have made it possible to prepare many proteins with the amino acid sequences found in humans. This is done by isolating the gene that encodes the protein and then incorporating the gene into a bacterium or into eukaryotic cells grown in tissue culture. Recent work in the cloning of cows and sheep has been justified in part as a potential source of human proteins that might be produced in the animals' milk. In the process of establishing systems for the commercial production of any protein product, it is always easier to isolate and purify the proteins if they are present in the extracellular fluid surrounding the cells that produced them.

Frequently, bacterial systems are used to produce human proteins. Bacteria are very cheap to grow and easy to work with. Experimental systems using eukaryotic cells are much more difficult and expensive, in spite of the fact that eukaryotic cells offer the advantage of being able to add sugar residues to glycoproteins and perhaps help the proteins fold into their biologically active conformations. Still, using bacteria and recombinant DNA for protein production has its complications. Growing proteins in bacteria requires the addition of the Shine–Dalgarno sequence (Section 12.4) to the mRNA to ensure its binding to the ribosome. The genetic manipulations necessary to make the product in a foreign host often result in a protein that is slightly modified, usually by having a few extra amino acids on the N-terminus. The presence of the extra amino acids complicates the process of approval by the FDA, because there is more concern about side effects.

Some notable successes in human protein production include the following:

1. Insulin (see text) is traditionally isolated from horses and swine, and it has been used to treat type I diabetes mellitus. However, after many years of use, about 5% of diabetics develop a severe allergy to the foreign protein. These people can now be treated with bacteria-derived human insulin, which costs only about 10% more than the animal-derived hormones.
2. In the past, human growth hormone (HGH) has been obtained only by extraction from the pituitary glands of cadavers, a practice that carries the risk of the cells being contaminated with HIV or other diseases. HGH is used for therapy in genetic dwarfs and for muscle-wasting diseases, including AIDS. HGH is a relatively large protein hormone, with more than 300 amino acids, but it is a simple protein with no sugar residues, so it was relatively easy to clone into bacteria. It is interesting to note that the availability of increased amounts of safe HGH has resulted in the emergence of a black market in the sale of the hormone for muscle building in athletes.
3. Two proteins, tissue plasminogen activator (TPA) and enterokinase (EK), are known to dissolve blood clots. If they are injected into the body within a critical period after a heart attack or stroke, either of these two proteins can prevent or minimize the disastrous effects of blood clots in the heart or brain. Without genetic recombination, enough of these proteins would never be available for this treatment to be of practical

use. Currently, two different companies are producing the two proteins.

4. Erythropoietin (EPO) is a hormone that stimulates the bone marrow to produce erythrocytes, commonly referred to as red blood cells (RBCs). This relatively small protein is lost during kidney dialysis; a healthy kidney also filters this hormone from the blood but then reabsorbs it back into the body. People with chronic kidney failure who are on dialysis while awaiting a kidney transplant thus suffer an additional problem of being chronically anemic from having too few RBCs. Such individuals must receive regular blood transfusions, which carries the risk of disease and possible allergic reaction. A genetically engineered erythropoietin, Epogen, is available from Amgen; it is the most commercially successful example of a human hormone coming from genetic research. EPO is also one of the most successful examples of a recombinant protein that has had its original purpose subverted. Because some endurance athletes have used EPO to boost their levels of RBCs, giving them a big advantage over their competitors, a black market for EPO exists in the sports industry. The situation came to a head during the 1998 Tour de France when a team doctor was arrested crossing the border into France while carrying countless vials of EPO. The EPO was eventually traced to a theft from a hospital. The team was ejected from the race, and the international governing body of cycling has invested hundreds of thousands of dollars to combat EPO usage, including developing assays for recombinant EPO.



■ **Epogen.** Recombinant erythropoietin is a human peptide hormone produced in bacteria through cloning and expression. This drug is used to boost red blood cells in patients who have lost blood through surgery or kidney dialysis. It has, unfortunately, also become one of the drugs of choice for illegally enhancing athletic performance.

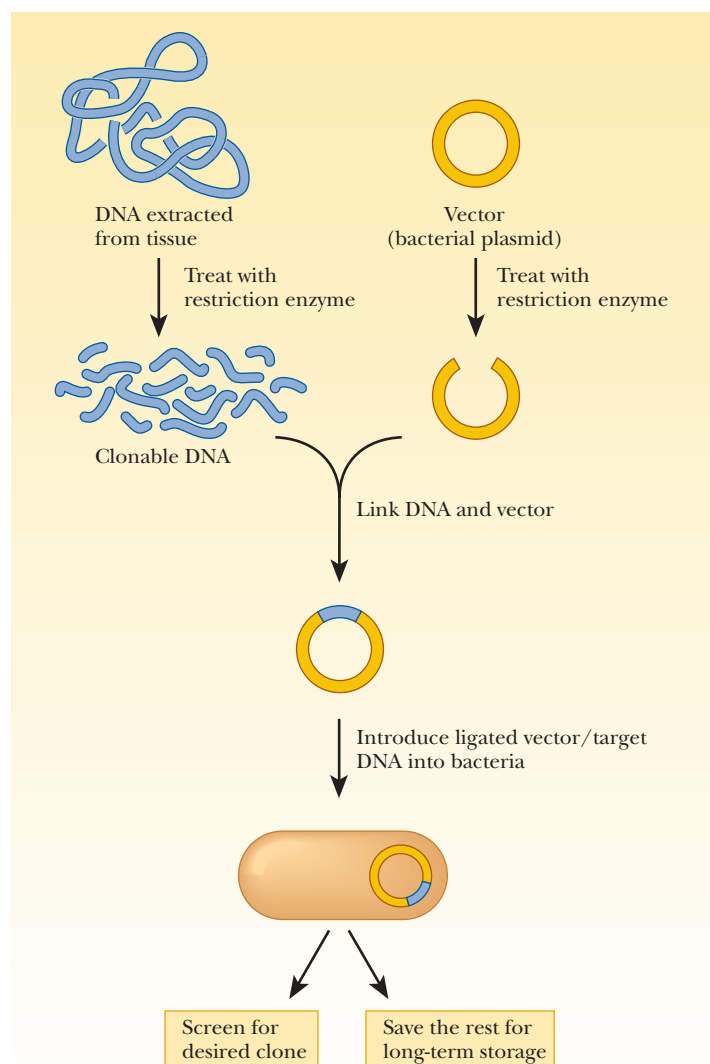
© AP Photo/Ed Betz

to create a super breed of insect accidentally that is immune to the effect of the toxin. On the positive side, fields planted with Bt cotton plants can sometimes use up to 80% less pesticide than fields planted with ordinary cotton.

## 13.5 DNA Libraries

Because methods exist for selecting regions of DNA from the genome of an organism of interest and cloning those regions in suitable vectors, a question that immediately comes to mind is whether we can take *all* the DNA of an organism (the total genome) and clone it in chunks of reasonable size. The answer is that we can do this, and the result is a **DNA library**.

Let's say that we want to construct a library of the human genome. A diploid human cell (a cell that has a set of chromosomes from both parents) has six billion base pairs. If we consider that 20,000 base pairs is a reasonable size for a cloned insert, we need a minimum of 300,000 different recombinant DNAs. It is quite a feat to achieve this number of different recombinants, and, in practice, we would need several times this minimum to ensure full representation and to account for the vector molecules that do not acquire an insert. For the purposes of this discussion, we shall use a bacterial plasmid as the vector to construct a suitable number of recombinant DNA molecules by the methods described in Section 13.3 (Figure 13.20).



■ **FIGURE 13.20** Steps involved in the construction of a DNA library. All the DNA of a given organism is extracted and treated with a restriction enzyme. The DNA fragments are incorporated into bacterial plasmids. Specific clones can be selected. The remaining clones are saved for future use. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)



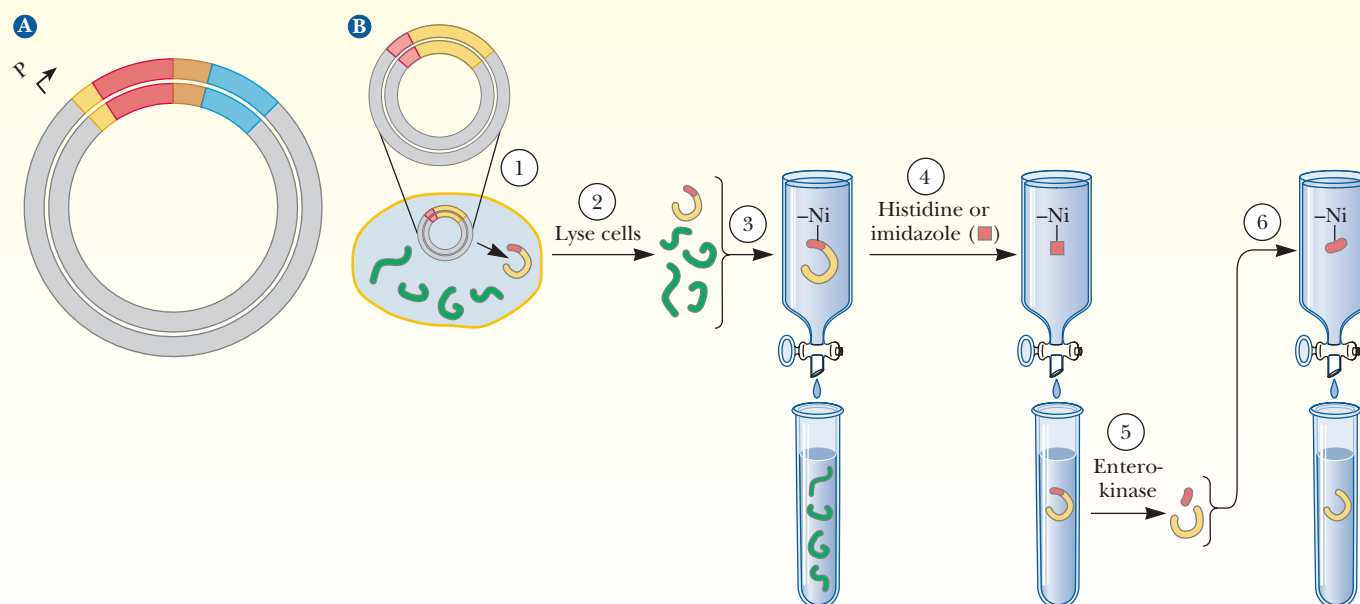
# Biochemical Connections

## ANALYTICAL CHEMISTRY (CHROMATOGRAPHY)

### Fusion Proteins and Fast Purifications

Affinity chromatography was introduced in Chapter 5 as an example of a powerful technique for protein purification. Molecular biologists have taken the idea one step further and have incorporated affinity chromatography ligand-binding sites directly into a protein to be expressed. A protein is created that contains not only the amino acid sequence of the desired polypeptide but also some extra amino acids at the N-terminus or C-terminus. These new proteins are called **fusion proteins**. The figure indicates how this might work. An expression vector that has a promoter for T7 polymerase, followed by a start sequence ATG, is used. A *his-tag* sequence that follows the ATG codes for six histidine residues. Following the his-tag is a sequence that is specific for a proteolytic enzyme called *enterokinase*. Finally comes the MCS, where the gene of interest can be cloned. After the desired gene is cloned

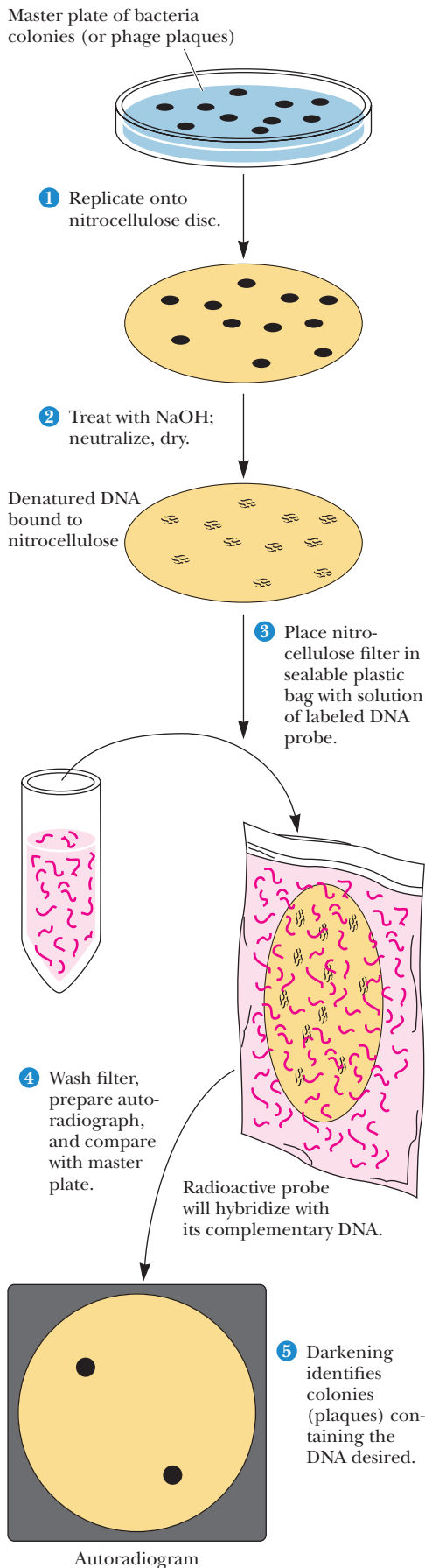
into this vector, it is transformed into bacteria and expressed. The fusion proteins that are translated have the initial methionine, six histidines, the enterokinase-specific amino acid sequence, and then the desired protein. Remember from Chapter 5 that an affinity chromatography resin uses a ligand that binds specifically to a protein of interest. This technique is used with a his-tagged fusion protein. A nickel affinity column is set up, which is very specific for histidine residues. The cells are lysed and passed over the column. All the proteins pass through except the fusion protein, which binds tightly to the nickel column. The fusion protein can be eluted with imidazole, a histidine analog. Enterokinase is then added to cleave off the his-tag, leaving the desired protein. Under ideal circumstances, this can be an almost perfect, one-step purification of the protein.



■ **Affinity chromatography and fusion proteins combine to purify a protein efficiently.** (a) An expression vector is used that inserts six histidines at the N-terminus just after the ATG start site. Following the histidines is a sequence that codes for an enterokinase site, after which is the protein you are trying to purify. (b) In Step 1, the plasmid is cloned and expressed to form cell proteins, including the fusion protein (red and yellow). In Step 2, the cells are lysed and, in Step 3, the lysate is run over a nickel affinity column. Nickel attracts the histidine tag on the fusion protein, and the other cell components are washed out. In Step 4, high concentrations of histidine or imidazole (red rectangles) are used to elute the fusion protein from the nickel affinity column. In Step 5, enterokinase is then used to cut out the histidines, leaving (in Step 6) the desired protein. (Adapted from Molecular Biology, by R. F. Weaver, McGraw-Hill, 1999.)

The next step is to separate the individual members of the population of plasmid DNA molecules by cloning them. The group of clones that has acquired a plasmid with an insert constitutes the recombinant library. The whole library can be stored for future use, or a single clone can be selected for further study. The process of constructing a DNA library can be quite laborious, leading many researchers to obtain previously constructed libraries from other laboratories or from commercial sources. Some journals require that libraries and individual clones that have been discussed in articles they publish be freely available to other laboratories.





## How do we find the piece of DNA we want in a library?

### Finding an Individual Clone in a DNA Library

Imagine that, after a DNA library has been constructed, someone wants to find a single clone (for example, one that contains a gene responsible for an inherited disease) out of the hundreds of thousands, or possibly millions, in the library. This degree of selectivity requires specialized techniques. One of the most useful of these techniques depends on separating and annealing complementary strands. An imprint is taken of the petri dish on which the bacterial colonies (or phage plaques) have grown. A nitrocellulose disc is placed on the dish and then removed. Some of each colony or plaque is transferred to the disc, and *the position of each is the same as it was on the dish*. The rest of the original colonies or plaques remain on the dish and can be stored for future use (Figure 13.21).

The nitrocellulose disc is treated with a denaturing agent to unwind all the DNA on it. (The DNA has become accessible by disruption [lysis] of the bacterial cells or phage.) After it is denatured, the DNA is permanently fixed to the disc by treatment with heat or ultraviolet light. The next step is to expose the disc to a solution that contains a single-stranded DNA (or RNA) probe that has a sequence complementary to one of the strands in the clone of interest (Figure 13.21). The probe anneals to the DNA of interest and only to that DNA. Any excess solution is washed off the nitrocellulose disc. In the case of radioactive probes, the disc is placed in contact with X-ray film (Section 13.1). Only the spots on the disc in which some of the probe has annealed to the DNA already there are radioactive, and only those spots expose the X-ray film. Because the original petri dish has been saved, the desired clone can be picked off the plate and allowed to reproduce.

If the nucleotide sequence of the desired DNA segment is not known and no probe is available, a complication arises. If the gene of interest directs the synthesis of a given protein, one chooses a vector that allows cloned genes to be transcribed and translated. If the presence of the desired protein can be detected by its function, that serves as the basis for detecting it. Alternatively, labeled antibodies can be used as a basis for protein detection.

RNA libraries are not constructed and cloned as such. Rather, the RNA of interest (usually mRNA) is used as the template for synthesis of complementary DNA (cDNA) in a reaction catalyzed by reverse transcriptase. The cDNA is incorporated into a vector (Figure 13.22). Ligating the cDNA to a vector requires the use of a synthetic linker. From this point on, the process of producing a **cDNA library** is virtually identical to that for constructing a genomic DNA

### ■ FIGURE 13.21 Screening a genomic library by colony hybridization (or plaque hybridization).

Host bacteria transformed with a plasmid-based genomic library or infected with a bacteriophage-based genomic library are plated on a petri dish and incubated overnight to allow bacterial colonies (or phage plaques) to form. A replica of the bacterial colonies (or phage plaques) is then obtained by overlaying the plate with a nitrocellulose disc (1). Nitrocellulose strongly binds nucleic acids; single-stranded nucleic acids are bound more tightly than double-stranded nucleic acids. Once the nitrocellulose disc has taken up an impression of the bacterial colonies (or phage plaques), it is removed and the petri dish is set aside and saved. The disc is treated with 2 M NaOH, neutralized, and dried (2). NaOH both lyses any bacteria (or phage particles) and dissociates the DNA strands. When the disc is dried, the DNA strands become immobilized on the filter. The dried disc is placed in a sealable plastic bag, and a solution containing heat-denatured (single-stranded), labeled probe is added (3). The bag is incubated to allow annealing of the probe DNA to any target DNA sequences that might be present on the nitrocellulose. The filter is then washed, dried, and placed on a piece of X-ray film to obtain an autoradiograph (4). The position of any spots on the X-ray film reveals where the labeled probe has hybridized with target DNA (5). The location of these spots can be used to recover the genomic clone from the bacteria (or phage plaques) on the original petri dish.

library. For a given organism, a genomic DNA library is the same no matter what the tissue source, and the DNA represents both expressed and unexpressed DNA. In contrast, a cDNA library is different depending on the tissue used and the expression profile of the cells.

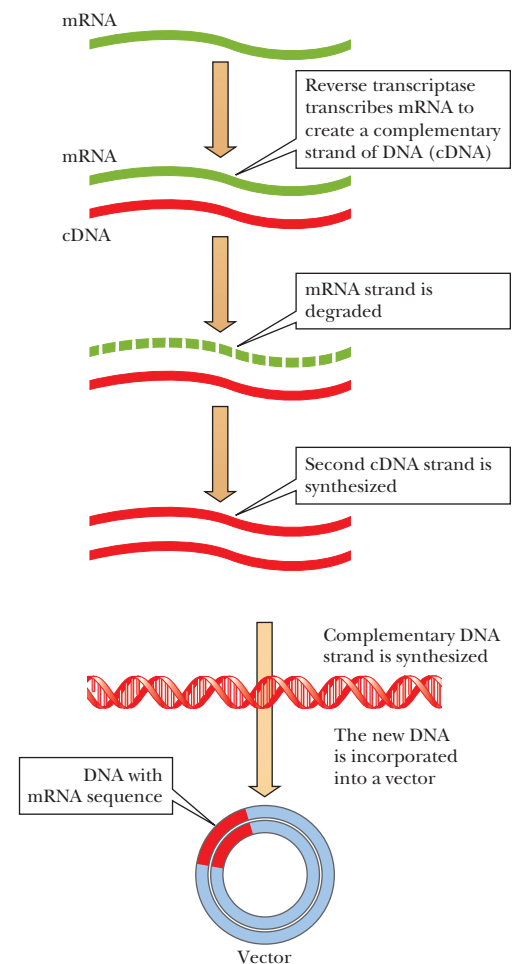
## 13.6 The Polymerase Chain Reaction

It is possible to increase the amount of a given DNA many times over without cloning that DNA. The method that makes this amplification possible is the **polymerase chain reaction (PCR)**. Any chosen DNA can be amplified, and it need not be separated from the rest of the DNA in a sample before the procedure is applied. PCR copies both complementary strands of the desired DNA sequence. Scientists had long wished for a cell-free, automated way of synthesizing DNA, but any system that could be automated would need to function at high temperatures so that the DNA strands could be separated physically without the need for the many enzymes found in DNA replication, such as topoisomerases and helicases. Unfortunately, such temperatures (around 90°C) would denature and inactivate the DNA polymerases that would be making the DNA strands. What made the process possible was the discovery of bacteria that live around deep-sea hydrothermal vents, under extreme pressures, and at temperatures higher than 100°C. If bacteria can live under those conditions, then their enzymes must be able to function at those temperatures. The bacterium, *Thermus aquaticus*, from which a heat resistant polymerase is extracted, is one of those that live in hot springs. The enzyme is called *Taq* polymerase. We saw in Chapter 1 that the biotechnology industry eagerly searches for organisms that live under extreme conditions, and here we have an example of why it does so.

At the start of the process, the two DNA strands are separated by heating, after which short oligonucleotide primers are added in large excess and, via cooling, are allowed to anneal to the DNA strands. These primers are complementary to the ends of the DNA chosen for amplification and serve the same purpose as the RNA primers in normal replication. Once the primers have annealed to the DNA, the temperature is raised again to optimize the activity of *Taq* polymerase, which begins synthesizing the new DNA from the 3' end of the primer. The two complementary strands grow in the 5' to 3' direction (Figure 13.23), and the *Taq* polymerase is allowed to work until the desired length of DNA has been synthesized. This first round doubles the amount of the desired DNA. The process of unwinding the two strands, annealing the primers, and copying the complementary strand is repeated, bringing about a second doubling of the selected double-stranded DNA. It is not necessary to add more primer because it is present in large excess. The whole process is automated. Control of the temperature to which the strands are heated to separate them is crucial, as is the temperature chosen for annealing the primers.

The amount of DNA continues to double in subsequent rounds of amplification. After about an hour, and 25 to 40 cycles of replication, one obtains millions to hundreds of millions of copies of the desired DNA segment, usually a few hundred to a few thousand base pairs long (Figure 13.23). Other DNA sequences are not amplified and do not interfere with the reaction or subsequent use of the amplified DNA.

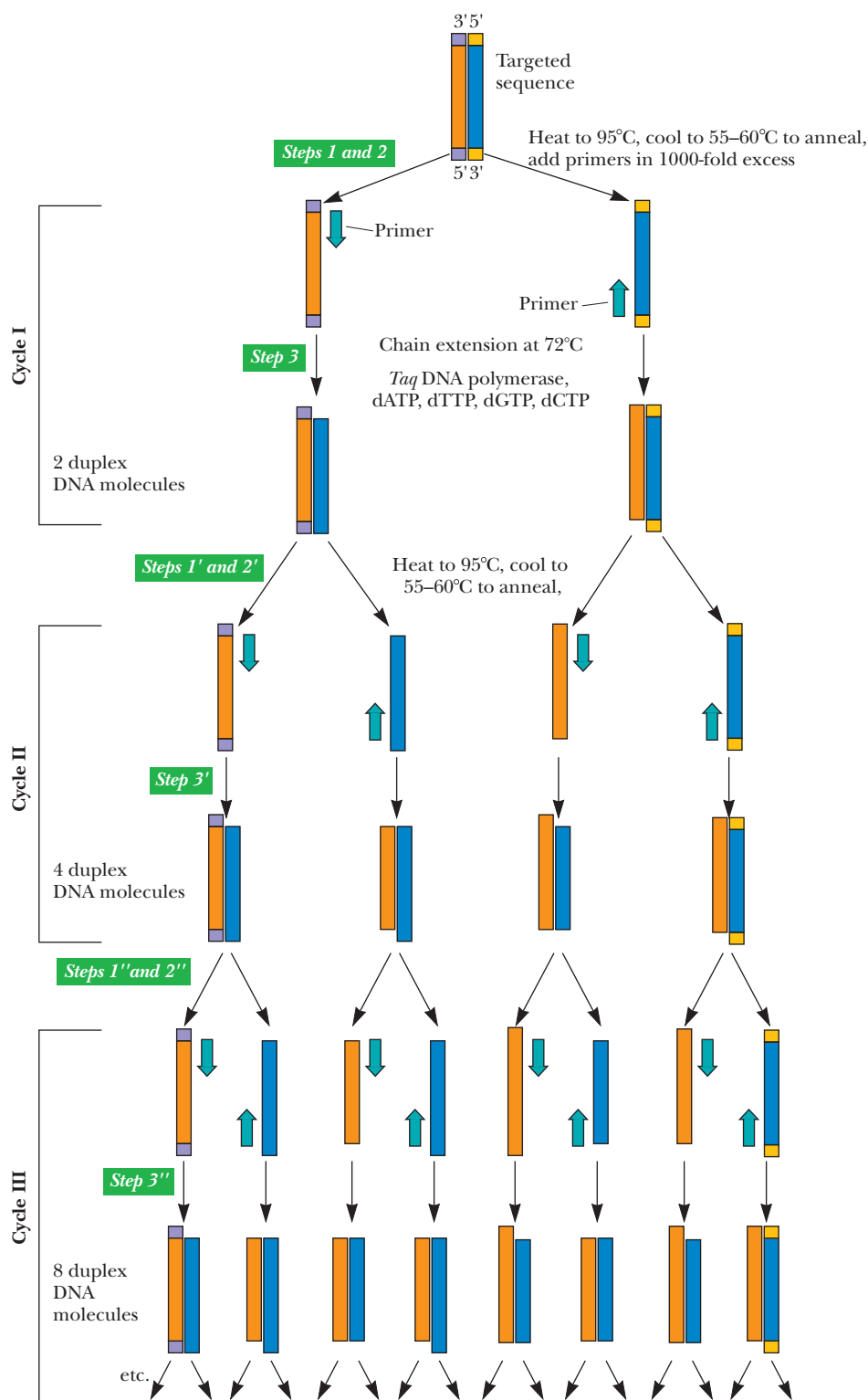
The most important part of the science behind PCR is the design of the primers. They have to be sufficiently long to be specific for the target sequence but not so long that they are too expensive. Usually, the primers are 18 to 30 bases long. They must also have optimal binding properties, such as an amount of G and C that is sufficient to allow them to anneal before the entire DNA re-natures. In addition, the two primers should contain similar amounts of G and C so that they have the same melting temperature. The sequence of the primer



**FIGURE 13.22 Formation of cDNA.** Reverse transcriptase catalyzes the synthesis of a strand of complementary DNA (cDNA) on a template of mRNA. The cDNA directs the synthesis of a second strand, which is then incorporated into a vector. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)



**Kary B. Mullis, inventor of the polymerase chain reaction and 1993 Nobel laureate in chemistry.**



■ **FIGURE 13.23 Polymerase chain reaction (PCR).** Oligonucleotides complementary to a given DNA sequence prime the synthesis of only that sequence. Heat-stable *Taq* DNA polymerase survives many cycles of heating. Theoretically, the amount of the specific primed sequence is doubled in each cycle.

must not lead to secondary structures within a primer or between the two different primers; otherwise, the primers bind to themselves instead of to the DNA being amplified. For example, if a primer had the sequence AAAAATTTTT, it would form a hairpin loop with itself and would not be available to bind the DNA. Section 13.7 explains how the primer sequence can be controlled to change the DNA that is being amplified.

### What are the advantages of PCR?

Amplification of the amounts of DNA in extremely small samples has made it possible to obtain accurate analyses that were not possible earlier. Forensic applications of the technique have resulted in positive identifications of crime victims and suspects. Even minuscule amounts of ancient DNA, such as those available from Egyptian mummies, can now be researched after amplification. The following Biochemical Connections box describes some forensic uses of DNA technology.

### Quantitative PCR allows sensitive measurement of DNA samples

One of the most recent innovations in biotechnology expands on the PCR technique to give quantitative results. Standard PCR is used for producing large quantities of DNA, so the point of the experiment is to let the assay run to its conclusion and harvest the DNA of interest. A newer technique is called **Quantitative PCR (qPCR)**. This allows the PCR reaction to generate time-point data

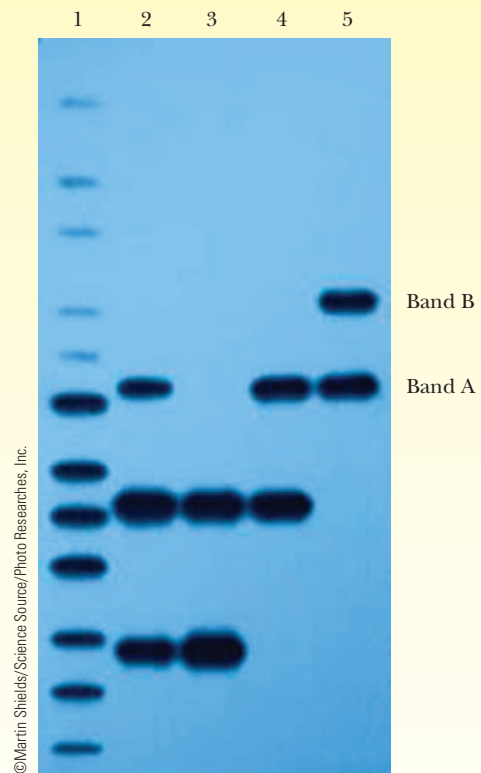
## Biochemical Connections FORENSICS

### CSI: Biochemistry—Forensic Uses of DNA Testing

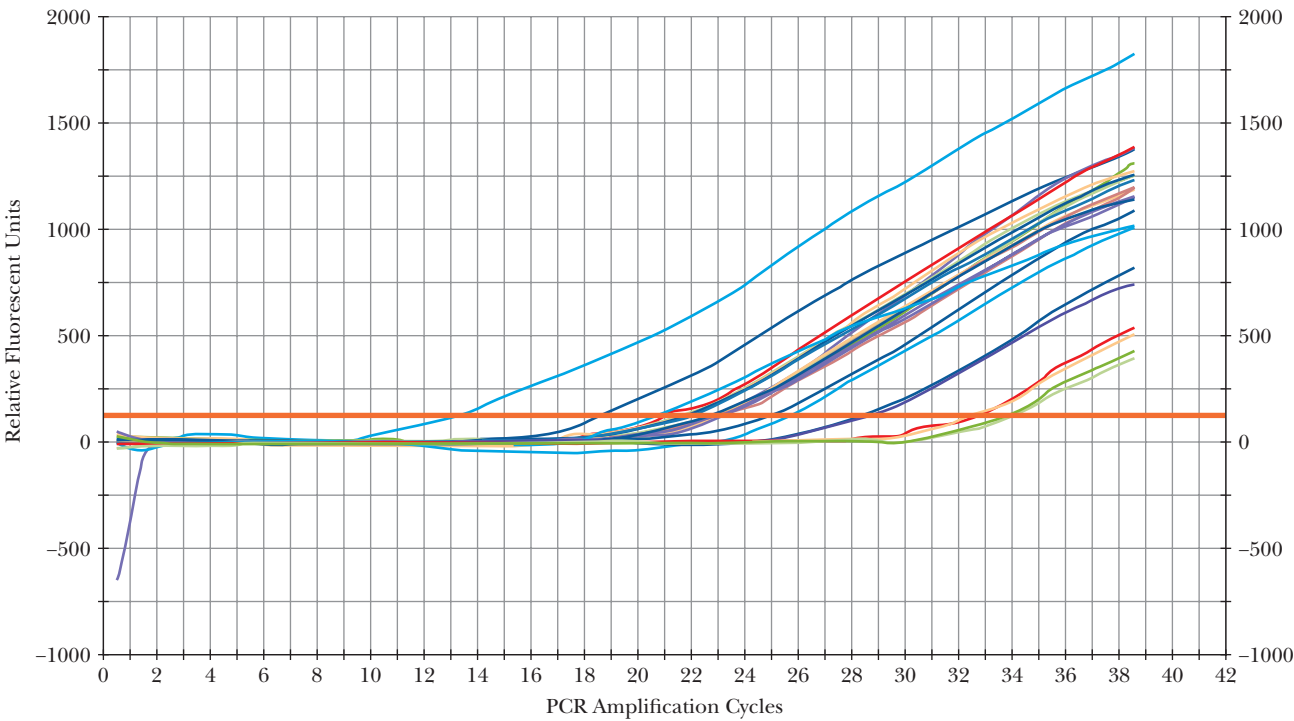
It has been suggested that tissue samples be taken from all convicted felons to allow for identification of their presence at future crime sites by DNA fingerprinting (Section 13.7). This may seem very desirable, but some ethical and constitutional questions must first be answered. Nevertheless, DNA testing has been used to determine whether people currently in custody might have been involved in unsolved crimes. One crime frequently examined in this way is rape, because body fluids are usually left behind. Many unsolved crimes are being solved, and some convicted prisoners are being found innocent. In at least one bizarre case, a felon was found innocent of the crime for which he was in jail, only to be rearrested in a few weeks because his DNA matched samples found on rape victims in three other cases.

The power of DNA testing cannot be overemphasized. In the trial of O. J. Simpson, the identity of the DNA was never questioned, although the defense successfully raised questions as to whether the evidence had been planted. In most cases, DNA evidence results in the release of innocent suspects in a crime. When the DNA matches, however, the rate of plea bargaining goes up, and the crimes are solved without long and expensive court trials.

Establishing paternity is a natural application for DNA testing. The DNA markers found in a child must arise from either the mother or the father. A man who is suspected of being the father of a particular child is immediately excluded from consideration if that child's DNA contains markers that are not found in the DNA of the mother or the suspect father. It is more difficult to prove conclusively that a person is the father of a particular child because it is usually too expensive to test enough markers to provide conclusive proof. However, in cases in which there are only two or three candidate fathers, it is usually possible to determine the correct one.



■ **Paternity Test DNA fingerprint.** This gel shows possible results of a DNA fingerprint from a paternity test. Lane 1 shows a DNA size ladder used as a control. Lane 2 is a DNA pattern from a mother. Lanes 3 and 4 are DNA patterns from possible fathers. Lane 5 is the DNA pattern from the child. Every band in the child's DNA must have a corresponding band in either the mother or the father's DNA. Band A in the child's DNA matches a band in the mother's DNA and in the prospective father's DNA in lane 4. However, band B in the child's DNA is not seen in the mother's DNA, nor is it seen in either of the prospective fathers' DNA. This excludes both possible fathers from being the father of the child.

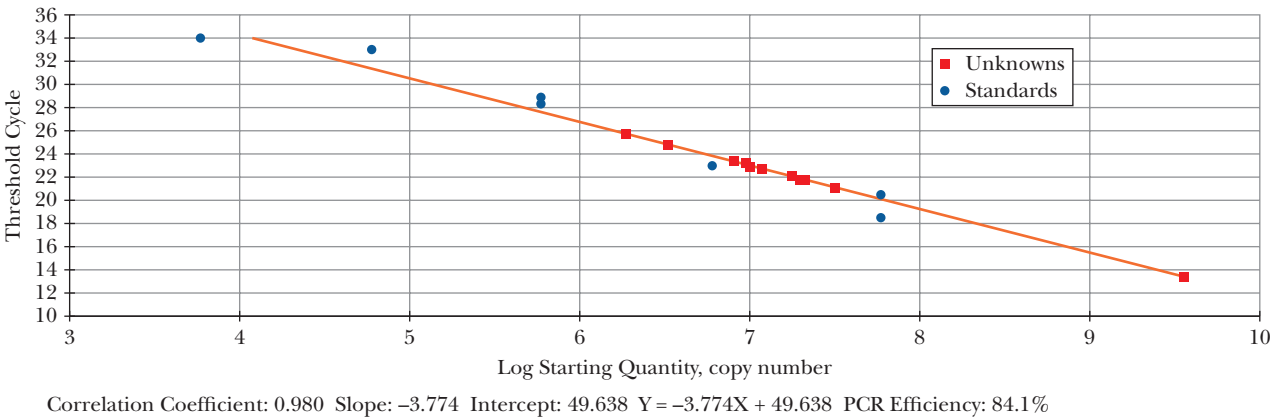


■ **FIGURE 13.24** Quantitative PCR. Cycles with different amounts of initial target DNA are shown as different colors.

that can be used to determine how much of the DNA was in the cell originally. Figure 13.24 shows how this works.

PCR is run on multiple samples containing the DNA sequence of interest. As the number of cycles increases, a threshold value is reached where the amount of DNA can be measured. For this purpose, the DNA is labeled with fluorescent markers. The more DNA in the sample, the earlier in the process the results can be seen (shown by the different colored lines).

To get the quantity of DNA, known standards are run along with unknowns. The time it took for a given sample to reach a defined value (shown as the horizontal orange line), can then be plotted on a standard curve, as shown in Figure 13.25.

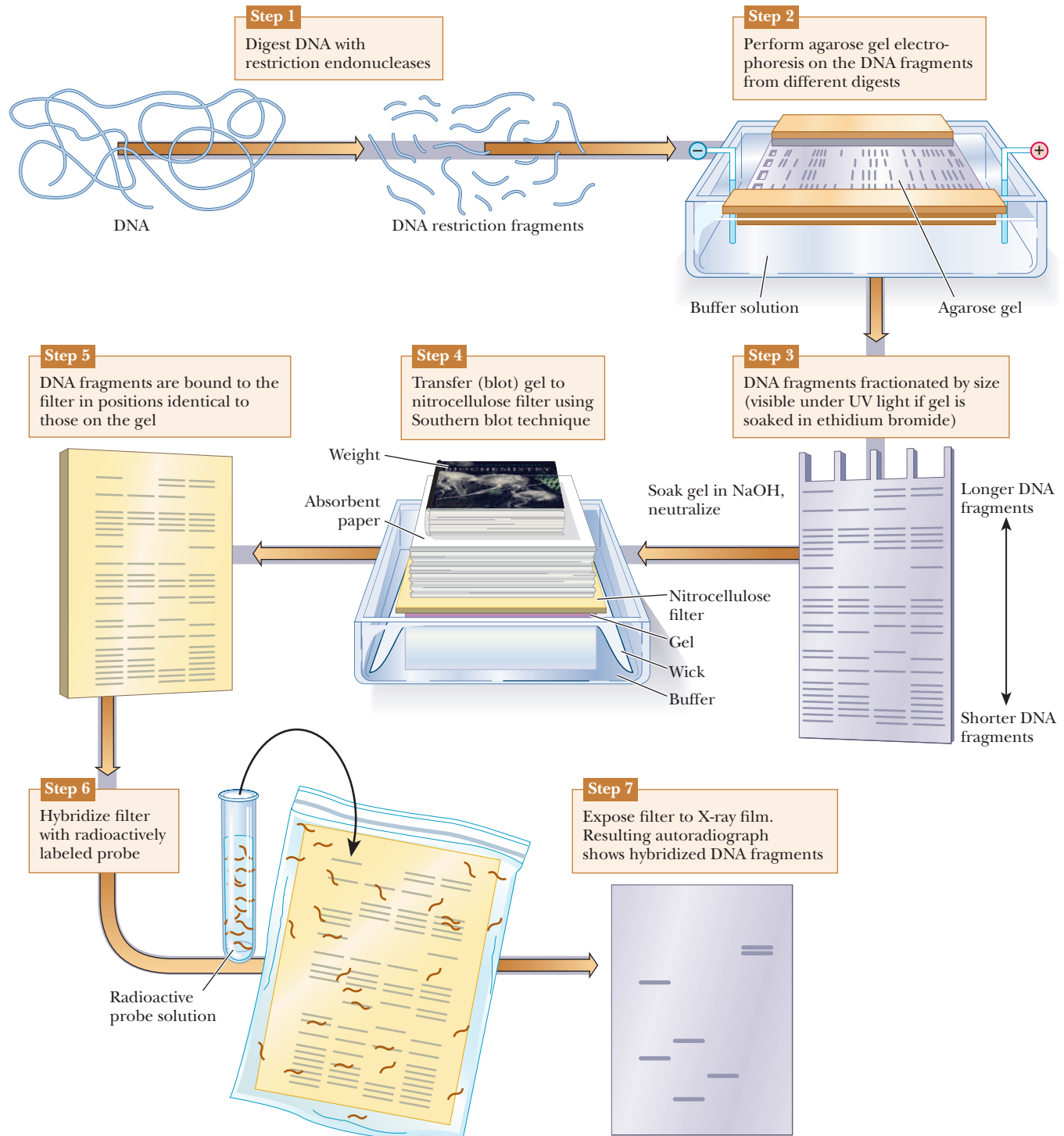


■ **FIGURE 13.25** Standard curve created from qPCR.



## 13.7 DNA Fingerprinting

DNA samples can be studied and compared using a technique called *DNA fingerprinting*. The DNA is digested with restriction enzymes and then run on an agarose gel (Figure 13.26). The DNA fragments can be seen directly on the gel if it is soaked in ethidium bromide and viewed under ultraviolet light (Section 13.1). As shown in Figure 13.26, Step 3, this gives bands of varying sizes, depending

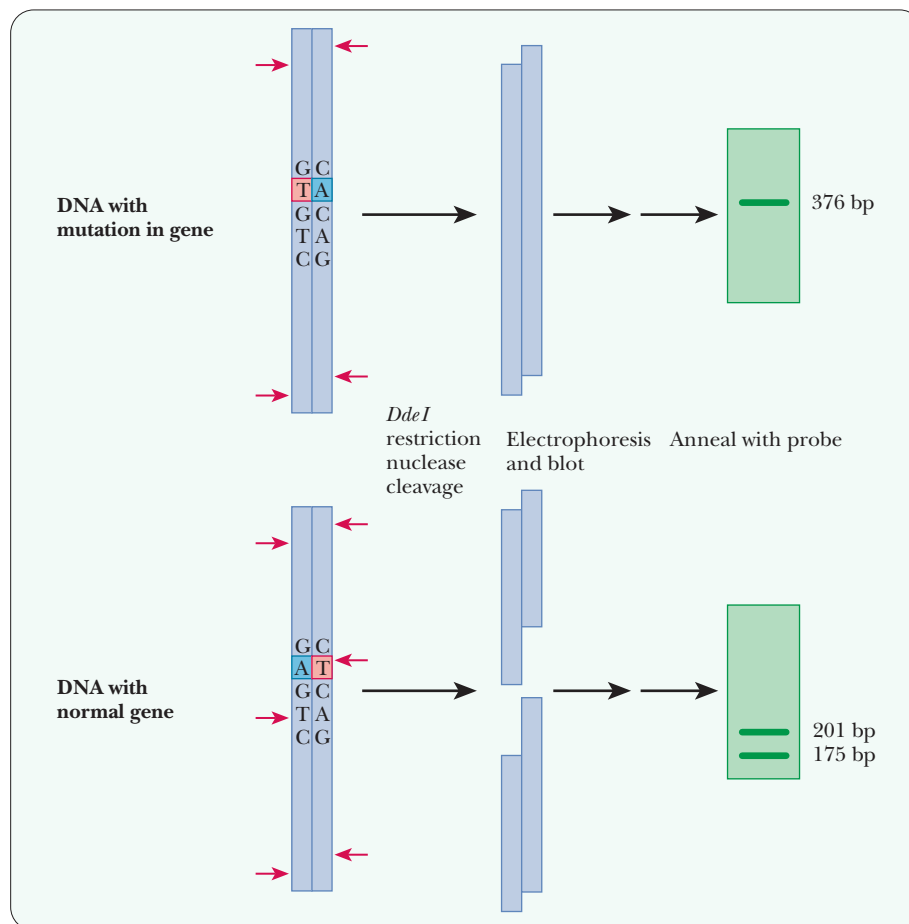


■ **FIGURE 13.26** The Southern blot. Electrophoretically separated DNA fragments are transferred to a nitrocellulose sheet. A radioactively labeled probe for a DNA sequence of interest is bound to the nitrocellulose, and bands are visualized with an autoradiograph.

on the nature of the DNA and the restriction enzymes used. If greater sensitivity is needed, or if the number of fragments would be too great to distinguish the bands, this technique can be modified to visualize only selected DNA sequences. The first step would be to transfer the DNA to a nitrocellulose membrane in a procedure called a **Southern blot**, after its inventor, E. M. Southern. The agarose gel is soaked in NaOH to denature the DNA because only single-stranded DNA binds to the nitrocellulose. The membrane is placed on the agarose gel, which is on top of a filter-paper wick placed in buffer. Dry absorbent paper is placed on top of the nitrocellulose. Wicking action carries the buffer from the buffer chamber up through the gel and nitrocellulose into the dry paper. The DNA bands move out of the gel and stick to the nitrocellulose. The next step is to visualize the bands on the nitrocellulose. A specific DNA probe that is labeled with  $^{32}\text{P}$  is incubated with the nitrocellulose membrane. The DNA probe binds to DNA fragments that are complementary. The membrane is then placed on photographic paper to produce an autoradiograph (Section 13.1). The use of the specific probe greatly reduces the number of bands seen and isolates desired DNA sequences.

### Restriction-Fragment Length Polymorphisms: A Powerful Method for Forensic Analysis

In organisms (such as humans) with two sets of chromosomes, a given gene on one chromosome may differ slightly from the corresponding gene on the paired chromosome. In the language of genetics, these genes are **alleles**. When they are the same on the paired chromosomes, the organism is **homozygous** for that gene; when they differ, the organism is **heterozygous**. A difference between alleles, even a change in one base pair, can mean that one allele has a recognition site for a restriction endonuclease and the other does not. Restriction fragments of different sizes are obtained on treatment with the endonuclease (Figure 13.27); they



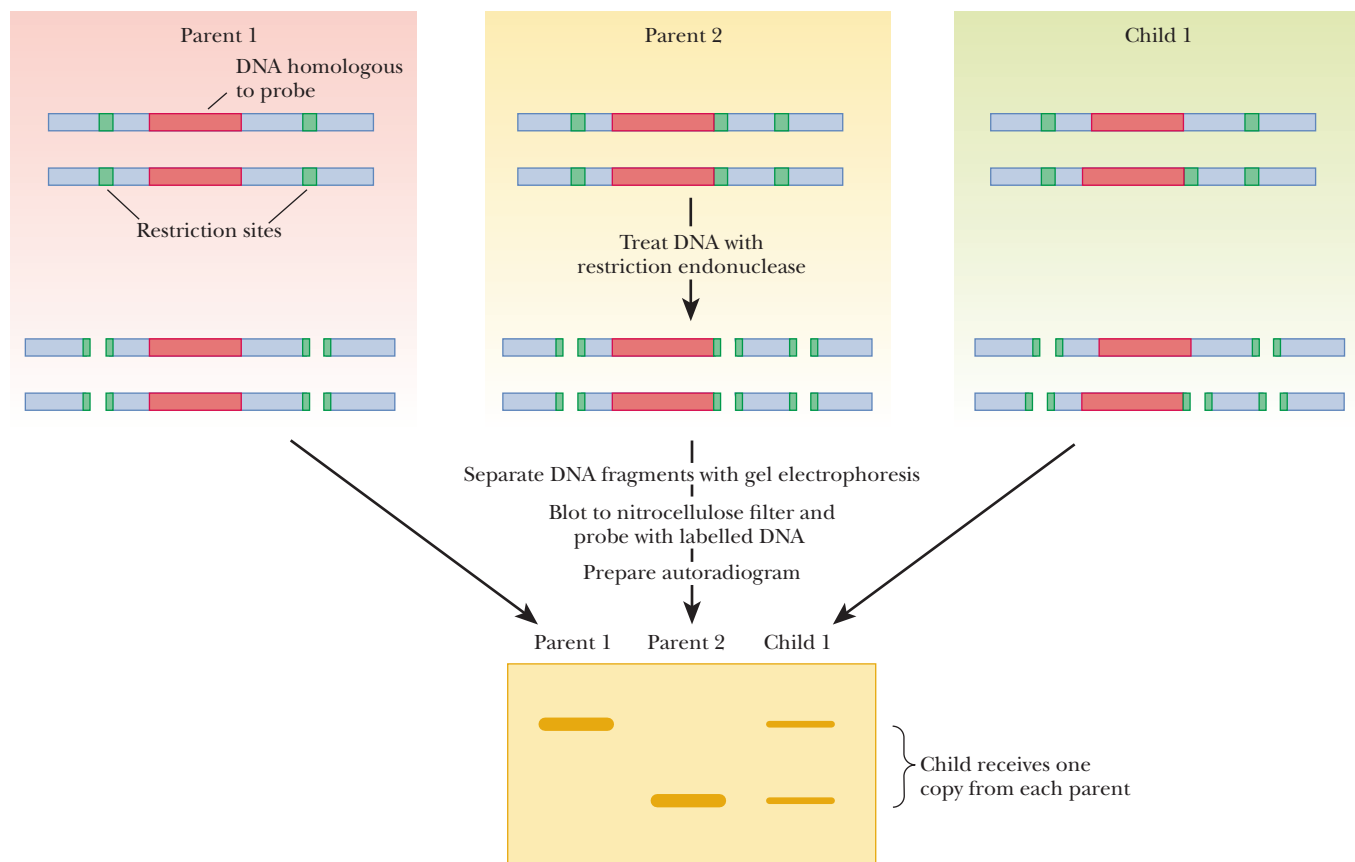
■ **FIGURE 13.27** The basis for restriction-fragment length polymorphism. A change of one base pair eliminates a restriction nuclease cleavage site. A portion of DNA that codes for a protein has a cleavage site for the restriction nuclease *DdeI*. The corresponding DNA with a mutation does not have this cleavage site. The difference can be detected by electrophoresis, followed by blotting and the annealing of a probe specific for this fragment. (The abbreviation *bp* stands for *base pairs*.) (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)

are called **restriction-fragment length polymorphisms**, or **RFLPs** (pronounced “riflips”) for short.

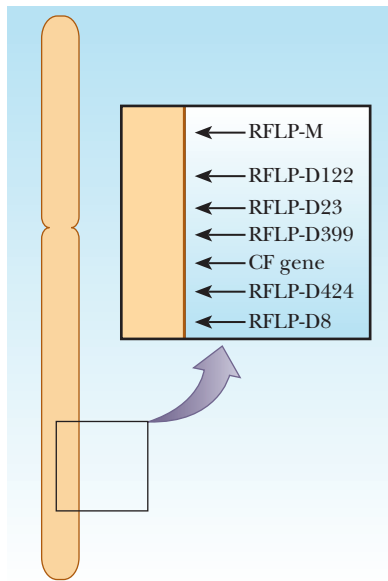
These *polymorphisms* (a word meaning “many shapes”) are analyzed via gel electrophoresis to separate the fragments by size, followed by blotting and the annealing of a probe for a specific sequence.

### How can differences in DNA from individuals be seen?

Research has shown that polymorphisms are quite common, much more so than the mutations in traits such as eye color and inherited diseases that were used for earlier genetic mapping. RFLPs can be used as markers for heredity in the same way as mutations in visible traits because they are inherited in the manner predicted by classical genetics (Figure 13.28). However, because they are much more abundant than mutations that lead to phenotypic variations, they have provided many more markers for detailed genetic mapping. Figure 13.28 shows how RFLP analysis can be used in paternity testing. A child gets one of each allele from each parent, so every fragment that the child has must also be present in one of the parents. Thus, it would be easy to eliminate a prospective father undergoing a test for paternity; if the child had an RFLP band that neither the mother nor the prospective father has, then the prospective father is excluded. The same kinds of analyses are done with evidence found at crime scenes. Suspects in criminal cases may be exonerated if their DNA



■ **FIGURE 13.28 Paternity testing.** Restriction-fragment length polymorphisms (RFLPs) can be detected by probes for homologous DNA. Parent 1 has DNA with two restriction sites near the DNA homologous to the probe. Because Parent 1 has the same pattern of restriction sites on both DNA strands, one large restriction fragment is detected. Parent 2 has three restriction sites on each DNA strand. A single smaller restriction fragment is detected by the homologous probe for Parent 2. Their child has inherited one copy of each DNA strand from each parent. The child's DNA produces one fragment of each size. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)



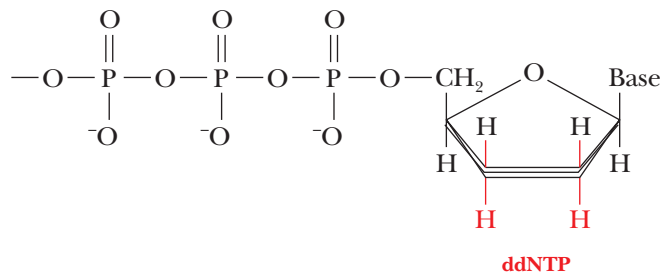
■ **FIGURE 13.29** Localization of the gene associated with cystic fibrosis (CF) on human chromosome 7. The RFLP markers are given arbitrary names. The location of the CF gene is given relative to the RFLP markers. (Adapted with permission from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)

samples do not match those found at the scene. RFLP analysis was used extensively in the process of locating the altered gene that causes cystic fibrosis, a prevalent genetic disease. Once the gene was located on chromosome 7, a series of RFLP markers was used to help map its exact position (Figure 13.29). Then the gene was isolated from restriction endonuclease digests and cloned, and its protein product was then characterized. The protein in question is involved in the transport of chloride ion ( $\text{Cl}^-$ ) through membranes. If this protein is defective, chloride ions remain in the cells and take up water by osmosis from the surrounding mucus. The mucus thickens as a result. In the lungs, the thickened mucus favors infections, particularly pneumonia. The results of this disease can be tragic, leading to a short life span in those who are affected by it. This information deepens our insight into the nature of cystic fibrosis and provides approaches to new treatment.

## 13.8 Sequencing DNA

We have already seen that the primary structure of a protein determines its secondary and tertiary structures. The same is true of nucleic acids; the nature and order of monomer units determine the properties of the whole molecule. Base pairing in both RNA and DNA depends on a series of complementary bases, whether these bases are on different polynucleotide strands, as in DNA, or on the same strand, as is frequently the case in RNA. Sequencing of nucleic acids is now fairly routine, and this relative ease would have amazed the scientists of the 1950s and 1960s.

The method devised by Sanger and Coulson for determining the base sequences of nucleic acids depends on selective interruption of oligonucleotide synthesis. A single-stranded DNA fragment whose sequence is to be determined is used as a template for the synthesis of a complementary strand. The new strand grows from the 5' end to the 3' end. This unique direction of growth is true for all nucleic acid synthesis (Chapter 10). The synthesis is interrupted at every possible site in a population of molecules. The interruption of synthesis depends on the presence of 2',3'-dideoxynucleoside triphosphates (ddNTPs).



The 3'-hydroxyl group of deoxyribonucleoside triphosphates (the usual monomer unit for DNA synthesis) has been replaced by a hydrogen.

### How do dideoxynucleotides allow us to sequence DNA?

These ddNTPs can be incorporated in a growing DNA chain, but they lack a 3'-hydroxyl group to form a bond to another nucleoside triphosphate. The incorporation of a ddNTP into the growing chain causes termination at that point. The presence of small amounts of ddNTPs in a replicating mixture causes random termination of chain growth.

# Biochemical Connections MOLECULAR BIOLOGY

## RNA Interference—The Newest Way to Study Genes

RNA interference (RNAi) was first discovered in a nematode worm (*Caenorhabditis elegans*). Double-stranded RNA (dsRNA) was found to cause gene silencing in a sequence-specific way.

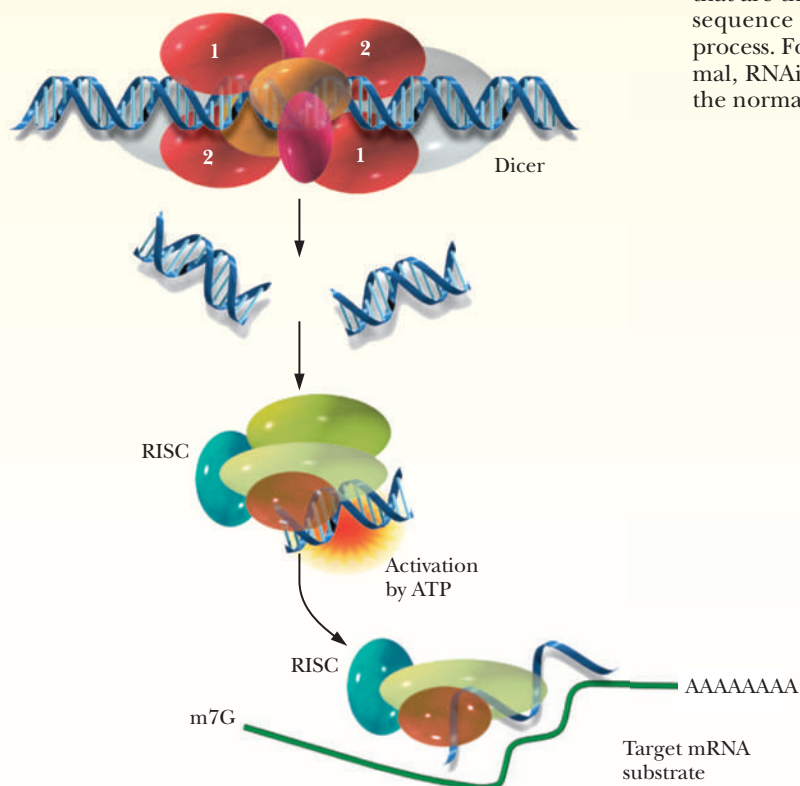
Researchers had long thought that RNA would be the perfect way to control gene expression, because the right sequence of RNA should bind to DNA and interfere with its transcription. While checking on the efficiency of antisense RNA (Section 11.1) as a suppressor of gene expression, researchers discovered that dsRNA was more than 10 times as effective at shutting off transcription of a gene. In addition, suppression by RNAi was shown to be transmissible to other cells of the organism and to the progeny in *C. elegans*. RNAi has since been found as a natural phenomenon in plants as a way of targeting viral RNAs for destruction. It has also been seen in *Drosophila* and, most recently, in mammals. It is now believed that RNAi is a natural regulatory mechanism for controlling gene expression. It may also be a protection mechanism against oncogenes that produce too much of a harmful product.

The process starts with an enzyme that is a member of the RNase III class, called **Dicer**. This enzyme binds to dsRNA and cleaves it into small interfering RNA (siRNA) of between 22 and 25 nucleotides, as shown in the figure. The siRNAs then bind to a protein complex called **RNA-induced silencing complex (RISC)**. The siRNA–RISC complex then binds to target mRNA that has the same sequence as part of the siRNA and degrades it. In this way, the sequence of the dsRNA controls the degradation of an mRNA target. We are still learning how this may have

arisen naturally. Perhaps an endogenous RNA-dependent RNA polymerase senses that too much of a particular mRNA is being made. It could then create the opposite strand, thereby forming the dsRNA. This would then trigger the eventual destruction of the mRNA. The same logic could apply to a defense mechanism against viruses.

Regardless of RNAi's natural purpose, it has become the fastest-growing new field in molecular biology. Many companies have sprung up overnight that produce RNAi kits and dsRNAs to use to initiate reactions. The technique is quickly becoming the newest way to knock out specific genes to see what then happens to the organism. All it takes is a knowledge of the gene sequence, and the correct dsRNA can then be used to produce siRNAs to shut off the gene. Researchers using RNAi have been better able to map the thousands of genes in certain organisms, such as *C. elegans*. As an example of the power of the technique, researchers just mapped the gene for an enzyme involved in the metabolism of vitamin K. This enzyme, **vitamin K epoxide reductase (VKOR)**, is the target of a widely prescribed anticoagulant. The enzyme has been known for 40 years, but attempts to purify it and locate it on the human chromosome had met with no success. Using information from the Human Genome Project, researchers were able to map human chromosomes and to come up with 13 genes that had characteristics that could mean that they were the VKOR gene. Using RNAi, the correct one was isolated in a few weeks, a major breakthrough for researchers that had been studying this enzyme for decades.

Medical researchers are very hopeful about the possibilities for this technique. Given the specificity of RNAi, mutant alleles that are the basis of diseases could be knocked out with the right sequence of dsRNA given as a trigger for the Dicer enzyme to process. For example, if one allele is mutated but the other is normal, RNAi could knock out the mutated one while not affecting the normal one.

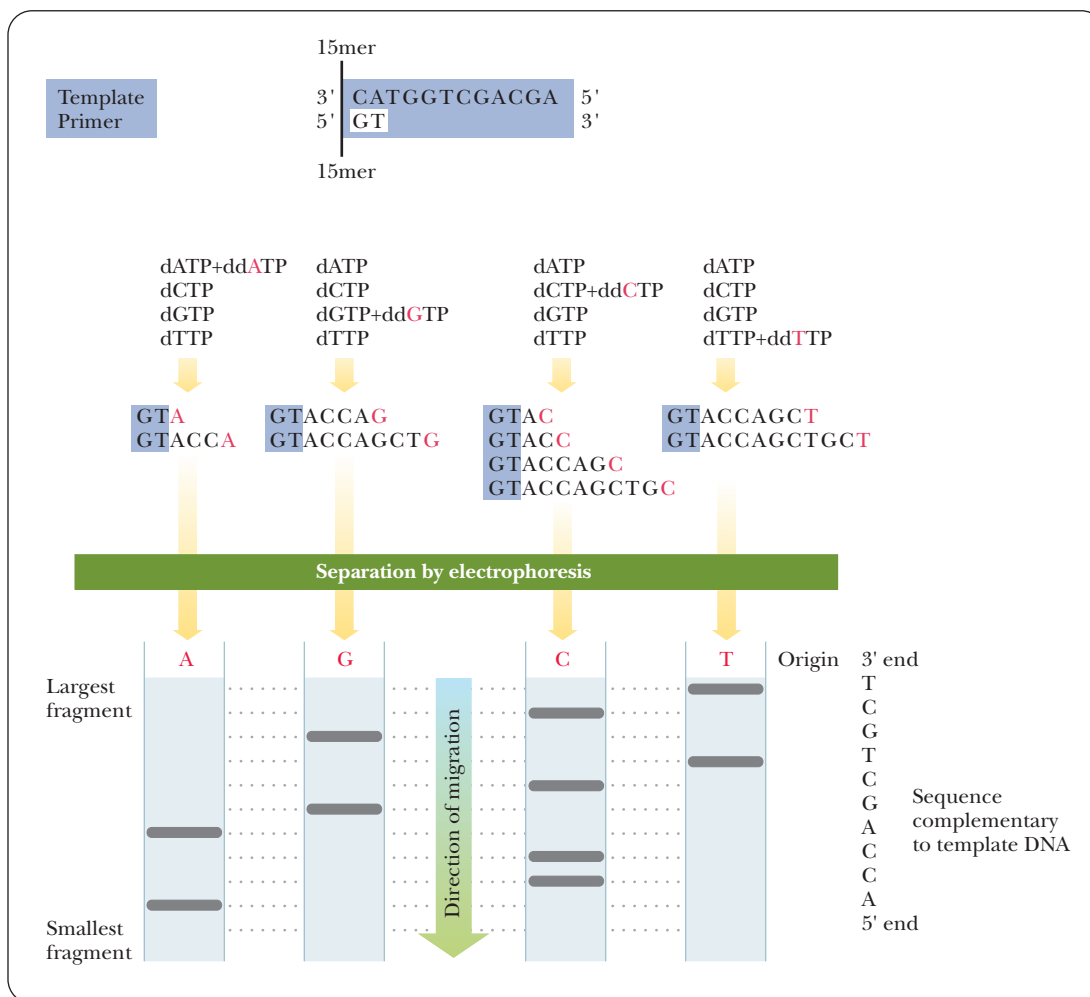


■ **RNA interference.** An enzyme complex called Dicer cleaves double-stranded RNA into small interfering RNAs (siRNA). These then bind to a protein complex called RNA-induced silencing complex (RISC). RISC unwinds the siRNA and one strand binds to the complementary mRNA, which is subsequently degraded.

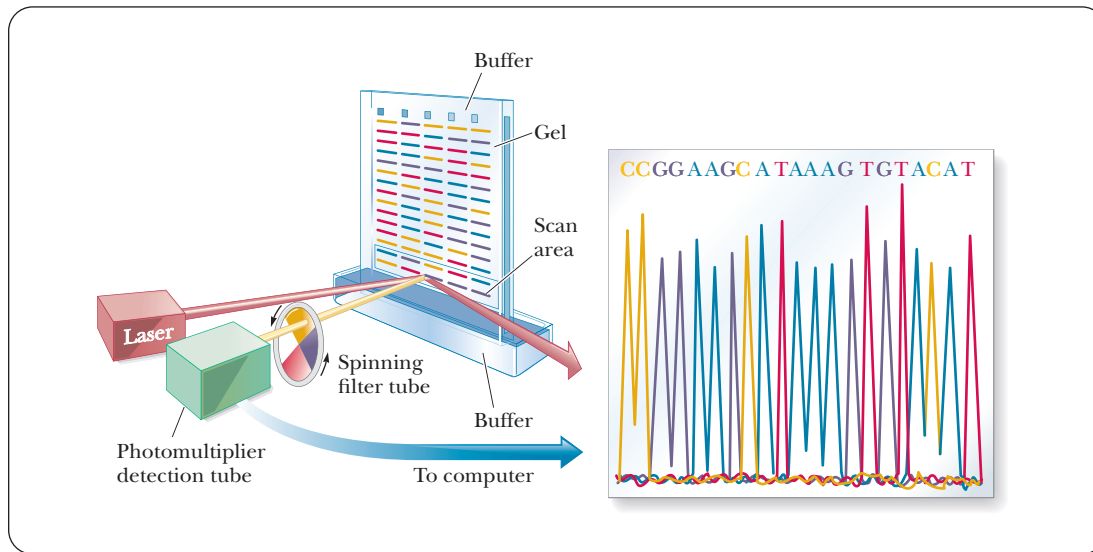


The DNA to be sequenced is mixed with a short oligonucleotide that serves as a primer for synthesis of the complementary strand. The primer is hydrogen-bonded toward the 3' end of the DNA to be sequenced. The DNA with primer is divided into four separate reaction mixtures. Each reaction mixture contains all four deoxyribonucleoside triphosphates (dNTPs), one of which is labeled to allow the newly synthesized fragments to be visualized by autoradiography or by fluorescence, as described in Section 13.1. In addition, each of the reaction mixtures contains one of the four ddNTPs. Synthesis of the chain is allowed to proceed in each of the four reaction mixtures. In each mixture, chain termination occurs at all possible sites for that nucleotide.

When gel electrophoresis is performed on each reaction mixture, a band corresponding to each position of chain termination appears. The sequence of the newly formed strand, which is complementary to that of the template DNA, can be “read” directly from the sequencing gel (Figure 13.30). A variation on this method is to use a single reaction mixture with a different fluorescent label on each of the four ddNTPs. Each fluorescent label can be detected by its characteristic spectrum, requiring only a single gel electrophoresis experiment. The use of fluorescent labels makes it possible to automate DNA sequencing,



**FIGURE 13.30** The Sanger–Coulson method for sequencing DNA. A primer at least 15 residues long is hydrogen-bonded to the 3' end of the DNA to be sequenced. Four reaction mixtures are prepared; each contains the four dNTPs and one of the four possible ddNTPs. In each reaction mixture, synthesis takes place, but in a given population of molecules, synthesis is interrupted at every possible site. A mixture of oligonucleotides of varying lengths is produced. The components of the mixture are separated by gel electrophoresis.



■ **FIGURE 13.31** Fluorescent labeling and automated sequencing of DNA. Four reactions are set up, one for each base, and the primer in each is end labeled with one of four different fluorescent dyes; the dyes color-code the base-specific sequencing protocol (a unique dye is used in each dideoxy nucleotide reaction). The four reaction mixtures are then combined and run in one lane. Thus, each lane in the gel represents a different sequencing experiment. As the differently sized fragments pass down the gel, a laser beam excites the dye in the scan area. The emitted energy passes through a rotating color filter and is detected by a fluorometer. The color of the emitted light identifies the final base in the fragment. (Applied Biosystems, Inc., Foster City, CA.)

with the whole process under computer control. Commercial kits are available for these sequencing methods (Figure 13.31).

When RNA is to be sequenced, the method of choice is not to analyze the RNA itself but to use the methods of DNA sequencing on a DNA complementary (cDNA) to the RNA in question. The cDNA, in turn, is generated by using the enzyme reverse transcriptase, which catalyzes the synthesis of DNA from an RNA template.

## 13.9 Genomics and Proteomics

With more and more full DNA sequences becoming available, it is tempting to compare those sequences to see whether patterns emerge from genes that encode proteins with similar functions. The amount of data makes use of a computer essential for the process. Databases on genome and protein sequences are so extensive as to require information technology at its best to solve problems. Knowing the full DNA sequence of the human genome, for example, allows us to address the causes of disease in a way that was not possible until now. That prospect was one of the main incentives for undertaking the Human Genome Project. The website of the National Human Genome Research Institute, which is a part of the National Institutes of Health (NIH) has useful information; the URL for this site is <http://www.genome.gov>.

A number of genomes are available online, along with software for sequence comparisons. An example is the material available from the Sanger Institute (<http://www.sanger.ac.uk>). In November 2003, the researchers at this institute announced that they had sequenced 2 billion bases from the DNA of several organisms (human, mouse, zebrafish, yeasts, and the roundworm *Caenorhabditis elegans*, among others). If this amount of DNA were the size of a spiral staircase, it would reach from the Earth to the Moon.

A question implicit in the determination of the genome of any organism is that of assignment of sequences to the chromosome in which they belong. This is a challenging task, and only suitable computer algorithms make it possible. Once

this has been achieved, one can compare genomes to see what changes have occurred in the DNA of complex organisms compared with those of simpler ones.

Beyond this application, challenging though it may be, lies the application to medicine, which is leading to a number of surprises. Two closely related genes (*BRCA1* and *BRCA2*) involved in the development of breast cancer interact with other genes and proteins, and this is a topic of feverish research. The connection between these genes and a number of seemingly unrelated cancers is only starting to be unraveled. Clearly, there is need to determine not only the genetic blueprint but the manner in which an organism puts it into action.

The **proteome** is the protein version of the genome. In all organisms for which sequence information is available, **proteomics** (the study of interactions among all the proteins in a cell) is assuming an important place in the life sciences. If the genome is the script, the proteome puts the play on stage. The genetically determined amino acid sequence of proteins determines their structure and how they interact with each other. Those interactions determine how they behave in a living organism. The potential medical applications of the human proteome are apparent, but these have not yet been realized. Proteomic information does exist, for eukaryotes such as yeast and the fruit fly *Drosophila melanogaster*, and the methods that have been developed for those experiments will be useful in unraveling the human proteome.

This is truly the “omics” era! As the technology to study massive quantities of data has improved, scientists have been able to focus on many different aspects of biochemistry. Genomics was just the beginning with the elucidation of the human genome. Now scientists are studying such concepts as the “transcriptome,” the sum of all genes being transcribed into RNAs in the cell, the “metabolome,” the sum of all the components of metabolic pathways, or the “kinome,” the sum of those proteins involved in kinase reactions.

### The Power of Microarrays—Robotic Technology Meets Biochemistry

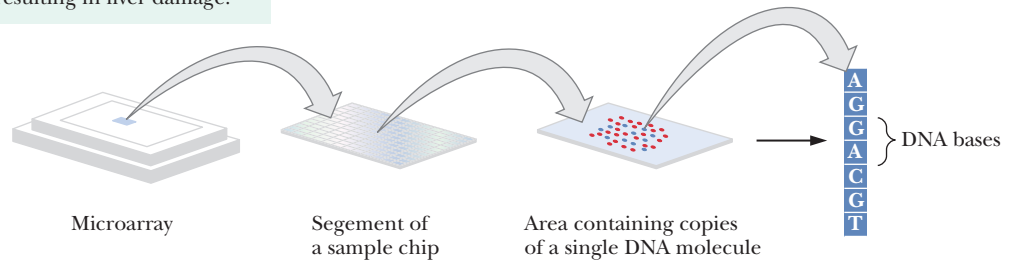
Thousands of genes and their products (i.e., RNA and proteins) in a given living organism function in a complicated and harmonious way. Unfortunately, traditional methods in molecular biology have always focused on analyzing one gene per experiment. In the past several years, a new technology, called **DNA microarray** (**DNA chip** or **gene chip**), has attracted tremendous interest among molecular biologists. Microarrays allow for the analysis of an entire genome in one experiment and are used to study gene expression, the transcription rates of the genome in vivo. The genes that are being transcribed at any particular time are known as the **transcriptome**. The principle behind the microarray is the placement of specific nucleotide sequences in an ordered array, which then base-pair with complementary sequences of DNA or RNA that have been labeled with fluorescent markers of different colors. The locations where binding occurred and the colors observed are then used to quantify the amount of DNA or RNA bound. Microarray chips are manufactured by high-speed robotics, which can put thousands of samples on a glass slide with an area of about 1 cm<sup>2</sup>. The diameter of an individual sample might be 200  $\mu$ m or less. Several different methods are used for implanting the DNA to be studied on the chip, and many companies make microarray chips.

#### How do microarrays work?

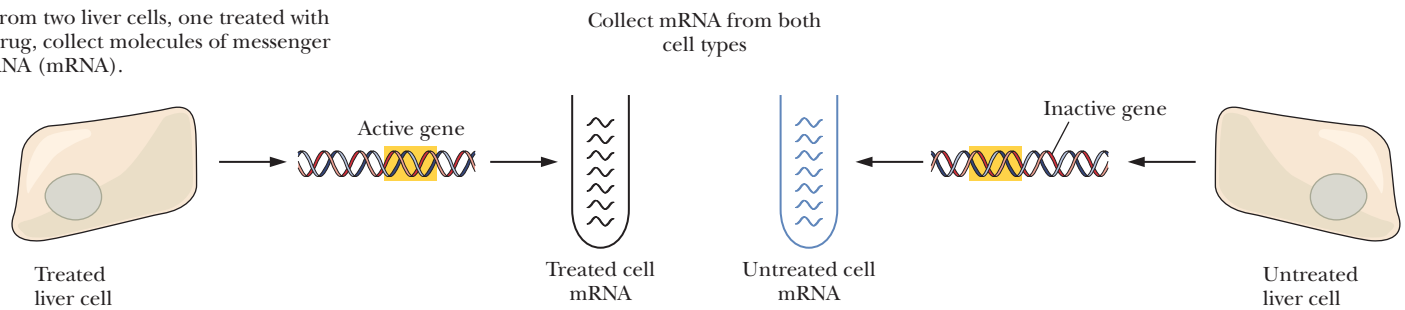
Figure 13.32 shows an example of how microarrays could be used to determine whether a potential new drug would be harmful to liver cells. In Step 1, a microarray is purchased or constructed that has single-stranded DNA representing thousands of different genes, each applied at a specific spot on the microarray chip. In Step 2, different populations of liver cells are collected, one treated with the potential drug and the other untreated. The mRNA being transcribed in these cells is then collected. In Step 3, the mRNA is converted to cDNA.

**Question—** to test whether a new drug will harm the liver by causing the genes in liver cells to alter their activity resulting in liver damage.

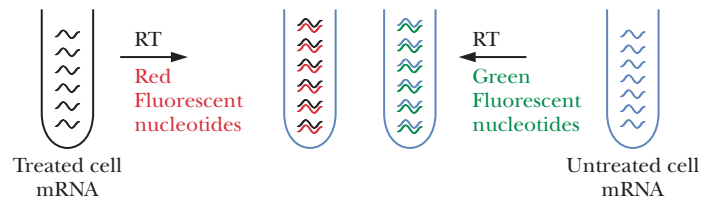
**1** Construct or buy a microarray



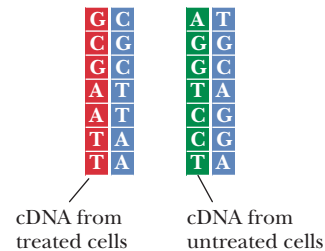
**2** From two liver cells, one treated with drug, collect molecules of messenger RNA (mRNA).



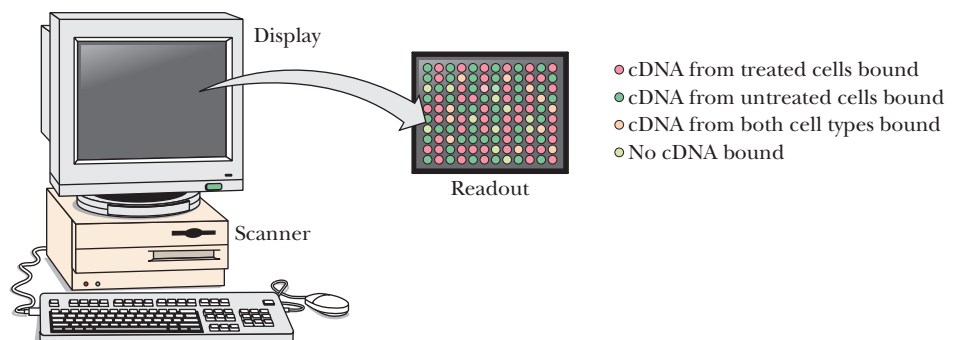
**3** Use reverse transcriptase (RT) to make cDNA.



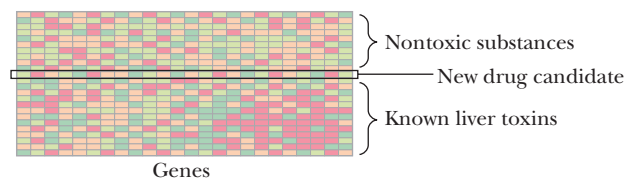
**4** Apply the green and red labeled cDNAs to the chip. When cDNA binds to its complementary sequence it means that the DNA from the sample was active.



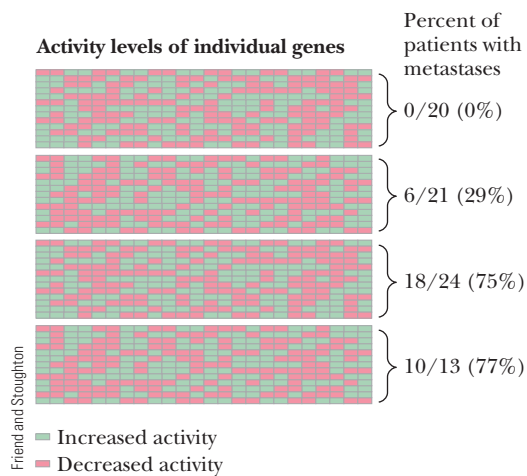
**5** Place the sample chip in a scanner; a computer can calculate the ratio of red to green spots and generate a color-coded readout.



**6** Close similarity would indicate new drug candidate was probably toxic.



**FIGURE 13.32** How arrays work.



**FIGURE 13.33** Expression profiles. Four microarrays show different expression patterns for cancer patients. These are compared to the percentage of patients that later developed metastases.

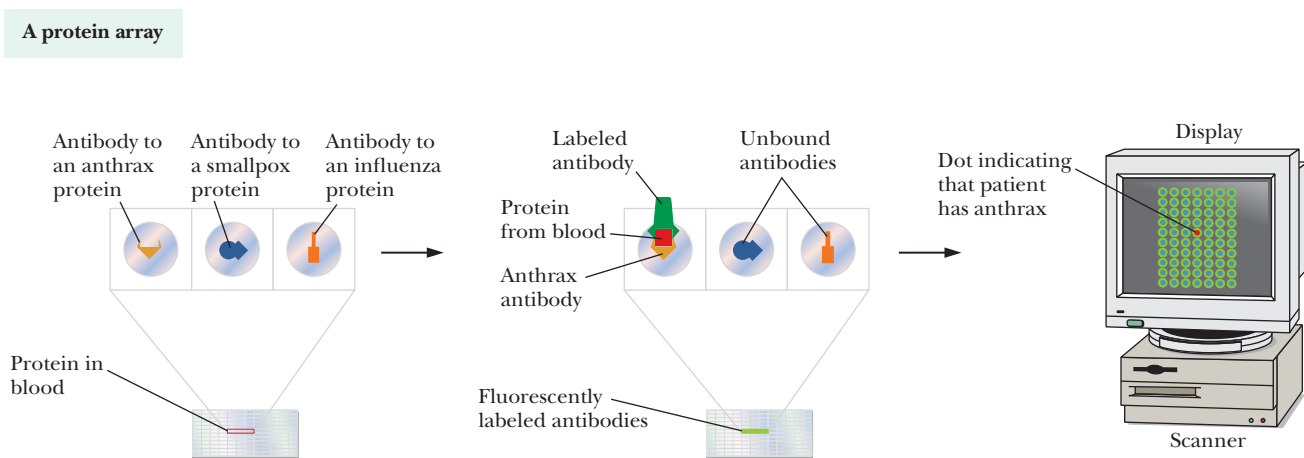
Green fluorescent labels are added to the cDNA from the untreated cells, and red fluorescent labels are added to the cDNA from the treated cells. In Step 4, the labeled cDNAs are added to the chip. The cDNAs bind to the chip if they find their complementary sequences in the single-stranded DNAs loaded on the chip. The expanded portion of Step 4 shows what is happening at the molecular level. The black sequences represent the DNA bound to the chip. Green sequences represent the cDNA from the untreated cells that bind to their target sequences. The red sequences are cDNA from treated cells. Some of the DNA sequences on the chip bind nothing. Some bind only the red sequences while others bind only the green ones. Some sequences on the chip bind both.

In Step 5, the chip is scanned and a computer analyzes the fluorescence. The results appear as a series of colored dots. A red dot indicates a DNA sequence on the chip that bound to the cDNA from the treated cells. This indicates an mRNA that was being expressed in treated cells. A green dot indicates RNA produced in untreated but not treated cells. A yellow dot would indicate an mRNA that was produced equally well in treated or untreated cells. Black spaces indicate DNA sequences on the chip for which no mRNA was produced in either situation. To answer the question about whether the potential drug is toxic to liver cells, the results from the microarray would be compared to controls run with liver cells and drugs known to be toxic versus those known to be nontoxic, as shown in Step 6.

Figure 13.33 shows the results of a study designed to scan cells from cancer patients and correlate microarray patterns with prognoses. The four different patterns are compared to the percentage of patients who developed metastases. Information like this could be critical to treatment of cancer. Doctors often have to choose between different strategies. If they had access to data like this from their patients, they would be able to predict the likelihood of the patient's developing more serious forms of cancer, and thereby able to choose a more appropriate treatment.

## Protein Arrays

Another type of microchip uses bound proteins instead of DNA. These protein arrays are based on interactions between proteins and antibodies (Chapter 14). For example, antibodies to known diseases can be bound to the microarray. A sample of a patient's blood can then be put on the microarray. If the patient has a particular disease, proteins specific to that disease bind to the appropriate antibodies. Fluorescently labeled antibodies are then added and the microarray scanned. The results look similar to the DNA microarrays discussed previously. Figure 13.34 shows how this would work to identify that a patient had anthrax. This technique is growing in popularity and power, but is limited by whether purified antibodies have been created for a particular disease.



**FIGURE 13.34** Protein arrays in action.



## SUMMARY

**How are nucleic acids separated?** Two of the primary necessities for successful experiments with nucleic acids are to separate the components of a mixture and to detect the presence of nucleic acids. DNA can be cut into pieces with restriction enzymes and then separated with gel electrophoresis.

**How can we visualize DNA?** DNA can be visualized with a variety of dyes. The most standard is to react digested DNA in a gel with ethidium bromide. Bands of DNA fluoresce orange under UV light and can be seen easily.

**Why are sticky ends important?** Restriction endonucleases play a large role in the manipulation of DNA. These enzymes produce short, single-stranded stretches, called sticky ends, at the ends of cleaved DNA. The sticky ends provide a way to link DNAs from different sources, even to the point of inserting eukaryotic DNA into bacterial genomes.

**What is cloning?** DNA samples from different sources can be selectively cut by using restriction endonucleases and then spliced together with DNA ligase to produce recombinant DNA. DNA from another source can be introduced into the genome of a virus or a bacterium. In bacteria, the foreign DNA is usually introduced into a plasmid, a smaller circular DNA separate from the main bacterial chromosome. The growth of the virus or bacterium also produces large amounts of the other DNA by the process of cloning.

**What are plasmids?** Plasmids are small circular pieces of DNA of bacterial origin. They usually carry several useful genes, such as an antibiotic resistance gene. Once inside a bacterium, they replicate when the bacterial DNA replicates.

**What is the purpose of genetic engineering?** Once the DNA is successfully cloned, it can be expressed using an expression vector and cell line. This allows for the production of eukaryotic proteins quickly and cheaply in bacterial hosts. Genetically altered organisms, such as mice and corn, have been engineered for both pure and applied scientific purposes, and many more changes are to come.

**How can human proteins be made by bacteria?** If a human gene is inserted into a plasmid and the plasmid is transformed into a bacterial cell, when the cell divides and replicates, so does the plasmid. If the plasmid is an expression vector, then it can also transcribe and translate the protein from the human gene being carried by the plasmid. In this way, bacteria can produce human proteins for science and medicine.

**What is an expression vector?** An expression vector is a plasmid that contains a ribosomal binding site and termination

site so that mRNA produced from a gene insert it carries is translated in the host cell.

**How do we find the piece of DNA we want in a library?** Once a DNA library is established on a series of plates, colonies are transferred to filter discs. These discs can be reacted with specific DNA probes for a sequence of interest. The probe has  $^{32}\text{P}$  in it, so when the probe binds, its location can be determined with autoradiography. The colony at that position can then be isolated and amplified for study.

**What are the advantages of PCR?** An alternative method of producing large amounts of a given DNA, called the polymerase chain reaction, depends only on enzymatic reactions and does not require viral or bacterial hosts. The procedure is automated and relies on a heat-stable form of DNA polymerase. The advantage is that PCR is faster, produces more DNA, and does not require plasmids and bacterial growth to generate results. Quantitative PCR allows for sensitive measurements of the amount of target DNA that was in a sample.

**How can differences in DNA from individuals be seen?** The ability to analyze DNA fragments is important for basic research and for forensic science. DNA fingerprinting allows for the identification of individuals from their DNA samples. DNA is digested with restriction enzymes, and a banding pattern is seen upon electrophoresis of the digest. No two individuals have the same pattern, just as no two people have the same fingerprints. This technique is often used for paternity tests and for identification of criminals.

**How do dideoxy nucleotides allow us to sequence DNA?** The Sanger–Coulson method allows for the determination of a DNA sequence by using dideoxy nucleotides. These nucleotides cause chain termination during DNA synthesis. By running parallel reactions with a dideoxy version of one each of the four deoxynucleotides, we can see a banding pattern on a gel that allows us to read the DNA sequence.

**How do microarrays work?** Robotic technology is used to load thousands of samples of DNA or protein onto a microchip. Biological samples are then overlaid on the chip and the binding is evaluated through the use of fluorescence. In the case of a DNA chip, for example, the mRNA from a cell is labeled with a red fluorescent marker. When put on the chip, the location of the red dots tells which DNA sequences were matched by the mRNA. Another cell sample taken under different conditions may have its mRNA labeled with a green fluorescent marker. Comparing where the red and green dots are then tells the researcher the differences in mRNA expression under the two circumstances.

## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

### 13.1 Purification and Detection of Nucleic Acids

- Recall** What advantages does fluorescent labeling offer over radioactive methods of labeling DNA?
- Recall** What methods are used to visualize radioactively labeled nucleic acids?
- Reflect and Apply** When proteins are separated using native gel electrophoresis, size, shape, and charge control their rate of migration on the gel. Why does DNA separate based on size, and why do we not worry much about shape or charge?

### 13.2 Restriction Endonucleases

- Recall** How does the use of restriction endonucleases of different specificities aid in the sequencing of DNA?
- Recall** What is the importance of methylation in the activity of restriction endonucleases?
- Recall** Why do restriction endonucleases not hydrolyze DNA from the organism that produces it?
- Recall** What role did restriction endonucleases play in localizing the gene associated with cystic fibrosis?
- Recall** Where did restriction endonucleases get their name?
- Recall** What do the following have in common? MOM; POP; NOON; MADAM, I'M ADAM; A MAN, A PLAN, A CANAL: PANAMA.
- Recall** Give three examples of DNA palindromes.
- Recall** What are three differences between the sites recognized by *Hae*III and those recognized by *Bam*HI?
- Recall** What are sticky ends? What is their importance in recombinant DNA technology?
- Recall** What would be an advantage of using *Hae*III for a cloning experiment? What would be a disadvantage?

### 13.3 Cloning

- Recall** Describe the cloning of DNA.
- Recall** What vectors can be used for cloning?
- Recall** Describe the method you would use to test for the uptake of a plasmid with a DNA insert.
- Recall** What is blue/white screening? What is the key feature of a plasmid that is used for it?
- Reflect and Apply** What are some general “requirements” for recombinant DNA technology?
- Reflect and Apply** What are some of the dangers of (and precautions against) recombinant DNA technology?

### 13.4 Genetic Engineering

- Recall** What are the purposes of genetic engineering in agriculture?
- Recall** What human proteins have been produced by genetic engineering?
- Recall** You go for a drive in the country with some friends and pass a cornfield with a sign. They do not understand the cryptic message on the sign, with the letters “Bt” followed by some numbers. You are able to enlighten them on the basis of information from this chapter. What is that information?
- Reflect and Apply** Using information we have seen about lactate dehydrogenase, how could you clone and express human lactate dehydrogenase 3 (LDH 3) in bacteria?
- Reflect and Apply** What are the requirements for an expression vector?

- Reflect and Apply** What is a fusion protein? How are fusion proteins involved in cloning and expression?
- Reflect and Apply** A friend tells you that she doesn't want to feed her baby high-production milk because she is afraid that the BST will interfere with the baby's growth by overstimulation. What do you tell her?
- Reflect and Apply** The genes for both the  $\alpha$ - and  $\beta$ -globin chains of hemoglobin contain introns (i.e., they are split genes). How would this fact affect your plans if you wanted to introduce the gene for  $\alpha$ -globin into a bacterial plasmid and have the bacteria produce  $\alpha$ -globin?
- Reflect and Apply** Outline the methods you would use to produce human growth hormone (a substance used in the treatment of dwarfism) in bacteria.
- Reflect and Apply** Bacteria and yeast are known not to have prions (Chapter 4). What does this fact have to do with the popularity of expressing mammalian proteins using bacterial vectors?

### 13.5 DNA Libraries

- Recall** What are the differences between a DNA library and a cDNA library?
- Reflect and Apply** Why is it a large undertaking to construct a DNA library?
- Reflect and Apply** Why do some journals require the authors of articles describing DNA libraries to make those libraries available to other researchers?

### 13.6 The Polymerase Chain Reaction

- Recall** Why is temperature control so important in the polymerase chain reaction?
- Recall** Why is the use of temperature-stable DNA polymerase an important factor in the polymerase chain reaction?
- Recall** What are the criteria for “good” primers in a PCR reaction?
- Reflect and Apply** What difficulties arise in the polymerase chain reaction if there is contamination of the DNA that is to be copied?
- Reflect and Apply** Each of the following pairs of primers has a problem with it. Tell why the primers would not work well.
  - Forward primer 5' GCCTCCGAGACCCATTGG 3'  
Reverse primer 5' TTCTAAGAACTGTTAAGG 3'
  - Forward primer 5' GGGGCCCCCTCACTCGGGGCCCC 3'  
Reverse primer 5' TCGGCGCCGTGGCCGAGGCAG 3'
  - Forward primer 5' TCGAATTGCCAATGAAGGTCCG 3'  
Reverse primer 5' CGGACCTTCATTGGCAATTCTGA 3'
- Recall** What is qPCR?
- Recall** What is the functional difference between regular PCR and qPCR?

### 13.7 DNA Fingerprinting

- Reflect and Apply** Suppose that you are a prosecuting attorney. How has the introduction of the polymerase chain reaction changed your job?
- Reflect and Apply** Why is DNA evidence more useful as exclusionary evidence than for positive identification of a suspect?

### 13.8 Sequencing DNA

- Reflect and Apply** Give the DNA sequence for the template strand that gives rise to the following sequence gel, prepared using

the Sanger method with a radioactive label at the 5' end of the primer.

A	C	G	T
	—		
—	—		—
		—	
—		—	
	—		—

43. **Reflect and Apply** Although techniques are available for determining the sequences of amino acids in proteins, it is becoming more and more common to sequence proteins indirectly by determining the base sequence of the gene for the protein and then inferring the amino acid sequence from the genetic-code relationships. Suggest why the latter technique is being used for proteins.
44. **Reflect and Apply** Sometimes knowing the DNA sequence of a gene that codes for a protein does not tell you the amino acid sequence. Suggest several reasons why this is so.
45. **Reflect and Apply** This is a conjectural question—there is no single “right” answer—that is good for discussion over tea and crumpets.

In what ways might it be possible to prevent genetic discrimination due to information made available by the Human Genome Project?

46. **Reflect and Apply** A recent television commercial featuring seven-time Tour de France winner Lance Armstrong talked about the possibility of people carrying a DNA genotype card with them that would contain all of the information necessary to predict future diseases. This could, therefore, be used to help prescribe drugs to stop a medical condition before it became apparent. Give a couple of specific examples of how this ability could be used for the benefit or the detriment of humankind.

### 13.9 Genomics and Proteomics

47. **Recall** What is the difference between the genome and the proteome?
48. **Recall** Has proteomic analysis been done on multicellular eukaryotes?
49. **Recall** Explain how a DNA microchip works.
50. **Reflect and Apply** If you wanted to study the nature of transcription in yeast under aerobic versus anaerobic conditions, how could you use DNA microarrays to accomplish this?
51. **Recall** How are DNA microarrays used to screen a patient's cells for a cancer prognosis?
52. **Recall** What are the key differences between DNA microarrays and protein microarrays, and how they are used in research?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# Viruses, Cancer, and Immunology

# 14

Human Immunodeficiency Virus (HIV) spreads from cell to cell. The viral RNA is packaged in the red cone structure and is surrounded by a membrane, part of which comes from the host cell. This picture shows the virus bursting free of the infected helper T cell. It will then find another cell to bind to and infect.

## 14.1 Viruses

Viruses have always been difficult to classify according to normal taxonomy. Many have argued over whether they should be considered living things. They cannot reproduce independently, and they cannot make proteins or generate energy independently, so they do not meet all of the requirements for life as we have traditionally defined it. But if they are not life forms, what are they? The simplest definition would be a relatively small amount of genetic material surrounded by a protein envelope. Most viruses have only one type of nucleic acid, either DNA or RNA. Depending on the virus, this nucleic acid could be either single-stranded or double-stranded.

### Why are viruses important?

Viruses are known for the diseases they cause. They are pathogens of bacteria, plants, and animals. Some viruses are deadly, such as the fast-acting **Ebola virus**, which can have a mortality rate of more than 85%, and the slow-acting but equally deadly **human immunodeficiency virus (HIV)**, which causes **acquired immunodeficiency syndrome (AIDS)**. Other viruses might be simply annoying, such as **rhinovirus**, which causes common colds.

### What is the structure of a virus?

Viruses are very small particles composed of nucleic acid and protein. The entire virus particle is called the **virion**. At the center of the virion is the nucleic acid. Surrounding this is the **capsid**, which is a protein coat. The combination of the nucleic acid and the capsid is called the **nucleocapsid**, and, for some viruses, such as the rhinovirus, that is the extent of the particle. Many other viruses, including HIV, have a **membrane envelope** surrounding the nucleocapsid. Many viruses also have **protein spikes** that help them attach to their host cell. Figure 14.1 shows the main features of a virus.

The overall shape of a virus varies. The classic viral shape most often seen in the literature has a hexagonal capsid with a rod sticking out of it that attaches to the host cell and acts like a syringe to inject the nucleic acid. Figure 14.2 shows the T2 bacteriophage of *E. coli*, a classic example of a virus of this shape. Tobacco mosaic virus (TMV), on the other hand, has a rod shape, as shown in Figure 14.3.

## Families of Viruses

Although many characteristics distinguish viruses, most are organized by whether they have a genome of DNA or of RNA and whether they have an envelope. In addition, the nature of the nucleic acid (linear vs. circular, small vs. large, single-stranded vs. double-stranded) and the mode of incorporation (nucleic acid remains separate vs. nucleic acid joins with host chromosome) distinguish the different virus types. Table 14.1 shows some of the known viral diseases and the families of viruses that cause them.

## Chapter Outline

### 14.1 Viruses

- Why are viruses important?
- What is the structure of a virus?
- How does a virus infect a cell?

### 14.2 Retroviruses

- Why are retroviruses important?

### 14.3 The Immune System

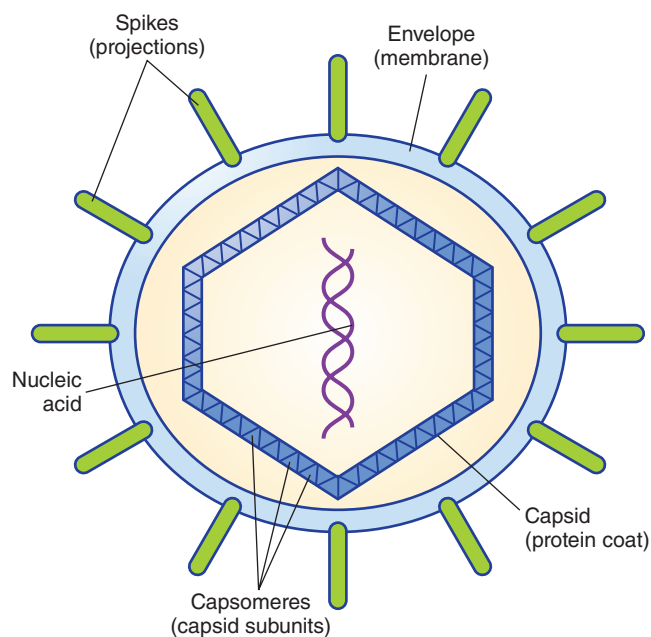
- How does the immune system work?
- What do T cells and B cells do?
- What are antibodies?

### 14.4 Cancer

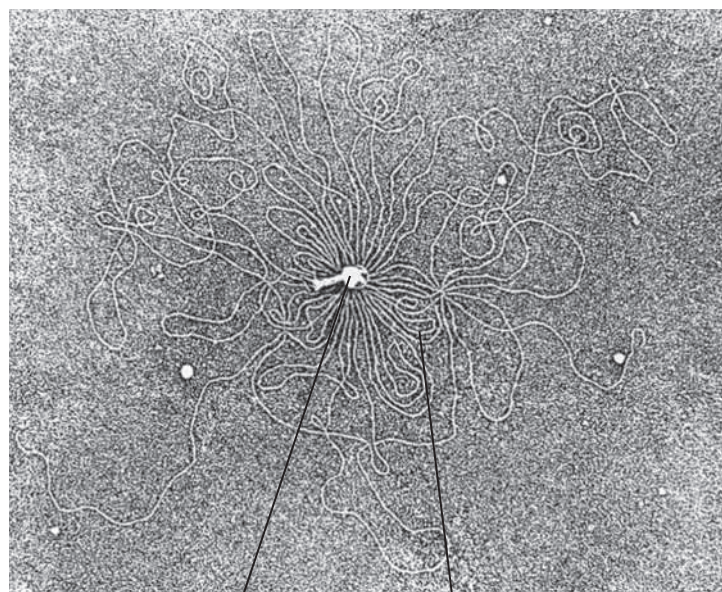
- What characterizes a cancer cell?
- What causes cancer?
- How do we fight cancer?

**OWL** Online homework for this chapter may be assigned in OWL.





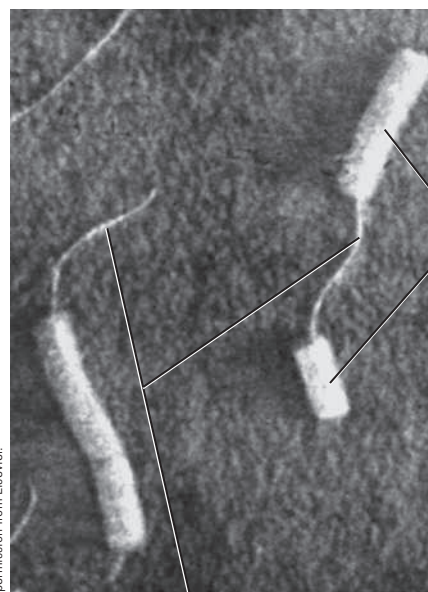
■ **FIGURE 14.1** The architecture of a typical virus particle. The nucleic acid is in the middle, surrounded by a protein coat called the capsid. Many viruses also have an envelope membrane that is usually covered with protein spikes.



The late Prof. Emeritus Albrecht K. Kleinschmidt, Universität Ulm, Ulm Germany, *Biochem. Biophys. Acta*, 61: 859, with permission from Elsevier Science.

Capsid DNA

■ **FIGURE 14.2** An electron micrograph of a hexagonal virus. The bacteriophage T2 virus was gently disrupted, releasing the DNA, which can be seen as many loops outside the virus.



Reprinted from *Virology*, vol 22, 1964, pp. 539-543, Dr. K.M. Corbett. "Tobacco Mosaic Virus." With permission from Elsevier.

RNA

■ **FIGURE 14.3** An electron micrograph of the rod-shaped tobacco mosaic virus.

## Virus Life Cycles

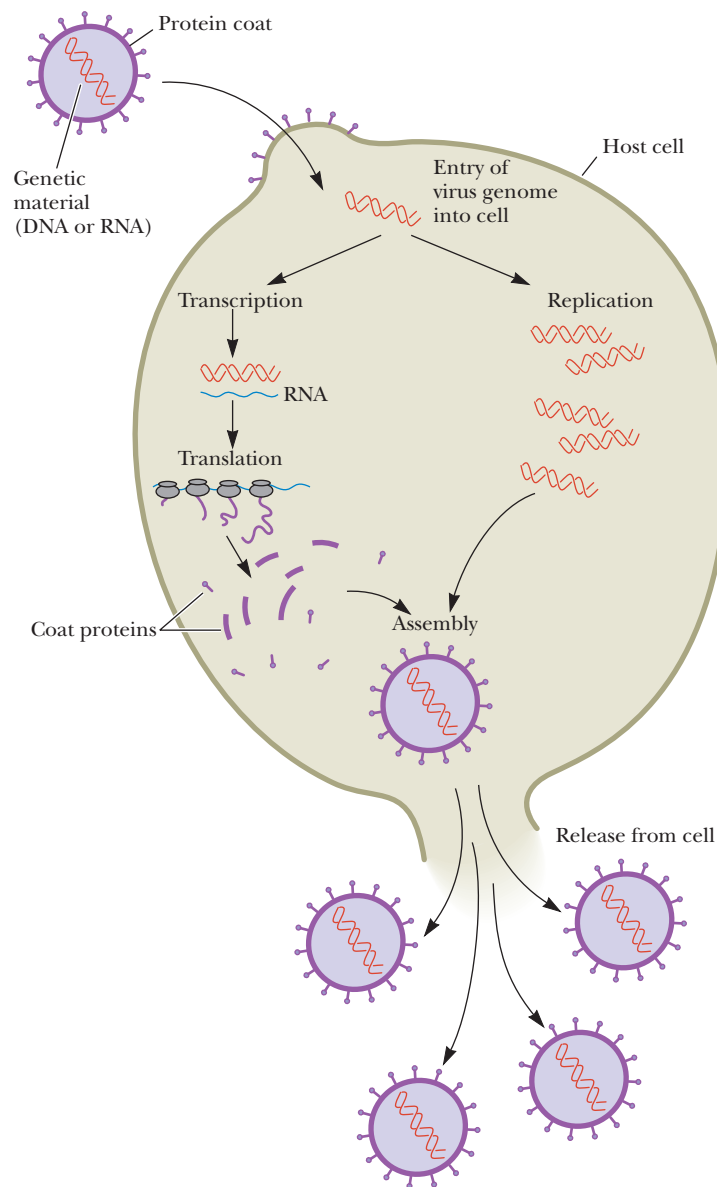
Most viruses cannot survive for long periods outside cells, so they must quickly gain access to a cell. There are several mechanisms for gaining access, and preventing access to the cell has been a major focus of pharmaceutical companies trying to develop antiviral drugs. Figure 14.4 shows a generic example of a virus infecting a cell. The virus binds to the cell membrane and releases its DNA into the cell. The DNA is then replicated by host DNA polymerases and transcribed

TABLE 14.1

Vertebrate Viruses and Diseases They Cause			
DNA-Containing Viruses, Nonenveloped		RNA-Containing, Enveloped	
Adenoviruses	Respiratory and gastrointestinal diseases	Arenaviruses Lassa virus	Lassa fever
Circoviruses	Anemia in chickens	Arteriviruses Equine arteritis virus	Equine viral arteritis
Iridoviruses	Various diseases of insects, fish, and frogs	Bunyaviruses California encephalitis virus	California encephalitis
Papovaviruses Papillomaviruses Human papillomavirus	Warts Cervical cancer	Hantaviruses	Epidemic hemorrhagic fever or pneumonia
Parvoviruses Human parvovirus B19 Canine parvovirus	Fifth disease (a childhood rash) Viral gastroenteritis in dogs	Coronaviruses	Respiratory disease, possibly gastroenteritis
DNA-Containing Viruses, Enveloped		Filoviruses Ebola virus Marburg virus	Ebola disease Marburg disease
African swine fever virus	African swine fever (rarely in humans)	Flaviviruses Dengue virus Hepatitis C virus St. Louis encephalitis virus Yellow fever virus	Dengue fever Hepatitis C St. Louis encephalitis Yellow fever
Hepadnaviruses	Hepatitis B	Orthomyxoviruses	Influenza
Herpesviruses Cytomegalovirus Epstein–Barr virus (EBV)	Birth defects Infectious mononucleosis and Burkitt's lymphoma	Paramyxoviruses Measles virus Mumps virus Respiratory syncytial virus (RSV)	Measles Mumps Pneumonia, bronchitis
Herpes simplex virus (HSV) type 1	Cold sores	Retroviruses Human T-cell lymphotropic viruses (HTLVs)	Leukemia, lymphoma
Herpes simplex virus (HSV) type 2	Genital herpes	Human immuno-deficiency virus (HIV)	Acquired immune deficiency syndrome (AIDS)
Varicella-zoster virus (herpes zoster virus)	Chickenpox and shingles	Rhabdoviruses Rabies virus	Rabies
Poxviruses Monkeypox virus Variola major virus	Monkeypox Smallpox	Togaviruses Rubella virus Eastern equine encephalomyelitis (EEE) virus	Rubella (German measles) Encephalomyelitis
RNA-Containing Viruses, Nonenveloped			
Astroviruses	Gastroenteritis		
Birnaviruses	Various diseases of birds, fish, and insects		
Calciviruses	Norwalk gastroenteritis		
Picornaviruses Hepatitis A virus Polioviruses Rhinoviruses	Hepatitis A Poliomyelitis Common cold		
Reoviruses Rotaviruses	Infantile gastroenteritis		

by host RNA polymerases. The transcription and translation of the mRNA leads to the proteins that are necessary to make the coat proteins of the capsid. New virions are produced and then released from the cell. This is called the **lytic** pathway, as the host cells are lysed by this process.

Viruses do not always lyse their host cells, however. A separate process called **lysogeny** involves the incorporation of the viral DNA into the host chromosome. **Simian virus 40 (SV40)** is an example of a DNA virus. It appears to be spherical, but it is actually an icosahedron, a geometric shape with 20 faces that are equilateral triangles, as shown in Figure 14.5. The genome of this virus is a closed circle of double-stranded DNA, with genes that encode the amino acid sequences of five proteins. Three of the five proteins are coat proteins. Of the



■ **FIGURE 14.4** The virus life cycle. Viruses are mobile bits of genetic information encapsulated in a protein coat. The genetic material may be either DNA or RNA. Once this genetic material gains entry into the host cell, it takes over the host machinery for macromolecular synthesis and subverts it to the synthesis of viral specific nucleic acids and proteins. These virus components are assembled into mature virus particles, which are then released from the cell. Often, this parasitic cycle of virus infection leads to cell death and disease.

remaining two proteins, one, the large-T protein, is involved in the development of the virus when it infects a cell. The function of the fifth protein, the small-T protein, is not known.

The outcome of infection by SV40 depends on the organism infected. When simian cells are infected, the virus enters the cell and loses its protein coat. The viral DNA is expressed first as mRNAs and then as proteins. The large-T protein is the first one made (Figure 14.6), triggering the replication of viral DNA, followed by viral coat proteins. The virus takes over the cellular machinery for both replication of DNA and protein synthesis. New virus particles are assembled, and eventually the infected cell bursts, releasing the new virus particles to infect other cells.

The results are different when SV40 infects rodent cells. The process is the same as far as the production of the large-T protein, but replication of the viral DNA does not take place. The SV40 DNA already present in the cell can be lost or can be integrated into the DNA of the host cell. If the SV40 DNA is lost, the infection has no apparent result. If it becomes integrated into the DNA of



the host cell, the infected cell loses control of its own growth. As a result of the accumulation of large-T protein, the infected cell behaves like a cancer cell. The large-T gene is an **oncogene**, one that causes cancer. Its mechanism is a subject of active research. The relationship between viruses and cancer will be looked at in more detail in Section 14.4.

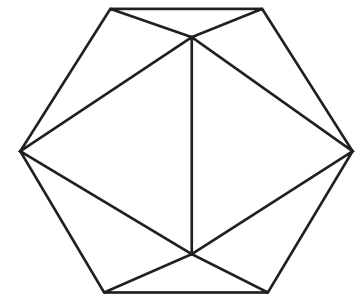
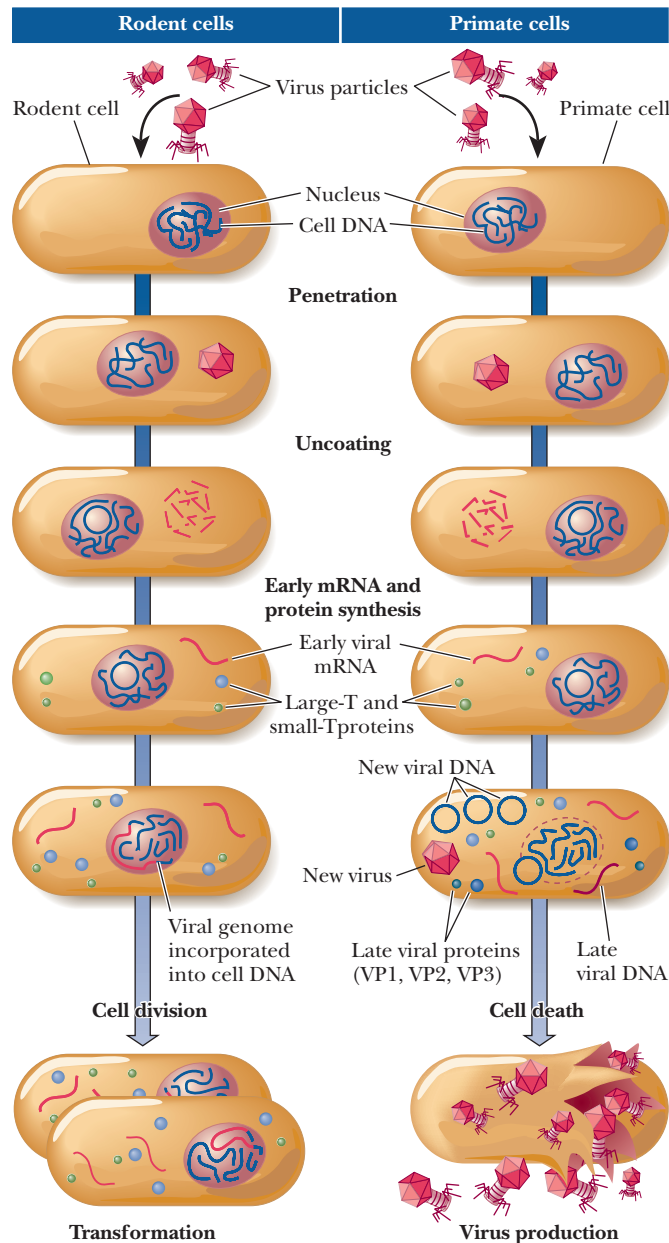
### How does a virus infect a cell?

A virus must attach to a host cell before it can penetrate, which is why so much research is involved in studying the exact mechanisms of viral attachment. A common method of attachment involves the binding of one of the spike proteins on the envelope of the virus to a specific receptor on the host cell. Figure 14.7 shows an example of HIV attachment. A specific spike protein called *gp120* binds to a CD4 receptor on helper T cells. After this happens, a co-receptor complexes with CD4 and *gp120*. Another spike protein, *gp41*, then punctures the cell so that the capsid can enter.



Dr. Jack Griffith/Lineberger Comprehensive Cancer Center, UNC

**A** Virus particles appear almost spherical in electron micrographs, but, on closer examination, they can be seen to have an icosahedral shape.



**B** The geometry of an icosahedron. This regular polyhedron has 20 faces, all of which are equilateral triangles of identical size.

**FIGURE 14.5** The architecture of simian virus 40 (SV40).

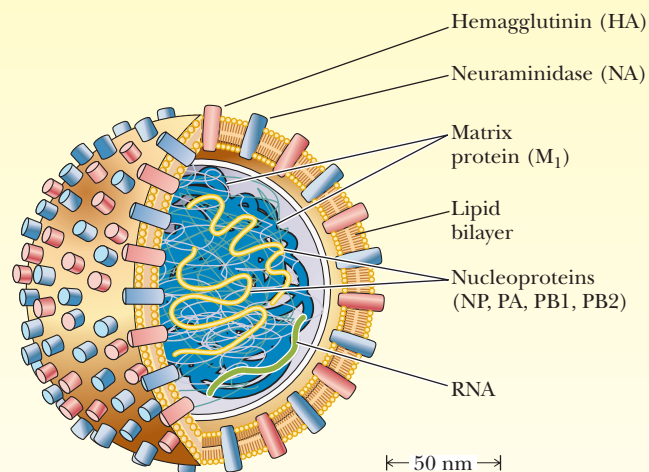
**FIGURE 14.6** Simian virus 40 life cycle. The outcome of infection of cells by simian virus 40 depends on the nature of the cells. When primate cells are infected, the large-T protein is produced, and its presence ultimately leads to the production of new viral DNA and coat proteins. New virus particles are assembled and released; the death of the host cell takes place when the new virions are released. When rodent cells are infected, the viral genome is incorporated into the cell DNA. (Adapted from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)

# Biochemical Connections EPIDEMIOLOGY

## A Little Swine Goes a Long Way

In the fall of 2009, a common phrase heard in the schoolyard was, “he’s got the swine,” referring to the outbreak of swine flu that had begun in the spring of that year. Certainly anyone reading this book has had influenza—the flu—a disease that most people take for granted as an annoying fact of life. There are yearly epidemics around the world, with some being very serious. In 1918, there was a worldwide flu pandemic that led to the deaths of 50 million people, one of the worst epidemics in history, surpassing even the black plague of the Middle Ages. By comparison, there are only about 40 million people today living with the HIV virus, and it has taken 30 years to get to that point. The flu virus has been with us for thousands of years and has never been fully controlled by modern medicine.

A single particle of the influenza virus (a virion) is a single-stranded RNA template strand genome with a protein coat that protrudes through a lipid bilayer envelope. The figure to the right shows the structural features of the influenza virus. There are three major types, designated A, B, and C, depending on differences in the proteins. Influenza viruses cause infections of the upper respiratory tract that lead to fever, muscle pain, headaches, nasal congestion, sore throat, and coughing. One of the biggest problems is that people who catch the flu often get secondary infections, including pneumonia, which is what makes the flu potentially lethal. We are going to talk about the influenza A virus because, of the three, it is responsible for most human illness. The most prominent features of the virus envelope are two spike proteins. One is called hemagglutinin (HA), which gets its name because it causes erythrocytes to clump together. The second is neuraminidase (NA), an enzyme that catalyzes the hydrolysis of a linkage of sialic acid to galactose or galactosamine (see Chapter 16). HA is believed to help the virus in recognizing target cells. NA is believed to help the virus get through mucous membranes and enter cells. Sixteen subtypes of HA are known (designated H1–H16), and nine subtypes of neuraminidase (designated N1–N9) have been cataloged. H1, H2, H3, N1, and N2 occur in most of the known viruses that affect humans. Individual influenza A viruses are named by giving the subtypes of HA and NA—for example, H1N1 or H3N2. The virus that causes the avian influenza that has been in the news is H5N1. The presence of the H5 protein affects humans, but so far to a lesser extent than the other HA subtypes. It does, of course, affect birds, with many fatalities among chickens, ducks,

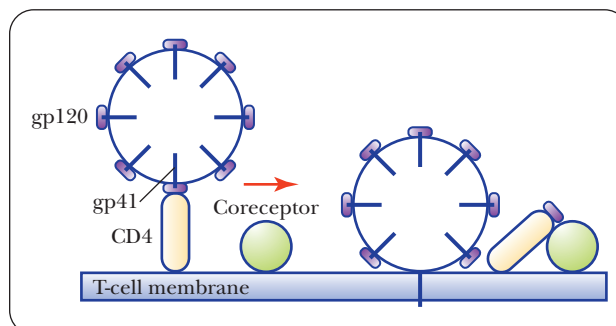


■ **A cutaway diagram of the influenza virion.** The HA and NA spikes are embedded in a lipid bilayer that forms the virion’s outer envelope. A matrix protein, M1, coats the inside of this membrane. The virion core contains the eight single-stranded segments that constitute its genome in a complex with the proteins NP, PA, PB1, and PB2 to form helical structures called neocapsids. (Reprinted with permission from the Estate of Bunji Tagawa.)

and geese. Great concern has arisen about widespread infection of humans by this virus.

The nature of the virus subtype determines its effect on humans. The relevant factors to epidemiologists are the transmissibility and the mortality. For example, there have been only a few hundred cases of people contracting the avian flu worldwide, so its transmissibility is relatively low. However, of those that have gotten the virus, more than 60% have died, so it is still a big concern. In contrast, the current swine flu is the H1N1 variety, and it is more transmissible, but far less deadly to those who get it. In many cases its symptoms are no worse than those of any common flu, and there have been few fatalities. One of the author’s sons goes to a small school that had 32 cases of swine flu in a single day, only two less than would have been necessary to close the school for a week. While not feeling particularly well, they were

■ **FIGURE 14.7 HIV attachment to a helper T cell.** A specific spike protein called gp120 binds to a CD4 receptor on helper T cells. After this happens, a co-receptor complexes with CD4 and gp120. Another spike protein, gp41, then punctures the cell so that the capsid can enter.





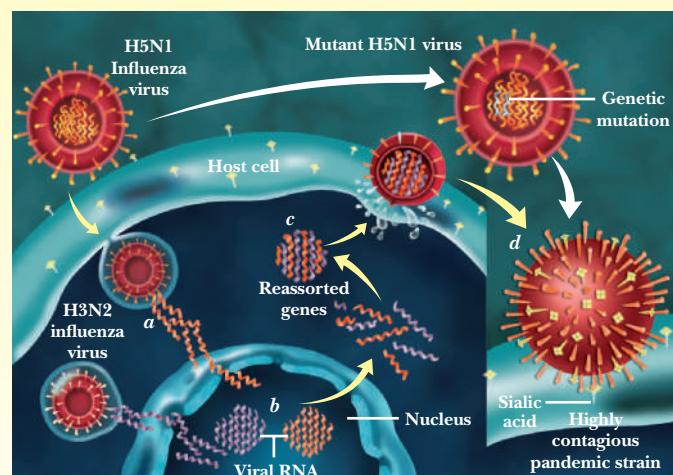
## Biochemical Connections (CONTINUED)

sent to school that day, a fact said author will not be allowed to forget as it cost them a week off of school. One of them was later diagnosed with the swine flu.

While the flu has been with us for millennia, it is always changing, and it is the possibility of such changes that worries agencies responsible for public health, such as the Centers for Disease Control (CDC) and the World Health Organization (WHO). Mutations occur frequently with viruses, and the biggest worry is that a strain with a high mortality could mutate into one that is also very transmissible. The figure to the right shows how the deadly avian strain could potentially change. In one possibility, the virus (H5N1 in this example) mutates and changes its surface proteins, making it more able to bind to human cells and infect them (pink path). The other possibility is that two viruses might infect the same cell (H5N1 and H3N2 in this example, yellow path). The viral RNAs could get mixed and produce reassorted genes, leading to different capabilities in a mutated new strain.

The deadly flu of 1918 was also an H1N1 swine flu. Clues that the 2009 flu was not completely new came from the fact that young people were hit much harder than old people, whereas old people are typically targets for new flu viruses. This indicated that people that had been around for several decades must have had some immunity to the 2009 swine flu, and this clue helped lead to a quicker identification of the flu type. In 1997, a flu that was mostly human in origin was found in North American pigs. A year later researchers found another version that combined genes from human, avian, and swine sources, a **triple reassortant**. The 2009 swine flu is also a triple reassortant, which combines pieces from three different sources. Such combinations demonstrate that flu viruses do not stay contained in one species for long. This is the main reason that scientists worry about what the next jumbling of flu genes will do. It is also why the CDC and WHO take every case of flu seriously. A combination that had the mortality of the avian flu with the transmissibility of the 2009 swine flu could lead to the next plague. Fortunately, it has not happened as of this writing.

Besides watching the annual flu seasons and hoping that new deadly strains will not develop, government agencies also worry about having enough of the vaccines necessary to protect people against the flu. Many people get annual flu shots, and there is a reservoir of standard flu vaccines. Unfortunately, there is not nearly enough vaccine to cover everyone, especially if a new,



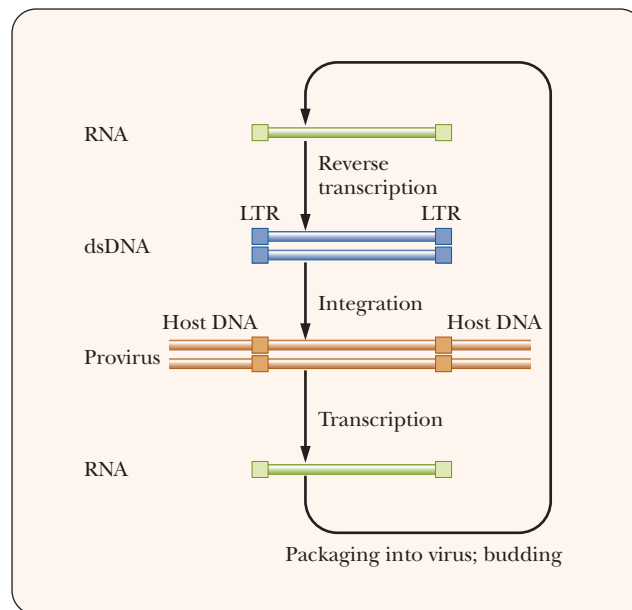
■ **Two Possible Strategies for Mutating Viruses.** The H5N1 strain might undergo a mutation that would make it bind more easily to the cell and therefore be more infective (pink path). The H5N1 and H3N2 strains might both bind to the same cell and then mix their RNA to form reassorted genes (yellow path). (Reprinted with permission. Copyright © 2005 Scientific American, a division of Nature America, Inc. All rights reserved.)

more virulent strain comes about. The 2009 swine flu was another reminder of this problem because the spread of the disease had peaked long before enough vaccine could be made. One of the reasons is that few companies are willing to make vaccines anymore due to fear of litigation. In the mid-1970s there were 40 or 50 companies worldwide making flu vaccines, but now there are just a few. In 1976 the government prepared for a pandemic for a new strain of flu and suggested a massive vaccine program. While there was an epidemic, a pandemic never materialized, and there were people who sued the companies over side effects of the vaccines. Nowadays, governments commission the production of flu vaccines and take the litigation risk upon themselves.

While much of the response to the 1976 epidemic was seen as a disaster, at the same time what was learned about the virus did help lead to more rapid responses to modern flu strains, such as the avian flu and the 2009 swine flu.

## 14.2 Retroviruses

A retrovirus gets its name from the fact that its replication is backward compared to the central dogma of molecular biology: it makes DNA from RNA. The genome of a retrovirus is single-stranded RNA. Once it infects the cell, this RNA is used as a template to make a double-stranded DNA. The enzyme that does this is a virally encoded reverse transcriptase. One of the unique features of the retrovirus life cycle is that the DNA produced by reverse transcription must be incorporated into the host DNA. This occurs because the ends of the DNA produced contain long terminal repeats (LTRs). LTRs are well known in DNA recombination events, and they allow the viral DNA to combine with the host's DNA. Figure 14.8 shows the replication cycle of a retrovirus.

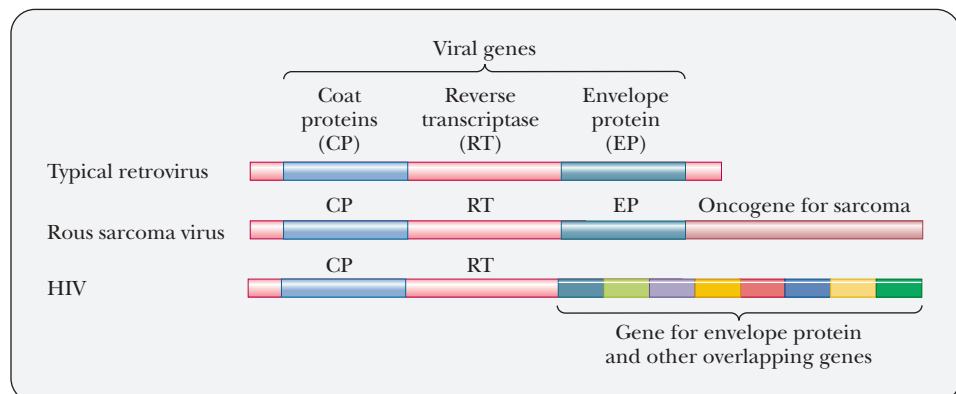


■ **FIGURE 14.8** The life cycle of a retrovirus. Viral RNA is released into the host cell, where the viral reverse transcriptase synthesizes double-stranded DNA from it. The DNA then incorporates into the host's DNA via recombination using the long terminal repeat (LTR) sequences. Eventually, the DNA is transcribed to RNA, which is then packaged into new virus particles.

### Why are retroviruses important?

Retroviruses are the subject of extensive research in virology these days for several reasons. The first is that retroviruses have been linked to cancer, and more aspects of the relationship between viruses and cancer are discovered every day. The second is that **human immunodeficiency virus (HIV)** is a retrovirus. HIV is the causative agent of the disease **acquired immunodeficiency syndrome (AIDS)**. AIDS treatments and a definitive cure are among the primary goals of retroviral research. The third is that retroviruses can be used in gene therapy, as described in the following Biochemical Connections box.

All retroviruses have certain genes in common. There is a gene for proteins of the nucleocapsid, often called **coat proteins (CP)**. They all have a gene for **reverse transcriptase (RT)**, and they all have genes for **envelope proteins (EP)**. Figure 14.9 shows a schematic of the RNA genomes of common retroviruses. In the case of the Rous sarcoma virus, the genome also contains an oncogene that causes tumors (see Section 14.4).



■ **FIGURE 14.9** Typical retrovirus genes. The RNA genomes of all retroviruses have genes for coat proteins (CP), for reverse transcriptase (RT), and for envelope protein (EP). In addition to these essential genes, the Rous sarcoma virus carries the sarcoma oncogene. The HIV genome is more complex, with a number of overlapping genes for envelope proteins and other proteins. (Adapted from *Dealing with Genes: The Language of Heredity*, by Paul Berg and Maxine Singer, © 1992 by University Science Books.)

# Biochemical Connections MEDICINE

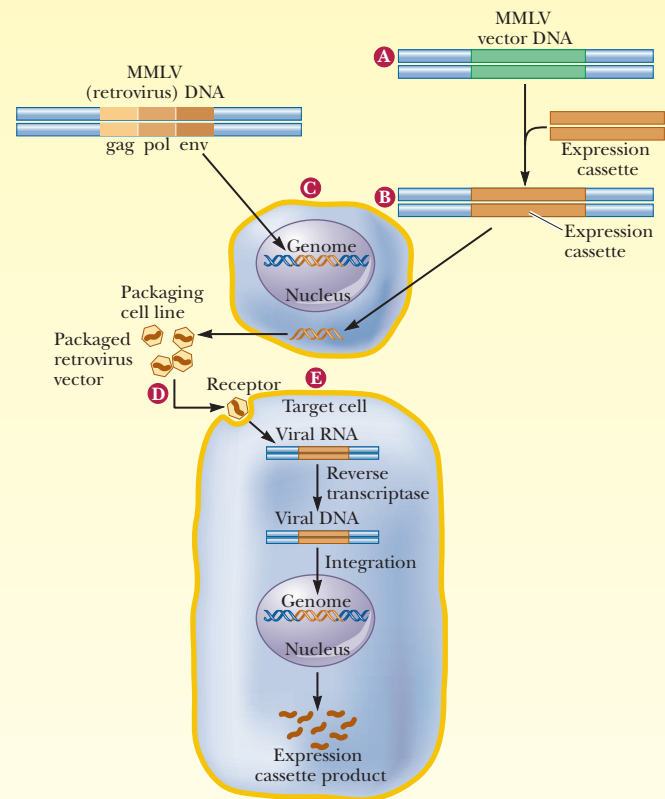
## Viruses Are Used for Gene Therapy

Although viruses have usually been seen as problems for humans, there is one field now in which they are being used for good. Viruses can be used to make alterations in somatic cells, whereby a genetic disease is treated by the introduction of a gene for a missing protein. This is called **gene therapy**. The most successful form of gene therapy to date involves the gene for **adenosine deaminase (ADA)**, an enzyme involved in purine catabolism (Section 23.8). If this enzyme is missing, dATP builds up in tissues, inhibiting the action of the enzyme ribonucleotide reductase. This results in a deficiency of the other three deoxyribonucleoside triphosphates (dNTPs). The dATP (in excess) and the other three dNTPs (deficient) are precursors for DNA synthesis. This imbalance particularly affects DNA synthesis in lymphocytes, on which much of the immune response depends. Individuals who are homozygous for adenosine deaminase deficiency develop **severe combined immune deficiency (SCID)**, the “bubble-boy” syndrome. These individuals are highly prone to infection because of their highly compromised immune systems. The ultimate goal of the planned gene therapy is to take bone marrow cells from affected individuals; introduce the gene for adenosine deaminase into the cells using a virus as a vector; and then reintroduce the bone marrow cells in the body, where they will produce the desired enzyme. The first clinical trials for ADA-SCID were simple enzyme-replacement therapies begun in 1982. The patients were given injections of ADA. Later clinical trials focused on correction of the gene in mature T cells. In 1990, transformed T cells were given to recipients via transfusions.

In trials at the National Institutes of Health (NIH), two girls, ages 4 and 9 at the start of treatment, showed improvement to the extent that they could attend regular public schools and have no more than the average number of infections. Administration of bone marrow stem cells in addition to T cells was the next step; clinical trials of this procedure were undertaken with two infants, ages 4 months and 8 months, in 2000. After 10 months, the children were healthy and had restored immune systems.

There are two types of delivery methods in human gene therapy. The first, called *ex vivo*, is the type used to combat SCID. *Ex vivo* delivery means that somatic cells are removed from the patient, altered with the gene therapy, and then given back to the patient. The most common vector for this is the retrovirus **Maloney murine leukemia virus (MMLV)**. The figure shows how the virus is used for gene therapy. Some of the MMLV is altered to remove the *gag*, *pol*, and *env* genes, rendering the virus unable to replicate. These genes are replaced with an **expression cassette**, which contains the gene being administered, such as the ADA gene, along with a suitable promoter (Chapter 11). This mutated virus is used to infect a packaging cell line. Normal MMLV is also used to infect the packaging cell line, which is not susceptible to the MMLV. The normal MMLV does not replicate in the packaging cell line. Its *gag*, *pol*, and *env* genes restore the mutated virus's ability to replicate, but only in this cell line. These controls are necessary to keep mutant viruses from escaping to other tissues. The mutated virus particles are collected from the packaging cell line and used to infect the target cells, such as bone marrow cells in the case of SCID. MMLV is a retrovirus, so it infects the target cell and produces DNA from its RNA genome, and this DNA can then incorporate into the host genome, along with the promoter and ADA gene. In this way, the target cells that were collected have been transformed and produce ADA. These cells are then put back into the patient.

The second delivery method, called *in vivo*, means that the virus is used to directly infect the patient's tissues. The most common vector for this delivery is the **adenovirus** (which is a DNA virus). A particular vector can be chosen based on specific receptors on the target tissue. Adenovirus has receptors in lung and



■ **Gene therapy via retroviruses.** The Maloney murine leukemia virus (MMLV) is used for *ex vivo* gene therapy. (a) Essential genes (*gag*, *pol*, *env*) are removed from the virus and (b) replaced with an expression cassette containing the gene being replaced with gene therapy. Removal of the essential viral genes renders the viruses unable to replicate. (c) The altered virus is then grown in a packaging cell line that allows replication. (d) Viruses are collected and used to infect cultured target cells from the patient needing the gene therapy. (e) The altered virus produces RNA, which then produces DNA via reverse transcriptase. The DNA integrates in the patient's cells' genome, and his or her cells produce the desired protein. The cultured cells are then given back to the patient. (Adapted from Figure 1 in Crystal, R. G., 1995. *Transfer of genes to humans: Early lessons and obstacles to success*. *Science* 270, 404.)

liver cells, and it has been used in clinical trials for gene therapy of cystic fibrosis and ornithine transcarbamoylase deficiency.

Clinical trials using gene therapy to combat cystic fibrosis and certain tumors in humans began almost 20 years ago, and new diseases are attacked by this technique all the time. In 2009, a fatal demyelinating brain disease called X-linked adrenoleukodystrophy (ALD) was shown to respond well to gene therapy. In mice, gene therapy has been successful in fighting diabetes. The field of gene therapy is exciting and full of promise, but there are many obstacles to success in humans. There are also many risks, such as the risk of a dangerous immunological response to the vector carrying the gene or the danger of a gene becoming incorporated into the host chromosome at a location that activates a cancer-causing gene. This possibility will be discussed further in Section 14.4.

Returning to the case of cystic fibrosis, the disease has been a great frustration to researchers. After much hope (and hype) surrounding the potential for a gene therapy cure, the reality has not matched the promise. Gene therapy simply hasn't worked for a variety of reasons. However, although a cure has remained elusive, the prognosis for CF patients is much better today than it was 20 years ago. New drugs and treatments have improved the quality of life for CF patients.



## 14.3 The Immune System

The immune system is a general term for many cellular and enzymatic processes that allow organisms to defend themselves from viruses, bacteria, and parasites. All vertebrates have an immune system, and we shall discuss several aspects of this critical system in this section.

One distinctive characteristic of the immune system is its ability to *distinguish self from nonself*. This ability enables the cells and molecules responsible for immunity to recognize and destroy pathogens (disease-causing agents, such as viruses and bacteria) when they invade the body—or even one’s own cells when they become cancerous. Because infectious diseases can be fatal, the operation of the immune system can be a matter of life and death. Striking confirmation of this last point is apparent in the lives of those who have AIDS. This disease so weakens the immune system that those who have it become prey to infections that proceed unchecked, with ultimately fatal consequences. Suppression of the immune system can save lives as well as take them. The development of drugs that suppress the immune system has made *organ transplants* possible. Recipients of hearts, lungs, kidneys, or livers tolerate the transplanted organs without rejecting them because these drugs have thwarted the way in which the immune system tries to attack the grafts. However, the immune suppression also makes transplant recipients more susceptible to infections.

The immune system can also go awry in distinguishing self from nonself. The result is **autoimmune disease**, in which the immune system attacks the body’s own tissues. Examples include rheumatoid arthritis, insulin-dependent diabetes, and multiple sclerosis. A significant portion of research on the immune system is directed toward developing approaches for treating these diseases. **Allergies** are another example of improper functioning of the immune system. Millions have asthma as a result of allergies to plant pollens and to other allergens (substances that trigger allergic attacks). Food allergies can evoke violent reactions that may be life-threatening.

Over the years, researchers have unraveled some of the mysteries of the immune system and have used its properties as a therapeutic aid. The first **vaccine**, that against smallpox, was developed about 200 years ago. Since that time it has been used so effectively as a preventive measure that smallpox has been eradicated. The action of vaccines of this sort depends on exposure to the infectious agent in a weakened form. The immune system mounts an attack, and *the immune system retains “memory” of the exposure*. In subsequent encounters with the same pathogen, the immune system can mount a quick and effective defense. This ability to retain “memory” is another major characteristic of the immune system. It is hoped that current research can be carried to the point of developing vaccines that can treat AIDS in people who are already infected. Other strategies are directed at finding treatments for autoimmune diseases. Still others are attempting to use the immune system to attack and destroy cancer cells.

### How does the immune system work?

There are two important aspects to the immunity process: those that operate on the cellular level and those that operate on the molecular level. In addition, we have to look at whether the immune system is acquired or whether it is always present. We shall discuss these two aspects in turn.

A major component of the immune system is the class of cells called **leukocytes**, otherwise known as white blood cells. Like all blood cells, they arise from common precursor cells (stem cells) in the bone marrow. Unlike other blood cells, however, they can leave the blood vessels and circulate in the lymphatic system. Lymphoid tissues (such as lymph nodes, the spleen, and, above all, the thymus gland) play important roles in the workings of the immune system.



Image copyright MANDY GODBEHEAR, 2010. Used under license from Shutterstock.com.

■ Allergic reactions arise when the immune system attacks innocuous substances. Allergies to plant pollens are common, producing well-known symptoms such as sneezing.

## Innate Immunity—The Front Lines of Defense

When one considers the tremendous numbers of bacteria, viruses, parasites, and toxins that our bodies have to deal with, it is a wonder that we are not continually sick. Most students learn about antibodies in high school, and these days everyone learns about T cells, because of their relationship to AIDS. However, there are many more weapons of immune defense than T cells and antibodies. In reality, you discover that you are sick only after a pathogen has managed to beat your front-line defense, which is called **innate immunity**.

Innate immunity has several parts. One part includes physical barriers, such as skin, mucus, and tears. All of these hinder penetration by pathogens and do not require specialized cells to fight the pathogen. However, if a pathogen, be it a bacterium, virus, or parasite, breaches this outer layer of defense, the cellular warriors of the innate system come into play. The cells of the innate immune system that we will discuss are **dendritic cells**, **macrophages**, and **natural killer (NK) cells**. One of the first and most important cells to join the fight are the dendritic cells, so called because of their dendrites, which are long, tentacle-like projections (see Figure 14.10). Dendritic cells are found in the skin, the mucous membranes, the lungs, and the spleen, and are the first cells of the innate system to have a crack at any virus or bacterium that wanders across their path. Using suction-cup-like receptors, they grab onto invaders and then engulf them by endocytosis. These cells then chop up the devoured pathogens and bring parts of their proteins to the surface. Here the protein fragments are displayed on a protein called a **major histocompatibility complex (MHC)**. The dendritic cells travel through the lymph to the spleen, where they present these antigens to other cells of the immune system, the **helper T cells** ( $T_H$  cells). Dendritic cells are members of a class of cells referred to as **antigen-presenting cells (APCs)**, and they are the starting point in most of the responses that are traditionally associated with the immune system. Once the dendritic cells present their antigens to the helper T cells, the latter release chemicals called **cytokines** that stimulate other members of the immune system, such as **killer T cells** (also called cytotoxic T cells or  $T_C$  cells) and **B cells**. Figure 14.11 shows the basics of the relationship between dendritic cells and other immune cells. There are two classes of MHC proteins (I and II), based on their structure and on what they bind to. MHC I binds to killer T cells, while MHC II binds to helper T cells. Besides the link between the two cells based on the MHC, another link (or perhaps two) is always necessary before cell proliferation occurs. The double signal is a trademark of most immune-cell responses, and it is thought to be one mechanism for making sure that the immune system is not activated in error.

Besides their basic role in presenting antigens to T cells and B cells, dendritic cells have recently been very popular with companies that are trying to generate antibodies to help fight cancer. Dendritic cells do have a downside, however. It was recently found that HIV uses a receptor on dendritic cells to hitch a ride in the lymph system until it can find a  $T_H$  cell. Some labs are working on chemicals to block this interaction in the hopes of slowing down HIV's travel through the body.

Another important cell type in the innate immunity system is the **natural killer (NK) cell** (Figure 14.12), which is a member of a class of leukocytes called **lymphocytes**, because it is derived from a type of stem cell called a *lymphoid stem cell*. NK cells kill off cells that have been infected by viruses or that are cancerous, and they secrete cytokines that call up other cells, such as macrophages, another innate-immunity cell type that destroys microbes. They also work with dendritic cells, in a sense. If an infection is small, the NK cells may end up killing off the infected dendritic cells before the rest of the immune system is activated. Thus, NK cells help decide whether the acquired-immunity system needs to be activated. NK cells are also important in fighting cancer. They are

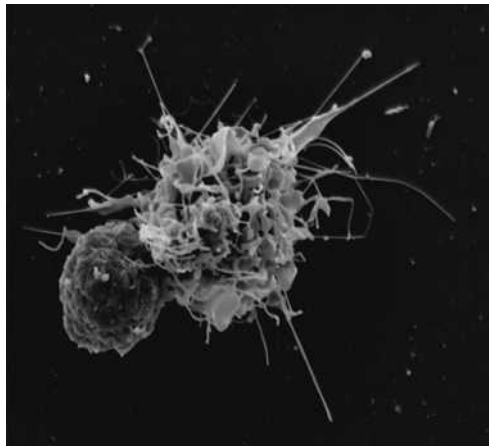


■ Edward Jenner developed the world's first vaccine in 1796. It was a safe and effective way to prevent smallpox and has led to eradication of this disease.



■ **FIGURE 14.10** Dendritic cells get their name from their tentacle-like arms. The one shown is from a human.





Courtesy of Steve Haley, William Bowers, and Richard Hunt, University of South Carolina Medical School

■ **FIGURE 14.11** Dendritic cells and the other cells of the immune system. This figure shows a rat dendritic cell interacting with a T cell. Through these interactions, the dendritic cells teach the acquired-immunity system what to attack.

stimulated by interferon, an antiviral glycoprotein, which was employed as one of the first treatments for cancer and the first protein to be cloned and expressed for human use (Chapter 13). Macrophages and other cells of the innate immune system have unfortunately also been identified as some of the big players in cancer, where they are a type of double-edged sword. Their presence can be a direct attack against cancer cells, but they also lead to inflammation, which has been shown to promote the progression of cancerous cells from a premalignant state to full proliferation.

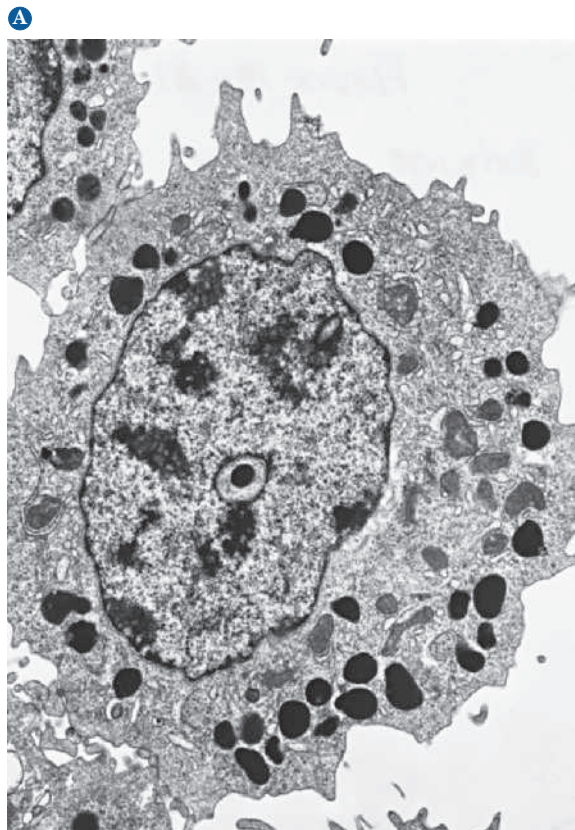
### Acquired Immunity: Cellular Aspects

Acquired immunity is dependent on two other types of lymphocytes: T cells and B cells. **T cells** develop primarily in the thymus gland and **B cells** develop primarily in the bone marrow, accounting for their names (Figure 14.13). Much of the cellular aspect of acquired immunity is the province of the T cells, whereas much of the molecular aspect depends on the activities of the B cells.

### What do T cells and B cells do?

#### T-Cell Functions

T cells can have a number of functions. As T cells differentiate, each becomes specialized for one of the possible functions. The first of these possibilities, that of the **killer T cells**, involves **T-cell receptors (TCRs)** on their surfaces that recognize and bind to **antigens**, the foreign substances that trigger the immune response. The antigens are presented to the T cell by antigen-presenting cells (APCs), such as macrophages and dendritic cells. The APCs ingest and process antigens, and then present them to the T cells. The processed antigen takes the form of a short peptide bound to an MHC I protein on the surface of the APC. Figure 14.14 shows how this works for macrophages. The macrophage

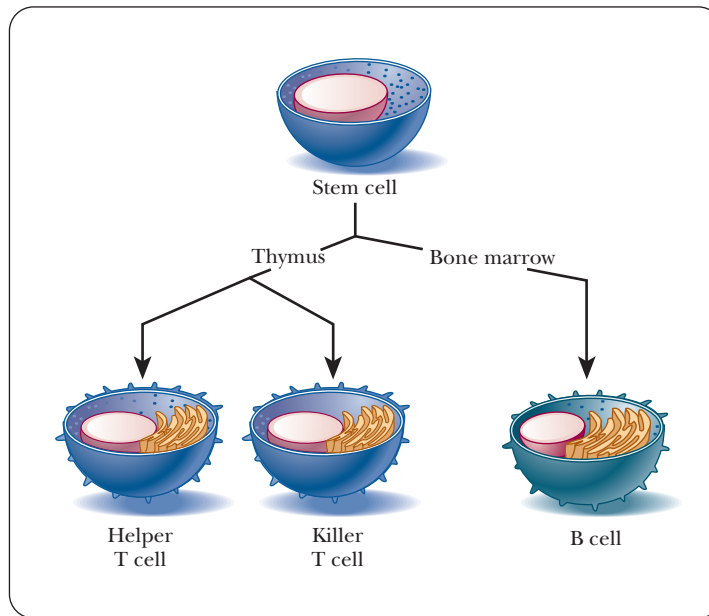


© Eye of Science/Photo Researchers, Inc.

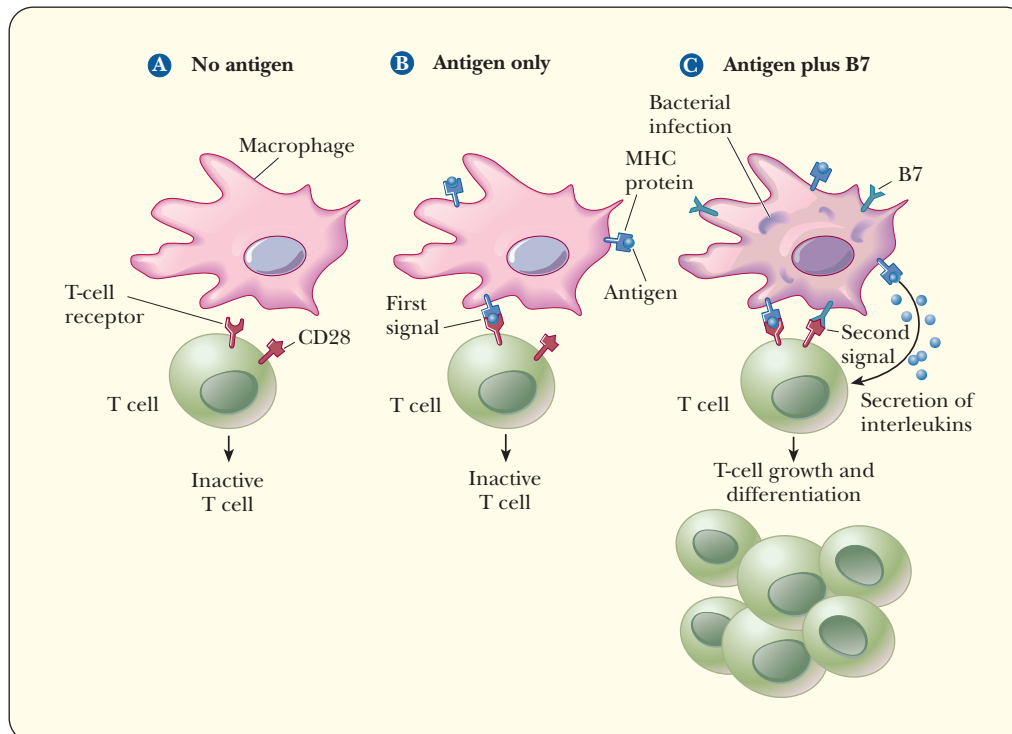


Courtesy of Ann M. Dvorak, M.D., Beth Israel Deaconess Medical Center, Boston, MA

■ **FIGURE 14.12** Natural killer (NK) cells are among the first cells to be involved in an immune response. They are nonphagocytic cells that can interact with and destroy other cells, such as those infected with a virus, or cancer cells. (a) An electron micrograph. (b) A high-resolution photomicrograph.



■ **FIGURE 14.13** The development of lymphocytes. All lymphocytes are ultimately derived from the stem cells of the bone marrow. In the thymus, two kinds of T cells develop: helper T cells and killer T cells. B cells develop in the bone marrow.



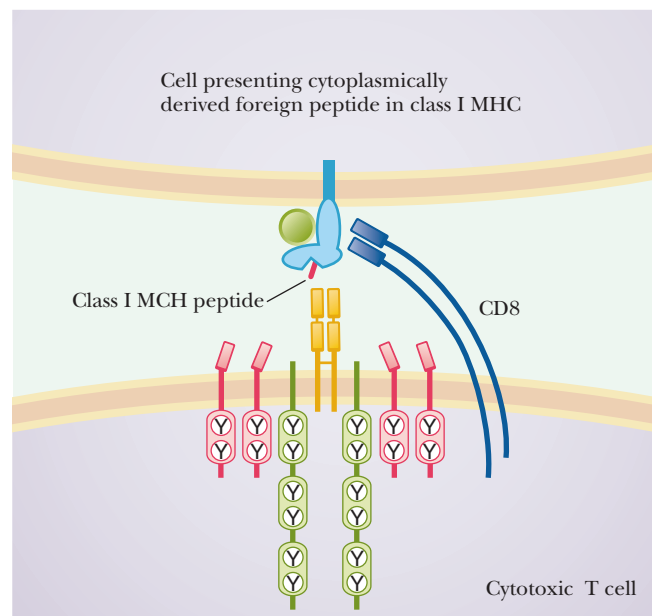
■ **FIGURE 14.14** A two-stage process leads to the growth and differentiation of T cells. (a) In the absence of antigen, proliferation of T cells does not take place. (b) In the presence of antigen alone, the T-cell receptor binds to antigen presented on the surface of a macrophage cell by the MHC protein. There is still no proliferation of T cells because the second signal is missing. In this way the body can avoid an inappropriate response to its own antigens. (c) When an infection takes place, a B7 protein is produced in response to the infection. The B7 protein on the surface of the infected cell binds to a CD28 protein on the surface of the immature T cell, giving the second signal that allows it to grow and proliferate. (Adapted from "How the Immune System Recognizes Invaders," by Charles A. Janeway, Jr.; illustration by Ian Warpole. *Sci. Amer.* 269 [3] [1993].)

also presents another molecule, a protein of a family known as B7, which binds to another T-cell surface protein called CD28; the exact nature of this B7 protein is a subject of active research. The combination of the two signals leads to T-cell growth and differentiation, producing killer T cells. Proliferation of killer T cells is also triggered when macrophages bound to T cells produce small proteins called **interleukins**. The T cells make an interleukin-receptor protein as long as they are bound to the macrophage but do not do so when they are no longer bound. Interleukins are part of a class of substances called cytokines. When we discussed innate immunity, we saw that this term refers to soluble protein factors produced by one cell that specifically affect another cell. In this way, T cells do not proliferate in uncontrolled fashion. A killer T cell also has another membrane protein called CD8, which helps it dock to the MHC of the antigen-presenting cell, as shown in Figure 14.15. In fact, the CD8 protein is such a distinguishing characteristic that many researchers use the term *CD8 cells* instead of *killer T cells*.

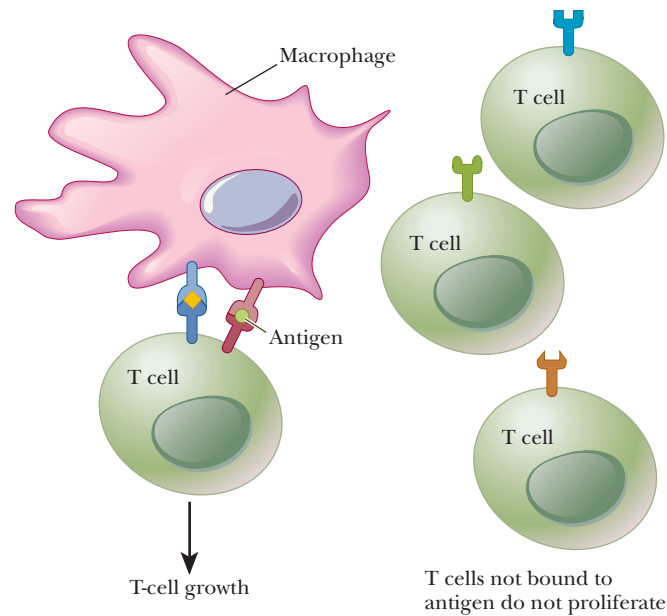
T cells that bind to a given antigen and *only to that antigen* grow when these conditions are fulfilled. Note the specificity of which the immune system is capable. Many substances, including ones that do not exist in nature, can be antigens. The remarkable adaptability of the immune system in dealing with so many possible challenges is another of its main features. The process by which only the cells that respond to a given antigen grow in preference to other T cells is called **clonal selection** (Figure 14.16). The immune system can thus be versatile in its responses to the challenges it meets. This clonal selection is the basis of the definition of acquired immunity. Most of the response of T cells stems from the rapid proliferation of cells once they are selected—the organism, in essence, acquiring these cells only when necessary. However, there must be at least one cell with the proper TCR to recognize the antigen and to bind to it. These receptors are not generated because there is a need; rather, they are generated randomly when stem cells differentiate into T cells. Fortunately, the diversity of T-cell receptors is so great that there are millions of TCR specificities.

The division of T cells during the peak of the immune response is very rapid, often reaching three or four divisions per day. This would lead to more than a thousandfold increase in the number of selected T cells in a few days.

As their name implies, killer T cells destroy antigen-infected cells. They do so by binding to them and by releasing a protein that perforates the plasma



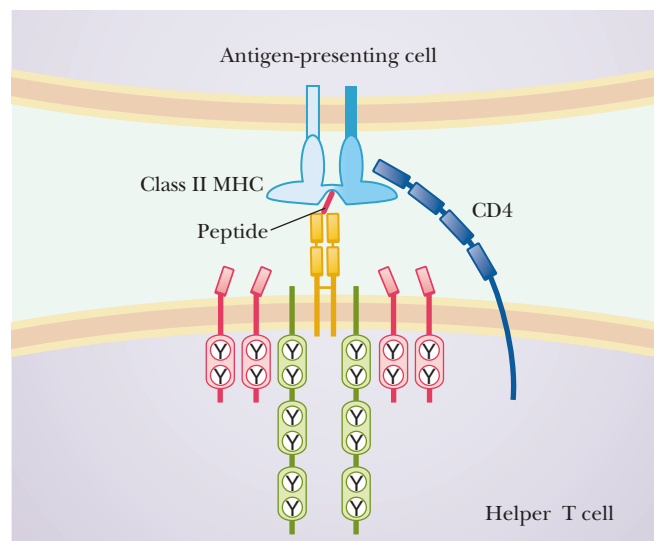
■ **FIGURE 14.15** Interaction between cytotoxic T cells (killer T cells) and antigen-presenting cells. Foreign peptides that are derived from the cytoplasm of infected cells are displayed on the surface by MHC I proteins. These bind to the T-cell receptor of a killer T cell. A docking protein called CD8 helps link the two cells.



■ **FIGURE 14.16 Clonal selection.** Clonal selection allows the immune system to be both versatile and efficient in responding to a wide range of possible antigens. Many different types of cells can be produced by the immune system, allowing it to deal with almost any possible challenge. Only cells that respond to an antigen that is actually present are produced in quantity; this is an efficient use of resources.

membranes of the infected cells. This aspect of the immune system is particularly effective in preventing the spread of viral infection by killing virus-infected host cells. In a situation such as this, the antigen can be considered all or part of the coat protein of the virus. When the infection subsides, some memory cells remain, conferring immunity against later attacks from the same virus.

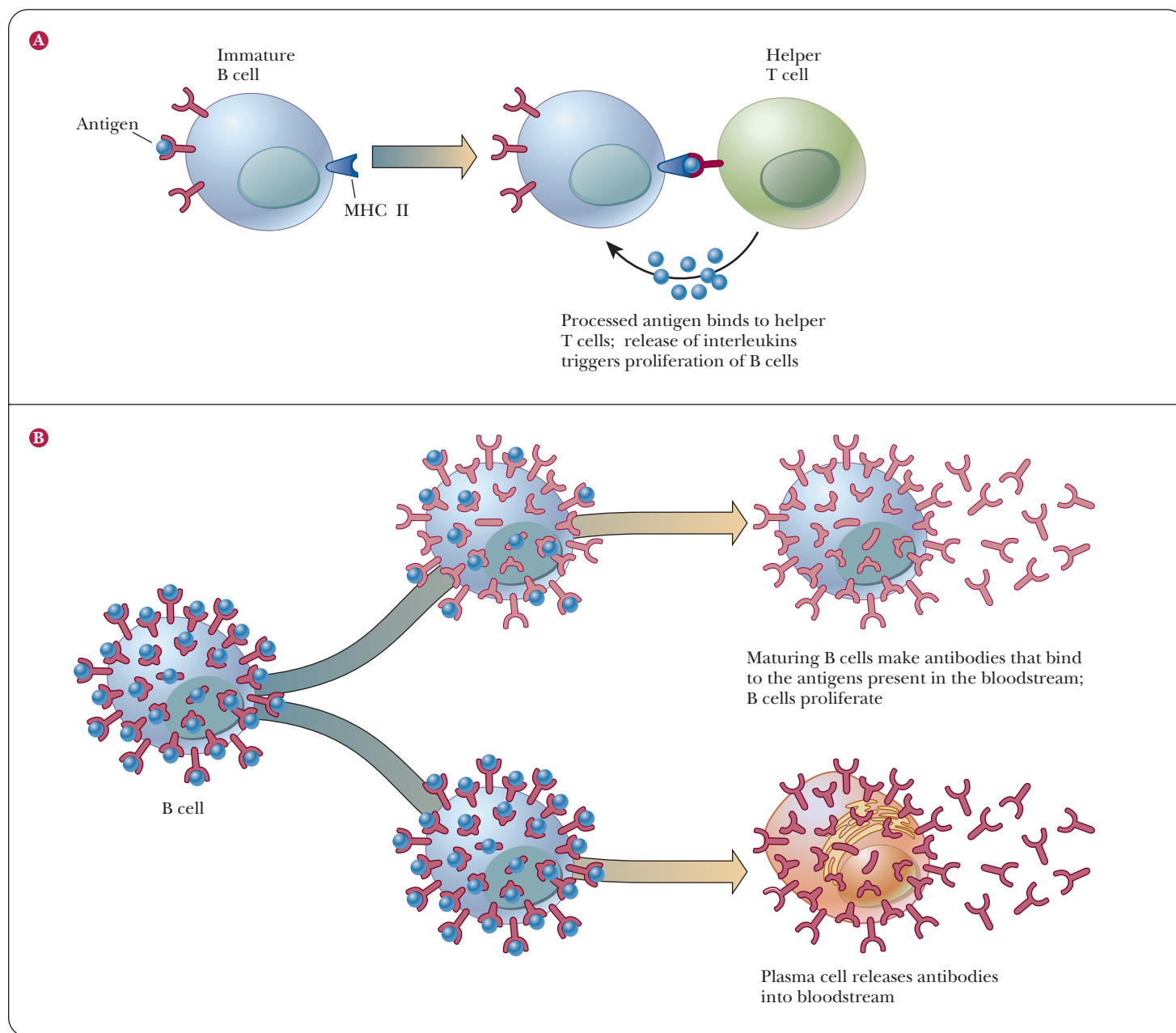
T cells play another role in the immune system. Another class of T cells develops receptors for a different group of antigen-presenting MHC proteins, in this case MHC II. These become **helper T cells**, which develop in much the same way as killer T cells. Helper T cells are also referred to as *CD4 cells*, because of the presence of that unique membrane protein. CD4 helps the cell dock to the MHC of the antigen-presenting cell, as shown in Figure 14.17. The function of helper T cells is primarily to aid in the stimulation of B cells. Maturing B cells display the MHC II protein, with processed antigen, on their surfaces. Note particularly that the MHC proteins play a key role in the immune system. This property has led to a considerable amount of research to determine their structure, including determination by X-ray crystallography. The MHC II of the B cells is the binding site for helper T cells. The binding of helper T cells



■ **FIGURE 14.17 Interaction between helper T cells and antigen-presenting cells.** Foreign peptides are displayed on the surface by MHC II proteins. These bind to the T-cell receptor of a helper T cell. A docking protein called CD4 helps link the two cells.



to B cells releases interleukins (IL-2 and IL-4) and triggers the development of B cells into plasma cells (Figure 14.18). Both B cells and plasma cells produce **antibodies** (also known as **immunoglobulins**), the proteins that will occupy most of our time as we discuss the molecular aspects of the immune response. B cells display antibodies on their surfaces in addition to the MHC II proteins. The antibodies recognize and bind to antigens. This property allows B cells to absorb antigens for processing. Plasma cells release circulating antibodies into the bloodstream, where they bind to antigen, marking it for destruction by the immune system. Helper T cells also help stimulate killer T cells and antigen-presenting cells via release of interleukins.



■ **FIGURE 14.18** Helper T cells aid in the development of B cells. (a) A helper T cell has a receptor for the MHC II protein on the surfaces of immature B cells. When the helper T cells bind to the processed antigen presented by the MHC II protein, they release interleukins and trigger the maturation and proliferation of B cells. (b) B cells have antibodies on their surfaces, which allow them to bind to antigens. The B cells with antibodies for the antigens that are present grow and develop. When B cells develop into plasma cells, they release circulating antibodies into the bloodstream. (Adapted from "How the Immune System Develops," by Irving L. Weissman and Max D. Cooper; illustrated by Jared Schneidman. *Sci. Amer.* September [1993].)



## T-Cell Memory

One of the major characteristics of the acquired-immunity system is that it exhibits memory. Although the system is slow to respond the first time it encounters an antigen, it is much quicker the next time. The process of generating T-cell memory involves the death of most of the T cells that were generated by the first infection with a particular antigen. Only a small percentage (5%–10%) of the original cells survive as memory cells. Still, this represents a much larger number than was present before the initial encounter with the antigen. These memory cells have a higher reproductive rate even in the absence of antigen than does a naïve T cell (one that has never encountered the antigen).

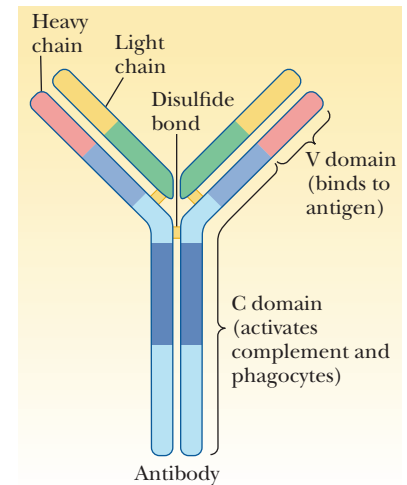
Several of the interleukins play key roles in these processes. Interleukin 7 is involved in maintenance of naïve killer T cells at low levels. When stimulated by antigens, the proliferation of T<sub>C</sub> cells is stimulated by interleukin 2. Memory T<sub>C</sub> cells, on the other hand, are maintained by interleukin 15.

T-cell memory is one place where killer T cells and helper T cells come together. Several researchers have recently shown that CD8 cells expand when confronted with the correct antigen in the absence of CD4 cells. However, CD8 cells that were clonally expanded without CD4 cells were unable later to make memory cells that were as active.

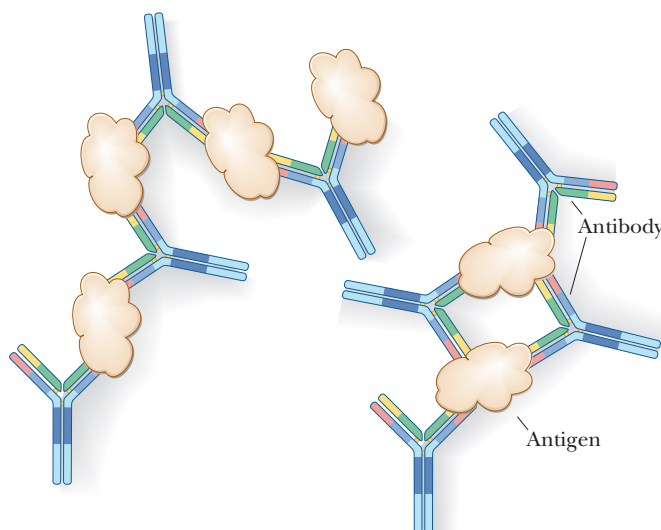
## The Immune System: Molecular Aspects

### What are antibodies?

Antibodies are Y-shaped molecules consisting of two identical heavy chains and two identical light chains held together by disulfide bonds (Figure 14.19). They are glycoproteins, with oligosaccharides linked to their heavy chains. There are different classes of antibodies, based on differences in the heavy chains. In some of these classes, heavy chains are linked to form dimers, trimers, or pentamers. Each light chain and each heavy chain has a constant region and a variable region. The variable region (also called the *V domain*) is found at the prongs of the Y and is the part of the antibody that binds to the antigen (Figure 14.20). The binding sites for the antibody on the antigen are called **epitopes**. Most antigens have several such binding sites, so that the immune system has several possible avenues of attack for naturally occurring antigens. Each antibody can bind to two antigens, and each antigen usually has multiple binding sites for antibodies, giving rise to a precipitate that is the basis of experimental methods for immunological research. The constant region (the *C domain*) is located at the hinge and the stem of the Y; this part of the antibody is recognized by phagocytes and by the complement system (the portion of the immune system that destroys antibody-bound antigen).



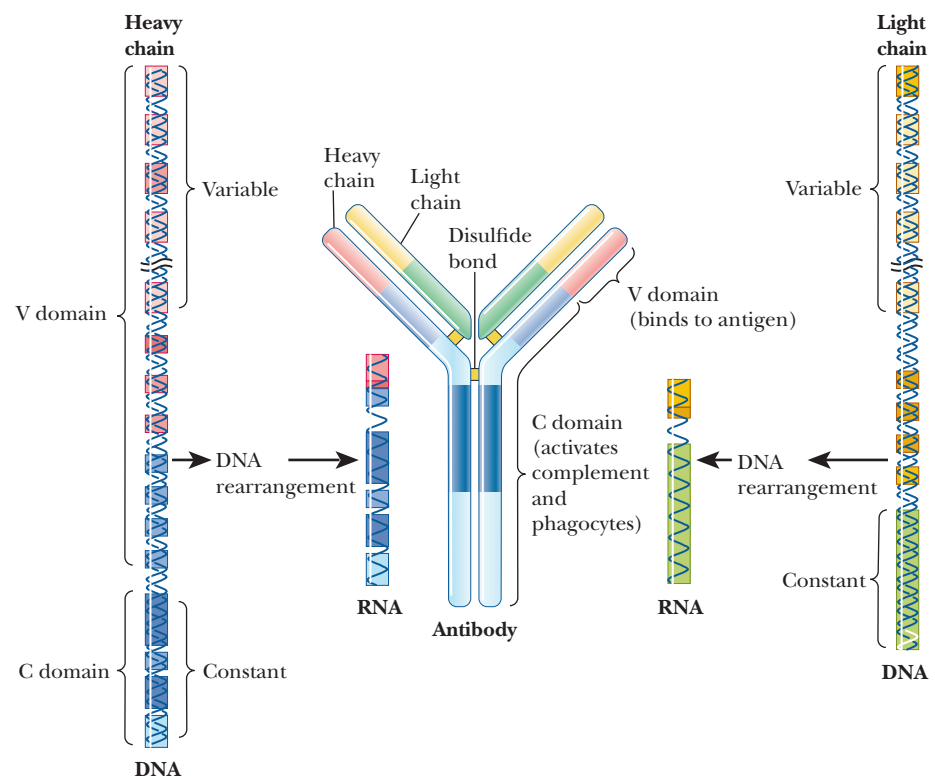
■ **FIGURE 14.19 Antibodies.** A typical antibody molecule is a Y-shaped molecule consisting of two identical light chains and two identical heavy chains linked by disulfide bonds. Each light chain and each heavy chain has a variable region and a constant region. The variable region, which is at the prongs of the Y, binds to antigen. The constant region, toward the stem of the Y, activates phagocytes and complement, the parts of the immune system that destroy antibody-bound antigen. (Adapted from “How the Immune System Recognizes Invaders,” by Charles A. Janeway, Jr.; illustration by Ian Warpole. *Sci. Amer.* September [1993].)



■ **FIGURE 14.20 An antigen–antibody reaction forms a precipitate.** An antigen, such as a bacterium or virus, typically has several binding sites for antibodies. Each variable region of an antibody (each prong of the Y) can bind to a different antigen. The aggregate thus forms a precipitate and is attacked by phagocytes and the complement system.

How does the body produce so many highly diverse antibodies to respond to essentially any possible antigen? The number of possible antibodies is virtually unlimited, as is the number of words in the English language. In a language, the letters of the alphabet can be arranged in countless ways to give a variety of words, and the same possibility for enormous numbers of rearrangements exists with the gene segments that code for portions of antibody chains. Antibody genes are inherited as small fragments that join to form a complete gene in individual B cells as they develop (Figure 14.21). When gene segments are joined, the enzymes that catalyze the process add random DNA bases to the ends of segments being spliced, allowing for the wide variety observed experimentally. This rearrangement process takes place in the genes for both the light and the heavy chains. (Bear in mind that the exon splicing and mRNA processing that we discussed in Chapter 11 still takes place as well.) In addition to these factors, B lymphocytes have a particularly high rate of somatic mutation, in which changes in the base sequence of DNA occur as the cell develops. Changes outside the germ cells apply only to the organism in which they take place and are not passed on to succeeding generations.

Each B cell (and each progeny plasma cell) produces only one kind of antibody. In principle, each such cell should be a source of a supply of homogeneous antibody by cloning. This is not possible in practice because lymphocytes do not grow continuously in culture. In the late 1970s, Georges Köhler and César Milstein developed a method to circumvent this problem, a feat for which they received the Nobel Prize in physiology in 1984. The technique requires fusing lymphocytes that make the desired antibody with mouse myeloma cells. The resulting **hybridoma** (hybrid myeloma), like all cancer cells, can be cloned in culture (Figure 14.22) and produces the desired antibody. Because the clones are the progeny of a single cell, they produce homogeneous **monoclonal antibodies**. In this way, it is possible to produce antibodies to almost any antigen in quantity. Monoclonal antibodies can be used to assay for biological substances that can act as antigens. A striking example of their usefulness is in



■ **FIGURE 14.21** Antibody heavy and light chains. The heavy and light chains of antibodies are encoded by genes that consist of a number of DNA segments. These segments rearrange and, in the process, give rise to genes for different chains in each B cell. Because the joining is highly variable, comparatively few gene segments give rise to millions of distinct antibodies. (*Adapted from "How the Immune System Recognizes Invaders," by Charles A. Janeway, Jr.; illustration by Ian Warpole. Sci. Amer. September [1993].*)

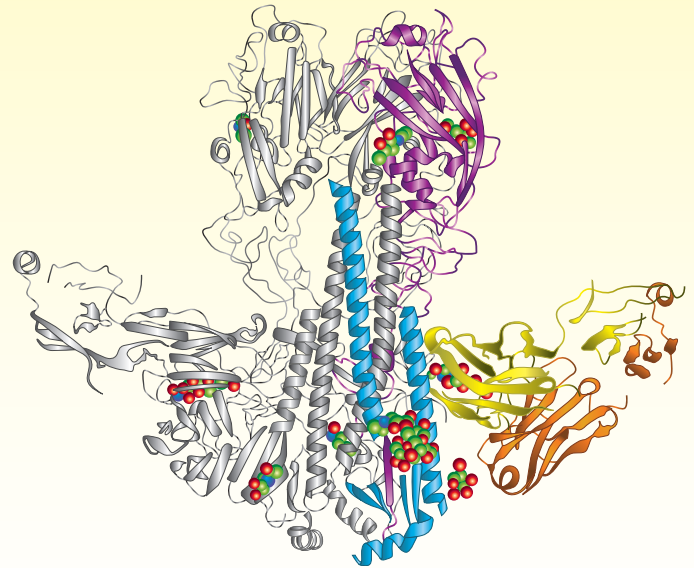
# Biochemical Connections EPIDEMIOLOGY

## Modern Science Takes on the Flu Virus

Besides generating vaccines against the flu virus (see Biochemical Connections on page 400), scientists are also trying to generate monoclonal antibodies that will fight the disease. Monoclonal antibodies are expensive to make, and in many cases the attempt is frustrated by the quickly changing nature of the virus. Therefore, researchers have been looking for an antibody that will attack part of the virus that does not change. In early 2009 two independent teams reported that they had created antibodies that would react with a portion of the flu virus's hemagglutinin (HA) protein. The good news was that the part of the hemagglutinin protein bound by the antibodies is relatively constant and does not change between strains.

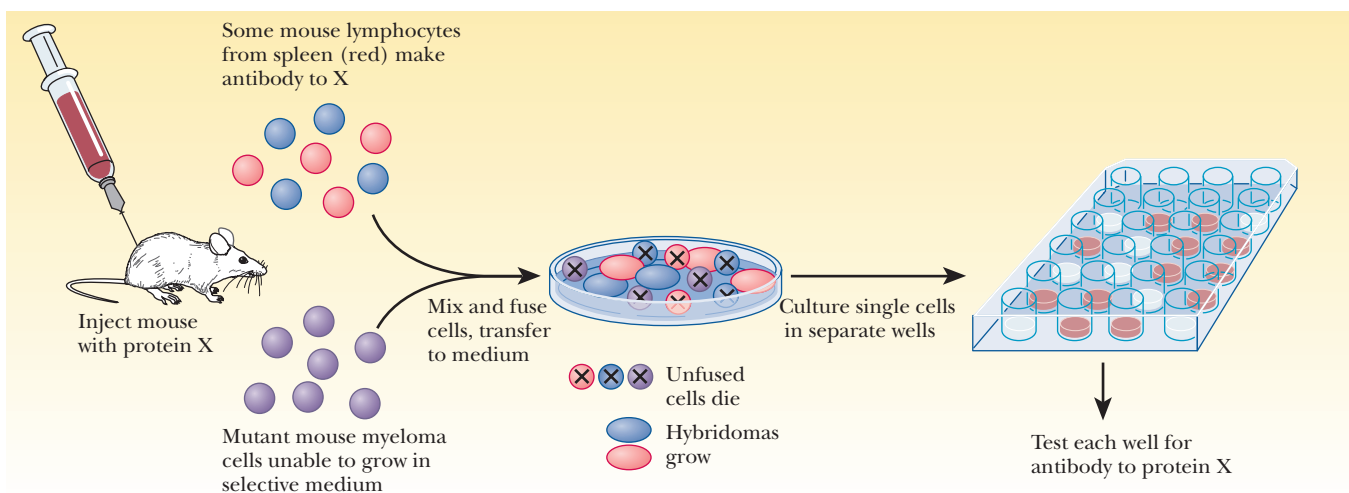
The teams identified 10 different antibodies that recognized the H5 subtype found in the avian flu, and found they would also block 8 of the 15 other HA types. They tested their antibodies in mice both before and after they were dosed with lethal quantities of avian flu. Most of the rodents survived, indicating that these antibodies would work as prevention or as cure.

The medical community is excited about the prospects of adding another weapon to the eternal fight against the flu. Such antibodies can be used to provide an immediate passive immunity to people who do not respond well to vaccines, such as the elderly or those whose immune systems are compromised. They will also allow a strong countermeasure to an impending pandemic. The usual downside to such new discoveries is the cost. There are not enough cheap vaccines to meet the needs around the world, and many economically deprived countries have trouble getting them, especially an expensive monoclonal antibody.



(From *Flu Antibodies Stir New Hope for Treatment* by Mitch Leslie (27 February 2009) *Science* 323 (5918), 1160. Used with permission of AAAS.)

Still, many feel governments that can afford them would be wise to stockpile some of these new antibodies as protection against the next pandemic.



■ **FIGURE 14.22** A procedure for producing monoclonal antibodies against a protein antigen X. A mouse is immunized against the antigen X, and some of its spleen lymphocytes produce antibody. The lymphocytes are fused with mutant myeloma cells that cannot grow in a given medium because they lack an enzyme found in the lymphocytes. Unfused cells die because lymphocytes cannot grow in culture, and the mutant myeloma cells cannot survive in this medium. The individual cells are grown in culture in separate wells and are tested for antibody to protein X.

testing blood for the presence of HIV; this procedure has become routine to protect the public blood supply.

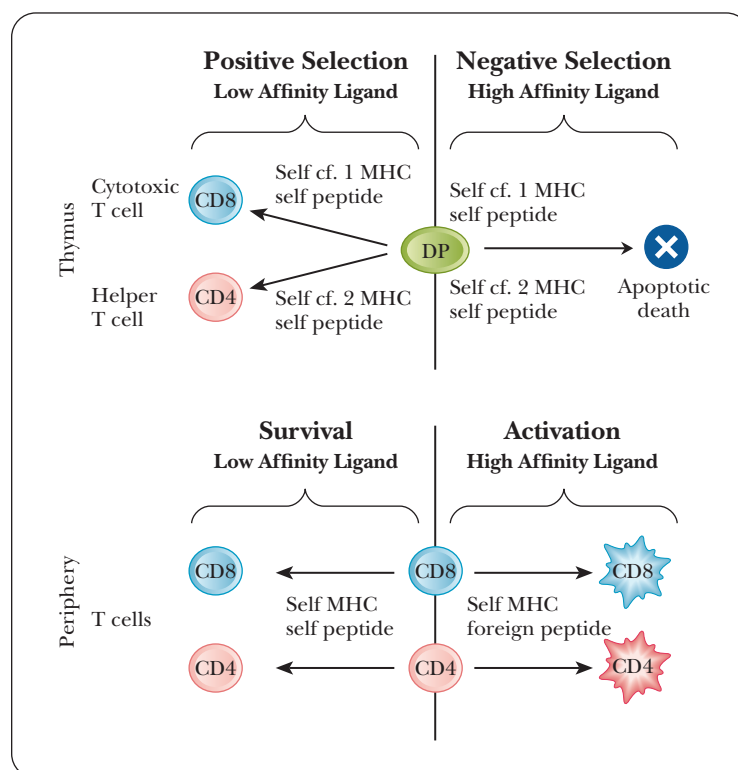
Recently, scientists have gone against the dogma of “one antigen—one antibody” and created designer antibodies that recognize two different proteins. These “two in one” antibodies can be useful tools against diseases that do not respond well to single treatments. For example, cancer and AIDS are often attacked with multiple treatments anyway, and creating such designer antibodies adds another weapon to the arsenal.

## Distinguishing Self from Nonself

With all the power the immune system has to attack foreign invaders, it must also do so with discretion, because we have our own cells that display proteins and other macromolecules on their surfaces. How the immune system knows not to attack these cells is a complicated and fascinating topic. When the body makes a mistake and attacks one of its own cells, the result is an **autoimmune disease**, examples of which are rheumatoid arthritis, lupus, multiple sclerosis, scleroderma, Crohn’s disease, and some forms of diabetes.

T cells and B cells have a wide variety of receptors on their surfaces. The affinities for a given antigen vary greatly. Below a certain threshold, an encounter between a lymphocyte receptor and an antigen is not sufficient to trigger that cell to become active and begin to multiply. These same cells also have stages of development. They mature in the bone marrow or the thymus and go through an early stage in which receptors first begin to appear on their surfaces.

In the case of T cells, a precursor form called a **DP cell** has both the CD4 and the CD8 protein. This cell is the turning point for the fate of its progeny. If the receptors of the DP cell do not recognize anything, including self-antigens or self-MHC proteins, then it dies by neglect. If the receptors recognize self-antigens or MHC but with low affinity, then the cell undergoes positive selection and differentiates into a killer T cell or a helper T cell, as shown in Figure 14.23. On the other hand, if the cell’s receptors encounter self-antigens



■ **FIGURE 14.23** Differentiation of T cells. A precursor to T cells called a DP cell is the turning point in the fate of T-cell progeny. If the DP cell reacts to nothing, including self-antigens or MHCs, then it dies by neglect (not shown). If it recognizes self-antigens or MHCs with high affinity, then it is programmed for apoptosis to avoid an autoimmune response. If it recognizes self-antigens or MHCs with low affinity, it differentiates into killer T cells and helper T cells. (Reprinted with permission from “Signaling Life and Death in the Thymus: Timing Is Everything,” by G. Werlen, B. Hausmann, D. Naeher, and E. Palmer. *Science* **299**, 1859–1863. Copyright © 2003 AAAS.)

that are recognized with high affinity, it undergoes a process called **negative selection** and is programmed for apoptosis, or cell death.

By the time the lymphocytes leave their tissue of origin, they have therefore already been stripped of the most dangerous individual cells that would tend to react to self-antigens. Some individual cells will still have a receptor with very low affinity for a self-antigen. If these slip out of the bone marrow or thymus, they do not initiate an immune response because their affinity is below the minimum threshold, and there is always the requirement for a secondary signal. They would need to have another cell, such as a macrophage, also present them with an antigen. In the case of a B cell, besides binding an antigen to its receptor, it would need to receive an interleukin 2 from a helper T cell that had also been stimulated by the same antigen.

## Biochemical Connections VIROLOGY

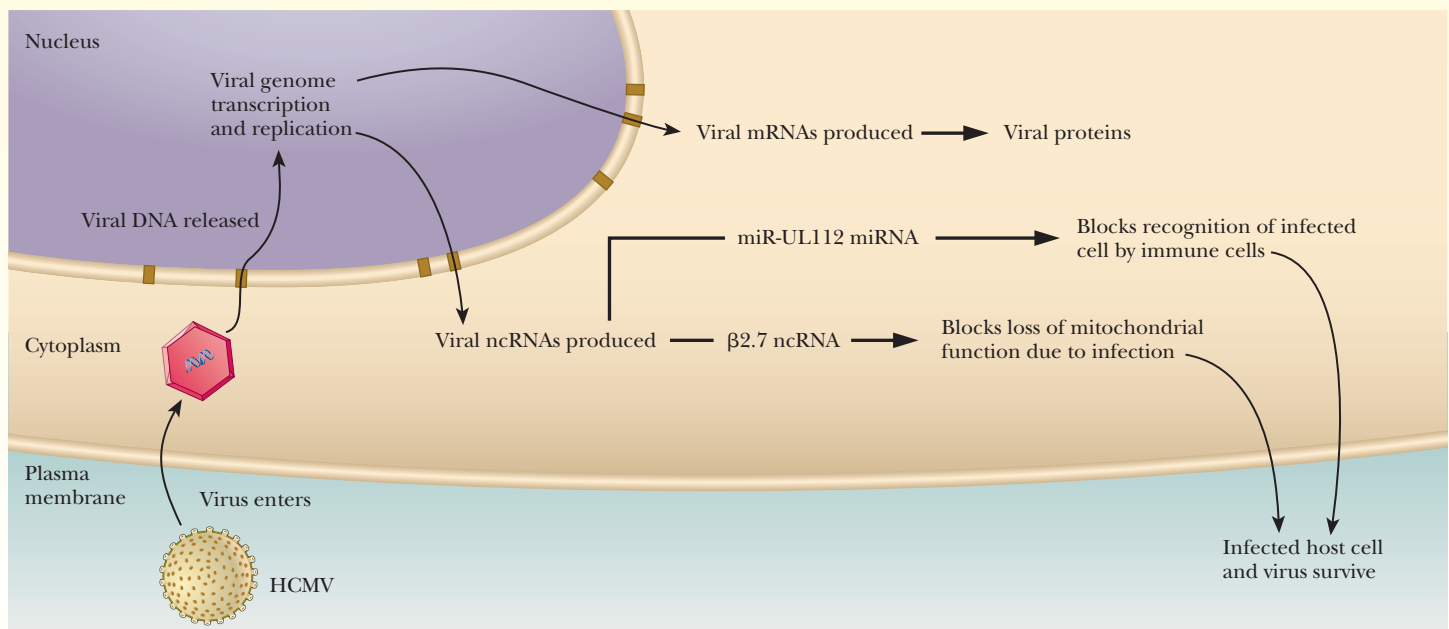
### Viral RNAs Outwit the Immune System

Herpes viruses are pathogenic DNA viruses known for establishing long-term latent infections. Some forms of herpes are sexually transmitted diseases and lead to genital lesions. Others are known for leading to “fever blisters” around the lips. Once the virus infects the cell, its viral genome is transcribed into viral mRNA and other smaller RNA molecules, collectively known as noncoding RNA or ncRNA. Some of the ncRNAs are very small, micro RNAs (Chapter 9), while others are longer at about 100 nucleotides. As we saw in Chapter 11, miRNA can selectively inhibit gene expression. Until recently, the function of the longer ncRNAs was less clear.

Researchers studying a herpes virus called human cytomegalovirus (HCMV) recently showed that one of the functions of the longer ncRNA from the virus was to allow the infected cells to evade the innate immunity response, as shown in the figure. RNAs make excellent weapons against the immune system. They are fast-acting because they do not need to be translated, and

they are poor targets for the adaptive immune response. HCMV is a herpes virus that causes severe disease in newborn infants and immunocompromised individuals. It encodes at least 2 long ncRNAs and 11 miRNAs. Current research suggests that these two ncRNAs are immune-response inhibitors.

Within hours of HCMV infection, a 2.7-kb viral ncRNA ( $\beta 2.7$ ) accumulates, reaching 20% of the total viral RNA. This RNA binds to components of the mitochondrial respiratory chain complex I (MRCC-I), stabilizing its function as a countermeasure to apoptosis. Thus,  $\beta 2.7$  prevents the premature death of the infected cell and the steady production of ATP during the viral life cycle. The virus also produces an miRNA, miR-UL112. This miRNA inhibits the production of mRNA for a cell surface ligand that attracts natural killer cells. Thus, these two RNA molecules inhibit the destruction of the cell by NK cells as well as promoting the general strength of the host cell to prevent cell death.



■ **Viral RNA shuts down the immune response.** Infection by HCMV induces a state of metabolic stress that would normally trigger host immune responses, including cell lysis or apoptosis (cell death). HCMV prevents these immune responses using two viral ncRNAs. The  $\beta 2.7$  ncRNA binds to components of the mitochondrial respiratory chain complex I (MRCC-I) and stabilizes mitochondrial energy production. miR-UL112 miRNA blocks the expression of cell surface ligand, MICB, which attracts natural killer cells and hence prevents lysis by these immune cells that would recognize MICB. (Reprinted by permission of Science magazine from “Outwitted by Viral RNAs,” Science 317, 329 [2007] by Bryan R. Cullen.)



All of these safeguards lead to the delicate balance that must be maintained by the immune system, a system that simultaneously has the diversity to bind to almost any molecule in the Universe but does not react to the myriad proteins that are recognized as self.

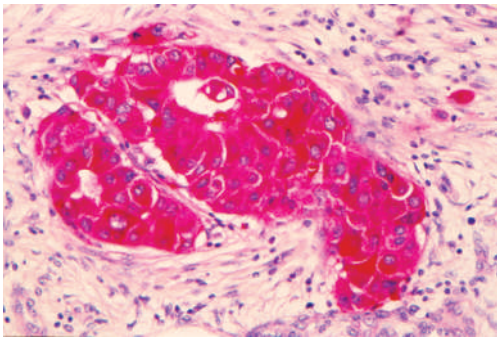
Autoimmune diseases have a wide range of effects on the lives of the afflicted. Patients with lupus and type 1 diabetes have a high rate of survival and the chance to lead a relatively normal life. Others, such as those with multiple sclerosis or scleroderma, experience much more tragic effects, and the prospects of long-term survival are not as good. Recently, doctors have come up with some radical techniques, almost desperation attempts, at cures for such diseases. One of the most radical is to replace the patient's immune system completely. Intense chemotherapy and radiation are used to destroy the immune system. Then the patient is given bone marrow transplants from matching donors in an attempt to "reboot" the immune system. To date about 1500 people have received such treatments. So far the results have been mixed. About one third of the participants went into remission and no longer needed continued treatments. This had not been possible with more standard approaches to their autoimmune disease. Another third benefited, but only for a year or two before suffering a relapse, and about a third did not respond at all. Between 1% and 5% of the participants died from the procedure.

## 14.4 Cancer

Cancer is one of the leading causes of death in humans, leading to 1500 deaths per day in the United States alone. It is characterized by cells that grow and divide out of control, often spreading to other tissues and causing them to become cancerous. Some estimates suggest that a third of all humans will get cancer during their lifetimes, so it is clearly a disease that is important for everyone to understand. However, the older a person gets, the more likely he or she is to get cancer. A 70-year-old is about 100 times more likely to get cancer than a 20-year-old.

### What characterizes a cancer cell?

All life-threatening cancers have at least six characteristics in common, and multiple problems must occur in a cell before it becomes cancerous. That may be why, even though cancer is common, most people still grow to old age and do not get cancer. First, cancer cells continue to grow and divide in situations in which normal cells do not. Most cells must receive a growth-chemical signal, but cancer cells manage to keep growing without such signals. Second, cancer cells continue to grow even when the neighboring cells send out "stop-growth" signals. For example, normal cells stop growing when compressed by other cells. Somehow, tumors manage to avoid this. (Figure 14.24 shows a tumor cell expanding and squeezing against neighboring tissue.) Third, cancer cells manage to keep going and avoid a "self-destruct" signal that usually occurs when DNA damage has occurred. Fourth, they can co-opt the body's vascular system, causing the growth of new blood vessels to supply the cancerous cells with nutrients. Fifth, they are essentially immortal. Normal cells can divide only for a finite number of times, usually in the 50–70 range. However, cancer cells and tumors can divide far more than that. The sixth characteristic is the most lethal: Although cells that exhibit the first five characteristics can be a problem, the fact that cancer cells have the ability to break loose, travel to other parts of the body, and create new tumors makes them lethal. This process is called **metastasis**. Stationary tumors can often be removed by surgery. However, once a cancer starts spreading, it is almost impossible to stop. Of every 10 deaths due to cancer—including a high percentage of lung, colon, and breast cancers—9 of them are due to cancers that metastasized.



© Michael Abbey/Photo Researchers, Inc.

■ **FIGURE 14.24** **Tumor cells.** Normal cells stop growing when they are squeezed by other cells. This tumor cell continues dividing and growing even though it is being squeezed by the adjacent tissue.

## Biochemical Connections GENETICS

### Cancer: The Dark Side of the Human Genome

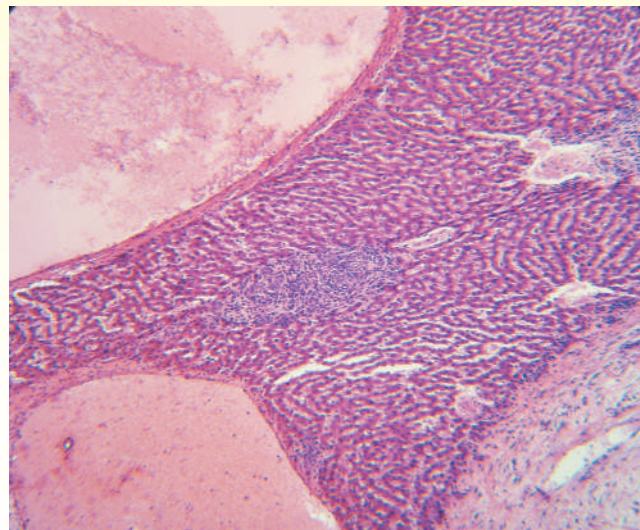
In a very real sense, cancer is the most common genetic disease. Somatic mutations (ones that do not affect the germ cells passed on to following generations) constantly take place in the human body, and they accumulate as a person ages. When these mutations reach a critical number in essential locations on the genome, cancer develops. The number of such mutations can easily number well into the thousands.

In December 2009, scientists at the Wellcome Trust Sanger Institute in the United Kingdom announced that they had completed sequencing the genome of melanoma (skin cancer) and lung cancer. They found more than 30,000 mutations in the melanoma genome and more than 23,000 in lung cancer. Sequencing of the breast cancer genome is in progress, with groups around the world working on cancers of the liver, stomach, brain, ovary, and pancreas.

In the words of one of the scientists involved, “This is a fundamental moment in cancer research.” It will now be possible to determine the factors that led to the development of the cancer. As an example, it is definite that most of the mutations associated with melanoma arise from excessive exposure to the sun. Likewise, smoking causes most of the DNA errors in lung cancer. Researchers estimate that a new mutation occurs with every 15 cigarettes smoked. Most of the mutations do not cause harm, but some will be in DNA sequences where alterations lead to cancer.

This information will make it possible to diagnose cancer much earlier and lead to more effective treatment. For individual patients, it will be possible to see which drugs are likely to be effective in treating the cancer and which ones are not. Even partial knowledge of the breast cancer genome has already led to the drug Herceptin, a chemotherapeutic agent that deals with the results of a gene that is overexpressed. The gene in question, HER2, encodes a receptor protein on the cell surface. In normal cells, the receptor binds to a signaling compound to activate

processes within a cell. In cancer, activation takes place without a signal, leading to the uncontrolled cell proliferation characteristic of cancer. Needless to say, Herceptin has no effect in cancers in which the gene in question is not overexpressed. It can even be harmful because of its toxicity, especially to the heart. This drug is but one example of the benefits that will come from knowledge of the cancer genome. Many breakthroughs can be expected in the next few years.



■ Electron micrograph of a cancer cell.

Image copyright Jubal Harshaw, 2010. Used under license from Shutterstock.com.

### What causes cancer?

One often hears of many things that cause cancer. Smoking or eating grilled meat cause cancer. Radiation or asbestos cause cancer. However, these things cannot truly be the ultimate cause, although they may play a role. The real cause may be a combination of insults to the cell that leads to it turning malignant. Cancer is ultimately a DNA disease. It has its roots in changes in the DNA inside a cell. Somehow these changes cause the loss of control of division and the other characteristics described above.

The changes in the DNA cause changes to specific proteins that are responsible for controlling the cell cycle. Most mutations of DNA affect two types of genes. The first is called a **tumor suppressor**, a gene that makes a protein that restricts the cell's ability to divide. If a mutation damages the gene for a tumor suppressor, then the cell has lost its brakes and divides out of control. The second type of gene, called an **oncogene**, is one whose protein product stimulates growth and cell division. Mutations of the oncogene cause it to be permanently active. Scientists are still looking for changes to genes that are direct causes of cancer. So far, more than 100 oncogenes and 15 tumor-suppressor genes have been linked to cancer. Current research involves compiling databases of all the known DNA sequences that are found to be mutated in many varieties of cancer. This approach has identified about 350 cancer-related genes. The task now is to pare this number down to the mutations that are important to the

development of the disease, because it is probable that many of these mutations are coincidental rather than causal.

**Oncogenes**

An oncogene is a gene that has been implicated in cancer. The root word, *onco*, means “cancer.” In 1911, a scientist named Peyton Rous demonstrated that solutions taken from chicken carcinomas could infect other cells. This was the first discovery of tumor viruses, and Rous was given the Nobel Prize in 1966 for his discovery. The virus was called the **Rous sarcoma virus**, and it was the first retrovirus shown to cause cancer. The gene that was specific to the cancer is called **v-src**, for *viral sarcoma*. This gene encodes a protein that causes transformation of the host cell into a cancer cell. Thus, the gene was given the name *oncogene*. The protein was called **pp60<sup>src</sup>**, which stands for a phosphoprotein of 60,000-Da molecular weight from the sarcoma virus (*src*).

However, it was later found that the sequence of the gene was very similar to that of a normal gene in eukaryotes. These genes are called **proto-oncogenes**. Many proto-oncogenes are normal and necessary for proper growth and development in eukaryotic cells. However, some transforming event causes the proto-oncogene to lose control. Sometimes this is due to a viral infection. In other cases, the event that causes a proto-oncogene to become an oncogene is not known. Table 14.2 shows some proto-oncogenes implicated in human tumors. Many of these genes are involved in signal-transduction pathways that affect the transcription of genes that speed up cell division. In Chapter 11, we looked at the control of transcription in eukaryotes and noted that many signaling pathways were routed through the CBP/p300 coactivator (see Figure 11.23). One of these pathways involved **mitogen activated protein kinase (MAPK)** and a transcription factor called **AP-1**. To understand the nature of many of the oncogenes shown in Table 14.2, we must take another look at this pathway.

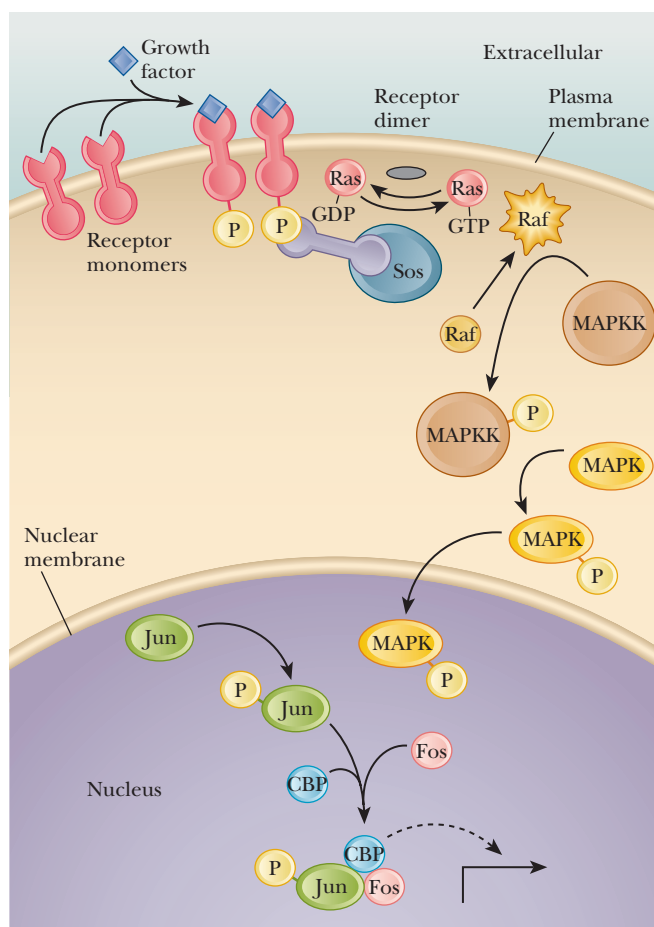
The process starts when an extracellular signal binds to a receptor on the cell membrane (see Figure 14.25). This receptor is a tyrosine kinase that dimerizes, and then each part phosphorylates the other. Once phosphorylated, the receptors are bound by an adaptor molecule, a protein called **GRB2** (pronounced “grab two”), which has a phosphotyrosine binding domain that is very similar to a domain found in the pp60<sup>src</sup> protein. The other end of GRB2 binds to a protein called **Sos**.

At this point, there is an interaction with a very important 21-kDa protein. This protein, called **p21<sup>ras</sup>** or just **Ras**, is involved in about 30% of human tumors. The designation *Ras* comes from *Rat sarcoma*, the original tissue in which it was discovered. The Ras family of proteins are GTP-binding proteins. In their resting

**TABLE 14.2**

A Representative List of Proto-Oncogenes Implicated in Human Tumors	
Proto-Oncogene	Neoplasm(s)
<i>abl</i>	Chronic myelogenous leukemia
<i>erbB-1</i>	Squamous cell carcinoma; astrocytoma
<i>erbB-2 (neu)</i>	Adenocarcinoma of breast, ovary, and stomach
<i>myc</i>	Burkitt's lymphoma carcinoma of lung, breast, and cervix
H- <i>ras</i>	Carcinoma of colon, lung, and pancreas; melanoma
N- <i>ras</i>	Carcinoma of genitourinary tract and thyroid; melanoma
<i>ros</i>	Astrocytoma
<i>src</i>	Carcinoma of colon
<i>jun</i>	} Several
<i>fos</i>	

Adapted from Bishop, J. M. 1991, Molecular themes in oncogenesis, *Cell* **64**: 235–248.



■ **FIGURE 14.25 MAP kinase signal transduction.** Signal transduction starts when a growth factor (blue) binds to a receptor monomer (red) on the cell membrane. The receptor is a tyrosine kinase, which then phosphorylates its partner receptor. The phosphorylated receptor is then recognized by GRB2 (light purple), which binds to the Ras exchanger Sos (blue). Sos is activated to exchange GDP for GTP on Ras (pink), activating it. Ras moves Raf (tan) to the cell membrane, where it becomes active. Raf phosphorylates MAP kinase kinase, which then phosphorylates MAP kinase (yellow). MAP kinase (MAPK) enters the nucleus and phosphorylates Jun (light green). Phosphorylated Jun binds to Fos and CBP and transcription is activated. (Reprinted by permission from *Molecular Biology*, by R. F. Weaver, 2nd ed., p. 375, McGraw-Hill.)

state, they are bound to GDP. After the cell signal, the Sos replaces the GDP for GTP. Intrinsic hydrolysis of the GTP returns the protein to its inactive state, but this process is slow. Proteins known as GTPase-activating proteins (GAPs) speed up this hydrolysis and are involved in the control of the Ras proteins. GAPs inactivate Ras by accelerating the hydrolysis of GTP. Oncogenic forms of Ras have impaired GTPase activity and are insensitive to GAPs, thus leaving them bound to GTP, which causes them to stimulate cell division continually.

Although Ras mutations have been some of the most-studied mutations leading to cancer, we can see that Ras is found rather early in the process that ultimately leads to cell division. Activated Ras attracts another protein called **Raf**, which then phosphorylates serines and threonines on **mitogen-activated protein kinase kinase (MAPKK)**. As one can guess from its name, this enzyme then phosphorylates mitogen-activated protein kinase (MAPK). This enzyme enters the nucleus and phosphorylates a transcription factor called **Jun**. Jun binds to another transcription factor called **Fos**. Together, Jun and Fos make up the transcription factor that we saw before called AP-1, which binds to CBP and stimulates the transcription of genes that lead to rapid cell division. As we can see in the table, *jun* and *fos* oncogenes code for these proteins. In 2002, researchers screening 20 different genes in 378 cancer samples found that the Raf gene was mutated in 70% of the malignant melanoma samples.

## Tumor Suppressors

Many human genes produce proteins called **tumor suppressors**. Tumor suppressors inhibit transcription of genes that would cause increased replication. When a mutation occurs in any of these suppressors, replication and division become uncontrolled and tumors result. Table 14.3 lists some human tumor-suppressor genes.



TABLE 14.3

Representative Tumor-Suppressor Genes Implicated in Human Tumors	
Tumor-Suppressor Gene	Neoplasm(s)
<i>RBI</i>	Retinoblastoma; osteosarcoma; carcinoma of breast, bladder, and lung
<i>p53</i>	Astrocytoma; carcinoma of breast, colon, and lung; osteosarcoma
<i>WT1</i>	Wilms' tumor
<i>DCC</i>	Carcinoma of colon
<i>NF1</i>	Neurofibromatosis type 1
<i>FAP</i>	Carcinoma of colon
<i>MEN-1</i>	Tumors of parathyroid, pancreas, pituitary, and adrenal cortex

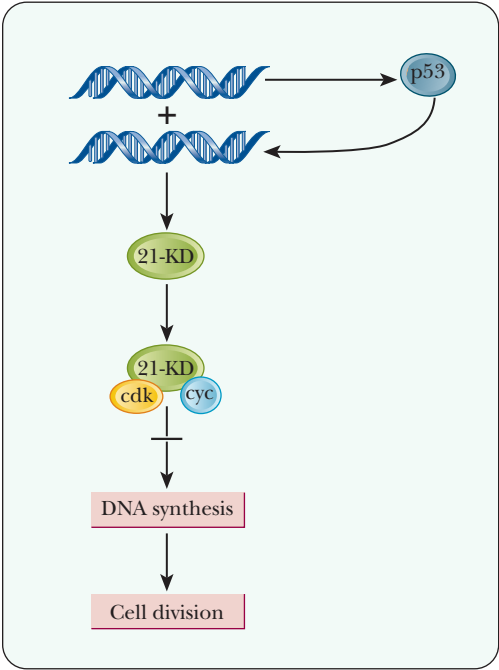
Adapted from Bishop, J. M. 1991. Molecular themes in oncogenesis. *Cell* 64, 235–248.

A 53-kDa protein designated **p53** has become the focus of feverish activity in cancer research. Mutations in the gene that codes for p53 are found in more than half of all human cancers. When the gene is operating normally, it acts as a tumor suppressor; when it is mutated, it is involved in a wide variety of cancers. By the end of 1993, mutations in the *p53* gene had been found in 51 types of human tumors. The role of p53 is to slow down cell division and to promote cell death (apoptosis) under certain circumstances, including when DNA is damaged or when cells are infected by viruses.

It is known that p53 binds to the basal transcription machinery (one of the TAFs bound to TFIID; see Chapter 11). When cancer-causing mutations occur in p53, it can no longer bind to DNA in a normal fashion. The mode of action of p53 as a tumor suppressor is twofold. As shown in Figure 14.26, it is an activator of RNA transcription; it “turns on” the transcription and translation of several genes. One of them, *Pic1*, encodes a 21-kDa protein, **P21**, that is a key regulator of DNA synthesis and thus of cell division. The P21 protein, which is present in normal cells but is missing from (or mutated in) cancer cells, binds to the enzymes known as *cyclin-dependent protein kinases* (CDKs), which, as their name implies, become active only when they associate with proteins called cyclins. Recall from Section 10.6 that cell division depends on the activity of cyclin-dependent kinases. Some of the oncogenes seen above work in such a way that the result is an overproduction of the CDK proteins, which keeps the cells dividing continuously. Normal levels of p53 protein cannot turn these genes off in cancer cells, but they could do this in normal cells. In normal cells, the result is that the cell cycle remains in the state between mitosis (in which cells divide) and the replication of DNA for the next cell division. DNA repair can take place at this stage. If the attempts at DNA repair fail, the p53 protein may trigger apoptosis, the programmed cell death characteristic of normal cells, but not of cancer cells.

The important point is that two different mechanisms are operating here. One is analogous to the brakes failing in your car (inadequate or defective p53 protein) and the other (overproduction of CDKs) is equivalent to the accelerator sticking in the open position—two opposite mechanisms with the same result: the car crashes.

A number of factors come together in explaining the variety of diseases we call cancer. Mutations of DNA lead to changes in the proteins that control cell growth, by either directly causing cell division or allowing it to occur by default. Still other mutations interfere with DNA repair. The possibility of finding new cancer therapies—and perhaps even cancer cures—is enhanced by understanding these contributing factors and how they affect each other.



■ **FIGURE 14.26** Action of p53. The p53 protein turns on the production of a 21-kDa protein. This protein binds to complexes of cyclin-dependent kinases (CDKs) and cyclins. The result of binding is inhibition of DNA synthesis and cell growth. (Adapted from Science, Figure 1, Vol. 262, 1993, p. 1644, by K. Suttiff, © 1993 by the AAAS.)



## Viruses and Cancer

The original work by Rous showed how viruses could cause cancer in certain situations. The close homology between the oncogene sequence found in some viruses and the proto-oncogene sequences in the mammalian genome has led many researchers to theorize that the oncogenes may have been of mammalian origin. In the course of repeated infections and travels, the virus may pick up pieces of DNA from a host and deliver another piece of DNA to a host. In the course of the rapid mutation that occurs in retroviruses, these proto-oncogenes could be mutated to a form that is oncogenic.

Retroviruses that cause cancer in humans are known; some forms of leukemia (caused by HTLV-I and HTLV-II, which infect T cells of the immune system) are well-known examples, as well as cervical cancer caused by cervical papillomavirus. Theoretically, any retrovirus that inserts its DNA into the host chromosome could accidentally disable a tumor-suppressor gene or enable an oncogene by insertion of a strong promoter sequence near a proto-oncogene. One of the biggest fears of using *in vivo* delivery techniques for human gene therapy (see Section 14.2) is that the viral DNA inserted into a human chromosome might become incorporated into an otherwise healthy tumor-suppressor gene. This would potentially solve one of the individual's problems by providing a functional gene he or she was missing while causing an even greater problem. This happened, unfortunately, in 2003, when researchers in France were using viral-gene therapy to treat patients with X-linked SCID (see Section 14.2). In 9 of 11 cases, the viral-gene therapy restored the immune systems of the patients. However, two patients developed leukemia. It was later found that the virus had inserted itself, in each case, near a gene that has been found to be a leukemia oncogene. This was a tragic setback in viral-gene therapy, and now many government agencies are discussing the future of such therapy.

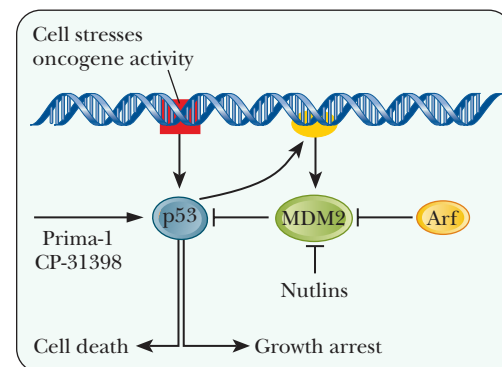
## How do we fight cancer?

Cancer can be treated in a variety of ways. The more traditional approaches include surgeries to remove tumors, radiation and chemotherapy to kill cancerous cells, and treatment with monoclonal antibodies that target specific tumors.

One of the most current focus points for research is the attempt to reactivate p53 in the cancerous tissues that have lost this function. Because this single gene was found to be the culprit in so many types of cancer, this became an obvious strategy. Clinical trials with mice have shown that in tumors that have lost p53 function, restoring p53 activity stops tumor growth and even shrinks the tumors. Remember that p53 has a two-pronged attack on tumor cells—it arrests the growth of the cell and it promotes cell death (apoptosis) of the cell. Many of the early trials involved specific delivery of an active p53 gene via gene therapy (Section 14.2). However, such delivery is impractical for human patients in many cases. The current focus is to find drugs that increase the levels of p53. Figure 14.27 shows the points of attack of the most recent drug candidates. Two drugs, *Prima-1* and *CP-31398*, reactivate mutant p53, perhaps by helping it fold more correctly than its mutated form. Another type of drug, called *nutlins*, inhibits a protein called MDM2, which is a natural inhibitor of p53. As has often been the case with cancer research, scientists and doctors must be very careful when tinkering with processes involved in cell growth. In some studies, reactivating p53 in certain laboratory animals had fatal effects as the reactivation caused a generalized cell death that went beyond the targeted tumors.

## Viruses Helping Cure Cancer

As we have seen in this chapter, viruses come in many types and cause many diseases. Viruses can be very specific to a single cell type because they rely on a protein receptor on the cell to gain entry. Liver cells display receptors that nerve cells do not, and vice versa. Oncologists (doctors who treat cancer) have



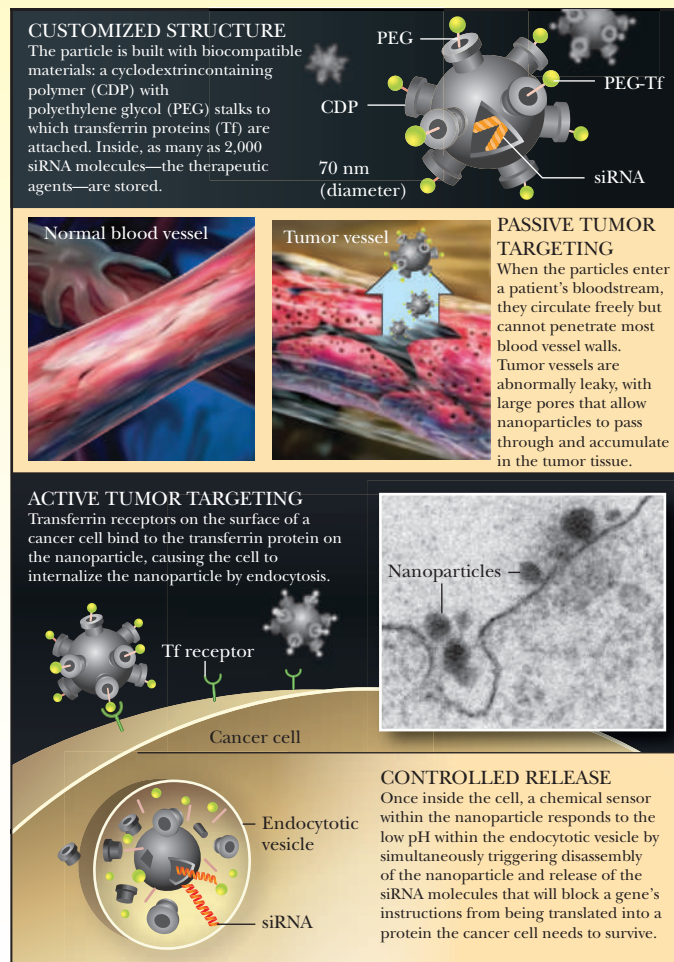
**FIGURE 14.27 Drug targets in the p53 pathway.** Some of the drugs being developed to fight cancer, including Prima 1 and CP-31398, aim to reactivate mutant p53 proteins, possibly by helping them fold more normally. Others, such as nutlins, were designed to foster p53 activity by preventing its interaction with its natural inhibitor MDM2. (Reprinted by permission of Science magazine, "Recruiting the Cell's Own Guardian for Cancer Therapy," *Science* **315**, 1211–1213 [2007] by Jean Marx.)

# Biochemical Connections BIOTECHNOLOGY

## Nanotech Tackles Cancer

The explosion of technology in the last 10 years has made it possible to envision medical procedures that are faster, cheaper, and more effective than more traditional treatments. One of the goals of cancer treatments is to minimize the damage done to healthy tissues and cells while being able to target the cancer cells. With traditional radiation or chemotherapy, this has been difficult. As described at the end of this section, using viruses to deliver treatment to specific targets is one approach that is very promising. Another approach that is being studied and is in clinical trials for several cancers is the use of **nanoparticles**. These artificial particles can be made in various sizes and materials, and can be filled with the specific molecule to be delivered to the target. The interesting sizes are those between 10 and 100 nm. Particles smaller than 10 nm are rapidly removed from the body. Particles over 100 nm cannot access the cells we wish to target. However, particles constructed to be in between in size can access cancer cells better than healthy cells due to the nature of cancer. Most tumors and cancer cells have vasculature that is damaged and porous. Therefore, as nanoparticles travel through the bloodstream, they will continually pass by healthy tissue, but then can leak out the pores into the cancerous tissues, as shown in the figure to the right.

The nanoparticles are created to have molecules on their surface that will be attracted to the cancer cells. In this case it is a polyethylene glycol molecule linked to a transferin protein. The particle is made up of a synthetic polymer and filled with a specific siRNA. The particles move through the blood and enter the cancer cells as shown. The cancer cells have receptors for the transferin protein and bind to them. The nanoparticles are then brought inside by endocytosis, whereupon they degrade and release their siRNA contents. Clinical trials are underway with several different nanoparticle types, and this technology shows tremendous promise for improving the quality of life of cancer patients undergoing treatment.

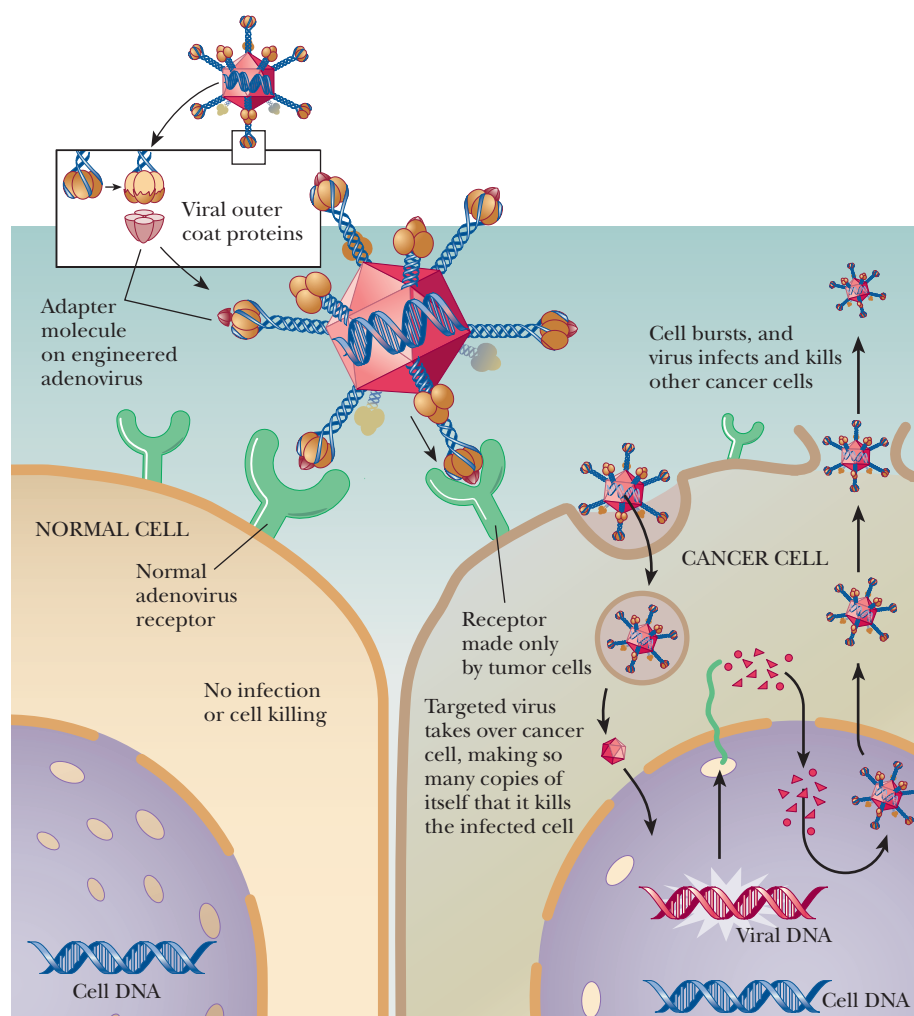


(Reprinted with permission. Copyright © 2009 Scientific American, a division of Nature America, Inc. All rights reserved.)

treated cancer for years with techniques such as radiation therapy and chemotherapy. These techniques attempt to target cancer cells, but, in the end, destroy other cells as well. In some sense, the goal of chemotherapy is to kill the cancer before the treatment kills the patient. If doctors could come up with a treatment that would be completely specific to cancer cells, it would go a long way, both toward stopping the cancer and toward making the patient's life more comfortable during the treatment. This was another chance for researchers to find something helpful about viruses.

In the 1990s a new type of treatment for cancer, called **virotherapy**, was begun. This technique was shown to target human tumor cells grafted onto mice. The treatment eliminated the human tumors. The virus of choice was an adenovirus, which we saw in the section on gene therapy. There are two strategies for virotherapy. One is to use the virus to attack and kill the cancer cell directly. The second is to have the virus ferry in a gene to the cancer cell that makes the cell more susceptible to a chemotherapy agent.

One of the biggest challenges in virotherapy is to make sure the virus specifically targets the cancer cell. The common adenovirus is not specific for cancer cells, so, in order to use it for virotherapy, other techniques must also be



■ **FIGURE 14.28** Transductional targeting in virotherapy. Viruses, such as adenovirus, are used to infect and destroy cancer cells selectively. Spikes on the adenovirus are mutated so that they recognize unique receptors on cancer cells. The virus selectively infects and lyses the cancer cells. (© 2003 Terese Winslow.)

employed. One of these is called **transductional targeting**. In this technique, antibodies are attached to the virus. These antibodies are created so that they target the cancer cell (see Figure 14.28). In this way, the normally indiscriminant adenovirus attacks only cancer cells. Once inside, the virus reproduces and eventually lyses the cell.

Another approach is called **transcriptional targeting**, shown in Figure 14.29. With this technique, the replication genes for adenovirus are placed after a promoter that is specific for a cancer cell. For example, skin cells make much more of the pigment melanin than other cells. Therefore, the genes for enzymes that make melanin are turned on more often in skin cells than in other cells (see Figure 14.29). Adenovirus can be engineered to have the promoter for the melanin-producing enzyme near the genes for virus replication. In cancerous skin cells, these promoters are triggered more often, so adenovirus replicates much quicker in skin-cancer cells, killing them specifically. Similar techniques have been used to target liver-cancer cells and prostate-cancer cells.

The other basic strategy is to have the virus ferry in a gene that makes the cancer cell more susceptible to chemotherapy. One such system uses a virus that targets rapidly dividing cells. Inside these cells, and only in these cells, the gene carried by the virus converts an innocuous pro-drug into an anticancer drug. These viruses are sometimes called “smart viruses” for their ability to select only the cancer cells. They then allow drugs to be used that are not harmful to normal cells.

The Biochemical Connections box on next page describes one other approach to fighting cancer.

# Biochemical Connections

## IMMUNOLOGY AND ONCOLOGY

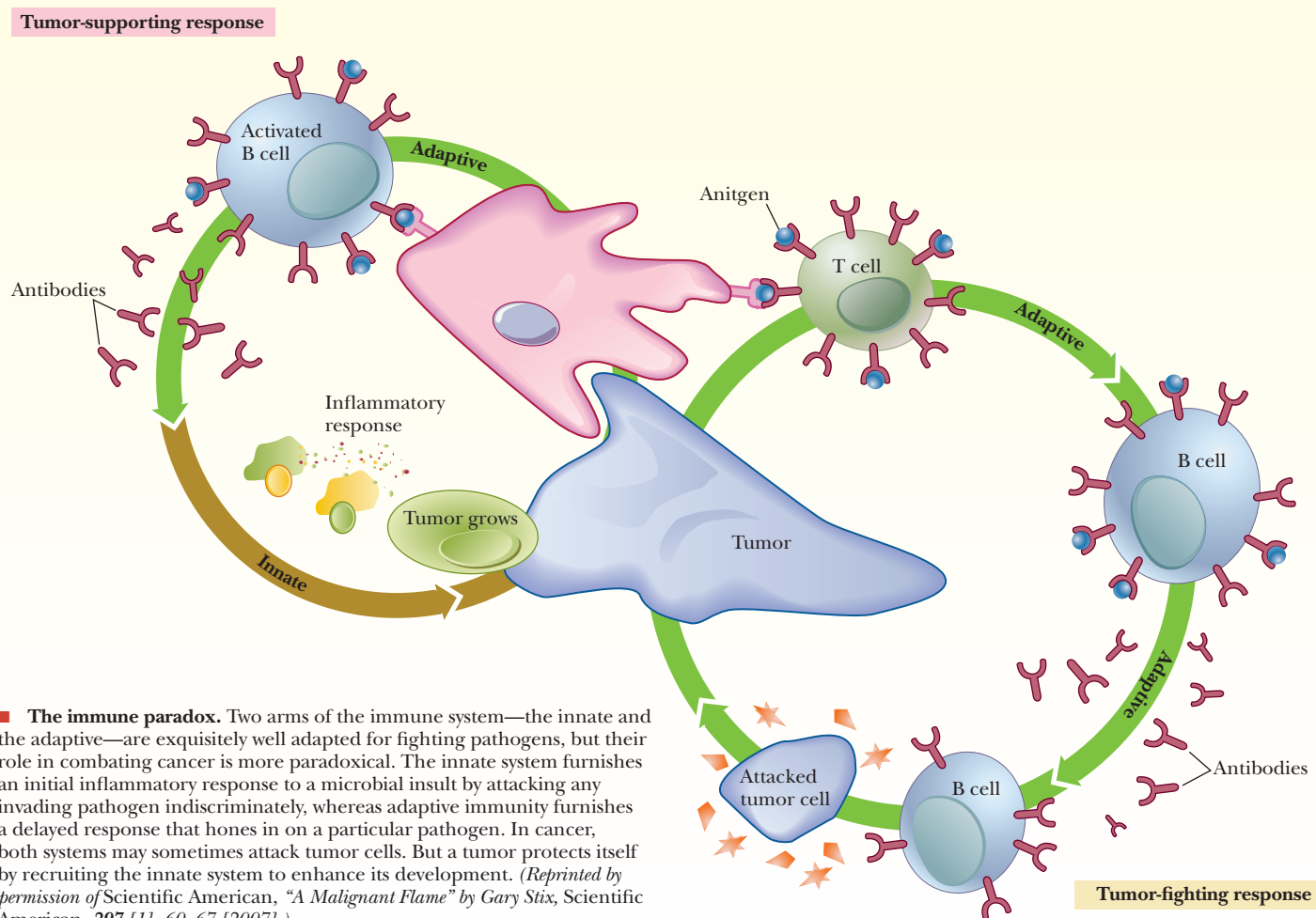
### Attacking the Symptoms instead of the Disease?

In the previous section we looked at innate immunity. This aspect of the immune system was the poor cousin to acquired immunity for decades, but has recently graduated to a much higher status because of its relationship to inflammation. Many of the processes involving macrophages and other innate immunity cell types lead to inflammation. Inflammation is known to be the underlying contributor to almost every chronic human disease known, including arthritis, Crohn's disease, diabetes, heart disease, Alzheimer's disease, and stroke. It has recently been shown that inflammation is also linked to cancer.

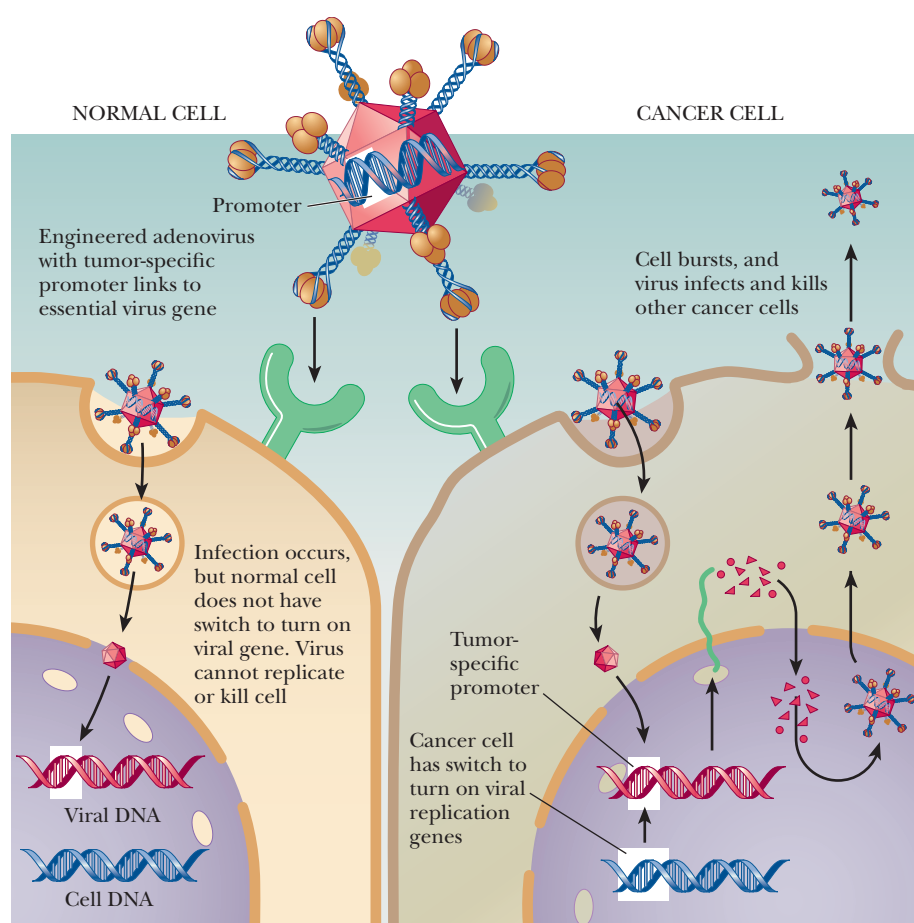
Cancer begins with a series of genetic changes that leads to cells overreplicating and then progressing to a stage where cells break off and set up new cancerous colonies at other sites. However, the stages in development are controlled by many factors, and researchers are now realizing that invasive cancers require cells of the innate immune system. These cells normally come to our rescue to protect against damage and disease, but in cancer they can be subverted for the good of the tumor and the ill of the patient. As shown in the figure, cells of the innate system can be both beneficial and detrimental to the fight against cancer.

As antigen-presenting cells, dendritic cells or macrophages stimulate T cells and B cells, which lead to an attack on a tumor cell. However, they also lead to an inflammatory response that leads to tumor growth through production of cytokines and growth factors. The evidence indicates that the inflammation surrounding macrophages encourages the conversion of precancerous tissue to full malignancy.

This realization has led researchers to question the dogmatic search for a cancer cure. Perhaps more time and money should be spent on treating the symptoms. Although no cure for AIDS has been found, much progress has been made in increasing the life span and quality of life of people with AIDS. Cancer researchers have always looked for cures, but if they can treat the inflammation symptom, then they might be able to stop the progression of the disease, essentially turning cancer into a manageable long-term disease. Patients rarely die from the primary cancer; rather they die from metastasis. Current research is looking at drug therapies including simple use of nonsteroidal anti-inflammatory drugs (NSAIDs) such as aspirin, more selective inhibition of prostaglandin  $E_2$ , and specific inhibition of cytokines such as IL-6 and IL-8.







■ **FIGURE 14.29 Transcriptional targeting in virotherapy.** A tumor-specific promoter is engineered into the adenovirus. The adenovirus infects many cells, but it is activated to replicate only in the cancer cell. (© 2003 Terese Winslow.)

## SUMMARY

**Why are viruses important?** Viruses are known to cause many diseases, and they may be very specific to a particular species and cell type. Viruses enter the cell by binding to specific receptors on the cell. Once inside the cell, the virus may replicate, form new viruses, and burst the cell. The virus may also hide its DNA by incorporating it into the host's DNA. Viruses have been linked to cancer, and they can be used as delivery agents in gene therapy. The most infamous virus these days is HIV, which causes AIDS.

**What is the structure of a virus?** At the center of a virus is its nucleic acid. This is surrounded by a protein coat called a capsid. The combination of the nucleic acid and the capsid is called the nucleocapsid. Many viruses also have a membrane envelope surrounding the nucleocapsid. Some also have protein spikes that help the virus attach to a host cell. Viruses have various shapes. Some are rod shaped, like the tobacco mosaic virus. Others have a hexagonal shape, like the bacteriophage T2 virus.

**How does a virus infect a cell?** A common method of attachment involves the binding of one of the spike proteins on the envelope of the virus to a specific receptor on the host cell.

**Why are retroviruses important?** Retroviruses have a genome based on RNA. When they infect a cell, their RNA is turned into DNA. The DNA is then incorporated into the host's DNA genome as part of the replication cycle for the virus. The most infamous retrovirus is human immunodeficiency virus, HIV, the causative agent of the disease AIDS. Retroviruses have also been linked to certain cancers. Retroviruses are also used in several forms of gene therapy.

**How does the immune system work?** One type of immunity, called innate immunity, consists of physical barriers, such as skin, and cellular warriors, such as dendritic cells. This system is always present and waiting to attack invading organisms or even cancerous cells. Another type of immunity, called acquired immunity, is based on two types of T cells (killer T cells and helper T cells) and on B cells. These cells are generated randomly with receptors that can be specific for an unimaginable number of antigens. When these cells encounter their specific antigens, they are stimulated to multiply, exponentially increasing the number of cells that can fight the invading organism. Acquired immune cells also leave behind memory cells so that if the same pathogen is seen again, the body is faster to eliminate it.



Immune cells must also be able to recognize self from nonself. T cells and B cells are conditioned, in their early stages of development, not to recognize proteins from that individual. In some cases, this system breaks down, and a person may be attacked by his or her own immune system, which may lead to an autoimmune disease.

**What do T cells and B cells do?** T cells have a number of functions. As they differentiate, they form killer T cells or helper T cells. Killer T cells bind to antigen-presenting cells, such as macrophages. This causes the T cell to proliferate. The T cell also secretes chemicals that destroy the antigen-presenting cell. Helper T cells also bind to antigen-presenting cells, but instead of destroying these cells, they stimulate B cells. B cells produce soluble antibodies that attack foreign antigens.

**What are antibodies?** Antibodies are Y-shaped proteins composed of two chains, a light chain and a heavy chain. There is a constant region and a variable region. The variable region is the part that binds specifically to an antigen. Antibodies bind to antigens, creating a reaction that forms a precipitate of the antibody–antigen complexes. This precipitate is then attacked by phagocytes and the complement system.

**What characterizes a cancer cell?** Cancer cells amplify external growth signals or produce their own. They are insensitive to antigrowth signals emitted by other cells. Cancer cells can

avoid apoptosis when other cells would be triggered to self-destruct. They can replicate indefinitely; normal cells can replicate 50–70 times before dying. Cancer cells emit chemical signals that promote blood vessel growth in their area to deliver oxygen to the cancerous tissues. They defy multiple signals that normally hold cells in place, thus traveling and thriving in other locations in the body.

**What causes cancer?** Cancer is a genetic disease. The ultimate cause is a mutation in a DNA sequence, whether this is caused by an inherited condition or tissue damage from chemicals or radiation. Most cancers have been linked to specific genes called oncogenes or to tumor-suppressor genes. When these genes are mutated, the cell loses the ability to control its replication. Although DNA damage starts the cancer, evidence suggests that inflammation associated with the innate immune system triggers many cancers to progress.

**How do we fight cancer?** There are many classical ways to fight cancer, such as radiation therapy and chemotherapy. Both of these are very hard on healthy cells and, therefore, on the patient. Novel techniques using viruses are now being tried to target cancer cells more directly, and some of these are showing tremendous promise. Because more than half of the known cancers involve a mutated and nonfunctional p53 gene, restoring the activity of p53 is a focus of cancer research today.

## REVIEW EXERCISES

 Interactive versions of these problems are assignable in OWL

### 14.1 Viruses

- Recall** What is the genetic material of a virus?
- Recall** Define the following:
  - virion
  - capsid
  - nucleocapsid
  - protein spike
- Recall** What determines the family in which a virus is categorized?
- Recall** How does a virus infect a cell?
- Recall** What is the difference between the lytic pathway and the lysogenic pathway?
- Reflect and Apply** Is there a correlation between the speed of a viral infection and its potential mortality rate? Explain.
- Reflect and Apply** If you were going to design a drug to fight a virus, what would be likely targets for the drug design?
- Reflect and Apply** Some viruses can undergo lysis or lysogeny even in the same host. What might be a reason for this? Under what conditions might the virus favor the one strategy over the other?
- Reflect and Apply** What might be the characteristics of cells of a human who is immune to HIV infection?
- Biochemical Connections** Why can the flu be considered a more deadly disease than AIDS?
- Biochemical Connections** What are the two major proteins that we study on a flu virus?
- Biochemical Connections** Is it a correct statement that H1N1 is less dangerous than H5N1?

- Biochemical Connections** How can flu viruses change? What worries the CDC and WHO with respect to flu viruses?

### 14.2 Retroviruses

- Recall** What is unique about the life cycle of a retrovirus?
- Recall** What enzyme is responsible for the production of viral DNA from a retrovirus?
- Recall** What are three reasons that retroviruses are studied so much these days?
- Recall** What is meant by gene therapy?
- Recall** What are the two types of gene therapy?
- Recall** What types of viruses are used for gene therapy, and how are they manipulated to make them useful?
- Recall** What are the potential hazards of gene therapy?
- Reflect and Apply** What are the considerations for choice of a vector in gene therapy?
- Reflect and Apply** Both ADA-SCID and type I diabetes are diseases based on lack of a particular protein. Why has the pioneering work on gene therapy focused on SCID instead of on diabetes?

### 14.3 The Immune System

- Recall** What health conditions are linked to malfunctioning immune systems?
- Recall** What is innate immunity? What is acquired immunity?
- Recall** What are the components of innate immunity?
- Recall** What are the components of acquired immunity?
- Recall** What is the purpose of a major histocompatibility complex?

28. **Recall** What is clonal selection?
  29. **Reflect and Apply** Describe the relationship between the innate immunity system and the acquired-immunity system.
  30. **Reflect and Apply** One of the first human proteins cloned was interferon. Why would it be important to be able to produce interferon in a lab?
  31. **Reflect and Apply** Describe how the cells of the acquired-immunity system develop so that they do not recognize self-antigens but do recognize foreign antigens.
  32. **Recall** What part of the immune system has been linked to progression of cancer?
  33. **Biochemical Connections** What type of viral RNAs from herpes viruses confound the immune system?
  34. **Biochemical Connections** Explain the mode of action of the herpes viral RNAs and how they confound the immune system.
  35. **Biochemical Connections** What is special about the recently made antibodies to the hemagglutinin portion of the flu virus?
  36. **Biochemical Connections** How did the study on mice using antibodies against the flu virus demonstrate its efficiency?
- 14.4 Cancer**
37. **Recall** What characteristics are shown by cancer cells?
  38. **Recall** What is a tumor suppressor? What is an oncogene?
  39. **Recall** Why are the proteins called p53 and Ras studied so much these days?
  40. **Recall** How are viruses related to cancer?
  41. **Reflect and Apply** What is virotherapy?
  42. **Reflect and Apply** Why is it inaccurate to say, “Smoking causes cancer”?
  43. **Reflect and Apply** Describe the difference between a tumor suppressor and an oncogene with respect to the actual causes of cancer.
  44. **Reflect and Apply** Describe the relationships between Ras, Jun, and Fos.
  45. **Reflect and Apply** Describe the nature of p53 reactivation as a cancer-fighting strategy.
  46. **Reflect and Apply** What is the difference between Prima-1 and nutlins in the way they would fight cancer?
  47. **Reflect and Apply** Describe different techniques that might restore p53 to a cell lacking it.
  48. **Biochemical Connections** Describe the positive and negative effects of the innate immune system on cancer cells.
  49. **Biochemical Connections** Explain why some researchers believe science should focus its attention on the inflammation associated with cancer progression instead of looking for a cure.
  50. **Biochemical Connections** What were the results of sequencing the human melanoma and lung cancer genomes?
  51. **Biochemical Connections** What causes the majority of the mutations found in human melanomas and lung cancer?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# The Importance of Energy Changes and Electron Transfer in Metabolism

# 15

The potential energy of the water at the top of a waterfall is transformed into kinetic energy in spectacular fashion.

## 15.1 Standard States for Free-Energy Changes

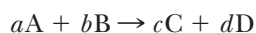
The word *bioenergetics* comes up many times in the study of biochemistry. In Chapter 1 we saw how the lowering of energy, which really means dispersal on the molecular level, is spontaneous in the thermodynamic sense. We have looked at some examples of energy changes, but we have not gone into detail about them. An example is the transport of substances into and out of cells across the cell membrane, such as the glucose transport system and the sodium-potassium pump described in Section 8.6. In this chapter, we will examine how energetic considerations apply to metabolism. We will compare so many different processes that it will be useful to have a benchmark against which to make those comparisons. Most of Chapter 20 will be devoted to the details of the transport of protons (hydrogen ion,  $H^+$ ) across membranes and its role in energy production in aerobic cells.

### What are standard states?

We can define *standard conditions* for any process and then use those standard conditions as the basis for comparing reactions. The choice of standard conditions is arbitrary. For a process under standard conditions, all substances involved in the reaction are in their **standard states**, in which case they are also said to be at *unit activity*. For pure solids and pure liquids, the standard state is the pure substance itself. For gases, the standard state is usually taken as a pressure of 1.00 atmosphere of that gas. For solutes, the standard state is usually taken as 1.00 molar concentration. Strictly speaking, these definitions for gases and for solutes are approximations, but they are valid for all but the most exacting work.

### What do standard states have to do with free-energy changes?

For any general reaction



we can write an equation that relates the free-energy change ( $\Delta G$ ) for the reaction under *any* conditions to the free-energy change under *standard* conditions ( $\Delta G^\circ$ ); the superscript  $^\circ$  refers to standard conditions. This equation is

$$\Delta G = \Delta G^\circ + RT \ln \frac{[C]^c [D]^d}{[A]^a [B]^b}$$

In this equation, the square brackets indicate molar concentrations,  $R$  is the gas constant ( $8.31 \text{ J mol}^{-1} \text{ K}^{-1}$ ), and  $T$  is the absolute temperature. The notation  $\ln$  refers to natural logarithms (to the base  $e$ ) rather than logarithms to the base 10, for which the notation is  $\log$ . This equation holds under all circumstances; the reaction does not have to be at equilibrium. The value of  $\Delta G$  under a given set of conditions depends on the value of  $\Delta G^\circ$  and on the concentration of reactants and products (given by the second term in the equation). Most biochemical reactions are described in terms of  $\Delta G^\circ$ , which is the  $\Delta G$  under standard conditions (1.00  $M$  concentration for solutes). There is only one  $\Delta G^\circ$  for a reaction at a given temperature.

## Chapter Outline

### 15.1 Standard States for Free-Energy Changes

- What are standard states?
- What do standard states have to do with free-energy changes?

### 15.2 A Modified Standard State for Biochemical Applications

- Why do we need a modified standard state for biochemical applications?

### 15.3 The Nature of Metabolism

- What is metabolism?

### 15.4 The Role of Oxidation and Reduction in Metabolism

- How are oxidation and reduction involved in metabolism?

### 15.5 Coenzymes in Biologically Important Oxidation–Reduction Reactions

- What are the reactions of key oxidation–reduction coenzymes?

### 15.6 Coupling of Production and Use of Energy

- How do energy-producing reactions allow energy-requiring reactions to take place?

### 15.7 Coenzyme A in Activation of Metabolic Pathways

- Why is coenzyme A such a good example of activation?

**OWL** Online homework for this chapter may be assigned in OWL.

When the reaction is at equilibrium,  $\Delta G = 0$ , and thus

$$0 = \Delta G^\circ + RT \ln \frac{[C]^c [D]^d}{[A]^a [B]^b}$$

$$\Delta G^\circ = -RT \ln \frac{[C]^c [D]^d}{[A]^a [B]^b}$$

The concentrations are now equilibrium concentrations, and this equation can be rewritten

$$\Delta G^\circ = -RT \ln K_{\text{eq}}$$

where  $K_{\text{eq}}$  is the equilibrium constant for the reaction. We now have a relationship between the equilibrium concentrations of reactants and products and the standard free-energy change. After we have determined the equilibrium concentrations of reactants by any convenient method, we can calculate the equilibrium constant,  $K_{\text{eq}}$ . We can then calculate the standard free-energy change,  $\Delta G^\circ$ , from the equilibrium constant.

## 15.2 A Modified Standard State for Biochemical Applications

### *Why do we need a modified standard state for biochemical applications?*

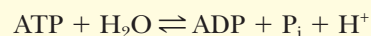
We have just seen that the calculation of standard free-energy changes includes the stipulation that all substances be in standard states, which for solutes can be approximated as a concentration of 1 *M*. If the hydrogen-ion concentration of a solution is 1 *M*, the pH is zero. (Recall that the logarithm of 1 to any base is zero.) The interior of a living cell is, in many respects, an aqueous solution of the cellular components, and the pH of such a system is normally in the neutral range. Biochemical reactions in the laboratory are usually carried out in buffers that are also at or near neutral pH. For this reason, it is convenient to define, for biochemical practice, a modified standard state, one that differs from the original standard state only by the change in hydrogen-ion concentration from 1 *M* to  $1 \times 10^{-7}$  *M*, implying a pH of 7.

When free-energy changes are calculated on the basis of this modified standard state, they are designated by the symbol  $\Delta G'^\circ$  (spoken “delta G zero prime”). The Biochemical Connections box describes other specific applications of thermodynamics to living organisms.

### Apply Your Knowledge

#### Use of Equilibrium Constants to Determine $\Delta G'^\circ$

Let us assume that the relative concentrations of reactants have been determined for a reaction carried out at pH 7 and 25°C (298 K). Such concentrations can be used to calculate an equilibrium constant,  $K_{\text{eq}}$ , which, in turn, can be used to determine the standard free-energy change,  $\Delta G'^\circ$ , for the reaction. A typical reaction to which this kind of calculation can be applied is the hydrolysis of ATP at pH 7, yielding ADP, monohydrogen phosphate ion (written as  $P_i$ ), and  $H^+$  (the reverse of a reaction we have already seen):



$$K'_{\text{eq}} = \frac{[\text{ADP}][P_i][H^+]}{[\text{ATP}]} \text{ pH 7, 25}^\circ\text{C}$$

The concentrations of the solutes are used to approximate their activities, and the activity of the water is one. The value for  $K_{\text{eq}}$  for this reaction is



determined in the laboratory; it is  $2.23 \times 10^5$ . Once we have this information, we can determine the standard free-energy change by substituting in the equation  $\Delta G^\circ = -RT \ln K_{\text{eq}}$ . The key point is to choose the correct quantities to substitute and to keep track of units. Substituting  $R = 8.31 \text{ J mol}^{-1} \text{ K}^{-1}$ ,  $T = 298 \text{ K}$ , and  $\ln K_{\text{eq}} = 12.32$ ,

$$\Delta G^\circ = -RT \ln K'_{\text{eq}}$$

$$\Delta G^\circ = (8.31 \text{ J mol}^{-1} \text{ K}^{-1})(298 \text{ K})(12.32)$$

$$\Delta G^\circ = -3.0500 \times 10^4 \text{ J mol}^{-1} = -30.5 \text{ kJ mol}^{-1} = -7.29 \text{ kcal mol}^{-1}$$

$$1 \text{ kJ} = 0.239 \text{ kcal}$$

In addition to illustrating the usefulness of a modified standard state for biochemical work, the negative value of  $\Delta G^\circ$  indicates that the reaction of hydrolysis of ATP to ADP is a spontaneous process in which energy is released.

## Biochemical Connections THERMODYNAMICS

### Living Things Need Energy—How Do They Use It?

Gibbs free energy,  $\Delta G$ , is perhaps the most suitable way to measure energy changes in living systems because it measures *the energy available to do work at constant temperature and pressure*, which describes the living state. Even cold-blooded organisms are at constant temperature and pressure at any given point in time; any temperature and pressure changes are slow enough not to affect measurements of  $\Delta G$ .

#### Spontaneity and Reversibility

The concept of spontaneity can be confusing, but it merely means that a reaction can occur without added energy. This is similar to water held behind a dam at the top of a hill, which has the potential energy to flow downhill, but does not do so unless someone opens the dam. Because water flows only downhill, that is the direction with a negative value of the free-energy change ( $-\Delta G$ ); pumping water uphill is nonspontaneous (requires energy) and has a positive value of the free-energy change ( $+\Delta G$ ). If the free-energy change is only  $1 \text{ kcal mol}^{-1}$  (about  $4 \text{ kJ mol}^{-1}$ ) in either direction, then the reaction is considered freely reversible. The reaction can readily go in either direction. If one adds reactants or removes products, the reaction shifts to the right; if one removes reactants or adds products, the reaction shifts to the left. This is a key aspect of a number of metabolic pathways; many reactions in the middle of the pathway are likely to be freely reversible. This means that the same enzymes can be used whether the pathway is in the process of breaking down a substance or of forming the substance.

In reversible metabolic pathways, often just the reactions at the ends are irreversible, and these reactions can be turned on or off to turn the whole pathway on or off, or even to reverse it.

#### Driving Endergonic Reactions

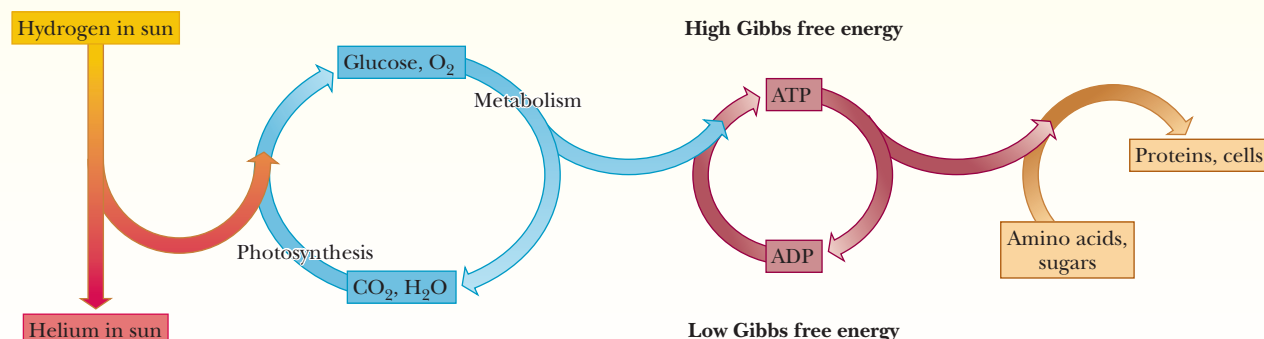
Reactions can sometimes be coupled together. This occurs when the phosphorylation of glucose is coupled to the hydrolysis of one phosphate group of ATP. Of course, there are not really two reactions going on; the enzyme merely transfers the phosphate from the ATP directly to the glucose (see Section 15.6). We can think of the phosphorylation of glucose and the hydrolysis of ATP as two parts of the same reaction. We can then add them together to determine the overall energy change and make sure that, overall, it is exergonic.

#### Energy Balance

It is important to remind ourselves of the Law of Conservation of Energy. The energy taken up by an organism is used in some way and never disappears. This is *energy balance*, which can be represented by an equation

$$\text{Energy intake} = \text{internal heat produced} + \text{work} + \text{storage}$$

Heat and work include not only the mechanical work of motion but anabolic processes and all the activities of an organism. In organisms, storage refers to chemical energy, which can be converted to other forms of energy as needed. Stored chemical energy can also be accessed as needed to drive endergonic reactions, as we have just seen. Here we will explore the implications of these statements.



- The synthesis of glucose and other sugars in plants, the production of ATP from ADP, and the elaboration of proteins and other biological molecules are all processes in which the Gibbs free energy of the system must increase. They occur only through coupling to other processes in which the Gibbs free energy decreases by an even larger amount. There is a local decrease in entropy at the expense of higher entropy of the Universe.

## 15.3 The Nature of Metabolism

### What is metabolism?

Until now, we have discussed some basic chemical principles and investigated the natures of the molecules of which living cells are composed. We have yet to discuss the bulk of chemical reactions of biomolecules themselves, which constitute **metabolism**, the biochemical basis of all life processes. The molecules of carbohydrates, fats, and proteins taken into an organism are processed in a variety of ways (Figure 15.1). The breakdown of larger molecules to smaller ones is called **catabolism**. Small molecules are used as the starting points of a variety of reactions to produce larger and more complex molecules, including proteins and nucleic acids; this process is called **anabolism**. Catabolism and anabolism are separate pathways; they are not simply the reverse of each other.

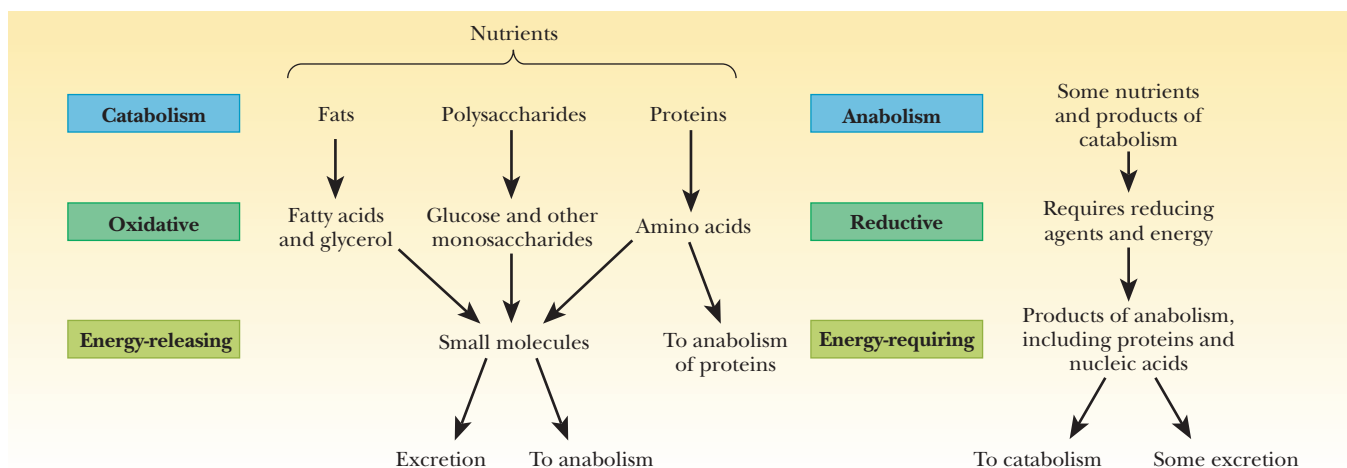
Catabolism is an oxidative process that releases energy; anabolism is a reductive process that requires energy. We shall need several chapters to explore some of the implications of this statement. In this chapter, we discuss oxidation and reduction (electron-transfer reactions) and their relation to the use of energy by living cells. The Biochemical Connections box on page 433 deals with another aspect of the unique energetics of living things.

Keep in mind that Figure 15.1 gives a very general outline of metabolism. In the specific case of photosynthetic organisms, the energy source is radiant energy from the Sun, the ultimate energy source for the Earth. Some bacteria can oxidize inorganic materials, including nitrates, sulfites, and hydrogen sulfide, producing ATP. In this text, we are going to concentrate on general cases, but it is important to bear in mind that life has many variations.

## 15.4 The Role of Oxidation and Reduction in Metabolism

### How are oxidation and reduction involved in metabolism?

**Oxidation–reduction reactions**, also referred to as *redox* reactions, are those in which electrons are transferred from a donor to an acceptor. **Oxidation** is the loss of electrons, and **reduction** is the gain of electrons. The substance that loses electrons (the electron donor)—that is, the one that is oxidized—is called the **reducing agent** or reductant. The substance that gains electrons (the electron acceptor)—the one that is reduced—is called the **oxidizing agent** or oxidant.



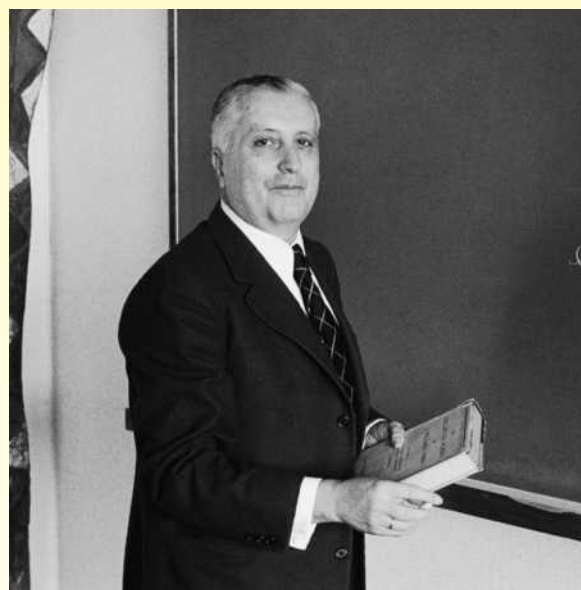
■ FIGURE 15.1 A comparison of catabolism and anabolism.

# Biochemical Connections THERMODYNAMICS

## Living Things Are Unique Thermodynamic Systems

Questions arise frequently about whether living organisms obey the laws of thermodynamics. The short answer is that they most definitely do. Most classical treatments of thermodynamics deal with closed systems at equilibrium. A closed system can exchange energy, but not matter, with its surroundings.

A living organism is obviously not a closed system, but an open system that can exchange both matter and energy with its surroundings. Because living organisms are open systems, they cannot be at equilibrium as long as they are alive, as shown in the following figure. They can, however, achieve a *steady state*, which is a stable condition. It is the state in which living things can operate at maximum thermodynamic efficiency. This point was established by Ilya Prigogine, winner of the 1977 Nobel Prize in chemistry for his work on nonequilibrium thermodynamics. He showed that for systems not at equilibrium, ordered structures can arise from disordered ones. This treatment of thermodynamics is quite advanced and highly mathematical, but the results are more directly applicable to biological systems than those of classical thermodynamics. This approach applies not only to living organisms but to the growth of cities and to predictions of auto traffic.

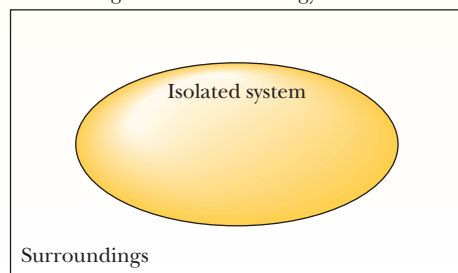


© Pierre Vauthey

■ Ilya Prigogine (1917–2003). Ilya Prigogine was born in Moscow in 1917. His family moved to Germany to escape the Russian revolution and subsequently moved to Belgium. He studied at the Université Libre in Brussels and remained there as a faculty member to conduct research on nonequilibrium thermodynamics. He was also associated with the University of Texas, which found a unique way to mark his receiving the Nobel Prize: a tower on the Texas campus is illuminated when one of the university's sports teams wins a championship. It was also illuminated at the time of the announcement of his Nobel Prize.

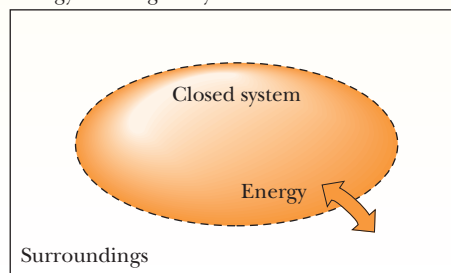
### Isolated system:

No exchange of matter or energy



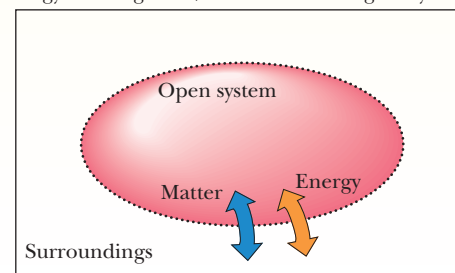
### Closed system:

Energy exchange may occur



### Open system:

Energy exchange and/or matter exchange may occur

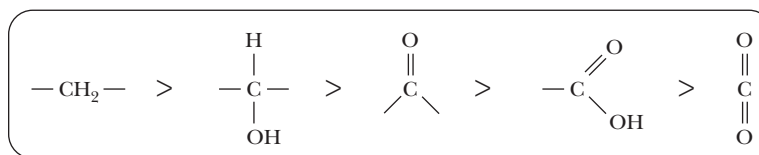


■ The characteristics of isolated, closed, and open systems. Isolated systems exchange neither matter nor energy with their surroundings. Closed systems may exchange energy, but not matter, with their surroundings. Open systems may exchange either matter or energy with the surroundings.

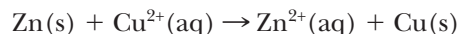
Both an oxidizing agent and a reducing agent are necessary for the transfer of electrons (an oxidation–reduction reaction) to take place.

An example of an oxidation–reduction reaction is the one that occurs when a strip of metallic zinc is placed in an aqueous solution containing copper ions. Although both zinc and copper ions play roles in life processes, this particular reaction does not occur in living organisms. However, it is a good place to start our discussion of electron transfer because, in this comparatively simple reaction, it is fairly easy to follow where the electrons are going. (It is not always

■ **FIGURE 15.2** Comparison of the state of reduction of carbon atoms in biomolecules:  $-\text{CH}_2-$  (fats)  $>$   $-\text{CHOH}-$  (carbohydrates)  $>$   $\text{C}=\text{O}$  (carbonyls)  $>$   $-\text{COOH}$  (carboxyls)  $>$   $\text{CO}_2$  (carbon dioxide, the final product of catabolism).



quite as easy to keep track of the details in biological redox reactions.) The experimental observation is that the zinc metal disappears and zinc ions go into solution, while copper ions are removed from the solution and copper metal is deposited. The equation for this reaction is



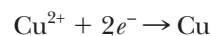
The notation (s) signifies a solid and (aq) signifies a solute in aqueous solution.

In the reaction between zinc metal and copper ion, the Zn lost two electrons to become the  $\text{Zn}^{2+}$  ion and was oxidized. A separate equation can be written for this part of the overall reaction, and it is called the **half reaction** of oxidation:



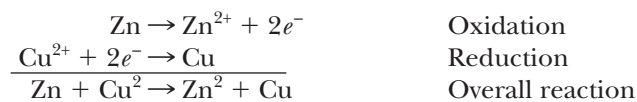
Zn is the reducing agent (it loses electrons; it is an electron donor; it is oxidized).

Likewise, the  $\text{Cu}^{2+}$  ion gained two electrons to form Cu and was reduced. An equation can also be written for this part of the overall reaction and is called the half reaction of reduction.



$\text{Cu}^{2+}$  is the oxidizing agent (it gains electrons; it is an electron acceptor; it is reduced).

If the two equations for the half reactions are combined, the result is an equation for the overall reaction:



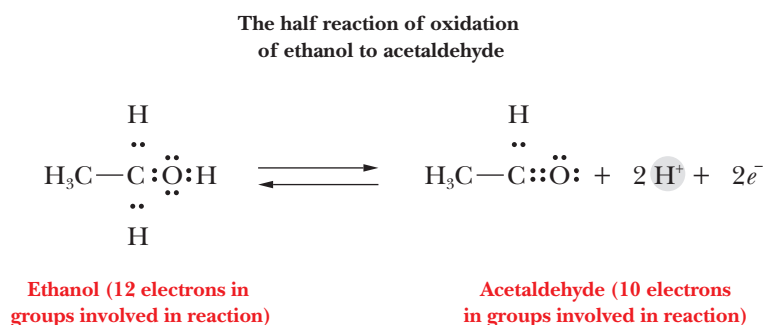
This reaction is a particularly clear example of electron transfer. It will be useful to keep these basic principles in mind when we examine the flow of electrons in the more complex redox reactions of aerobic metabolism. In many of the biological redox reactions we will encounter, the oxidation state of a carbon atom changes. Figure 15.2 shows the changes that occur as carbon in its most reduced form (an alkane) becomes oxidized to an alcohol, an aldehyde, a carboxylic acid, and ultimately carbon dioxide. Each of these oxidations requires the loss of two electrons.

## 15.5 Coenzymes in Biologically Important Oxidation–Reduction Reactions

### What are the reactions of key oxidation–reduction coenzymes?

Oxidation–reduction reactions are discussed at length in textbooks of general and inorganic chemistry, but the oxidation of nutrients by living organisms to provide energy requires its own special treatment. The description of redox reactions in terms of oxidation numbers, which is widely used with inorganic compounds, can be used to deal with the oxidation of carbon-containing

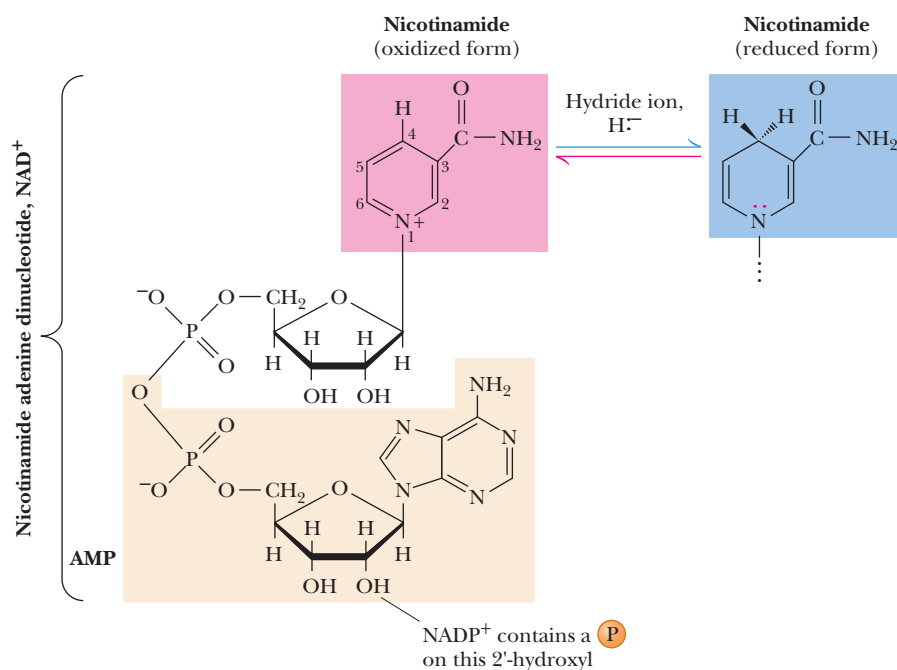
molecules. However, our discussion will be more pictorial and easier to follow if we write equations for the half reactions and then concentrate on the functional groups of the reactants and products and on the number of electrons transferred. An example is the oxidation half reaction for the conversion of ethanol to acetaldehyde.



Writing the Lewis electron-dot structures for the functional groups involved in the reaction helps us keep track of the electrons being transferred. In the oxidation of ethanol, there are 12 electrons in the part of the ethanol molecule involved in the reaction and 10 electrons in the corresponding part of the acetaldehyde molecule; two electrons are transferred to an electron acceptor (an oxidizing agent). This type of “bookkeeping” is useful for dealing with biochemical reactions. Many biological oxidation reactions, like this example, are accompanied by the transfer of a proton ( $\text{H}^+$ ). The oxidation half reaction has been written as a reversible reaction because the occurrence of oxidation or reduction depends on the other reagents present.

Another example of an oxidation half reaction is that for the conversion of NADH, the reduced form of nicotinamide adenine dinucleotide, to the oxidized form,  $\text{NAD}^+$ . This substance is an important **coenzyme** in many reactions.

Figure 15.3 shows the structure of  $\text{NAD}^+$  and NADH; the nicotinamide portion, the functional group involved in the reaction, is indicated in red and blue.

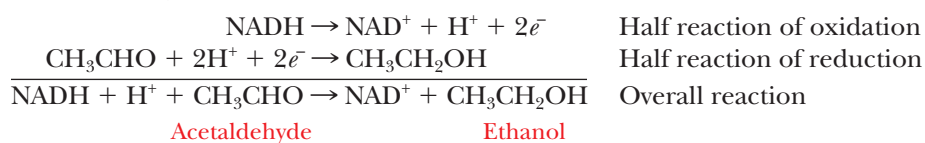


**FIGURE 15.3** The structures and redox states of the nicotinamide coenzymes. Hydride ion (a proton with two electrons) transfers to  $\text{NAD}^+$  to produce NADH.



Nicotinamide is a derivative of nicotinic acid (also called niacin), one of the B-complex vitamins (see Section 7.8). A similar compound is NADPH (for which the oxidized form is  $\text{NADP}^+$ ). It differs from NADH by having an additional phosphate group; the site of attachment of this phosphate group to ribose is also indicated in Figure 15.3. To simplify writing the equation for the oxidation of NADH, only the nicotinamide ring is shown explicitly, with the rest of the molecule designated as R. The two electrons that are lost when NADH is converted to  $\text{NAD}^+$  can be considered to come from the bond between carbon and the lost hydrogen, with the nitrogen lone-pair electrons becoming involved in a bond. Note that the loss of a hydrogen and two electrons can be considered as the loss of a hydride ion ( $\text{H}^-$ ) by NADH and is sometimes written that way.

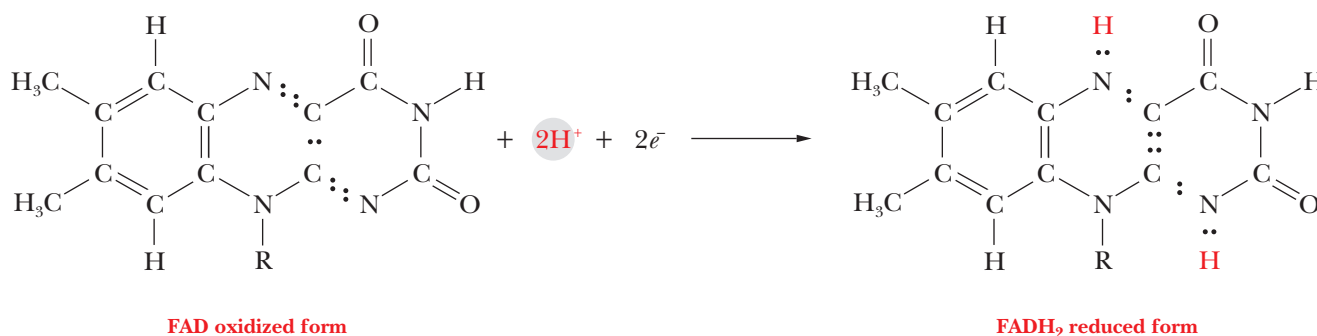
The equations for both the reaction of NADH to  $\text{NAD}^+$  and that of ethanol to acetaldehyde have been written as oxidation half reactions. If ethanol and NADH were mixed in a test tube, no reaction could take place because there would be no electron acceptor. If, however, NADH were mixed with acetaldehyde, which is an oxidized species, a transfer of electrons could take place, producing ethanol and  $\text{NAD}^+$ . (This reaction would take place very slowly in the absence of an enzyme to catalyze it. Here we have an excellent example of the difference between the thermodynamic and kinetic aspects of reactions. The reaction is spontaneous in the thermodynamic sense but very slow in the kinetic sense.)



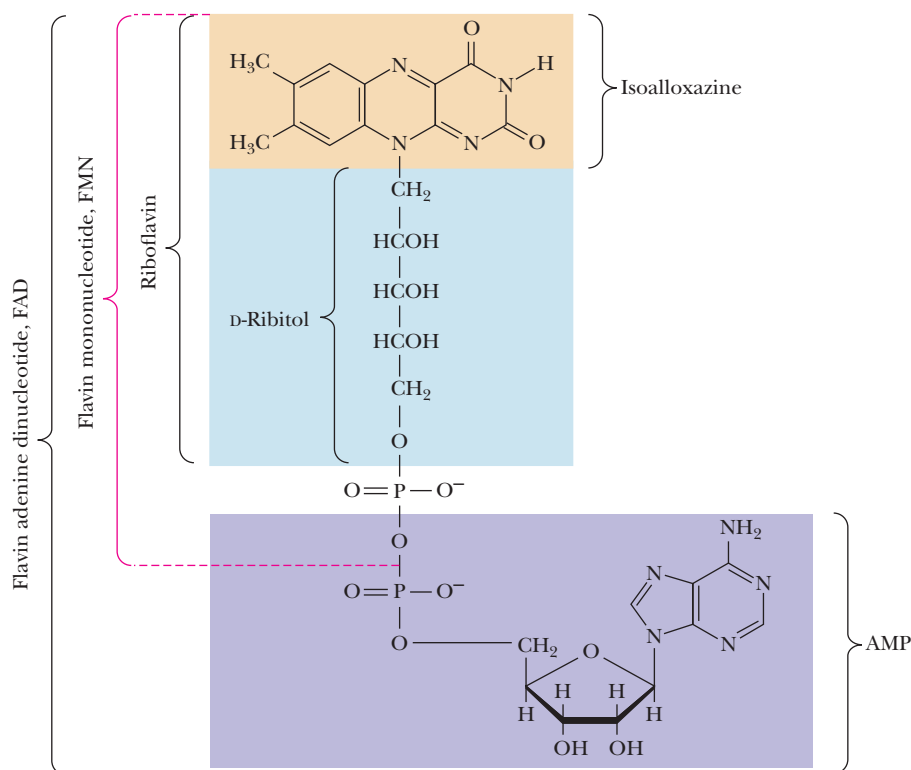
Such a reaction does take place in some organisms as the last step of alcoholic fermentation. The NADH is oxidized while the acetaldehyde is reduced.

Another important electron acceptor is FAD (flavin adenine dinucleotide) (Figure 15.4), which is the oxidized form of  $\text{FADH}_2$ . The symbol  $\text{FADH}_2$  explicitly recognizes that protons (hydrogen ions) as well as electrons are accepted by FAD. The structures shown in this equation again point out the electrons that are transferred in the reaction. Several other coenzymes contain the flavin group; they are derived from the vitamin riboflavin (vitamin  $\text{B}_2$ ).

The half reaction of reduction of FAD to  $\text{FADH}_2$



Oxidation of nutrients to provide energy for an organism cannot take place without reduction of some electron acceptor. The ultimate electron acceptor in aerobic oxidation is oxygen; we shall encounter intermediate electron acceptors as we discuss metabolic processes. Reduction of metabolites plays a significant role in living organisms in anabolic processes. Important biomolecules are synthesized in organisms by many reactions in which a metabolite is reduced while the reduced form of a coenzyme is oxidized.

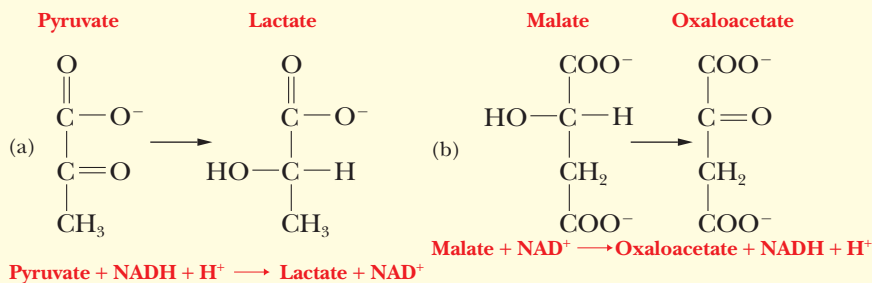


■ **FIGURE 15.4** The structures of riboflavin, flavin mononucleotide (FMN), and flavin adenine dinucleotide (FAD). Even in organisms that rely on the nicotinamide coenzymes (NADH and NADPH) for many of their oxidation–reduction cycles, the flavin coenzymes fill essential roles. Flavins are stronger oxidizing agents than  $\text{NAD}^+$  and NADP. They can be reduced by both one-electron and two-electron pathways and can be reoxidized easily by molecular oxygen. Enzymes that use flavins to carry out their reactions—flavoenzymes—are involved in many kinds of oxidation–reduction reactions.

## Apply Your Knowledge

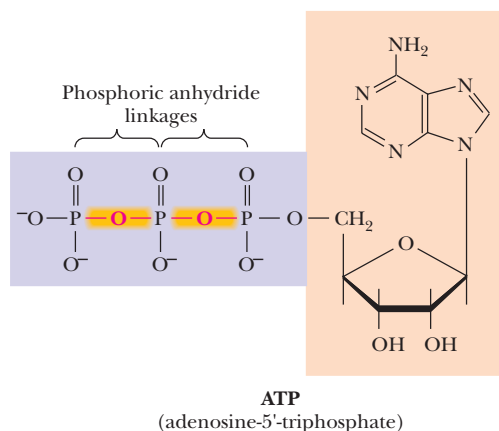
### Oxidation and Reduction

In the following reactions, identify the substance oxidized, the substance reduced, the oxidizing agent, and the reducing agent.



### Solution

The way to approach this question is to recall that NADH is a reduced form of the coenzyme. It will be oxidized and will serve as a reducing agent.  $\text{NAD}^+$  is the oxidized form. It will be reduced and thus will serve as the oxidizing agent. In the first reaction, pyruvate is reduced and NADH is oxidized; pyruvate is the oxidizing agent, and NADH is the reducing agent. In the second reaction, malate is oxidized, and  $\text{NAD}^+$  is reduced;  $\text{NAD}^+$  is the oxidizing agent, and malate is the reducing agent.



■ **FIGURE 15.5** The phosphoric anhydride bonds in ATP are “high-energy” bonds, referring to the fact that they require or release convenient amounts of energy, depending on the direction of the reaction.

## 15.6 Coupling of Production and Use of Energy

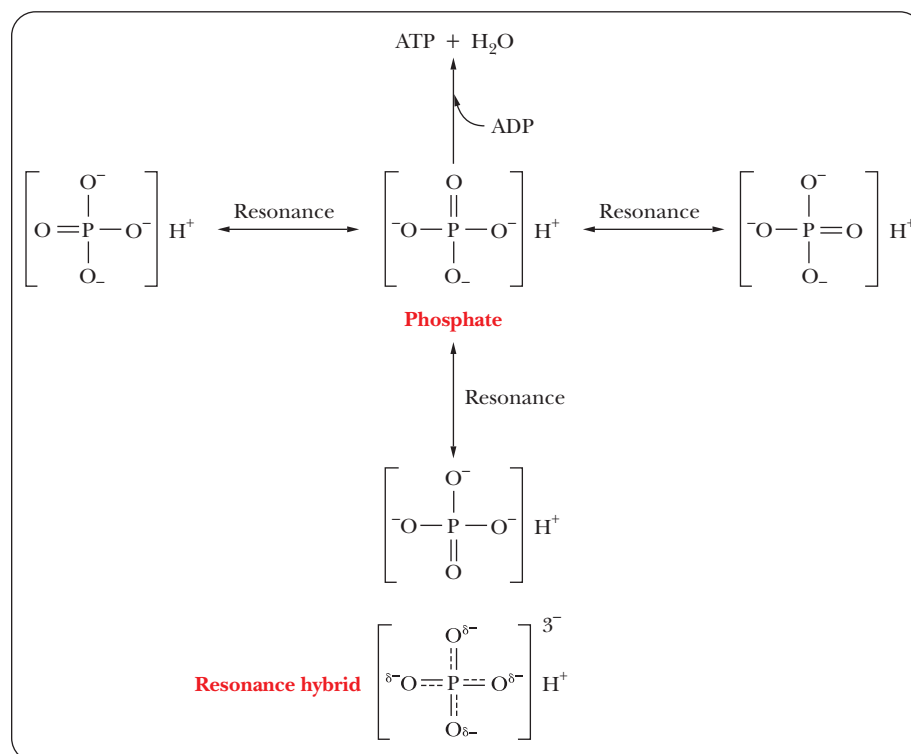
Another important question about metabolism is: “How is the energy released by the oxidation of nutrients trapped and used?” This energy cannot be used directly; it must be shunted into an easily accessible form of chemical energy.

In Section 1.11, we saw that several phosphorus-containing compounds, such as ATP, can be hydrolyzed easily, and that the reaction releases energy. Formation of ATP is intimately linked with the release of energy from oxidation of nutrients. The coupling of energy-producing reactions and energy-requiring reactions is a central feature in the metabolism of all organisms.

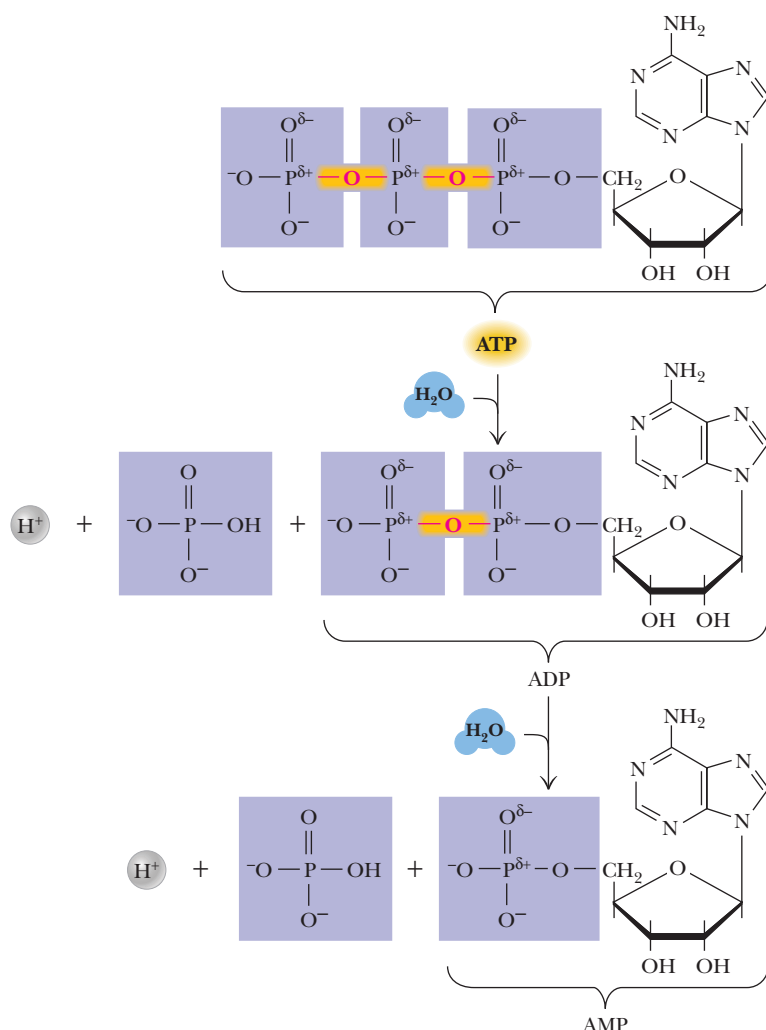
### How do energy-producing reactions allow energy-requiring reactions to take place?

The phosphorylation of ADP (adenosine diphosphate) to produce ATP (adenosine triphosphate) requires energy, which can be supplied by the oxidation of nutrients. Conversely, the hydrolysis of ATP to ADP releases energy (Figure 15.5).

The forms of ADP and ATP shown in this section are in their ionization states for pH 7. The symbol  $P_i$  for phosphate ion comes from its name in biochemical jargon, “inorganic phosphate.” Note that there are four negative charges on ATP and three on ADP; electrostatic repulsion makes ATP less stable than ADP. Energy must be expended to put an additional negatively charged phosphate group on ADP by forming a covalent bond to the phosphate group being added. In addition, there is an entropy loss when ADP is phosphorylated to ATP. Inorganic phosphate can adopt multiple resonance structures, and the loss of these potential structures results in a decrease in entropy when the phosphate is attached to ADP (Figure 15.6). The  $\Delta G^\circ$  for the reaction refers to the usual biochemical convention of pH 7 as the standard state for hydrogen ion (Section 15.2). Note, however, that there is a marked decrease in electrostatic repulsion on hydrolysis of ATP to ADP (Figure 15.7).

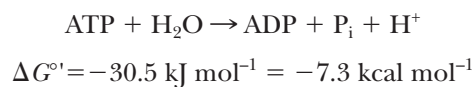


■ **FIGURE 15.6** Loss of a resonance-stabilized phosphate ion in production of ATP. When ADP is phosphorylated to ATP, there is a loss of the resonance-stabilized phosphate ion, resulting in a decrease in entropy. ( $\delta^-$  denotes a partial negative charge.)



■ **FIGURE 15.7** Decrease in electrostatic repulsion on hydrolysis of ATP. Hydrolysis of ATP to ADP (and/or hydrolysis of ADP to AMP) leads to relief of electrostatic repulsion.

The reverse reaction, the hydrolysis of ATP to ADP and phosphate ion, releases  $30.5 \text{ kJ mol}^{-1}$  ( $7.3 \text{ kcal mol}^{-1}$ ) when energy is needed:



The bond that is hydrolyzed when this reaction takes place is sometimes called a “high-energy bond,” which is shorthand terminology for a reaction in which hydrolysis of a specific bond releases a useful amount of energy. Another way of indicating such a bond is ~P. Numerous organophosphate compounds with high-energy bonds play roles in metabolism, but ATP is by far the most important (Table 15.1). In some cases, the free energy of hydrolysis of organophosphates is higher than that of ATP and is thus able to drive the phosphorylation of ADP to ATP.

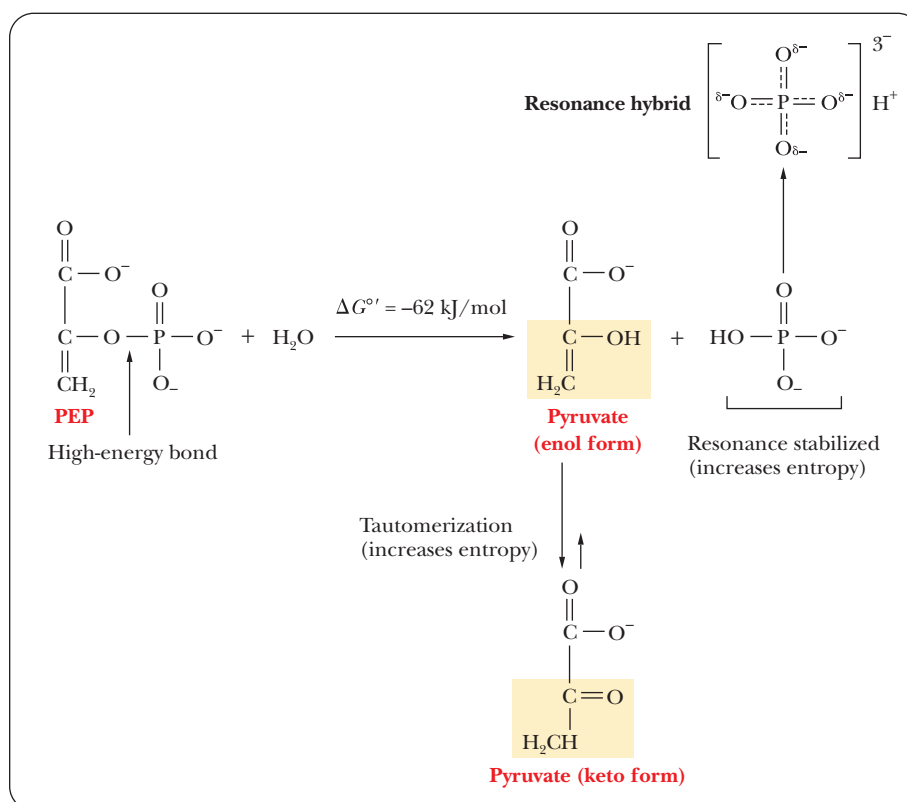
Phosphoenolpyruvate (PEP), a molecule we shall encounter when we look at glycolysis, tops the list. It is a very high-energy compound because of the resonance stabilization of the liberated phosphate when it is hydrolyzed (the same effect as that seen with ATP) and because keto–enol tautomerization of pyruvate is a possibility. Both effects increase the entropy upon hydrolysis (Figure 15.8).

The energy of hydrolysis of ATP is not stored energy, just as an electric current does not represent stored energy. Both ATP and electric current must be

TABLE 15.1

Compound	$\Delta G^{\circ'}$	
	$\text{kJ mol}^{-1}$	$\text{kcal mol}^{-1}$
Phosphoenolpyruvate	-61.9	-14.8
Carbamoyl phosphate	-51.4	-12.3
Creatine phosphate	-43.1	-10.3
Acetyl phosphate	-42.2	-10.1
ATP (to ADP)	-30.5	-7.3
Glucose-1-phosphate	-20.9	-5.0
Glucose-6-phosphate	-12.5	-3.0
Glycerol-3-phosphate	-9.7	-2.3

produced when they are needed—by organisms or by a power plant, as the case may be. The cycling of ATP and ADP in metabolic processes is a way of shunting energy from its production (by oxidation of nutrients) to its uses (in processes such as biosynthesis of essential compounds or muscle contraction) when it is needed. The oxidation processes take place when the organism needs the energy that can be generated by the hydrolysis of ATP. When chemical energy is stored, it is usually in the form of fats and carbohydrates, which are metabolized as needed. Certain small biomolecules, such as creatine phosphate, can also serve as vehicles for storing chemical energy. The energy that must be supplied for the many endergonic reactions in life processes comes directly from the hydrolysis of ATP and indirectly from the oxidation of nutrients. The latter produces the energy needed to phosphorylate ADP to ATP



■ **FIGURE 15.8** Increase in entropy on hydrolysis of phosphoenolpyruvate. When phosphoenolpyruvate is hydrolyzed to pyruvate and phosphate, it results in an increase in entropy. Both the formation of the keto form of pyruvate and the resonance structures of phosphate lead to the increase in entropy.



(Figure 15.9). In addition to its key role in bioenergetics, ATP can do many other things, especially in the realm of intercellular communication. The following Biochemical Connections box gives some details.

Let us examine some biological reactions that release energy and see how some of that energy is used to phosphorylate ADP to ATP. The multistep

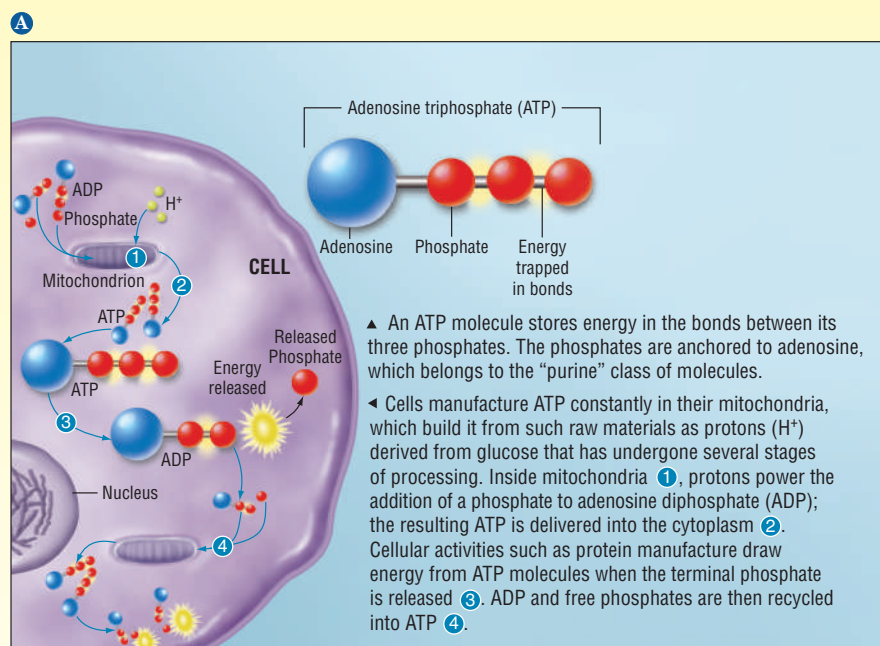
## Biochemical Connections PHYSIOLOGY

### ATP in Cell Signaling

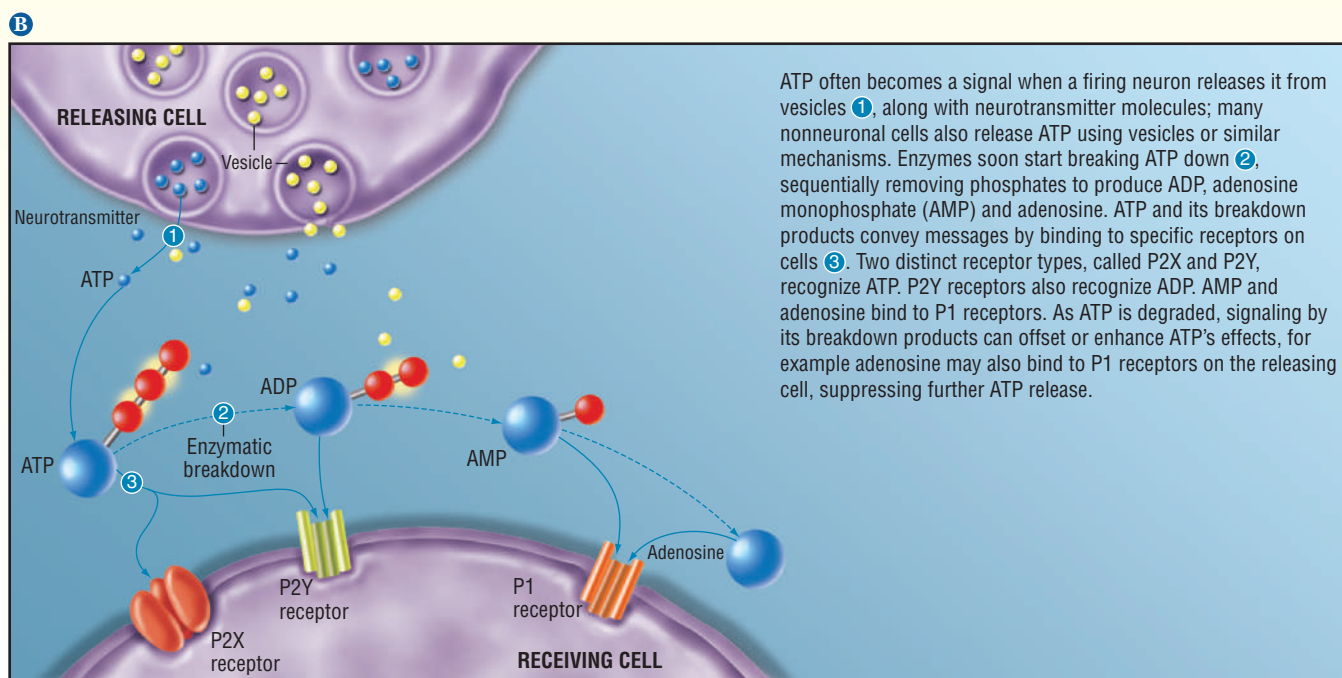
The role of ATP in generating and using energy within cells is a fundamental and well known fact. Cells produce ATP in their mitochondria and use it to provide energy for their activities, such as protein synthesis. It is less well known that ATP is widely involved in chemical signaling between cells. For example, during the transmission of nerve signals, ATP is released along with neurotransmitters.

It is hydrolyzed by enzymes to ADP, AMP, and adenosine. Specific receptors exist on the surface of receiving cells: P2X receptors for ATP itself, P2Y receptors for ATP and ADP, and P1 receptors for AMP and adenosine.

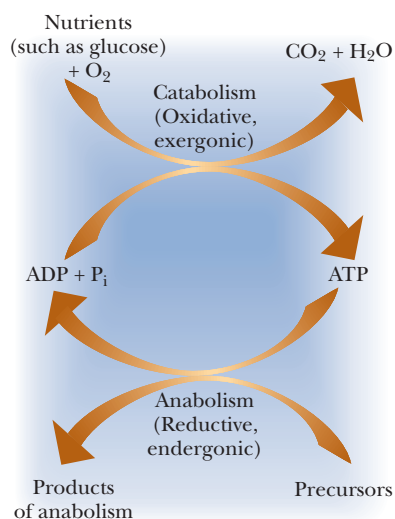
The binding of ATP and its degradation products to their individual receptors sets off a response within the cell. The effects can vary widely both in nature and duration. In many cases, a short-term effect, such as hydrolysis of a chemical bond, triggers a long-term effect. Release of calcium ions from an intracellular reservoir, for example, frequently leads to activation of key enzymes. In Chapter 24, we will see many examples of the responses of cells to extracellular triggers as a result of receptor binding.



■ (A) ATP in generation and use of energy.

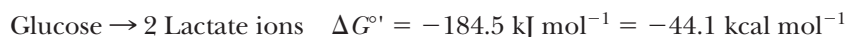


■ (B) ATP in cell signaling. (From Khakh, B., and G. Burnstock, *The Double Life of ATP*, Sci. Amer. 301 (6) 84–92 (2009).)



■ **FIGURE 15.9** The role of ATP as energy currency in processes that release energy and processes that use energy.

conversion of glucose to lactate ions is an exergonic and anaerobic process. Two molecules of ADP are phosphorylated to ATP for each molecule of glucose metabolized. The basic reactions are the production of lactate, which is exergonic,

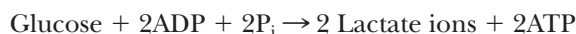


and the phosphorylation of two moles of ADP for each mole of glucose, which is endergonic.



$$\Delta G^{\circ} = 61.0 \text{ kJ mol}^{-1} = 14.6 \text{ kcal mol}^{-1}$$

(In the interest of simplicity, we shall write the equation for phosphorylation of ADP in terms of ADP, P<sub>i</sub>, and ATP only.) The overall reaction is

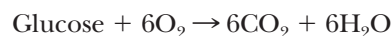


$$\Delta G^{\circ} \text{ overall} = -184.5 + 61.0 = -123.5 \text{ kJ mol}^{-1} = -29.5 \text{ kcal mol}^{-1}$$

Not only can we add the two chemical reactions to obtain an equation for the overall reaction, we can also add the free-energy changes for the two reactions to find the overall free-energy change. We can do this because the free-energy change is a **state function**; it depends only on the initial and final states of the system under consideration, not on the path between those states. The exergonic reaction provides energy, which drives the endergonic reaction. This phenomenon is called **coupling**. The percentage of the released energy that is used to phosphorylate ADP is the efficiency of energy use in anaerobic metabolism; it is  $(61.0/184.5) \times 100$ , or about 33%. The number 61.0 comes from the number of kilojoules required to phosphorylate 2 moles of ADP to ATP, and the number 184.5 is the number of kilojoules released when 1 mole of glucose is converted to 2 moles of lactate.

The breakdown of glucose goes further under aerobic conditions than under anaerobic conditions. The end products of aerobic oxidation are 6 molecules of carbon dioxide and 6 molecules of water for each molecule of glucose. Up to 32 molecules of ADP can be phosphorylated to ATP when 1 molecule of glucose is broken down completely to carbon dioxide and water.

The exergonic reaction for the complete oxidation of glucose is



$$\Delta G^{\circ} = -2867 \text{ kJ mol}^{-1} = -685.9 \text{ kcal mol}^{-1}$$

The endergonic reaction for phosphorylation is



$$\Delta G^{\circ} = 976 \text{ kJ} = 233.5 \text{ kcal}$$

The net reaction is



$$\Delta G^{\circ} = -2867 + 976 = -1891 \text{ kJ mol}^{-1} = -452.4 \text{ kcal mol}^{-1}$$

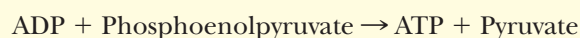
Note that, once again, we add the two reactions and their respective free-energy changes to obtain the overall reaction and its free-energy change. The efficiency of aerobic oxidation of glucose is  $(976/2867) \times 100$ , about 34%. (We performed this calculation in the same way that we did with the example of anaerobic oxidation of glucose.) More ATP is produced by the coupling process in aerobic oxidation of glucose than by the coupling process in anaerobic

oxidation. The hydrolysis of ATP produced by breakdown (aerobic or anaerobic) of glucose can be coupled to endergonic processes, such as muscle contraction in exercise. As any jogger or long-distance swimmer knows, aerobic metabolism involves large quantities of energy, processed in a highly efficient fashion. We have now seen two examples of coupling of exergonic and endergonic processes—aerobic oxidation of glucose and anaerobic fermentation of glucose—involving different amounts of energy.

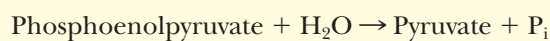
### Apply Your Knowledge

#### Predicting Reactions: Calculations of Free Energies

We shall use values from Table 15.1 to calculate  $\Delta G^\circ$  for the following reaction and decide whether it is spontaneous. The most important point here is that we add algebraically. In particular, we have to remember to change the sign of the  $\Delta G^\circ$  when we reverse the direction of the reaction.

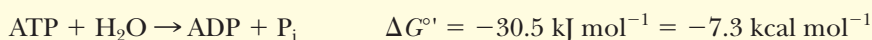


From Table 15.1,

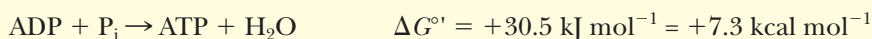


$$\Delta G^\circ = -61.9 \text{ kJ mol}^{-1} = -14.8 \text{ kcal mol}^{-1}$$

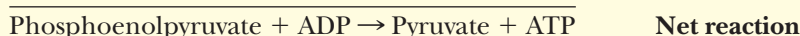
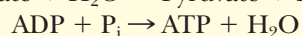
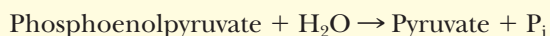
Also,



We want the reverse of the second reaction:



We now add the two reactions and their free-energy changes:



$$\Delta G^\circ = -61.9 \text{ kJ mol}^{-1} + 30.5 \text{ kJ mol}^{-1} = -31.4 \text{ kJ mol}^{-1}$$

$$\Delta G^\circ = -14.8 \text{ kcal mol}^{-1} + 7.3 \text{ kcal mol}^{-1} = -7.5 \text{ kcal mol}^{-1}$$

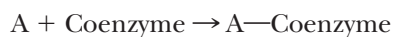
The reaction is spontaneous, as indicated by the  $\Delta G^\circ$ . However, remember that even though this reaction is spontaneous, nothing would happen if you simply put these chemicals in a test tube. Biochemical reactions require enzymes to catalyze them.

## 15.7 Coenzyme A in Activation of Metabolic Pathways

The metabolic oxidation of glucose that we saw in Section 15.6 does not take place in one step. The anaerobic breakdown of glucose requires many steps, and the complete aerobic oxidation of glucose to carbon dioxide and water has still more steps. One of the most important points about the multistep nature of all metabolic processes, including the oxidation of glucose, is that the many stages allow for efficient production and use of energy. The electrons produced by the oxidation of glucose are passed along to oxygen, the ultimate electron acceptor, by intermediate electron acceptors. Many of the intermediate stages of the oxidation of glucose are coupled to ATP production by phosphorylation of ADP.

### Why is coenzyme A such a good example of activation?

A step frequently encountered in metabolism is the process of **activation**. In a reaction of this sort, a metabolite (a component of a metabolic pathway) is bonded to some other molecule, such as a coenzyme, and the free-energy change for breaking this new bond is negative. In other words, the next reaction in the metabolic pathway is exergonic. For example, if substance A is the metabolite and reacts with substance B to give AB, the following series of reactions might take place:



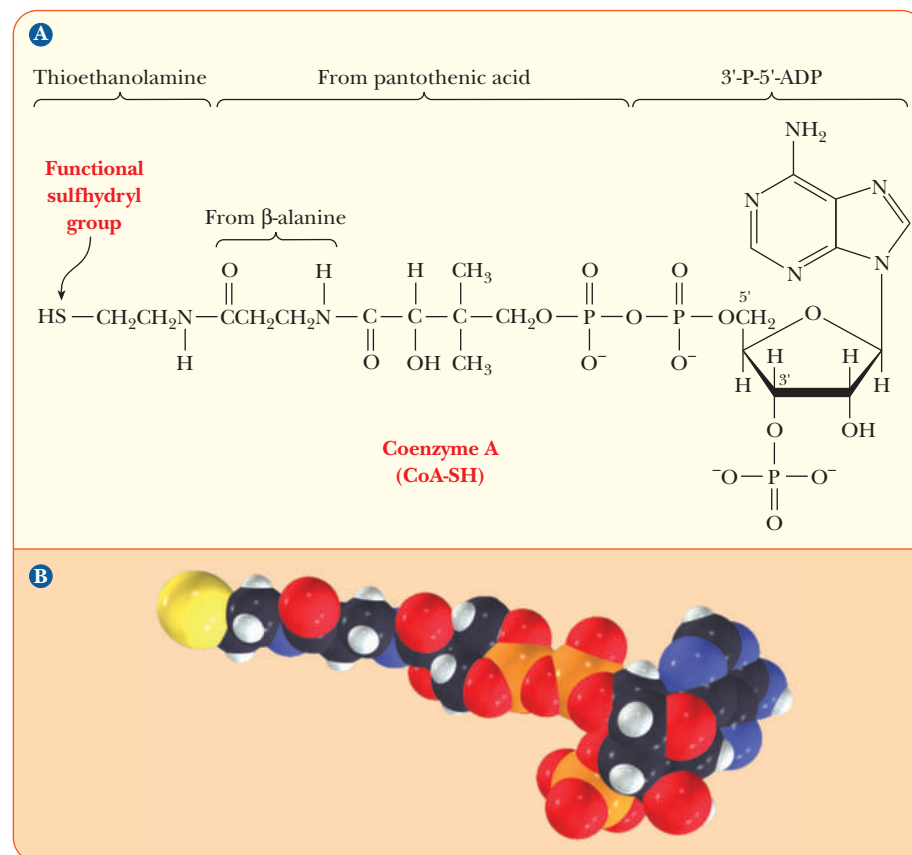
**Activation step**



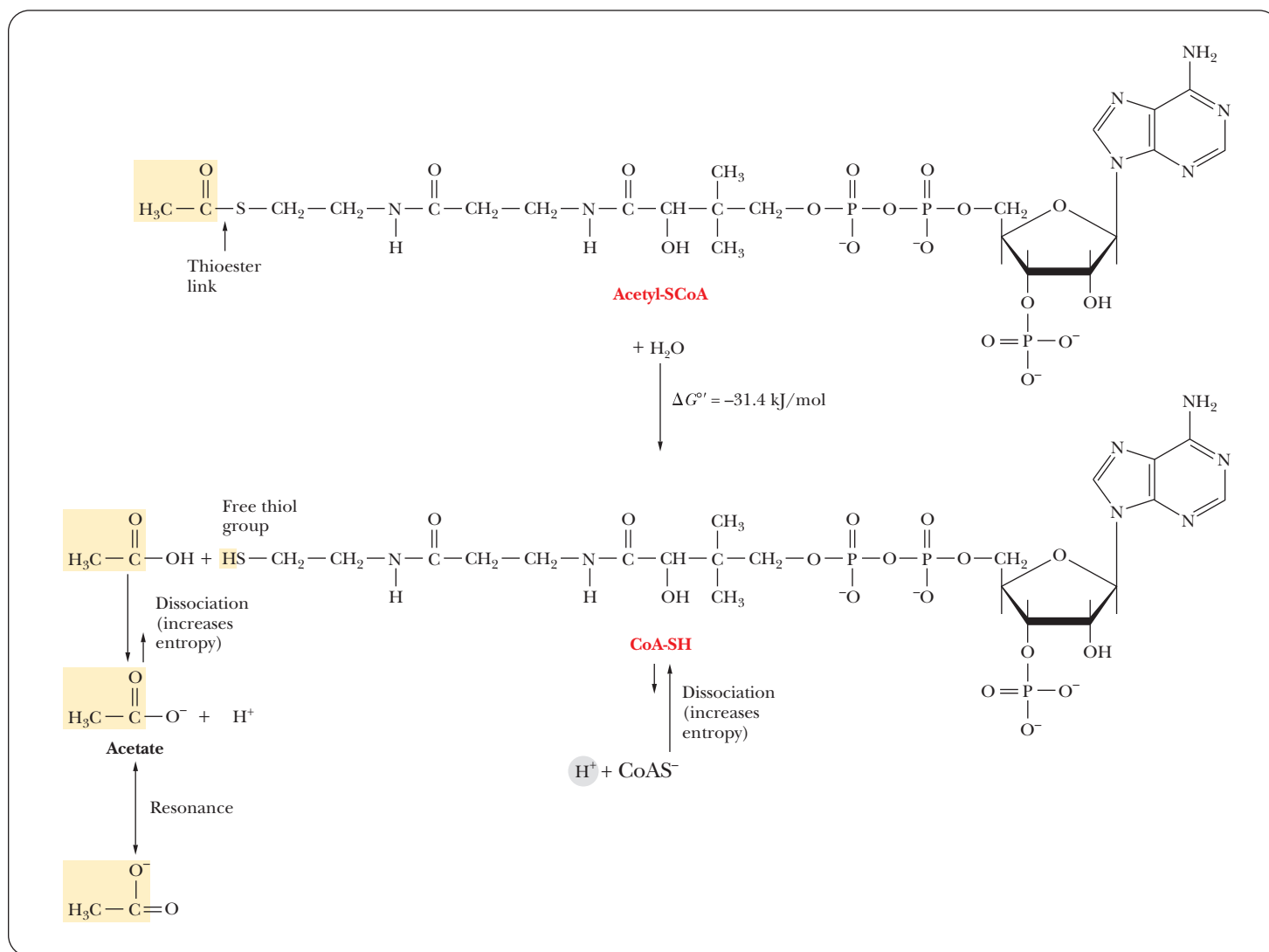
The formation of a more reactive substance in this fashion is called activation.

There are many examples of activation in metabolic processes. We can discuss one of the most useful examples now. It involves forming a covalent bond to a compound known as coenzyme A (CoA).

The structure of CoA is complex. It consists of several smaller components linked together covalently (Figure 15.10). One part is 3'-P-5'-ADP, a derivative of adenosine with phosphate groups esterified to the sugar, as shown in the structure. Another part is derived from the vitamin pantothenic acid, and the part of the molecule involved in activation reactions contains a thiol group. In fact, coenzyme A is frequently written as CoA-SH to emphasize that the thiol group is the reactive portion of the molecule. For example, carboxylic acids form thioester linkages with CoA-SH. The metabolically active form of a carboxylic acid is the corresponding acyl-CoA thioester, in which the thioester linkage is a high-energy bond (Figure 15.11). Thioesters are high-energy compounds



■ **FIGURE 15.10** Two ways of looking at coenzyme A. (a) The structure of coenzyme A. (b) Space-filling model of coenzyme A.



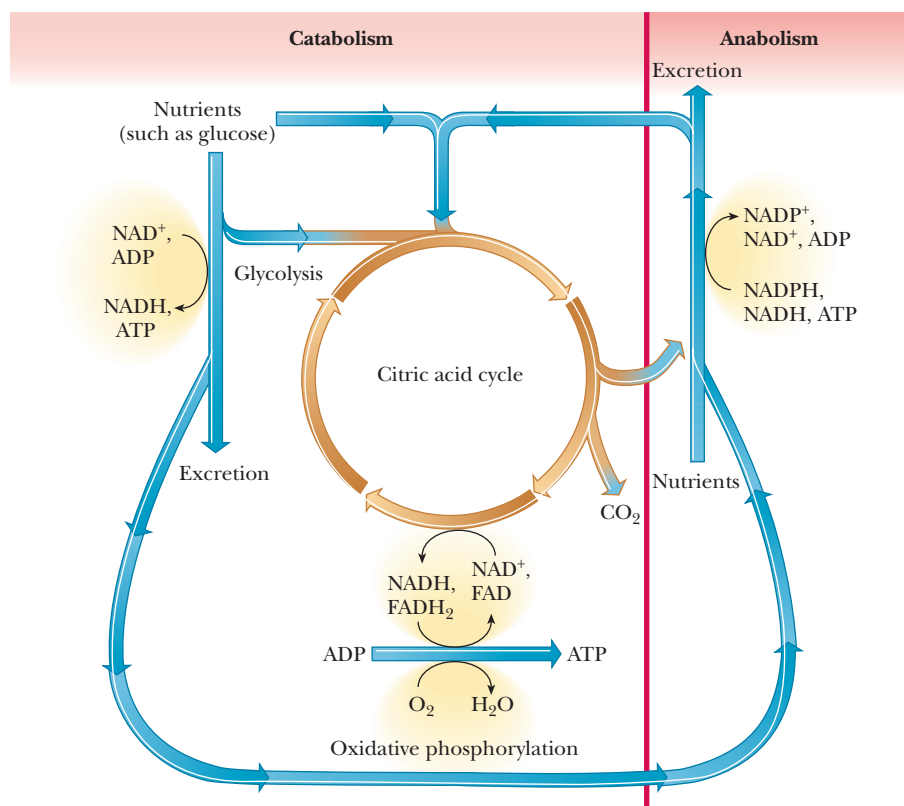
■ **FIGURE 15.11** The hydrolysis of acetyl-CoA. The products are stabilized by resonance and by dissociation.

because of the possible dissociation of the products after hydrolysis and resonance structures of the products. For example, when acetyl-CoA is hydrolyzed, the  $-\text{SH}$  at the end of the molecule can dissociate slightly to form  $\text{H}^+$  and  $\text{CoA-S}^-$ . The acetate released by the hydrolysis is stabilized by resonance. Acetyl-CoA is a particularly important metabolic intermediate; other acyl-CoA species figure prominently in lipid metabolism.

The important coenzymes we have met in this chapter— $\text{NAD}^+$ ,  $\text{NADP}^+$ ,  $\text{FAD}$ , and coenzyme A—share an important structural feature: all contain ADP. In  $\text{NADP}^+$ , there is an additional phosphate group at the 2' position of the ribose group of ADP. In CoA, the additional phosphate group is at the 3' position.

Like catabolism, anabolism proceeds in stages. Unlike catabolism, which releases energy, anabolism requires energy. The ATP produced by catabolism is hydrolyzed to release the needed energy. Reactions in which metabolites are reduced are part of anabolism; they require reducing agents, such as  $\text{NADH}$ ,  $\text{NADPH}$ , and  $\text{FADH}_2$ , all of which are the reduced forms of coenzymes mentioned in this chapter. In their oxidized forms, these coenzymes serve as the intermediate oxidizing agents needed in catabolism. In their reduced form,





■ **FIGURE 15.12** The role of electron transfer and ATP production in metabolism. NAD<sup>+</sup>, FAD, and ATP are constantly recycled.

the same coenzymes provide the “reducing power” needed for the anabolic processes of biosynthesis; in this case, the coenzymes act as reducing agents.

We are now in a position to expand on our earlier statements about the natures of anabolism and catabolism. Figure 15.12 is an outline of metabolic pathways that explicitly points out two important features of metabolism: the role of electron transfer and the role of ATP in the release and utilization of energy. Even though this outline is more extended than the one in Figure 15.1, it is still very general. The more important specific pathways have been studied in detail, and some are still the subjects of active research. We shall discuss some of the most important metabolic pathways in the remainder of this textbook.

## SUMMARY

**What are standard states?** A change in free energy under any set of conditions can be compared to the free-energy change under standard conditions ( $\Delta G^\circ$ ), in which the concentration of all reactants in solution is set at 1 *M*.

**What do standard states have to do with free-energy changes?** Free-energy changes under standard conditions can be related to the equilibrium constant of a reaction by the equation  $\Delta G^\circ = -RT \ln K_{\text{eq}}$ .

**Why do we need a modified standard state for biochemical applications?** Since biochemical reactions do not occur naturally at a hydrogen-ion concentration of 1 *M*, a biochemical

standard state ( $\Delta G'^\circ$ ) is often used, where the  $[\text{H}^+]$  is set at  $1 \times 10^{-7} \text{ M}$  (pH = 7.0).

**What is metabolism?** The reactions of the biomolecules in the cell constitute metabolism. The breakdown of larger molecules to smaller ones is called catabolism. The reaction of small molecules to produce larger and more complex molecules is called anabolism. Catabolism and anabolism are separate pathways, not the reverse of each other. Metabolism is the biochemical basis of all life processes.

**How are oxidation and reduction involved in metabolism?** Catabolism is an oxidative process that releases energy;

anabolism is a reductive process that requires energy. Oxidation–reduction (redox) reactions are those in which electrons are transferred from a donor to an acceptor. Oxidation is the loss of electrons, and reduction is the gain of electrons.

### What are the reactions of key oxidation–reduction coenzymes?

Many biologically important redox reactions involve coenzymes, such as NADH and FADH<sub>2</sub>. These coenzymes appear in many reactions as one of the half reactions that can be written for a redox reaction.

### How do energy-producing reactions allow energy-requiring reactions to take place?

The coupling of energy-producing reactions and energy-requiring reactions is a central feature in the metabolism of all organisms. In catabolism, oxidative

reactions are coupled to the endergonic production of ATP by phosphorylation of ADP. Aerobic metabolism is a more efficient means of using the chemical energy of nutrients than is anaerobic metabolism. In anabolism, the exergonic hydrolysis of the high-energy bond of ATP releases the energy needed to drive endergonic reductive reactions.

### Why is coenzyme A such a good example of activation?

Metabolism proceeds in stages, and the many stages allow for the efficient production and use of energy. The process of activation, producing high-energy intermediates, occurs in many metabolic pathways. The formation of thioester linkages by the reaction of carboxylic acids with coenzyme A is an example of the activation process that occurs in several pathways.

## REVIEW EXERCISES

 Interactive versions of these problems are assignable in OWL

### 15.1 Standard States for Free-Energy Changes

- Recall** Is there a connection between the free-energy change for a reaction and its equilibrium constant? If there is a connection, what is it?
- Reflect and Apply** What do the following indicators tell you about whether a reaction can proceed as written?
  - The standard free-energy change is positive.
  - The free-energy change is positive.
  - The reaction is exergonic.
- Reflect and Apply** Consider the reaction



$$K_{\text{eq}} = \frac{[\text{glucose}][\text{P}_i]}{[\text{glucose-6-P}]}$$

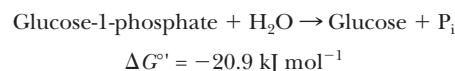
The  $K_{\text{eq}}$  at pH 8.5 and 38°C is 122. Can you determine the rate of the reaction from this information?

### 15.2 A Modified Standard State for Biochemical Applications

- Recall** Why is it necessary to define a modified standard state for biochemical applications of thermodynamics?
- Recall** Which of the following statements is (are) true about the modified standard state for biochemistry? For each, explain why or why not.
  - $[\text{H}^+] = 1 \times 10^{-7} \text{ M}$ , not 1 M.
  - The concentration of any solute is  $1 \times 10^{-7} \text{ M}$ .
- Recall** How can you tell if the standard Gibbs free energy given for a reaction is for chemical standard states or biological standard states?
- Recall** Can the thermodynamic property  $\Delta G^\circ$  be used to predict the speed of a reaction in a living organism? Why or why not?
- Mathematical** Calculate  $\Delta G^\circ$  for the following values of  $K_{\text{eq}}$ :  $1 \times 10^4$ ,  $1$ ,  $1 \times 10^{-6}$ .
- Mathematical** For the hydrolysis of ATP at 25°C (298 K) and pH 7,  $\text{ATP} + \text{H}_2\text{O} \rightarrow \text{ADP} + \text{P}_i + \text{H}^+$ , the standard free energy of hydrolysis ( $\Delta G^\circ$ ) is  $-30.5 \text{ kJ mol}^{-1}$  ( $-7.3 \text{ kcal mol}^{-1}$ ), and the standard enthalpy change ( $\Delta H^\circ$ ) is  $-20.1 \text{ kJ mol}^{-1}$  ( $-4.8 \text{ kcal mol}^{-1}$ ).

Calculate the standard entropy change ( $\Delta S^\circ$ ) for the reaction, in both joules and calories. Why is the positive sign of the answer to be expected in view of the nature of the reaction? *Hint:* You may want to review some material from Chapter 1.

- Mathematical** Consider the reaction  $\text{A} \rightleftharpoons \text{B} + \text{C}$ , where  $\Delta G^\circ = 0.00$ .
  - What is the value of  $\Delta G$  (not  $\Delta G^\circ$ ) when the initial concentrations of A, B, and C are 1 M,  $10^{-3} \text{ M}$ , and  $10^{-6} \text{ M}$ ?
  - Try the same calculations for the reaction  $\text{D} + \text{E} \rightleftharpoons \text{F}$ , for the same relative order of concentrations.
  - Try the same calculations for the reaction  $\text{G} \rightleftharpoons \text{H}$ , if the concentrations are 1 M and  $10^{-3} \text{ M}$  for G and H, respectively.
- Mathematical** Compare your answers for parts (a) and (b) with that for part (c) in Question 10. What do your answers to parts (a), (b), and (c) say about the influence of concentrations of reactants and products on reactions?
- Mathematical** The  $\Delta G^\circ$  for the reaction  $\text{Citrate} \rightarrow \text{Isocitrate}$  is  $+6.64 \text{ kJ mol}^{-1} = +1.59 \text{ kcal mol}^{-1}$ . The  $\Delta G^\circ$  for the reaction  $\text{Isocitrate} \rightarrow \alpha\text{-Ketoglutarate}$  is  $-267 \text{ kJ mol}^{-1} = -63.9 \text{ kcal mol}^{-1}$ . What is the  $\Delta G^\circ$  for the conversion of citrate to  $\alpha$ -ketoglutarate? Is that reaction exergonic or endergonic, and why?
- Mathematical** If a reaction can be written  $\text{A} \rightarrow \text{B}$ , and the  $\Delta G^\circ$  is  $20 \text{ kJ mol}^{-1}$ , what would the substrate/product ratio have to be for the reaction to be thermodynamically favorable?
- Mathematical** All the organophosphate compounds listed in Table 15.1 undergo hydrolysis reactions in the same way as ATP. The following equation illustrates the situation for glucose-1-phosphate.



Using the free-energy values in Table 15.1, predict whether the following reactions will proceed in the direction written, and calculate the  $\Delta G^\circ$  for the reaction, assuming that the reactants are initially present in a 1:1 molar ratio.

- $\text{ATP} + \text{Creatine} \rightarrow \text{Creatine phosphate} + \text{ADP}$
- $\text{ATP} + \text{Glycerol} \rightarrow \text{Glycerol-3-phosphate} + \text{ADP}$
- $\text{ATP} + \text{Pyruvate} \rightarrow \text{Phosphoenolpyruvate} + \text{ADP}$
- $\text{ATP} + \text{Glucose} \rightarrow \text{Glucose-6-phosphate} + \text{ADP}$

15. **Reflect and Apply** Can you use the equation  $\Delta G^\circ = -RT \ln K_{eq}$  to get the  $\Delta G^\circ$  from the information in Question 3?
16. **Reflect and Apply** Why are  $\Delta G^\circ$  values not rigorously applicable to biochemical systems?

### 15.3 The Nature of Metabolism

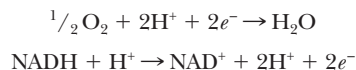
17. **Recall** Organize the following words into two related groups: catabolism, energy-requiring, reductive, anabolism, oxidative, energy-yielding.
18. **Biochemical Connections** Comment on the statement that the existence of life is a violation of the second law of thermodynamics, adding concepts from this chapter to those we saw in Chapter 1.
19. **Reflect and Apply** Would you expect the production of sugars by plants in photosynthesis to be an exergonic or endergonic process? Give the reason for your answer.
20. **Reflect and Apply** Would you expect the biosynthesis of a protein from the constituent amino acids in an organism to be an exergonic or endergonic process? Give the reason for your answer.
21. **Reflect and Apply** Adult humans synthesize large amounts of ATP in the course of a day, but their body weights do not change significantly. In the same time period, the structures and compositions of their bodies also do not change appreciably. Explain this apparent contradiction.

### 15.4 The Role of Oxidation and Reduction in Metabolism

22. **Recall** Identify the molecules oxidized and reduced in the following reactions and write the half reactions.
- (a)  $\text{CH}_3\text{CH}_2\text{CHO} + \text{NADH} \rightarrow \text{CH}_3\text{CH}_2\text{CH}_2\text{OH} + \text{NAD}^+$
- (b)  $\text{Cu}^{2+}(\text{aq}) + \text{Fe}^{2+}(\text{aq}) \rightarrow \text{Cu}^+(\text{aq}) + \text{Fe}^{3+}(\text{aq})$
23. **Recall** For each of the reactions in Question 22, give the oxidizing agent and reducing agents.

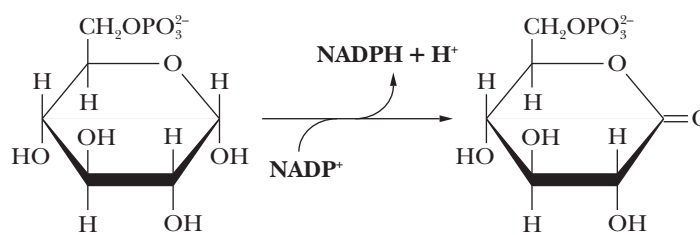
### 15.5 Coenzymes in Biologically Important Oxidation-Reduction Reactions

24. **Recall** What structural feature do  $\text{NAD}^+$ ,  $\text{NADP}^+$ , and  $\text{FAD}$  have in common?
25. **Recall** What is the structural difference between  $\text{NADH}$  and  $\text{NADPH}$ ?
26. **Recall** How does the difference between  $\text{NADH}$  and  $\text{NADPH}$  affect the reactions in which they are involved?
27. **Reflect and Apply** Which of the following statements are true? For each, explain why or why not.
- (a) All coenzymes are electron-transfer agents.
- (b) Coenzymes do not contain phosphorus or sulfur.
- (c) Generating ATP is a way of storing energy.
28. **Reflect and Apply** A biochemical reaction transfers  $60 \text{ kJ mol}^{-1}$  ( $15 \text{ kcal mol}^{-1}$ ) of energy. What general process most likely would be involved in this transfer? What cofactor (or cosubstrate) likely would be used? Which cofactor probably would not be used?
29. **Reflect and Apply** The following half reactions play important roles in metabolism.



Which of these two is a half reaction of oxidation? Which one is a half reaction of reduction? Write the equation for the overall reaction. Which reagent is the oxidizing agent (electron acceptor)? Which reagent is the reducing agent (electron donor)?

30. **Reflect and Apply** Draw  $\text{NAD}^+$  and  $\text{FAD}$  showing where the electrons and hydrogens go when the molecules are reduced.
31. **Reflect and Apply** There is a reaction in carbohydrate metabolism in which glucose-6-phosphate reacts with  $\text{NADP}^+$  to give 6-phosphoglucono- $\delta$ -lactone and  $\text{NADPH}$ .

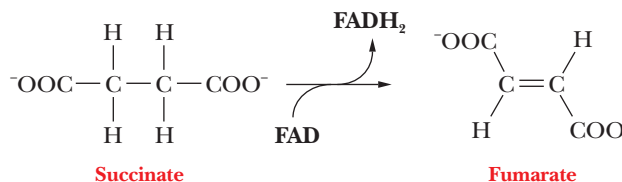


Glucose-6-phosphate

6-Phosphoglucono- $\delta$ -lactone

In this reaction, which substance is oxidized, and which is reduced? Which substance is the oxidizing agent, and which is the reducing agent?

32. **Reflect and Apply** There is a reaction in which succinate reacts with  $\text{FAD}$  to give fumarate and  $\text{FADH}_2$ .



Succinate

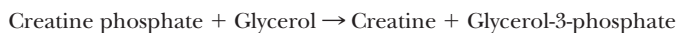
Fumarate

In this reaction, which substance is oxidized, and which is reduced? Which substance is the oxidizing agent, and which is the reducing agent?

33. **Reflect and Apply** Suggest a reason that catabolic pathways generally produce  $\text{NADH}$  and  $\text{FADH}_2$ , whereas anabolic pathways generally use  $\text{NADPH}$ .

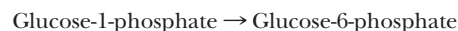
### 15.6 Coupling of Production and Use of Energy

34. **Mathematical** What substrate concentrations would be necessary to make the reaction in part (c) of Question 14 a favorable reaction?
35. **Mathematical** Using the data in Table 15.1, calculate the value of  $\Delta G^\circ$  for the following reaction.



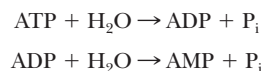
*Hint:* This reaction proceeds in stages. ATP is formed in the first step, and the phosphate group is transferred from ATP to glycerol in the second step.

36. **Mathematical** Using information from Table 15.1, calculate the value of  $\Delta G^\circ$  for the following reaction.

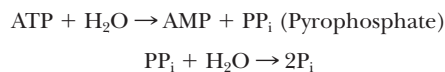


37. **Mathematical** Show that the hydrolysis of ATP to AMP and  $2\text{P}_i$  releases the same amount of energy by either of the two following pathways.

*Pathway 1*



*Pathway 2*



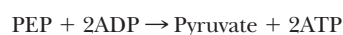
38. **Mathematical** The standard free-energy change for the reaction



is  $+1.7 \text{ kJ mol}^{-1}$ . From this information and that in Table 15.1, calculate the  $\Delta G^\circ$  for the reaction



39. **Reflect and Apply** What are the usual ionic forms of ATP and ADP in typical cells? Does this information have any bearing on the free-energy change for the conversion of ATP to ADP?
40. **Reflect and Apply** Comment on the free energy of hydrolysis of the phosphate bond of ATP ( $-30.5 \text{ kJ mol}^{-1}$ ;  $-7.3 \text{ kcal mol}^{-1}$ ) relative to those of other organophosphates (e.g., sugar phosphates, creatine phosphate).
41. **Reflect and Apply** A friend has seen creatine supplements for sale in a health-food store and asks why. What do you tell your friend?
42. **Reflect and Apply** Would you expect an increase or a decrease of entropy to accompany the hydrolysis of phosphatidylcholine to the constituent parts (glycerol, two fatty acids, phosphoric acid, and choline)? Why?
43. **Reflect and Apply** Explain and show why phosphoenolpyruvate is a high-energy compound.
44. **Reflect and Apply** A very favorable reaction is the production of ATP and pyruvate from ADP and phosphoenolpyruvate. Given the standard free-energy change for this coupled reaction, why does the following reaction not occur?



45. **Reflect and Apply** Short periods of exercise, such as sprints, are characterized by lactic acid production and the condition known as oxygen debt. Comment on this fact in light of the material discussed in this chapter.

### 15.7 Coenzyme A in Activation of Metabolic Pathways

46. **Reflect and Apply** Why is the process of activation a useful strategy in metabolism?
47. **Reflect and Apply** What is the molecular logic that makes a pathway with a number of comparatively small energy changes more likely than a single reaction with a large energy change?
48. **Reflect and Apply** Why are thioesters considered high-energy compounds?
49. **Reflect and Apply** Explain why several biochemical pathways start by putting a coenzyme A onto the molecule that initiates the pathway.
50. **Reflect and Apply** This is a conjectural question: If the reactive part of coenzyme A is the thioester, why is the molecule so complicated?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





# Carbohydrates

# 16

Like bread and pasta, fruits and vegetables are sources of carbohydrates, which provide energy.

© Walter Wick/Getty Images

## 16.1 Sugars: Their Structures and Stereochemistry

When the word *carbohydrate* was coined, it originally referred to compounds of the general formula  $C_n(H_2O)_n$ . However, only the simple sugars, or **monosaccharides**, fit this formula exactly. The other types of carbohydrates, oligosaccharides and polysaccharides, are based on the monosaccharide units and have slightly different general formulas. **Oligosaccharides** are formed when a few (Greek *oligos*) monosaccharides are linked; polysaccharides are formed when many (Greek *polys*) monosaccharides are bonded together. The reaction that adds monosaccharide units to a growing carbohydrate molecule involves the loss of one  $H_2O$  for each new link formed, accounting for the difference in the general formula.

Many commonly encountered carbohydrates are polysaccharides, including glycogen, which is found in animals, and starch and cellulose, which occur in plants. Carbohydrates play a number of important roles in biochemistry. First, they are major energy sources (Chapters 17 through 20 are devoted to carbohydrate metabolism). Second, oligosaccharides play a key role in processes that take place on the surfaces of cells, particularly in cell–cell interactions and immune recognition. In addition, polysaccharides are essential structural components of several classes of organisms. Cellulose is a major component of grass and trees, and other polysaccharides are major components of bacterial cell walls.

### What is unique about the structures of sugars?

The building blocks of all carbohydrates are the simple sugars called **monosaccharides**. A monosaccharide can be a polyhydroxy aldehyde (**aldose**) or a polyhydroxy ketone (**ketose**). The simplest monosaccharides contain three carbon atoms and are called trioses (*tri* meaning “three”). *Glyceraldehyde* is the aldose with three carbons (an aldotriose), and *dihydroxyacetone* is the ketose with three carbon atoms (a ketotriose). Figure 16.1 shows these molecules.

Aldoses with four, five, six, and seven carbon atoms are called aldotetroses, aldopentoses, aldohexoses, and aldohexoses, respectively. The corresponding ketoses are ketotetroses, ketopentoses, ketohexoses, and ketoheptoses. Six-carbon sugars are the most abundant in nature, but two five-carbon sugars, ribose and deoxyribose, occur in the structures of RNA and DNA, respectively. Four-carbon and seven-carbon sugars play roles in photosynthesis and other metabolic pathways.

We have already seen (in Section 3.1) that some molecules are not superimposable on their mirror images and that these mirror images are **optical isomers (stereoisomers)** of each other. A chiral (asymmetric) carbon atom is the usual source of optical isomerism, as was the case with amino acids. The simplest carbohydrate that contains a chiral carbon is glyceraldehyde, which can exist in two isomeric forms that are mirror images of each other [Figure 16.1(2) and (3)]. Note that the two forms differ in the position of the hydroxyl group bonded to the central carbon. (Dihydroxyacetone does not contain a

## Chapter Outline

### 16.1 Sugars: Their Structures and Stereochemistry

- What is unique about the structures of sugars?
- What happens if a sugar forms a cyclic molecule?

### 16.2 Reactions of Monosaccharides

- What are some oxidation–reduction reactions of sugars?
- What are some important esterification reactions of sugars?
- What are glycosides, and how do they form?
- What are some other important derivatives of sugars?

### 16.3 Some Important Oligosaccharides

- What makes sucrose an important compound?
- Are any other disaccharides important to us?

### 16.4 Structures and Functions of Polysaccharides

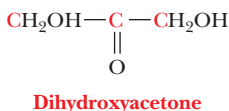
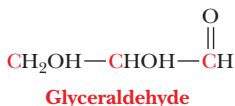
- How do cellulose and starch differ from one another?
- Is there more than one form of starch?
- How is glycogen related to starch?
- What is chitin?
- What role do polysaccharides play in the structure of cell walls?
- Do polysaccharides play any specific roles in connective tissue?

### 16.5 Glycoproteins

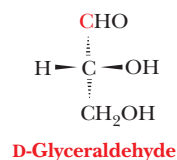
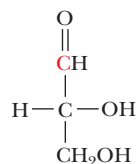
- How are carbohydrates important in the immune response?

**OWL** Online homework for this chapter may be assigned in OWL.

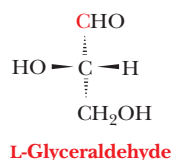
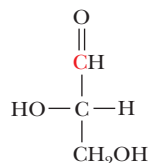
- 1 A comparison of glyceraldehyde (an aldotriose) and dihydroxyacetone (a ketotriose).



- 2 The structure of D-glyceraldehyde and a space-filling model of D-glyceraldehyde.



- 3 The structure of L-glyceraldehyde and a space-filling model of L-glyceraldehyde.

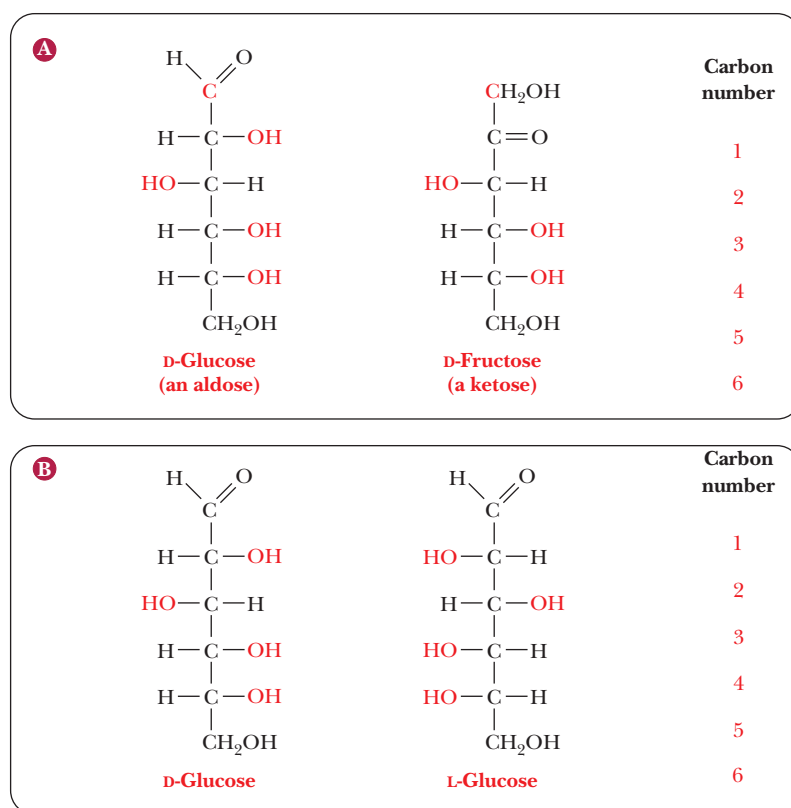


■ **FIGURE 16.1** The structures of the simplest carbohydrates, the trioses. (Leonard Lessin/Waldo Feng/Mt. Sinai CORE)

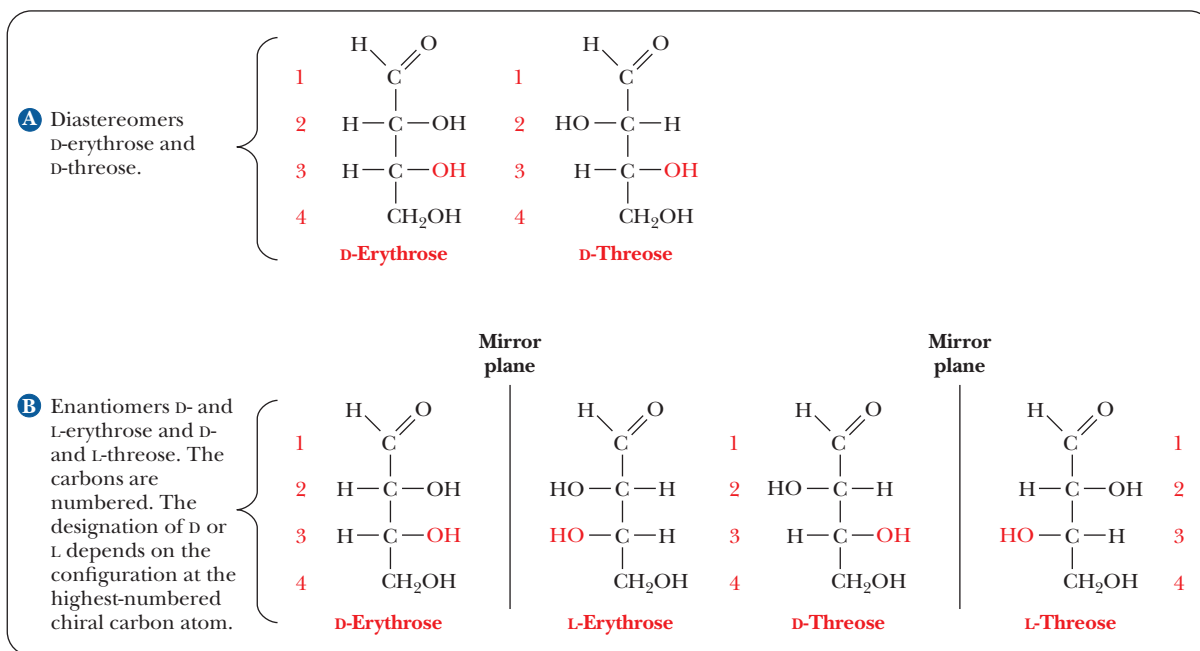
chiral carbon atom and does not exist in nonsuperimposable mirror-image forms.) The two forms of glyceraldehyde are designated D-glyceraldehyde and L-glyceraldehyde. Mirror-image stereoisomers are also called **enantiomers**, and D-glyceraldehyde and L-glyceraldehyde are enantiomers of each other. Certain conventions are used for two-dimensional drawings of the three-dimensional structures of stereoisomers. The dashed wedges represent bonds directed away from the viewer, below the plane of the paper, and the solid wedges represent bonds directed oppositely, toward the viewer and out of the plane of the paper. The **configuration** is the three-dimensional arrangement of groups around a chiral carbon atom, and stereoisomers differ from each other in configuration. The *D,L* system to denote stereochemistry is widely used by biochemists. Organic chemists tend to use a more recent one, designated the *R,S* system. There is not a one-to-one correspondence between the two systems. For example, some D-isomers are *R*, and some are *S*.

The two enantiomers of glyceraldehyde are the only possible stereoisomers of three-carbon sugars, but the possibilities for stereoisomerism increase as the number of carbon atoms increases. To show the structures of the resulting molecules, we need to say more about the convention for a two-dimensional

perspective of the molecular structure, which is called the **Fischer projection** method, after the German chemist Emil Fischer, who established the structures of many sugars. We shall use some common six-carbon sugars for purposes of illustration. In a Fischer projection, bonds written vertically on the two-dimensional paper represent bonds directed *behind* the paper in three dimensions, whereas bonds written horizontally represent bonds directed *in front of* the paper in three dimensions. Figure 16.2 shows that the most highly oxidized carbon—in this case, the one involved in the aldehyde group—is written at the “top” and is designated carbon 1, or C-1. In the ketose shown, the ketone group becomes C-2, the carbon atom next to the “top.” Most common sugars are aldoses rather than ketoses, so our discussion will focus mainly on aldoses. The other carbon atoms are numbered in sequence from the “top.” The designation of the configuration as L or D depends on the arrangement at the chiral carbon with the highest number. In the cases of both glucose and fructose, this is C-5. In the Fischer projection of the D configuration, the hydroxyl group is on the right of the highest-numbered chiral carbon, whereas the hydroxyl group is on the left of the highest-numbered chiral carbon in the L configuration. Let us see what happens as another carbon is added to glyceraldehyde to give a four-carbon sugar. In other words, what are the possible stereoisomers for an aldotetrose? The aldotetroses (Figure 16.3) have two chiral carbons, C-2 and C-3, and there are  $2^2$ , or four, possible stereoisomers. Two of the isomers have the D configuration, and two have the L configuration. The two D isomers have the same configuration at C-3, but they differ in configuration (arrangement of the —OH group) at the other chiral carbon, C-2. These two isomers are called D-erythrose and D-threose. They are not superimposable on each other, but neither are they mirror images of each other. Such nonsuperimposable, non-mirror-image stereoisomers are called **diastereomers**. The two L isomers are L-erythrose and L-threose. L-Erythrose is the enantiomer (mirror image) of D-erythrose, and L-threose is the enantiomer of D-threose. L-Threose is a diastereomer of both D- and L-erythrose, and L-erythrose is a diastereomer of both D- and L-threose. Diastereomers that differ from each other in the configuration



■ **FIGURE 16.2** Numbering of carbon atoms in sugars. (a) Examples of an aldose (D-glucose) and a ketose (D-fructose), showing the numbering of carbon atoms. (b) A comparison of the structures of D-glucose and L-glucose.



■ **FIGURE 16.3** Stereoisomers of an aldotetrose.

at only one chiral carbon are called **epimers**; D-erythrose and D-threose are epimers.

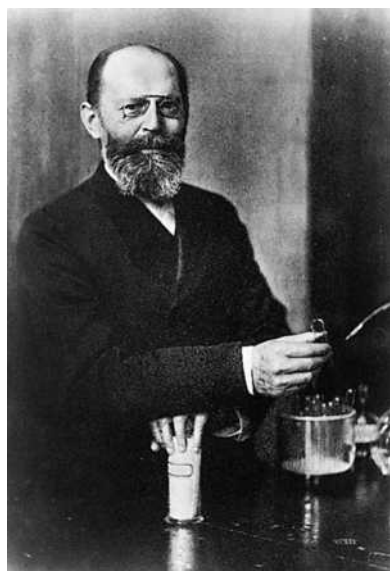
Aldopentoses have three chiral carbons, and there are  $2^3$ , or 8, possible stereoisomers—four D forms and four L forms. Aldohexoses have four chiral carbons and  $2^4$ , or 16, stereoisomers—eight D forms and eight L forms (Figure 16.4). Some of the possible stereoisomers are much more common in nature than others, and most biochemical discussion centers on the common, naturally occurring sugars. For example, D sugars, rather than L sugars, predominate in nature. Most of the sugars we encounter in nature, especially in foods, contain either five or six carbon atoms. We shall discuss D-glucose (an aldohexose) and D-ribose (an aldopentose) far more than many other sugars. Glucose is a ubiquitous energy source, and ribose plays an important role in the structure of nucleic acids.

### What happens if a sugar forms a cyclic molecule?

Sugars, especially those with five or six carbon atoms, normally exist as cyclic molecules rather than as the open-chain forms we have shown so far. The cyclization takes place as a result of interaction between the functional groups on distant carbons, such as C-1 and C-5, to form a cyclic **hemiacetal** (in aldohexoses). Another possibility (Figure 16.5) is interaction between C-2 and C-5 to form a cyclic **hemiketal** (in ketohexoses). In either case, the carbonyl carbon becomes a new chiral center called the **anomeric carbon**. The cyclic sugars can take either of two different forms, designated  $\alpha$  and  $\beta$ , and are called **anomers** of each other.

The Fischer projection of the  $\alpha$ -anomer of a D sugar has the anomeric hydroxyl group to the right of the anomeric carbon (C—OH), and the  $\beta$ -anomer of a D sugar has the anomeric hydroxyl group to the left of the anomeric carbon (Figure 16.6). The free carbonyl species can readily form either the  $\alpha$ - or  $\beta$ -anomer, and the anomers can be converted from one form to another through the free carbonyl species. In some biochemical molecules, any anomer of a given sugar can be used, but, in other cases, only one anomer occurs. For example, in living organisms, only  $\beta$ -D-ribose and  $\beta$ -D-deoxyribose are found in RNA and DNA, respectively.

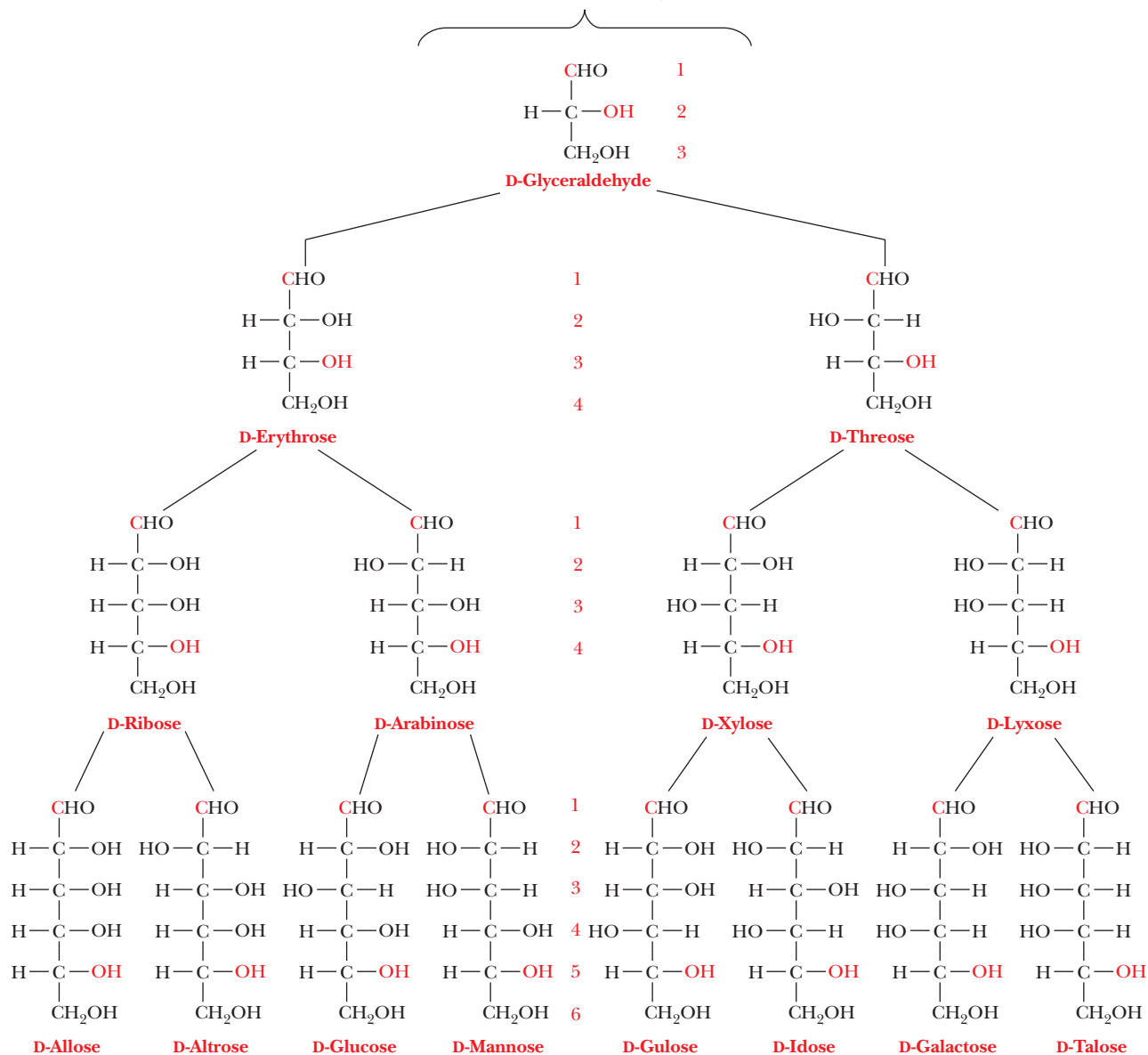
Fischer projection formulas are useful for describing the stereochemistry of sugars, but their long bonds and right-angle bends do not give a realistic picture of the bonding situation in the cyclic forms, nor do they accurately represent the



© Corbis

■ Emil Fischer (1852–1919) was a German-born scientist who won the Nobel Prize in chemistry in 1902 for his studies on sugars, purine derivatives, and peptides.

- A** Aldoses containing from three to six carbon atoms, with the numbering of the carbon atom shown. Note that the figure shows only half the possible isomers. For each isomer shown, there is an enantiomer that is not shown, the L series.

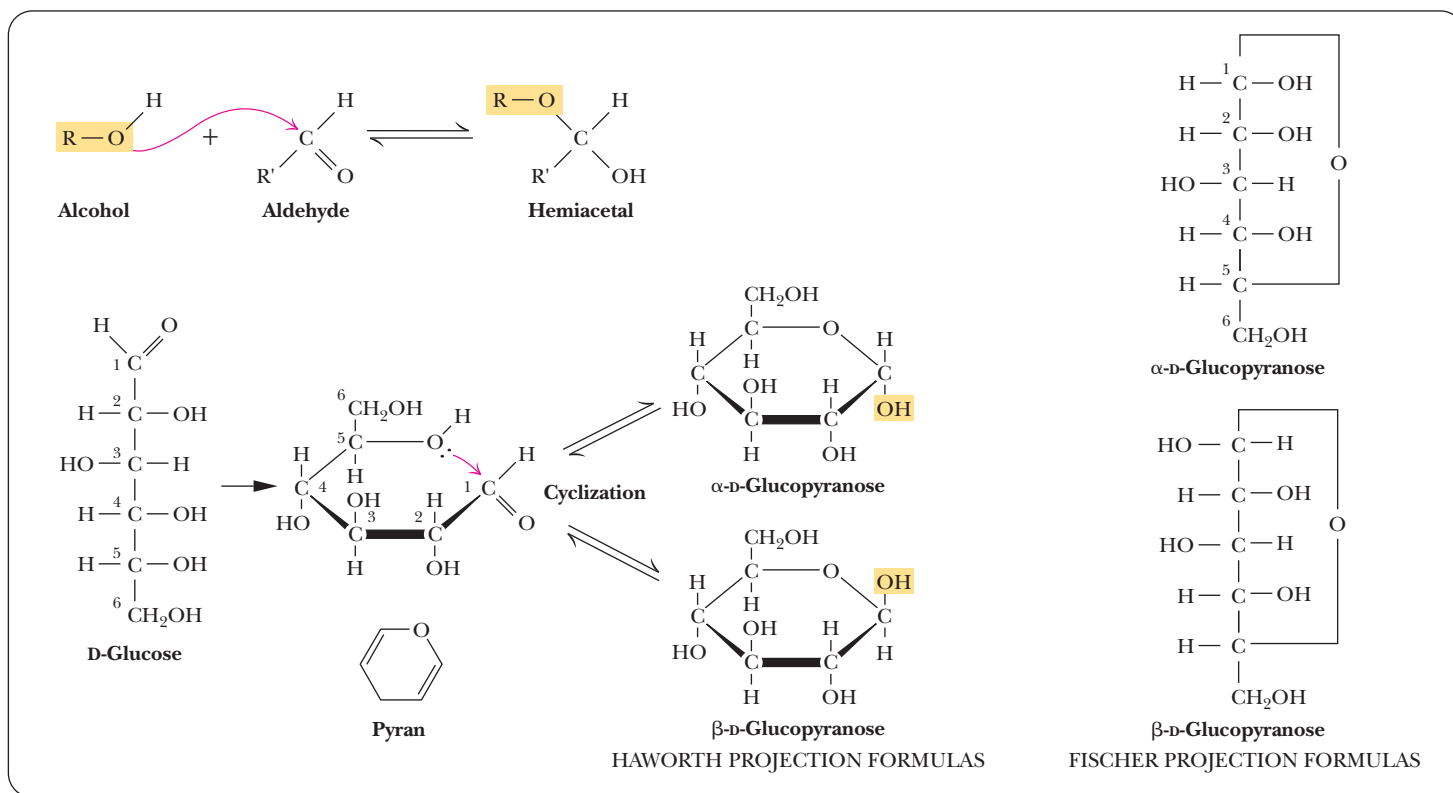


- B** The relationship between mirror images is of interest to mathematicians as well as to chemists. Lewis Carroll (C. L. Dodgson), the author of *Alice's Adventures in Wonderland*, was a contemporary of Emil Fischer.

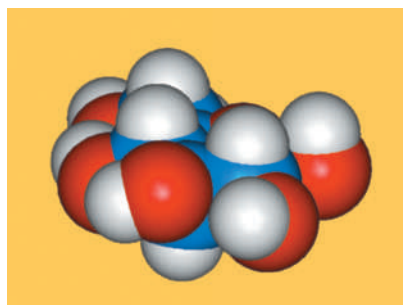


**FIGURE 16.4** Stereochemical relationships among monosaccharides.





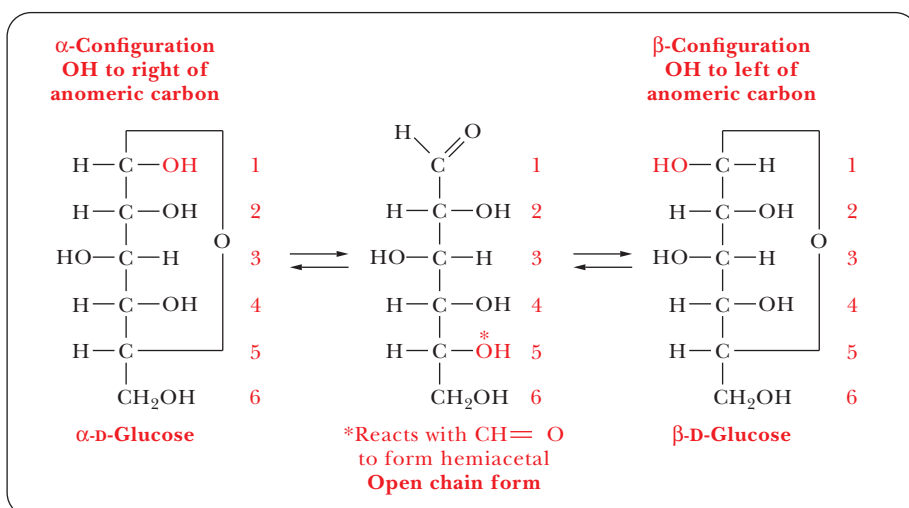
■ **FIGURE 16.5** The linear form of D-glucose undergoes an intramolecular reaction to form a cyclic hemiacetal.



$\beta$ -D-Glucopyranose

overall shape of the molecules. **Haworth projection formulas** are more useful for those purposes. In Haworth projections, the cyclic structures of sugars are shown in perspective drawings as planar five- or six-membered rings viewed nearly edge on. A five-membered ring is called a **furanose** because of its resemblance to furan; a six-membered ring is called a **pyranose** because of its resemblance to pyran [Figure 16.7(a) and (b)]. These cyclic formulas approximate the shapes of the actual molecules better for furanoses than for pyranoses. The five-membered rings of furanoses are in reality very nearly planar, but the six-membered rings of pyranoses actually exist in solution in the chair conformation [Figure 16.7(c)]. The chair conformation is widely shown in textbooks of organic chemistry. This

■ **FIGURE 16.6** Fischer projection formulas of three forms of glucose. Note that the  $\alpha$  and  $\beta$  forms can be converted to each other through the open-chain form. The configuration at carbon 5 determines the D designation.



kind of structure is particularly useful in discussions of molecular recognition. The chair conformation and the Haworth projections are alternative ways of expressing the same information. Even though the Haworth formulas are approximations, they are useful shorthand for the structures of reactants and products in many reactions that we are going to see. The Haworth projections represent the stereochemistry of sugars more realistically than do the Fischer projections, and the Haworth scheme is adequate for our purposes. That is why biochemists use it, even though organic chemists prefer the chair form. We shall continue to use Haworth projections in our discussion of sugars.

For a D sugar, any group that is written to the *right* of the carbon in a Fischer projection has a *downward* direction in a Haworth projection; any group that is written to the *left* in a Fischer projection has an *upward* direction in a Haworth projection. The terminal  $\text{—CH}_2\text{OH}$  group, which contains the carbon atom with the highest number in the numbering scheme, is shown in an upward direction. The structures of  $\alpha$ - and  $\beta$ -D-glucose, which are both pyranoses, and of  $\beta$ -D-ribose, which is a furanose, illustrate this point (Figure 16.8).

## Biochemical Connections NUTRITION AND HEALTH

### Low-Carbohydrate Diets

In the 1970s, the diets that were supposed to be the healthiest were low in fat and high in carbohydrates. “Carbo-loading” was the craze for athletes of all types, genders, and ages, as well as for the average, sedentary person. Over thirty years later, things have changed so much that you can buy a burger wrapped in a piece of lettuce instead of a bun. Why did a macromolecule once thought to be healthy become something people want to avoid? The answer has to do with how glucose, the primary monosaccharide of life, is metabolized. Rising glucose levels in the blood cause a subsequent rise in levels of the hormone insulin. Insulin stimulates cells to take up glucose from the blood so that the cells get the energy and blood-glucose levels remain stable. We now know that insulin also has the unfortunate effect of stimulating fat synthesis and storage and inhibiting fat burning.

Some popular recent diets, such as the Zone Diet and the Atkins Diet, are based on keeping the carbohydrate levels low so that insulin levels do not rise and stimulate this fat storage. Current popular diet systems, such as NutriSystem and Weight Watchers, are also marketing their products by focusing on the type and

quantity of carbohydrates using a “glycemic index” to distinguish between “good carbs” and “bad carbs.”

In the case of athletes, however, little evidence suggests that a low-carbohydrate diet is effective for athletic performance, because of the extended time needed to replenish muscle and liver glycogen when the athlete is not on a high-carbohydrate diet.

Early in 2010, a study found that, for overall weight loss, a low-carbohydrate diet was as effective as a regimen that combined weight-loss drugs with a low-fat diet. Useful as this result was, it was accompanied by another finding with wide-ranging implications. Many of the participants in this study had chronic health problems such as high blood pressure or diabetes in addition to overweight or obesity. The group on the low-carbohydrate diet had a significant drop in blood pressure compared to the comparison group on a low-fat diet with weight-loss drugs. Weight loss in and of itself usually produces a drop in blood pressure, with a reduction or an elimination of the need for medication. However, only 21% of the low-fat group in this study achieved a lowering of blood pressure to this extent, compared to nearly half of those in the low-carbohydrate group.



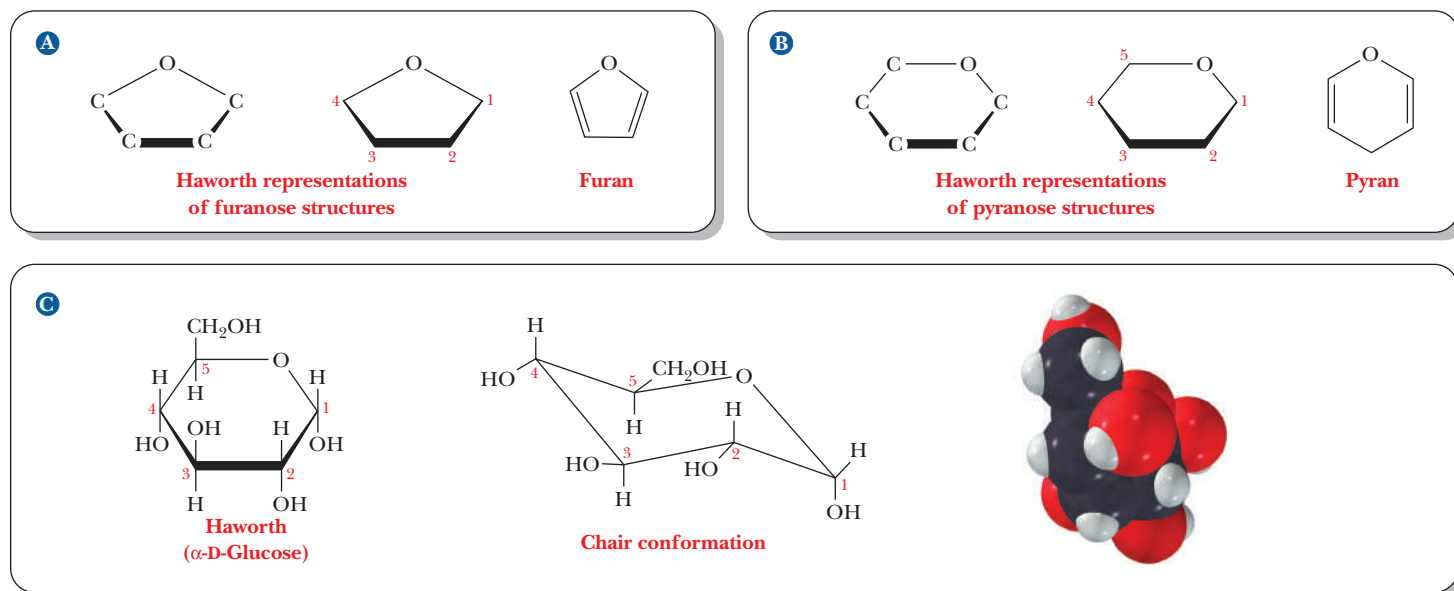
© Cengage Learning/Charles D. Winters

■ Some food products make a point of their low glycemic index.

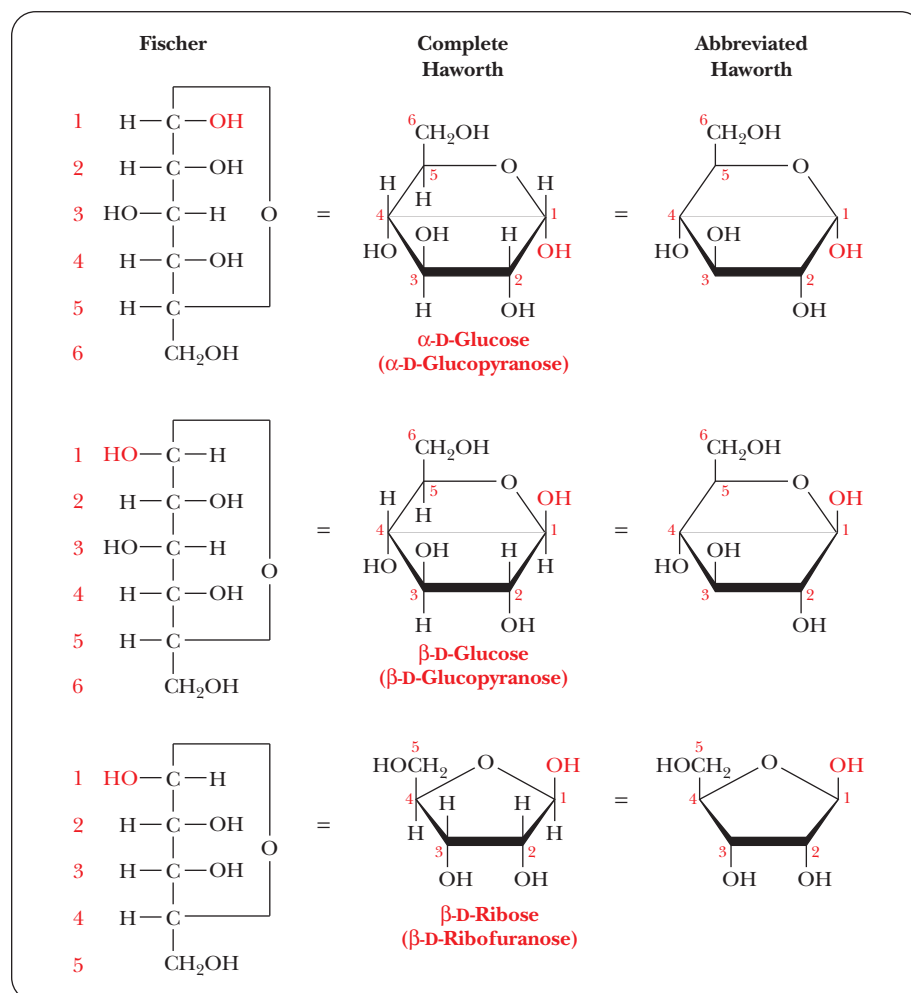


Image copyright Kokhanchikov, 2010. Used under license from Shutterstock.com.

■ Weight loss on low-carbohydrate diets can lead to a marked drop in blood pressure.



■ **FIGURE 16.7** Haworth representations of sugar structures. (a) A comparison of the structure of furan with Haworth representations of furanoses. (b) A comparison of the structure of pyran with Haworth representations of pyranoses. (c)  $\alpha$ -D-Glucopyranose in the Haworth representation (left), in the chair conformation (middle), and as a space-filling model (right). (Leonard Lessin/Waldo Feng/Mt. Sinai CORE)



■ **FIGURE 16.8** A comparison of the Fischer, complete Haworth, and abbreviated Haworth representations of  $\alpha$ - and  $\beta$ -D-glucose (glucopyranose) and  $\beta$ -D-ribose (ribofuranose). In the Haworth representation, the  $\alpha$ -anomer is represented with the OH group (red) downward, and the  $\beta$ -anomer is represented with the OH group (red) upward.

Note that, in the  $\alpha$ -anomer, the hydroxyl on the anomeric carbon is on the opposite side of the ring from the terminal  $\text{—CH}_2\text{OH}$  group (i.e., pointing down). In the  $\beta$ -anomer, it is on the same side of the ring (pointing up). The same convention holds for  $\alpha$ - and  $\beta$ -anomers of furanoses.

## 16.2 Reactions of Monosaccharides

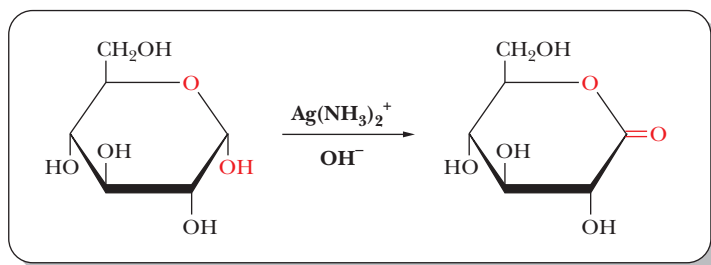
### What are some oxidation–reduction reactions of sugars?

Oxidation and reduction reactions of sugars play key roles in biochemistry. Oxidation of sugars provides energy for organisms to carry out their life processes; the highest yield of energy from carbohydrates occurs when sugars are completely oxidized to  $\text{CO}_2$  and  $\text{H}_2\text{O}$  in aerobic processes. The reverse of complete oxidation of sugars is the reduction of  $\text{CO}_2$  and  $\text{H}_2\text{O}$  to form sugars, a process that takes place in photosynthesis.

Several oxidation reactions of sugars are of some importance in laboratory practice because they can be used to identify sugars. Aldehyde groups can be oxidized to give the carboxyl group that is characteristic of acids, and this reaction is the basis of a test for the presence of aldoses. When the aldehyde is oxidized, some oxidizing agent must be reduced. Aldoses are called **reducing sugars** because of this type of reaction; ketoses can also be reducing sugars because they isomerize to aldoses. In the cyclic form, the compound produced by oxidation of an aldose is a *lactone* (a cyclic ester linking the carboxyl group and one of the sugar alcohols, as shown in Figure 16.9). A lactone of considerable importance to humans is discussed in the Biochemical Connections box on page 461.

Two types of reagent are used in the laboratory to detect the presence of reducing sugars. The first of these is Tollens reagent, which uses the silver ammonia complex ion,  $\text{Ag}(\text{NH}_3)_2^+$ , as the oxidizing agent. A silver mirror is deposited on the wall of the test tube if a reducing sugar is present, as a result of the  $\text{Ag}^+$  in the complex ion being reduced to free silver metal (Figure 16.10). A more recent method for detection of glucose, but not other reducing sugars, is based on the use of the enzyme glucose oxidase, which is specific for glucose.

In addition to oxidized sugars, there are some important reduced sugars. In *deoxy sugars*, a hydrogen atom is substituted for one of the hydroxyl groups of the sugar. One of these deoxy sugars is L-fucose (L-6-deoxygalactose), which is found in the carbohydrate portions of some glycoproteins (Figure 16.11), including the ABO blood-group antigens. The name *glycoprotein* indicates that these substances are conjugated proteins that contain some carbohydrate group (*glykos* is Greek for “sweet”) in addition to the polypeptide chain. An

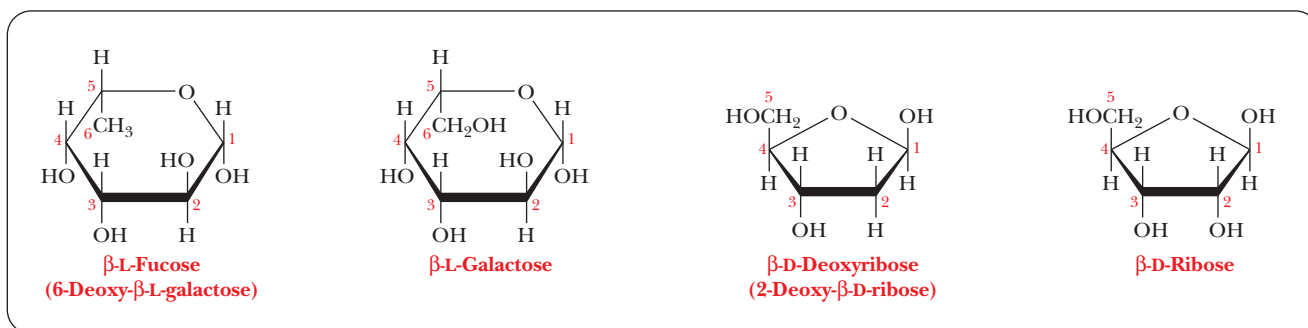


■ **FIGURE 16.9** Oxidation of a sugar to a lactone. An example of an oxidation reaction of sugars: oxidation of  $\alpha$ -D-glucose hemiacetal to give a lactone. Deposition of free silver as a silver mirror indicates that the reaction has taken place.



© Cengage Learning/Charles D. Winters

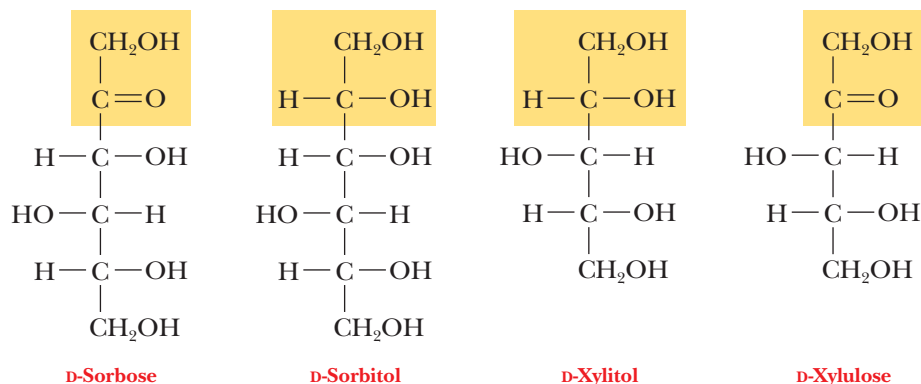
■ **FIGURE 16.10** A silver mirror produced by an aldehyde. After the addition of Tollens reagent to an aldehyde, a silver mirror has been deposited in the inside of this flask.



■ **FIGURE 16.11** Structures of two deoxy sugars. The structures of the parent sugars are shown for comparison.

even more important example of a deoxy sugar is D-2-deoxyribose, the sugar found in DNA (Figure 16.11).

When the carbonyl group of a sugar is reduced to a hydroxyl group, the resulting compound is one of the polyhydroxy alcohols known as *alditols*. Two compounds of this kind, xylitol and sorbitol, derivatives of the sugars xylulose and sorbose, respectively, have commercial importance as sweeteners in sugarless chewing gum and candy.



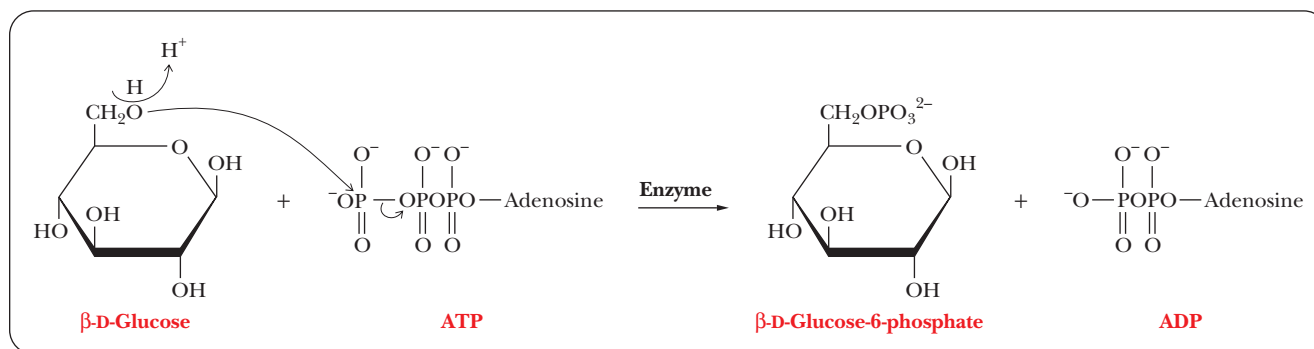
### What are some important esterification reactions of sugars?

The hydroxyl groups of sugars behave exactly like all other alcohols in the sense that they can react with acids and derivatives of acids to form esters. The phosphate esters are particularly important because they are the usual intermediates in the breakdown of carbohydrates to provide energy. Phosphate esters are frequently formed by transfer of a phosphate group from ATP (adenosine triphosphate) to give the phosphorylated sugar and ADP (adenosine diphosphate), as shown in Figure 16.12. Such reactions play an important role in the metabolism of sugars (Section 17.2).

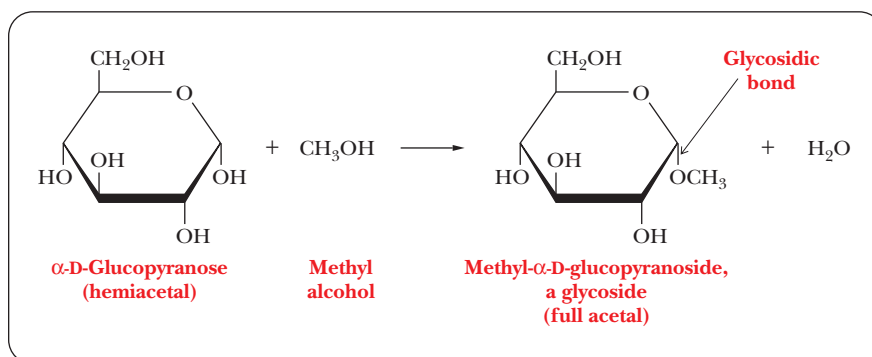
### What are glycosides, and how do they form?

It is possible for a sugar hydroxyl group (ROH) bonded to the anomeric carbon to react with another hydroxyl (R'—OH) to form a glycosidic linkage (R'—O—R). A glycosidic linkage is *not* an ether (the R'—O—R notation is misleading) because glycosides can be hydrolyzed to the original alcohols. This type of reaction involves the anomeric carbon of the sugar in its cyclic form. (Recall that the anomeric carbon is the carbonyl carbon of the open-chain form of the sugar and is the one that becomes a chiral center in the cyclic form.) Stated in a slightly different way, a hemiacetal carbon can react with an alcohol such as methyl alcohol to give a *full acetal*, or **glycoside** (Figure 16.13).





■ **FIGURE 16.12** The formation of a phosphate ester of glucose. ATP is the phosphate group donor. The enzyme specifies the interaction with  $\text{—CH}_2\text{OH}$  on carbon 6.



■ **FIGURE 16.13** An example of the formation of a glycoside. Methyl alcohol ( $\text{CH}_3\text{OH}$ ) and an  $\alpha\text{-D-glucopyranose}$  react to form the corresponding glycoside.

The newly formed bond is called a **glycosidic bond**. The glycosidic bonds discussed in this chapter are *O*-glycosides, with each sugar bonded to an oxygen atom of another molecule. (We encountered *N*-glycosides in Chapter 9 when we discussed nucleosides and nucleotides, in which the sugar is bonded to a nitrogen atom of a base.) Glycosides derived from furanoses are called **furanosides**, and those derived from pyranoses are called **pyranosides**.

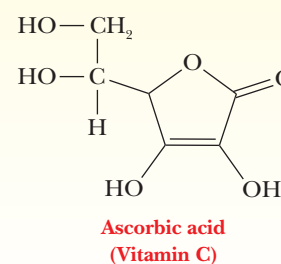
## Biochemical Connections NUTRITION

### Vitamin C Is Related to Sugars

Vitamin C (ascorbic acid) is an unsaturated lactone with a five-membered ring structure. Each carbon is bonded to a hydroxyl group, except for the carboxyl carbon that is involved in the cyclic ester bond. Most animals can synthesize vitamin C; the exceptions are guinea pigs and primates, including humans. As a result, guinea pigs and primates must acquire vitamin C in their diet. Air oxidation of ascorbic acid, followed by hydrolysis of the ester bond, leads to loss of activity as a vitamin. Consequently, a lack of fresh food can cause vitamin C deficiencies, which, in turn, can lead to the disease scurvy (Section 4.3). In this disease, defects in collagen structure cause skin lesions and fragile blood vessels. The presence of hydroxyproline is necessary for collagen stability because of hydrogen-bonded cross-links between collagen strands. Ascorbic acid, in turn, is essential for the activity of prolyl hydroxylase, which converts proline residues in collagen to hydroxyproline. Lack of ascorbic acid eventually leads to the fragile collagen responsible for the symptoms of scurvy.

The British navy introduced citrus fruit into the diet of sailors in the 18th century to prevent scurvy during long sea voyages, and many people still consume citrus fruit for its vitamin C.

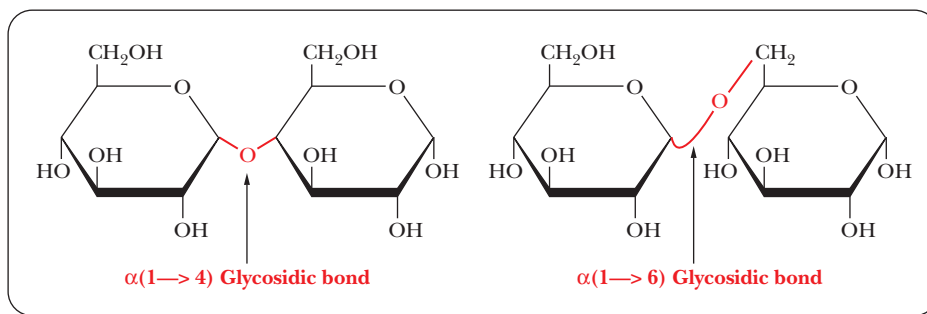
Potatoes are another important source of vitamin C, not because potatoes contain a *high* concentration of ascorbic acid but because we eat so many potatoes.



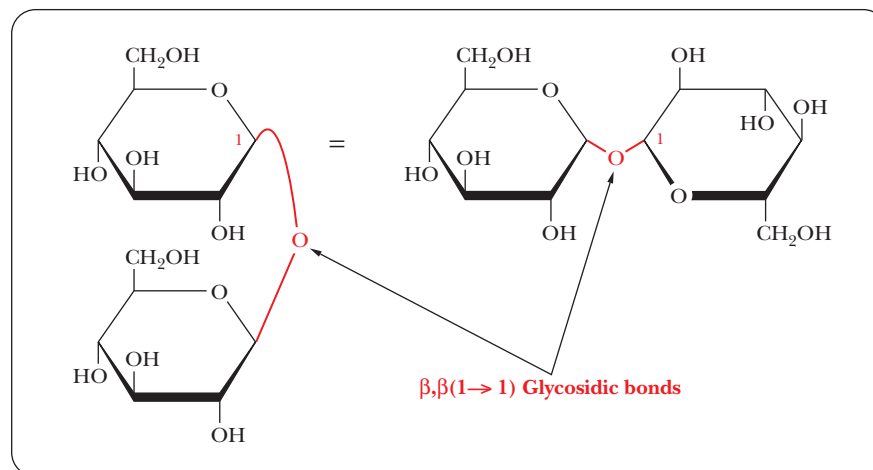
Glycosidic bonds between monosaccharide units are the basis for the formation of oligosaccharides and polysaccharides. Glycosidic linkages can take various forms; the anomeric carbon of one sugar can be bonded to any one of the —OH groups on a second sugar to form an  $\alpha$ - or  $\beta$ -glycosidic linkage. Many different combinations are found in nature. The —OH groups are numbered so that they can be distinguished, and the numbering scheme follows that of the carbon atoms. The notation for the glycosidic linkage between the two sugars specifies which anomeric form of the sugar is involved in the bond and also specifies which carbon atoms of the two sugars are linked together. Two ways in which two  $\alpha$ -D-glucose molecules can be linked together are  $\alpha(1 \rightarrow 4)$  and  $\alpha(1 \rightarrow 6)$ . In the first example, the  $\alpha$ -anomeric carbon (C-1) of the first glucose molecule is joined in a glycosidic bond to the fourth carbon atom (C-4) of the second glucose molecule; the C-1 of the first glucose molecule is linked to the C-6 of the second glucose molecule in the second example (Figure 16.14). Another possibility of a glycosidic bond, this time between two  $\beta$ -D-glucose molecules, is a  $\beta, \beta(1 \rightarrow 1)$  linkage. The anomeric forms at both C-1 carbons must be specified because the linkage is between the two anomeric carbons, each of which is C-1 (Figure 16.15).

When oligosaccharides and polysaccharides form as a result of glycosidic bonding, their chemical natures depend on which monosaccharides are linked together and also on the particular glycosidic bond formed (i.e., which anomers and which carbon atoms are linked together). The difference between cellulose and starch depends on the glycosidic bond formed between glucose monomers. Because of the variation in glycosidic linkages, both linear and branched-chain polymers can be formed. If the internal monosaccharide residues that are incorporated in a polysaccharide form only two glycosidic bonds, the polymer will be linear. (Of course, the end residues will be involved in only one glycosidic linkage.) Some internal residues can form three glycosidic bonds, leading to the formation of branched-chain structures (Figure 16.16).

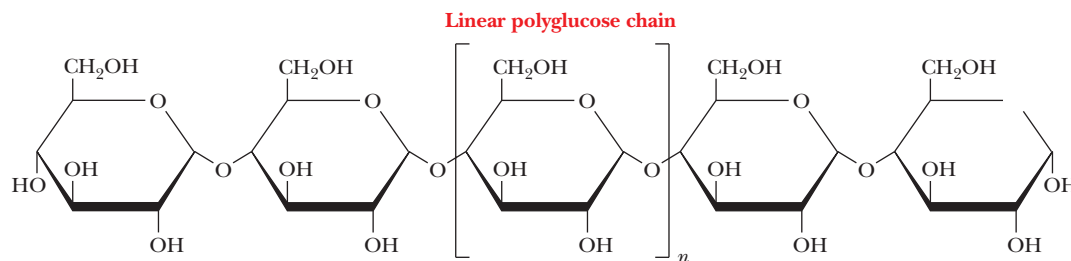
■ **FIGURE 16.14** Two different disaccharides of  $\alpha$ -D-glucose. These two chemical compounds have different properties because one has an  $\alpha(1 \rightarrow 4)$  linkage and the other has an  $\alpha(1 \rightarrow 6)$  linkage.



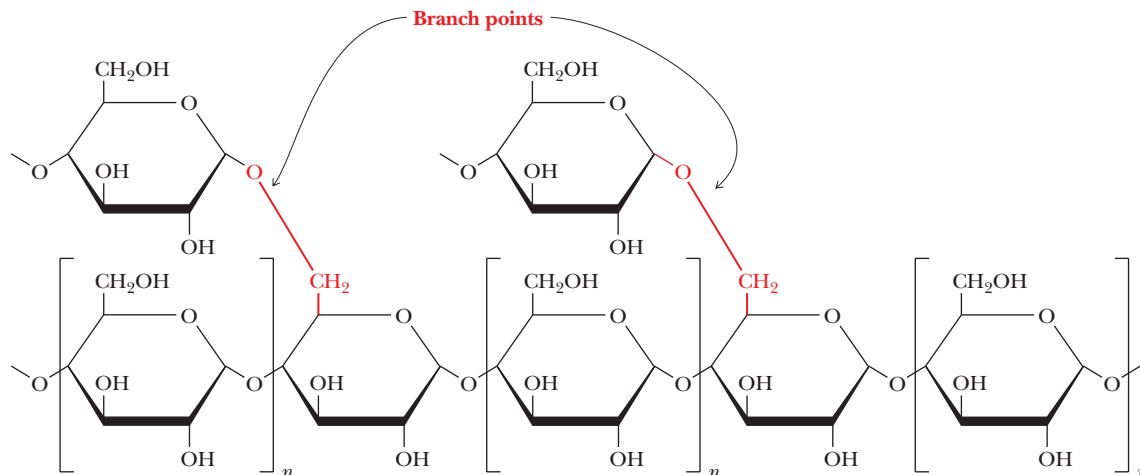
■ **FIGURE 16.15** A disaccharide of  $\beta$ -D-glucose. Both anomeric carbons (C-1) are involved in the glycosidic linkage.



- A** The linear polyglucose chain occurs in amylose. All glycosidic bonds are  $\alpha(1 \rightarrow 4)$ .



- B** The branched-chain polymer occurs in amylopectin and glycogen. Branched-polyglucose-chain glycosidic bonds are  $\alpha(1 \rightarrow 6)$  at branch points, but all glycosidic bonds along the chain are  $\alpha(1 \rightarrow 4)$ .



**FIGURE 16.16** Linear and branched-chain polymers of  $\alpha$ -D-glucose. The linear polyglucose chain.

Another point about glycosides is worth mentioning. We have already seen that the anomeric carbon is frequently involved in the glycosidic linkage, and also that the test for the presence of sugars—specifically for reducing sugars—requires a reaction of the group at the anomeric carbon. The internal anomeric carbons in oligosaccharides are not free to give the test for reducing sugars. Only if the end residue is a free hemiacetal rather than a glycoside will there be a positive test for a reducing sugar (Figure 16.17). The level of detection can be important for such a test. A sample that contains only a few molecules of a large polysaccharide, each molecule with a single reducing end, might well produce a negative test because there are not enough reducing ends to detect. The Biochemical Connections box on page 466 describes some interesting compounds that contain glycosidic bonds.

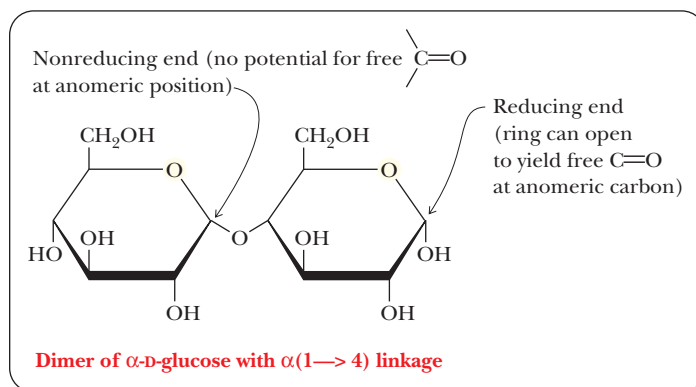
### What are some other important derivatives of sugars?

**Amino sugars** are an interesting class of compounds related to the monosaccharides. We shall not go into the chemistry of their formation, but it will be useful to have some acquaintance with them when we discuss polysaccharides. In sugars of this type, an amino group ( $-\text{NH}_2$ ) or one of its derivatives is substituted for the hydroxyl group of the parent sugar. In *N*-acetyl amino sugars, the amino group itself carries an acetyl group ( $\text{CH}_3-\text{CO}-$ ) as a substituent.

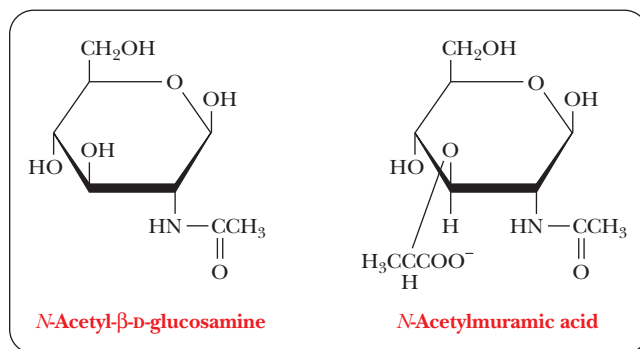
Two particularly important examples are *N*-acetyl- $\beta$ -D-glucosamine and its derivative *N*-acetyl- $\beta$ -muramic acid, which has an added carboxylic acid side chain (Figure 16.18). These two compounds are components of bacterial cell walls. We did not specify whether *N*-acetylmuramic acid belongs to the L or the D series of configurations, and we did not specify the  $\alpha$ - or  $\beta$ -anomer. This type

■ **FIGURE 16.17** Reducing sugars.

A disaccharide with a free hemiacetal end is a reducing sugar because of the presence of a free anomeric aldehyde carbonyl or potential aldehyde group.



■ **FIGURE 16.18** The structures of *N*-acetyl- $\beta$ -D-glucosamine and *N*-acetylmuramic acid.



of shorthand is the usual practice with  $\beta$ -D-glucose and its derivatives; the D configuration and the  $\beta$ -anomeric form are so common that we need not specify them all the time unless we want to make some specific point. The position of the amino group is also left unspecified because discussion of amino sugars usually centers on a few compounds whose structures are well known.

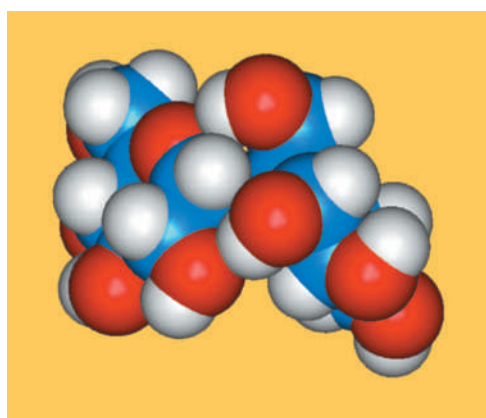
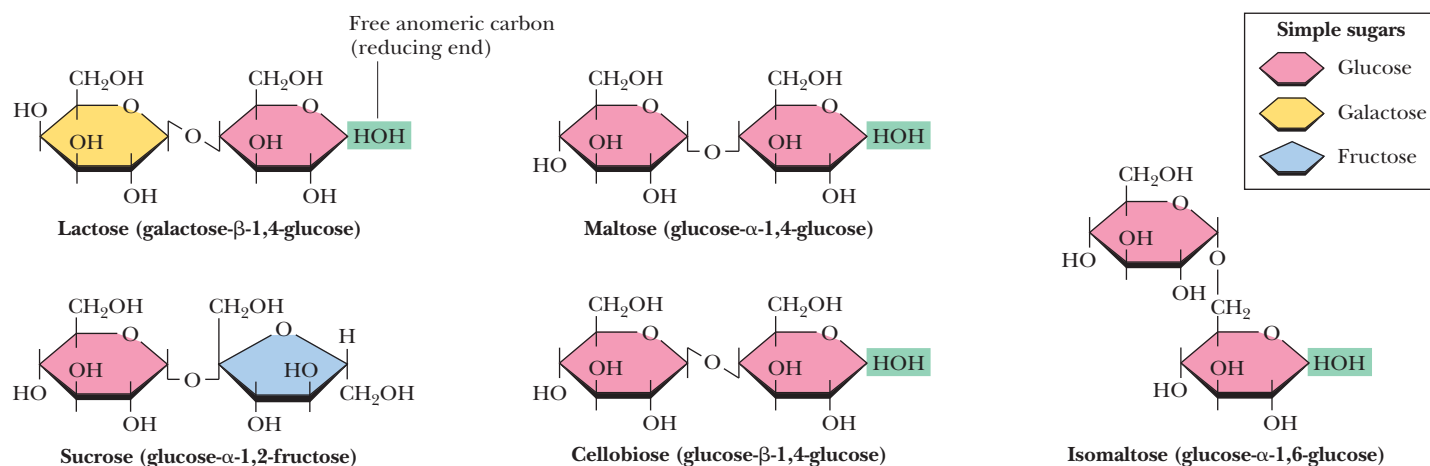
## 16.3 Some Important Oligosaccharides

Oligomers of sugars frequently occur as **disaccharides**, formed by linking two monosaccharide units by glycosidic bonds. Three of the most important examples of oligosaccharides are disaccharides. They are sucrose, lactose, and maltose (Figure 16.19). Two other disaccharides, isomaltose and cellobiose, are shown for comparison.

### What makes sucrose an important compound?

**Sucrose** is common table sugar, which is extracted from sugarcane and sugar beets. The monosaccharide units that make up sucrose are  $\alpha$ -D-glucose and  $\beta$ -D-fructose. Glucose (an aldohexose) is a pyranose, and fructose (a ketohexose) is a furanose. The  $\alpha$  C-1 carbon of the glucose is linked to the  $\beta$  C-2 carbon of the fructose (Figure 16.19) in a glycosidic linkage that has the notation  $\alpha,\beta(1 \rightarrow 2)$ . Sucrose is not a reducing sugar because both anomeric groups are involved in the glycosidic linkage. Free glucose is a reducing sugar, and free fructose can also give a positive test, even though it is a ketone rather than an aldehyde in the open-chain form. Fructose and ketoses in general can act as reducing sugars because they can isomerize to aldoses in a rather complex rearrangement reaction. (We need not concern ourselves with the details of this isomerization.)

When animals consume sucrose, it is hydrolyzed to glucose and fructose, which are then degraded by metabolic processes to provide energy. Humans

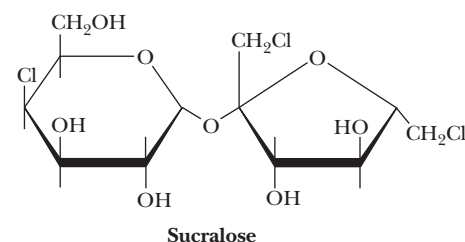


Sucrose

■ **FIGURE 16.19** The structures of several important disaccharides. Note that the notation —HOH means that the configuration can be either  $\alpha$  or  $\beta$ . When a D sugar is drawn in this orientation, if the —OH group is above the ring, the configuration is termed  $\beta$ . The configuration is termed  $\alpha$  if the —OH group is below the ring. Also note that sucrose has no free anomeric carbon atoms.

consume large quantities of sucrose, and excess consumption can contribute to health problems; this fact has led to a search for other sweetening agents. One that has been proposed is fructose itself. It is sweeter than sucrose; therefore, a smaller amount (by weight) of fructose than sucrose can produce the same sweetening effect with fewer calories. Consequently, high-fructose corn syrup is frequently used in food processing. The presence of fructose changes the texture of food, and the reaction to the change tends to depend on the preference of the consumer. Artificial sweeteners have been produced in the laboratory and have frequently been suspected of having harmful side effects; the ensuing controversies bear eloquent testimony to the human craving for sweets. Saccharin, for example, has been found to cause cancer in laboratory animals, as have cyclamates, but the applicability of these results to human carcinogenesis has been questioned by some. Aspartame (NutraSweet; Section 3.5) has been suspected of causing neurological problems, especially in individuals whose metabolisms cannot tolerate phenylalanine.

Another artificial sweetener is a derivative of sucrose. This substance, sucralose, which is marketed under the trade name Splenda, differs from sucrose in two ways (Figure 16.20). The first difference is that three of the hydroxyl groups have been replaced with three chlorine atoms. The second is that the configuration at carbon atom 4 of the six-membered pyranose ring of glucose has been inverted, producing a galactose derivative. The three hydroxyl groups that have been replaced by chlorine atoms are those bonded to carbon atoms 1 and 6 of the fructose moiety and to carbon atom 4 of the galactose moiety.



■ **FIGURE 16.20** The structure of sucralose. Note that sucralose (marketed under the trade name Splenda) differs from sucrose in the substitution of chlorine for three hydroxyls.

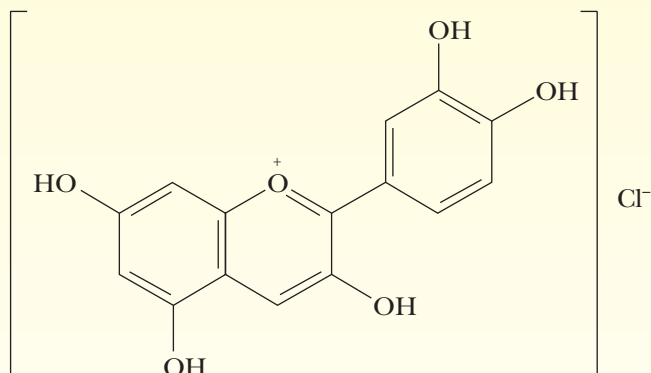


# Biochemical Connections PLANT SCIENCE

## Fruits, Flowers, Striking Colors, and Medicinal Uses Too

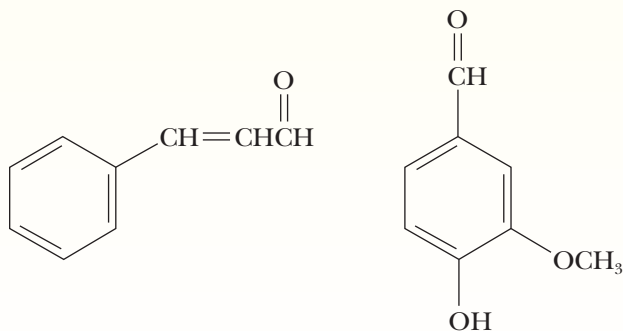
In sucrose, starches, and other sugar polymers, the *O*-glycoside bonds attach sugars to sugars. Other major categories of glycosides are known in which the sugar binds to some other type of molecule. Probably the most common example is the structure of nucleotides (Section 9.2), *N*-glycosides, in which the sugar binds to the nitrogenous, aromatic base, as found in ATP, many vitamins, DNA, and RNA. In glycolipids (Section 8.2) and glycoproteins (Section 16.5), carbohydrates are attached to both lipids and proteins, respectively, by glycoside linkages.

The red and blue colors of some flowers are sugar derivatives, often called anthocyanins. These pigments involve various sugars bonded to the compound cyanidin and its derivatives. These compounds are water soluble because of the polar groups they possess. You may have done an acid–base titration of the pigment from red cabbage or from blueberry juice in a chemistry lab. In contrast, orange, yellow, and green plant pigments tend to be lipid in composition and insoluble in water.



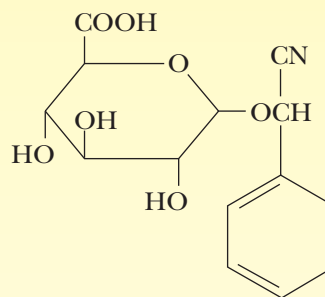
**Cyanidin chloride**

Many flavors involve sugar glycosides. Two familiar ones are cinnamon and vanilla, in which the sugars bond to cinnamaldehyde (3-phenyl-2-propenal) and vanillin, respectively. Both of these compounds are aromatic aldehydes. The distinctive taste of the kernel in a peach or apricot pit (a bitter-almond flavor) is due to laetrile, a controversial substance suggested as a cancer treatment by some.



**Cinnamaldehyde**

**Vanillin**



**Laetrile**

Many medically important substances have a glycosidic linkage as a part of their structure. Digitalis, prescribed for irregular heart-beat, is a mixture of several steroid complexes with sugars attached. Laetrile, a benzaldehyde derivative with a glycosidic linkage to glucuronic acid, was once thought to fight cancer, possibly because the cyanide moiety would poison the fast-growing cancer cells. This treatment is not approved in the United States, and it is likely that the cyanide causes more problems than it solves. The National Cancer Institute maintains a website at <http://www.cancer.gov>; use the search function there to find information about laetrile.



Photodisc Green/Getty Images

■ The foxglove plant produces the important cardiac medication digitalis.

## Biochemical Connections NUTRITION

### Lactose Intolerance: Why Do So Many People Not Want to Drink Milk?

Humans can be intolerant of milk and milk products for several reasons. Sugar intolerance results from the inability either to digest or to metabolize certain sugars. This problem differs from a food allergy, which involves an immune response (Section 14.5). A negative reaction to sugars in the diet usually involves intolerance, whereas proteins, including those found in milk, tend to cause allergies. Most sugar intolerance is due to missing or defective enzymes, so this is another example of inborn errors of metabolism.

Lactose is sometimes referred to as milk sugar because it occurs in milk. In some adults, a deficiency of the enzyme lactase in the intestinal villi causes a buildup of the disaccharide when milk products are ingested. This is because lactase is necessary to degrade lactose to galactose and glucose so that it can be absorbed into the bloodstream from the villi. Without the enzyme,

an accumulation of lactose in the intestine can be acted on by the lactase of intestinal bacteria (as opposed to the desirable lactase of the villi), producing hydrogen gas, carbon dioxide, and organic acids. The products of the bacterial lactase reaction lead to digestive problems, such as bloating and diarrhea, as does the presence of undegraded lactose. In addition, the by-products of the extra bacterial growth draw water into the intestine, thus aggravating the diarrhea. This disorder affects only about one-tenth of the Caucasian population of the United States, but it is more common among African-Americans, Asians, Native Americans, and Hispanics.

Even if the enzyme lactase is present so that lactose can be broken down by the body, other problems can occur. A different but related problem can occur in the further metabolism of galactose. If the enzyme that catalyzes a subsequent reaction in the pathway is missing and galactose builds up, a condition known as galactosemia can result. This is a severe problem in infants because the nonmetabolized galactose accumulates within cells and is converted to the hydroxy sugar galactitol, which cannot escape. Water is drawn into these cells and the swelling and edema causes damage. The critical tissue is the brain, which is not fully developed at birth. The swelling cells can crush the brain tissue, resulting in severe and irreversible retardation. The clinical test for this disorder is inexpensive and is required by law in all states.

The dietary therapy for these two problems is quite different. Lactose-intolerant individuals must avoid lactose throughout their lives. Fortunately, tablets like Lactaid are available to add to regular milk, as are lactose- and galactose-free formulas for feeding infants. True fermented food products such as yogurt and many cheeses (especially aged ones) have had their lactose degraded during fermentation. However, many foods are not processed in this way, so lactose-intolerant individuals need to exercise caution in their food choices.

There is no way to treat milk to make it safe for people who have galactosemia, so affected individuals must avoid milk during childhood. Fortunately, a galactose-free diet is easy to achieve simply by avoiding milk. After puberty, the development of other metabolic pathways for galactose alleviates the problem in most afflicted individuals. For people who want to avoid milk, there are plenty of milk substitutes, such as soy milk or rice milk. You can even get your latte or mocha made with soy milk at Starbucks nowadays.



© Cengage Learning/Charles D. Winters

■ **Dairy substitutes for the lactose intolerant.**  
These products help those with lactose intolerance meet their calcium needs.

Sucralose is not metabolized by the body, and, consequently, it does not provide calories. Tests conducted so far, as well as anecdotal evidence, indicate that it is a safe sugar substitute. It is likely to find wide use in the near future. We can safely predict that the search for nonfattening sweeteners will continue and that it will be accompanied by controversy.

### Are any other disaccharides important to us?

**Lactose** (see the Biochemical Connections box above) is a disaccharide made up of  $\beta$ -D-galactose and D-glucose. Galactose is the C-4 epimer of glucose. In other words, the only difference between glucose and galactose is inversion of configuration at C-4. The glycosidic linkage is  $\beta(1 \rightarrow 4)$ , between the anomeric

carbon C-1 of the  $\beta$  form of galactose and the C-4 carbon of glucose (Figure 16.19). Since the anomeric carbon of glucose is not involved in the glycosidic linkage, it can be in either the  $\alpha$  or the  $\beta$  form. The two anomeric forms of lactose can be specified, and the designation refers to the glucose residue; galactose must be present as the  $\beta$ -anomer, since the  $\beta$  form of galactose is required by the structure of lactose. Lactose is a reducing sugar because the group at the anomeric carbon of the glucose portion is not involved in a glycosidic linkage, so it is free to react with oxidizing agents.

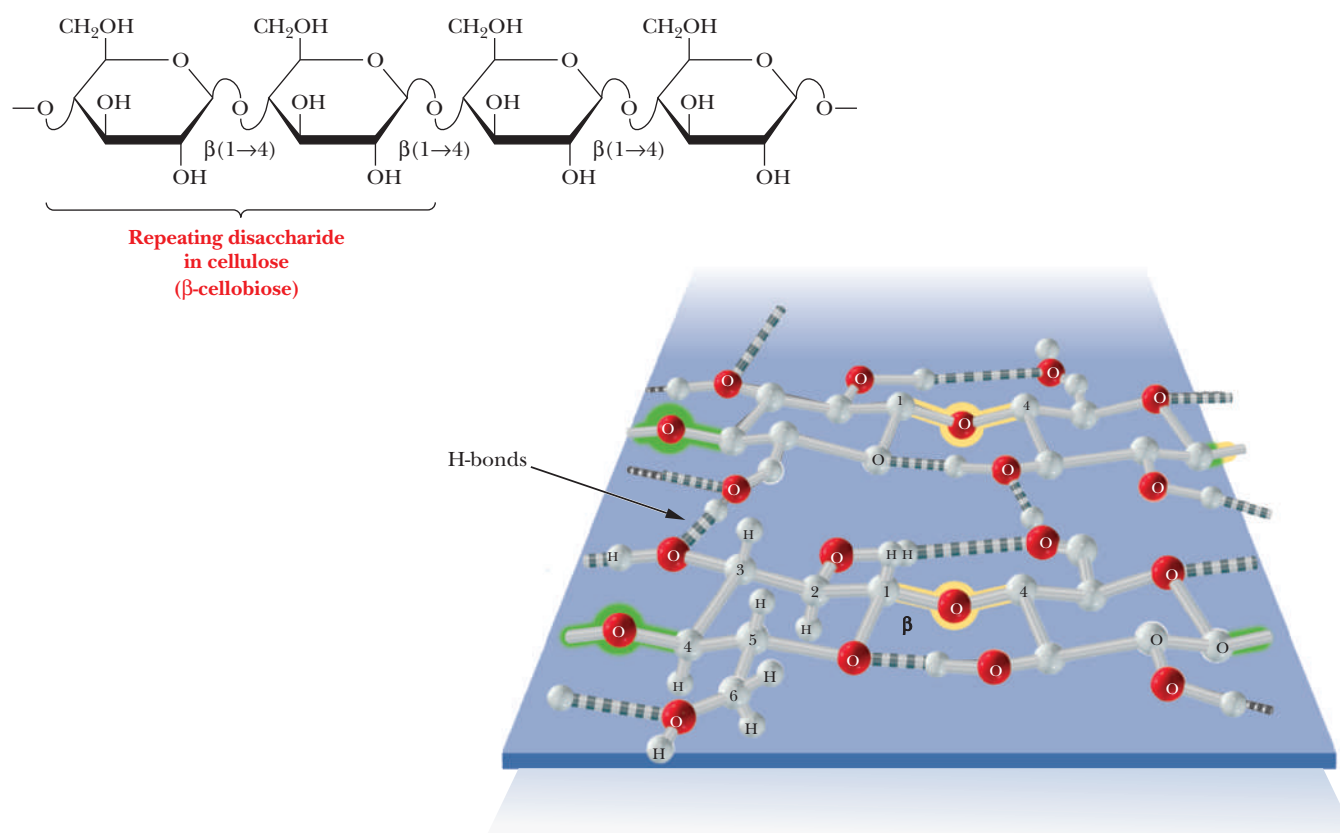
*Maltose* is a disaccharide obtained from the hydrolysis of starch. It consists of two residues of D-glucose in an  $\alpha(1 \rightarrow 4)$  linkage. Maltose differs from *cellobiose*, a disaccharide that is obtained from the hydrolysis of cellulose, only in the glycosidic linkage. In cellobiose, the two residues of D-glucose are bonded together in a  $\beta(1 \rightarrow 4)$  linkage (Figure 16.19). Mammals can digest maltose, but not cellobiose. Yeast, specifically brewer's yeast, contains enzymes that hydrolyze the starch in sprouted barley (barley malt) first to maltose and then to glucose, which is fermented in the brewing of beer. Maltose is also used in other beverages, such as malted milk.

## 16.4 Structures and Functions of Polysaccharides

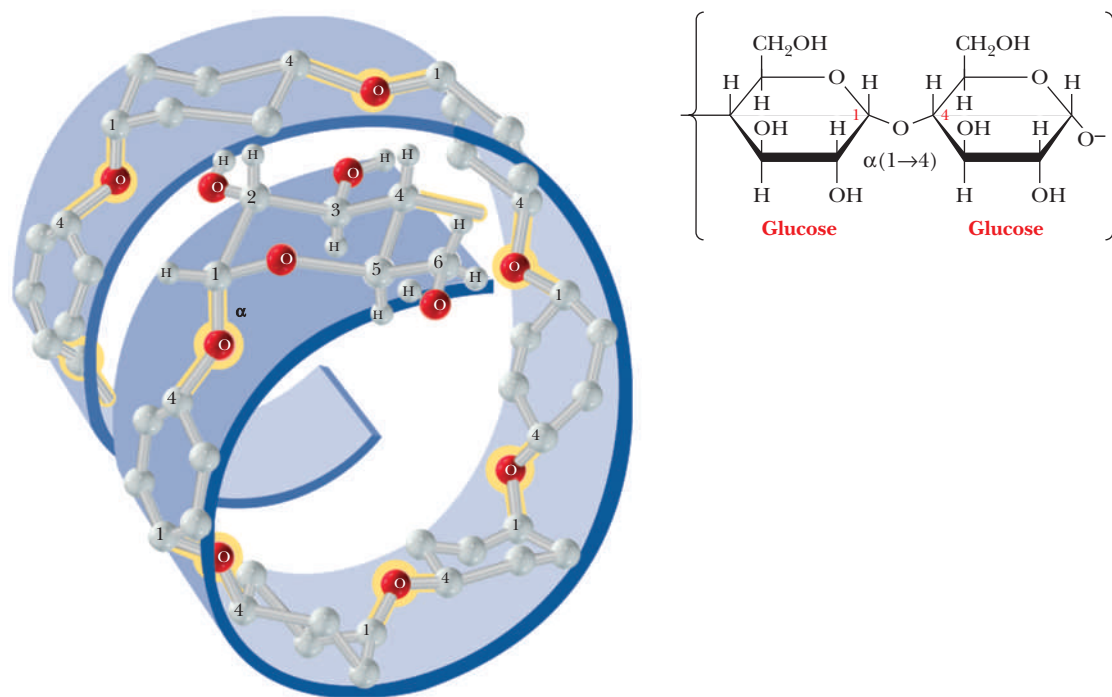
When many monosaccharides are linked together, the result is a **polysaccharide**. Polysaccharides that occur in organisms are usually composed of a very few types of monosaccharide components. A polymer that consists of only one type of monosaccharide is a *homopolysaccharide*; a polymer that consists of more than one type of monosaccharide is a *heteropolysaccharide*. Glucose is the most common monomer. When there is more than one type of monomer, frequently only two types of molecules occur in a repeating sequence. A complete characterization of a polysaccharide includes specification of which monomers are present and, if necessary, the sequence of monomers. It also requires that the type of glycosidic linkage be specified. We shall see the importance of the type of glycosidic linkage as we discuss different polysaccharides, since the nature of the linkage determines function. Cellulose and chitin are polysaccharides with  $\beta$ -glycosidic linkages, and both are structural materials. Starch and glycogen, also polysaccharides, have  $\alpha$ -glycosidic linkages, and they serve as carbohydrate-storage polymers in plants and animals, respectively.

### How do cellulose and starch differ from one another?

Cellulose is the major structural component of plants, especially of wood and plant fibers. It is a linear homopolysaccharide of  $\beta$ -D-glucose, and all residues are linked in  $\beta(1 \rightarrow 4)$  glycosidic bonds (Figure 16.21). Individual polysaccharide chains are hydrogen-bonded together, giving plant fibers their mechanical strength. Animals lack the enzymes, called *cellulases*, that hydrolyze cellulose to glucose. Such enzymes attack the  $\beta$ -linkages between glucoses, which is common to structural polymers; the  $\alpha$ -linkage between glucoses, which animals can digest, is characteristic of energy-storage polymers such as starch (Figure 16.22). Cellulases are found in certain bacteria, including the bacteria that inhabit the digestive tracts of insects, such as termites, and grazing animals, such as cattle and horses. The presence of these bacteria explains why cows and horses can live on grass and hay but humans cannot. The damage done by termites to the wooden parts of buildings arises from their ability to use cellulose in wood as a nutrient—owing to the presence of suitable bacteria in their digestive tracts.



■ **FIGURE 16.21** The polymeric structure of cellulose.  $\beta$ -Cellobiose is the repeating disaccharide. The monomer of cellulose is the  $\beta$ -anomer of glucose, which gives rise to long chains that can hydrogen-bond to one another.



■ **FIGURE 16.22** The structure of starch is based on the  $\alpha$ -anomer of glucose. The monomer of starch is the  $\alpha$ -anomer of glucose, which gives rise to a chain that folds into a helical form. The repeating dimer has  $\alpha(1 \rightarrow 4)$  linkages throughout.





Albert Grant/Photo Researchers, Inc.



William J. Weber/Visuals Unlimited

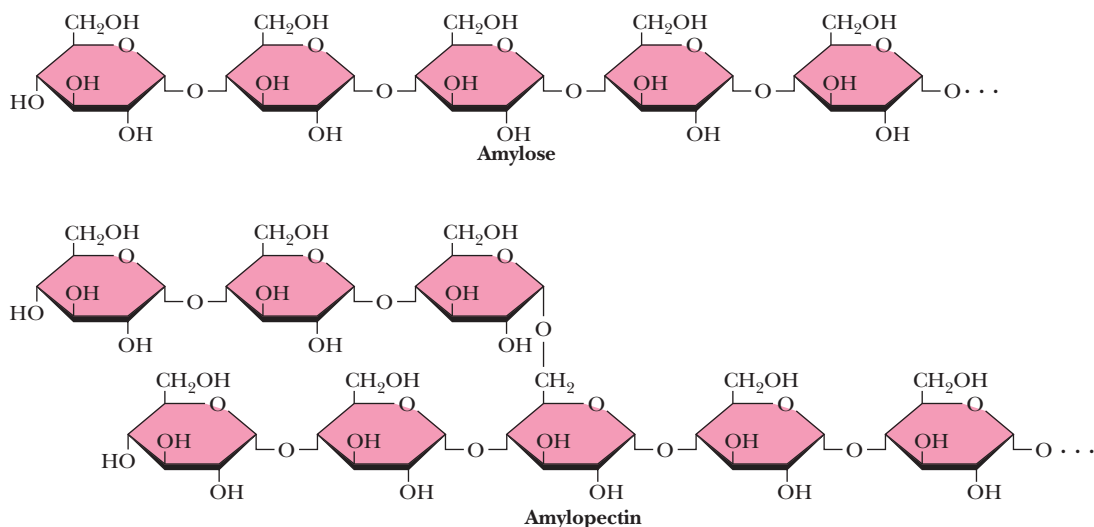
■ **Termites and cattle eating cellulose-containing foodstuffs.** Termites can digest the cellulose in wood, and cattle can digest the cellulose in grass, because bacteria in their digestive tracts produce the enzyme cellulase, which hydrolyzes the  $\beta$ -glycosidic linkage in cellulose.

### Is there more than one form of starch?

The importance of carbohydrates as energy sources suggests that there is some use for polysaccharides in metabolism. We shall now discuss in more detail some polysaccharides, such as starches, that serve as vehicles for storage of glucose.

**Starches** are polymers of  $\alpha$ -D-glucose that occur in plant cells, usually as starch granules in the cytosol. Note that there is an  $\alpha$ -linkage in starch, in contrast with the  $\beta$ -linkage in cellulose. The types of starches can be distinguished from one another by their degrees of chain branching. Amylose is a linear polymer of glucose, with all the residues linked together by  $\alpha(1 \rightarrow 4)$  bonds. Amylopectin is a branched chain polymer, with the branches starting at  $\alpha(1 \rightarrow 6)$  linkages along the chain of  $\alpha(1 \rightarrow 4)$  linkages (Figure 16.23). The most usual conformation of amylose is a helix with six residues per turn. Iodine molecules can fit inside the helix to form a starch-iodine complex, which has a characteristic dark-blue color (Figure 16.24). The formation of this complex is a well-known test for the presence of starch. If there is a preferred conformation for amylopectin, it is not yet known. (It is known that the color of the product obtained when amylopectin and glycogen react with iodine is red-brown, not blue.)

Because starches are storage molecules, there must be a mechanism for releasing glucose from starch when the organism needs energy. Both plants and animals contain enzymes that hydrolyze starches. Two of these enzymes, known as  $\alpha$ - and  $\beta$ -amylase (the  $\alpha$  and  $\beta$  do not signify anomeric forms in this case), attack  $\alpha(1 \rightarrow 4)$  linkages.  $\beta$ -amylase is an *exoglycosidase* that cleaves from the nonreducing end of the polymer. Maltose, a dimer of glucose, is the product of reaction. The other enzyme,  $\alpha$ -amylase, is an *endoglycosidase*, which can hydrolyze a glycosidic linkage anywhere along the chain to produce glucose and maltose. Amylose can be completely degraded to glucose and maltose by the two amylases, but amylopectin is not completely degraded because the branching linkages are not attacked. However, *debranching enzymes* occur in both plants and animals; they degrade the  $\alpha(1 \rightarrow 6)$  linkages. When these enzymes are combined with the amylases, they contribute to the complete degradation of both forms of starch.



■ **FIGURE 16.23** Amylose and amylopectin are the two forms of starch. Note that the linear linkages are  $\alpha(1 \rightarrow 4)$ , but the branches in amylopectin are  $\alpha(1 \rightarrow 6)$ . Branches in polysaccharides can involve any of the hydroxyl groups on the monosaccharide components. Amylopectin is a highly branched structure, with branches occurring at every 12 to 30 residues.

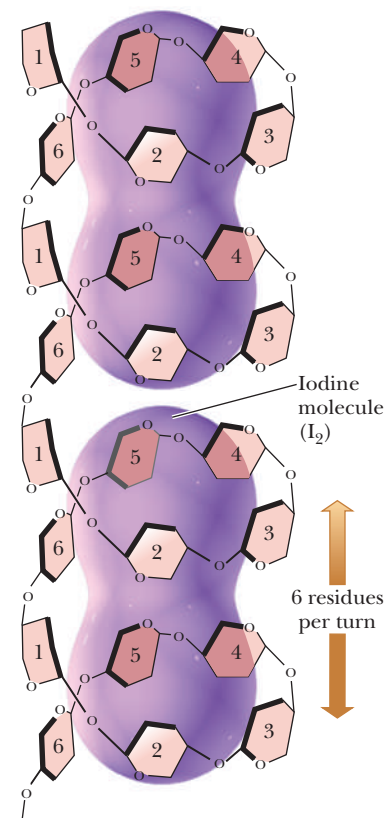


### How is glycogen related to starch?

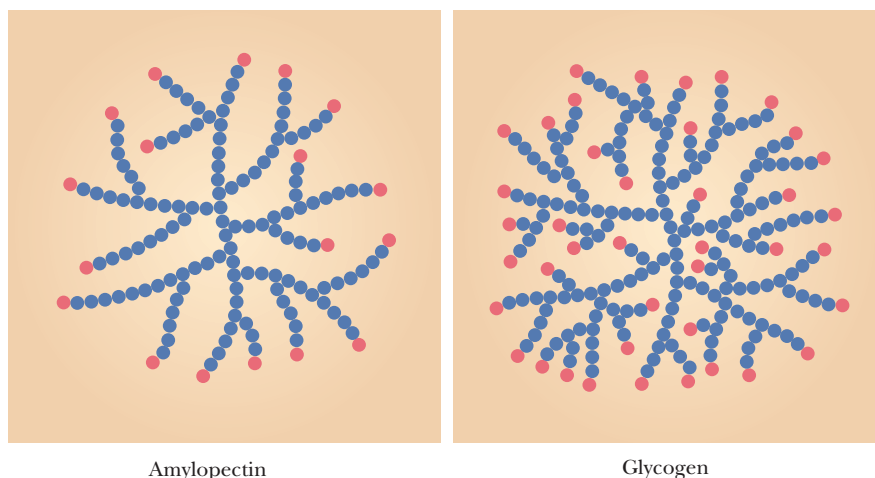
Although starches occur only in plants, there is a similar carbohydrate storage polymer in animals. **Glycogen** is a branched-chain polymer of  $\alpha$ -D-glucose, and in this respect it is similar to the amylopectin fraction of starch. Like amylopectin, glycogen consists of a chain of  $\alpha(1 \rightarrow 4)$  linkages with  $\alpha(1 \rightarrow 6)$  linkages at the branch points. The main difference between glycogen and amylopectin is that glycogen is more highly branched (Figure 16.25). Branch points occur about every 10 residues in glycogen and about every 25 residues in amylopectin. In glycogen, the average chain length is 13 glucose residues, and there are 12 layers of branching. At the heart of every glycogen molecule is a protein called glycogenin, which is discussed in Section 18.1. Glycogen is found in animal cells in granules similar to the starch granules in plant cells. Glycogen granules are observed in well-fed liver and muscle cells, but they are not seen in some other cell types, such as brain and heart cells under normal conditions. Some athletes, particularly long-distance runners, try to build up their glycogen reserves before a race by eating large amounts of carbohydrates. When the organism needs energy, various degradative enzymes remove glucose units (Section 18.1). Glycogen phosphorylase is one such enzyme; it cleaves one glucose at a time from the nonreducing end of a branch to produce glucose-1-phosphate, which then enters the metabolic pathways of carbohydrate breakdown. Debranching enzymes also play a role in the complete breakdown of glycogen. The number of branch points is significant for two reasons. First, a more branched polysaccharide is more water soluble. This may not be as important for a plant, but, for a mammal, the amount of glycogen in solution is. There are glycogen-storage diseases caused by lower-than-normal levels of branching enzymes. The glycogen products resemble starch and can fall out of solution, forming glycogen crystals in the muscles and liver. Second, when an organism needs energy quickly, the glycogen phosphorylase has more potential targets if there are more branches, allowing a quicker mobilization of glucose. Again, this is not as important to a plant, so there was no evolutionary pressure to make starch highly branched.

### What is chitin?

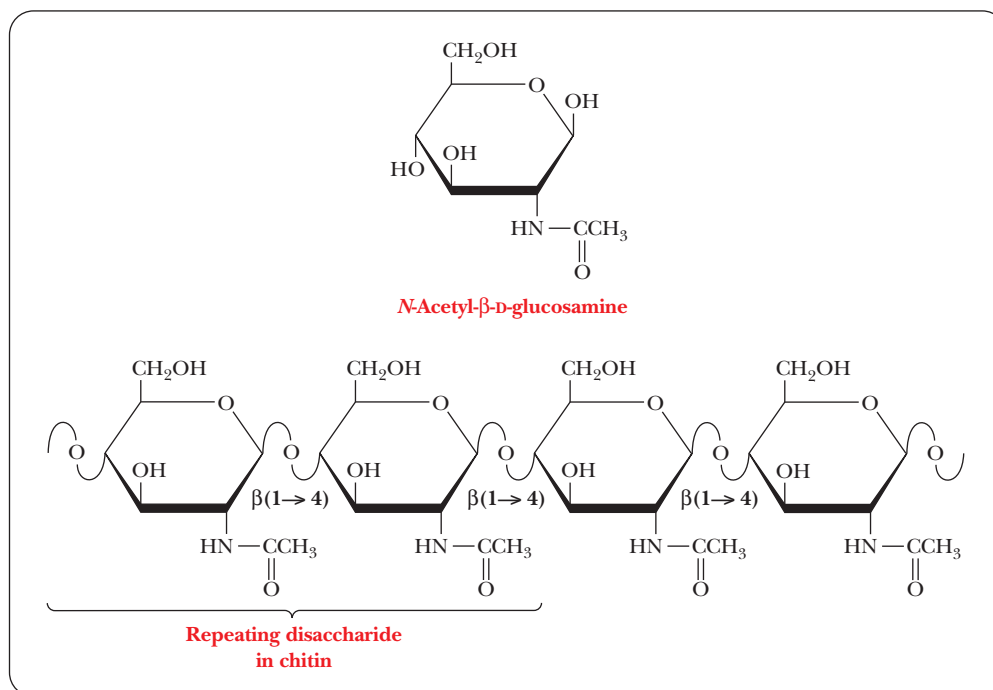
A polysaccharide that is similar to cellulose in both structure and function is chitin, which is also a linear homopolysaccharide with all the residues linked in  $\beta(1 \rightarrow 4)$  glycosidic bonds. Chitin differs from cellulose in the nature of the monosaccharide unit; in cellulose, the monomer is  $\beta$ -D-glucose; in chitin, the monomer is *N*-acetyl- $\beta$ -D-glucosamine. The latter compound differs from glucose only in the substitution of the *N*-acetyl amino group ( $-\text{NH}-\text{CO}-\text{CH}_3$ ) for the hydroxyl group ( $-\text{OH}$ ) on carbon C-2 (Figure 16.26). Like cellulose,



■ **FIGURE 16.24** The starch-iodine complex. Amylose occurs as a helix with six residues per turn. In the starch-iodine complex, the iodine molecules are parallel to the long axis of the helix. Four turns of the helix are shown here. Six turns of the helix, containing 36 glycosyl residues, are required to produce the characteristic blue color of the complex.



■ **FIGURE 16.25** A comparison of the degrees of branching in amylopectin and glycogen.



■ **FIGURE 16.26** The polymeric structure of **chitin**. *N*-Acetylglucosamine is the monomer, and a dimer of *N*-acetylglucosamine is the repeating disaccharide.

chitin plays a structural role and has a fair amount of mechanical strength because the individual strands are held together by hydrogen bonds. It is a major structural component of the exoskeletons of invertebrates such as insects and crustaceans (a group that includes lobsters and shrimp), and it also occurs in cell walls of algae, fungi, and yeasts.

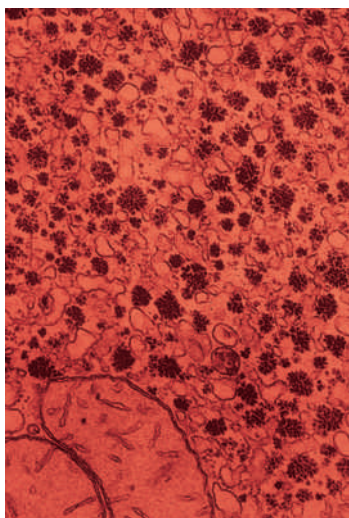
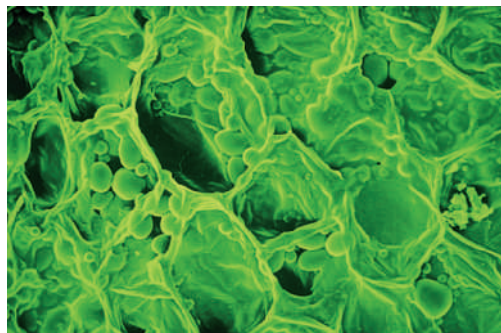
### What role do polysaccharides play in the structure of cell walls?

In organisms that have cell walls, such as bacteria and plants, the walls consist largely of polysaccharides. The cell walls of bacteria and plants have biochemical differences, however.

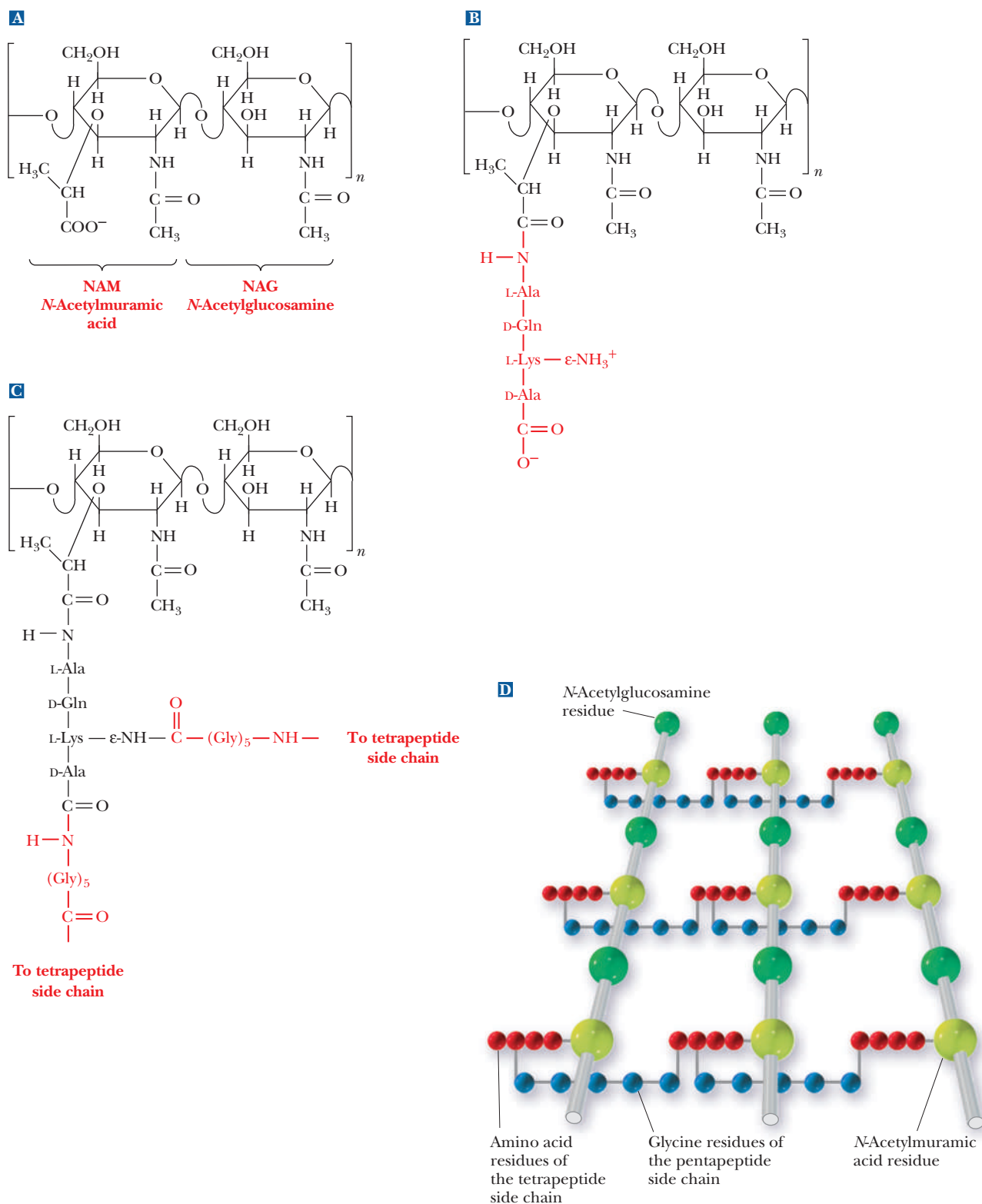
Heteropolysaccharides are major components of *bacterial* cell walls. A distinguishing feature of prokaryotic cell walls is that the polysaccharides are cross-linked by peptides. The repeating unit of the polysaccharide consists of two residues held together by  $\beta$ (1 $\rightarrow$ 4) glycosidic links, as was the case in cellulose and chitin. One of the two monomers is *N*-acetyl-D-glucosamine, which occurs in chitin, and the other monomer is *N*-acetylmuramic acid (Figure 16.27a). The structure of *N*-acetylmuramic acid differs from that of *N*-acetylglucosamine by the substitution of a lactic acid side chain [ $-\text{O}-\text{CH}(\text{CH}_3)-\text{COOH}$ ] for the hydroxyl group ( $-\text{OH}$ ) on carbon 3. *N*-Acetylmuramic acid is found only in prokaryotic cell walls; it does not occur in eukaryotic cell walls.

The cross-links in bacterial cell walls consist of small peptides. We shall use one of the best-known examples as an illustration. In the cell wall of the bacterium *Staphylococcus aureus*, an oligomer of four amino acids (a tetramer) is bonded to *N*-acetylmuramic acid, forming a side chain (Figure 16.27b). The tetrapeptides are themselves cross-linked by another small peptide, in this case consisting of five amino acids.

The carboxyl group of the lactic acid side chain of *N*-acetylmuramic acid forms an amide bond with the N-terminal end of a tetrapeptide that has the sequence L-Ala-D-Gln-L-Lys-D-Ala. Recall that bacterial cell walls are one of the few places where D-amino acids occur in nature. The occurrence of D-amino acids and *N*-acetylmuramic acid in bacterial cell walls but not in plant cell walls shows a biochemical as well as structural difference between prokaryotes and eukaryotes.



■ Electron micrographs of starch granules in a plant and glycogen granules in an animal.



■ **FIGURE 16.27** The structure of the peptidoglycan of the bacterial cell wall of *Staphylococcus aureus*. (a) The repeating disaccharide. (b) The repeating disaccharide with the tetrapeptide side chain (shown in red). (c) Adding the pentaglycine cross-links (shown in red). (d) Schematic diagram of the peptidoglycan. The sugars are the larger spheres. The red spheres are the amino acid residues of the tetrapeptide, and the blue spheres are the glycine residues of the pentapeptide.

## Biochemical Connections ALLIED HEALTH

### Why Is Dietary Fiber So Good for You?

Fiber in the diet is colloquially called roughage. It is principally made of complex carbohydrates, may have some protein components, and is moderately to fully insoluble. The health benefits of fiber are just beginning to be fully realized. We have known for a long time that roughage stimulates peristaltic action and thus helps move the digested food through the intestines, decreasing the transit time through the gut.

Potentially toxic substances in food and in bile fluid bind to fiber and are exported from the body, thus preventing them

from damaging the lower intestine or being reabsorbed there. Statistical evidence indicates that high fiber also reduces colon and other cancers, precisely because fiber binds suspected carcinogens. It is also plausible that the benefit is due to a lack of other items in the high-fiber diet. People on high-fiber diets also tend to take in less fat and fewer calories. Any difference in heart disease or cancer may be due to these other differences.

There has been much publicity about fiber in the diet reducing cholesterol. Fiber does bind cholesterol, and it certainly causes some decrease in the amount in the blood. The reduction, expressed as a percentage, is higher in cases in which the original level of cholesterol is higher. There is, however, no definitive evidence that lowering cholesterol via the ingestion of fiber results in less heart disease.

Fiber comes in two forms: soluble and insoluble. The most common insoluble fiber is cellulose, which is found in lettuce, carrots, bean sprouts, celery, brown rice, most other vegetables, many fruit skins, and pumpernickel bread. Insoluble fiber binds various molecules but otherwise merely forms bulk in the lower intestine. Soluble fibers include amylopectin and other pectins, as well as complex starches. Uncooked and mildly processed foods contain a higher proportion of this type of fiber.

Because of increased surface area, these fibers seem to be more beneficial. Good sources include bran (especially oat bran), barley, and fresh fruits (with skin), brussels sprouts, potatoes with skin, beans, and zucchini. Soluble fiber binds water very well, increasing satiety by helping to fill the stomach.

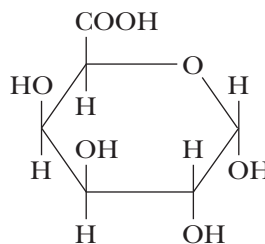


© Foodcollection.com/Alamy

■ Many breakfast cereals advertise their high-fiber content. The presence of fruit adds to the fiber content of such a breakfast.

The tetrapeptide forms two cross-links, both of them to a pentapeptide that consists of five glycine residues, (Gly)<sub>5</sub>. The glycine pentamers form peptide bonds to the C-terminal end and to the side-chain  $\epsilon$ -amino group of the lysine in the tetrapeptide [Figure 16.27(c)]. This extensive cross-linking produces a three-dimensional network of considerable mechanical strength, which is why bacterial cell walls are extremely difficult to disrupt. The material that results from the cross-linking of polysaccharides by peptides is a **peptidoglycan** [Figure 16.25(d)], so named because it has both peptide and carbohydrate components.

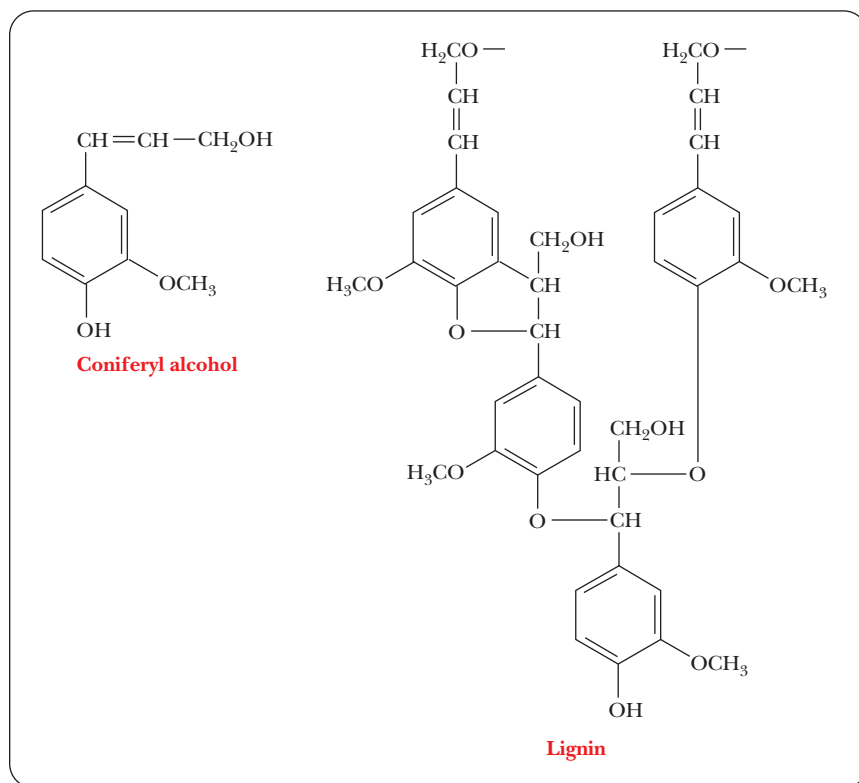
*Plant cell walls* consist largely of **cellulose**. The other important polysaccharide component found in plant cell walls is **pectin**, a polymer made up mostly of D-galacturonic acid, a derivative of galactose in which the hydroxyl group on carbon C-6 has been oxidized to a carboxyl group.



D-Galacturonic acid

Pectin is extracted from plants because it has commercial importance in the food-processing industry as a gelling agent in yogurt, fruit preserves, jams, and jellies. The major nonpolysaccharide component in plant cell walls, especially in woody plants, is **lignin** (Latin *lignum*, “wood”). Lignin is a polymer of coniferyl





■ **FIGURE 16.28** The structure of lignin, a polymer of coniferyl alcohol.

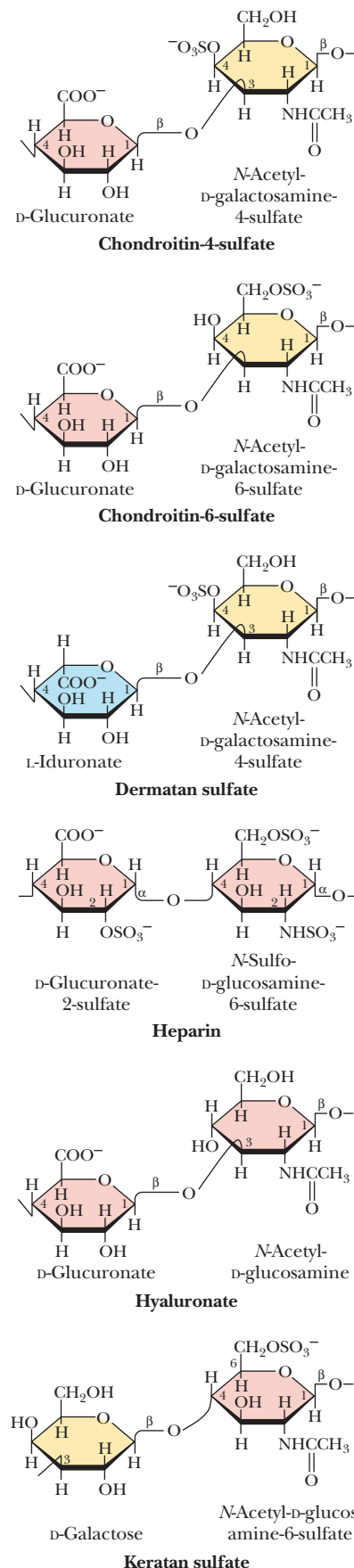
alcohol, and it is a very tough and durable material (Figure 16.28). Unlike bacterial cell walls, plant cell walls contain comparatively little peptide or protein.

### Do polysaccharides play any specific roles in connective tissue?

**Glycosaminoglycans** are a type of polysaccharide based on a repeating disaccharide in which one of the sugars is an amino sugar and at least one of them has a negative charge owing to the presence of a sulfate group or a carboxyl group. These polysaccharides are involved in a wide variety of cellular functions and tissues. Figure 16.29 shows the disaccharide structure of the most common ones. Heparin is a natural anticoagulant that helps prevent blood clots. Hyaluronic acid is a component of the vitreous humor of the eye and of the lubricating fluid of joints. The chondroitin sulfates and keratan sulfate are components of connective tissue. Glucosamine sulfate and chondroitin sulfate are sold in large quantities as over-the-counter drugs used to help repair frayed or otherwise damaged cartilage, especially in knees. Many people who are advised that they need knee surgery for damaged ligaments look for improvement first with a two- or three-month regimen of these glycosaminoglycans. Questions exist about the efficacy of this treatment, so it will be interesting to see what future it may have.

## 16.5 Glycoproteins

Glycoproteins contain carbohydrate residues in addition to the polypeptide chain (Chapter 4). Some of the most important examples of glycoproteins are involved in the immune response; for example, **antibodies**, which bind to and immobilize antigens (the substances attacking the organism), are glycoproteins. Carbohydrates also play an important role as **antigenic determinants**, the portions of an antigenic molecule that antibodies recognize and to which they bind.



■ **FIGURE 16.29** Glycosaminoglycans, which are formed from repeating disaccharide units, often occur as components of the proteoglycans.

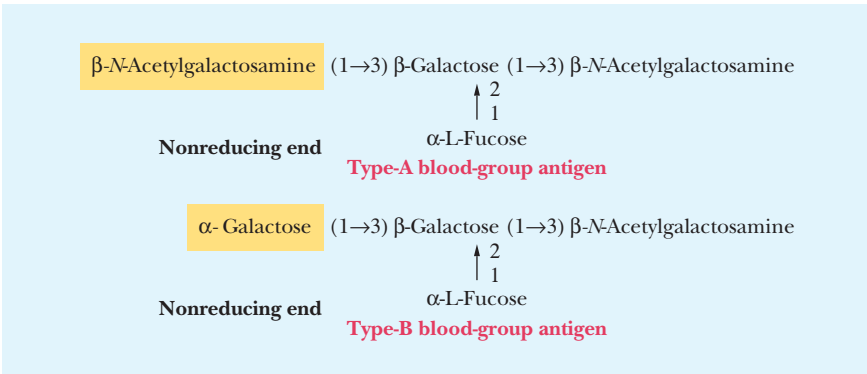


How are carbohydrates important in the immune response?

An example of the role of the oligosaccharide portion of glycoproteins as antigenic determinants is found in human blood groups. There are four human blood groups: A, B, AB, and O (see the Biochemical Connections box below). The distinctions between the groups depend on the oligosaccharide portions of the glycoproteins on the surfaces of the blood cells called erythrocytes. In all blood types, the oligosaccharide portion of the molecule contains the sugar L-fucose, mentioned earlier in this chapter as an example of a deoxy sugar. *N*-Acetylgalactosamine is found at the nonreducing end of the oligosaccharide in the type-A blood-group antigen. In type-B blood,  $\alpha$ -D-galactose takes the place of *N*-acetylgalactosamine. In type-O blood, neither of these terminal residues is present, and, in type-AB blood, both kinds of oligosaccharide are present (Figure 16.30).

Glycoproteins also play an important role in eukaryotic cell membranes. The sugar portions are added to the protein as it passes through the Golgi on its way to the cell surface. Those glycoproteins with an extremely high carbohydrate content (85%–95% by weight) are classified as **proteoglycans**. (Note the similarity of this term to the word *peptidoglycan*, which we met in Section 16.4.) Proteoglycans are constantly being synthesized and broken down. If there is a lack of the lysosomal enzymes that degrade them, proteoglycans accumulate, with tragic consequences. One of the most striking consequences is the genetic disease known as Hurler’s syndrome, in which the material that accumulates includes large amounts of amino sugars (Section 16.2). This disease leads to skeletal deformities, severe mental retardation, and death in early childhood.

■ FIGURE 16.30 The structures of the blood-group antigenic determinants.



Biochemical Connections ALLIED HEALTH

Glycoproteins and Blood Transfusions

If a blood transfusion is attempted with incompatible blood types, as when blood from a type-A donor is given to a type-B recipient, an antigen–antibody reaction takes place because the type-B recipient has antibodies to the type-A blood. The characteristic oligosaccharide residues of type-A blood cells serve as the antigen. A cross-linking reaction occurs between antigens and antibodies, and the blood cells clump together. In the case of a transfusion of type-B blood to a type-A recipient, antibodies to type-B blood produce the same result. Type-O blood has neither antigenic determinant, and so people with type-O blood are considered universal donors. However, these people have antibodies to both type-A and type-B blood, and so they are not universal acceptors. Type-AB people have both antigenic determinants. As a result,

they do not produce either type of antibody; they are universal acceptors.

Transfusion Relationships			
Blood Type	Makes Antibodies Against	Can Receive From	Can Donate To
O	A, B	O	O, A, B, AB
A	B	O, A	A, AB
B	A	O, B	B, AB
AB	None	O, A, B, AB	AB

## SUMMARY

**What is unique about the structures of sugars?** The simplest examples of carbohydrates are monosaccharides, compounds that each contain a single carbonyl group and two or more hydroxyl groups. Monosaccharides frequently encountered in biochemistry are sugars that contain from three to seven carbon atoms. Sugars contain one or more chiral centers; the configurations of the possible stereoisomers can be represented by Fischer projection formulas.

**What happens if a sugar forms a cyclic molecule?** Sugars exist predominantly as cyclic molecules rather than in an open-chain form. Haworth projection formulas are more realistic representations of the cyclic forms of sugars than are Fischer projection formulas. Many stereoisomers are possible for five- and six-carbon sugars, but only a few of the possibilities are encountered frequently in nature.

**What are some oxidation–reduction reactions of sugars?** Monosaccharides can undergo various reactions. Oxidation reactions make up one important group.

**What are some important esterification reactions of sugars?** Esterification of sugars to phosphoric acid plays an important role in metabolism.

**What are glycosides, and how do they form?** The most important reaction of sugars by far is the formation of glycosidic linkages, which give rise to oligosaccharides and polysaccharides.

**What are some other important derivatives of sugars?** Amino sugars are the basis of cell wall structures.

**What makes sucrose an important compound?** Three important examples of oligosaccharides are the disaccharides sucrose, lactose, and maltose. Sucrose is common table sugar. It is a disaccharide formed when a glycosidic bond forms between glucose and fructose.

**Are any other disaccharides important to us?** Lactose occurs in milk, and maltose is obtained via the hydrolysis of starch.

**How do cellulose and starch differ from one another?** In polysaccharides, the repeating unit of the polymer is frequently limited to one or two kinds of monomer. Cellulose and starch differ in the anomeric form of their glycosidic bonds: the  $\alpha$  form in starch and the  $\beta$  form in cellulose.

**Is there more than one form of starch?** Starch exists in two polymeric forms, the linear amylose and the branched amylopectin.

**How is glycogen related to starch?** Starch, found in plants, and glycogen, which occurs in animals, differ from each other in the degree of branching in the polymer structure.


**What is chitin?** Cellulose and chitin are polymers based on single kinds of monomer units—glucose and *N*-acetylglucosamine, respectively. Both polymers play structural roles in organisms.

**What role do polysaccharides play in the structure of cell walls?** In bacterial cell walls, polysaccharides are cross-linked to peptides. Plant cell walls consist primarily of glucose.

**Do polysaccharides play any specific roles in connective tissue?** Glycosaminoglycans are a type of polysaccharide based on a repeating disaccharide in which one of the sugars is an amino sugar and at least one of them has a negative charge owing to the presence of a sulfate group or a carboxyl group. They play a role in joint lubrication and also in the blood clotting process.

**How are carbohydrates important in the immune response?** In glycoproteins, carbohydrate residues are covalently linked to the polypeptide chain. Such glycoproteins can play a role in the recognition sites of antigens. A common example is the ABO blood group, in which the three major blood types are distinguished by sugar molecules attached to the protein.

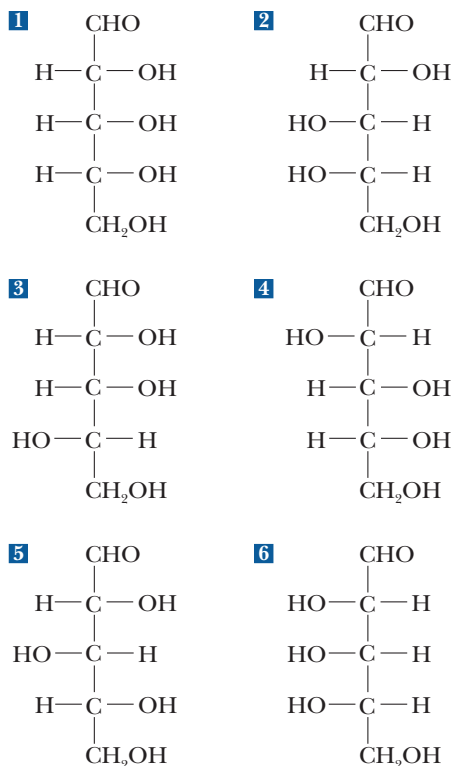
## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

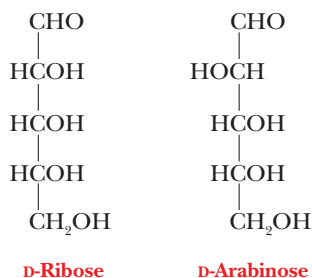
### 16.1 Sugars: Their Structures and Stereochemistry

- Recall** Define the following terms: polysaccharide, furanose, pyranose, aldose, ketose, glycosidic bond, oligosaccharide, glycoprotein.
- Recall** Name which, if any, of the following are epimers of D-glucose: D-mannose, D-galactose, D-ribose.
- Recall** Name which, if any, of the following groups are *not* aldose–ketose pairs: D-ribose and D-ribulose, D-glucose and D-fructose, D-glyceraldehyde and dihydroxyacetone.
- Recall** What is the difference between an enantiomer and a diastereomer?
- Recall** How many possible epimers of D-glucose exist?
- Recall** Why are furanoses and pyranoses the most common cyclic forms of sugars?
- Recall** How many chiral centers are there in the open-chain form of glucose? In the cyclic form?
- Reflect and Apply** Following are Fischer projections for a group of five-carbon sugars, all of which are aldopentoses. Identify the pairs

that are enantiomers and the pairs that are epimers. (The sugars shown here are not all of the possible five-carbon sugars.)



9. **Reflect and Apply** The sugar alcohol often used in “sugarless” gums and candies is L-sorbitol. Much of this alcohol is prepared by reduction of D-glucose. Compare these two structures and explain how this can be.
10. **Reflect and Apply** Consider the structures of arabinose and ribose. Explain why nucleotide derivatives of arabinose, such as ara-C and ara-A, are effective metabolic poisons.



11. **Reflect and Apply** Two sugars are epimers of each other. Is it possible to convert one to the other without breaking covalent bonds?
12. **Reflect and Apply** How does the cyclization of sugars introduce a new chiral center?

## 16.2 Reactions of Monosaccharides

13. **Recall** What is unusual about the structure of N-acetylmuramic acid (Figure 16.18) compared with the structures of other carbohydrates?
14. **Recall** What is the chemical difference between a sugar phosphate and a sugar involved in a glycosidic bond?
15. **Recall** Define the term *reducing sugar*.
16. **Biochemical Connections** What are the structural differences between vitamin C and sugars? Do these structural differences play a role in the susceptibility of this vitamin to air oxidation?

## 16.3 Some Important Oligosaccharides

17. **Recall** Name two differences between sucrose and lactose. Name two similarities.
18. **Reflect and Apply** Draw a Haworth projection for the disaccharide gentibiose, given the following information:
- It is a dimer of glucose.
  - The glycosidic linkage is  $\beta(1 \rightarrow 6)$ .
  - The anomeric carbon not involved in the glycosidic linkage is in the  $\alpha$  configuration.
19. **Biochemical Connections** What is the metabolic basis for the observation that many adults cannot ingest large quantities of milk without developing gastric difficulties?
20. **Reflect and Apply** Draw Haworth projection formulas for dimers of glucose with the following types of glycosidic linkages:
- A  $\beta(1 \rightarrow 4)$  linkage (both molecules of glucose in the  $\beta$  form)
  - An  $\alpha, \alpha(1 \rightarrow 1)$  linkage
  - A  $\beta(1 \rightarrow 6)$  linkage (both molecules of glucose in the  $\beta$  form)
21. **Biochemical Connections** A friend asks you why some parents at her child’s school want a choice of beverages served at lunch, rather than milk alone. What do you tell your friend?

## 16.4 Structures and Functions of Polysaccharides

22. **Recall** What are some of the main differences between the cell walls of plants and those of bacteria?
23. **Recall** How does chitin differ from cellulose in structure and function?
24. **Recall** How does glycogen differ from starch in structure and function?
25. **Recall** What is the main structural difference between cellulose and starch?
26. **Recall** What is the main structural difference between glycogen and starch?
27. **Recall** How do the cell walls of bacteria differ from those of plants?
28. **Reflect and Apply** Pectin, which occurs in plant cell walls, exists in nature as a polymer of D-galacturonic acid methylated at carbon 6 of the monomer. Draw a Haworth projection for a repeating disaccharide unit of pectin with one methylated and one unmethylated monomer unit in  $\alpha(1 \rightarrow 4)$  linkage.
29. **Reflect and Apply** Advertisements for a food supplement to be taken by athletes claimed that the energy bars contained the two best precursors of glycogen. What were they?
30. **Reflect and Apply** Explain how the minor structural difference between  $\alpha$ - and  $\beta$ -glucose is related to the differences in structure and function in the polymers formed from these two monomers.
31. **Reflect and Apply** All naturally occurring polysaccharides have one terminal residue, which contains a free anomeric carbon. Why do these polysaccharides *not* give a positive chemical test for a reducing sugar?
32. **Reflect and Apply** An amylose chain is 5000 glucose units long. At how many places must it be cleaved to reduce the average length to 2500 units? To 1000 units? To 200 units? What percentage of the glycosidic links are hydrolyzed in each case? (Even partial hydrolysis can drastically alter the physical properties of polysaccharides and thus affect their structural role in organisms.)
33. **Reflect and Apply** Suppose that a polymer of glucose with alternating  $\alpha(1 \rightarrow 4)$  and  $\beta(1 \rightarrow 4)$  glycosidic linkages has just been discovered. Draw a Haworth projection for a repeating tetramer (two repeating dimers) of such a polysaccharide. Would you expect this polymer to have primarily a structural role or an energy-storage role in organisms? What sort of organisms, if any, could use this polysaccharide as a food source?
34. **Reflect and Apply** Glycogen is highly branched. What advantage, if any, does this provide an animal?

35. **Reflect and Apply** No animal can digest cellulose. Reconcile this statement with the fact that many animals are herbivores that depend heavily on cellulose as a food source.
36. **Reflect and Apply** How does the presence of  $\alpha$ -bonds versus  $\beta$ -bonds influence the digestibility of glucose polymers by humans? *Hint: There are two effects.*
37. **Reflect and Apply** How do the sites of cleavage of starch differ from one another when the cleavage reaction is catalyzed by  $\alpha$ -amylase and  $\beta$ -amylase?
38. **Biochemical Connections** What is the benefit of fiber in the diet?
39. **Reflect and Apply** How would you expect the active site of a cellulase to differ from the active site of an enzyme that degrades starch?
40. **Reflect and Apply** Would you expect cross-linking to play a role in the structure of polysaccharides? If so, how would the cross-links be formed?
41. **Reflect and Apply** Compare the information in the sequence of monomers in a polysaccharide with that in the sequence of amino acid residues in a protein.
42. **Reflect and Apply** Why is it advantageous that polysaccharides can have branched chains? How do they achieve this structural feature?
43. **Reflect and Apply** Why is the polysaccharide chitin a suitable material for the exoskeleton of invertebrates such as lobsters? What other sort of material can play a similar role?
44. **Reflect and Apply** Could bacterial cell walls consist largely of protein? Why or why not?
45. **Reflect and Apply** Some athletes eat diets high in carbohydrates before an event. Suggest a biochemical basis for this practice.
46. **Reflect and Apply** You are a teaching assistant in a general chemistry lab. The next experiment is to be an oxidation–reduction titration involving iodine. You get a starch indicator from the stockroom. Why do you need it?
47. **Reflect and Apply** Blood samples for research or medical tests sometimes have heparin added. Why is this done?
48. **Reflect and Apply** Based on what you know about glycosidic bonds, propose a scheme for formation of covalent bonds between the carbohydrate and protein portions of glycoproteins.

### 16.5 Glycoproteins

49. **Recall** What are glycoproteins? What are some of their biochemical roles?
50. **Biochemical Connections** Briefly indicate the role of glycoproteins as antigenic determinants for blood groups.

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





# Glycolysis

# 17

© Dimitri Lunati/Corbis

For athletes, efficient use of carbohydrates can provide the margin of victory.

## 17.1 The Overall Pathway of Glycolysis

The first stage of glucose metabolism in organisms from bacteria to humans is called **glycolysis**, and it was the first biochemical pathway elucidated. In glycolysis, one molecule of glucose (a six-carbon compound) is converted to fructose-1,6-bisphosphate (also a six-carbon compound), which eventually gives rise to two molecules of pyruvate (a three-carbon compound) (Figure 17.1). The glycolytic pathway (also called the Embden–Meyerhoff pathway) involves many steps, including the reactions in which metabolites of glucose are oxidized. Each reaction in the pathway is catalyzed by an enzyme specific for that reaction. In each of two reactions in the pathway, one molecule of ATP is hydrolyzed for each molecule of glucose metabolized; the energy released in the hydrolysis of these two ATP molecules makes coupled endergonic reactions possible. In each of two other reactions, two molecules of ATP are produced by phosphorylation of ADP for each molecule of glucose, giving a total of four ATP molecules produced. A comparison of the number of ATP molecules used by hydrolysis (two) and the number produced (four) shows that there is a net gain of two ATP molecules for each molecule of glucose processed in glycolysis (Section 15.10). Glycolysis plays a key role in the way organisms extract energy from nutrients.

### What are the possible fates of pyruvate in glycolysis?

When pyruvate is formed, it can have one of several fates (Figure 17.1). In aerobic metabolism (in the presence of oxygen), pyruvate loses carbon dioxide. The remaining two carbon atoms become linked to coenzyme A (Section 15.11) as an acetyl group to form acetyl-CoA, which then enters the citric acid cycle (Chapter 19). There are two fates for pyruvate in anaerobic metabolism (in the absence of oxygen). In organisms capable of alcoholic fermentation, pyruvate loses carbon dioxide, this time producing acetaldehyde, which, in turn, is reduced to produce ethanol (Section 17.4). The more common fate of pyruvate in anaerobic metabolism is reduction to lactate, called **anaerobic glycolysis** to distinguish it from conversion of glucose to pyruvate, which is simply called glycolysis. Anaerobic metabolism is the only energy source in mammalian red blood cells, as well as in several species of bacteria, such as *Lactobacillus* in sour milk and *Clostridium botulinum* in tainted canned foods. The Biochemical Connections box on page 484 discusses an important practical application of fermentation.

In all these reactions, the conversion of glucose to product is an oxidation reaction, requiring an accompanying reduction reaction in which  $\text{NAD}^+$  is converted to NADH, a point to which we shall return when we discuss the pathway in detail. The breakdown of glucose to pyruvate can be summarized as follows:

Glucose (**Six carbon atoms**)  $\rightarrow$  2 Pyruvate (**Three carbon atoms**)

$2\text{ATP} + 4\text{ADP} + 2\text{P}_i \rightarrow 2\text{ADP} + 4\text{ATP}$  (**Phosphorylation**)

Glucose +  $2\text{ADP} + 2\text{P}_i \rightarrow 2$  Pyruvate +  $2\text{ATP}$  (**Net reaction**)

## Chapter Outline

### 17.1 The Overall Pathway of Glycolysis

- What are the possible fates of pyruvate in glycolysis?
- What are the reactions of glycolysis?

### 17.2 Conversion of Six-Carbon Glucose to Three-Carbon Glyceraldehyde-3-Phosphate

- What reactions convert glucose-6-phosphate to glyceraldehyde-3-phosphate?

### 17.3 Glyceraldehyde-3-Phosphate Is Converted to Pyruvate

- What reactions convert glyceraldehyde-3-phosphate to pyruvate?
- Where are the control points in the glycolytic pathway?

### 17.4 Anaerobic Metabolism of Pyruvate

- How does the conversion of pyruvate to lactate take place in muscle?
- How does alcoholic fermentation take place?

### 17.5 Energy Production in Glycolysis

- What is the energy yield from glycolysis?

**OWL** Online homework for this chapter may be assigned in OWL.

■ **FIGURE 17.1** One molecule of glucose is converted to two molecules of pyruvate. Under aerobic conditions, pyruvate is oxidized to  $\text{CO}_2$  and  $\text{H}_2\text{O}$  by the citric acid cycle (Chapter 19) and oxidative phosphorylation (Chapter 20). Under anaerobic conditions, lactate is produced, especially in muscle. Alcoholic fermentation occurs in yeast. The NADH produced in the conversion of glucose to pyruvate is reoxidized to  $\text{NAD}^+$  in the subsequent reactions of pyruvate.

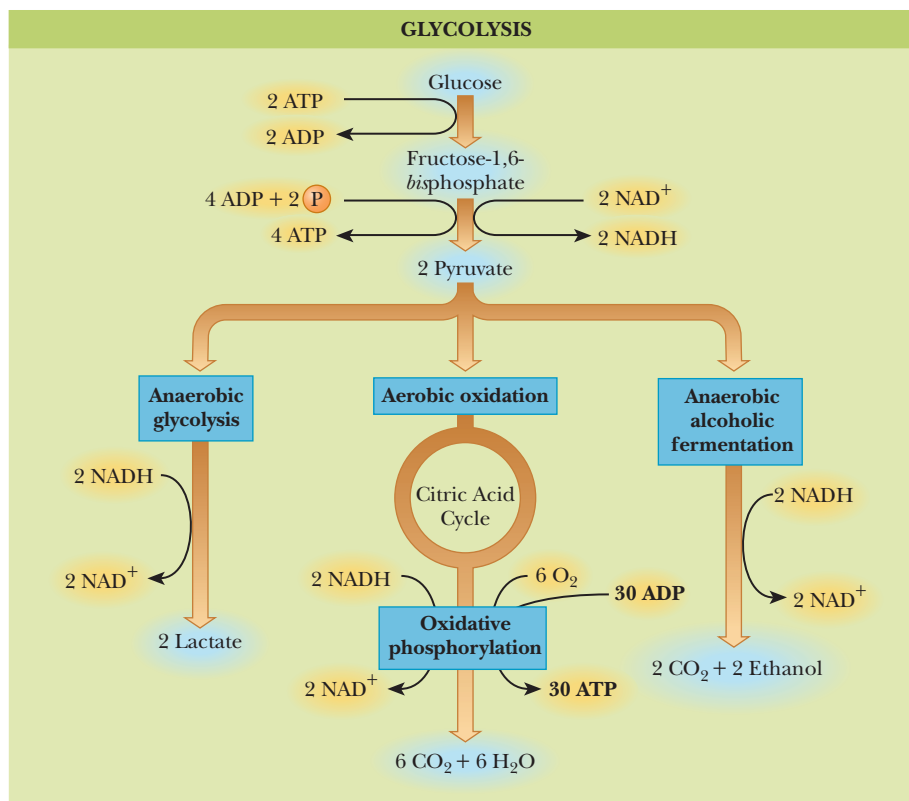


Figure 17.2 shows the reaction sequence with the names of the compounds. All sugars in the pathway have the D configuration; we shall assume this point throughout this chapter.

### What are the reactions of glycolysis?

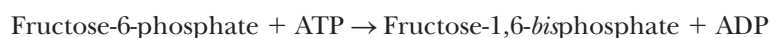
**Step 1. Phosphorylation** of glucose to give glucose-6-phosphate (ATP is the source of the phosphate group). (See Equation 17.1, page 485.)



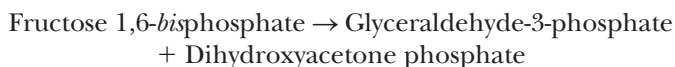
**Step 2. Isomerization** of glucose-6-phosphate to give fructose-6-phosphate. (See Equation 17.2, page 489.)



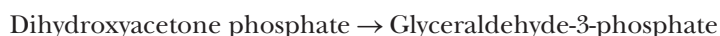
**Step 3. Phosphorylation** of fructose-6-phosphate to give fructose-1,6-bisphosphate (ATP is the source of the phosphate group). (See Equation 17.3, page 489.)

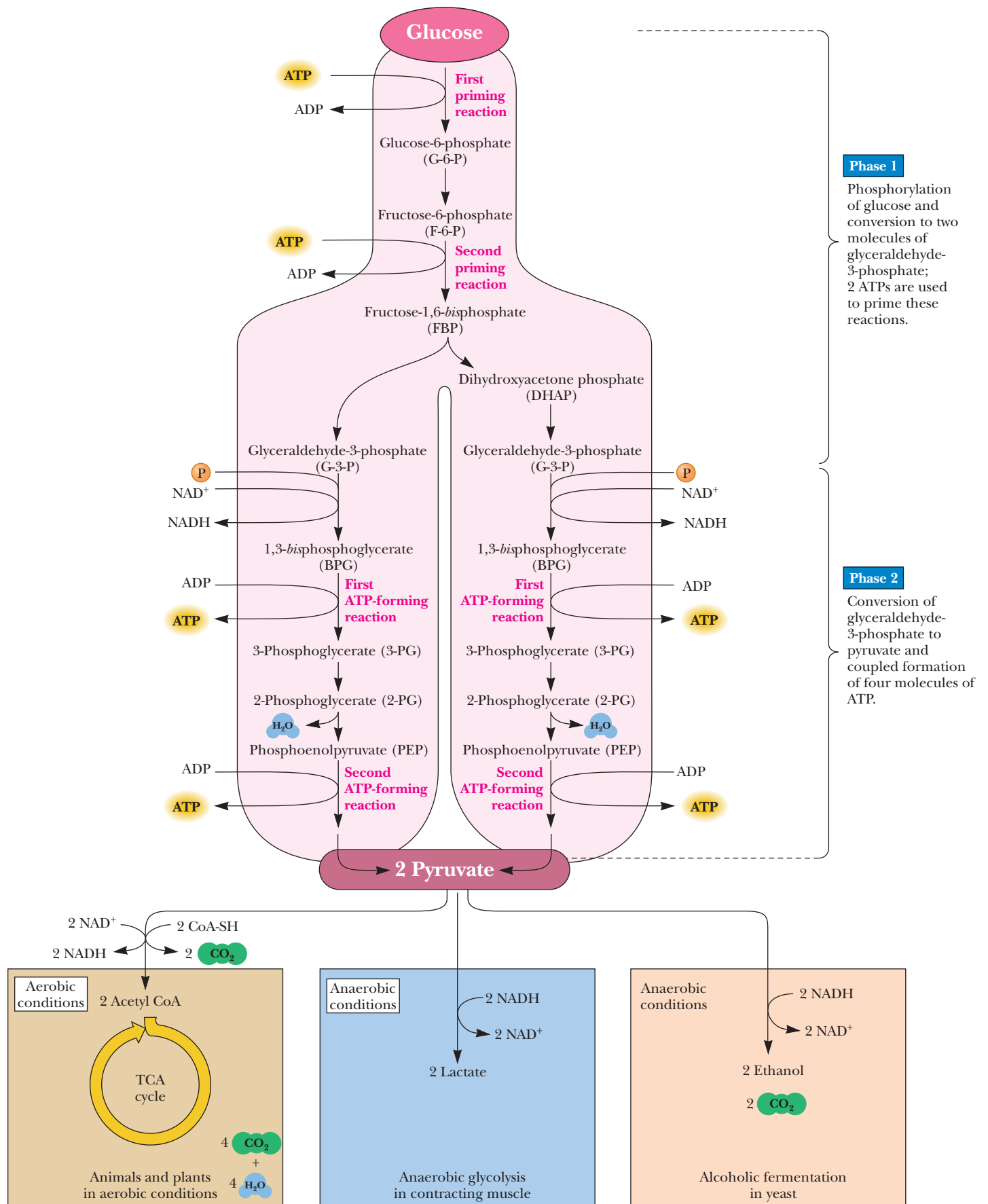


**Step 4. Cleavage** of fructose-1,6-bisphosphate to give two 3-carbon fragments, glyceraldehyde-3-phosphate and dihydroxyacetone phosphate. (See Equation 17.4, page 490.)



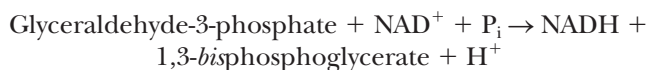
**Step 5. Isomerization** of dihydroxyacetone phosphate to give glyceraldehyde-3-phosphate. (See Equation 17.5, page 490.)





■ **FIGURE 17.2** The glycolytic pathway.

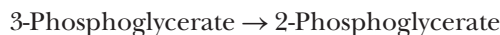
**Step 6. Oxidation** (and phosphorylation) of glyceraldehyde-3-phosphate to give 1,3-bisphosphoglycerate. (See Equation 17.6, page 491.)



**Step 7. Transfer of a phosphate group** from 1,3-bisphosphoglycerate to ADP (phosphorylation of ADP to ATP) to give 3-phosphoglycerate. (See Equation 17.7, page 495.)



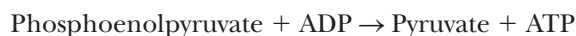
**Step 8. Isomerization** of 3-phosphoglycerate to give 2-phosphoglycerate. (See Equation 17.8, page 497.)



**Step 9. Dehydration** of 2-phosphoglycerate to give phosphoenolpyruvate. (See Equation 17.9, page 497.)



**Step 10. Transfer of a phosphate group** from phosphoenolpyruvate to ADP (phosphorylation of ADP to ATP) to give pyruvate. (See Equation 17.10, page 497.)



Note that only one of the 10 steps in this pathway involves an electron-transfer reaction. We shall now look at each of these reactions in detail.

## Biochemical Connections ENVIRONMENTAL SCIENCE

### Biofuels from Fermentation

The grave concern about depletion of fossil fuels, especially petroleum-based ones, has led to interest in developing renewable energy sources. Most kinds of organic matter can be used as fuel. Wood and animal dung have been used for centuries, and continue to be used in many parts of the world. Still, many kinds of engines are designed to operate with liquid fuels, such as gasoline in cars. Ethanol is a common liquid product of fermentation of carbohydrates, and its use as a fuel is being discussed widely.

Ethanol is not normally used as the sole fuel in an internal combustion engine but is usually mixed with gasoline. The specific

formulation is designated with an E number, such as E10, E25, or E85, where the number refers to the percentage of ethanol in the mixture. E10 is frequently used in the United States, and the use of E25 is mandatory in Brazil. E85 is used in some European countries, particularly Sweden, where it is the standard fuel for the flexible-fuel vehicles that can run on gasoline or on ethanol-containing mixtures. E85 is becoming more available in the United States too. Only certain vehicles, labeled “flex-fuel,” can use it. Just as diesel pumps have a special color code (green), the E85 has the color code yellow. When the E85 pump is located on the same machine as regular gasoline, it comes with a large warning sign that says, “Warning, this is not gasoline,” to help people avoid making a costly mistake.

It is possible to use any number of carbohydrate sources to obtain ethanol. Paper mill waste, peanut shells, sawdust, and garbage have all been used in the past. Note that paper and sawdust are primarily cellulose, which is a polymer of glucose. Almost any kind of plant material can be fermented to produce ethanol, which is then recovered by distillation. Perhaps the most widely used source of ethanol is corn. It is certainly a widely grown crop, especially in the United States. Sorghum and soybeans have also been suggested as sources, and they are also widely grown crops. Many farming communities in the Midwest have welcomed this trend as an economic boon. On the other hand, concerns have been raised about diversion of food crops for energy, and many economists are predicting an increase in prices of many food items as the supply of food crops decreases to make way for biofuel crops. This point is the subject of a number of editorials and letters to the editor in a wide spectrum of both science and business news media. The use of biofuels, particularly in the form of ethanol, is in its infancy. We can expect to hear much more discussion of the subject in the next few years.



© Grant Heilman Photography/Alamy

## 17.2 Conversion of Six-Carbon Glucose to Three-Carbon Glyceraldehyde-3-Phosphate

The first steps of the glycolytic pathway prepare for the electron transfer and the eventual phosphorylation of ADP; these reactions make use of the free energy of hydrolysis of ATP. Figure 17.3 summarizes this part of the pathway, which is often called the *preparation phase* of glycolysis.

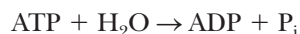
### What reactions convert glucose-6-phosphate to glyceraldehyde-3-phosphate?

**Step 1.** Glucose is phosphorylated to give glucose-6-phosphate. The phosphorylation of glucose is an endergonic reaction.



$$\Delta G^{\circ'} = 13.8 \text{ kJ mol}^{-1} = 3.3 \text{ kcal mol}^{-1}$$

The hydrolysis of ATP is exergonic.

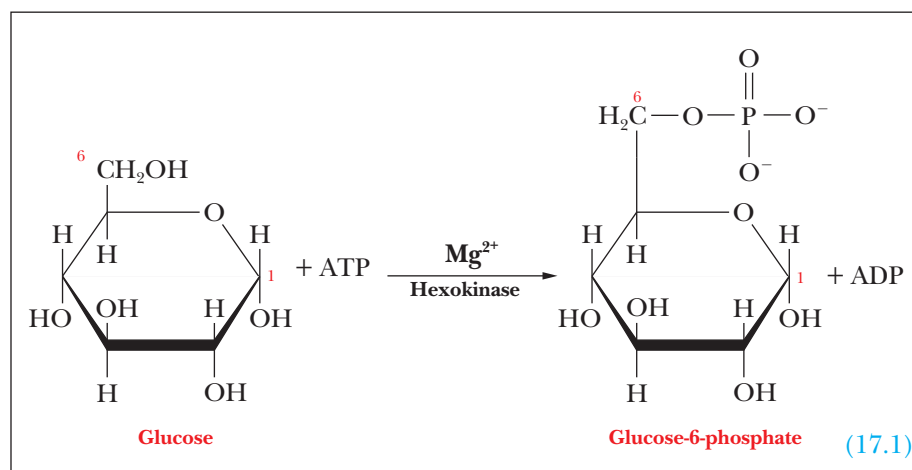


$$\Delta G^{\circ'} = -30.5 \text{ kJ mol}^{-1} = -7.3 \text{ kcal mol}^{-1}$$

These two reactions are coupled, so the overall reaction is the sum of the two and is exergonic.



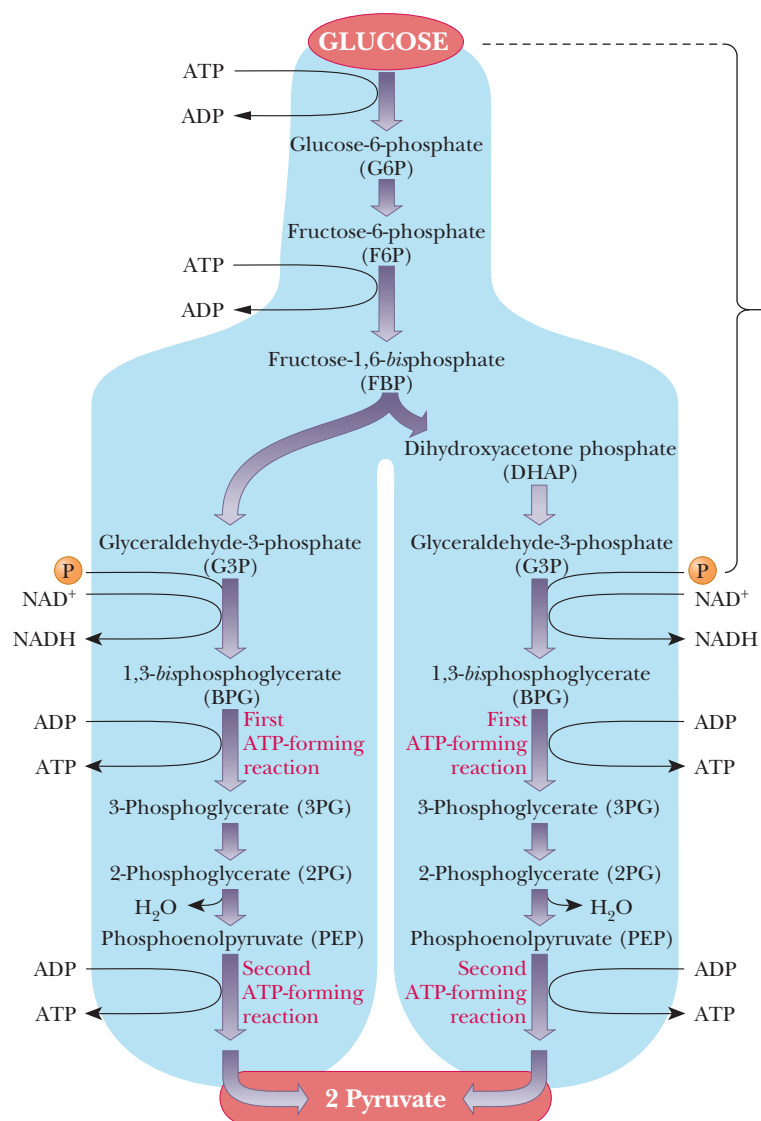
$$\Delta G^{\circ'} = (13.8 + -30.5) \text{ kJ mol}^{-1} = -16.7 \text{ kJ mol}^{-1} = -4.0 \text{ kcal mol}^{-1}$$



Recall that  $\Delta G^{\circ'}$  is calculated under standard states with the concentration of all reactants and products at 1 *M* except hydrogen ion. If we look at the actual free-energy change in the cell, the number varies depending on cell type and metabolic state, but a typical value for this reaction is  $-33.9 \text{ kJ mol}^{-1}$  or  $-8.12 \text{ kcal mol}^{-1}$ . Thus the reaction is typically even more favorable under cellular conditions. Table 17.1 gives the  $\Delta G^{\circ'}$  and  $\Delta G$  values for all the reactions of anaerobic glycolysis in erythrocytes.

This reaction illustrates the use of chemical energy originally produced by the oxidation of nutrients and ultimately trapped by phosphorylation of ADP to ATP. Recall from Section 15.10 that ATP does not represent stored energy, just as an electric current does not represent stored energy. The chemical energy of nutrients is released by oxidation and is made available for immediate use on demand by being trapped as ATP.





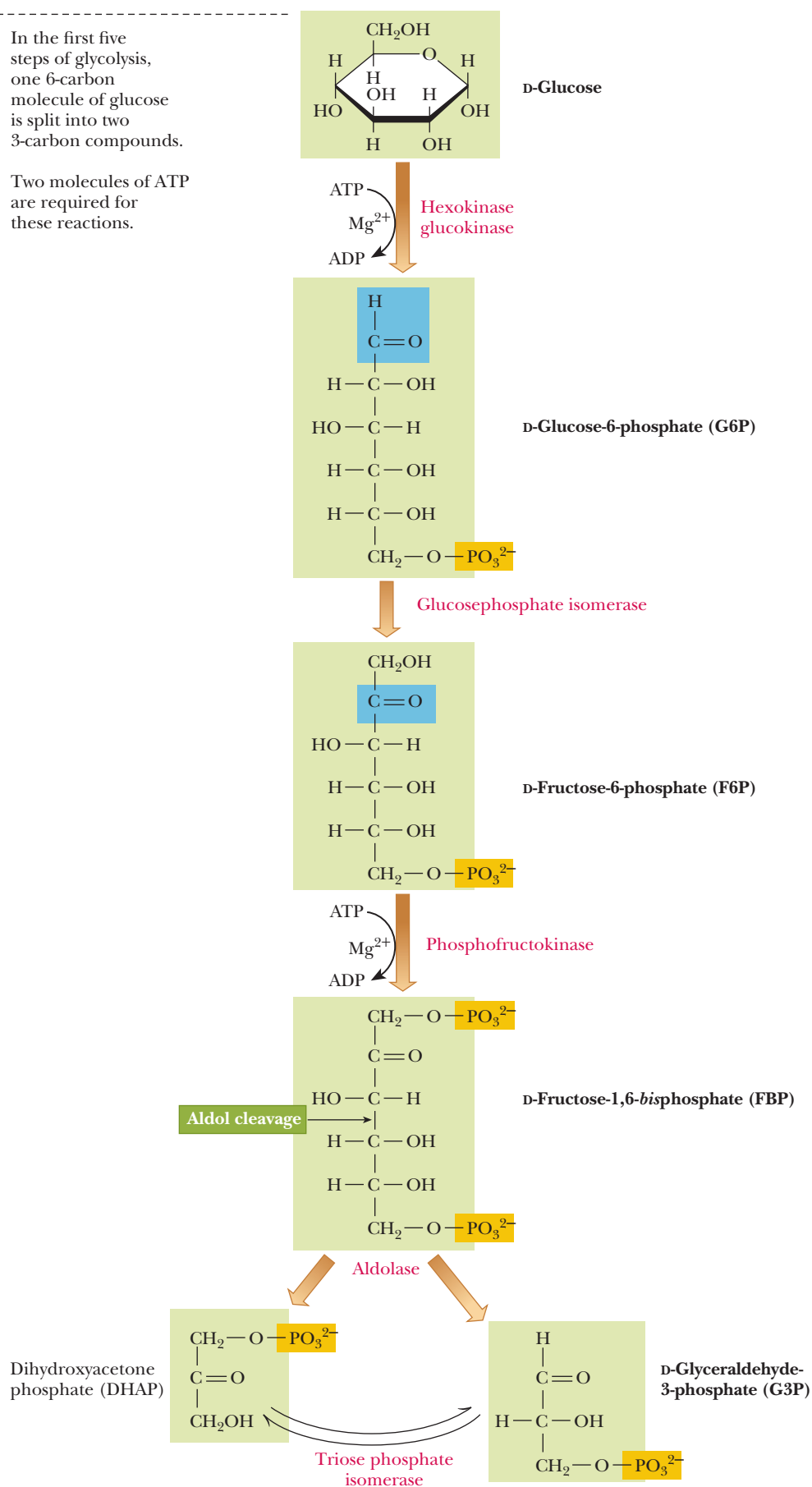
■ **FIGURE 17.3** Conversion of pyruvate to glyceraldehyde-3-phosphate. In the first phase of glycolysis, five reactions convert a molecule of glucose to two molecules of glyceraldehyde-3-phosphate. The structures of reaction components are shown on the facing page.

The enzyme that catalyzes this reaction is **hexokinase**. The term *kinase* is applied to the class of ATP-dependent enzymes that transfer a phosphate group from ATP to a substrate. The substrate of hexokinase is not necessarily glucose; rather, it can be any one of a number of hexoses, such as glucose, fructose, and mannose. Glucose-6-phosphate inhibits the activity of hexokinase; this is a control point in the pathway. Some organisms or tissues contain multiple isozymes of hexokinase. One isoform of hexokinase found in the human liver, called glucokinase, lowers blood glucose levels after one has eaten a meal. Liver glucokinase requires a much higher substrate level to achieve saturation than hexokinase does. Because of this, when glucose levels are high, the liver can metabolize glucose via glycolysis preferentially over the other tissues. When glucose levels are low, hexokinase is still active in all tissues (see the Biochemical Connections box on page 157).

A large conformational change takes place in hexokinase when substrate is bound. It has been shown by X-ray crystallography that, in the

In the first five steps of glycolysis, one 6-carbon molecule of glucose is split into two 3-carbon compounds.

Two molecules of ATP are required for these reactions.



**TABLE 17.1**

The Reactions of Glycolysis and Their Standard Free-Energy Changes					
Step	Reaction	Enzyme	$\Delta G^{\circ\prime}$		$\Delta G^{**}$
			$\text{kJ mol}^{-1}$	$\text{kcal mol}^{-1}$	$\text{kJ mol}^{-1}$
1	Glucose + ATP → Glucose-6-phosphate + ADP	Hexokinase/ Glucokinase	−16.7	−4.0	−33.9
2	Glucose-6-phosphate → Fructose-6-phosphate	Glucose phosphate isomerase	+1.67	+0.4	−2.92
3	Fructose-6-phosphate + ATP → Fructose-1,6-bisphosphate + ADP	Phosphofructokinase	−14.2	−3.4	−18.8
4	Fructose-1,6-bisphosphate → Dihydroxyacetone phosphate + Glyceraldehyde-3-phosphate	Aldolase	+23.9	+5.7	−0.23
5	Dihydroxyacetone phosphate → Glyceraldehyde-3-phosphate	Triose phosphate isomerase	+7.56	+1.8	+2.41
6	2(Glyceraldehyde-3-phosphate + NAD <sup>+</sup> + P <sub>i</sub> → 1,3-bisphosphoglycerate + NADH + H <sup>+</sup> )	Glyceraldehyde-3-P dehydrogenase	2(+6.20)	2(+1.5)	2(−1.29)
7	2(1,3-bisphosphoglycerate + ADP → 3-Phosphoglycerate + ATP)	Phosphoglycerate kinase	2(−18.8)	2(−4.5)	2(+0.1)
8	2(3-Phosphoglycerate → 2-Phosphoglycerate)	Phosphoglyceromutase	2(+4.4)	2(+1.1)	2(+0.83)
9	2(2-Phosphoglycerate → Phosphoenolpyruvate + H <sub>2</sub> O)	Enolase	2(+1.8)	2(+0.4)	2(+1.1)
10	2(Phosphoenolpyruvate + ADP → Pyruvate + ATP)	Pyruvate kinase	2(−31.4)	2(−7.5)	2(−23.0)
Overall	Glucose + 2ADP + 2P <sub>i</sub> + NAD <sup>+</sup> →	Lactate dehydrogenase	−73.3	−17.5	−98.0
	2 Pyruvate → 2ATP + NADH + H <sup>+</sup>		2(−25.1)	2(−6.0)	2(−14.8)
	2(Pyruvate + NADH + H <sup>+</sup> → Lactate + NAD <sup>+</sup> )		−123.5	−29.5	−127.6
	Glucose + 2ADP + 2P <sub>i</sub> → 2 Lactate + 2ATP				

\*  $\Delta G^{\circ\prime}$  values are assumed to be the same at 25°C and 37°C and are calculated for standard-state conditions (1 M concentration of reactants and products pH 7.0).  
\*\*  $\Delta G$  values are calculated at 310 K (37°C) using steady-state concentrations of these metabolites found in erythrocytes.

# Biochemical Connections ALLIED HEALTH

## Dolphins as a Model for Humans with Diabetes

Dolphins have fascinated humans since at least the time of the ancient Greeks. The level of interest can only increase with the announcement in February 2010 that dolphins can share with humans symptoms typical of type 2 diabetes.

For a number of years, veterinarians have monitored the health of 52 dolphins owned by the U.S. Navy. Blood samples reveal that fasting dolphins have high levels of blood glucose, similar to diabetic humans. Blood glucose levels return to normal after a meal. The dolphins have been observed to maintain glucose levels on a high-protein diet. When the animals have high levels of glucose, they also have high levels of iron and triglycerides in the blood, typical of diabetes. Dolphins and humans respond to low-carbohydrate, high-protein diets with similar metabolic changes. In addition, the central nervous system of both species has a high demand for glucose, in view of the high brain-to-body-mass ratio common to both. These similarities, combined with the way that the symptoms can be turned on and off with diet in dolphins, make them a model for humans with diabetes.

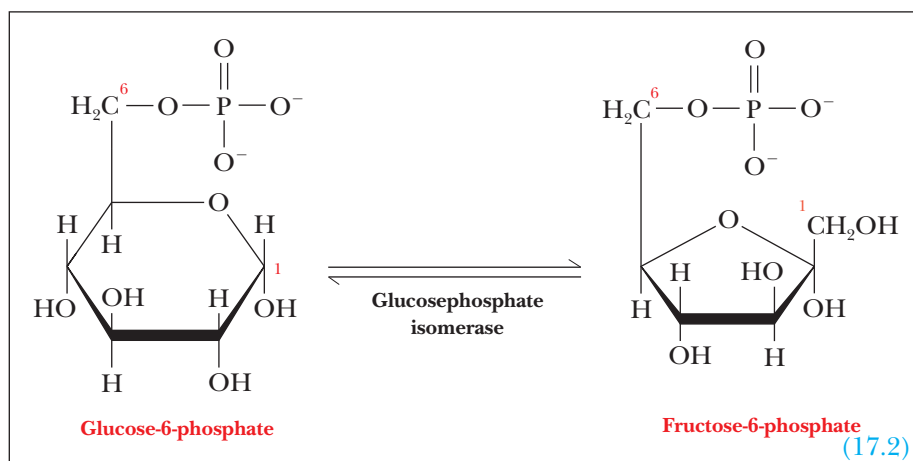


■ Dolphins can share symptoms of type 2 diabetes with humans.

absence of substrate, two lobes of the enzyme that surround the binding site are quite far apart. When glucose is bound, the two lobes move closer together, and the glucose becomes almost completely surrounded by protein (Figure 17.4).

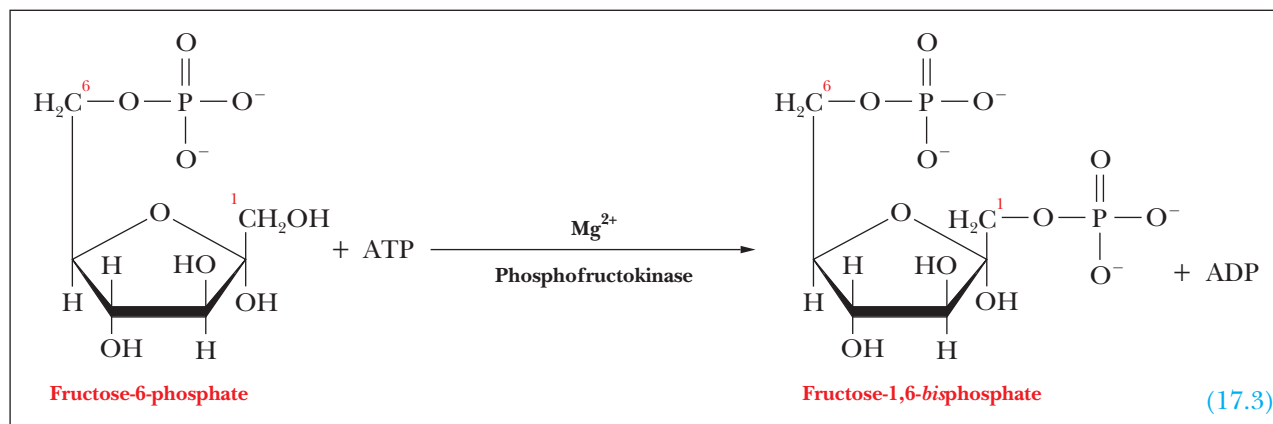
This type of behavior is consistent with the induced-fit theory of enzyme action (Section 6.4). In all kinases for which the structure is known, a cleft closes when substrate is bound.

**Step 2.** Glucose-6-phosphate isomerizes to give fructose-6-phosphate. **Glucose-phosphate isomerase** is the enzyme that catalyzes this reaction. The C-1 aldehyde group of glucose-6-phosphate is reduced to a hydroxyl group, and the C-2 hydroxyl group is oxidized to give the ketone group of fructose-6-phosphate, with no net oxidation or reduction. (Recall from Section 16.1 that glucose is an aldose, a sugar whose open-chain, noncyclic structure contains an aldehyde group, while fructose is a ketose, a sugar whose corresponding structure contains a ketone group.) The phosphorylated forms, glucose-6-phosphate and fructose-6-phosphate, are an aldose and a ketose, respectively.

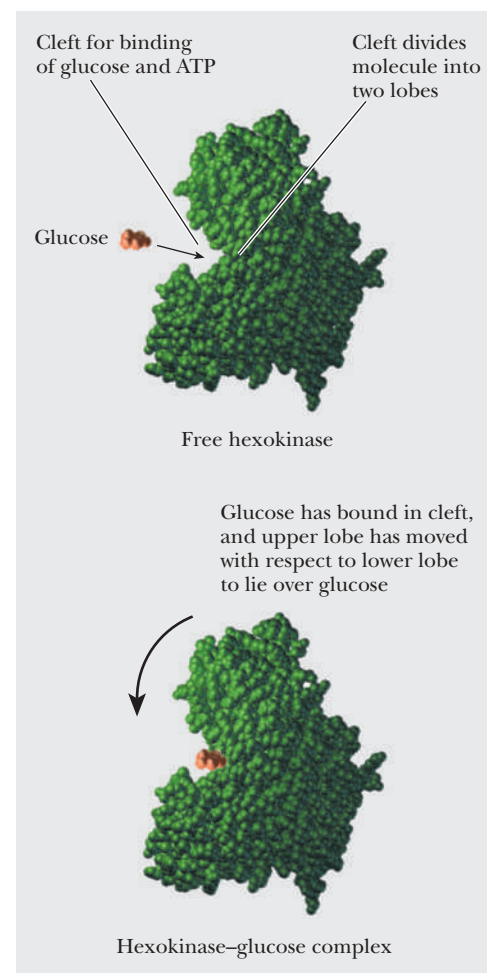


**Step 3.** Fructose-6-phosphate is further phosphorylated, producing fructose-1,6-bisphosphate.

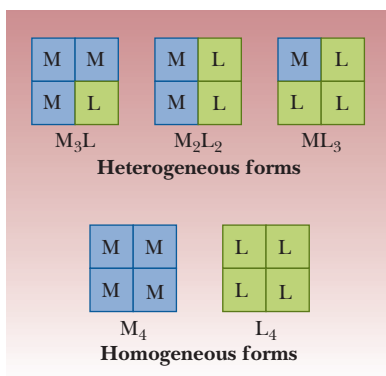
As in the reaction in Step 1, the endergonic reaction of phosphorylation of fructose-6-phosphate is coupled to the exergonic reaction of hydrolysis of ATP, and the overall reaction is exergonic. See Table 17.1.



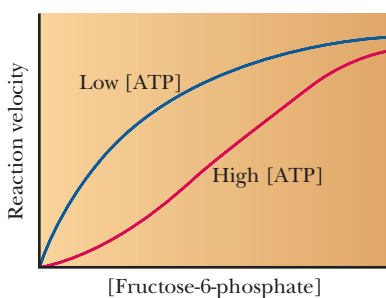
The reaction in which fructose-6-phosphate is phosphorylated to give fructose-1,6-bisphosphate is the one in which the sugar is committed to



■ **FIGURE 17.4** A comparison of the conformations of hexokinase and the hexokinase-glucose complex.



■ **FIGURE 17.5** The possible isozymes of phosphofructokinase.



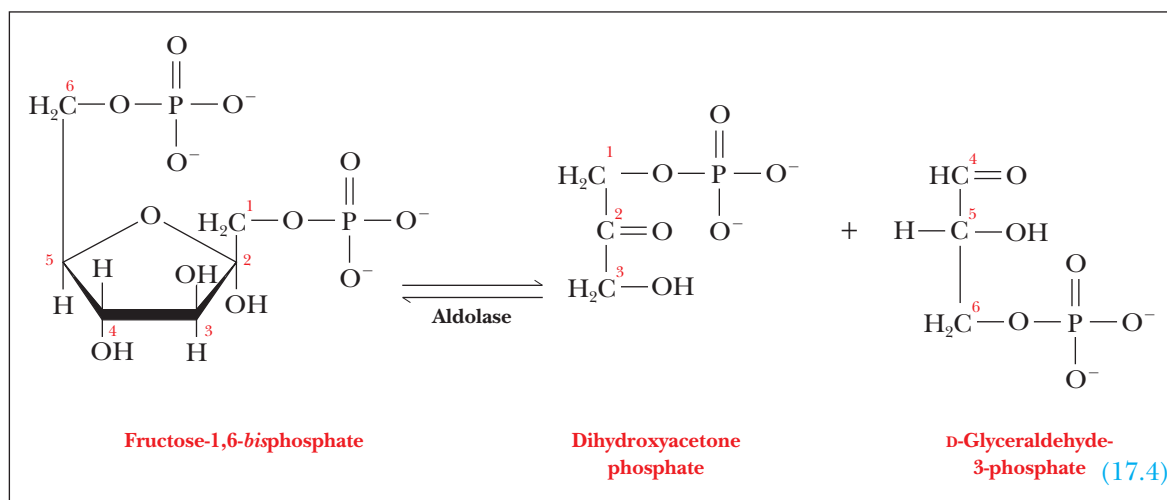
■ **FIGURE 17.6** Allosteric effects in phosphofructokinase. At high [ATP], phosphofructokinase behaves cooperatively, and the plot of enzyme activity versus [fructose-6-phosphate] is sigmoidal. High [ATP] thus inhibits PFK, decreasing the enzyme's affinity for fructose-6-phosphate.

glycolysis. Glucose-6-phosphate and fructose-6-phosphate can play roles in other pathways, but fructose-1,6-bisphosphate does not. After fructose-1,6-bisphosphate is formed from the original sugar, no other pathways are available, and the molecule must undergo the rest of the reactions of glycolysis. The phosphorylation of fructose-6-phosphate is highly exergonic and irreversible, and **phosphofructokinase**, the enzyme that catalyzes it, is the key regulatory enzyme in glycolysis.

Phosphofructokinase is a tetramer that is subject to allosteric feedback regulation of the type we discussed in Chapter 7. There are two types of subunits, designated M and L, that can combine into tetramers to give different permutations ( $M_4$ ,  $M_3L$ ,  $M_2L_2$ ,  $ML_3$ , and  $L_4$ ). These combinations of subunits are referred to as **isozymes**, and they have subtle physical and kinetic differences (Figure 17.5). The subunits differ slightly in amino acid composition, so the two isozymes can be separated from each other by electrophoresis (Chapter 5). The tetrameric form that occurs in muscle is designated  $M_4$ , while that in liver is designated  $L_4$ . In red blood cells, several of the combinations can be found. Individuals who lack the gene that directs the synthesis of the M form of the enzyme can carry on glycolysis in their livers but experience muscle weakness because they lack the enzyme in muscle.

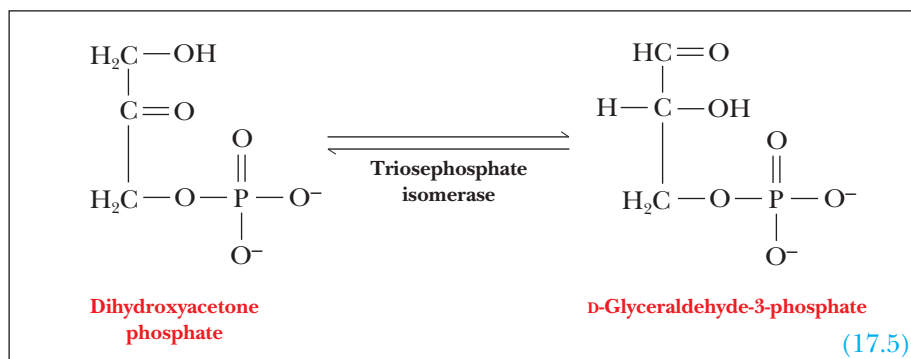
When the rate of the phosphofructokinase reaction is observed at varying concentrations of substrate (fructose-6-phosphate), the sigmoidal curve typical of allosteric enzymes is obtained. ATP is an allosteric effector in the reaction. High levels of ATP depress the rate of the reaction, and low levels of ATP stimulate the reaction (Figure 17.6). When there is a high level of ATP in the cell, a good deal of chemical energy is immediately available from hydrolysis of ATP. The cell does not need to metabolize glucose for energy, so the presence of ATP inhibits the glycolytic pathway at this point. There is also another, more potent, allosteric effector of phosphofructokinase. This effector is fructose-2,6-bisphosphate; we shall discuss its mode of action in Section 18.3 when we consider general control mechanisms in carbohydrate metabolism.

**Step 4.** Fructose-1,6-bisphosphate is split into two three-carbon fragments. The cleavage reaction here is the reverse of an aldol condensation; the enzyme that catalyzes it is called **aldolase**. In the enzyme isolated from most animal sources (the one from muscle is the most extensively studied), the basic side chain of an essential lysine residue plays the key role in catalyzing this reaction. The thiol group of a cysteine also acts as a base here.



**Step 5.** The dihydroxyacetone phosphate is converted to glyceraldehyde-3-phosphate.





The enzyme that catalyzes this reaction is **triosephosphate isomerase**. (Both dihydroxyacetone and glyceraldehyde are trioses.)

One molecule of glyceraldehyde-3-phosphate has already been produced by the aldolase reaction; we now have a second molecule of glyceraldehyde-3-phosphate, produced by the triosephosphate isomerase reaction. The original molecule of glucose, which contains six carbon atoms, has now been converted to two molecules of glyceraldehyde-3-phosphate, each of which contains three carbon atoms.

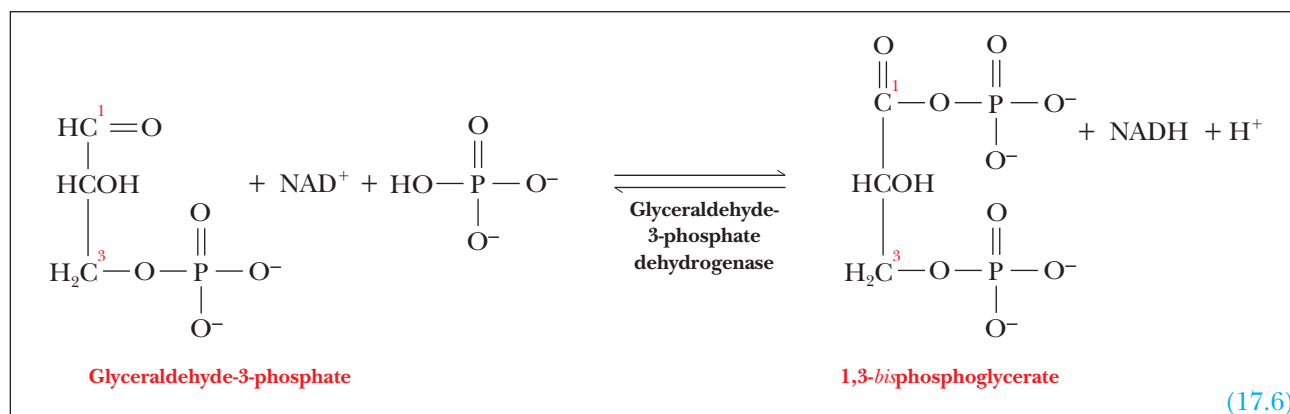
The  $\Delta G$  value for this reaction under physiological conditions is slightly positive ( $+2.41 \text{ kJ mol}^{-1}$  or  $+0.58 \text{ kcal mol}^{-1}$ ). It might be tempting to think that the reaction would not occur and that glycolysis would be halted at this step. We must remember that just as coupled reactions involving ATP hydrolysis add their  $\Delta G$  values together for the overall reaction, glycolysis is composed of many reactions that have very negative  $\Delta G$  values that can drive the reaction to completion. A few reactions in glycolysis have small, positive  $\Delta G$  values (see Table 17.1), but four reactions have very large, negative values, so that the  $\Delta G$  for the whole process is negative.

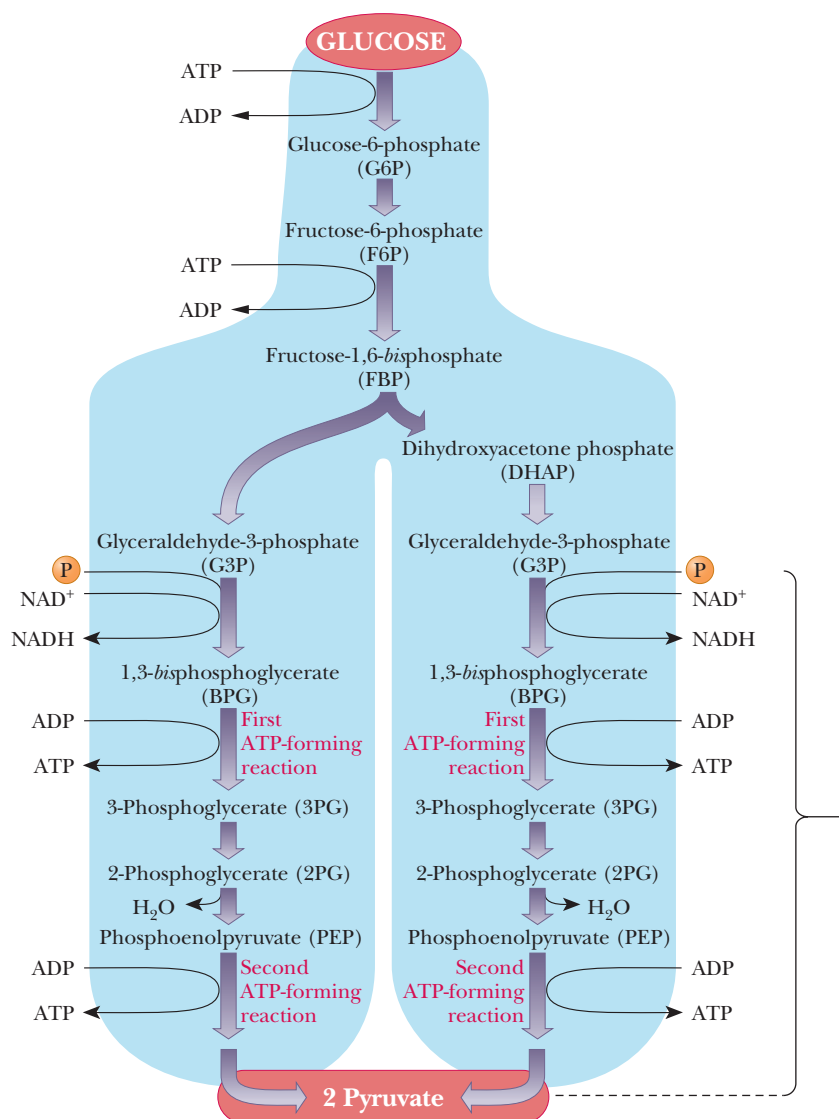
## 17.3 Glyceraldehyde-3-Phosphate Is Converted to Pyruvate

At this point, a molecule of glucose (a six-carbon compound) that enters the pathway has been converted to two molecules of glyceraldehyde-3-phosphate. We have not seen any oxidation reactions yet, but now we shall encounter them. Keep in mind that in the rest of the pathway two molecules of each of the three-carbon compounds take part in every reaction for each original glucose molecule. Figure 17.7 summarizes the second part of the pathway, which is often referred to as the *payoff phase* of glycolysis, since ATP is produced instead of used in this phase.

### What reactions convert glyceraldehyde-3-phosphate to pyruvate?

**Step 6.** Glyceraldehyde-3-phosphate is oxidized to 1,3-bisphosphoglycerate.





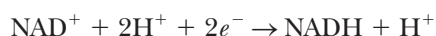
■ **FIGURE 17.7** The second phase of glycolysis. Structures are shown on the facing page.

This reaction, *the* characteristic reaction of glycolysis, should be looked at more closely. It involves the addition of a phosphate group to glyceraldehyde-3-phosphate as well as an electron-transfer reaction, from glyceraldehyde-3-phosphate to NAD<sup>+</sup>. We will simplify the discussion by considering the two parts separately.

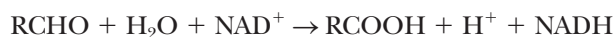
The half reaction of oxidation is that of an aldehyde to a carboxylic acid group, in which water can be considered to take part in the reaction.



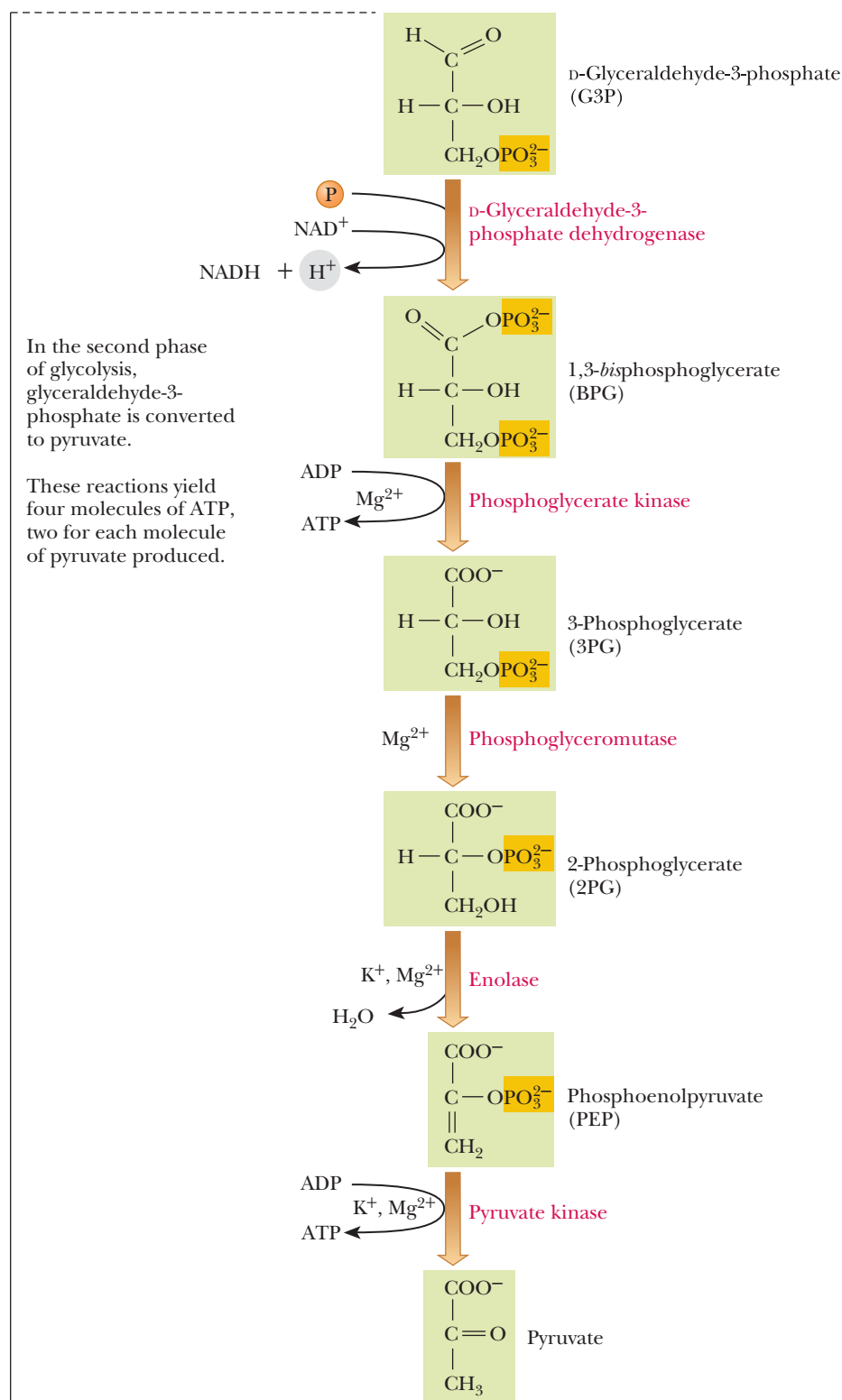
The half reaction of reduction is that of NAD<sup>+</sup> to NADH (Section 15.9).



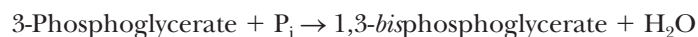
The overall redox reaction is thus



in which R indicates the portions of the molecule other than the aldehyde and carboxylic acid groups, respectively. The oxidation reaction is exergonic under standard conditions ( $\Delta G^\circ = -43.1 \text{ kJ mol}^{-1} = -10.3 \text{ kcal mol}^{-1}$ ), but oxidation is only part of the overall reaction.

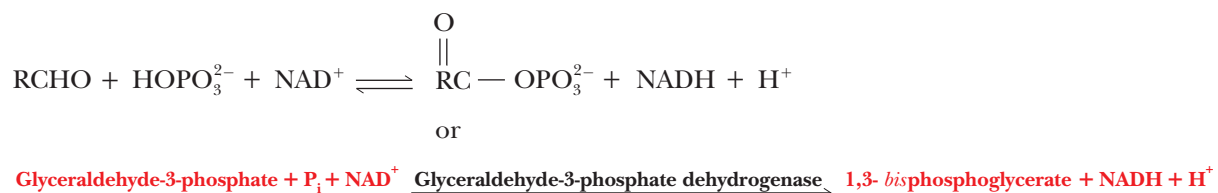


The phosphate group that is linked to the carboxyl group does not form an ester, since an ester linkage requires an alcohol and an acid. Instead, the carboxylic acid group and phosphoric acid form a mixed anhydride of two acids by loss of water (Section 2.2),



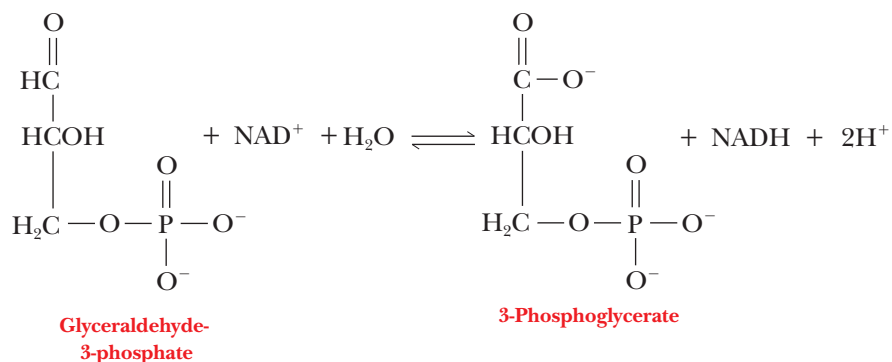
in which the substances involved in the reaction are in the ionized form appropriate at pH 7. Note that ATP and ADP do not appear in the equation. The source of the phosphate group is phosphate ion itself, rather than ATP. The phosphorylation reaction is endergonic under standard conditions ( $\Delta G^{\circ'} = 49.3 \text{ kJ mol}^{-1} = 11.8 \text{ kcal mol}^{-1}$ ).

The overall reaction, including electron transfer and phosphorylation, is

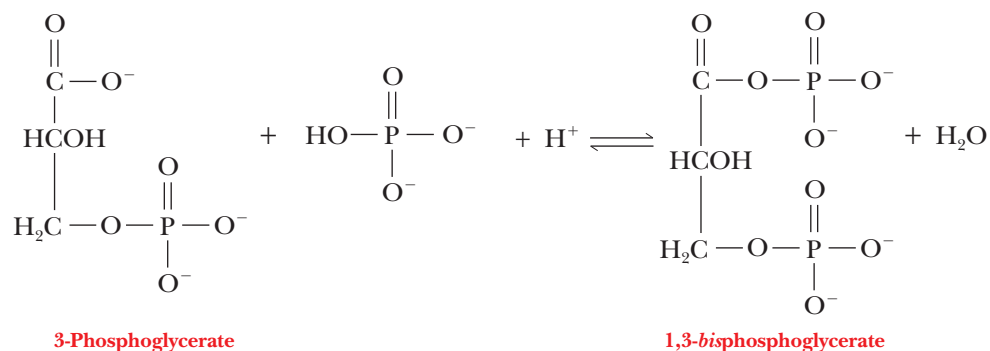


Let's show the two reactions that make up this reaction.

**1. Oxidation of glyceraldehyde-3-phosphate ( $\Delta G^{\circ'} = -43.1 \text{ kJ mol}^{-1} = -10.3 \text{ kcal mol}^{-1}$ )**

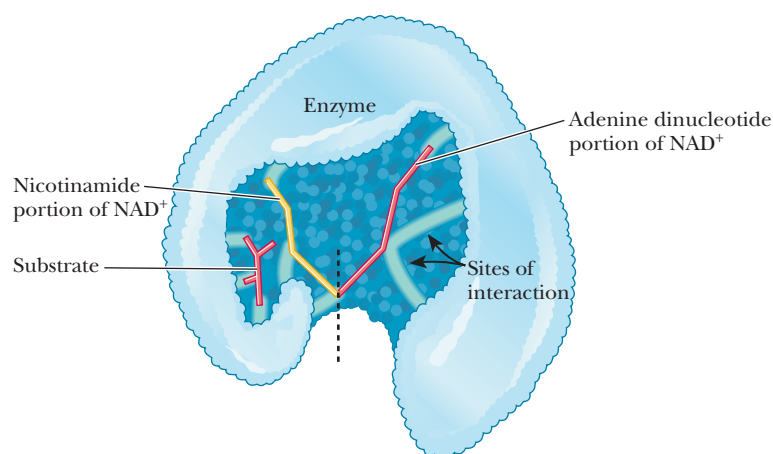


**2. Phosphorylation of 3-phosphoglycerate ( $\Delta G^{\circ'} = 49.3 \text{ kJ mol}^{-1} = 11.8 \text{ kcal mol}^{-1}$ )**



The standard free-energy change for the overall reaction is the sum of the values for the oxidation and phosphorylation reactions. The overall reaction is not far from equilibrium, being only slightly endergonic.

$$\begin{aligned} \Delta G^{\circ'}_{\text{overall}} &= \Delta G^{\circ'}_{\text{oxidation}} + \Delta G^{\circ'}_{\text{phosphorylation}} \\ &= (-43.1 \text{ kJ mol}^{-1}) + (49.3 \text{ kJ mol}^{-1}) \\ &= 6.2 \text{ kJ mol}^{-1} = 1.5 \text{ kcal mol}^{-1} \end{aligned}$$

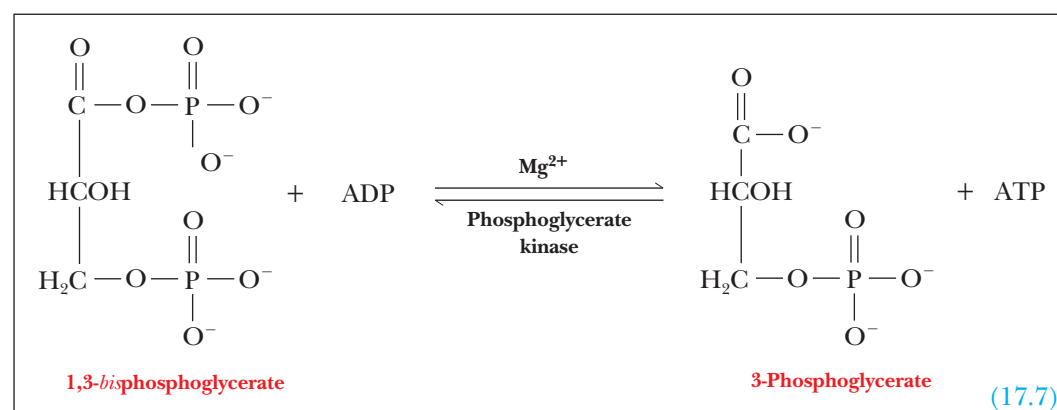


■ **FIGURE 17.8** Schematic view of the binding site of an NADH-linked dehydrogenase. There are specific binding sites for the adenine nucleotide portion of the coenzyme (shown in red to the right of the dashed line) and for the nicotinamide portion of the coenzyme (shown in yellow to the left of the dashed line), in addition to the binding site for the substrate. Specific interactions with the enzyme hold the substrate and coenzyme in the proper positions. Sites of interaction are shown as a series of pale green lines.

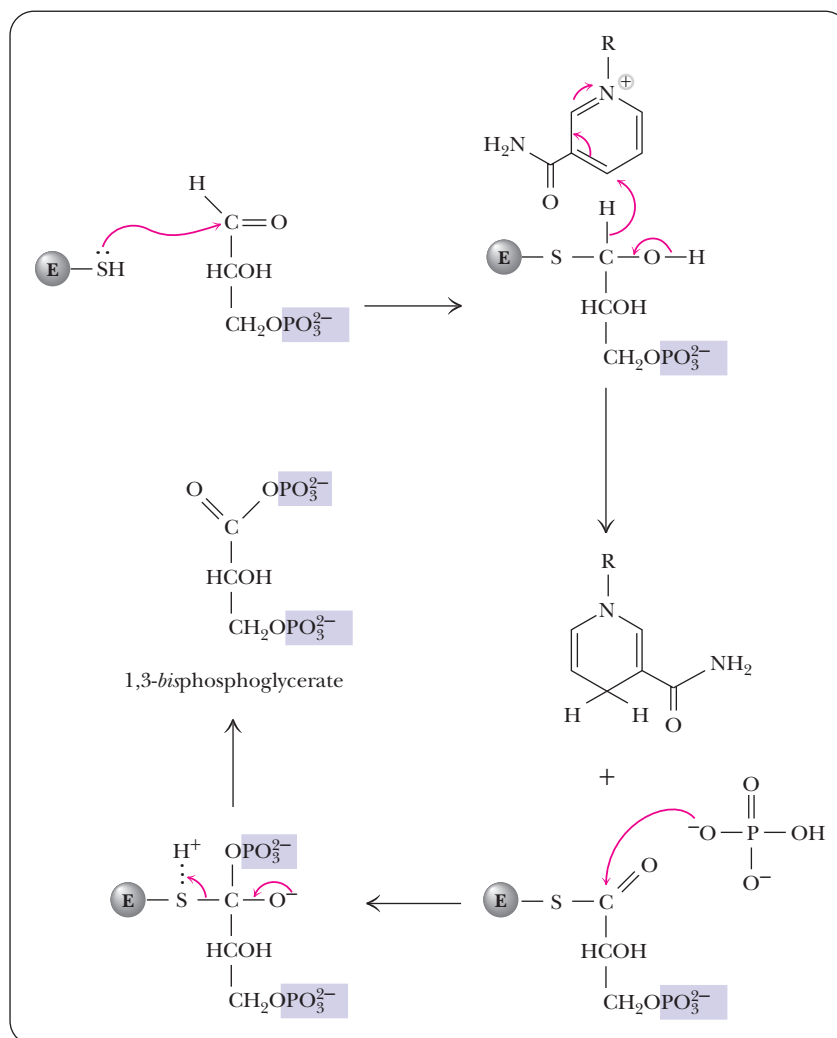
This value of the standard free-energy change is for the reaction of one mole of glyceraldehyde-3-phosphate; the value must be multiplied by 2 to get the value for each mole of glucose ( $\Delta G^\circ = 12.4 \text{ kJ mol}^{-1} = 3.0 \text{ kcal mol}^{-1}$ ). The  $\Delta G$  under cellular conditions is slightly negative ( $-1.29 \text{ kJ mol}^{-1}$  or  $-0.31 \text{ kcal mol}^{-1}$ ) (Table 17.1). The enzyme that catalyzes the conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate is **glyceraldehyde-3-phosphate dehydrogenase**. This enzyme is one of a class of similar enzymes, the NADH-linked dehydrogenases. The structures of a number of dehydrogenases of this type have been studied via X-ray crystallography. The overall structures are not strikingly similar, but the structure of the binding site for NADH is quite similar in all these enzymes (Figure 17.8). (The oxidizing agent is  $\text{NAD}^+$ ; both oxidized and reduced forms of the coenzyme bind to the enzyme.) One portion of the binding site is specific for the nicotinamide ring, and one portion is specific for the adenine ring.

The molecule of glyceraldehyde-3-phosphate dehydrogenase is a tetramer, consisting of four identical subunits. Each subunit binds one molecule of  $\text{NAD}^+$ , and each subunit contains an essential cysteine residue. A thioester involving the cysteine residue is the key intermediate in this reaction. In the phosphorylation step, the thioester acts as a high-energy intermediate (see Chapter 15 for a discussion of thioesters). Phosphate ion attacks the thioester, forming a mixed anhydride of the carboxylic and phosphoric acids, which is also a high-energy compound (Figure 17.9). This compound is 1,3-bisphosphoglycerate, the product of the reaction. Production of ATP requires a high-energy compound as starting material. The 1,3-bisphosphoglycerate fulfills this requirement and transfers a phosphate group to ADP in a highly exergonic reaction (i.e., it has a high phosphate-group transfer potential).

**Step 7.** The next step is one of the two reactions in which ATP is produced by phosphorylation of ADP.



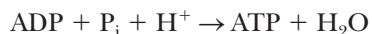




■ **FIGURE 17.9** The role of the active cysteine residue in glyceraldehyde-3-phosphate dehydrogenase. Phosphate ion attacks the thioester derivative of glyceraldehyde-3-phosphate dehydrogenase to produce 1,3-bisphosphoglycerate and to regenerate the thiol group of cysteine.

The enzyme that catalyzes this reaction is **phosphoglycerate kinase**. By now the term *kinase* should be familiar as the generic name for a class of ATP-dependent phosphate-group transfer enzymes. The most striking feature of the reaction has to do with energetics of the phosphate-group transfer. In this step in glycolysis, a phosphate group is transferred from 1,3-bisphosphoglycerate to a molecule of ADP, producing ATP, the first of two such reactions in the glycolytic pathway. We already mentioned that 1,3-bisphosphoglycerate can easily transfer a phosphate group to other substances. Note that a substrate, namely 1,3-bisphosphoglycerate, has transferred a phosphate group to ADP. This transfer is typical of **substrate-level phosphorylation**. It is to be distinguished from oxidative phosphorylation (Sections 20.1 through 20.5), in which transfer of phosphate groups is linked to electron-transfer reactions in which oxygen is the ultimate electron acceptor. The only requirement for substrate-level phosphorylation is that the standard free energy of the hydrolysis reaction is more negative than that for hydrolysis of the new phosphate compound being formed. Recall that the standard free energy of hydrolysis of 1,3-bisphosphoglycerate is  $-49.3 \text{ kJ mol}^{-1}$ . We have already seen that the standard free energy

of hydrolysis of ATP is  $-30.5 \text{ kJ mol}^{-1}$ , and we must change the sign of the free-energy change when the reverse reaction occurs:



$$\Delta G^\circ = 30.5 \text{ kJ mol}^{-1} = 7.3 \text{ kcal mol}^{-1}$$

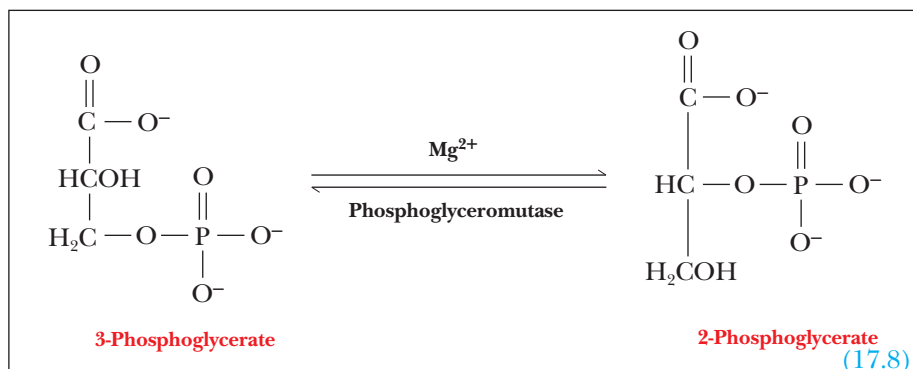
The net reaction is



$$\Delta G^\circ = -49.3 \text{ kJ mol}^{-1} + 30.5 \text{ kJ mol}^{-1} = -18.8 \text{ kJ mol}^{-1} = -4.5 \text{ kcal mol}^{-1}$$

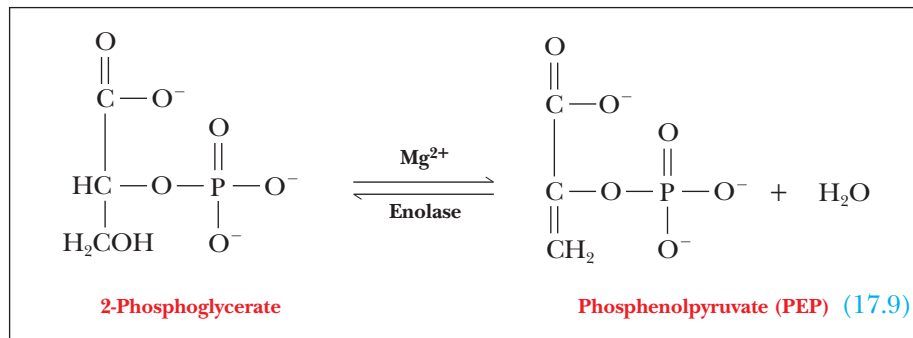
Two molecules of ATP are produced by this reaction for each molecule of glucose that enters the glycolytic pathway. In the earlier stages of the pathway, two molecules of ATP were invested to produce fructose-1,6-bisphosphate, and now they have been recovered. At this point, the balance of ATP use and production is exactly even. The next few reactions will bring about the production of two more molecules of ATP for each original molecule of glucose, leading to the net gain of two ATP molecules in glycolysis.

**Step 8.** The phosphate group is transferred from carbon 3 to carbon 2 of the glyceric acid backbone, setting the stage for the reaction that follows.

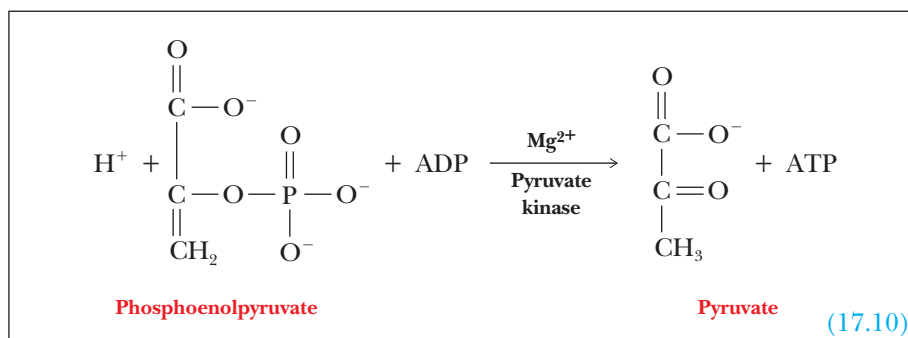


The enzyme that catalyzes this reaction is **phosphoglyceromutase**.

**Step 9.** The 2-phosphoglycerate molecule loses one molecule of water, producing phosphoenolpyruvate. This reaction does not involve electron transfer; it is a dehydration reaction. **Enolase**, the enzyme that catalyzes this reaction, requires  $\text{Mg}^{2+}$  as a cofactor. The water molecule that is eliminated binds to  $\text{Mg}^{2+}$  in the course of the reaction.



**Step 10.** Phosphoenolpyruvate transfers its phosphate group to ADP, producing ATP and pyruvate.



The double bond shifts to the oxygen on carbon 2 and a hydrogen shifts to carbon 3. Phosphoenolpyruvate is a high-energy compound with a high phosphate-group transfer potential. The free energy of hydrolysis of this compound is more negative than that of ATP ( $-61.9 \text{ kJ mol}^{-1}$  versus  $-30.5 \text{ kJ mol}^{-1}$ , or  $-14.8 \text{ kcal mol}^{-1}$  versus  $-7.3 \text{ kcal mol}^{-1}$ ). The reaction that occurs in this step can be considered to be the sum of the hydrolysis of phosphoenolpyruvate and the phosphorylation of ADP. This reaction is another example of substrate-level phosphorylation.

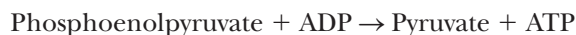


$$\Delta G^\circ = -61.9 \text{ kJ mol}^{-1} = -14.8 \text{ kcal mol}^{-1}$$



$$\Delta G^\circ = 30.5 \text{ kJ mol}^{-1} = 7.3 \text{ kcal mol}^{-1}$$

The net reaction is



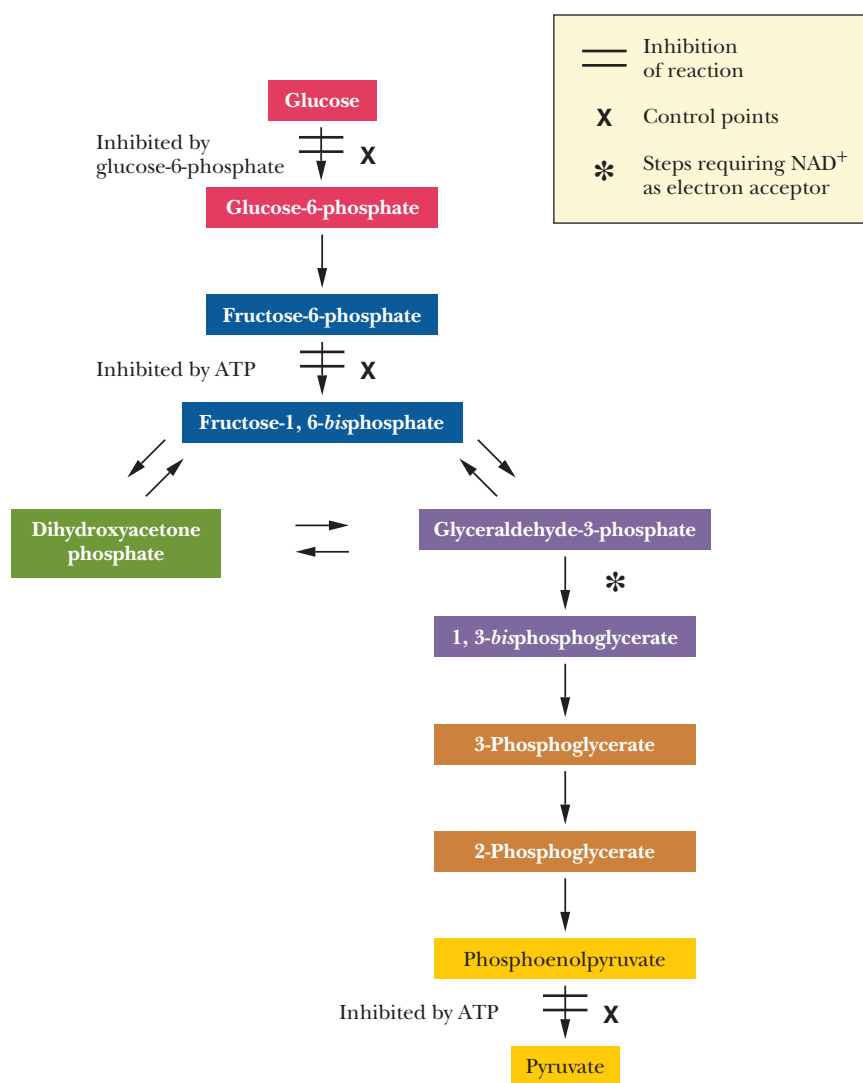
$$\Delta G^\circ = -31.4 \text{ kJ mol}^{-1} = -7.5 \text{ kcal mol}^{-1}$$

Since two moles of pyruvate are produced for each mole of glucose, twice as much energy is released for each mole of starting material.

**Pyruvate kinase** is the enzyme that catalyzes this reaction. Like phosphofructokinase, it is an allosteric enzyme consisting of four subunits of two different types (M and L), as we saw with phosphofructokinase. Pyruvate kinase is inhibited by ATP. The conversion of phosphoenolpyruvate to pyruvate slows down when the cell has a high concentration of ATP—that is to say, when the cell does not have a great need for energy in the form of ATP. Because of the different isozymes of pyruvate kinase found in liver versus muscle, the control of glycolysis is handled differently in these two tissues, which we will look at in detail in Chapter 18.

### Where are the control points in the glycolytic pathway?

One of the most important questions that we can ask about any metabolic pathway is, at which points is control exercised? Pathways can be “shut down” if an organism has no immediate need for their products, which saves energy for the organism. In glycolysis, three reactions are control points. The first is the reaction of glucose to glucose-6-phosphate, catalyzed by hexokinase; the second, which is the production of fructose-1,6-bisphosphate, is catalyzed by phosphofructokinase; and the last is the reaction of PEP to pyruvate, catalyzed by pyruvate kinase (Figure 17.10). It is frequently observed that control is exercised near the start and end of a pathway, as well as at points involving key intermediates such as fructose-1,6-bisphosphate. When we have learned more about the metabolism of carbohydrates, we can return to the role of phosphofructokinase and fructose-1,6-bisphosphate in the regulation of several pathways of carbohydrate metabolism (Section 18.3).

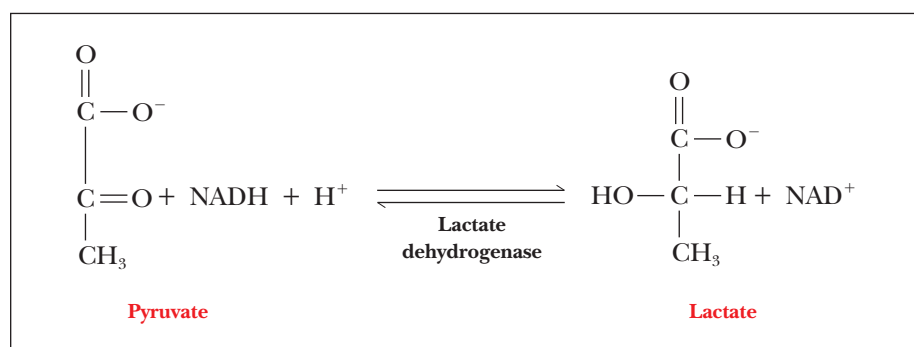


■ **FIGURE 17.10** Control points in glycolysis.

## 17.4 Anaerobic Metabolism of Pyruvate

### *How does the conversion of pyruvate to lactate take place in muscle?*

The final reaction of anaerobic glycolysis is the reduction of pyruvate to lactate.

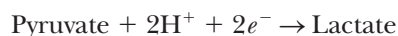


This reaction is also exergonic ( $\Delta G^\circ' = -25.1 \text{ kJ mol}^{-1} = -6.0 \text{ kcal mol}^{-1}$ ); as before, we need to multiply this value by 2 to find the energy yield for each

molecule of glucose that enters the pathway. Lactate is a dead end in muscle metabolism, but it can be recycled in the liver to form pyruvate and even glucose by a pathway called gluconeogenesis (“new synthesis of glucose”), which we will discuss in Section 18.2.

**Lactate dehydrogenase** (LDH) is the enzyme that catalyzes this reaction. Like glyceraldehyde-3-phosphate dehydrogenase, LDH is an NADH-linked dehydrogenase and consists of four subunits. There are two kinds of subunits, designated M and H, which vary slightly in amino acid composition. The quaternary structure of the tetramer can vary according to the relative amounts of the two kinds of subunits, yielding five possible isozymes. In human skeletal muscle, the homogeneous tetramer of the M<sub>4</sub> type predominates, and in the heart the other homogeneous possibility, the H<sub>4</sub> tetramer, is the predominant form. The heterogeneous forms—M<sub>3</sub>H, M<sub>2</sub>H<sub>2</sub>, and MH<sub>3</sub>—occur in blood serum. A very sensitive clinical test for heart disease is based on the existence of the various isozymic forms of this enzyme. The relative amounts of the H<sub>4</sub> and MH<sub>3</sub> isozymes in blood serum increase drastically after myocardial infarction (heart attack) compared with normal serum. The different isozymes have slightly different kinetic properties due to their subunit compositions. The H<sub>4</sub> isozyme (also called LDH 1) has a higher affinity for lactate as a substrate. The M<sub>4</sub> isozyme (LDH 5) is allosterically inhibited by pyruvate. These differences reflect the isozymes’ general roles in metabolism. The skeletal muscle is a highly anaerobic tissue, whereas the heart is not.

At this point, one might ask why the reduction of pyruvate to lactate (a waste product in aerobic organisms) is the last step in anaerobic glycolysis, a pathway that provides energy for the organism by oxidation of nutrients. There is another point to consider about the reaction, one that involves the relative amounts of NAD<sup>+</sup> and NADH in a cell. The half reaction of reduction can be written



and the half reaction of oxidation is

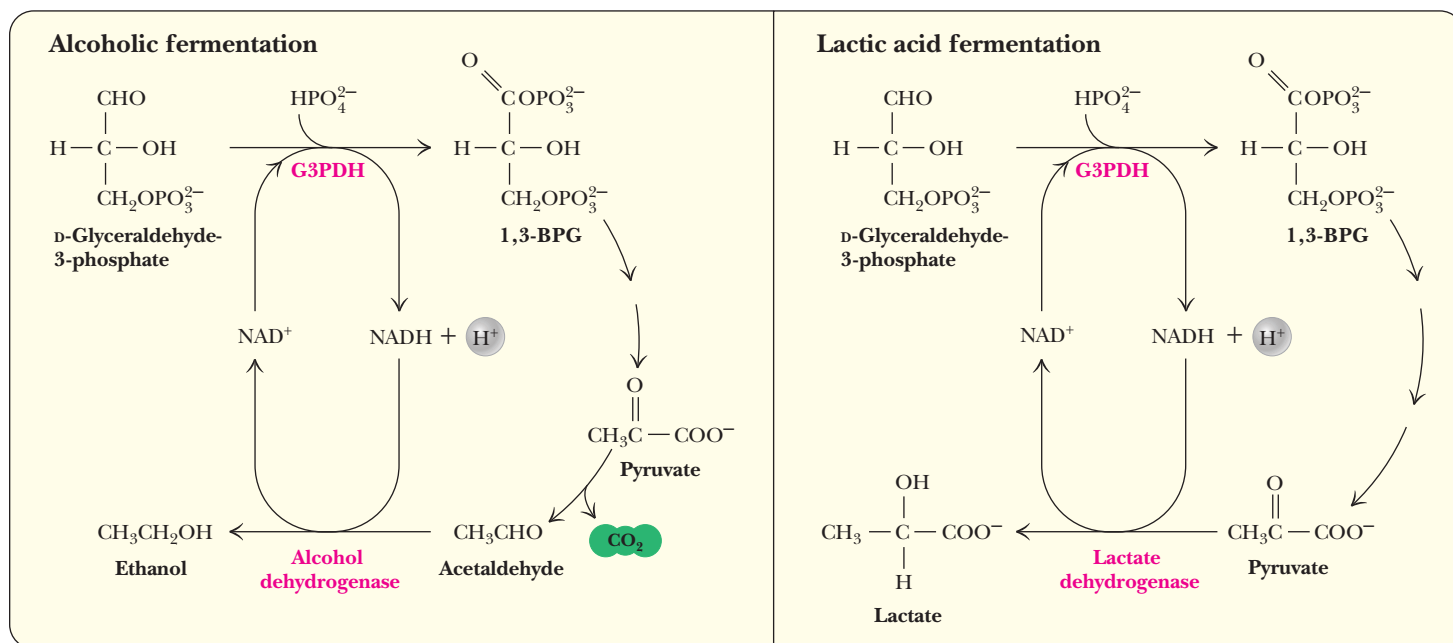


The overall reaction is, as we saw earlier,



The NADH produced from NAD<sup>+</sup> by the earlier oxidation of glyceraldehyde-3-phosphate is used up with no net change in the relative amounts of NADH and NAD<sup>+</sup> in the cell (Figure 17.11). This regeneration is needed under anaerobic conditions in the cell so that NAD<sup>+</sup> will be present for further glycolysis to take place. Without this regeneration, the oxidation reactions in anaerobic organisms would soon come to a halt because of the lack of NAD<sup>+</sup> to serve as an oxidizing agent in fermentative processes. The production of lactate buys time for the organism experiencing anaerobic metabolism and shifts some of the load away from the muscles and onto the liver, in which gluconeogenesis can reconvert lactate to pyruvate and glucose (Chapter 18). The same considerations apply in alcoholic fermentation (which will be discussed next). On the other hand, NADH is a frequently encountered reducing agent in many reactions, and it is lost to the organism in lactate production. Aerobic metabolism makes more efficient use of reducing agents (“reducing power”) such as NADH because the conversion of pyruvate to lactate does not occur in aerobic metabolism. The NADH produced in the stages of glycolysis leading to the production of pyruvate is available for use in other reactions in which a reducing agent is needed.



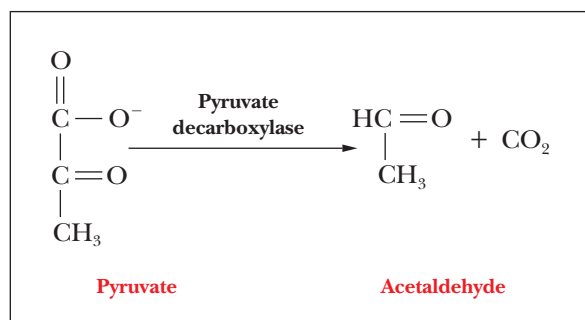


■ **FIGURE 17.11** The recycling of NAD<sup>+</sup> and NADH in anaerobic glycolysis.

### How does alcoholic fermentation take place?

Two other reactions related to the glycolytic pathway lead to the production of ethanol by *alcoholic fermentation*. This process is one of the alternative fates of pyruvate (Section 17.1). In the first of the two reactions that lead to the production of ethanol, pyruvate is decarboxylated (loses carbon dioxide) to produce acetaldehyde. The enzyme that catalyzes this reaction is *pyruvate decarboxylase*.

This enzyme requires Mg<sup>2+</sup> and a cofactor we have not seen before, **thiamine pyrophosphate** (TPP). (Thiamine itself is vitamin B<sub>1</sub>.) In TPP the carbon atom between the nitrogen and the sulfur in the thiazole ring (Figure 17.12) is highly reactive. It forms a carbanion (an ion with a negative charge on a carbon atom) quite easily, and the carbanion, in turn, attacks the carbonyl group of pyruvate to form an adduct. Carbon dioxide splits off, leaving a two-carbon fragment covalently bonded to TPP. There is a shift of electrons, and the two-carbon fragment splits off, producing acetaldehyde (Figure 17.13). The two-carbon fragment bonded to TPP is sometimes called activated acetaldehyde, and TPP can be found in several reactions that are decarboxylations.



The carbon dioxide produced is responsible for the bubbles in beer and in sparkling wines. Acetaldehyde is then reduced to produce ethanol, and, at the same time, one molecule of NADH is oxidized to NAD<sup>+</sup> for each molecule of ethanol produced.

## Biochemical Connections ALLIED HEALTH (DENTISTRY)

### What Is the Connection between Anaerobic Metabolism and Dental Plaque?

Dental caries, or tooth decay, is one of the most prevalent diseases in the United States and possibly in the world, although modern treatments such as fluoride and flossing have greatly reduced its



©Mark Thomas/Photo Researchers, Inc.

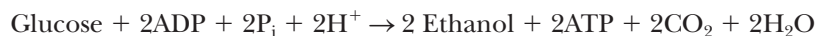
incidence in young people. Contributing factors in tooth decay are a combination of a diet high in refined sugars, the development of dental plaque, and anaerobic metabolism.

The high-sugar diet allows for rapid growth of bacteria in the mouth, and sucrose is perhaps the most efficiently used sugar because the bacteria can make their polysaccharide “glue” more efficiently from this nonreducing sugar. The bacteria grow in expanding sticky colonies, forming plaque on the tooth surface. The bacteria growing under the surface of the plaque must utilize anaerobic metabolism because oxygen does not diffuse readily through the waxy surface of dental plaque. The two predominant by-products, lactate and pyruvate, are relatively strong organic acids, and these acid products actually destroy the enamel surface. The bacteria, of course, grow rapidly in the pock holes. If the enamel is eaten all the way through, the bacteria grow even more readily in the softer dentin layer beneath the enamel.

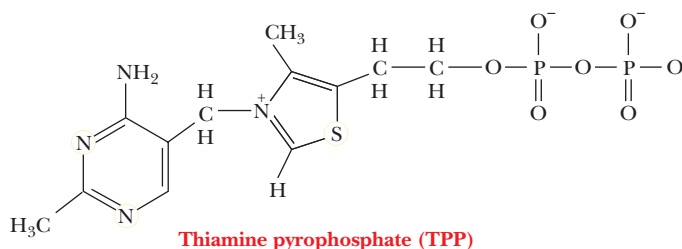
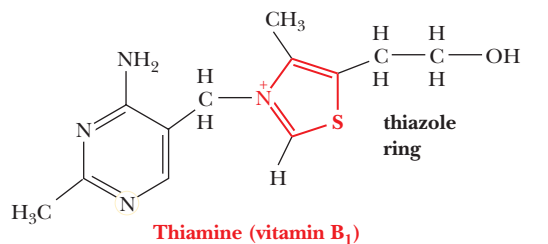
Fluoridation results in a much harder enamel surface, and the fluoride may inhibit the metabolism of the bacteria. Daily flossing disrupts the plaque, and the anaerobic conditions never get started.



The reduction reaction of alcoholic fermentation is similar to the reduction of pyruvate to lactate, in the sense that it provides for recycling of  $\text{NAD}^+$  and thus allows further anaerobic oxidation (fermentation) reactions. The net reaction for alcoholic fermentation is



$\text{NAD}^+$  and  $\text{NADH}$  do not appear explicitly in the net equation. It is essential that the recycling of  $\text{NADH}$  to  $\text{NAD}^+$  takes place here, just as it does when lactate is produced, so that there can be further anaerobic oxidation. **Alcohol dehydrogenase**, the enzyme that catalyzes the conversion of acetaldehyde to ethanol, is similar to lactate dehydrogenase in many ways. The most striking similarity is that both are  $\text{NADH}$ -linked dehydrogenases, and both are tetramers.

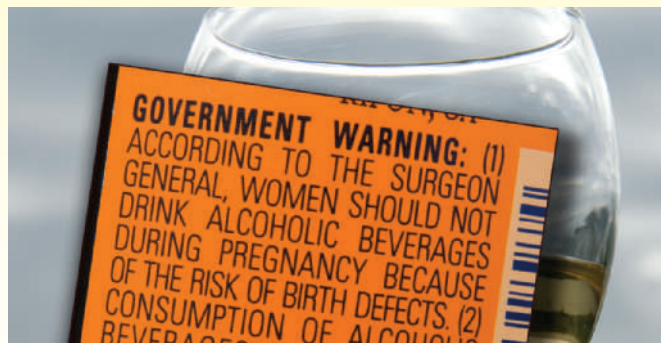


■ **FIGURE 17.12** The structures of thiamine (vitamin B<sub>1</sub>) and thiamine pyrophosphate (TPP), the active form of the coenzyme.

# Biochemical Connections ALLIED HEALTH

## Fetal Alcohol Syndrome

The variety of injuries to a fetus caused by maternal consumption of ethanol is called fetal alcohol syndrome. In catabolism of ethanol by the body, the first step is conversion to acetaldehyde—the

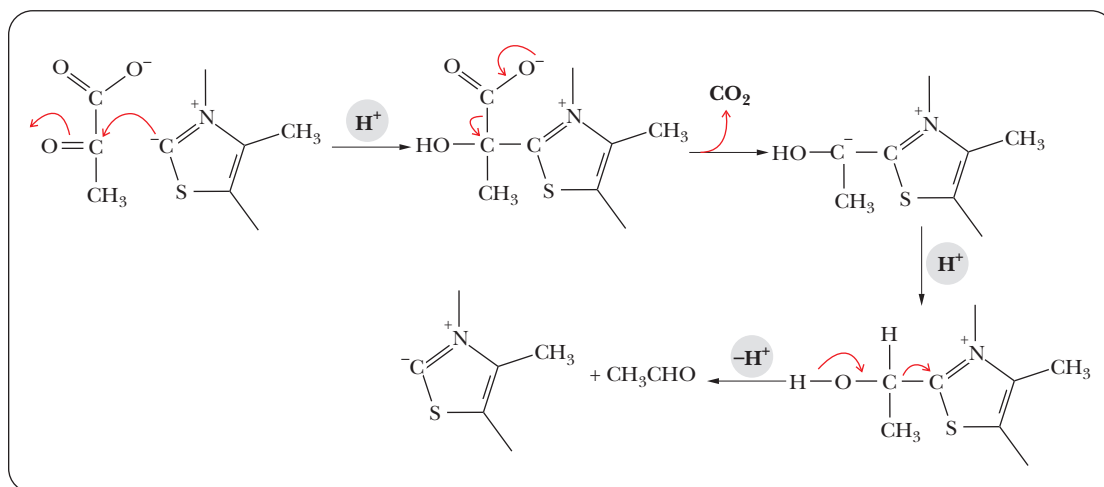


© Cengage Learning/Charles D. Winters

reverse of the last reaction of alcoholic fermentation. The level of acetaldehyde in the blood of a pregnant woman is the key to detecting fetal alcohol syndrome. It has been shown that the acetaldehyde is transferred across the placenta and that it accumulates in the liver of the fetus. Acetaldehyde is toxic, and this is one of the most important factors in fetal alcohol syndrome.

In addition to the toxic effects of acetaldehyde, consumption of ethanol during pregnancy harms the fetus in other ways. It depresses transfer of nutrients to the fetus, resulting in lower levels of sugars (hypoglycemia), vitamins, and essential amino acids. Lower levels of oxygen (hypoxia) also occur. This last effect is more drastic when the mother smokes during pregnancy, as well as consuming alcohol.

The labels of alcoholic beverages now include a warning against consumption during pregnancy. The American Medical Association has issued the unequivocal warning that “there is no known safe level of alcohol during pregnancy.”



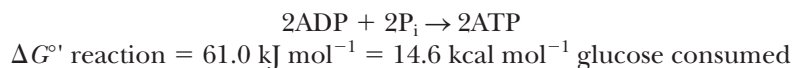
■ **FIGURE 17.13** The mechanism of the pyruvate decarboxylase reaction. The carbanion form of the thiazole ring of TPP is strongly nucleophilic. The carbanion attacks the carbonyl carbon of pyruvate to form an adduct. Carbon dioxide splits out, leaving a two-carbon fragment (activated acetaldehyde) covalently bonded to the coenzyme. A shift of electrons releases acetaldehyde, regenerating the carbanion.

## 17.5 Energy Production in Glycolysis

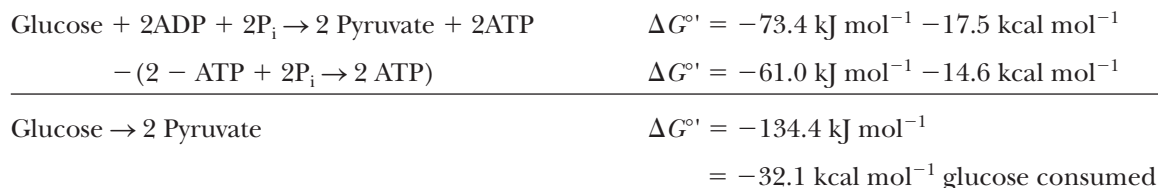
### What is the energy yield from glycolysis?

Now that we have seen the reactions of the glycolytic pathway, we can do some bookkeeping and determine the standard free-energy change for the entire pathway by using the data from Table 17.1.

The overall process of glycolysis is exergonic. We can calculate  $\Delta G^\circ$  for the entire reaction by adding up the  $\Delta G^\circ$  values from each of the steps. Remember that all of the reactions from triose phosphate isomerase to pyruvate kinase are doubled. This gives a final figure from glucose to two pyruvates of  $-74.0 \text{ kJ mol}^{-1}$  or  $-17.5 \text{ kcal mol}^{-1}$ . The energy released in the exergonic phases of the process drives the endergonic reactions. The net reaction of glycolysis explicitly includes an important endergonic process, that of phosphorylation of two molecules of ADP.



Without the production of ATP, the reaction of one molecule of glucose to produce two molecules of pyruvate would be even more exergonic. Thus, *subtracting* out the synthesis of ATP:



(The corresponding figure for the conversion of one mole of glucose to two moles of lactate is  $-184.6 \text{ kJ mol}^{-1} = -44.1 \text{ kcal mol}^{-1}$ .) Without production of ATP, the energy released by the conversion of glucose to pyruvate would be lost to the organism and dissipated as heat. The energy required to produce the two molecules of ATP for each molecule of glucose can be recovered by the organism when the ATP is hydrolyzed in some metabolic process. We discussed this point briefly in Chapter 15, when we compared the thermodynamic efficiency of anaerobic and aerobic metabolism. The percentage of the energy released by the breakdown of glucose to lactate that is “captured” by the organism when ADP is phosphorylated to ATP is the efficiency of energy use in glycolysis; it is  $(61.0/184.6) \times 100$ , or about 33%. Recall this percentage from Section 15.10. It comes from calculating the energy used to phosphorylate two moles of ATP as a percentage of the energy released by the conversion of one mole of glucose to two moles of lactate. The net release of energy in glycolysis,  $123.6 \text{ kJ}$  ( $29.5 \text{ kcal}$ ) for each mole of glucose converted to lactate, is dissipated as heat by the organism. Without the production of ATP to serve as a source of energy for other metabolic processes, the energy released by glycolysis would serve no purpose for the organism, except to help maintain body temperature in warm-blooded animals. A soft drink with ice can help keep you warm even on the coldest day of winter (if it is not a diet drink) because of its high sugar content.

The free-energy changes we have listed in this section are the standard values, assuming the standard conditions, such as  $1 \text{ M}$  concentrations of all solutes except hydrogen ion. Concentrations under physiological conditions can differ markedly from standard values. Fortunately, there are well-known methods (Section 15.3) for calculating the difference in the free-energy change. Also, large changes in concentrations frequently lead to relatively small differences in the free-energy change, about a few kilojoules per mole. Some of the free-energy changes may be different under physiological conditions from the values listed here for standard conditions, but the underlying principles and the conclusions drawn from them remain the same.

## SUMMARY

**What are the possible fates of pyruvate in glycolysis?** In glycolysis, one molecule of glucose gives rise, after a long series of reactions, to two molecules of pyruvate. Along the way, two net molecules of ATP and NADH are produced. In aerobic metabolism, pyruvate is further oxidized to carbon dioxide and water. In anaerobic metabolism, the product is lactate or, in organisms capable of alcoholic fermentation, it is ethanol.

**What are the reactions of glycolysis?** A series of 10 reactions converts one molecule of glucose to two molecules of

pyruvate. Four of these reactions transfer a phosphate group, three are isomerizations, one is a cleavage, one a dehydration, and only one an oxidation.

**What reactions convert glucose-6-phosphate to glyceraldehyde-3-phosphate?** In the first half of glycolysis, glucose is phosphorylated to glucose-6-phosphate, using an ATP in the process. Glucose-6-phosphate is isomerized to fructose-6-phosphate, which is then phosphorylated again to fructose-1,6-bisphosphate, utilizing another ATP. Fructose-1,6-bisphosphate

is a key intermediate, and the enzyme that catalyzes its formation, phosphofructokinase, is an important controlling factor in the pathway. Fructose 1,6-bisphosphate is then split into two three-carbon compounds, glyceraldehyde-3-phosphate and dihydroxyacetone phosphate, the latter of which is then also converted to glyceraldehyde-3-phosphate. The overall reaction in the first half of the pathway is the conversion of one molecule of glucose into two molecules of glyceraldehyde-3-phosphate at the expense of two molecules of ATP.

**What reactions convert glyceraldehyde-3-phosphate to pyruvate?** Glyceraldehyde-3-phosphate is oxidized to 1,3-bisphosphoglycerate and  $\text{NAD}^+$  is reduced to NADH. The 1,3-bisphosphoglycerate is then converted to 3-phosphoglycerate and ATP is produced. 3-Phosphoglycerate is converted in two steps to phosphoenolpyruvate, an important high-energy compound. Phosphoenolpyruvate is then converted to pyruvate and ATP is produced. The overall reaction of the second half of the pathway is that two molecules of glyceraldehyde-3-phosphate are converted to two molecules of pyruvate, and four molecules of ATP are produced.

**Where are the control points in the glycolytic pathway?** There are three control points in the glycolytic pathway. The first is at the beginning, where glucose is converted to glucose-6-phosphate. The second is at the committed step, the production of fructose 1,6-bisphosphate. The third is the conversion of phosphoenolpyruvate to pyruvate.

**How does the conversion of pyruvate to lactate take place in muscle?** There are two fates for pyruvate in anaerobic metabolism. The common outcome is reduction to lactate, catalyzed by the enzyme lactate dehydrogenase.  $\text{NAD}^+$  is recycled in the process.

**How does alcoholic fermentation take place?** In organisms capable of alcoholic fermentation, pyruvate loses carbon dioxide to produce acetaldehyde, which, in turn, is reduced to produce ethanol. Thiamine pyrophosphate is a coenzyme needed for the process.

**What is the energy yield from glycolysis?** In each of two reactions in the pathway, one molecule of ATP is hydrolyzed for each molecule of glucose metabolized. In each of two other reactions, two molecules of ATP are produced by phosphorylation of ADP for each molecule of glucose, giving a total of four ATP molecules produced. There is a net gain of two ATP molecules for each molecule of glucose processed in glycolysis. The anaerobic breakdown of glucose to lactate can be summarized as follows:



The overall process of glycolysis is exergonic.

Reaction	$\Delta G^\circ$	
	$\text{kJ mol}^{-1}$	$\text{kcal mol}^{-1}$
Glucose + 2ADP + 2P <sub>i</sub> → 2 Pyruvate + 2ATP	−73.3	−17.5
2(Pyruvate + NADH + H <sup>+</sup> → Lactate + NAD <sup>+</sup> )	−50.2	−12.0
Glucose + 2ADP + 2P <sub>i</sub> → 2 Lactate + 2ATP	−123.5	−29.5

Without production of ATP, glycolysis would be still more exergonic, but the energy released would be lost to the organism and dissipated as heat.

## REVIEW EXERCISES

**OWL** Interactive versions of these problems are assignable in OWL

### 17.1 The Overall Pathway of Glycolysis

- Recall** Which reaction or reactions that we have met in this chapter require ATP? Which reaction or reactions produce ATP? List the enzymes that catalyze the reactions that require and that produce ATP.
- Recall** Which reaction or reactions that we have met in this chapter require NADH? Which reaction or reactions require  $\text{NAD}^+$ ? List the enzymes that catalyze the reactions that require NADH and that require  $\text{NAD}^+$ .
- Recall** What are the possible metabolic fates of pyruvate?

### 17.2 Conversion of Six-Carbon Glucose to Three-Carbon Glyceraldehyde-3-Phosphate

- Recall** Explain the origin of the name of the enzyme aldolase.
- Recall** Define *isozymes* and give an example from the material discussed in this chapter.
- Recall** Why would enzymes be found as isozymes?
- Recall** Why is the formation of fructose-1,6-bisphosphate the committed step in glycolysis?
- Reflect and Apply** Show that the reaction  $\text{Glucose} \rightarrow 2 \text{ Glyceraldehyde-3-phosphate}$  is slightly endergonic ( $\Delta G^\circ = 2.2 \text{ kJ mol}^{-1} = 0.53 \text{ kcal mol}^{-1}$ ); that is, it is not too far from equilibrium. Use the data in Table 17.1.

- Reflect and Apply** What is the metabolic advantage of having both hexokinase and glucokinase to phosphorylate glucose?
- Reflect and Apply** What are the metabolic effects of not being able to produce the M subunit of phosphofructokinase?
- Reflect and Apply** In what way is the observed mode of action of hexokinase consistent with the induced-fit theory of enzyme action?
- Reflect and Apply** How does ATP act as an allosteric effector in the mode of action of phosphofructokinase?

### 17.3 Glyceraldehyde-3-Phosphate Is Converted to Pyruvate

- Recall** At what point in glycolysis are all the reactions considered doubled?
- Recall** Which of the enzymes discussed in this chapter are NADH-linked dehydrogenases?
- Recall** Define *substrate-level phosphorylation* and give an example from the reactions discussed in this chapter.
- Recall** Which reactions are the control points in glycolysis?
- Recall** Which molecules act as inhibitors of glycolysis? Which molecules act as activators?
- Recall** Many NADH-linked dehydrogenases have similar active sites. Which part of glyceraldehyde-3-phosphate dehydrogenase would be the most conserved between other enzymes?



19. **Recall** Several of the enzymes of glycolysis fall into classes that we will see often in metabolism. What reaction types are catalyzed by each of the following:
  - (a) Kinases
  - (b) Isomerases
  - (c) Aldolases
  - (d) Dehydrogenases
20. **Recall** What is the difference between an isomerase and a mutase?
21. **Reflect and Apply** Is the reaction of 2-phosphoglycerate to phosphoenolpyruvate a redox reaction? Give the reason for your answer.
22. **Reflect and Apply** Show the carbon atom that changes oxidation state during the reaction catalyzed by glyceraldehyde-3-phosphate dehydrogenase. What is the functional group that changes during the reaction?
23. **Reflect and Apply** Discuss the logic of the nature of the allosteric inhibitors and activators of glycolysis. Why would these molecules be used?
24. **Reflect and Apply** Many species have a third type of LDH subunit that is found predominantly in the testes. If this subunit, called C, were expressed in other tissues and could combine with the M and H subunits, how many LDH isozymes would be possible? What would their compositions be?
25. **Reflect and Apply** The M and H subunits of lactate dehydrogenase have very similar sizes and shapes but differ in amino acid composition. If the only difference between the two were that the H subunit had a glutamic acid in a position where the M subunit had a serine, how would the five isozymes of LDH separate on electrophoresis using a gel at pH 8.6? (See Chapter 5 for details on electrophoresis.)
26. **Reflect and Apply** Why is the formation of fructose-1,6-bisphosphate a step in which control is likely to be exercised in the glycolytic pathway?
27. **Reflect and Apply** High levels of glucose-6-phosphate inhibit glycolysis. If the concentration of glucose-6-phosphate decreases, activity is restored. Why?
28. **Reflect and Apply** Most metabolic pathways are relatively long and appear to be very complex. For example, there are 10 individual chemical reactions in glycolysis, converting glucose to pyruvate. Suggest a reason for the complexity.
29. **Reflect and Apply** The mechanism involved in the reaction catalyzed by phosphoglyceromutase is known to involve a phosphorylated enzyme intermediate. If 3-phosphoglycerate is radioactively labeled with  $^{32}\text{P}$ , the product of the reaction, 2-phosphoglycerate, does not have any radioactive label. Design a mechanism to explain these facts.

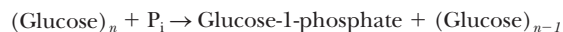
### 17.4 Anaerobic Metabolism of Pyruvate

30. **Recall** What does the material of this chapter have to do with beer? What does it have to do with tired and aching muscles?
31. **Recall** If lactic acid is the buildup product of strenuous muscle activity, why is sodium lactate often given to hospital patients intravenously?
32. **Recall** What is the metabolic purpose of lactic acid production?
33. **Reflect and Apply** Using the Lewis electron-dot notation, show explicitly the transfer of electrons in the following redox reactions.
  - (a)  $\text{Pyruvate} + \text{NADH} + \text{H}^+ \rightarrow \text{Lactate} + \text{NAD}^+$
  - (b)  $\text{Acetaldehyde} + \text{NADH} + \text{H}^+ \rightarrow \text{Ethanol} + \text{NAD}^+$
  - (c)  $\text{Glyceraldehyde-3-phosphate} + \text{NAD}^+ \rightarrow \text{3-Phosphoglycerate} + \text{NADH} + \text{H}^+$  (redox reaction only)

34. **Reflect and Apply** Briefly discuss the role of thiamine pyrophosphate in enzymatic reactions, using material from this chapter to illustrate your points.
35. **Reflect and Apply** What is unique about TPP that makes it useful in decarboxylation reactions?
36. **Biochemical Connections** Beriberi is a disease caused by a deficiency of vitamin B<sub>1</sub> (thiamine) in the diet. Thiamine is the precursor of thiamine pyrophosphate. In view of what you have learned in this chapter, why is it not surprising that alcoholics tend to develop this disease?
37. **Reflect and Apply** Most hunters know that meat from animals that have been run to death tastes sour. Suggest a reason for this observation.
38. **Reflect and Apply** What is the metabolic advantage in the conversion of glucose to lactate, in which there is no *net* oxidation or reduction?
39. **Biochemical Connections** Cancer cells grow so rapidly that they have a higher rate of anaerobic metabolism than most body tissues, especially at the center of a tumor. Can you use drugs that poison the enzymes of anaerobic metabolism in the treatment of cancer? Why, or why not?

### 17.5 Energy Production in Glycolysis

40. **Reflect and Apply** Show how the estimate of 33% efficiency of energy use in anaerobic glycolysis is derived.
41. **Recall** What is the net gain of ATP molecules derived from the reactions of glycolysis?
42. **Recall** How does the result in Question 41 differ from the gross yield of ATP?
43. **Recall** Which reactions in glycolysis are coupled reactions?
44. **Recall** Which steps in glycolysis are physiologically irreversible?
45. **Reflect and Apply** Show, by a series of equations, the energetics of phosphorylation of ADP by phosphoenolpyruvate.
46. **Reflect and Apply** What should be the net ATP yield for glycolysis when fructose, mannose, and galactose are used as the starting compounds? Justify your answer.
47. **Reflect and Apply** In the muscles, glycogen is broken down via the following reaction:



What would be the ATP yield per molecule of glucose in the muscle if glycogen were the source of the glucose?

48. **Reflect and Apply** Using Table 17.1, predict whether the following reaction is thermodynamically possible:
 
$$\text{Phosphoenolpyruvate} + \text{P}_i + 2\text{ADP} \rightarrow \text{Pyruvate} + 2\text{ATP}$$
49. **Reflect and Apply** Does the reaction shown in Question 48 occur in nature? If not, why not?
50. **Reflect and Apply** According to Table 17.1, several reactions have very positive  $\Delta G^\circ$  values. How can this be explained, given that these reactions do occur in the cell?
51. **Reflect and Apply** According to Table 17.1, four reactions have positive  $\Delta G$  values. How can this be explained?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).

# Storage Mechanisms and Control in Carbohydrate Metabolism

# 18

Control of carbohydrate metabolism is important in physical activity of all sorts.

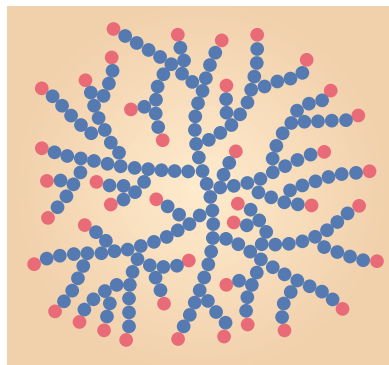
## 18.1 How Glycogen Is Produced and Degraded

When we digest a meal high in carbohydrates, we have a supply of glucose that exceeds our immediate needs. We store glucose as a polymer, glycogen (Section 16.4), that is similar to the starches found in plants; glycogen differs from starch only in the degree of chain branching. In fact, glycogen is sometimes called “animal starch” because of this similarity. A look at the metabolism of glycogen will give us some insights into how glucose can be stored in this form and made available on demand. In the degradation of glycogen, several glucose residues can be released simultaneously, one from each end of a branch, rather than one at a time as would be the case in a linear polymer.

This feature is useful to an organism in meeting short-term demands for energy by increasing the glucose supply as quickly as possible (Figure 18.1). Mathematical modeling has shown that the structure of glycogen is *optimized* for its ability to store and deliver energy quickly and for the longest amount of time possible. The key to this optimization is the average chain length of the branches (13 residues). If the average chain length were much greater or much shorter, glycogen would not be as efficient a vehicle for energy storage and release on demand. Experimental results support the conclusions reached from the mathematical modeling.

### How does the breakdown of glycogen take place?

Glycogen is found primarily in liver and muscle. The release of glycogen stored in the liver is triggered by low levels of glucose in blood. Liver glycogen breaks down to glucose-6-phosphate, which is hydrolyzed to give glucose. The release of glucose from the liver by this breakdown of glycogen replenishes the supply



Glycogen

■ **FIGURE 18.1** The branched-chain structure of glycogen. The highly branched structure of glycogen makes it possible for several glucose residues to be released at once to meet energy needs. This would not be possible with a linear polymer. The red dots indicate the terminal glucose residues that are released from glycogen. The more branch points there are, the more of these terminal residues are available at one time.

## Chapter Outline

### 18.1 How Glycogen Is Produced and Degraded

- How does the breakdown of glycogen take place?
- How is glycogen formed from glucose?
- How is glycogen metabolism controlled?

### 18.2 Gluconeogenesis Produces Glucose from Pyruvate

- Why is oxaloacetate an intermediate in gluconeogenesis?
- What is the role of sugar phosphates in gluconeogenesis?

### 18.3 Control of Carbohydrate Metabolism

- How does control of key enzymes control carbohydrate metabolism?
- How do different organs share carbohydrate metabolism?
- What roles do the first and last steps of glycolysis play in control of carbohydrate metabolism?

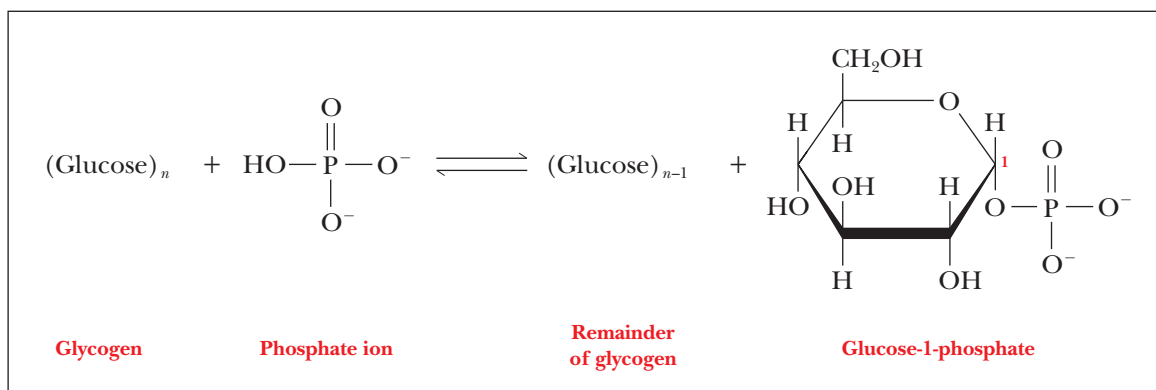
### 18.4 Glucose Is Sometimes Diverted through the Pentose Phosphate Pathway

- What are the oxidative reactions of the pentose phosphate pathway?
- What are the nonoxidative reactions of the pentose phosphate pathway, and why are they important?
- How is the pentose phosphate pathway controlled?

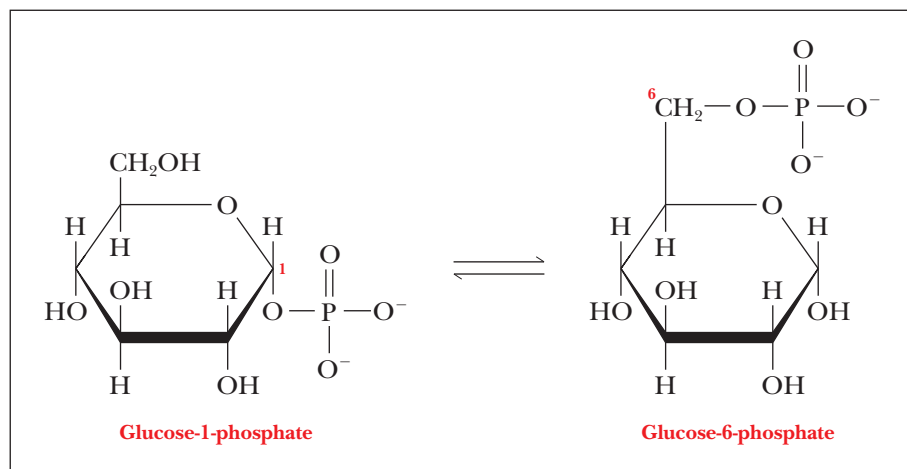
**OWL** Online homework for this chapter may be assigned in OWL.

of glucose in the blood. In muscle, glucose-6-phosphate obtained from glycogen breakdown enters the glycolytic pathway directly rather than being hydrolyzed to glucose and then exported to the bloodstream.

Three reactions play roles in the conversion of glycogen to glucose-6-phosphate. In the first reaction, each glucose residue cleaved from glycogen reacts with phosphate to give glucose-1-phosphate. Note particularly that this cleavage reaction is one of **phosphorolysis** rather than hydrolysis.



In a second reaction, glucose-1-phosphate isomerizes to give glucose-6-phosphate.



Complete breakdown of glycogen also requires a debranching reaction to hydrolyze the glycosidic bonds of the glucose residues at branch points in the glycogen structure. The enzyme that catalyzes the first of these reactions is **glycogen phosphorylase**; the second reaction is catalyzed by **phosphoglucomutase**.

Glycogen phosphorylase



Phosphoglucomutase



Glycogen phosphorylase cleaves the  $\alpha(1 \rightarrow 4)$  linkages in glycogen. Complete breakdown requires **debranching enzymes** that degrade the  $\alpha(1 \rightarrow 6)$  linkages. Note that no ATP is hydrolyzed in the first reaction. In the glycolytic pathway,

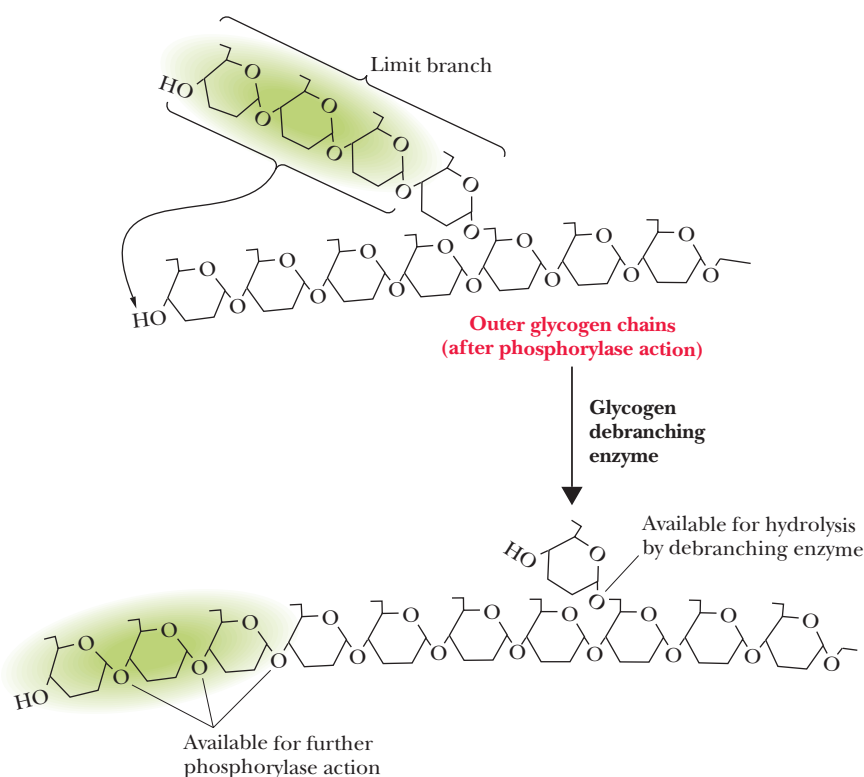
we saw another example of phosphorylation of a substrate directly by phosphate without involvement of ATP: the phosphorylation of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate. This is an alternative mode of entry to the glycolytic pathway that “saves” one molecule of ATP for each molecule of glucose because it bypasses the first step in glycolysis. When glycogen rather than glucose is the starting material for glycolysis, there is a net gain of three ATP molecules for each glucose monomer, rather than two ATP molecules, as when glucose itself is the starting point. Thus, glycogen is a more effective energy source than glucose. Of course, there is no “free lunch” in biochemistry and, as we shall see, it takes energy to put the glucoses together into glycogen.

The debranching of glycogen involves the transfer of a “limit branch” of three glucose residues to the end of another branch, where they are subsequently removed by glycogen phosphorylase. The same glycogen debranching enzyme then hydrolyzes the  $\alpha(1 \rightarrow 6)$  glycosidic bond of the last glucose residue remaining at the branch point (Figure 18.2).

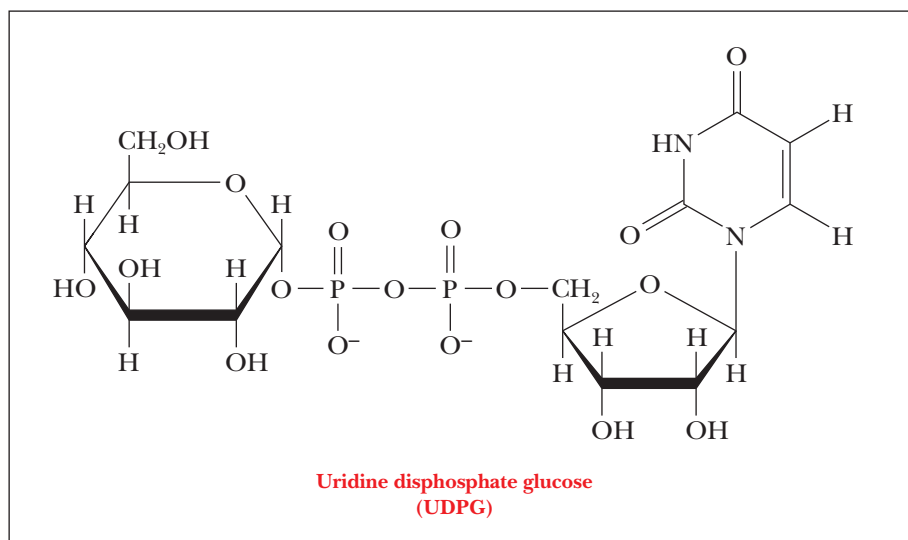
When an organism needs energy quickly, glycogen breakdown is important. Muscle tissue can mobilize glycogen more easily than fat and can do so anaerobically. With low-intensity exercise, such as jogging or long-distance running, fat is the preferred fuel, but as the intensity increases, muscle and liver glycogen becomes more important. Some athletes, particularly middle-distance runners and cyclists, try to build up their glycogen reserves before a race by eating large amounts of carbohydrates. The Biochemical Connections box on page 514 goes into more detail on this subject.

### How is glycogen formed from glucose?

The formation of glycogen from glucose is not the exact reversal of the breakdown of glycogen to glucose. The synthesis of glycogen requires energy, which is provided by the hydrolysis of a nucleoside triphosphate, UTP. In the first stage of glycogen synthesis, glucose-1-phosphate (obtained from glucose-6-phosphate by an isomerization reaction) reacts with UTP to produce uridine diphosphate glucose (also called UDP-glucose or UDPG) and pyrophosphate ( $PP_i$ ).

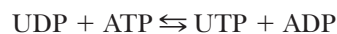


■ **FIGURE 18.2** The mode of action of the debranching enzyme in glycogen breakdown. The enzyme transfers three  $\alpha(1 \rightarrow 4)$ -linked glucose residues from a limit branch to the end of another branch. The same enzyme also catalyzes the hydrolysis of the  $\alpha(1 \rightarrow 6)$ -linked residue at the branch point.



The enzyme that catalyzes this reaction is *UDP-glucose pyrophosphorylase*. The exchange of one phosphoric anhydride bond for another has a free-energy change close to zero. The release of energy comes about when the enzyme inorganic pyrophosphatase catalyzes the hydrolysis of pyrophosphate to two phosphates, a strongly exergonic reaction.

It is common in biochemistry to see the energy released by the hydrolysis of pyrophosphate combined with the free energy of hydrolysis of a nucleoside triphosphate. The coupling of these two exergonic reactions to a reaction that is not energetically favorable allows an otherwise endergonic reaction to take place. The supply of UTP is replenished by an exchange reaction with ATP, which is catalyzed by nucleoside phosphate kinase:



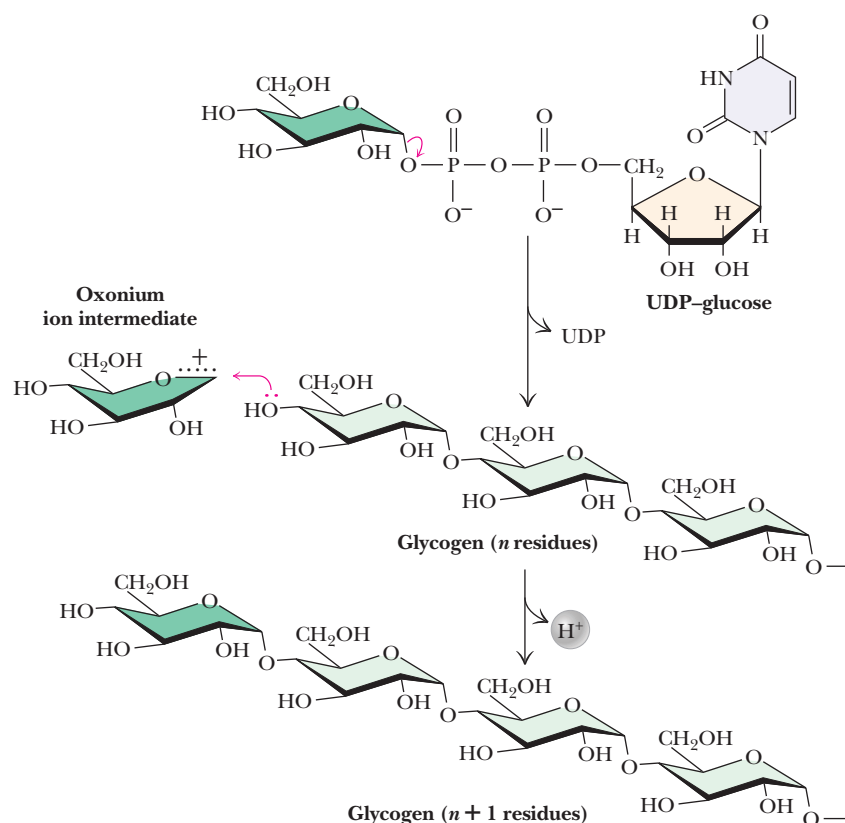
This exchange reaction makes the hydrolysis of any nucleoside triphosphate energetically equivalent to the hydrolysis of ATP.

The addition of UDPG to a growing chain of glycogen is the next step in glycogen synthesis. Each step involves formation of a new  $\alpha(1 \rightarrow 4)$  glycosidic bond in a reaction catalyzed by the enzyme **glycogen synthase** (Figure 18.3). This enzyme cannot simply form a bond between two isolated glucose molecules; it must add to an existing chain with  $\alpha(1 \rightarrow 4)$  glycosidic linkages. The initiation of glycogen synthesis requires a primer for this reason. The hydroxyl group of a specific tyrosine of the protein *glycogenin* (37,300 Da) serves this purpose. In the first stage of glycogen synthesis, a glucose residue is linked to this tyrosine hydroxyl, and glucose residues are successively added to this first one. The glycogenin molecule itself acts as the catalyst for addition of glucoses until there are about eight of them linked together. At that point, glycogen synthase takes over.

	$\Delta G^\circ$	
	$\text{kJ mol}^{-1}$	$\text{kcal mol}^{-1}$
Glucose-1-phosphate + UTP $\rightleftharpoons$ UDPG + $\text{PP}_i$	$\sim 0$	$\sim 0$
$\text{H}_2\text{O} + \text{PP}_i \rightarrow 2\text{P}_i$	$-30.5$	$-7.3$
Overall Glucose-1-phosphate + UTP $\rightarrow$ UDPG + $2\text{P}_i$	$-30.5$	$-7.3$

Synthesis of glycogen requires the formation of  $\alpha(1 \rightarrow 6)$  as well as  $\alpha(1 \rightarrow 4)$  glycosidic linkages. A **branching enzyme** accomplishes this task. It does so by transferring a segment about 7 residues long from the end of a growing chain





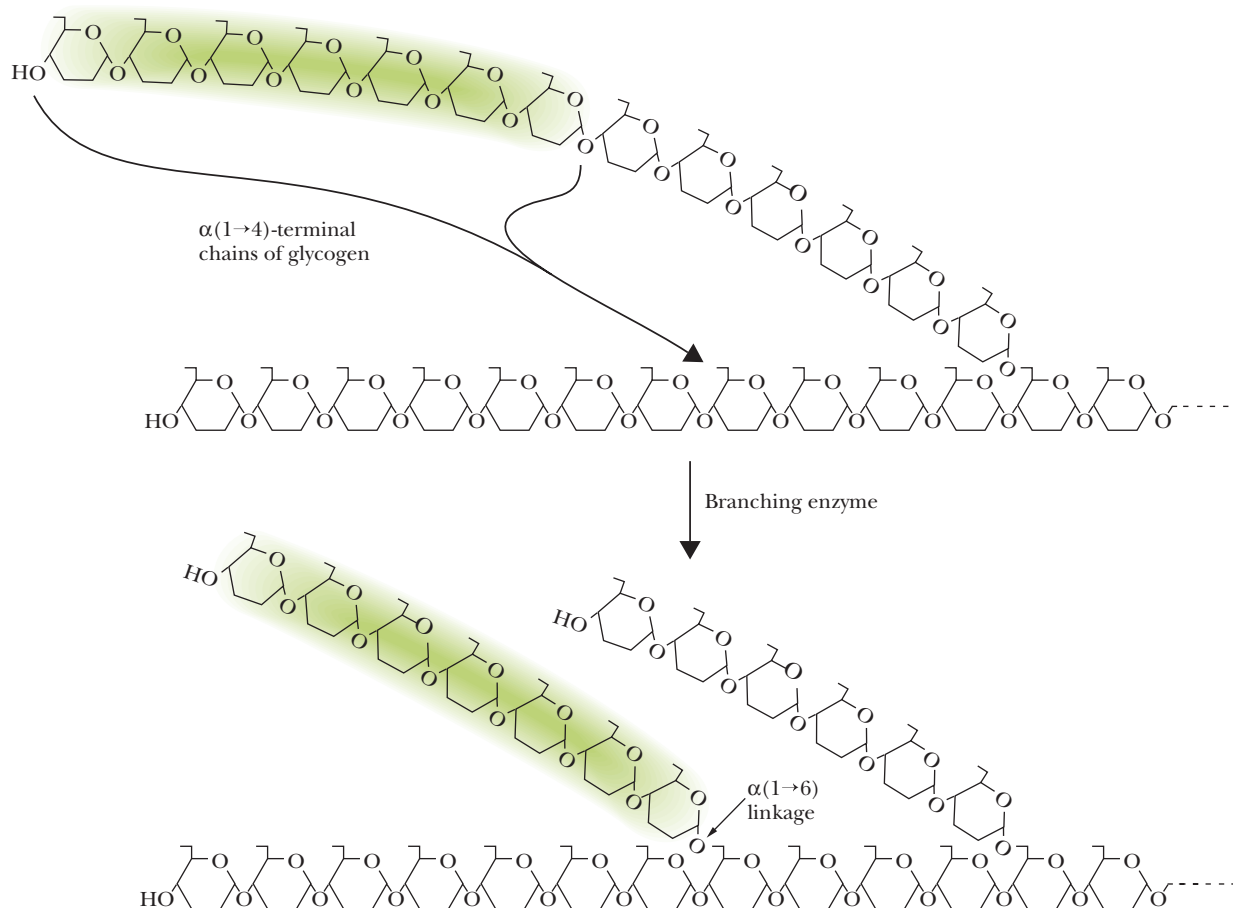
■ **FIGURE 18.3** The reaction catalyzed by glycogen synthase. A glucose residue is transferred from UDPG to the growing end of a glycogen chain in an  $\alpha(1 \rightarrow 4)$  linkage.

to a branch point where it catalyzes the formation of the required  $\alpha(1 \rightarrow 6)$  glycosidic linkage (Figure 18.4). Note that this enzyme has already catalyzed the breaking of an  $\alpha(1 \rightarrow 4)$  glycosidic linkage in the process of transferring the oligosaccharide segment. Each transferred segment must come from a chain at least 11 residues long; each new branch point must be at least 4 residues away from the nearest existing branch point.

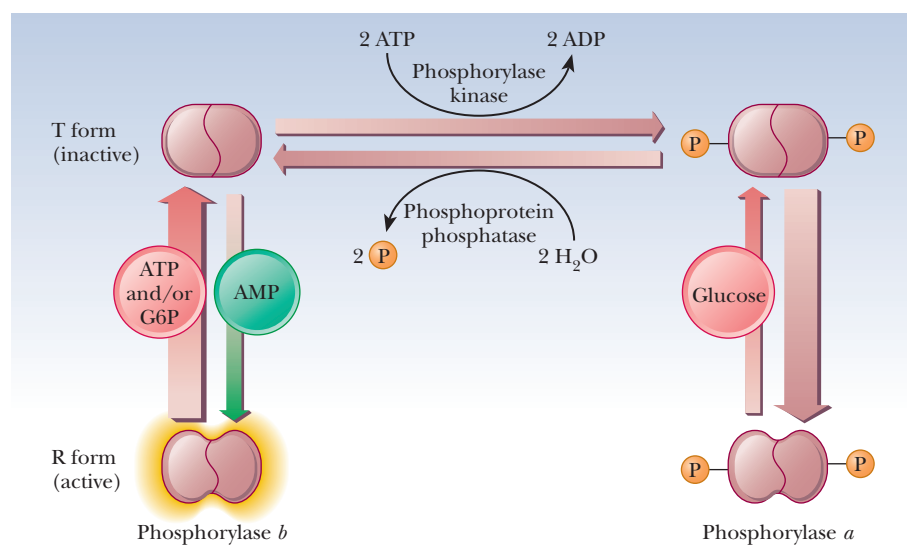
### How is glycogen metabolism controlled?

How does an organism ensure that glycogen synthesis and glycogen breakdown do not operate simultaneously? If this were to occur, the main result would be the hydrolysis of UTP, which would waste chemical energy stored in the phosphoric anhydride bonds. A major controlling factor lies in the behavior of glycogen phosphorylase. This enzyme is subject not only to allosteric control but also to another control feature: covalent modification. We saw an earlier example of this kind of control in the sodium–potassium pump in Section 8.6. In that example, phosphorylation and dephosphorylation of an enzyme determined whether it was active, and a similar effect takes place here.

Figure 18.5 summarizes some of the salient control features that affect glycogen phosphorylase activity. The enzyme is a dimer that exists in two forms, the inactive T (taut) form and the active R (relaxed) form. In the T form (and *only* in the T form), it can be modified by phosphorylation of a specific serine residue on each of the two subunits. The esterification of the serines to phosphoric acid is catalyzed by the enzyme *phosphorylase kinase*; the dephosphorylation is catalyzed by *phosphoprotein phosphatase*. The phosphorylated form of glycogen phosphorylase is called **phosphorylase a**, and the dephosphorylated form is called **phosphorylase b**. The switch from phosphorylase b to phosphorylase a is the major form of control over the activity of phosphorylase. The response time of the changes is on the order of seconds to minutes. Phosphorylase is



■ **FIGURE 18.4** The mode of action of the branching enzyme in glycogen synthesis. A segment seven residues long is transferred from a growing branch to a new branch point, where an  $\alpha(1 \rightarrow 6)$  linkage is formed.



■ **FIGURE 18.5** Glycogen phosphorylase activity is subject to allosteric control and covalent modification. Phosphorylation of the *a* form of the enzyme converts it to the *b* form. Only the *T* form is subject to modification and demodification. The *a* and *b* forms respond to different allosteric effectors (see text). ATP and glucose are allosteric inhibitors. AMP is an allosteric activator.

also controlled more quickly in times of urgency by allosteric effectors, with a response time of milliseconds.

In liver, glucose is an allosteric inhibitor of phosphorylase *a*. It binds to the substrate site and favors the transition to the T state. It also exposes the phosphorylated serines so that the phosphatase can hydrolyze them. This shifts the equilibrium to phosphorylase *b*. In muscle, the primary allosteric effectors are ATP, AMP, and glucose-6-phosphate (G6P). When the muscles use ATP to contract, AMP levels rise. This increase in AMP stimulates formation of the R state of phosphorylase *b*, which is active. When ATP is plentiful or glucose-6-phosphate builds up, these molecules act as allosteric inhibitors shifting the equilibrium back to the T form. These differences ensure that glycogen will be degraded when there is a need for energy, as is the case with high [AMP], low [G6P], and low [ATP]. When the reverse is true (low [AMP], high [G6P], and high [ATP]), the need for energy, and consequently for glycogen breakdown, is less. “Shutting down” glycogen phosphorylase activity is the appropriate response. The combination of covalent modification and allosteric control of the process allows for a degree of fine-tuning that would not be possible with either mechanism alone. Hormonal control also enters into the picture. When epinephrine is released from the adrenal gland in response to stress, it triggers a series of events, discussed more fully in Section 24.4, that suppress the activity of glycogen synthase and stimulate that of glycogen phosphorylase.

The activity of glycogen synthase is subject to the same type of covalent modification as glycogen phosphorylase. The difference is that the response is opposite. The inactive form of glycogen synthase is the phosphorylated form. The active form is unphosphorylated. The hormonal signals (glucagon or epinephrine) stimulate the phosphorylation of glycogen synthase via an enzyme called cAMP-dependent protein kinase (Chapter 24). After the glycogen synthase is phosphorylated, it becomes inactive at the same time the hormonal signal is activating phosphorylase. Glycogen synthase can also be phosphorylated by several other enzymes, including phosphorylase kinase and several enzymes called glycogen synthase kinases. Glycogen synthase is dephosphorylated by the same phosphoprotein phosphatase that removes the phosphate from phosphorylase. The phosphorylation of glycogen synthase is also more complicated in that there are multiple phosphorylation sites. As many as nine different amino acid residues have been found to be phosphorylated. As the progressive level of phosphorylation increases, the activity of the enzyme decreases.

Glycogen synthase is also under allosteric control. It is inhibited by ATP. This inhibition can be overcome by glucose-6-phosphate, which is an activator. However, the two forms of glycogen synthase respond very differently to glucose-6-phosphate. The phosphorylated (inactive) form is called **glycogen synthase D** (for “glucose-6-phosphate dependent”) because it is active only under very high concentrations of glucose-6-phosphate. In fact, the level necessary to give significant activity would be beyond the physiological range. The nonphosphorylated form is called **glycogen synthase I** (for “glucose-6-phosphate independent”) because it is active even with low concentrations of glucose-6-phosphate. Thus, even though purified enzymes can be shown to respond to allosteric effectors, the true control over the activity of glycogen synthase is by its phosphorylation state, which, in turn, is controlled by hormonal states.

The fact that two target enzymes, glycogen phosphorylase and glycogen synthase, are modified in the same way by the same enzymes links the opposing processes of synthesis and breakdown of glycogen even more intimately.

Finally, the modifying enzymes are themselves subject to covalent modification and allosteric control. This feature complicates the process considerably but adds the possibility of an amplified response to small changes in conditions. A small change in the concentration of an allosteric effector of a modifying enzyme can cause a large change in the concentration of an active, modified target enzyme; this amplification response is due to the fact that the substrate

## Biochemical Connections EXERCISE PHYSIOLOGY

### Why Do Athletes Go In for Glycogen Loading?

Glycogen is the primary energy source for a muscle that was at rest and then starts working vigorously. The energy of ATP hydrolysis derived from glycogen breakdown is initially produced *anaerobically*, with the lactic acid product being processed back to glucose in the liver. As an athlete becomes well conditioned, the muscle cells have more mitochondria, allowing for more *aerobic* metabolism of fats and carbohydrates for energy.

The switch to aerobic metabolism takes a few minutes, which is why athletes must warm up before an event. In long-distance events, athletes rely more on fat metabolism than they do in short-distance events, but in any race there is a final surge at the end in which the level of muscle glycogen may well determine the winner.

The idea behind glycogen loading is that if there is more available glycogen, then a person can carry out anaerobic metabolism for a longer period of time, either at the end of a distance event or for the entire event, if the effort level is high enough. This is probably true, but several questions come to mind: How long does the glycogen last? What is the best way to “load” glycogen? Is it safe? Theoretical calculations estimate that it takes 8 to 12 minutes to use all the glycogen in the skeletal muscle, although this range varies greatly depending on the intensity level. Allowing for loading of extra glycogen, it might last half an hour. Evidence shows that glycogen may be used more slowly in well-conditioned athletes because they exhibit higher fat utilization.

Early loading methods involved glycogen depletion for three days via a high-protein diet and extreme exercise, followed by loading from a high-carbohydrate diet and resting. This method yields a marked increase of glycogen, but some of it then is stored in the heart (which usually has little or no glycogen). The practice actually stresses the heart muscle. There is clearly some danger here. Dangers are also associated with the high-protein diet because too much protein often leads to a mineral imbalance, which also stresses the heart and the kidneys. Again, there is some danger. In addition, the training was often nonoptimal during the week because the athlete had trouble performing while on the low-carbohydrate diet and didn’t train much at all during the loading phase. Simple carbohydrate loading without previous extreme glycogen depletion does increase glycogen, but not as much; however, this increase does not risk potential stress to the heart.

Simple loading merely involves eating diets rich in pasta, starch, and complex carbohydrate for a few days before the strenuous exertion. It is not clear whether simple loading works.

It is certainly possible to increase the amount of glycogen in muscle, but a question remains about how long it will last during vigorous exercise. Ultimately, all diet considerations for athletes are very individual, and what works for one may not work for another.



Courtesy of Mount Holyoke College

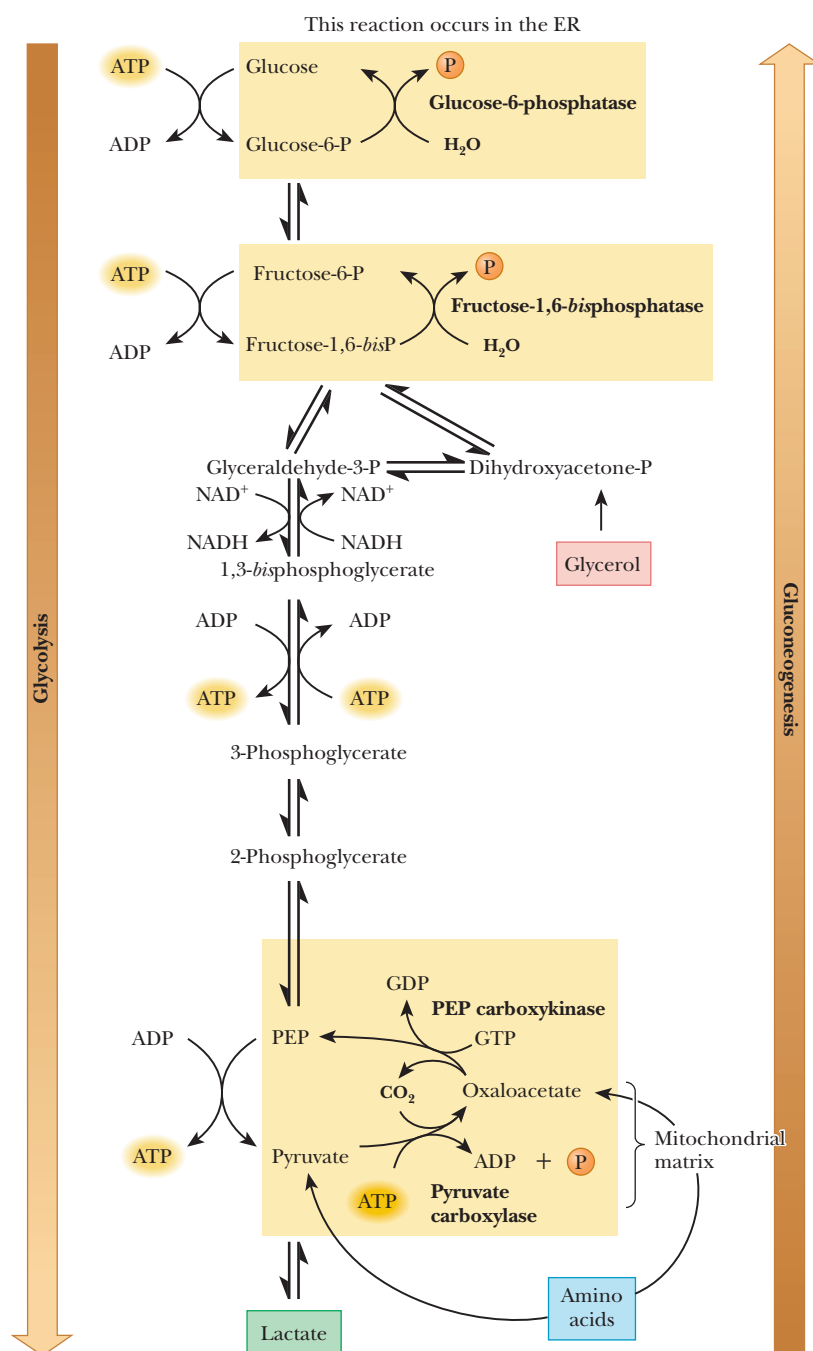
for the modifying enzyme is itself an enzyme. At this point, the situation has become very complex indeed, but it is a good example of how opposing processes of breakdown and synthesis can be controlled to the advantage of an organism. When we see in the next section how glucose is synthesized from lactate, we shall have another example, one that we can contrast with glycolysis to explore in more detail how carbohydrate metabolism is controlled.

## 18.2 Gluconeogenesis Produces Glucose from Pyruvate

The conversion of pyruvate to glucose occurs by a process called **gluconeogenesis**. Gluconeogenesis is not the exact reversal of glycolysis. We first met pyruvate as a product of glycolysis, but it can arise from other sources to be the starting point of the anabolism of glucose. Some of the reactions of glycolysis are essentially irreversible; these reactions are bypassed in gluconeogenesis.

An analogy is a hiker who goes directly down a steep slope but climbs back up the hill by an alternative, easier route. We shall see that the biosynthesis and the degradation of many important biomolecules follow different pathways.

Glycolysis involves three irreversible steps, and the differences between glycolysis and gluconeogenesis are found in these three reactions. The first of the glycolytic reactions is the production of pyruvate (and ATP) from phosphoenolpyruvate. The second is the production of fructose-1,6-bisphosphate from fructose-6-phosphate, and the third is the production of glucose-6-phosphate from glucose. Because the first of these reactions is exergonic, the reverse reaction is endergonic. Reversing the second and third reactions would require the production of ATP from ADP, which is also an endergonic reaction. The net result of gluconeogenesis includes the reversal of these three glycolytic reactions, but the pathway is different, with different reactions and different enzymes (Figure 18.6).

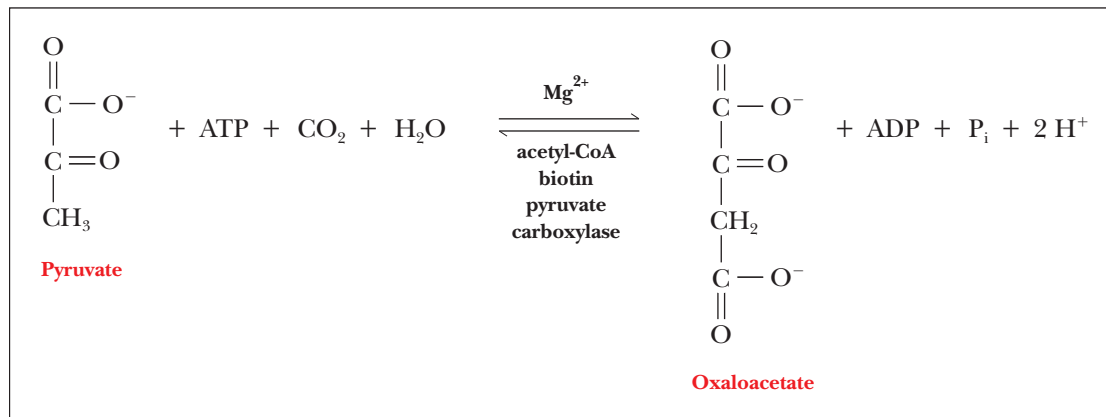


**FIGURE 18.6** The pathways of gluconeogenesis and glycolysis. Species in blue, green, and pink shaded boxes indicate other entry points for gluconeogenesis (in addition to pyruvate).



### Why is oxaloacetate an intermediate in gluconeogenesis?

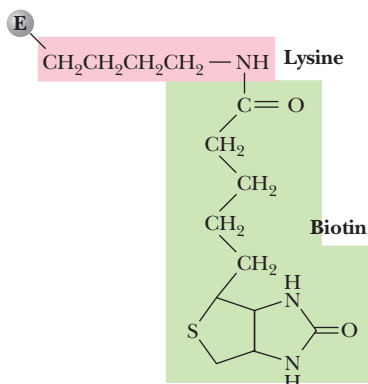
The conversion of pyruvate to phosphoenolpyruvate in gluconeogenesis takes place in two steps. The first step is the reaction of pyruvate and carbon dioxide to give oxaloacetate. This step requires energy, which is available from the hydrolysis of ATP.



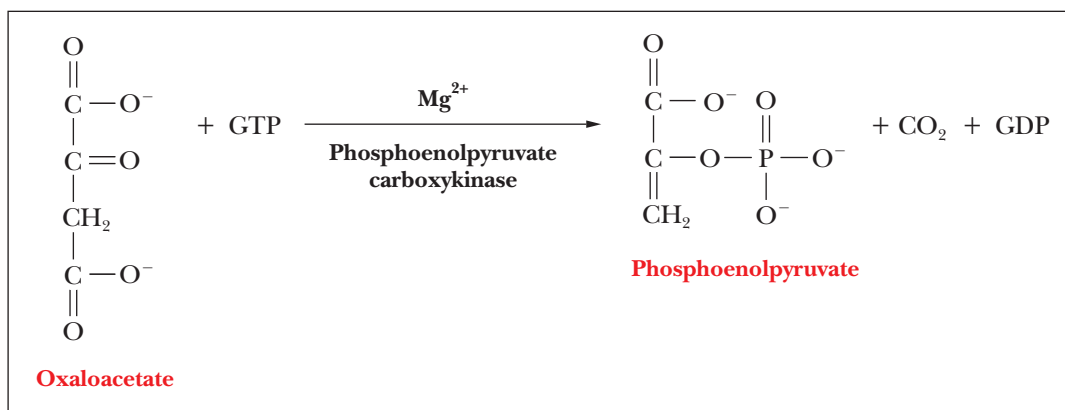
The enzyme that catalyzes this reaction is *pyruvate carboxylase*, an allosteric enzyme found in the mitochondria. Acetyl-CoA is an allosteric effector that activates pyruvate carboxylase. If high levels of acetyl-CoA are present (in other words, if there is more acetyl-CoA than is needed to supply the citric acid cycle), pyruvate (a precursor of acetyl-CoA) can be diverted to gluconeogenesis. (Oxaloacetate from the citric acid cycle can frequently be a starting point for gluconeogenesis as well.) Magnesium ion ( $\text{Mg}^{2+}$ ) and biotin are also required for effective catalysis. We have seen  $\text{Mg}^{2+}$  as a cofactor before, but we have not seen biotin, which requires some discussion.

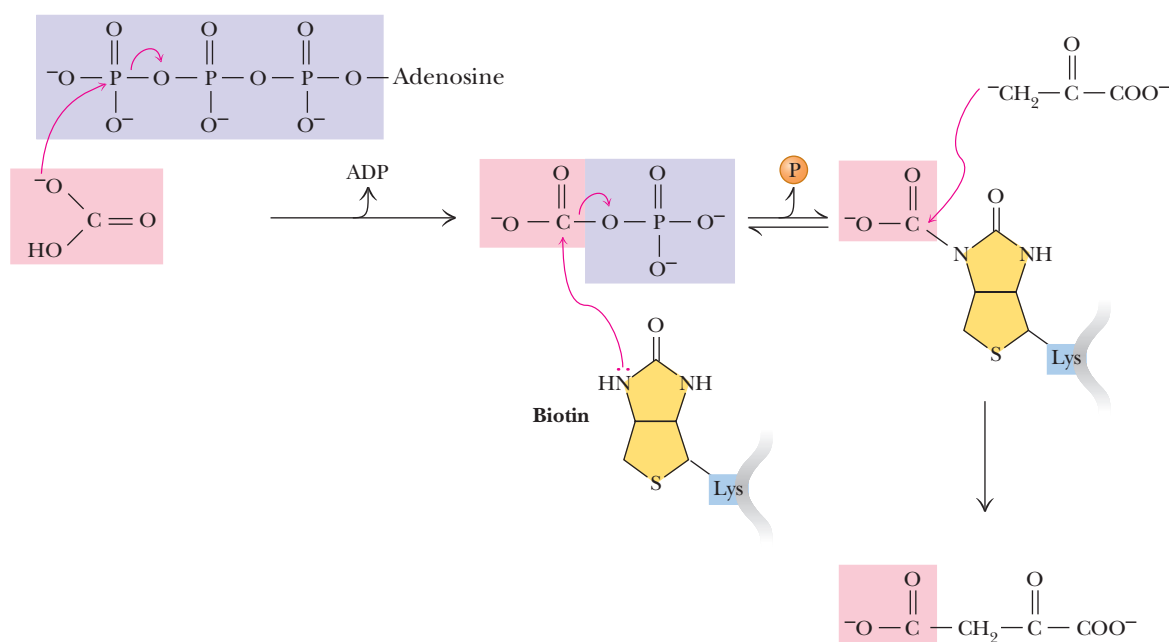
**Biotin** is a carrier of carbon dioxide; it has a specific site for covalent attachment of  $\text{CO}_2$  (Figure 18.7). The carboxyl group of the biotin forms an amide bond with the  $\epsilon$ -amino group of a specific lysine side chain of pyruvate carboxylase. The  $\text{CO}_2$  is attached to the biotin, which, in turn, is covalently bonded to the enzyme, and then the  $\text{CO}_2$  is shifted to pyruvate to form oxaloacetate (Figure 18.8). Note that ATP is required for this reaction.

The conversion of oxaloacetate to phosphoenolpyruvate is catalyzed by the enzyme *phosphoenolpyruvate carboxykinase* (*PEPCK*), which is found in the mitochondria and the cytosol. This reaction also involves hydrolysis of a nucleoside triphosphate—GTP, in this case, rather than ATP.



■ **FIGURE 18.7** The structure of biotin and its mode of attachment to pyruvate carboxylase.



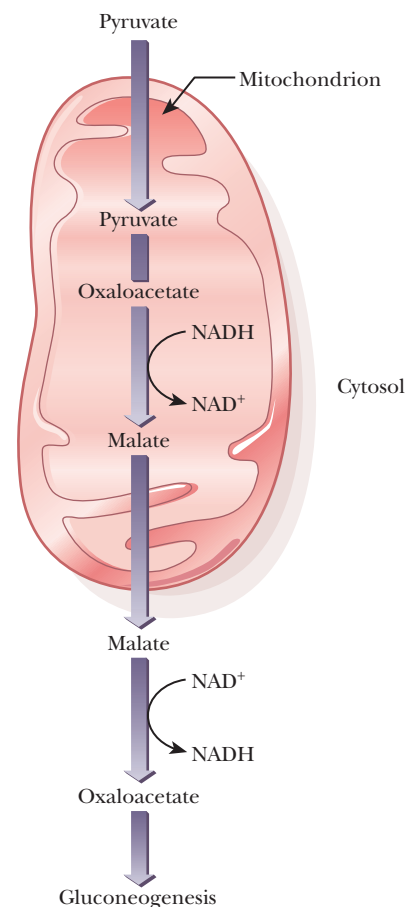


■ **FIGURE 18.8** The two stages of the pyruvate carboxylase reaction.  $\text{CO}_2$  is attached to the biotinylated enzyme.  $\text{CO}_2$  is transferred from the biotinylated enzyme to pyruvate, forming oxaloacetate. ATP is required in the first part of the reaction.

The successive carboxylation and decarboxylation reactions are both close to equilibrium (they have low values of their standard free energies); as a result, the conversion of pyruvate to phosphoenolpyruvate is also close to equilibrium ( $\Delta G^\circ = 2.1 \text{ kJ mol}^{-1} = 0.5 \text{ kcal mol}^{-1}$ ). A small increase in the level of oxaloacetate can drive the equilibrium to the right, and a small increase in the level of phosphoenolpyruvate can drive it to the left. A concept well known in general chemistry, the **law of mass action**, relates the concentrations of reactants and products in a system at equilibrium. Changing the concentration of reactants or products causes a shift to reestablish equilibrium. A reaction proceeds to the right on addition of reactants and to the left on addition of products.



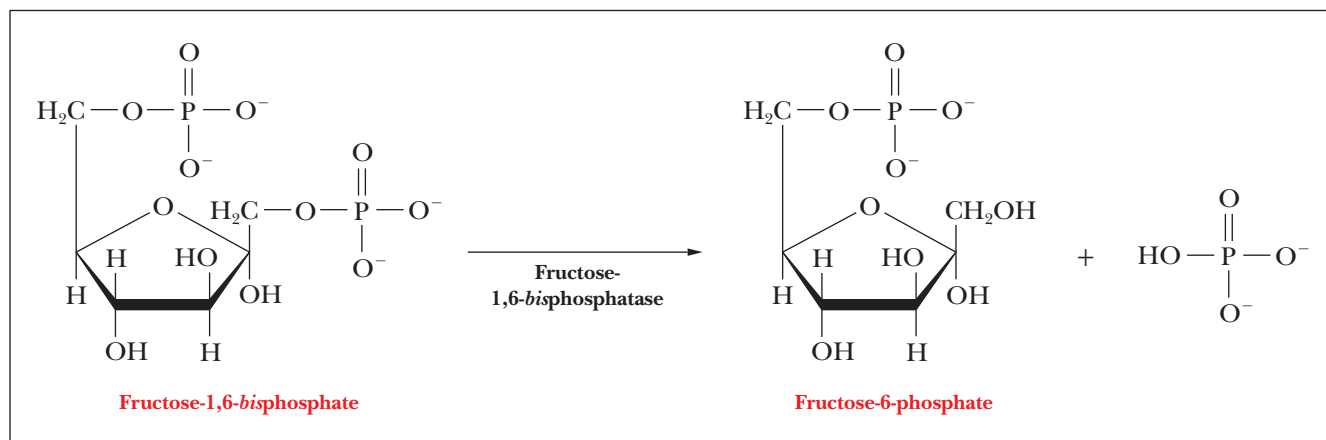
The oxaloacetate formed in the mitochondria can have two fates with respect to gluconeogenesis. It can continue to form PEP, which can then leave the mitochondria via a specific transporter to continue gluconeogenesis in the cytosol. The other possibility is that the oxaloacetate can be turned into malate via mitochondrial malate dehydrogenase, a reaction that uses NADH, as shown in Figure 18.9. Malate can then leave the mitochondria and have the reaction reversed by cytosolic malate dehydrogenase. The reason for this two-step process is that oxaloacetate cannot leave the mitochondria, but malate can. (The pathway involving malate is the one that takes place in the liver, where gluconeogenesis largely takes place.) You might wonder why these two paths exist to get PEP into the cytosol to continue gluconeogenesis. The answer brings us back to a familiar enzyme we saw in glycolysis, glyceraldehyde-3-phosphate dehydrogenase. Remember from Chapter 17 that the purpose of lactate dehydrogenase is to reduce pyruvate to lactate so that NADH could be oxidized to form  $\text{NAD}^+$ , which is needed to continue glycolysis. This reaction must be reversed in gluconeogenesis, and the cytosol has a low ratio of NADH to  $\text{NAD}^+$ . The purpose of the roundabout way of getting oxaloacetate out of the mitochondria via malate dehydrogenase is to produce NADH in the cytosol so that gluconeogenesis can continue.



■ **FIGURE 18.9** Pyruvate carboxylase catalyzes a compartmentalized reaction. Pyruvate is converted to oxaloacetate in the mitochondria. Because oxaloacetate cannot be transported across the mitochondrial membrane, it must be reduced to malate, transported to the cytosol, and then oxidized back to oxaloacetate before gluconeogenesis can continue.

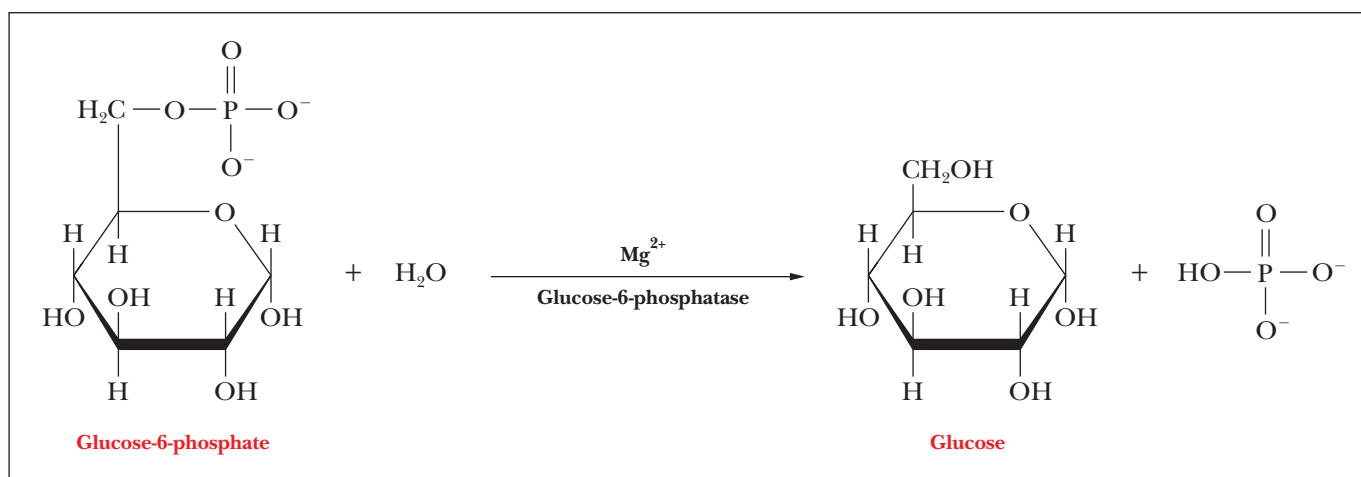
**What is the role of sugar phosphates in gluconeogenesis?**

The other two reactions in which gluconeogenesis differs from glycolysis are ones in which a phosphate-ester bond to a sugar-hydroxyl group is hydrolyzed. Both reactions are catalyzed by phosphatases, and both reactions are exergonic. The first reaction is the hydrolysis of fructose-1,6-bisphosphate to produce fructose-6-phosphate and phosphate ion ( $\Delta G^{\circ} = -16.7 \text{ kJ mol}^{-1} = -4.0 \text{ kcal mol}^{-1}$ ).



This reaction is catalyzed by the enzyme *fructose-1,6-bisphosphatase*, an allosteric enzyme strongly inhibited by adenosine monophosphate (AMP) but stimulated by ATP. Because of allosteric regulation, this reaction is also a control point in the pathway. When the cell has an ample supply of ATP, the formation rather than the breakdown of glucose is favored. This enzyme is inhibited by fructose-2,6-bisphosphate, a compound we met in Section 17.2 as an extremely potent activator of phosphofructokinase. We shall return to this point in the next section.

The second reaction is the hydrolysis of glucose-6-phosphate to glucose and phosphate ion ( $\Delta G^{\circ} = -13.8 \text{ kJ mol}^{-1} = -3.3 \text{ kcal mol}^{-1}$ ). The enzyme that catalyzes this reaction is *glucose-6-phosphatase*.



When we discussed glycolysis, we saw that both of the phosphorylation reactions, which are the reverse of these two phosphatase-catalyzed reactions, are endergonic. In glycolysis, the phosphorylation reactions must be coupled to the hydrolysis of ATP to make them exergonic and thus energetically allowed. In gluconeogenesis, the organism can make direct use of the fact that the

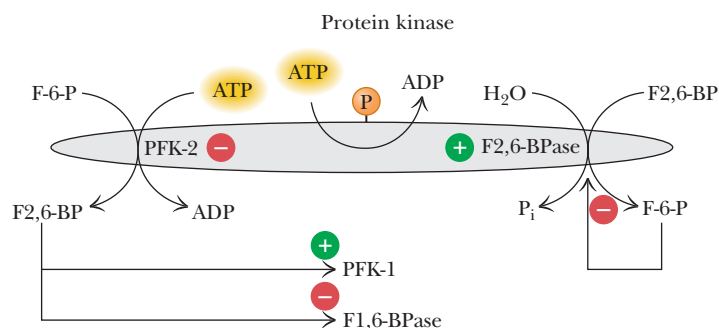
hydrolysis reactions of the sugar phosphates are exergonic. The corresponding reactions are not the reverse of each other in the two pathways. They differ from each other in whether they require ATP and in the enzymes involved. Hydrolysis of glucose-6-phosphate to glucose occurs in the endoplasmic reticulum. This is an example of an interesting pathway that requires three cellular locations (mitochondria, cytosol, endoplasmic reticulum).

## 18.3 Control of Carbohydrate Metabolism

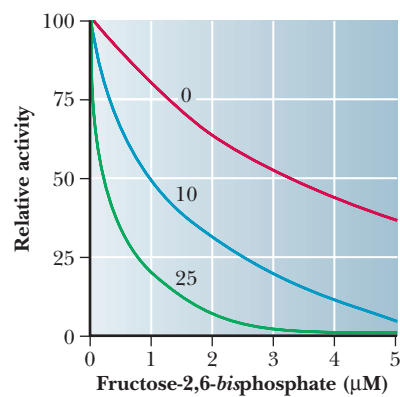
We have now seen several aspects of carbohydrate metabolism: glycolysis, gluconeogenesis, and the reciprocal breakdown and synthesis of glycogen. Glucose has a central role in all these processes. It is the starting point for glycolysis, in which it is broken down to pyruvate, and for the synthesis of glycogen, in which many glucose residues combine to give the glycogen polymer. Glucose is also the product of gluconeogenesis, which has the net effect of reversing glycolysis; glucose is also obtained from the breakdown of glycogen. Each of the opposing pathways, glycolysis and gluconeogenesis, on the one hand, and the breakdown and synthesis of glycogen, on the other hand, is not the exact reversal of the other, even though the net results are. In other words, a different path is used to arrive at the same place. It is time to see how all these related pathways are controlled.

### How does control of key enzymes control carbohydrate metabolism?

An important element in the control process involves fructose-2,6-bisphosphate (F2,6P). We mentioned in Section 17.2 that this compound is an important allosteric activator of phosphofructokinase (PFK), the key enzyme of glycolysis; it is also an inhibitor of fructose bisphosphate phosphatase (FBPase), which plays a role in gluconeogenesis. A high concentration of F2,6P stimulates glycolysis, whereas a low concentration stimulates gluconeogenesis. The concentration of F2,6P in a cell depends on the balance between its synthesis, catalyzed by *phosphofructokinase-2* (PFK-2), and its breakdown, catalyzed by *fructose-bisphosphatase-2* (FBPase-2). The enzymes that control the formation and breakdown of F2,6P are themselves controlled by a phosphorylation/dephosphorylation mechanism similar to what we have already seen in the case of glycogen phosphorylase and glycogen synthase (Figure 18.10). Both enzyme activities are located on the same protein (a dimer of about 100 kDa molecular mass). Phosphorylation of the dimeric protein leads to an increase in activity of FBPase-2 and a decrease in the concentration of F2,6P, ultimately stimulating gluconeogenesis. Dephosphorylation of the dimeric protein leads to an increase in PFK-2 activity and an increase in the concentration of F2,6P, ultimately stimulating glycolysis.



**FIGURE 18.10** The formation and breakdown of fructose-2,6-bisphosphate (F2,6P). These two processes are catalyzed by two enzyme activities on the same protein. These two enzyme activities are controlled by a phosphorylation/dephosphorylation mechanism. The green plus sign indicates phosphorylation, and the red minus sign, dephosphorylation. Phosphorylation activates the enzyme that degrades F2,6P whereas dephosphorylation activates the enzyme that produces it.



■ **FIGURE 18.11** Allosteric effects in the control of carbohydrate metabolism. The effect of AMP (0, 10, and 25  $\mu\text{M}$  [micromolar]) on the inhibition of fructose-1,6-bisphosphatase by fructose-2,6-bisphosphate. Activity was measured in the presence of 10  $\mu\text{M}$  fructose-1,6-bisphosphate. (Adapted from Van Schaftingen, E., and H. G. Hers, 1981. *Inhibition of fructose-1,6-bisphosphatase by fructose-2,6-bisphosphate*. Proc. Nat'l Acad. of Sci., U.S.A. 78:2861–2863.)

The net result is similar to the control of glycogen synthesis and breakdown that we saw in Section 18.1.

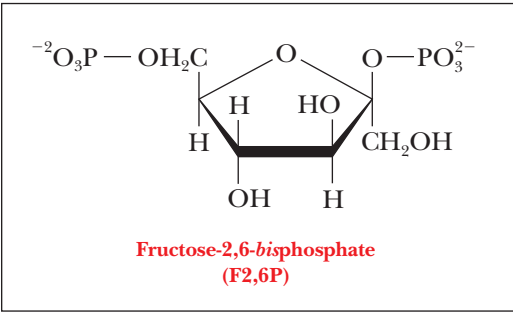


Figure 18.11 shows the effect of fructose-2,6-bisphosphate on the activity of FBPase. The inhibitor works by itself, but its effect is greatly increased by the presence of the allosteric inhibitor AMP.

Table 18.1 summarizes important mechanisms of metabolic control. Even though we discuss them here in the context of carbohydrate metabolism, they apply to all aspects of metabolism. Of the four kinds of control mechanisms listed in Table 18.1—allosteric control, covalent modification, substrate cycling, and genetic control—we have seen examples of allosteric control and covalent modification and, in Chapter 11, discussed genetic control using the *lac* operon as an example. Substrate cycling is a mechanism that we can profitably discuss here.

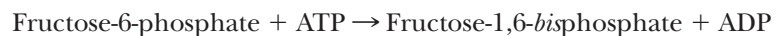
The term **substrate cycling** refers to the fact that opposing reactions can be catalyzed by different enzymes. Consequently, the opposing reactions can be independently regulated and have different rates. It would not be possible

**TABLE 18.1**

Mechanisms of Metabolic Control		
Type of Control	Mode of Operation	Examples
Allosteric	Effectors (substrates, products, or coenzymes) of a pathway inhibit or activate an enzyme. (Responds rapidly to external stimuli.)	ATCase (Section 7.2); Phosphofructokinase (Section 17.2)
Covalent modification	Inhibition or activation of enzyme depends on formation or breaking of a bond, frequently by phosphorylation or dephosphorylation. (Responds rapidly to external stimuli.)	Sodium–potassium pump (Section 8.6); Glycogen phosphorylase, glycogen synthase (Section 18.1)
Substrate cycles	Two opposing reactions, such as formation and breakdown of a given substance, are catalyzed by different enzymes, which can be activated or inhibited separately. (Responds rapidly to external stimuli.)	Glycolysis (Chapter 17) and gluconeogenesis (Section 18.2)
Genetic control	The amount of enzyme present is increased by protein synthesis. (Longer-term control than the other mechanisms listed here.)	Induction of $\beta$ -galactosidase (Section 11.2)



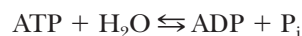
to have different rates with the same enzyme because a catalyst speeds up a reaction and the reverse of the reaction to the same extent (Section 6.2). We shall use the conversion of fructose-6-phosphate to fructose-1,6-bisphosphate and then back to fructose-6-phosphate as an example of a substrate cycle. In glycolysis, the reaction catalyzed by phosphofructokinase is highly exergonic under physiological conditions ( $\Delta G = -25.9 \text{ kJ mol}^{-1} = -6.2 \text{ kcal mol}^{-1}$ ).



The opposing reaction, which is part of gluconeogenesis, is also exergonic ( $\Delta G = -8.6 \text{ kJ mol}^{-1} = -2.1 \text{ kcal mol}^{-1}$  under physiological conditions) and is catalyzed by another enzyme, namely fructose-1,6-bisphosphatase.



Note that the opposing reactions are not the exact reverse of one another. If we add the two opposing reactions together, we obtain the net reaction



Hydrolysis of ATP is the energetic price that is paid for independent control of the opposing reactions.

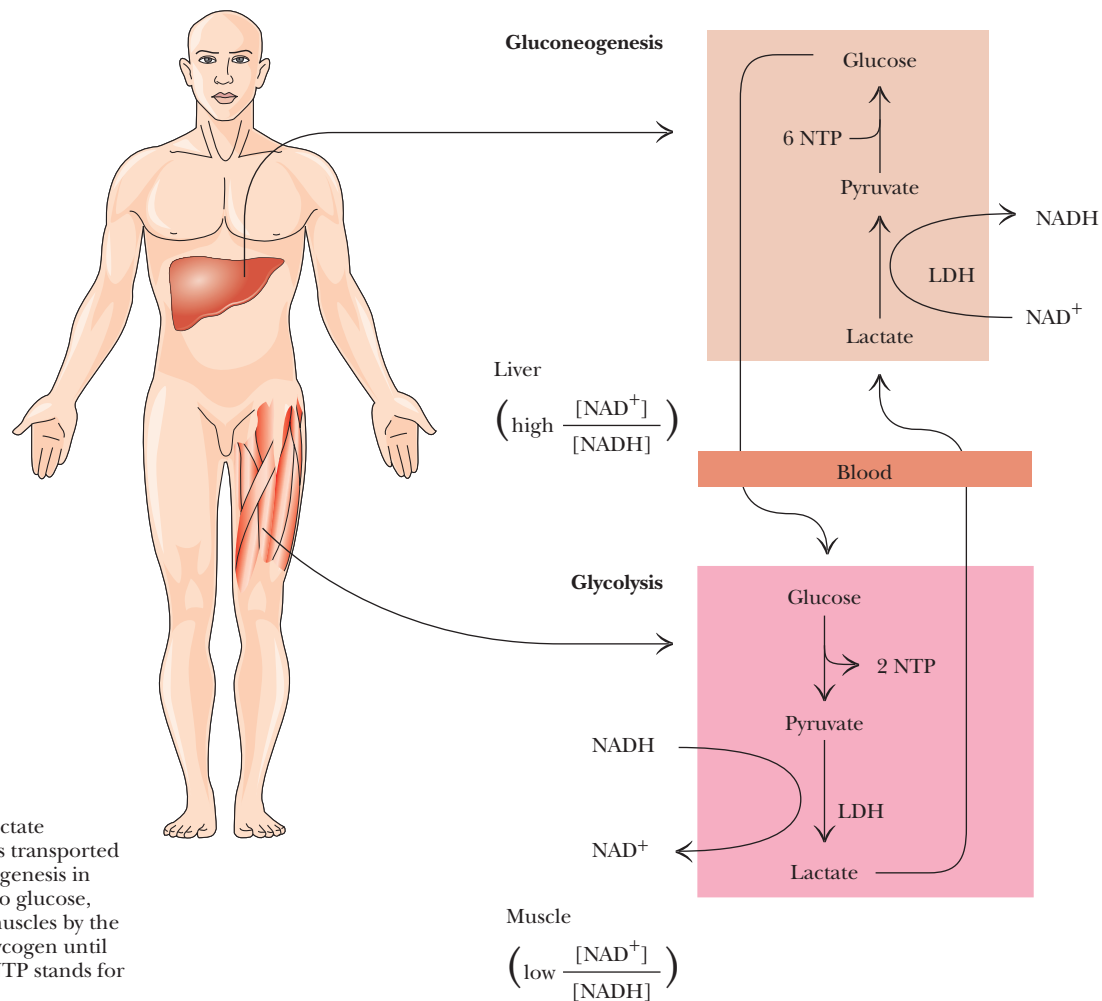
### How do different organs share carbohydrate metabolism?

Using combinations of these control mechanisms, an organism can set up a division of labor among tissues and organs to maintain control of glucose metabolism. A particularly clear example is found in the Cori cycle. Shown in Figure 18.12, the Cori cycle is named for Gerty and Carl Cori, who first described it. There is cycling of glucose due to glycolysis in muscle and gluconeogenesis in liver. Glycolysis in fast-twitch skeletal muscle produces lactate under conditions of oxygen debt, such as a sprint. Fast-twitch muscle has comparatively few mitochondria, so metabolism is largely anaerobic in this tissue. The buildup of lactate contributes to the muscular aches that follow strenuous exercise. Gluconeogenesis recycles the lactate that is produced (lactate is first oxidized to pyruvate). The process occurs to a great extent in the liver after the lactate is transported there by the blood. Glucose produced in the liver is transported back to skeletal muscle by the blood, where it becomes an energy store for the next burst of exercise. This is the main reason that athletes receive post-event massages and that they always cool down after the event. Cooling down keeps the blood flowing through the muscles and allows the lactate and other acids to leave the cells and enter the blood. Massages increase this movement from the cells to the blood. Note that we have a division of labor between two different types of organs—muscle and liver. In the same cell (of whatever type), these two metabolic pathways—glycolysis and gluconeogenesis—are not highly active simultaneously. When the cell needs ATP, glycolysis is more active; when there is little need for ATP, gluconeogenesis is more active. Because of the hydrolysis of ATP and GTP in the reactions of gluconeogenesis that differ from those of glycolysis, the overall pathway from two molecules of pyruvate back to one molecule of glucose is exergonic ( $\Delta G^\circ = -37.6 \text{ kJ mol}^{-1} = -9.0 \text{ kcal mol}^{-1}$ , for one mole of glucose). The conversion of pyruvate to lactate is exergonic, which means that the reverse reaction is endergonic. The energy released by the exergonic conversion of pyruvate to glucose by gluconeogenesis facilitates the endergonic conversion of lactate to pyruvate.

Note that the Cori cycle requires the net hydrolysis of two ATP and two GTP. ATP is produced by the glycolytic part of the cycle, but the portion involving gluconeogenesis requires yet more ATP in addition to GTP.

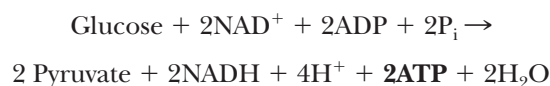


■ Gerty and Carl Cori, codiscoverers of the Cori cycle.

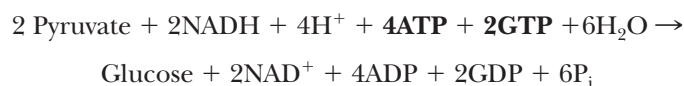


■ **FIGURE 18.12 The Cori cycle.** Lactate produced in muscles by glycolysis is transported by the blood to the liver. Gluconeogenesis in the liver converts the lactate back to glucose, which can be carried back to the muscles by the blood. Glucose can be stored as glycogen until it is degraded by glycogenolysis. (NTP stands for nucleoside triphosphate.)

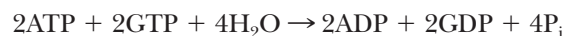
*Glycolysis:*



*Gluconeogenesis:*



*Overall:*



The hydrolysis of both ATP and GTP is the price of increased simultaneous control of the two opposing pathways.

### What roles do the first and last steps of glycolysis play in control of carbohydrate metabolism?

The final step of glycolysis is also a major control point in glucose metabolism. Pyruvate kinase (PK) is allosterically affected by several compounds. ATP and alanine both inhibit it. The ATP makes sense because there would be no reason to sacrifice glucose to make more energy if there is ample ATP.

The alanine may be less intuitive. Alanine is the amino version of pyruvate. In other words, it is one reaction away from pyruvate via an enzyme called a transaminase. Therefore, a high level of alanine indicates that a high level of pyruvate is already present, so the enzyme that would make more pyruvate can be shut down. Fructose-1,6-bisphosphate allosterically activates PK so that the incoming products of the first reactions of glycolysis can be processed.

Pyruvate kinase is also found as isozymes with three different types of subunits, M, L, and A. The M subunit predominates in muscle; the L, in liver; and the A, in other tissues. A native pyruvate kinase molecule has four subunits, similar to lactate dehydrogenase and phosphofructokinase. In addition to the allosteric controls mentioned earlier, the liver isozymes also are subject to covalent modification, as shown in Figure 18.13. Low levels of blood sugar trigger the release of glucagon, which leads to the production of a protein kinase, as we saw with glycogen phosphorylase. The protein kinase phosphorylates PK, which renders PK less active. In this way, glycolysis is shut down in the liver when blood glucose is low. The following Biochemical Connections box looks at a role for pyruvate kinase activation in the treatment of cancer.

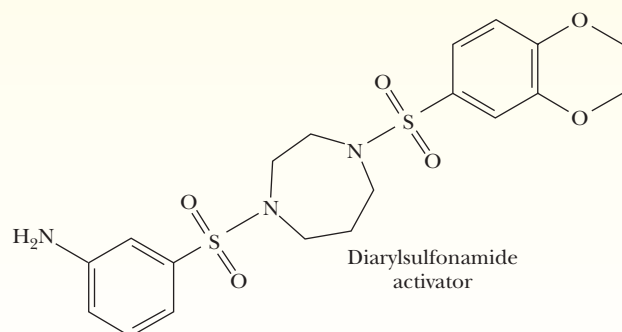
Hexokinase is inhibited by high levels of its product, glucose-6-phosphate. When glycolysis is inhibited through phosphofructokinase, glucose-6-phosphate

## Biochemical Connections CANCER RESEARCH

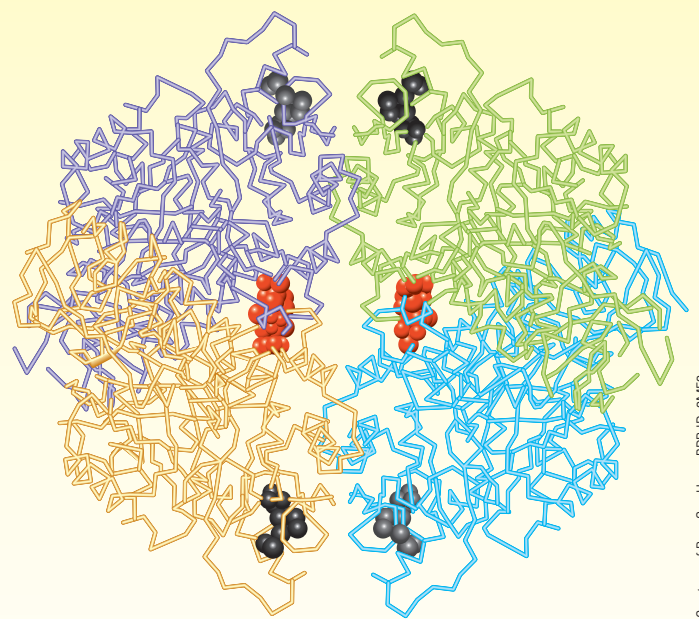
### Using Pyruvate Kinase Isozymes to Treat Cancer

It has long been known that the metabolism of cancer cells differs from that of normal cells. In carbohydrate metabolism, cancer cells rely on anaerobic metabolism, whereas normal cells tend mainly to utilize aerobic degradation of glucose to obtain energy. Two isozymes of pyruvate kinase play a key role in the difference. The PKM1 isozyme is found in normal cells, but the PKM2 form predominates in cancer cells. The PKM2 isozyme catalyzes the production of pyruvate less efficiently than PKM1. As a result, intermediate metabolites of glycolysis accumulate and, in turn, are used in the reactions that lead to a proliferation of cancer cells.

Researchers at the National Institutes of Health (NIH) set out to activate the PKM2 isozyme in hopes of switching the metabolism of cancer cells to that of normal cells. They were able to achieve their goal by combining the natural allosteric activator of pyruvate kinase, fructose-1,6-bisphosphate, with a second, synthetic activator, a member of the class of molecules known as diarylsulfonamides.



■ The structure of the synthetic diarylsulfonamide activator. (From Chemical and Engineering News, February 2010, p. 46. Reprinted by permission.)

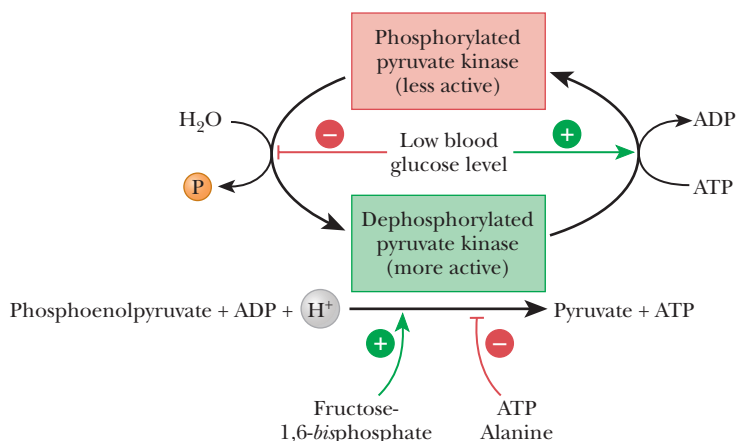


■ Tetramer of pyruvate kinase formed in the presence of the natural activator (shown in black) and the synthetic activator (shown in red). (From Chemical and Engineering News, February 15, 2010, page 46.)

PKM2 forms a tetramer with increased activity, similar to that of PKM1, in the presence of fructose-1,6-bisphosphate and the diarylsulfonamide. The structure of this tetramer has been determined by X-ray crystallography. In the figure above, two molecules of the diarylsulfonamide (red) are bound to the tetramer, in addition to one molecule of fructose-1,6-bisphosphate (black) for each subunit of the tetramer. This result is only one approach to metabolic treatments for cancer, but it may well lead the way to others.

Courtesy of Bum-Soo Hong, PDB ID: 3ME3

■ **FIGURE 18.13** Control of liver pyruvate kinase by phosphorylation. When blood glucose is low, phosphorylation of pyruvate kinase is favored. The phosphorylated form is less active, thereby slowing glycolysis and allowing pyruvate to produce glucose by gluconeogenesis.



builds up, shutting down hexokinase. This keeps glucose from being metabolized in the liver when it is needed in the blood and other tissues. However, the liver contains a second enzyme, glucokinase, that phosphorylates glucose. (See the Biochemical Connections box on page 157.) Glucokinase has a higher  $K_M$  for glucose than hexokinase, so it functions only when glucose is abundant. If there is an excess of glucose in the liver, glucokinase phosphorylates it to glucose-6-phosphate. The purpose of this phosphorylation is so that it can eventually be polymerized into glycogen.

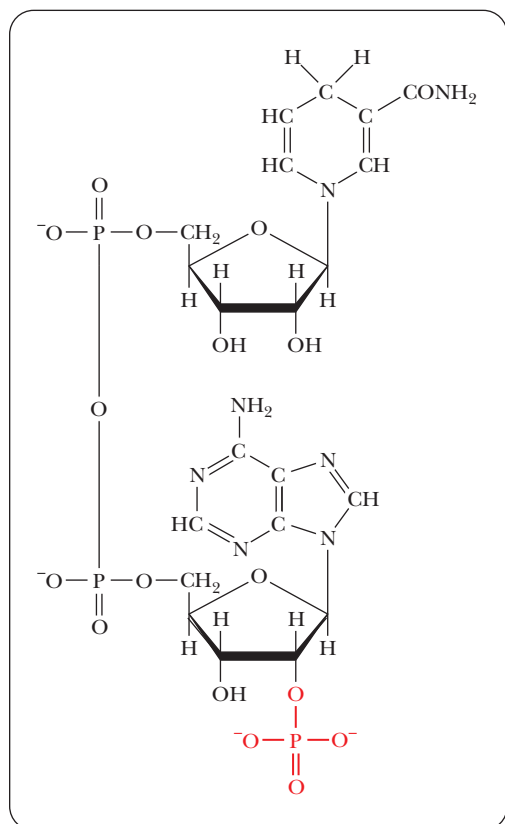
## 18.4 Glucose Is Sometimes Diverted through the Pentose Phosphate Pathway

The **pentose phosphate pathway** is an alternative to glycolysis and differs from it in several important ways. In glycolysis, one of our most important concerns was the production of ATP. In the pentose phosphate pathway, the production of ATP is not the crux of the matter. As the name of the pathway indicates, five-carbon sugars, including ribose, are produced from glucose. Ribose and its derivative deoxyribose play an important role in the structure of nucleic acids. Another important facet of the pentose phosphate pathway is the production of nicotinamide adenine dinucleotide phosphate (NADPH), a compound that differs from nicotinamide adenine dinucleotide (NADH) by having one extra phosphate group esterified to the ribose ring of the adenine nucleotide portion of the molecule (Figure 18.14). A more important difference is the way these two coenzymes function. NADH is produced in the oxidative reactions that give rise to ATP. NADPH is a reducing agent in biosynthesis, which, by its very nature, is a reductive process. For example, in Chapter 21, we shall see the important role that NADPH plays in the biosynthesis of lipids.

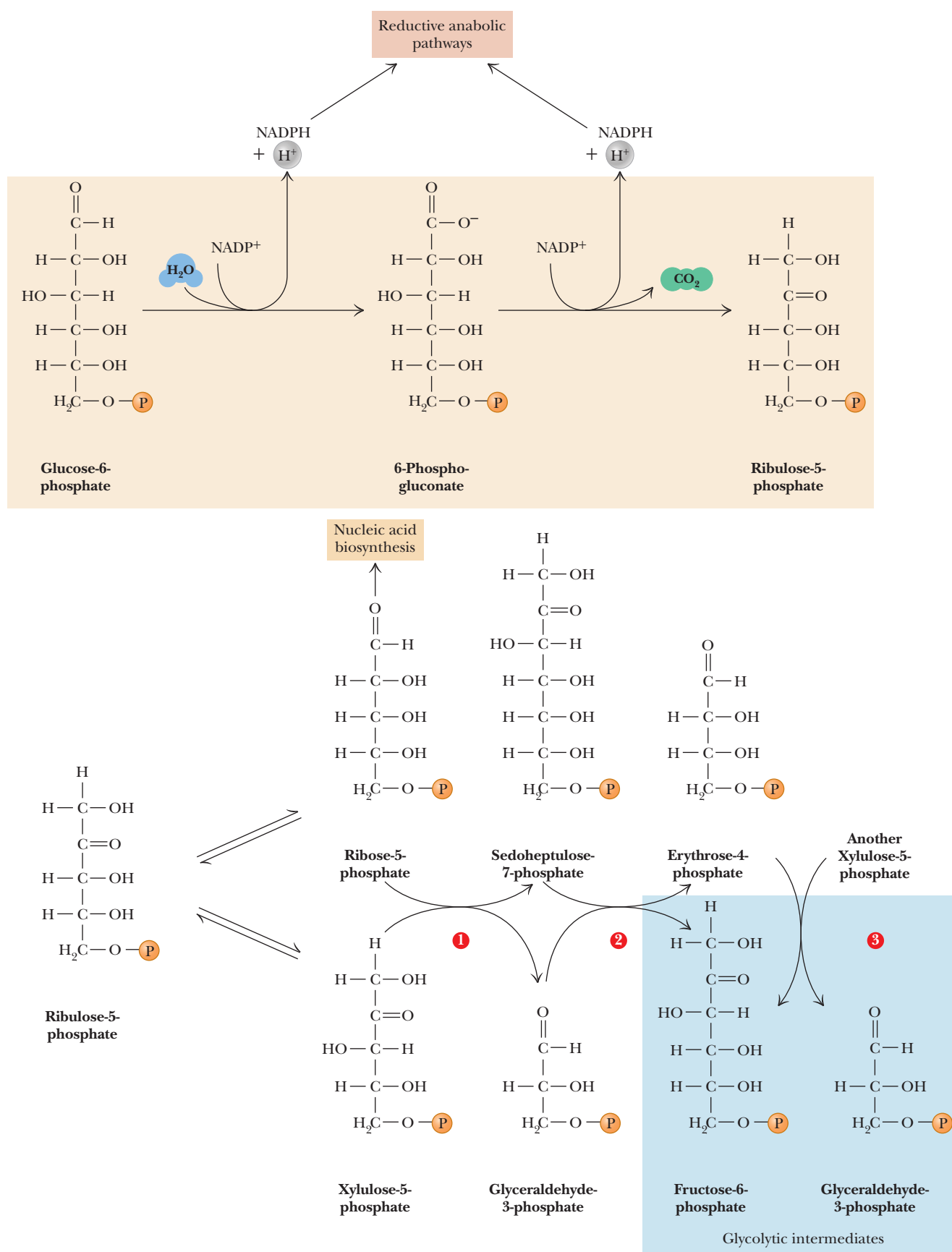
The pentose phosphate pathway begins with a series of oxidation reactions that produce NADPH and five-carbon sugars. The remainder of the pathway involves nonoxidative reshuffling of the carbon skeletons of the sugars involved. The products of these nonoxidative reactions include substances such as fructose-6-phosphate and glyceraldehyde-3-phosphate, which play a role in glycolysis. Some of these reshuffling reactions will reappear when we look at the production of sugars in photosynthesis.

### What are the oxidative reactions of the pentose phosphate pathway?

In the first reaction of the pathway, glucose-6-phosphate is oxidized to 6-phosphogluconate (Figure 18.15, *top*). The enzyme that catalyzes this reaction is *glucose-6-phosphate dehydrogenase*. Note that NADPH is produced by the reaction.



■ **FIGURE 18.14** The structure of reduced adenine dinucleotide phosphate (NADPH).



■ **FIGURE 18.15** The pentose phosphate pathway. The numerals in the red circles indicate steps discussed in the text.



The next reaction is an oxidative decarboxylation, and NADPH is produced once again. The 6-phosphogluconate molecule loses its carboxyl group, which is released as carbon dioxide, and the five-carbon keto-sugar (ketose) ribulose-5-phosphate is the other product. The enzyme that catalyzes this reaction is *6-phosphogluconate dehydrogenase*. In the process, the C-3 hydroxyl group of the 6-phosphogluconate is oxidized to form a  $\beta$ -keto acid, which is unstable and readily decarboxylates to form ribulose-5-phosphate.

**What are the nonoxidative reactions of the pentose phosphate pathway, and why are they important?**

In the remaining steps of the pentose phosphate pathway, several reactions involve transfer of two- and three-carbon units. To keep track of the carbon backbone of the sugars and their aldehyde and ketone functional groups, we shall write the formulas in the open-chain form.

There are two different reactions in which ribulose-5-phosphate isomerizes. In one of these reactions, catalyzed by *phosphopentose-3-epimerase*, there is an inversion of configuration around carbon atom 3, producing xylulose-5-phosphate, which is also a ketose (Figure 18.15, *bottom*). The other isomerization reaction, catalyzed by *phosphopentose isomerase*, produces a sugar with an aldehyde group (an aldose) rather than a ketone. In this second reaction, ribulose-5-phosphate isomerizes to ribose-5-phosphate (Figure 18.15, *bottom*). Ribose-5-phosphate is a necessary building block for the synthesis of nucleic acids and coenzymes such as NADH.

The group-transfer reactions that link the pentose phosphate pathway with glycolysis require the two five-carbon sugars produced by the isomerization of ribulose-5-phosphate. Two molecules of xylulose-5-phosphate and one molecule of ribose-5-phosphate rearrange to give two molecules of fructose-6-phosphate and one molecule of glyceraldehyde-3-phosphate. In other words, three molecules of pentose (with five carbon atoms each) give two molecules of hexose (with six carbon atoms each) and one molecule of a triose (with three carbon atoms). The total number of carbon atoms (15) does not change, but there is considerable rearrangement as a result of group transfer.

Two enzymes, *transketolase* and *transaldolase*, are responsible for the reshuffling of the carbon atoms of sugars such as ribose-5-phosphate and xylulose-5-phosphate in the remainder of the pathway, which consists of three reactions. Transketolase transfers a two-carbon unit. Transaldolase transfers a three-carbon unit. Transketolase catalyzes the first and third reactions in the rearrangement process, and transaldolase catalyzes the second reaction. The results of these transfers are summarized in Table 18.2. In the first of these reactions, a two-carbon unit from xylulose-5-phosphate (five carbons) is transferred to ribose-5-phosphate (five carbons) to give sedoheptulose-7-phosphate (seven carbons) and glyceraldehyde-3-phosphate (three carbons), as shown in Figure 18.15, *bottom*, red numeral 1.

TABLE 18.2

Group-Transfer Reactions in the Pentose Phosphate Pathway			
	Reactant	Enzyme	Products
Two-carbon shift	$C_5 + C_5$	Transketolase	$C_7 + C_3$
		$\rightleftharpoons$	
Three-carbon shift	$C_7 + C_3$	Transaldolase	$C_6 + C_4$
		$\rightleftharpoons$	
Two-carbon shift	$C_5 + C_4$	Transketolase	$C_6 + C_3$
Net reaction	$3C_5$	$\rightleftharpoons$	$2C_6 + C_3$

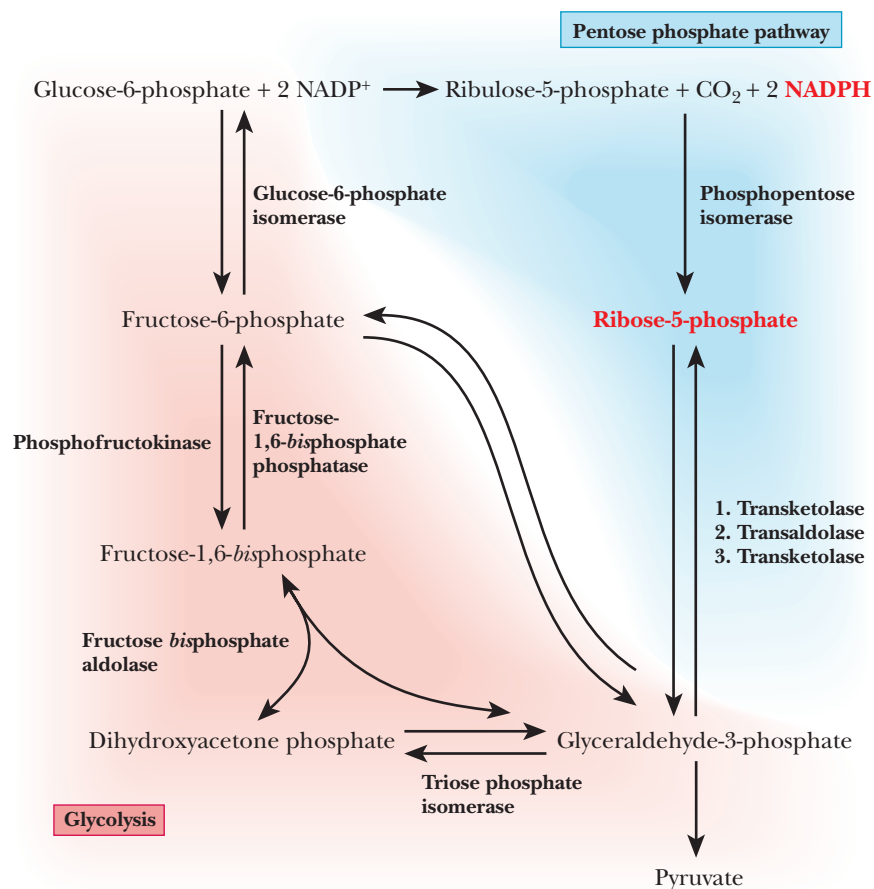
In the reaction catalyzed by transaldolase, a three-carbon unit is transferred from the seven-carbon sedoheptulose-7-phosphate to the three-carbon glyceraldehyde-3-phosphate (Figure 18.15, red numeral 2). The products of the reaction are fructose-6-phosphate (six carbons) and erythrose-4-phosphate (four carbons).

In the final reaction of this type in the pathway, xylulose-5-phosphate reacts with erythrose-4-phosphate. This reaction is catalyzed by transketolase. The products of the reaction are fructose-6-phosphate and glyceraldehyde-3-phosphate (Figure 18.15, red numeral 3).

In the pentose phosphate pathway, glucose-6-phosphate can be converted to fructose-6-phosphate and glyceraldehyde-3-phosphate by a means other than the glycolytic pathway. For this reason, the pentose phosphate pathway is also called the *hexose monophosphate shunt*, and this name is used in some texts. A major feature of the pentose phosphate pathway is the production of ribose-5-phosphate and NADPH. The control mechanisms of the pentose phosphate pathway can respond to the varying needs of organisms for either or both of these compounds.

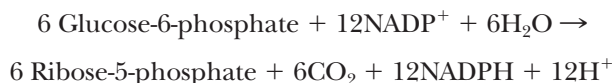
### How is the pentose phosphate pathway controlled?

As we have seen, the reactions catalyzed by transketolase and transaldolase are reversible, which allows the pentose phosphate pathway to respond to the needs of an organism. The starting material, glucose-6-phosphate, undergoes different reactions depending on whether there is a greater need for ribose-5-phosphate or for NADPH. The operation of the oxidative portion of the pathway depends strongly on the organism's requirement for NADPH. The need for ribose-5-phosphate can be met in other ways, since ribose-5-phosphate can be obtained from glycolytic intermediates without the oxidative reactions of the pentose phosphate pathway (Figure 18.16).



■ **FIGURE 18.16** Relationships between the pentose phosphate pathway and glycolysis. If the organism needs NADPH more than ribose-5-phosphate, the entire pentose phosphate pathway is operative. If the organism needs ribose-5-phosphate more than NADPH, the nonoxidative reactions of the pentose phosphate pathway, operating in reverse, produce ribose-5-phosphate (see text).

If the organism needs more NADPH than ribose-5-phosphate, the reaction series goes through the complete pathway just discussed. The oxidative reactions at the beginning of the pathway are needed to produce NADPH. The net reaction for the oxidative portion of the pathway is



The following Biochemical Connections box discusses a clinical manifestation of an enzyme malfunction in the pentose phosphate pathway.

If the organism has a greater need for ribose-5-phosphate than for NADPH, fructose-6-phosphate and glyceraldehyde-3-phosphate can give rise to ribose-5-phosphate by the successive operation of the transketolase and transaldolase reactions, bypassing the oxidative portion of the pentose phosphate pathway

## Biochemical Connections ALLIED HEALTH

### The Pentose Phosphate Pathway and Hemolytic Anemia

The pentose phosphate pathway is the only source of NADPH in red blood cells, which, as a result, are highly dependent on the proper functioning of the enzymes involved. A glucose-6-phosphate dehydrogenase deficiency leads to an NADPH deficiency, which can, in turn, lead to *hemolytic anemia* because of wholesale destruction of red blood cells.

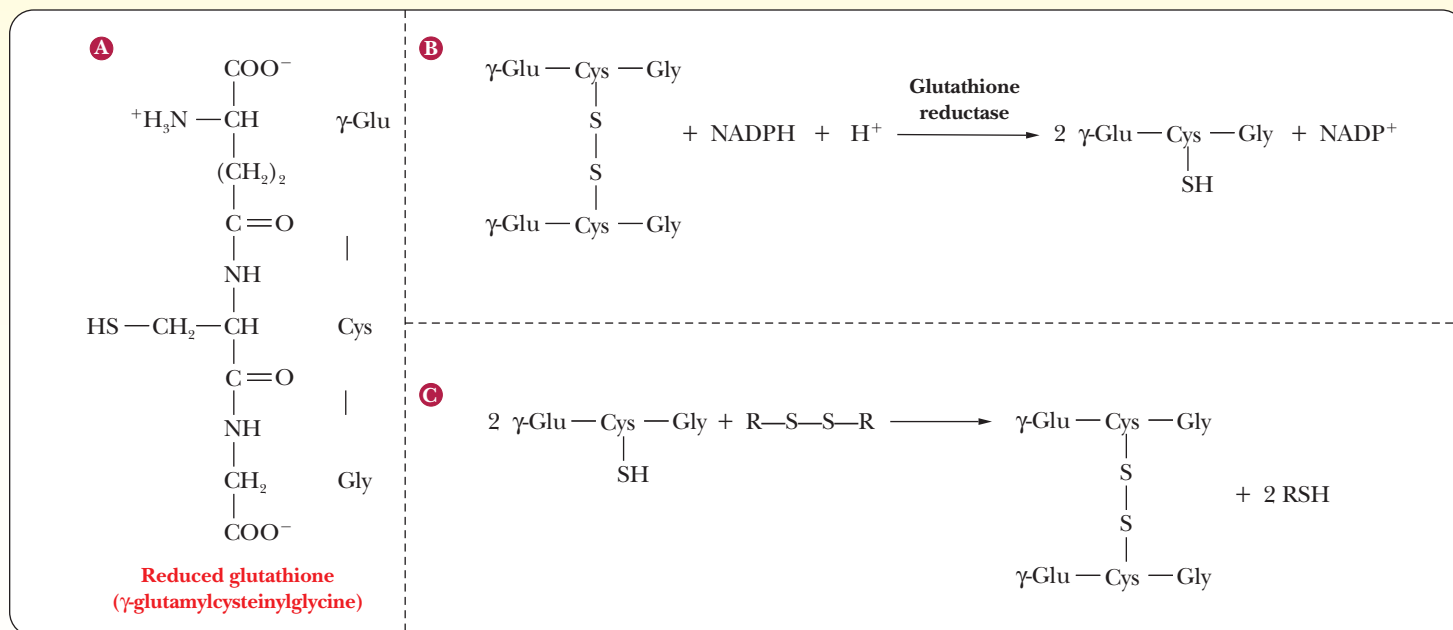
The relationship between NADPH deficiency and anemia is an indirect one. NADPH is required to reduce the peptide glutathione from the disulfide to the free thiol form. Mammalian red blood cells lack mitochondria, which host many redox reactions.

Consequently, these cells are limited in the ways in which they can deal with redox balance. A substance like glutathione, which

can take part in redox reactions, assumes greater importance than would be the case in cells with large numbers of mitochondria. The presence of the reduced form of glutathione is necessary for the maintenance of the sulfhydryl groups of hemoglobin and other proteins in their reduced forms, as well as for keeping the Fe(II) of hemoglobin in its reduced form.

Glutathione also maintains the integrity of red cells by reacting with peroxides that would otherwise degrade fatty-acid side chains in the cell membrane. About 11% of African-Americans are affected by glucose-6-phosphate dehydrogenase deficiency.

This condition, like the sickle-cell trait, leads to increased resistance to malaria, accounting for some of its persistence in the gene pool in spite of its otherwise deleterious consequences.



■ **Glutathione and its reactions.** (a) The structure of glutathione. (b) The role of NADPH in the production of glutathione. (c) The role of glutathione in maintaining the reduced form of protein sulfhydryl groups.

(follow the red shaded path down to glyceraldehyde-3-phosphate and then go up to ribose-5-phosphate) (Figure 18.16). The reactions catalyzed by transketolase and transaldolase are reversible, and this fact plays an important role in the organism's ability to adjust its metabolism to changes in conditions. We shall now look at the mode of action of these two enzymes.

Transaldolase has many features in common with the enzyme aldolase, which we met in the glycolytic pathway. Both an aldol cleavage and an aldol condensation occur at different stages of the reaction. We already saw the mechanism of aldol cleavage, involving the formation of a Schiff base, when we discussed the aldolase reaction in glycolysis, and we need not discuss this point further.

Transketolase resembles pyruvate decarboxylase, the enzyme that converts pyruvate to acetaldehyde (Section 17.4), in that it also requires  $\text{Mg}^{2+}$  and thiamine pyrophosphate (TPP). As in the pyruvate decarboxylase reaction, a carbanion plays a crucial role in the reaction mechanism, which is similar to that of the conversion of pyruvate to acetaldehyde.

## SUMMARY

**How does the breakdown of glycogen take place?** Glycogen can readily be broken down to glucose in response to energy needs. Glycogen phosphorylase uses phosphate to break an  $\alpha(1 \rightarrow 4)$  linkage, yielding glucose-1-phosphate and a glycogen molecule shorter by one glucose. Debranching enzyme aids in the degradation of the molecule around the  $\alpha(1 \rightarrow 6)$  linkages.

**How is glycogen formed from glucose?** When an organism has an available supply of extra glucose, more than is immediately needed as a source of energy extracted in glycolysis, it forms glycogen, a polymer of glucose. Glycogen synthase catalyzes the reaction between a glycogen molecule and UDP-glucose to add a glucose molecule to the glycogen via an  $\alpha(1 \rightarrow 4)$  linkage. Branching enzyme moves sections of a chain of glucoses so that there are  $\alpha(1 \rightarrow 6)$  branch points.

**How is glycogen metabolism controlled?** Control mechanisms ensure that both formation and breakdown of glycogen are not active simultaneously, a situation that would waste energy.

**Why is oxaloacetate an intermediate in gluconeogenesis?** The conversion of pyruvate (the product of glycolysis) to glucose takes place by a process called gluconeogenesis, which is not the exact reversal of glycolysis. Glycolysis involves three irreversible steps. One of these irreversible steps is the conversion of phosphoenolpyruvate to pyruvate. It is favorable to convert pyruvate to oxaloacetate to facilitate the conversion to phosphoenolpyruvate.

**What is the role of sugar phosphates in gluconeogenesis?** The hydrolysis of sugar phosphates is energetically favorable, so these steps have the effect of reversing the early, energy-requiring steps of glycolysis.

**How does control of key enzymes control carbohydrate metabolism?** Glycogen synthase and glycogen phosphorylase are

reciprocally controlled by phosphorylation. Glycolysis and gluconeogenesis are controlled at several points, with phosphofructokinase and fructose biphosphatase being the most important.

**How do different organs share carbohydrate metabolism?** In the same cell, glycolysis and gluconeogenesis are not highly active simultaneously. When the cell needs ATP, glycolysis is more active; when there is little need for ATP, gluconeogenesis is more active. Glycolysis and gluconeogenesis play roles in the Cori cycle. The division of labor between liver and muscle allows glycolysis and gluconeogenesis to take place in different organs to serve the needs of an organism.

**What roles do the first and last steps of glycolysis play in control of carbohydrate metabolism?** Hexokinase and pyruvate kinase, the enzymes that catalyze the first and last steps, respectively, in glycolysis are also important control points. They have the effect of slowing down the pathway when energy is not needed and speeding it up when there is a need.


**What are the oxidative reactions of the pentose phosphate pathway?** The pentose phosphate pathway is an alternative pathway for glucose metabolism. In this pathway five-carbon sugars, including ribose, are produced from glucose. In the oxidative reactions of the pathway, NADPH is also produced.

**What are the nonoxidative reactions of the pentose phosphate pathway, and why are they important?** The nonoxidative reactions of the pentose phosphate pathway produce five-carbon sugars, particularly ribose. They are important when an organism has less need for NADPH but needs the sugars.

**How is the pentose phosphate pathway controlled?** Control of the pathway allows the organism to adjust the relative levels of production of five-carbon sugars and of NADPH according to its needs.



## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

## 18.1 How Glycogen Is Produced and Degraded

- Recall** Why is it essential that the mechanisms that activate glycogen synthesis also deactivate glycogen phosphorylase?
- Recall** How does phosphorylase differ from hydrolysis?
- Recall** Why is it advantageous that breakdown of glycogen gives rise to glucose-6-phosphate rather than to glucose?
- Recall** Briefly outline the role of UDPG in glycogen biosynthesis.
- Recall** Name two control mechanisms that play a role in glycogen biosynthesis. Give an example of each.
- Reflect and Apply** Does the net gain of ATP in glycolysis differ when glycogen, rather than glucose, is the starting material? If so, what is the change?
- Reflect and Apply** In metabolism, glucose-6-phosphate (G6P) can be used for glycogen synthesis or for glycolysis, among other fates. What does it cost, in terms of ATP equivalents, to store G6P as glycogen, rather than to use it for energy in glycolysis? *Hint:* The branched structure of glycogen leads to 90% of glucose residues being released as glucose-1-phosphate and 10% as glucose.
- Reflect and Apply** How does the cost of storing glucose-6-phosphate (G6P) as glycogen differ from the answer you obtained in Question 7 if G6P were used for energy in aerobic metabolism?
- Biochemical Connections** You are planning to go on a strenuous hike and are advised to eat plenty of high-carbohydrate foods, such as bread and pasta, for several days beforehand. Suggest a reason for the advice.
- Biochemical Connections** Would eating candy bars, high in sucrose rather than complex carbohydrates, help build up glycogen stores?
- Biochemical Connections** Would it be advantageous to consume a candy bar with a high refined-sugar content *immediately* before you start the strenuous hike in Question 9?
- Reflect and Apply** The concentration of lactate in blood rises sharply during a sprint and declines slowly for about an hour afterward. What causes the rapid rise in lactate concentration? What causes the decline in lactate concentration after the run?
- Reflect and Apply** A researcher claims to have discovered a variant form of glycogen. The variation is that it has very few branches (every 50 glucose residues or so) and that the branches are only three residues long. Is it likely that this discovery will be confirmed by later work?
- Reflect and Apply** What is the source of the energy needed to incorporate glucose residues into glycogen? How is it used?
- Reflect and Apply** Why is it useful to have a primer in glycogen synthesis?
- Reflect and Apply** Is the glycogen synthase reaction exergonic or endergonic? What is the reason for your answer?
- Reflect and Apply** What is the effect on gluconeogenesis and glycogen synthesis of (a) increasing the level of ATP, (b) decreasing the concentration of fructose-1,6-bisphosphate, and (c) increasing the concentration of fructose-6-phosphate?
- Reflect and Apply** Briefly describe “going for the burn” in a workout in terms of the material in this chapter.
- Reflect and Apply** Suggest a reason why sugar nucleotides, such as UDPG, play a role in glycogen synthesis, rather than sugar phosphates, such as glucose-6-phosphate.

## 18.2 Gluconeogenesis Produces Glucose from Pyruvate

- Recall** What reactions in this chapter require acetyl-CoA or biotin?
- Recall** Which steps of glycolysis are irreversible? What bearing does this observation have on the reactions in which gluconeogenesis differs from glycolysis?

- Recall** What is the role of biotin in gluconeogenesis?
- Recall** How does the role of glucose-6-phosphate in gluconeogenesis differ from that in glycolysis?
- Reflect and Apply** Avidin, a protein found in egg whites, binds to biotin so strongly that it inhibits enzymes that require biotin. What is the effect of avidin on glycogen formation? On gluconeogenesis? On the pentose phosphate pathway?
- Reflect and Apply** How does the hydrolysis of fructose-1,6-bisphosphate bring about the reversal of one of the physiologically irreversible steps of glycolysis?

## 18.3 Control of Carbohydrate Metabolism

- Recall** Which reaction or reactions discussed in this chapter require ATP? Which reaction or reactions produce ATP? List the enzymes that catalyze the reactions that require and that produce ATP.
- Recall** How does fructose-2,6-bisphosphate play a role as an allosteric effector?
- Recall** How do glucokinase and hexokinase differ in function?
- Recall** What is the Cori cycle?
- Reflect and Apply** Earlier biochemists called substrate cycles “futile cycles.” Why might they have chosen such a name? Why is it something of a misnomer?
- Reflect and Apply** Why is it advantageous for two control mechanisms—allosteric control and covalent modification—to be involved in the metabolism of glycogen?
- Reflect and Apply** How can different time scales for response be achieved in control mechanisms?
- Reflect and Apply** How do the control mechanisms in glycogen metabolism lead to amplification of response to a stimulus?
- Reflect and Apply** Why would you expect to see that reactions of substrate cycles involve different enzymes for different directions?
- Reflect and Apply** Suggest a reason or reasons why the Cori cycle takes place in the liver and in muscle.
- Reflect and Apply** Explain how fructose-2,6-bisphosphate can play a role in more than one metabolic pathway.
- Reflect and Apply** How can the synthesis and breakdown of fructose-2,6-bisphosphate be controlled independently?
- Reflect and Apply** How is it advantageous for animals to convert ingested starch to glucose and then to incorporate the glucose into glycogen?

## 18.4 Glucose Is Sometimes Diverted through the Pentose Phosphate Pathway

- Recall** List three differences in structure or function between NADH and NADPH.
- Recall** What are four possible metabolic fates of glucose-6-phosphate?
- Biochemical Connections** What is the connection between material in this chapter and hemolytic anemia?
- Recall** Show how the pentose phosphate pathway, which is connected to the glycolytic pathway, can do the following.
  - Make both NADPH and pentose phosphates, in roughly equal amounts
  - Make mostly or only NADPH
  - Make mostly or only pentose phosphates
- Recall** What is a major difference between transketolase and transaldolase?
- Biochemical Connections** List two ways in which glutathione functions in red blood cells.



45. **Recall** Does thiamine pyrophosphate play a role in the reactions of the pentose phosphate pathway? If so, what is that role?
46. **Reflect and Apply** Using the Lewis electron-dot notation, show explicitly the transfer of electrons in the following redox reaction.



The lactone is a cyclic ester that is an intermediate in the production of 6-phosphogluconate.

47. **Reflect and Apply** Suggest a reason why a different reducing agent (NADPH) is used in anabolic reactions rather than NADH, which plays a role in catabolic ones.

48. **Reflect and Apply** Explain how the pentose phosphate pathway can respond to a cell's need for ATP, NADPH, and ribose-5-phosphate.
49. **Reflect and Apply** Why is it reasonable to expect that glucose-6-phosphate will be oxidized to a lactone (see Question 46) rather than to an open-chain compound?
50. **Reflect and Apply** How would it affect the reactions of the pentose phosphate pathway to have an epimerase and not an isomerase to catalyze the reshuffling reactions?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# The Citric Acid Cycle

The citric acid cycle is the central pathway of metabolism. It plays a pivotal role in the production of energy by a cell.

## 19.1 The Central Role of the Citric Acid Cycle in Metabolism

The evolution of aerobic metabolism, by which nutrients are oxidized to carbon dioxide and water, was an important step in the history of life on the Earth. Organisms can obtain far more energy from nutrients by aerobic oxidation than by anaerobic oxidation. (Even yeast—which is usually thought of in terms of the anaerobic reactions of alcoholic fermentation and is responsible for producing bread, beer, and wine—uses the citric acid cycle and aerobically degrades glucose to carbon dioxide and water.) We saw in Chapter 17 that glycolysis produces only two molecules of ATP for each molecule of glucose metabolized. In this chapter and the next, we shall see how 30 to 32 molecules of ATP can be produced from each molecule of glucose in complete aerobic oxidation to carbon dioxide and water. Three processes play roles in aerobic metabolism: the **citric acid cycle**, which we discuss in this chapter, and **electron transport** and **oxidative phosphorylation**, both of which we shall discuss in Chapter 20 (Figure 19.1). These three processes operate together in aerobic metabolism; separate discussion is a matter of convenience only.

Metabolism consists of catabolism, which is the oxidative breakdown of nutrients, and anabolism, which is reductive synthesis of biomolecules. The citric acid cycle is **amphibolic**, meaning that it plays a role in both catabolism and anabolism. Although the citric acid cycle is a part of the pathway of aerobic oxidation of nutrients (a catabolic pathway; see Section 19.7), some of the molecules that are included in this cycle are the starting points of biosynthetic (anabolic) pathways (see Section 19.8). Metabolic pathways operate simultaneously, even though we talk about them separately. We should always keep this point in mind.

The citric acid cycle has two other common names. One is the *Krebs cycle*, after Sir Hans Krebs, who first investigated the pathway (work for which he received a Nobel Prize in 1953). The other name is the *tricarboxylic acid cycle* (or *TCA cycle*), from the fact that some of the molecules involved are acids with three carboxyl groups. We shall start our discussion with a general overview of the pathway and then go on to discuss specific reactions.

## 19.2 The Overall Pathway of the Citric Acid Cycle

An important difference between glycolysis and the citric acid cycle is the part of the cell in which these pathways occur. In eukaryotes, glycolysis occurs in the cytosol, while the citric acid cycle takes place in mitochondria. Most of the enzymes of the citric acid cycle are present in the mitochondrial matrix.

## Chapter Outline

### 19.1 The Central Role of the Citric Acid Cycle in Metabolism

### 19.2 The Overall Pathway of the Citric Acid Cycle

- Where does the citric acid cycle take place in the cell?
- What are the key features of the citric acid cycle?

### 19.3 How Pyruvate Is Converted to Acetyl-CoA

- How many enzymes are needed to convert pyruvate to acetyl-CoA?

### 19.4 The Individual Reactions of the Citric Acid Cycle

### 19.5 Energetics and Control of the Citric Acid Cycle

- How does the pyruvate dehydrogenase reaction control the citric acid cycle?
- How is control exerted within the citric acid cycle?

### 19.6 The Glyoxylate Cycle: A Related Pathway

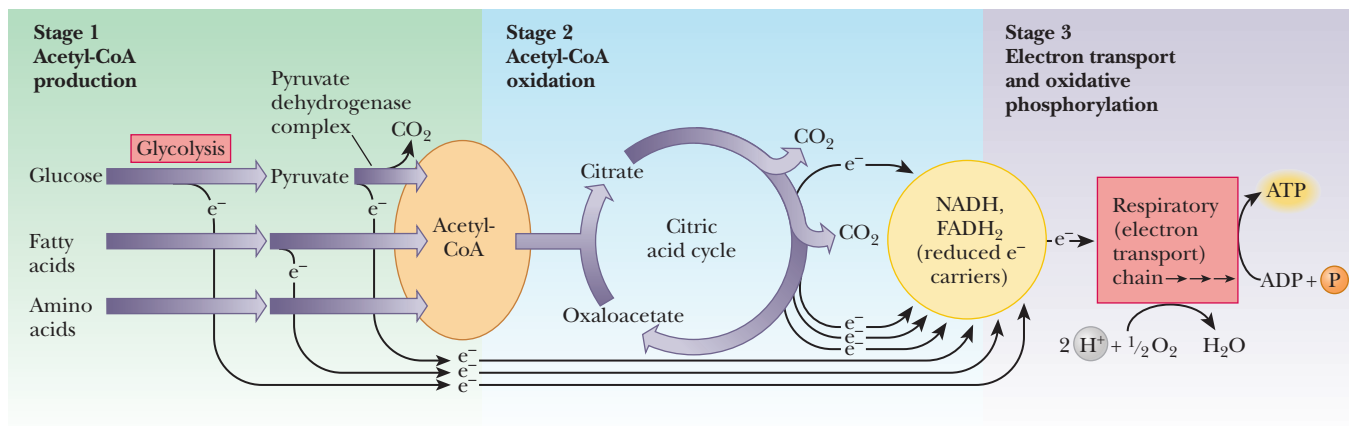
### 19.7 The Citric Acid Cycle in Catabolism

### 19.8 The Citric Acid Cycle in Anabolism

- How is lipid anabolism related to the citric acid cycle?
- How is amino acid metabolism related to the citric acid cycle?

### 19.9 The Link to Oxygen

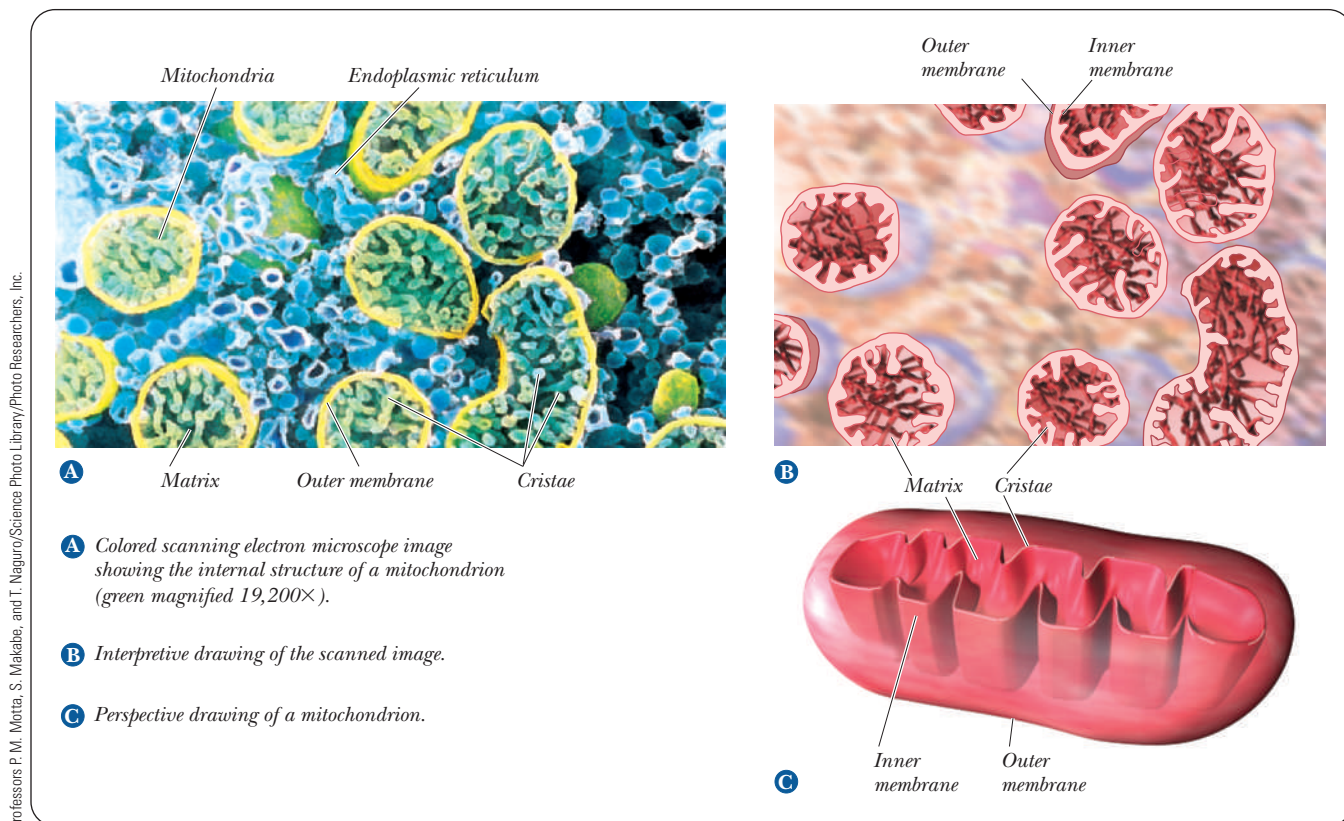
**OWL** Online homework for this chapter may be assigned in OWL.



■ **FIGURE 19.1** The central relationship of the citric acid cycle to catabolism. Amino acids, fatty acids, and glucose can all produce acetyl-CoA in stage 1 of catabolism. In stage 2, acetyl-CoA enters the citric acid cycle. Stages 1 and 2 produce reduced electron carriers (shown here as  $\text{e}^-$ ). In stage 3, the electrons enter the electron transport chain, which then produces ATP.

### Where does the citric acid cycle take place in the cell?

A quick review of some aspects of mitochondrial structure is in order here because we shall want to describe the exact location of each of the components of the citric acid cycle and the electron transport chain. Recall from Chapter 1 that a mitochondrion has an inner and an outer membrane (Figure 19.2). The region enclosed by the inner membrane is called the **mitochondrial matrix**, and an **intermembrane space** exists between the inner and outer membranes.



■ **FIGURE 19.2** The structure of a mitochondrion. (For an electron micrograph of mitochondrial structure, see Figure 1.13.)

The inner membrane is a tight barrier between the matrix and the cytosol, and very few compounds can cross this barrier without a specific transport protein (Section 8.4). The reactions of the citric acid cycle take place in the matrix, except for the one in which the intermediate electron acceptor is FAD. The enzyme that catalyzes the FAD-linked reaction is an integral part of the inner mitochondrial membrane and is linked directly to the electron transport chain (Chapter 20).

### What are the key features of the citric acid cycle?

The citric acid cycle is shown in schematic form in Figure 19.3. Under aerobic conditions, pyruvate produced by glycolysis is oxidized further, with carbon dioxide and water as the final products. First, the pyruvate is oxidized to one carbon dioxide molecule and to one acetyl group, which becomes linked to an intermediate, coenzyme A (CoA) (Section 15.7). The acetyl-CoA enters the citric acid cycle. In the citric acid cycle, two more molecules of carbon dioxide are produced for each molecule of acetyl-CoA that enters the cycle, and electrons are transferred in the process. The immediate electron acceptor in all cases but one is NAD, which is reduced to NADH. In the one case in which there is another intermediate electron acceptor, FAD (flavin adenine dinucleotide), which is derived from riboflavin (vitamin B<sub>2</sub>), takes up two electrons and two hydrogen ions to produce FADH<sub>2</sub>. The electrons are passed from NADH and FADH<sub>2</sub> through several stages of an electron transport chain with a different redox reaction at each step. The final electron acceptor is oxygen, with water as the product. Note that, starting from pyruvate, a three-carbon compound, three carbons are lost as CO<sub>2</sub> via the production of acetyl-CoA and one turn of the cycle. The cycle produces energy in the form of reduced electron equivalents (the NADH and FADH<sub>2</sub> that will enter the electron transport chain), but the carbon skeletons are effectively lost. The cycle also produces one high-energy compound directly, GTP (guanosine triphosphate).

In the first reaction of the cycle, the two-carbon acetyl group condenses with the four-carbon oxaloacetate ion to produce the six-carbon citrate ion. In the next few steps, the citrate isomerizes, and then it both loses carbon dioxide and is oxidized. This process, called **oxidative decarboxylation**, produces the five-carbon compound  $\alpha$ -ketoglutarate, which again is oxidatively decarboxylated to produce the four-carbon compound succinate. The cycle is completed by regeneration of oxaloacetate from succinate in several steps. We shall see many of these intermediates again in other pathways, especially  $\alpha$ -ketoglutarate, which is very important in amino acid and protein metabolism.

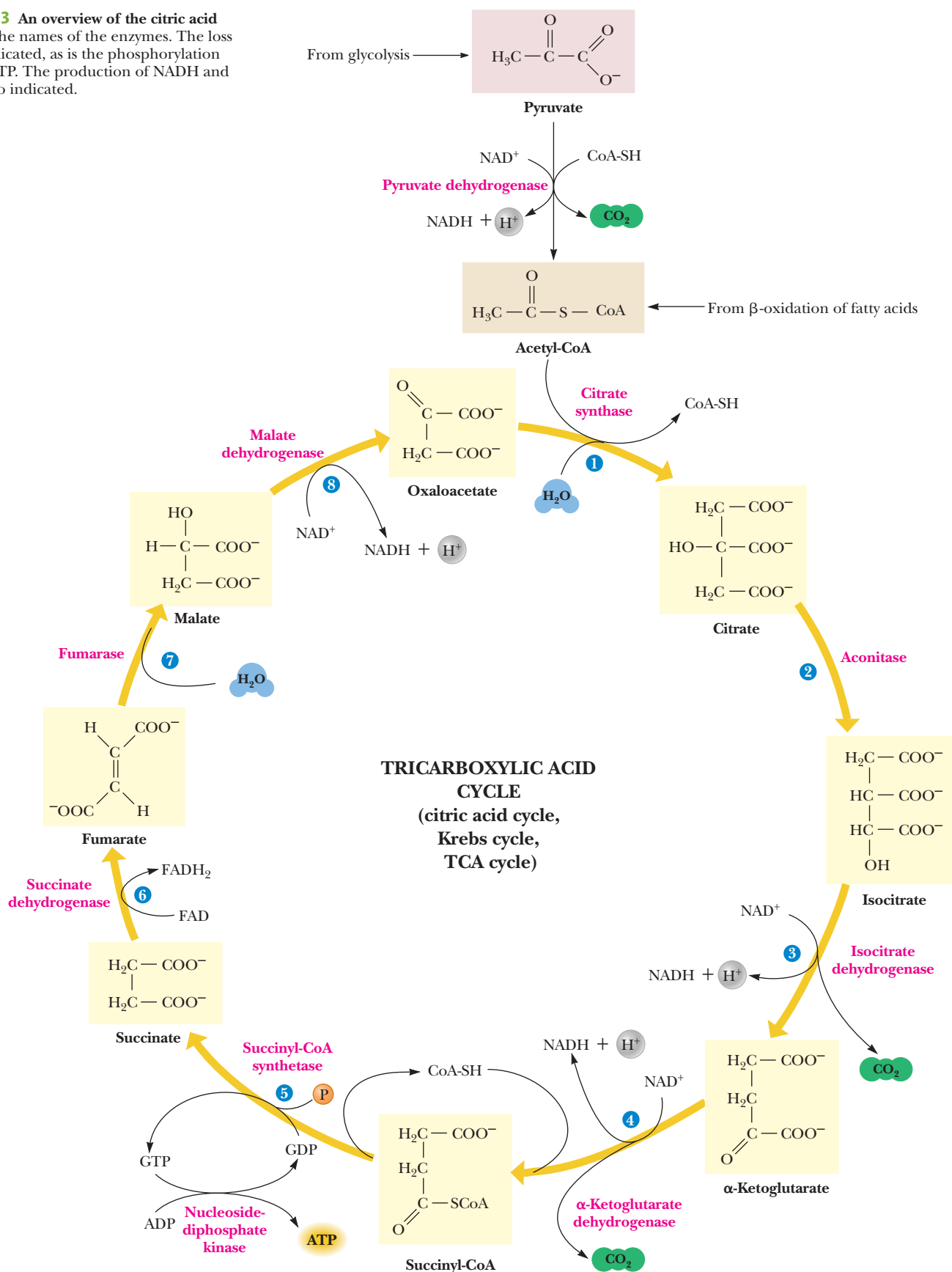
The citric acid cycle has eight steps, each catalyzed by a different enzyme. Four of the eight steps—Steps 3, 4, 6, and 8—are oxidation reactions (see Figure 19.3). The oxidizing agent is NAD<sup>+</sup> in all except Step 6, in which FAD plays the same role. In Step 5, a molecule of GDP (guanosine diphosphate) is phosphorylated to produce GTP. This reaction is equivalent to the production of ATP because the phosphate group is easily transferred to ADP, producing GDP and ATP.

## 19.3 How Pyruvate Is Converted to Acetyl-CoA

Pyruvate can come from several sources, including glycolysis, as we have seen. It moves from the cytosol into the mitochondrion via a specific transporter. There, an enzyme system called the **pyruvate dehydrogenase complex** is responsible for the conversion of pyruvate to carbon dioxide and the acetyl portion of acetyl-CoA. There is an —SH group at one end of the CoA molecule, which is the point at which the acetyl group is attached. As a result, CoA is frequently shown in equations as CoA-SH. Because CoA is a thiol (the sulfur [thio] analog



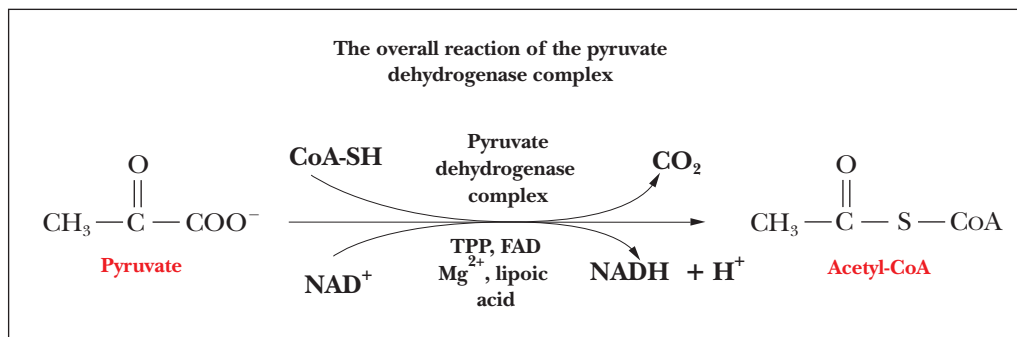
■ **FIGURE 19.3** An overview of the citric acid cycle. Note the names of the enzymes. The loss of  $\text{CO}_2$  is indicated, as is the phosphorylation of GDP to GTP. The production of NADH and  $\text{FADH}_2$  is also indicated.



of an alcohol), acetyl-CoA is a **thioester**, with a sulfur atom replacing an oxygen of the usual carboxylic ester. This difference is important, since thioesters are high-energy compounds (Chapter 15). In other words, the hydrolysis of thioesters releases enough energy to drive other reactions. An oxidation reaction precedes the transfer of the acetyl group to the CoA. The whole process involves several enzymes, all of which are part of the pyruvate dehydrogenase complex. The overall reaction



is exergonic ( $\Delta G^\circ = -33.4 \text{ kJ mol}^{-1} = -8.0 \text{ kcal mol}^{-1}$ ), and NADH can then be used to generate ATP via the electron transport chain (Chapter 20).



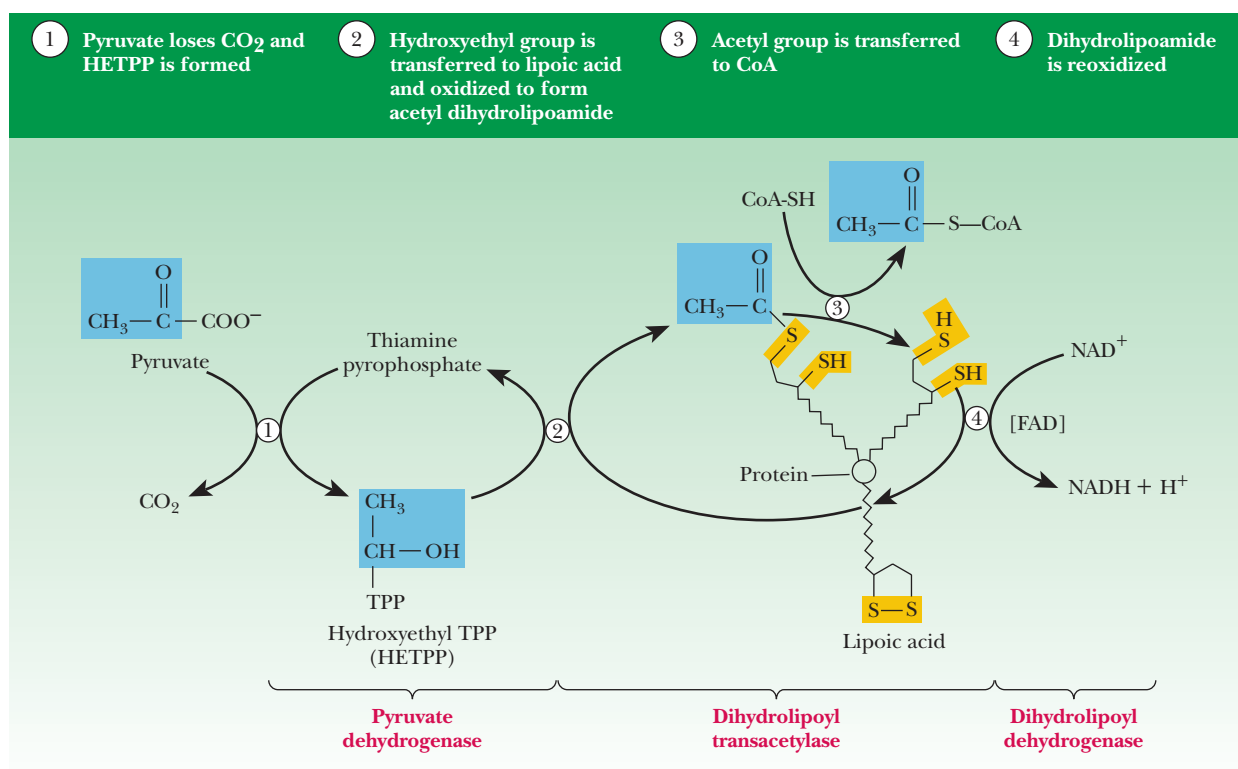
### How many enzymes are needed to convert pyruvate to acetyl-CoA?

Five enzymes make up the pyruvate dehydrogenase complex in mammals. They are *pyruvate dehydrogenase (PDH)*, *dihydrolipoyl transacetylase*, *dihydrolipoyl dehydrogenase*, *pyruvate dehydrogenase kinase*, and *pyruvate dehydrogenase phosphatase*. The first three are involved in the conversion of pyruvate to acetyl-CoA. The kinase and the phosphatase are enzymes used in the control of PDH (Section 19.5) and are present on a single polypeptide. The reaction takes place in five steps. Two enzymes catalyze reactions of *lipoic acid*, a compound that has a disulfide group in its oxidized form and two sulfhydryl groups in its reduced form.

Lipoic acid differs in one respect from other coenzymes. It is a vitamin, rather than a metabolite of a vitamin, as is the case with many other coenzymes (Table 7.3). (The classification of lipoic acid as a vitamin is open to question. There is no evidence of a requirement for it in the human diet, but it is required for the growth of some bacteria and protists.) Lipoic acid can act as an oxidizing agent; the reaction involves hydrogen transfer, which frequently accompanies biological oxidation–reduction reactions (Section 15.5). Another reaction of lipoic acid is the formation of a thioester linkage with the acetyl group before it is transferred to the acetyl-CoA. Lipoic acid can act simply as an oxidizing agent, or it can simultaneously take part in two reactions—a redox reaction and the shift of an acetyl group by transesterification.

The first step in the reaction sequence that converts pyruvate to carbon dioxide and acetyl-CoA is catalyzed by pyruvate dehydrogenase, as shown in Figure 19.4. This enzyme requires thiamine pyrophosphate (TPP; a metabolite of vitamin B<sub>1</sub>, or thiamine) as a coenzyme. The coenzyme is not covalently bonded to the enzyme; they are held together by noncovalent interactions. Mg<sup>2+</sup> is also required. We saw the action of TPP as a coenzyme in the conversion of pyruvate to acetaldehyde, catalyzed by pyruvate decarboxylase (Section 17.4). In the pyruvate dehydrogenase reaction, an  $\alpha$ -keto acid, pyruvate, loses carbon dioxide; the remaining two-carbon unit becomes covalently bonded to TPP.

The second step of the reaction is catalyzed by dihydrolipoyl transacetylase. This enzyme requires lipoic acid as a coenzyme. The lipoic acid is covalently



■ **FIGURE 19.4** The mechanism of the pyruvate dehydrogenase reaction. Decarboxylation of pyruvate occurs with formation of hydroxyethyl-TPP (Step 1). Transfer of the two-carbon unit to lipoic acid in Step 2 is followed by formation of acetyl-CoA in Step 3. Lipoic acid is reoxidized in Step 4 of the reaction.

bonded to the enzyme by an amide bond to the  $\epsilon$ -amino group of a lysine side chain. The two-carbon unit that originally came from pyruvate is transferred from the thiamine pyrophosphate to the lipoic acid, and, in the process, a hydroxyl group is oxidized to produce an acetyl group. The disulfide group of the lipoic acid is the oxidizing agent, which is itself reduced, and the product of the reaction is a thioester. In other words, the acetyl group is now covalently bonded to the lipoic acid by a thioester linkage (see Figure 19.4).

The third step of the reaction is also catalyzed by dihydrolipoyl transacetylase. A molecule of CoA-SH attacks the thioester linkage, and the acetyl group is transferred to it. The acetyl group remains bound in a thioester linkage; this time it appears as acetyl-CoA rather than esterified to lipoic acid. The reduced form of lipoic acid remains covalently bound to dihydrolipoyl transacetylase (see Figure 19.4). The reaction of pyruvate and CoA-SH has now reached the stage of the products, carbon dioxide and acetyl-CoA, but the lipoic acid coenzyme is in a reduced form. The rest of the steps regenerate the lipoic acid, so further reactions can be catalyzed by the transacetylase.

In the fourth step of the overall reaction, the enzyme dihydrolipoyl dehydrogenase reoxidizes the reduced lipoic acid from the sulfhydryl to the disulfide form. The lipoic acid still remains covalently bonded to the transacetylase enzyme. The dehydrogenase also has a coenzyme, FAD (Section 15.5), that is bound to the enzyme by noncovalent interactions. As a result, FAD is reduced to  $\text{FADH}_2$ .  $\text{FADH}_2$  is reoxidized in turn. The oxidizing agent is  $\text{NAD}^+$ , and NADH is the product along with reoxidized FAD. Enzymes such as pyruvate dehydrogenase are called flavoproteins because of their attached FADs.

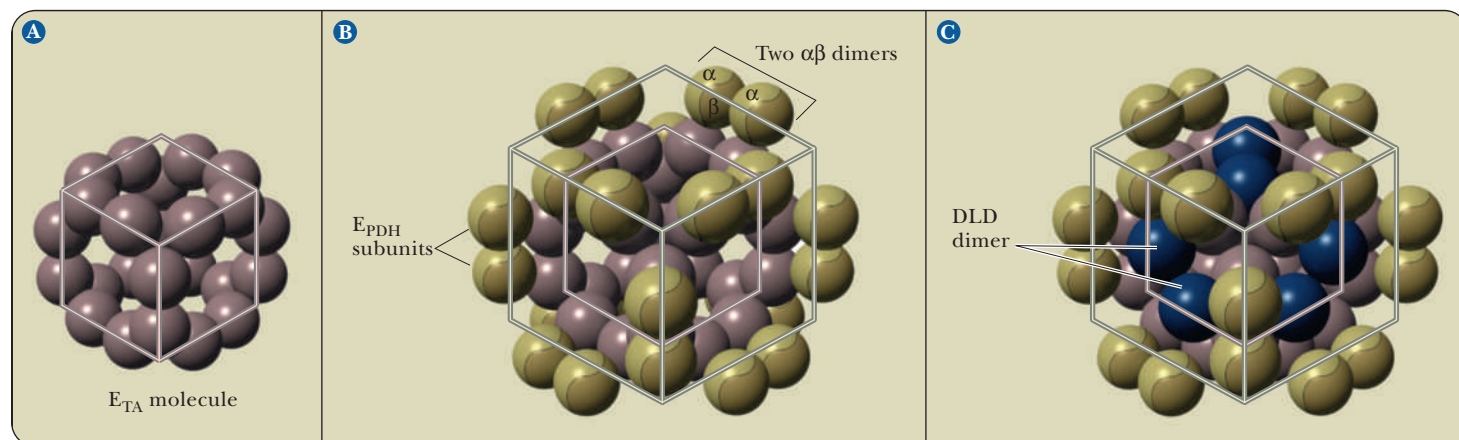
The reduction of  $\text{NAD}^+$  to NADH accompanies the oxidation of pyruvate to the acetyl group, and the overall equation shows that there has been a transfer

of two electrons from pyruvate to  $\text{NAD}^+$  (Equation 19.1). The electrons gained by  $\text{NAD}^+$  in generating NADH in this step are passed to the electron transport chain (the next step in aerobic metabolism). In Chapter 20, we shall see that the transfer of electrons from NADH ultimately to oxygen will give rise to 2.5 ATP. Two molecules of pyruvate are produced for each molecule of glucose, so that there will eventually be five ATP from each glucose from this step alone.

The reaction leading from pyruvate to acetyl-CoA is a complex one that requires three enzymes, each of which has its own coenzyme in addition to  $\text{NAD}^+$ . The spatial orientation of the individual enzyme molecules with respect to one another is itself complex. In the enzyme isolated from *E. coli*, the arrangement is quite compact, so that the various steps of the reaction can be thoroughly coordinated. There is a core of 24 dihydrolipoyl transacetylase molecules. The 24 polypeptide chains are arranged in eight trimers, with each trimer occupying the corner of a cube. There are 12  $\alpha\beta$  dimers of pyruvate dehydrogenase, and they occupy the edges of the cube. Finally, six dimers of dihydrolipoyl dehydrogenase lie on the six faces of the cube (Figure 19.5). Note that many levels of structure combine to produce a suitable environment for the conversion of pyruvate to acetyl-CoA. Each enzyme molecule in this array has its own tertiary structure, and the array itself has the cubical structure we have just seen.

A compact arrangement, such as the one in the pyruvate dehydrogenase multienzyme complex, has two great advantages over an arrangement in which the various components are more widely dispersed. First, the various stages of the reaction can take place more efficiently because the reactants and the enzymes are so close to each other. The role of lipoic acid is particularly important here. Recall that the lipoic acid is covalently attached to the transacetylase enzyme that occupies a central position in the complex. The lipoic acid and the lysine side chain to which it is bonded are long enough to act as a “swinging arm,” which can move to the site of each of the steps of the reaction (Figure 19.4). As a result of the swinging-arm action, the lipoic acid can move to the pyruvate dehydrogenase site to accept the two-carbon unit and then transfer it to the active site of the transacetylase. The acetyl group can then be transferred to CoA-SH from the lipoic acid. Finally, the lipoic acid can swing to the active site of the dehydrogenase so that the sulfhydryl groups can be reoxidized to a disulfide.

A second advantage of a multienzyme complex is that regulatory controls can be applied more efficiently in such a system than in a single enzyme



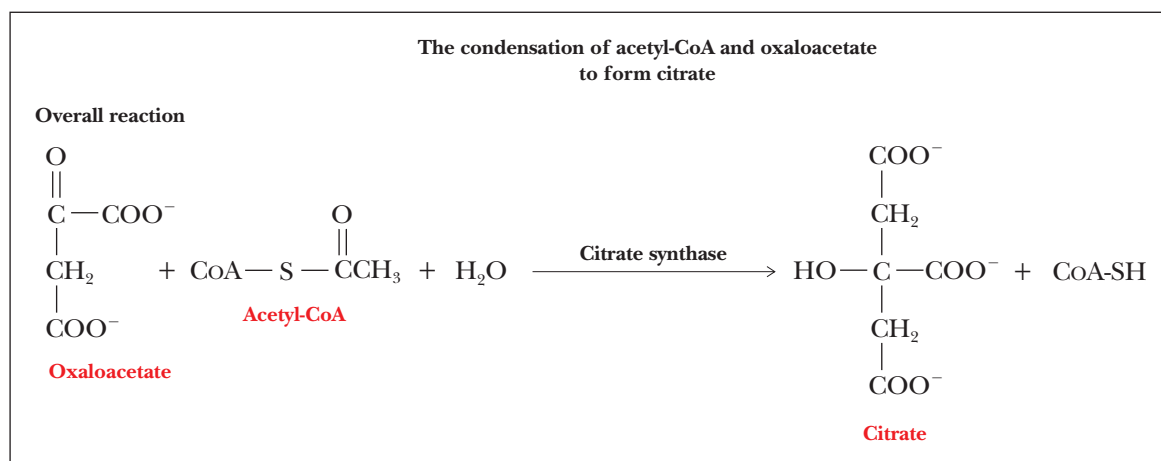
**FIGURE 19.5** The structure of the pyruvate dehydrogenase complex. (a) 24 dihydrolipoyl transacetylase (TA) subunits. (b) 24  $\alpha\beta$  dimers of pyruvate dehydrogenase are added to the cube (two per edge). (c) Addition of 12 dihydrolipoyl dehydrogenase subunits (two per face) completes the complex.

molecule. In the case of the pyruvate dehydrogenase complex, controlling factors are intimately associated with the multienzyme complex itself, which we shall study in Section 19.5.

## 19.4 The Individual Reactions of the Citric Acid Cycle

The reactions of the citric acid cycle proper and the enzymes that catalyze them are listed in Table 19.1. We shall now discuss each of these reactions in turn.

**Step 1. Formation of Citrate** The first step of the citric acid cycle is the reaction of acetyl-CoA and oxaloacetate to form citrate and CoA-SH. This reaction is called a condensation because a new carbon-carbon bond is formed. The condensation reaction of acetyl-CoA and oxaloacetate to form citryl-CoA takes place in the first stage of the reaction. The condensation is followed by the hydrolysis of citryl-CoA to give citrate and CoA-SH.



The reaction is catalyzed by the enzyme **citrate synthase**, originally called “condensing enzyme.” A synthase is an enzyme that makes a new covalent bond during the reaction, but it does not require the direct input of ATP. It is an exergonic reaction ( $\Delta G^{\circ} = -32.8 \text{ kJ mol}^{-1} = -7.8 \text{ kcal mol}^{-1}$ ) because the hydrolysis of a thioester releases energy. Thioesters are considered high-energy compounds.

**Step 2. Isomerization of Citrate to Isocitrate** The second reaction of the citric acid cycle, the one catalyzed by aconitase, is the isomerization of citrate to

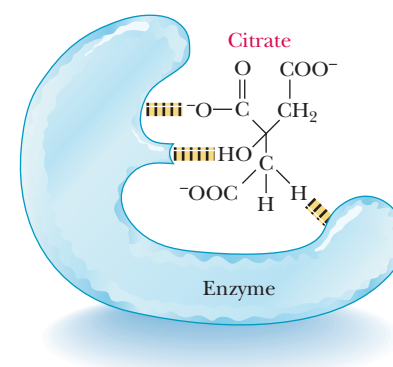
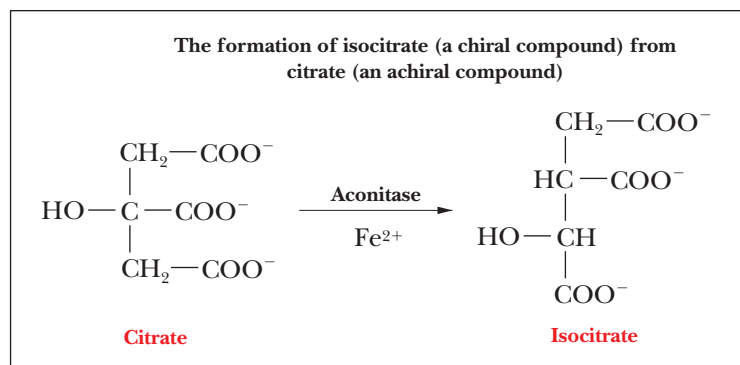
TABLE 19.1

The Reactions of the Citric Acid Cycle		
Step	Reaction	Enzyme
1	Acetyl-CoA + Oxaloacetate + H <sub>2</sub> O → Citrate + CoA-SH	Citrate synthase
2	Citrate → Isocitrate	Aconitase
3	Isocitrate + NAD <sup>+</sup> → α-Ketoglutarate + NADH + CO <sub>2</sub> + H <sup>+</sup>	Isocitrate dehydrogenase
4	α-Ketoglutarate + NAD <sup>+</sup> + CoA-SH → Succinyl-CoA + NADH + CO <sub>2</sub> + H <sup>+</sup>	α-Ketoglutarate dehydrogenase
5	Succinyl-CoA + GDP + P <sub>i</sub> → Succinate + GTP + CoA-SH	Succinyl-CoA synthetase
6	Succinate + FAD → Fumarate + FADH <sub>2</sub>	Succinate dehydrogenase
7	Fumarate + H <sub>2</sub> O → L-Malate	Fumarase
8	L-Malate + NAD <sup>+</sup> → Oxaloacetate + NADH + H <sup>+</sup>	Malate dehydrogenase



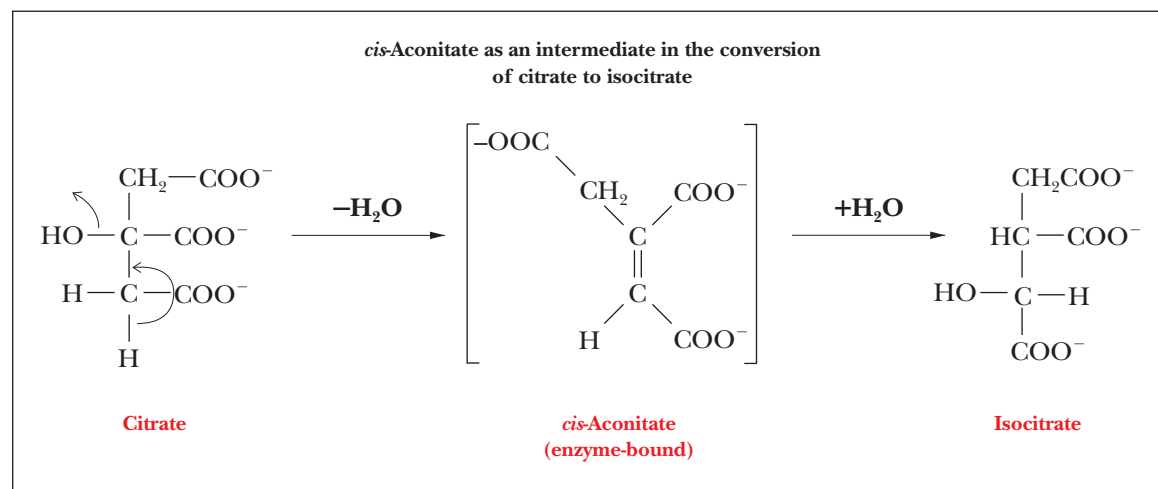
isocitrate. The enzyme requires  $\text{Fe}^{2+}$ . One of the most interesting features of the reaction is that citrate, a symmetrical (achiral) compound, is converted to isocitrate, a chiral compound, a molecule that cannot be superimposed on its mirror image.

It is often possible for a chiral compound to have several different isomers. Isocitrate has four possible isomers, but only one of the four is produced by this reaction. (We shall not discuss nomenclature of the isomers of isocitrate here. See Question 28 at the end of this chapter for a question about the other isomers.) Aconitase, the enzyme that catalyzes the conversion of citrate to isocitrate, can select one end of the citrate molecule in preference to the other.



■ **FIGURE 19.6** Three-point attachment to the enzyme aconitase makes the two  $\text{—CH}_2\text{—COO}^-$  ends of citrate stereochemically nonequivalent.

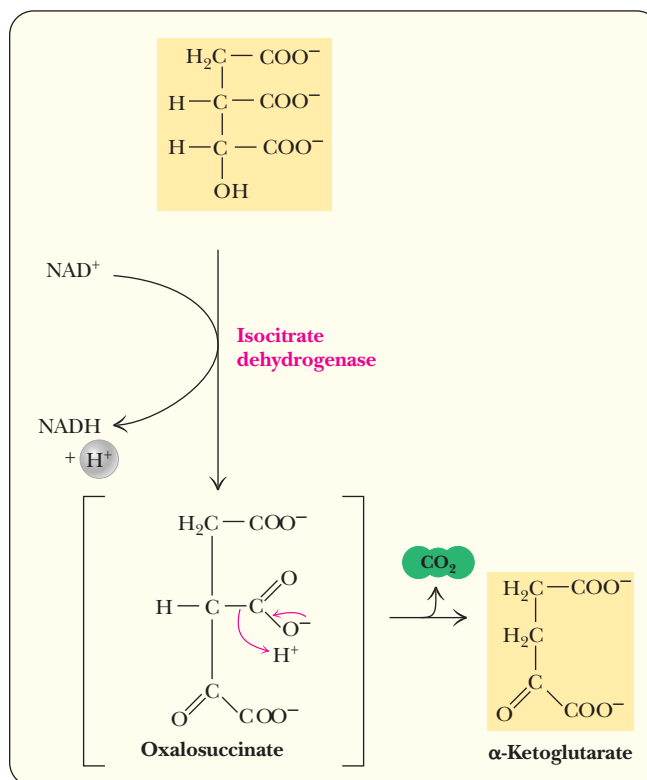
This type of behavior means that the enzyme can bind a symmetrical substrate in an unsymmetrical binding site. In Section 7.6, we mentioned that this possibility exists, and here we have an example of it. The enzyme forms an unsymmetrical three-point attachment to the citrate molecule (Figure 19.6). The reaction proceeds by removal of a water molecule from the citrate to produce *cis*-aconitate, and then water is added back to the *cis*-aconitate to give isocitrate.



The intermediate, *cis*-aconitate, remains bound to the enzyme during the course of the reaction. There is some evidence that the citrate is complexed to the  $\text{Fe(II)}$  in the active site of the enzyme in such a way that the citrate curls back on itself in a nearly circular conformation. Several authors have been unable to resist the temptation to call this situation the “ferrous wheel.”

**Step 3. Formation of  $\alpha$ -Ketoglutarate and  $\text{CO}_2$ —First Oxidation** The third step in the citric acid cycle is the oxidative decarboxylation of isocitrate to  $\alpha$ -ketoglutarate and carbon dioxide. This reaction is the first of two oxidative decarboxylations of the citric acid cycle; the enzyme that catalyzes it is **isocitrate dehydrogenase**. The reaction takes place in two steps (Figure 19.7). First, isocitrate is oxidized

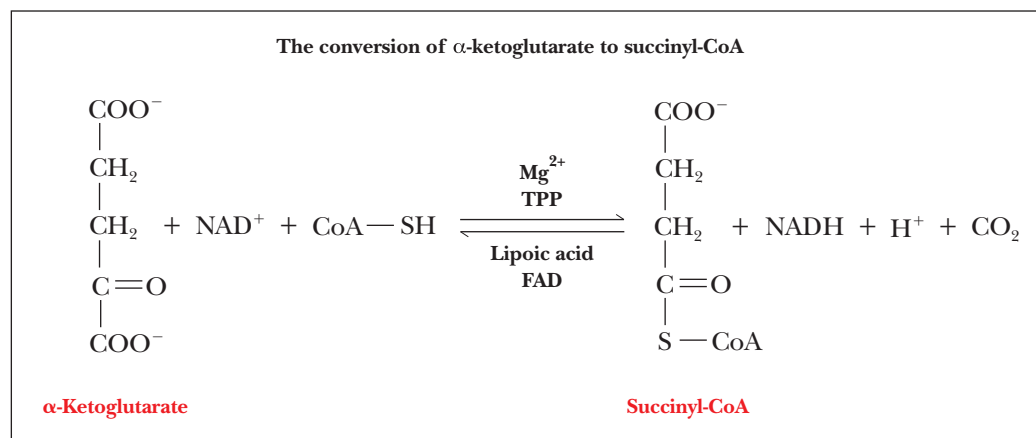
■ **FIGURE 19.7** The isocitrate dehydrogenase reaction.



to oxalosuccinate, which remains bound to the enzyme. Then oxalosuccinate is decarboxylated, and the carbon dioxide and α-ketoglutarate are released.

This is the first of the reactions in which NADH is produced. One molecule of NADH is produced from NAD<sup>+</sup> at this stage by the loss of two electrons in the oxidation. As we saw in our discussion of the pyruvate dehydrogenase complex, each NADH produced leads to the production of 2.5 ATP in later stages of aerobic metabolism. Recall also that there will be two NADH, equivalent to five ATP, for each original molecule of glucose.

**Step 4. Formation of Succinyl-CoA and CO<sub>2</sub>—Second Oxidation** The second oxidative decarboxylation takes place in Step 4 of the citric acid cycle, in which carbon dioxide and succinyl-CoA are formed from α-ketoglutarate and CoA.



This reaction is similar to the one in which acetyl-CoA is formed from pyruvate, with NADH produced from NAD<sup>+</sup>. Once again, each NADH eventually gives rise to 2.5 ATP, with five ATP from each original molecule of glucose.

## Biochemical Connections TOXICOLOGY

### Fluorine Compounds and Carbohydrate Metabolism

Compounds of fluorine are best known to the general public in the context of the heated controversy over the addition of fluoride ion (usually as sodium fluoride) to water supplies with the goal of preventing tooth decay. It is less well known that both ionic and covalent compounds of fluorine play roles in the metabolism of carbohydrates. In ionic form, fluoride ion is known to be an inhibitor of the glycolytic enzymes phosphoglyceromutase, enolase, and pyruvate kinase. Two compounds with covalent bonds, fluoroacetate and fluorocitrate, play a role in inhibition of the citric acid cycle. Of these two, fluoroacetate is noteworthy for its role in plant development; it is a precursor of *fluoroacetyl-CoA*, which is an inhibitor of citrate synthase.

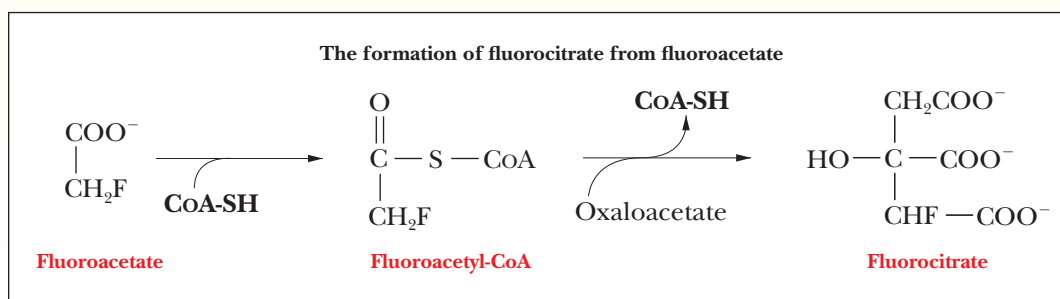
The source of the fluoroacetyl-CoA is fluoroacetate, which is found in the leaves of various types of poisonous plants, including locoweeds. Animals that ingest these plants form fluoroacetyl-CoA, which, in turn, is converted to fluorocitrate by their citrate synthase. Fluorocitrate, in turn, is a potent inhibitor of *aconitase*, the enzyme that catalyzes the next reaction of the citric acid cycle. These plants are poisonous because they produce a potent inhibitor of life processes.

The poison called Compound 1080 (pronounced “ten-eighty”) is sodium fluoroacetate. Ranchers who want to protect

their sheep from attacks by coyotes put the poison just outside the ranch fence. When the coyotes eat this poison, they die. The mechanism of poisoning by Compound 1080 is the same as that by plant poisons.



Stephen Krasemann/Photo Researchers, Inc.



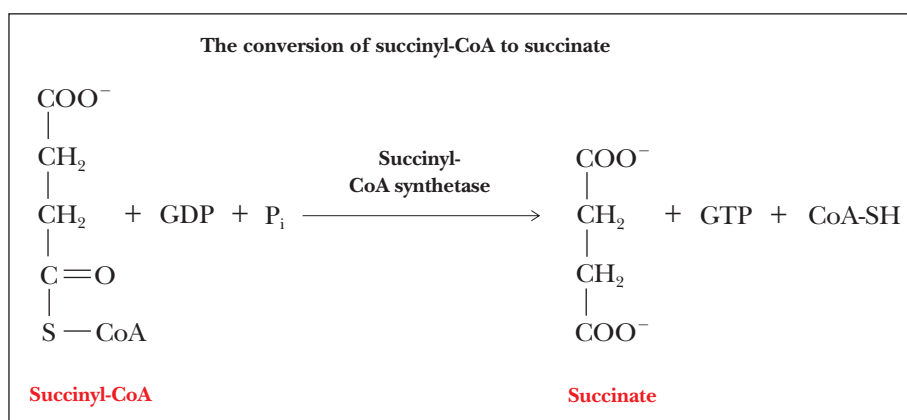
The reaction occurs in several stages and is catalyzed by an enzyme system called the  **$\alpha$ -ketoglutarate dehydrogenase complex**, which is very similar to the pyruvate dehydrogenase complex. Each of these multienzyme systems consists of three enzymes that catalyze the overall reaction. The reaction takes place in several steps, and there is again a requirement for thiamine pyrophosphate (TPP), FAD, lipoic acid, and  $\text{Mg}^{2+}$ . This reaction is highly exergonic ( $\Delta G^{\circ'} = -33.4 \text{ kJ mol}^{-1} = -8.0 \text{ kcal mol}^{-1}$ ), as is the one catalyzed by pyruvate dehydrogenase.

At this point, two molecules of  $\text{CO}_2$  have been produced by the oxidative decarboxylations of the citric acid cycle. Removal of the  $\text{CO}_2$  makes the citric acid cycle irreversible in vivo, although in vitro each separate reaction is reversible. One might suspect that the two molecules of  $\text{CO}_2$  arise from the two carbon atoms of acetyl-CoA. Labeling studies have shown that this is not the case, but a full discussion of this point is beyond the scope of this text.

The two  $\text{CO}_2$  arise from carbon atoms that were part of the oxaloacetate with which the acetyl group condensed. The carbons of this acetyl group are incorporated into the oxaloacetate that will be regenerated for the next round of the cycle. The release of the  $\text{CO}_2$  molecules has a profound influence on mammalian physiology, as will be discussed later in this chapter. We should also

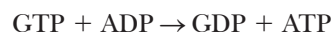
mention that the  $\alpha$ -ketoglutarate dehydrogenase complex reaction is the third one in which we have encountered an enzyme that requires TPP.

**Step 5. Formation of Succinate** In the next step of the cycle, the thioester bond of succinyl-CoA is hydrolyzed to produce succinate and CoA-SH; an accompanying reaction is the phosphorylation of GDP to GTP. The whole reaction is catalyzed by the enzyme **succinyl-CoA synthetase**. A synthetase is an enzyme that creates a new covalent bond and requires the direct input of energy from a high-energy phosphate. Recall that we met a synthase (citrate synthase) earlier. The difference between a synthase and a synthetase is that a synthase does not require energy from phosphate-bond hydrolysis, whereas a synthetase does. In the reaction mechanism, a phosphate group covalently bonded to the enzyme is directly transferred to the GDP. The phosphorylation of GDP to GTP is endergonic, as is the corresponding ADP-to-ATP reaction ( $\Delta G^{\circ} = 30.5 \text{ kJ mol}^{-1} = 7.3 \text{ kcal mol}^{-1}$ ).



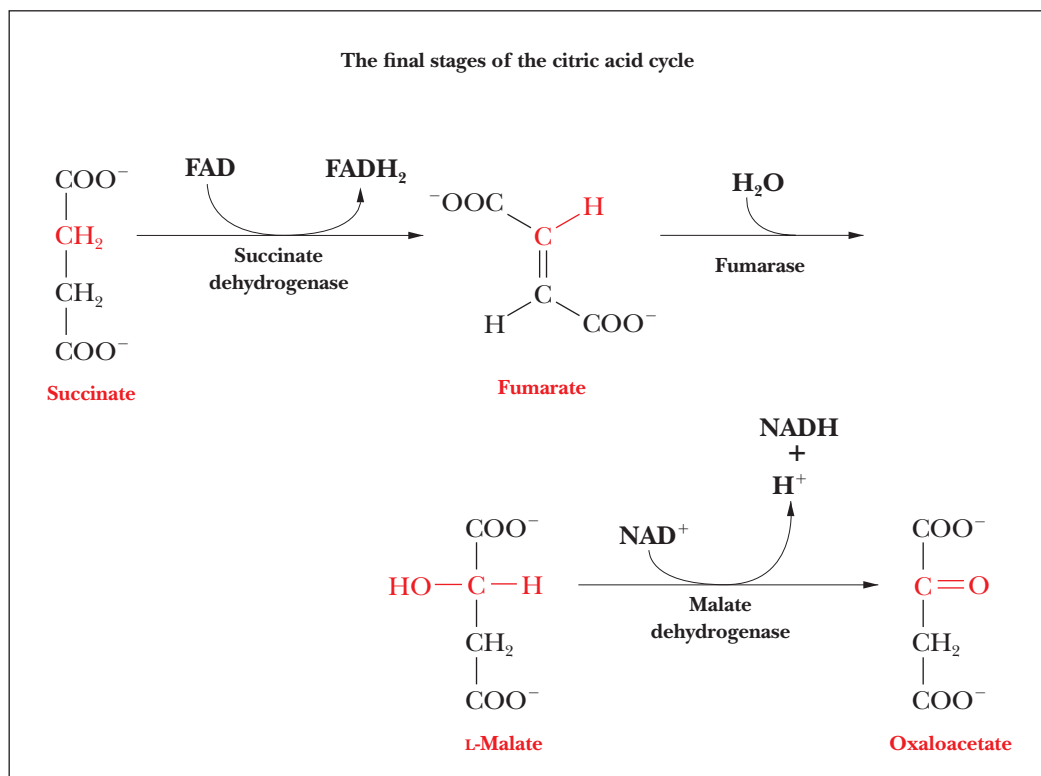
The energy required for the phosphorylation of GDP to GTP is provided by the hydrolysis of succinyl-CoA to produce succinate and CoA. The free energy of hydrolysis ( $\Delta G^{\circ}$ ) of succinyl-CoA is  $-33.4 \text{ kJ mol}^{-1}$  ( $-8.0 \text{ kcal mol}^{-1}$ ). The overall reaction is slightly exergonic ( $\Delta G^{\circ} = -3.3 \text{ kJ mol}^{-1} = -0.8 \text{ kcal mol}^{-1}$ ) and, as a result, does not contribute greatly to the overall production of energy by the mitochondrion. Note that the name of the enzyme describes the reverse reaction. Succinyl-CoA synthetase would produce succinyl-CoA while spending an ATP or another high-energy phosphate. This reaction is the opposite of that.

The enzyme **nucleosidediphosphate kinase** catalyzes the transfer of a phosphate group from GTP to ADP to give GDP and ATP.

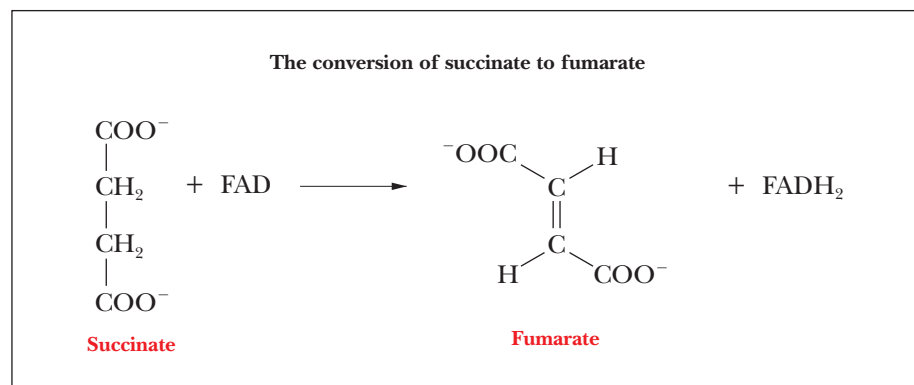


This reaction step is called substrate-level phosphorylation to distinguish it from the type of reaction for production of ATP that is coupled to the electron transport chain. The production of ATP in this reaction is the only place in the citric acid cycle in which chemical energy in the form of ATP is made available to the cell. Except for this reaction, the generation of ATP characteristic of aerobic metabolism is associated with the electron transport chain, the subject of Chapter 20. About 30 to 32 molecules of ATP can be obtained from the oxidation of a single molecule of glucose by the combination of anaerobic and aerobic oxidation, compared with only two molecules of ATP produced by anaerobic glycolysis alone. (This variation in the stoichiometry of ATP produced is the result of differences in metabolic state and mechanisms of transport in different tissues, as will be explained in Chapter 20.) The combined reactions that occur in mitochondria are of great importance to aerobic organisms.

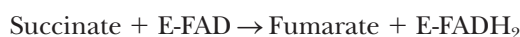
In the next three steps in the citric acid cycle (Steps 6 through 8), the four-carbon succinate ion is converted to oxaloacetate ion to complete the cycle.



**Step 6. Formation of Fumarate—FAD-Linked Oxidation** Succinate is oxidized to fumarate, a reaction that is catalyzed by the enzyme **succinate dehydrogenase**. This enzyme is an integral protein of the inner mitochondrial membrane. We shall have much more to say about the enzymes bound to the inner mitochondrial membrane in Chapter 20. The other individual enzymes of the citric acid cycle are in the mitochondrial matrix. The electron acceptor, which is FAD rather than NAD<sup>+</sup>, is covalently bonded to the enzyme; succinate dehydrogenase is also called a flavoprotein because of the presence of FAD with its flavin moiety. In the succinate dehydrogenase reaction, FAD is reduced to FADH<sub>2</sub> and succinate is oxidized to fumarate.



The overall reaction is

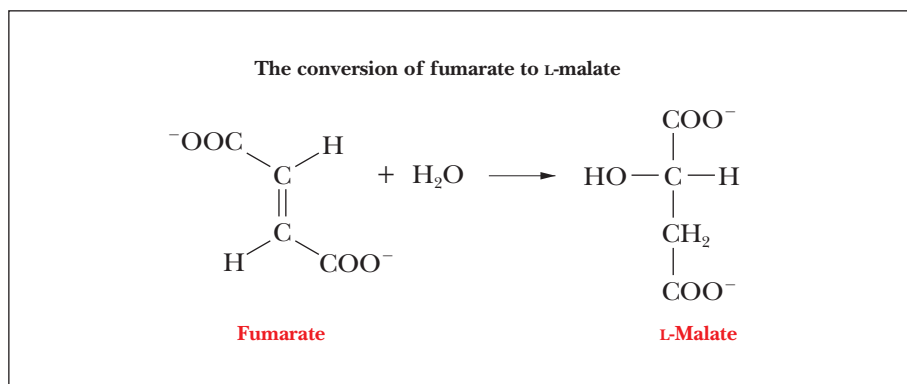


The E-FAD and E-FADH<sub>2</sub> in the equation indicate that the electron acceptor is covalently bonded to the enzyme. The FADH<sub>2</sub> group passes electrons on to the electron transport chain, and eventually to oxygen, and gives rise to 1.5 ATP, rather than 2.5, as is the case with NADH.

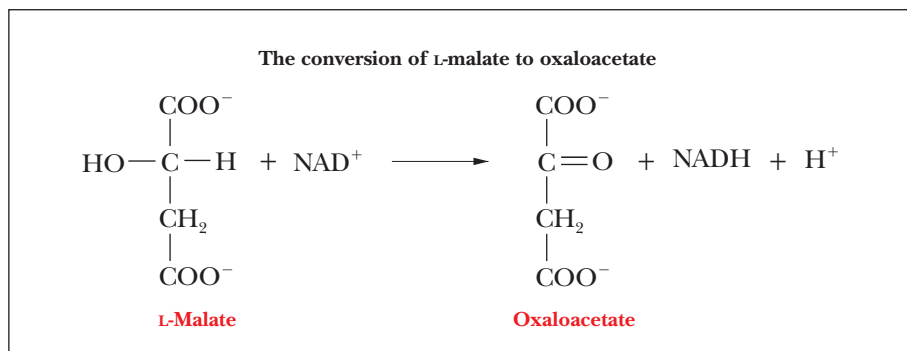


Succinate dehydrogenase contains iron atoms but does not contain a heme group; it is referred to as a **nonheme iron protein** or an *iron-sulfur protein*. The latter name refers to the fact that the protein contains several clusters that consist of four atoms each of iron and of sulfur.

**Step 7. Formation of L-Malate** In Step 7, which is catalyzed by the enzyme **fumarase**, water is added across the double bond of fumarate in a hydration reaction to give malate. Again, there is stereospecificity in the reaction. Malate has two enantiomers, L- and D-malate, but only L-malate is produced.



**Step 8. Regeneration of Oxaloacetate—Final Oxidation Step** Malate is oxidized to oxaloacetate, and another molecule of  $\text{NAD}^+$  is reduced to NADH. This reaction is catalyzed by the enzyme **malate dehydrogenase**. The oxaloacetate can then react with another molecule of acetyl-CoA to start another round of the cycle.

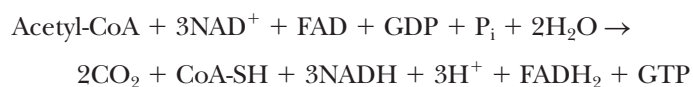


The oxidation of pyruvate by the pyruvate dehydrogenase complex and the citric acid cycle results in the production of three molecules of  $\text{CO}_2$ . As a result of these oxidation reactions, one molecule of GDP is phosphorylated to GTP, one molecule of FAD is reduced to  $\text{FADH}_2$ , and four molecules of  $\text{NAD}^+$  are reduced to NADH. Of the four molecules of NADH produced, three come from the citric acid cycle, and one comes from the reaction of the pyruvate dehydrogenase complex. The overall stoichiometry of the oxidation reactions is the sum of the pyruvate dehydrogenase reaction and the citric acid cycle. Note that only one high-energy phosphate, GTP, is produced *directly* from the citric acid cycle, but many more ATP will arise from reoxidation of NADH and  $\text{FADH}_2$ .

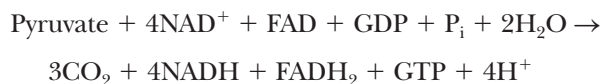
*Pyruvate dehydrogenase complex:*



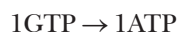
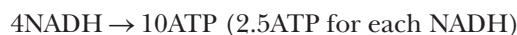
*Citric acid cycle:*



*Overall reaction:*



*Eventual ATP production per pyruvate:*



Total 12.5 ATP per pyruvate or 25 ATP per glucose

There were also two ATP produced per glucose in glycolysis and two NADH, which will give rise to another five ATP (seven more ATP total). In the next chapter, we shall say more about the subject of ATP production from the complete oxidation of glucose.

At this point, we would do well to recapitulate what we have said about the citric acid cycle (see Figure 19.3). When studying a pathway such as this, we might learn many details but also be able to see the big picture. The entire pathway is shown with the enzyme names outside the circle. The most important reactions can be identified by those that have important cofactors (NADH, FADH<sub>2</sub>, GTP). Also important are the steps where CO<sub>2</sub> is given off.

These important reactions also play a large role in the cycle's contribution to our metabolism. One purpose of the cycle is to produce energy. It does that by producing GTP directly and by producing reduced electron carriers (NADH and FADH<sub>2</sub>). The three decarboxylations mean that for every three carbons entering as pyruvate, three carbons are effectively lost during the cycle, a fact that has many implications to our metabolism, as we shall see later in the chapter.

## 19.5 Energetics and Control of the Citric Acid Cycle

The reaction of pyruvate to acetyl-CoA is exergonic, as we have seen ( $\Delta G^\circ = -33.4 \text{ kJ mol}^{-1} = -8.0 \text{ kcal mol}^{-1}$ ). The citric acid cycle itself is also exergonic ( $\Delta G^\circ = -44.3 \text{ kJ mol}^{-1} = -10.6 \text{ kcal mol}^{-1}$ ), and you will be asked in Question 38 to confirm this point. The standard free-energy changes for the individual reactions are listed in Table 19.2. Of the individual reactions of the cycle, only one is strongly endergonic: the oxidation of malate to

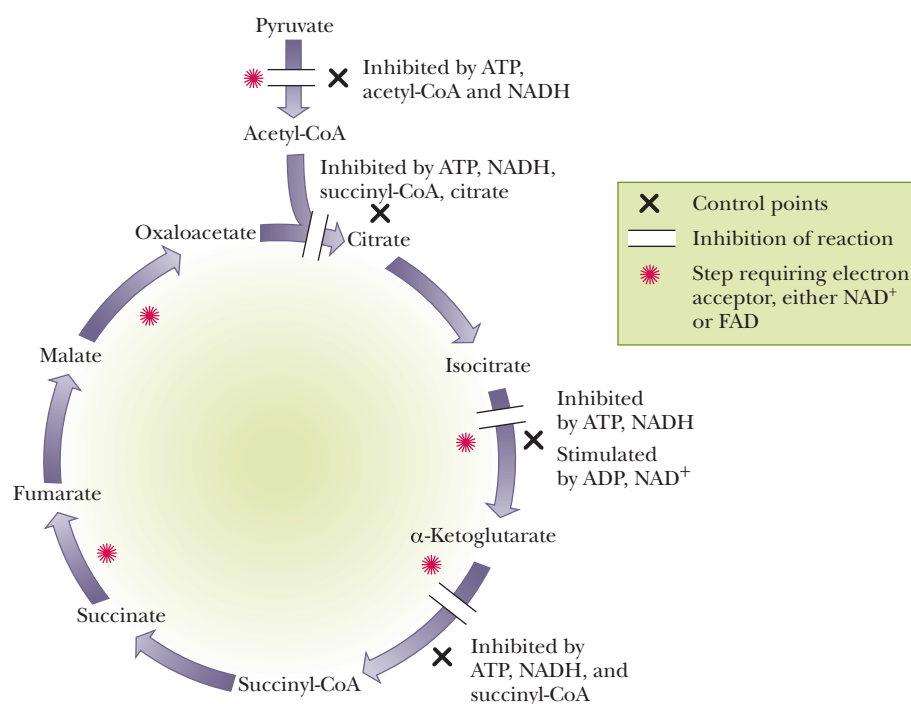
**TABLE 19.2**

The Energetics of Conversion of Pyruvate to CO <sub>2</sub>			
Step	Reaction	$\Delta G^\circ$	
		kJ mol <sup>-1</sup>	kcal mol <sup>-1</sup>
1	Pyruvate + CoA-SH + NAD <sup>+</sup> → Acetyl-CoA + NADH + CO <sub>2</sub>	-33.4	-8.0
	Acetyl-CoA + Oxaloacetate + H <sub>2</sub> O → Citrate + CoA-SH + H <sup>+</sup>	-32.2	-7.7
2	Citrate → Isocitrate	+6.3	+1.5
3	Isocitrate + NAD <sup>+</sup> → α-Ketoglutarate + NADH + CO <sub>2</sub> + H <sup>+</sup>	-7.1	-1.7
4	α-Ketoglutarate + NAD <sup>+</sup> + CoA-SH → Succinyl-CoA + NADH + CO <sub>2</sub> + H	-33.4	-8.0
5	Succinyl-CoA + GDP + P <sub>i</sub> → Succinate + GTP + CoA-SH	-3.3	-0.8
6	Succinate + FAD → Fumarate + FADH <sub>2</sub>	~0	~0
7	Fumarate + H <sub>2</sub> O → L-Malate	-3.8	-0.9
8	L-Malate + NAD <sup>+</sup> → Oxaloacetate + NADH + H <sup>+</sup>	+29.2	+7.0
Overall:	Pyruvate + 4NAD <sup>+</sup> + FAD + GDP + P <sub>i</sub> + 2H <sub>2</sub> O → CO <sub>2</sub> + 4NADH + FADH <sub>2</sub> + GTP + 4H <sup>+</sup>	-77.7	-18.6

oxaloacetate ( $\Delta G^{\circ} = +29.2 \text{ kJ mol}^{-1} = +7.0 \text{ kcal mol}^{-1}$ ). This endergonic reaction is, however, coupled to one of the strongly exergonic reactions of the cycle, the condensation of acetyl-CoA and oxaloacetate to produce citrate and coenzyme A ( $\Delta G^{\circ} = -32.2 \text{ kJ mol}^{-1} = -7.7 \text{ kcal mol}^{-1}$ ). (Recall that these values for the free-energy changes refer to standard conditions. The effect of concentrations of metabolites *in vivo* can change matters drastically.) In addition to the energy released by the oxidation reactions, there is more release of energy to come in the electron transport chain. When the four NADH and single  $\text{FADH}_2$  produced by the pyruvate dehydrogenase complex and citric acid cycle are reoxidized by the electron transport chain, considerable quantities of ATP are produced. *Control of the citric acid cycle is exercised at three points*; that is, three enzymes within the citric acid cycle play a regulatory role (Figure 19.8). There is also control of access to the cycle via pyruvate dehydrogenase.

### How does the pyruvate dehydrogenase reaction control the citric acid cycle?

The overall reaction is part of a pathway that releases energy. It is not surprising that the enzyme that initiates it is inhibited by ATP and NADH because both compounds are abundant when a cell has a good deal of energy readily available. The end products of a series of reactions inhibit the first reaction of the series, and the intermediate reactions do not take place when their products are not needed. Consistent with this picture, the pyruvate dehydrogenase (PDH) complex is activated by ADP, which is abundant when a cell needs energy. In mammals, the actual mechanism by which the inhibition takes place is the phosphorylation of pyruvate dehydrogenase. A phosphate group is covalently bound to the enzyme in a reaction catalyzed by the enzyme *pyruvate dehydrogenase kinase*. When the need arises for pyruvate dehydrogenase to be activated, the hydrolysis of the phosphate ester linkage (dephosphorylation) is catalyzed by another enzyme, *phosphoprotein phosphatase*. This latter enzyme is itself activated by  $\text{Ca}^{2+}$ . Both enzymes are associated with the mammalian pyruvate dehydrogenase complex, permitting effective control of the overall reaction from



**FIGURE 19.8** Control points in the conversion of pyruvate to acetyl-CoA and in the citric acid cycle.

pyruvate to acetyl-CoA. The PDH kinase and PDH phosphatase are found on the same polypeptide chain. High levels of ATP activate the kinase. Pyruvate dehydrogenase is also inhibited by high levels of acetyl-CoA. This makes a great deal of metabolic sense. When fats are plentiful and are being degraded for energy, their product is acetyl-CoA (Chapter 21). Thus, if acetyl-CoA is plentiful, there is no reason to send carbohydrates to the citric acid cycle. Pyruvate dehydrogenase is inhibited, and the acetyl-CoA for the TCA cycle comes from other sources.

### How is control exerted within the citric acid cycle?

Within the citric acid cycle itself, the three control points are the reactions catalyzed by citrate synthase, isocitrate dehydrogenase, and the  $\alpha$ -ketoglutarate dehydrogenase complex. We have already mentioned that the first reaction of the cycle is one in which regulatory control appears, as is to be expected in the first reaction of any pathway. Citrate synthase is an allosteric enzyme inhibited by ATP, NADH, succinyl-CoA, and its own product, citrate.

The second regulatory site is the isocitrate dehydrogenase reaction. In this case, ADP and NAD<sup>+</sup> are allosteric activators of the enzyme. We have called attention to the recurring pattern in which ATP and NADH inhibit enzymes of the pathway, and ADP and NAD<sup>+</sup> activate these enzymes.

The  $\alpha$ -ketoglutarate dehydrogenase complex is the third regulatory site. As before, ATP and NADH are inhibitors. Succinyl-CoA is also an inhibitor of this reaction. This recurring theme in metabolism reflects the way in which a cell can adjust to an active state or to a resting state.

When a cell is metabolically active it uses ATP and NADH at a great rate, producing large amounts of ADP and NAD<sup>+</sup> (Table 19.3). In other words, when the ATP/ADP ratio is low, the cell is using energy and needs to release more energy from stored nutrients. A low NADH/NAD<sup>+</sup> ratio is also characteristic of an active metabolic state. On the other hand, a resting cell has fairly high levels of ATP and NADH. The ATP/ADP ratio and the NADH/NAD<sup>+</sup> ratio are also high in resting cells, which do not need to maintain a high level of oxidation to produce energy.

When cells have low energy requirements (that is, when they have a high “energy charge”) with high ATP/ADP and NADH/NAD<sup>+</sup> ratios, the presence of so much ATP and NADH serves as a signal to “shut down” the enzymes responsible for oxidative reactions. When cells have a low energy charge, characterized by low ATP/ADP and NADH/NAD<sup>+</sup> ratios, the need to release more energy and to generate more ATP serves as a signal to “turn on” the oxidative enzymes. This relationship of energy requirements to enzyme activity is the basis for the overall regulatory mechanism exerted at a few key control points in metabolic pathways.

**TABLE 19.3**

#### Relationship between the Metabolic State of a Cell and the ATP/ADP and NADH/NAD<sup>+</sup> Ratios

##### Cells in a resting metabolic state

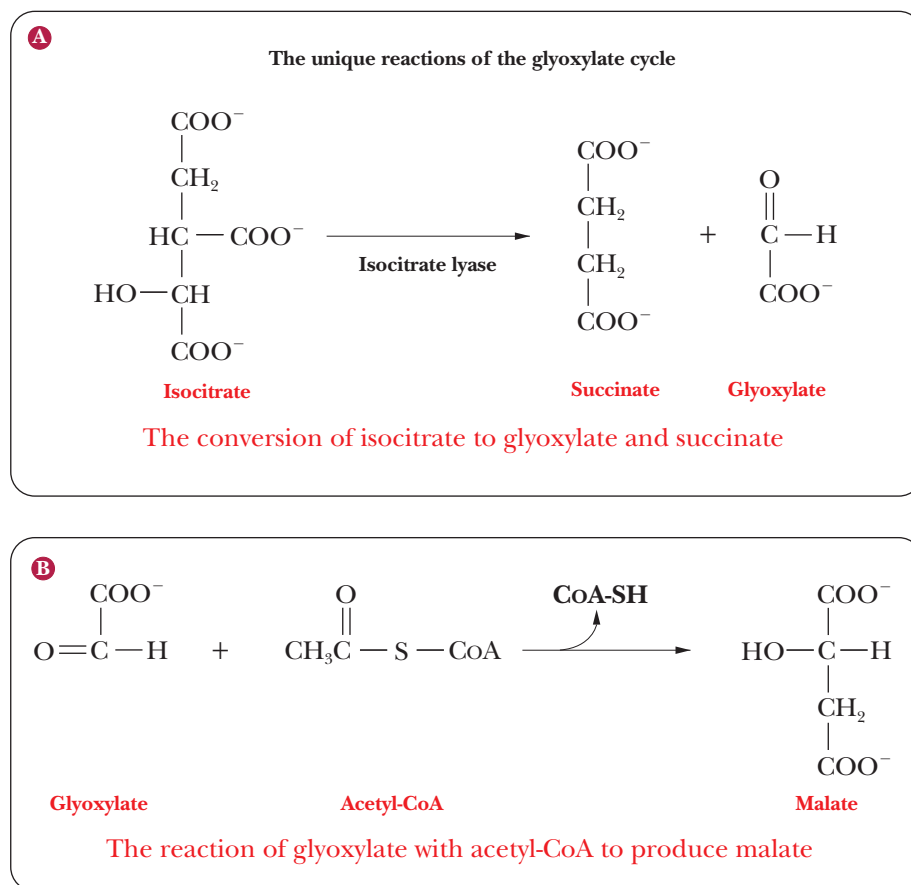
- Need and use comparatively little energy
- High ATP, low ADP levels imply high (ATP/ADP)
- High NADH, low NAD<sup>+</sup> levels imply high (NADH/NAD<sup>+</sup>)

##### Cells in a highly active metabolic state

- Need and use more energy than resting cells
- Low ATP, high ADP levels imply low (ATP/ADP)
- Low NADH, high NAD<sup>+</sup> levels imply low (NADH/NAD<sup>+</sup>)

## 19.6 The Glyoxylate Cycle: A Related Pathway

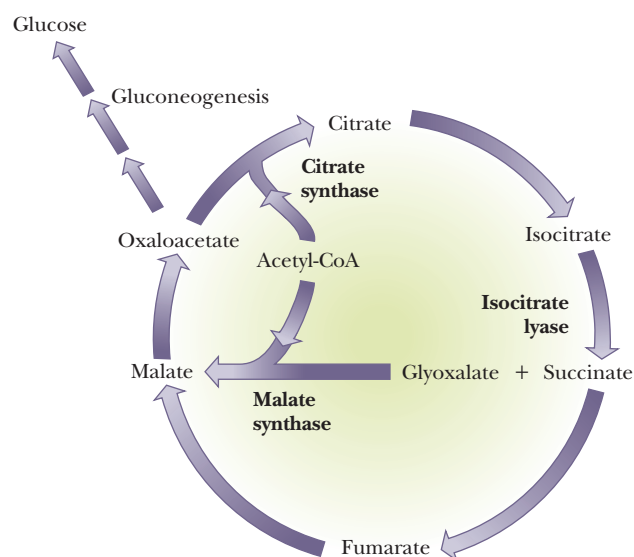
In plants and in some bacteria, but not in animals, acetyl-CoA can serve as the starting material for the biosynthesis of carbohydrates. Animals can convert carbohydrates to fats, but not fats to carbohydrates. (Acetyl-CoA is produced in the catabolism of fatty acids.) Two enzymes are responsible for the ability of plants and bacteria to produce glucose from fatty acids. **Isocitrate lyase** cleaves isocitrate, producing glyoxylate and succinate. **Malate synthase** catalyzes the reaction of glyoxylate with acetyl-CoA to produce malate.



These two reactions in succession bypass the two oxidative decarboxylation steps of the citric acid cycle. The net result is an alternative pathway, the **glyoxylate cycle** (Figure 19.9). Two molecules of acetyl-CoA enter the glyoxylate cycle; they give rise to one molecule of malate and eventually to one molecule of oxaloacetate. Two two-carbon units (the acetyl groups of acetyl-CoA) give rise to a four-carbon unit (malate), which is then converted to oxaloacetate (also a four-carbon compound). Glucose can then be produced from oxaloacetate by gluconeogenesis. This is a subtle, yet very important, distinction between the glyoxylate cycle and the citric acid cycle. The carbon skeletons that enter the citric acid cycle as acetyl-CoA are effectively lost by the decarboxylation steps. This means that if oxaloacetate (OAA) is drawn off to make glucose, there will be no OAA to continue the cycle. For this reason, fats cannot lead to a *net* production of glucose. With the glyoxylate cycle, the bypass reactions go around the decarboxylations, creating an *extra* four-carbon compound that can be drawn off to make glucose without depleting the citric acid cycle of its starting compound.

Specialized organelles in plants, called **glyoxysomes**, are the sites of the glyoxylate cycle. This pathway is particularly important in germinating seeds. The fatty acids stored in the seeds are broken down for energy during germination.





■ **FIGURE 19.9** The glyoxylate cycle. This pathway results in the net conversion of two acetyl-CoA to oxaloacetate. All the reactions are shown in purple. The unique reactions of the glyoxylate cycle are shown with a light green highlight in the center of the circle.

First, the fatty acids give rise to acetyl-CoA, which can enter the citric acid cycle and go on to release energy in the ways we have already seen. The citric acid cycle and the glyoxylate cycle can operate simultaneously. Acetyl-CoA also serves as the starting point for the synthesis of glucose and any other compounds needed by the growing seedling. (Recall that carbohydrates play an important structural, as well as energy-producing, role in plants.)

The glyoxylate cycle also occurs in bacteria. This point is far from surprising because many types of bacteria can live on very limited carbon sources. They have metabolic pathways that can produce all the biomolecules they need from quite simple molecules. The glyoxylate cycle is one example of how bacteria manage this feat.

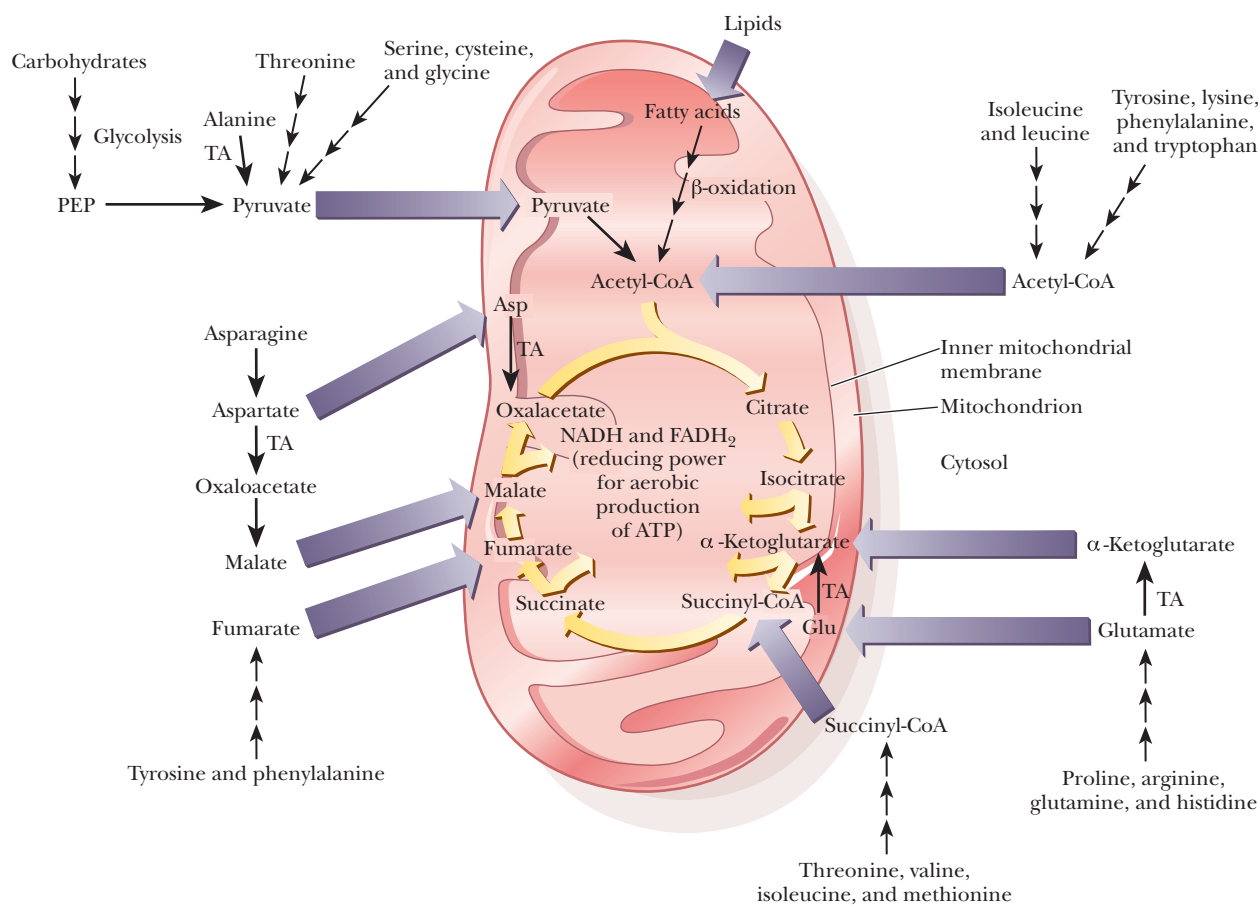
## 19.7 The Citric Acid Cycle in Catabolism

The nutrients taken in by an organism can include large molecules. This observation is especially true in the case of animals, which ingest polysaccharides and proteins, which are polymers, as well as lipids. Nucleic acids constitute a very small percentage of the nutrients present in foodstuffs, and we shall not consider their catabolism.

The first step in the breakdown of nutrients is the degradation of large molecules to smaller ones. Polysaccharides are hydrolyzed by specific enzymes to produce sugar monomers; an example is the breakdown of starch by amylases. Lipases hydrolyze triacylglycerols to give fatty acids and glycerol. Proteins are digested by proteases, with amino acids as the end products. Sugars, fatty acids, and amino acids then enter their specific catabolic pathways.

In Chapter 17, we discussed the glycolytic pathway by which sugars are converted to pyruvate, which then enters the citric acid cycle. In Chapter 21, we will see how fatty acids are converted to acetyl-CoA; we learned about the fate of acetyl-CoA in the citric acid cycle earlier in this chapter. Amino acids enter the cycle by various paths. We will discuss catabolic reactions of amino acids in Chapter 23.

Figure 19.10 shows schematically the various catabolic pathways that feed into the citric acid cycle. The catabolic reactions occur in the cytosol; the citric acid cycle takes place in mitochondria. Many of the end products of catabolism cross the mitochondrial membrane and then participate in the citric acid cycle. This figure also shows the outline of pathways by which amino acids are

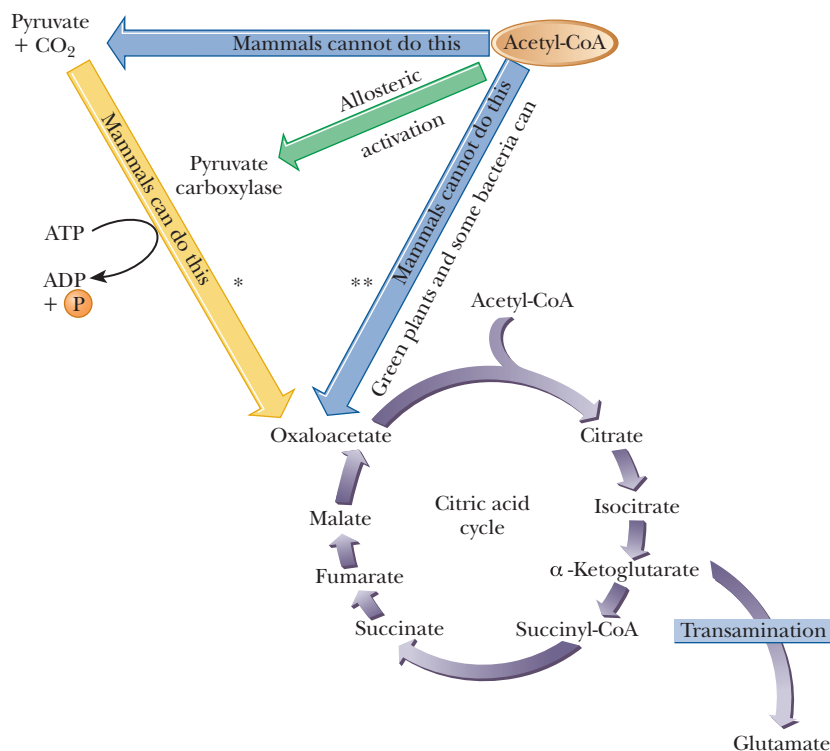


■ **FIGURE 19.10** A summary of catabolism, showing the central role of the citric acid cycle. Note that the end products of the catabolism of carbohydrates, lipids, and amino acids all appear. (PEP is phosphoenolpyruvate;  $\alpha$ -KG is  $\alpha$ -ketoglutarate; TA is transamination;  $\rightarrow\rightarrow\rightarrow$  is a multistep pathway.)

converted to components of the citric acid cycle. Be sure to notice that sugars, fatty acids, and amino acids are all included in this overall catabolic scheme. Just as “all roads lead to Rome,” all pathways lead to the citric acid cycle.

## 19.8 The Citric Acid Cycle in Anabolism

The citric acid cycle is a source of starting materials for the biosynthesis of many important biomolecules, but the supply of the starting materials that are components of the cycle must be replenished if the cycle is to continue operating. See the Biochemical Connections box on page 556. In particular, the oxaloacetate in an organism must be maintained at a level sufficient to allow acetyl-CoA to enter the cycle. A reaction that replenishes a citric acid cycle intermediate is called an **anaplerotic reaction**. In some organisms, acetyl-CoA can be converted to oxaloacetate and other citric acid cycle intermediates by the glyoxylate cycle (Section 19.6), but mammals cannot do this. In mammals, oxaloacetate is produced from pyruvate by the enzyme *pyruvate carboxylase* (Figure 19.11). We already encountered this enzyme and this reaction in the context of gluconeogenesis (see Section 18.2), and here we have another highly important role for this enzyme and the reaction it catalyzes. The supply of oxaloacetate would soon be depleted if there were no means of producing it from a readily available precursor.



**FIGURE 19.11** How mammals keep an adequate supply of metabolic intermediates. An anabolic reaction uses a citric acid cycle intermediate ( $\alpha$ -ketoglutarate is transaminated to glutamate in our example), competing with the rest of the cycle. The concentration of acetyl-CoA rises and signals the allosteric activation of pyruvate carboxylase to produce more oxaloacetate.  
 \*Anaplerotic reaction.  
 \*\*Part of glyoxylate pathway.

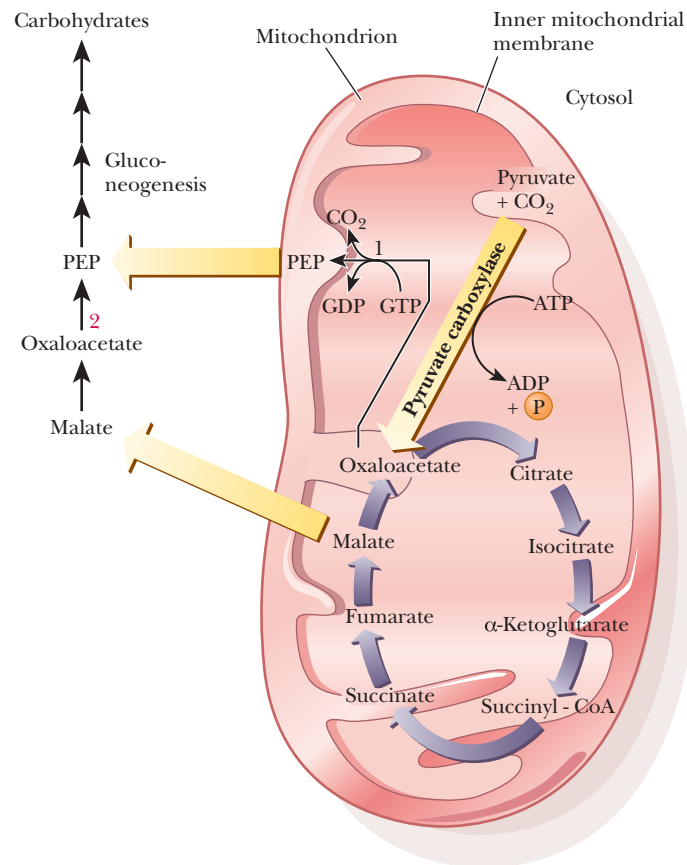
This reaction, which produces oxaloacetate from pyruvate, provides a connection between the amphibolic citric acid cycle and the anabolism of sugars by gluconeogenesis. On this same topic of carbohydrate anabolism, we should note again that pyruvate cannot be produced from acetyl-CoA in mammals. Because acetyl-CoA is the end product of catabolism of fatty acids, we can see that mammals could not exist with fats or acetate as the sole carbon source. The intermediates of carbohydrate metabolism would soon be depleted. Carbohydrates are the principal energy and carbon source in animals (Figure 19.11), and glucose is especially critical in humans because it is the preferred fuel for our brain cells. Plants can carry out the conversion of acetyl-CoA to pyruvate and oxaloacetate, so they can exist without carbohydrates as a carbon source. The conversion of pyruvate to acetyl-CoA does take place in both plants and animals (see Section 19.3).

The anabolic reactions of gluconeogenesis take place in the cytosol. Oxaloacetate is not transported across the mitochondrial membrane. Two mechanisms exist for the transfer of molecules needed for gluconeogenesis from mitochondria to the cytosol. One mechanism takes advantage of the fact that phosphoenolpyruvate can be formed from oxaloacetate in the mitochondrial matrix (this reaction is the next step in gluconeogenesis); phosphoenolpyruvate is then transferred to the cytosol, where the remaining reactions take place (Figure 19.12). The other mechanism relies on the fact that malate, which is another intermediate of the citric acid cycle, can be transferred to the cytosol. There is a *malate dehydrogenase* enzyme in the cytosol as well as in mitochondria, and malate can be converted to oxaloacetate in the cytosol.



Oxaloacetate is then converted to phosphoenolpyruvate, leading to the rest of the steps of gluconeogenesis (Figure 19.12).

Gluconeogenesis has many steps in common with the production of glucose in photosynthesis, but photosynthesis also has many reactions in common with



■ **FIGURE 19.12** Transfer of the starting materials of gluconeogenesis from the mitochondrion to the cytosol. Note that phosphoenolpyruvate (PEP) can be transferred from the mitochondrion to the cytosol, as can malate. Oxaloacetate is not transported across the mitochondrial membrane. (1 is PEP carboxykinase in mitochondria; 2 is PEP carboxykinase in cytosol; other symbols are as in Figure 19.10.)

the pentose phosphate pathway. Thus, nature has evolved common strategies to deal with carbohydrate metabolism in all its aspects.

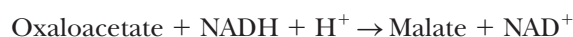
### How is lipid anabolism related to the citric acid cycle?

The starting point of lipid anabolism is acetyl-CoA. The anabolic reactions of lipid metabolism, like those of carbohydrate metabolism, take place in the cytosol; these reactions are catalyzed by soluble enzymes that are not bound to membranes. Acetyl-CoA is mainly produced in mitochondria, whether from pyruvate or from the breakdown of fatty acids. An indirect transfer mechanism exists for transfer of acetyl-CoA in which citrate is transferred to the cytosol (Figure 19.13). Citrate reacts with CoA-SH to produce citryl-CoA, which is then cleaved to yield oxaloacetate and acetyl-CoA. The enzyme that catalyzes this reaction requires ATP and is called ATP-citrate lyase. The overall reaction is

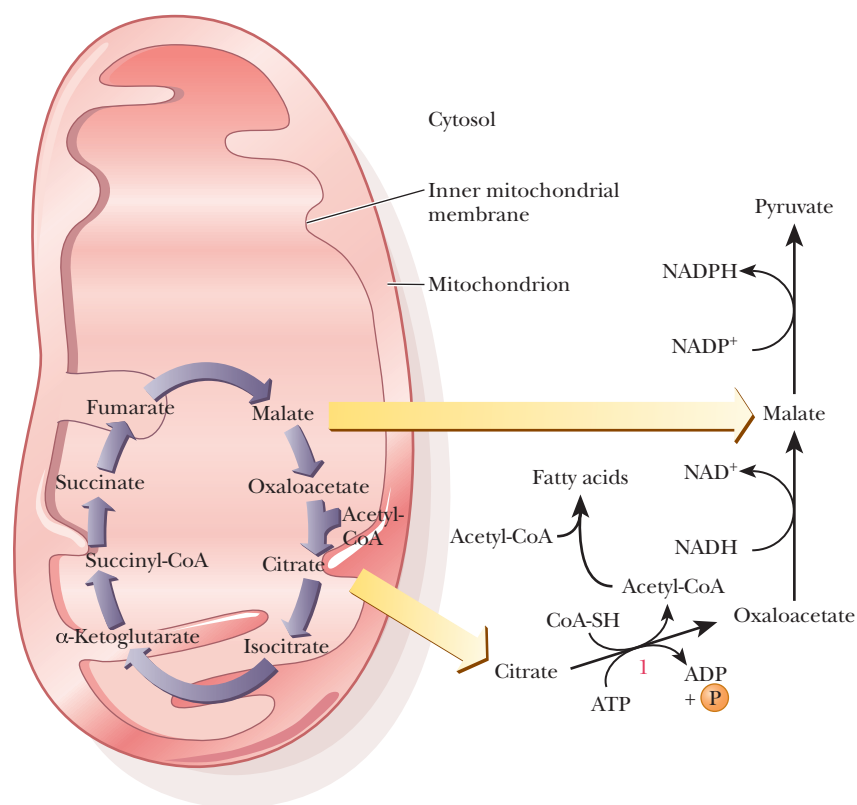


Acetyl-CoA is the starting point for lipid anabolism in both plants and animals. An important source of acetyl-CoA is the catabolism of carbohydrates. We have just seen that animals cannot convert lipids to carbohydrates, but they can convert carbohydrates to lipids. The efficiency of the conversion of carbohydrates to lipids in animals is a source of considerable chagrin to many humans (see the Biochemical Connections box on page 558).

Oxaloacetate can be reduced to malate by the reverse of a reaction we saw in the last section in the context of carbohydrate anabolism.



Malate can move into and out of mitochondria by active transport processes, and the malate produced in this reaction can be used again in the citric acid



■ **FIGURE 19.13** Transfer of the starting materials of lipid anabolism from the mitochondrion to the cytosol. (1 is ATP-citrate lyase; other symbols are as in Figure 19.10.) It is not definitely established whether acetyl-CoA is transported from the mitochondrion to the cytosol.

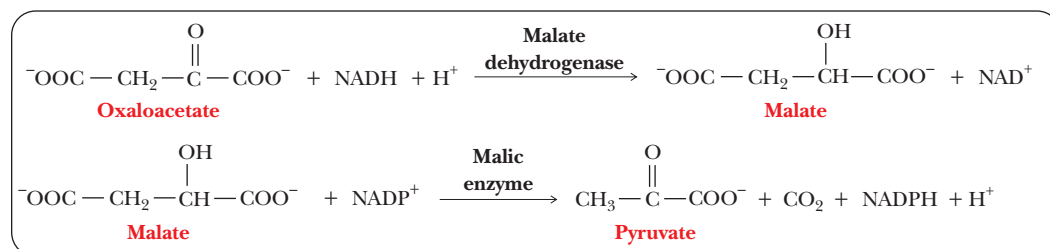
cycle. However, malate need not be transported back into mitochondria but can be oxidatively decarboxylated to pyruvate by *malic enzyme*, which requires  $\text{NADP}^+$ .



These last two reactions are a reduction reaction followed by an oxidation; there is *no net oxidation*. There is, however, a *substitution of NADPH for NADH*.

This last point is important because many of the enzymes of fatty acid synthesis require NADPH. The pentose phosphate pathway (Section 18.4) is the principal source of NADPH in most organisms, but here we have another source (Figure 19.14).

The two ways of producing NADPH clearly indicate that all metabolic pathways are related. The exchange reactions involving malate and citryl-CoA constitute a control mechanism in lipid anabolism, while the pentose phosphate pathway is part of carbohydrate metabolism. Both carbohydrates and lipids are important energy sources in many organisms, particularly animals.



■ **FIGURE 19.14** Reactions involving citric acid cycle intermediates that produce NADPH for fatty acid anabolism. Note that these reactions take place in the cytosol.



### How is amino acid metabolism related to the citric acid cycle?

The anabolic reactions that produce amino acids have, as a starting point, the intermediates of the citric acid cycle that can cross the mitochondrial membrane into the cytosol. We have already seen that malate can cross the mitochondrial membrane and give rise to oxaloacetate in the cytosol. Oxaloacetate can undergo a transamination reaction to produce aspartate, and aspartate, in turn, can undergo further reactions to form not only amino acids but also other nitrogen-containing metabolites, such as pyrimidines. Similarly, isocitrate can cross the mitochondrial membrane and produce  $\alpha$ -ketoglutarate in the cytosol. Glutamate arises from  $\alpha$ -ketoglutarate as a result of another transamination reaction, and glutamate undergoes further reactions to form still more amino acids. Succinyl-CoA gives rise not to amino acids but to the porphyrin ring of the heme group. Another difference is that the first reaction

## Biochemical Connections EVOLUTION

### Why Can't Animals Use All the Same Energy Sources as Plants and Bacteria?

The citric acid cycle is important not only as a source of energy during aerobic metabolism but also as a key pathway in the synthesis of important metabolic intermediates. We shall see in subsequent chapters that it is a source of starting materials for the production of amino acids, carbohydrates, vitamins, nucleotides, and heme. However, if these intermediates are used for the synthesis of other molecules, then they must be replenished to maintain the catalytic nature of this cycle. The term **anaplerotic** means “filling up,” and the reactions that replenish the citric acid cycle are called anaplerotic reactions. One source of needed compounds, available to all organisms, is the group of amino acids that can be converted to citric acid cycle intermediates in a single reaction. A simple reaction available to all organisms is to add carbon dioxide to the pyruvate and phosphoenolpyruvate generated from metabolism of sugars. Another source, important in bacteria and plants, is the glyoxylate cycle discussed in Section 19.6. This source is vital to the ability of plants to fix carbon dioxide to carbohydrates.

Some anaerobic organisms have developed only parts of the citric acid cycle, which they use exclusively to make the important precursors. These simple yet important reactions emphasize the truly connected nature of what we often artificially separate into “pathways.” They also illustrate the convergence of evolution to a few key molecules and metabolic steps.

Which molecule is arguably the most important metabolic intermediate? Acetyl-CoA is perhaps *the* central molecule of metabolism. When one plots a chart of all known metabolic pathways, acetyl-CoA ends up close to the center of that chart.

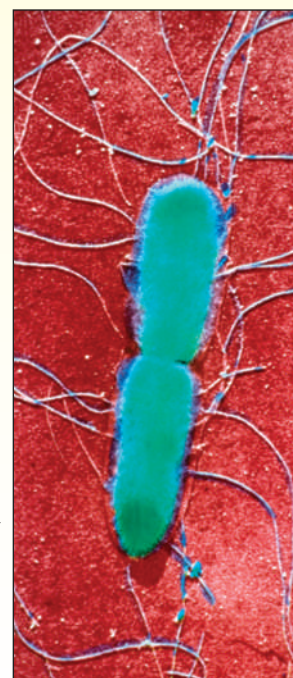
The reasons are quite simple. This important compound really links the metabolism of the three major classes of nutrients to each other. All sugars, all fatty acids, and many amino acids pass through acetyl-CoA on their way to becoming water and carbon dioxide. Equally important is the key use of this intermediate in the synthesis of the major biomolecules. Some, but not all, organisms can carry out all these conversions. Bacteria provide an example of organisms that can do so, whereas humans are an example of ones that cannot. Many bacteria can live off acetic acid as their sole carbon source; however, it is first converted to acetyl-CoA. Acetyl-CoA is converted to fatty acids, terpenes, and steroids. More important is the conversion of two molecules of acetyl-CoA to malate in plants and bacteria via the glyoxylate pathway. This key compound is the starting point for the synthesis of both amino acids and carbohydrates. It is interesting to note, as mentioned in Section 19.6, that this key glyoxylate reaction is missing in mammals.



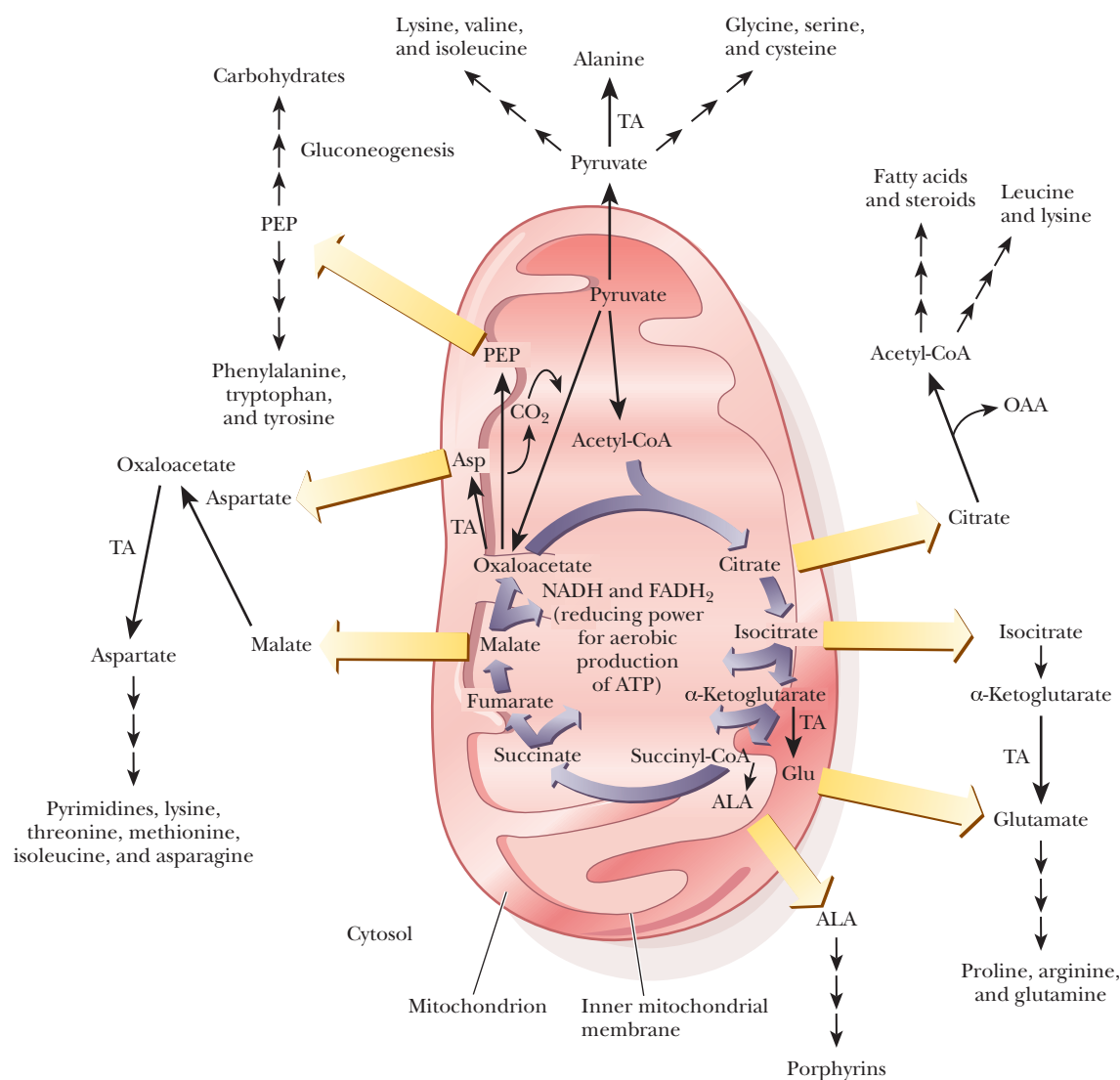
© Nial McDermid/Alamy



© Andrew Catterwall/Alamy



© Phototake/Alamy



■ **FIGURE 19.15** A summary of anabolism, showing the central role of the citric acid cycle. Note that there are pathways for the biosynthesis of carbohydrates, lipids, and amino acids. OAA is oxaloacetate, and ALA is  $\delta$ -aminolevulinic acid. Symbols are as in Figure 19.10.)

of heme biosynthesis, the condensation of succinyl-CoA and glycine to form  $\delta$ -aminolevulinic acid (see supplementary material on nitrogen metabolism on the website), takes place in the mitochondrial matrix, while the remainder of the pathway occurs in the cytosol.

The overall outline of anabolic reactions is shown in Figure 19.15. We used the same type of diagram in Figure 19.10 to show the overall outline of catabolism. The similarity of the two schematic diagrams points out that catabolism and anabolism, while not exactly the same, are closely related. The operation of any metabolic pathway, anabolic or catabolic, can be “speeded up” or “slowed down” in response to the needs of an organism by control mechanisms, such as feedback control. Regulation of metabolism takes place in similar ways in many different pathways.

## 19.9 The Link to Oxygen

The citric acid cycle is considered part of aerobic metabolism, but we have not encountered any reactions in this chapter in which oxygen takes part. The reactions of the citric acid cycle are intimately related to those of electron transport

## Biochemical Connections NUTRITION

### Why Is It So Hard to Lose Weight?

One of the great tragedies about being human is that it is far too easy to gain weight and far too difficult to lose it. If we had to analyze the specific chemical reactions that make this a reality, we would look very carefully at the citric acid cycle, especially the decarboxylation reactions.

As everybody knows, all food in excess can be stored as fat. This is true for carbohydrates, proteins, and, of course, fats. In addition, these molecules can be interconverted, with the exception that fats cannot give a net yield of carbohydrates, as we saw in Section 19.6. Why can fats not yield carbohydrates? The real answer lies in the fact that the only way a fat molecule would have to make glucose would be to enter the citric acid cycle as acetyl-CoA and then to be drawn off as oxaloacetate for gluconeogenesis. Unfortunately, the two carbons that enter are effectively lost by the decarboxylations. (We have already seen that, in one round of the citric acid cycle, it isn't really these same two carbons that are lost; nevertheless, a two-carbon loss is a two-carbon loss, regardless of which two carbons they were.) This leads to an imbalance in the catabolic pathways versus the anabolic pathways.

All roads lead to fat, but fat cannot lead back to carbohydrate. Humans are very sensitive to glucose levels in the blood because so much of our metabolism is geared toward protecting our brain cells, which prefer glucose as a fuel. If we eat more carbohydrates than we need, the excess carbohydrates turn to fat. As we know, it is very easy to put on fat, especially as we age. What about the reverse? Why don't we just stop eating? Won't that reverse the process? The answer is yes and no. When we start eating less, fat stores are mobilized for energy. Fat is an excellent source of energy because it forms acetyl-CoA and gives a steady influx for the citric acid cycle. Thus, we can lose some weight by reducing our caloric intake. Unfortunately, our blood sugar also drops as soon as our glycogen stores run out. We don't have very much stored glycogen that could maintain our blood glucose levels.

After the blood glucose drops, we become depressed, sluggish, and irritable. We start having negative thoughts like, "This dieting thing is really stupid. I should eat a pint of ice cream." If we continue the diet, and given that we cannot turn fats into carbohydrates, where does the blood glucose come from? Only one source remains: proteins. Proteins are degraded to amino acids, and they are eventually converted to pyruvate for gluconeogenesis. Thus, we begin to lose muscle as well as fat.

There is a bright side to all of this, however. Using our knowledge of biochemistry, we can see that there is a better way to lose weight than dieting—exercise! If you exercise correctly, you can train your body to use fats to supply acetyl-CoA for the citric acid cycle. If you maintain a normal diet, you maintain your blood

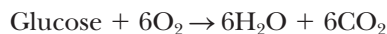
glucose and do not degrade proteins for that purpose; your ingested carbohydrates are sufficient to maintain blood glucose and carbohydrate stores. With the proper balance of exercise to food intake, and the proper balance of the right types of nutrients, we can increase the breakdown of fat without sacrificing our carbohydrate stores or our proteins. In essence, it is easier and healthier to train off the weight than to diet off the weight. This has been known for a long time. Now we are in a position to see why it is biochemically so.



© Blue Moon Stock/Alamy

and oxidative phosphorylation, which do eventually lead to oxygen. The citric acid cycle provides a vital link between the chemical energy of nutrients and the chemical energy of ATP. Many molecules of ATP can be generated as a result of coupling to oxygen, and we shall see that the number depends on the NADH and FADH<sub>2</sub> generated in the citric acid cycle.

Recall the classic equation for the aerobic oxidation of glucose:



We have seen the metabolism of glucose through glycolysis. Now we see where the CO<sub>2</sub> comes from—namely, the three decarboxylation reactions associated with the citric acid cycle. In the next chapter, we will see where the water and oxygen come from.



## SUMMARY

**The central role of the citric acid cycle in metabolism** The citric acid cycle plays a central role in metabolism. It is the first part of aerobic metabolism; it is also amphibolic (both catabolic and anabolic).

**Where does the citric acid cycle take place in the cell?** Unlike glycolysis, which takes place in the cytosol, the citric acid cycle occurs in mitochondria. Most of the enzymes of the citric acid cycle are in the mitochondrial matrix. Succinate dehydrogenase, the sole exception, is localized in the inner mitochondrial membrane.

**What are the key features of the citric acid cycle?** Pyruvate produced by glycolysis is transformed by oxidative decarboxylation into acetyl-CoA in the presence of coenzyme A. Acetyl-CoA then enters the citric acid cycle by reacting with oxaloacetate to produce citrate. The reactions of the citric acid cycle include two other oxidative decarboxylations, which transform the six-carbon compound citrate into the four-carbon compound succinate. The cycle is completed by regeneration of oxaloacetate from succinate in a multi-step process that includes two other oxidation reactions. The overall reaction, starting with pyruvate, is



NAD<sup>+</sup> and FAD are the electron acceptors in the oxidation reactions. The cycle is strongly exergonic.

**How many enzymes are needed to convert pyruvate to acetyl-CoA?** Pyruvate is produced by glycolysis in the cytosol of the cell. The citric acid cycle takes place in the matrix of the mitochondria, so the pyruvate must first pass through a transporter into this organelle. There, pyruvate will find pyruvate dehydrogenase, a large, multisubunit protein made up of three enzymes involved in the production of acetyl-CoA plus two enzyme activities involved in control of the enzymes. The reaction requires several cofactors, including FAD, lipoic acid, and TPP.

**The individual reactions of the citric acid cycle.** Acetyl-CoA condenses with oxaloacetate to give citrate, a six-carbon compound. Citrate isomerizes to isocitrate, which then undergoes an oxidative decarboxylation to  $\alpha$ -ketoglutarate, a five-carbon compound. This then undergoes another oxidative decarboxylation producing succinyl-CoA, a four-carbon compound. The two decarboxylation steps also produce NADH. Succinyl-CoA is converted to succinate with the concomitant production of GTP. Succinate is oxidized to fumarate, and FADH<sub>2</sub> is produced. Fumarate is converted to malate, which is then oxidized to oxaloacetate while another NADH is produced.

The overall pathway has a  $\Delta G^\circ$  of  $-77.7 \text{ kJ mol}^{-1}$ . During the course of the cycle, starting from pyruvate, four NADH molecules and one FADH<sub>2</sub> are produced. Between the GTP formed directly and the reoxidation of the reduced electron

carriers by the electron transport chain, the citric acid cycle produces 25 ATP. Control of the citric acid cycle is exercised at three points.

**How does the pyruvate dehydrogenase reaction control the citric acid cycle?** There is a control point outside the cycle, the reaction in which pyruvate produces acetyl-CoA.

**How is control exerted within the citric acid cycle?** Within the citric acid cycle, the three control points are the reactions catalyzed by citrate synthase, isocitrate dehydrogenase, and the  $\alpha$ -ketoglutarate dehydrogenase complex.

In general, ATP and NADH are inhibitors, and ADP and NAD<sup>+</sup> are activators of the enzymes at the control points.

In plants and bacteria, there is a pathway related to the citric acid cycle: the glyoxylate cycle. The two oxidative decarboxylations of the citric acid cycle are bypassed. This pathway plays a role in the ability of plants to convert acetyl-CoA to carbohydrates, a process that does not occur in animals.

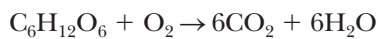
Like a giant traffic circle of life, the citric acid cycle has many routes entering it. Many members of the three basic nutrient types—proteins, fats, and carbohydrates—are metabolized to smaller molecules that can cross the mitochondrial membrane and enter the citric acid cycle as one of the intermediate molecules. In this way, the cycle allows us to get energy from the food we eat. Carbohydrates and many amino acids can enter the cycle either as pyruvate or as acetyl-CoA. Lipids enter as acetyl-CoA. Because of the transamination reaction possible with glutamate and  $\alpha$ -ketoglutarate, almost any amino acid can be transaminated to glutamate, producing  $\alpha$ -ketoglutarate that can enter the cycle. Several other pathways lead to amino acids entering the pathway as succinate, fumarate, or malate.

While the citric acid cycle takes place in mitochondria, many anabolic reactions take place in the cytosol. Oxaloacetate, the starting material for gluconeogenesis, is a component of the citric acid cycle. Malate, but not oxaloacetate, can be transported across the mitochondrial membrane. After malate from mitochondria is carried to the cytosol, it can be converted to oxaloacetate by malate dehydrogenase, an enzyme that requires NAD<sup>+</sup>. Malate, which crosses the mitochondrial membrane, plays a role in lipid anabolism, in a reaction in which malate is oxidatively decarboxylated to pyruvate by an enzyme that requires NADP<sup>+</sup>, producing NADPH.

**How is lipid anabolism related to the citric acid cycle?** The malate reaction is an important source of NADPH for lipid anabolism, with the pentose phosphate pathway the only other source.

**How is amino acid metabolism related to the citric acid cycle?** In addition, most of the intermediates have anabolic pathways leading to amino acids and fatty acids, as well as some that lead to porphyrins or pyrimidines.

Glycolysis and the citric acid cycle account for some of the overall equation for the oxidation of glucose:



The glucose is seen in glycolysis. The decarboxylation steps of the citric acid cycle account for the  $\text{CO}_2$ . However, the oxygen in the equation does not appear until the last

step of the electron transport chain. If insufficient oxygen is available, the electron transport chain will not be able to process the reduced electron carriers from the TCA cycle, and it will have to slow down as well. Continued activity under these circumstances will cause the pyruvate produced by glycolysis to be processed anaerobically to lactate.

## REVIEW EXERCISES

 Interactive versions of these problems are assignable in OWL

### 19.1 The Central Role of the Citric Acid Cycle in Metabolism

- Recall** Which pathways are involved in the anaerobic metabolism of glucose? Which pathways are involved in the aerobic metabolism of glucose?
- Recall** How many ATPs can be produced from one molecule of glucose anaerobically? Aerobically?
- Recall** What are the different names used to describe the pathway discussed in this chapter?
- Recall** What is meant by the statement that a pathway is amphibolic?

### 19.2 The Overall Pathway of the Citric Acid Cycle

- Recall** In what part of the cell does the citric acid cycle take place? Does this differ from the part of the cell where glycolysis occurs?
- Recall** How does pyruvate from glycolysis get to the pyruvate dehydrogenase complex?
- Recall** What electron acceptors play a role in the citric acid cycle?
- Recall** What three molecules produced during the citric acid cycle are an indirect or direct source of high-energy compounds?

### 19.3 How Pyruvate Is Converted to Acetyl-CoA

- Recall** How many enzymes are involved in mammalian pyruvate dehydrogenase? What are their functions?
- Recall** Briefly describe the dual role of lipoic acid in the pyruvate dehydrogenase complex.
- Recall** What is the advantage to the organization of the PDH complex?
- Recall** In the PDH reaction alone, we can see cofactors that come from four different vitamins. What are they?
- Reflect and Apply** Draw the structures of the activated carbon groups bound to thiamine pyrophosphate in three enzymes that contain this coenzyme. *Hint:* Keto–enol tautomerism may enter into the picture.
- Reflect and Apply** Prepare a sketch showing how the individual reactions of the three enzymes of the pyruvate dehydrogenase complex give rise to the overall reaction.

### 19.4 The Individual Reactions of the Citric Acid Cycle

- Recall** Why is the reaction catalyzed by citrate synthase considered a condensation reaction?
- Recall** What does it mean when an enzyme has the name *synthase*?
- Biochemical Connections** What is fluoroacetate? Why is it used?
- Recall** With respect to stereochemistry, what is unique about the reaction catalyzed by aconitase?
- Recall** In which steps of the aerobic processing of pyruvate is  $\text{CO}_2$  produced?
- Recall** In which steps of the aerobic processing of pyruvate are reduced electron carriers produced?
- Recall** What type of reaction is catalyzed by isocitrate dehydrogenase and  $\alpha$ -ketoglutarate dehydrogenase?

- Recall** What are the similarities and differences between the reactions catalyzed by pyruvate dehydrogenase and  $\alpha$ -ketoglutarate dehydrogenase?
- Recall** What does it mean when an enzyme is called a synthetase?
- Recall** Why can we say that production of a GTP is equivalent to production of an ATP?
- Recall** What are the major differences between the oxidations in the citric acid cycle that use  $\text{NAD}^+$  as an electron acceptor and the one that uses FAD?
- Recall** ATP is a competitive inhibitor of NADH binding to malate dehydrogenase, as are ADP and AMP. Suggest a structural basis for this inhibition.
- Recall** Is the conversion of fumarate to malate a redox (electron transfer) reaction? Give the reason for your answer.
- Reflect and Apply** We have seen one of the four possible isomers of isocitrate, the one produced in the aconitase reaction. Draw the configurations of the other three.
- Reflect and Apply** Show, by Lewis electron-dot structures of the appropriate portions of the molecule, where electrons are lost in the following conversions:
  - Pyruvate to acetyl-CoA
  - Isocitrate to  $\alpha$ -ketoglutarate
  - $\alpha$ -Ketoglutarate to succinyl-CoA
  - Succinate to fumarate
  - Malate to oxaloacetate

### 19.5 Energetics and Control of the Citric Acid Cycle

- Recall** Which steps of aerobic metabolism of pyruvate through the citric acid cycle are control points?
- Recall** Describe the multiple ways that PDH is controlled.
- Recall** What are the two most common inhibitors of steps of the citric acid cycle and the reaction catalyzed by pyruvate dehydrogenase?
- Reflect and Apply** How does an increase in the ADP/ATP ratio affect the activity of isocitrate dehydrogenase?
- Reflect and Apply** How does an increase in the  $\text{NADH}/\text{NAD}^+$  ratio affect the activity of pyruvate dehydrogenase?
- Reflect and Apply** Would you expect the citric acid cycle to be more or less active when a cell has a high ATP/ADP ratio and a high  $\text{NADH}/\text{NAD}^+$  ratio? Give the reason for your answer.
- Reflect and Apply** Would you expect  $\Delta G^\circ$  for the hydrolysis of a thioester to be (a) large and negative, (b) large and positive, (c) small and negative, or (d) small and positive? Give the reason for your answer.
- Reflect and Apply** Acetyl-CoA and succinyl-CoA are both high-energy thioesters, but their chemical energy is put to different uses. Elaborate.
- Reflect and Apply** Some reactions of the citric acid cycle are endergonic. Show how the overall cycle is exergonic. (See Table 19.2.)



39. **Reflect and Apply** How could the expression “milking it for all it’s worth” relate to the citric acid cycle?
40. **Reflect and Apply** Using the information in Chapters 17–19, calculate the amount of ATP that can be produced from one molecule of lactose metabolized aerobically through glycolysis and the citric acid cycle.

### 19.6 The Glyoxylate Cycle: A Related Pathway

41. **Recall** Which enzymes of the citric acid cycle are missing from the glyoxylate cycle?
42. **Recall** What are the unique reactions of the glyoxylate cycle?
43. **Biochemical Connections** Why is it possible for bacteria to survive on acetic acid as a sole carbon source, but not human beings?

### 19.7 The Citric Acid Cycle in Catabolism

44. **Recall** Describe the various purposes of the citric acid cycle.
45. **Reflect and Apply** The intermediates of glycolysis are phosphorylated, but those of the citric acid cycle are not. Suggest a reason why.

46. **Reflect and Apply** Discuss oxidative decarboxylation, using a reaction from this chapter to illustrate your points.
47. **Reflect and Apply** Many soft drinks contain citric acid as a significant part of their flavor. Is this a good nutrient?

### 19.8 The Citric Acid Cycle in Anabolism

48. **Recall** NADH is an important coenzyme in catabolic processes, whereas NADPH appears in anabolic processes. Explain how an exchange of the two can be effected.
49. **Biochemical Connections** What are the anaplerotic reactions in mammals?
50. **Reflect and Apply** Why is acetyl-CoA considered the central molecule of metabolism?

### 19.9 The Link to Oxygen

51. **Reflect and Apply** Why is the citric acid cycle considered part of aerobic metabolism, even though molecular oxygen does not appear in any reaction?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# Electron Transport and Oxidative Phosphorylation

# 20

Mitochondria, shown here, are the sites of the citric acid cycle, electron transport, and oxidative phosphorylation.

© Keith R. Porter/Photo Researchers, Inc.

## 20.1 The Role of Electron Transport in Metabolism

Aerobic metabolism is a highly efficient way for an organism to extract energy from nutrients. In eukaryotic cells, the aerobic processes (including conversion of pyruvate to acetyl-CoA, the citric acid cycle, and electron transport) all occur in the mitochondria, while the anaerobic process, glycolysis, takes place outside the mitochondria in the cytosol. We have not yet seen any reactions in which oxygen plays a part, but in this chapter we shall discuss the role of oxygen in metabolism as the final acceptor of electrons in the **electron transport chain**. The reactions of the electron transport chain take place in the inner mitochondrial membrane.

### What is the importance of mitochondrial structure in ATP production?

The energy released by the oxidation of nutrients is used by organisms in the form of the chemical energy of ATP. Production of ATP in the mitochondria is the result of **oxidative phosphorylation**, in which ADP is phosphorylated to give ATP. The production of ATP by oxidative phosphorylation (an endergonic process) is separate from electron transport to oxygen (an exergonic process), but the reactions of the electron transport chain are strongly linked to one another and are tightly coupled to the synthesis of ATP by phosphorylation of ADP. The operation of the electron transport chain leads to pumping of protons (hydrogen ions) across the inner mitochondrial membrane, creating a pH gradient (also called a **proton gradient**). This proton gradient represents stored potential energy and provides the basis of the coupling mechanism (Figure 20.1). *Chemiosmotic coupling* is the name given to this mechanism (Section 20.5). Oxidative phosphorylation gives rise to most of the ATP production associated with the complete oxidation of glucose.

The NADH and FADH<sub>2</sub> molecules generated in glycolysis and the citric acid cycle transfer electrons to oxygen in the series of reactions known collectively as the electron transport chain. The NADH and FADH<sub>2</sub> are oxidized to NAD<sup>+</sup> and FAD and can be used again in various metabolic pathways. Oxygen, the ultimate electron acceptor, is reduced to water; this completes the process by which glucose is completely oxidized to carbon dioxide and water.

We have already seen how carbon dioxide is produced from pyruvate, which in turn is produced from glucose by the pyruvate dehydrogenase complex and the citric acid cycle. In this chapter, we shall see how water is produced.

The complete series of oxidation–reduction reactions of the electron transport chain is presented in schematic form in Figure 20.2. A particularly noteworthy point about electron transport is that, on average, 2.5 moles of ATP are generated for each mole of NADH that enters the electron transport chain, and, on average, 1.5 moles of ATP are produced for each mole of FADH<sub>2</sub>. The general outline of the process is that NADH passes electrons to coenzyme Q, as does

## Chapter Outline

### 20.1 The Role of Electron Transport in Metabolism

- What is the importance of mitochondrial structure in ATP production?

### 20.2 Reduction Potentials in the Electron Transport Chain

- How can reduction potentials be used to predict the direction of electron transport?

### 20.3 Organization of Electron Transport Complexes

- What reactions take place in the respiratory complexes?
- What is the nature of the iron-containing proteins of electron transport?

### 20.4 The Connection between Electron Transport and Phosphorylation

- What is the coupling factor in oxidative phosphorylation?

### 20.5 The Mechanism of Coupling in Oxidative Phosphorylation

- What is chemiosmotic coupling?
- What is conformational coupling?

### 20.6 Respiratory Inhibitors Can Be Used to Study Electron Transport

- Do respiratory inhibitors have a connection with respiratory complexes?

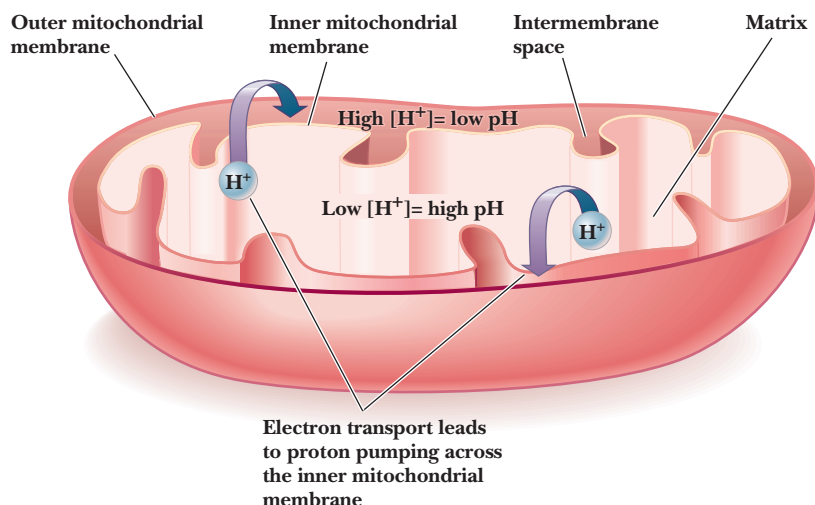
### 20.7 Shuttle Mechanisms

- How do shuttle mechanisms differ from one another?

### 20.8 The ATP Yield from Complete Oxidation of Glucose

**OWL** Online homework for this chapter may be assigned in OWL.

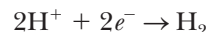
■ **FIGURE 20.1** A proton gradient is established across the inner mitochondrial membrane as a result of electron transport. Transfer of electrons through the electron transport chain leads to the pumping of protons from the matrix to the intermembrane space. The proton gradient (also called the pH gradient), together with the membrane potential (a voltage across the membrane), provides the basis of the coupling mechanism that drives ATP synthesis.



FADH<sub>2</sub>, providing an alternative mode of entry into the electron transport chain. Electrons are then passed from coenzyme Q to a series of proteins called cytochromes (which are designated by lowercase letters) and, eventually, to oxygen.

## 20.2 Reduction Potentials in the Electron Transport Chain

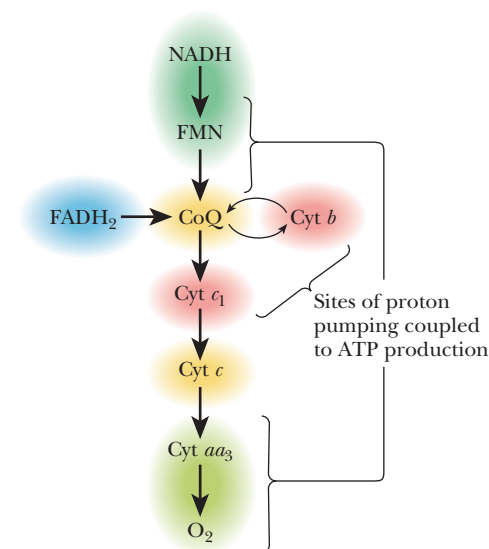
Up until now, most of the energy considerations we have had concerned phosphorylation potentials. In Section 15.6, we saw how the free-energy change associated with hydrolysis of ATP could be used to drive otherwise endergonic reactions. The opposite is also true—when a reaction is highly exergonic, it can drive the formation of ATP. When we look closely at the energy changes in electron transport, a more useful approach is to consider the change in energy associated with the movement of electrons from one carrier to another. Each carrier in the electron transport chain can be isolated and studied, and each can exist in an oxidized or a reduced form (Section 15.5). If we had two potential electron carriers, such as NADH and coenzyme Q (see Section 20.3), for example, how would we know whether electrons would be more likely to be transferred from the NADH to the coenzyme Q or the other way around? This is determined by measuring a **reduction potential** for each of the carriers. A molecule with a high reduction potential tends to be reduced if it is paired with a molecule with a lower reduction potential. This is measured by making a simple battery cell, as shown in Figure 20.3. The reference point is the half-cell on the right where hydrogen ion is in aqueous solution in equilibrium with hydrogen gas. The reduction of hydrogen ion to hydrogen gas



is the control and is considered to have a voltage ( $E$ ) of zero. The sample to be tested is in the other half-cell. The electric circuit is completed by bridge with a salt-containing agar gel.

### How can reduction potentials be used to predict the direction of electron transport?

Figure 20.3a shows what happens if ethanol and acetaldehyde are put into the sample half-cell. Electrons flow away from the sample cell and toward the reference cell. This means that the hydrogen ion is being reduced to hydrogen gas and the ethanol is being oxidized to acetaldehyde. Therefore, the hydrogen/H<sup>+</sup>



■ **FIGURE 20.2** Schematic representation of the electron transport chain, showing sites of proton pumping coupled to oxidative phosphorylation. FMN is the flavin coenzyme *flavin mononucleotide*, which differs from FAD in not having an adenine nucleotide. CoQ is coenzyme Q (see Figure 20.5). Cyt *b*, cyt *c*<sub>1</sub>, cyt *c*, and cyt *aa*<sub>3</sub> are the heme-containing proteins cytochrome *b*, cytochrome *c*<sub>1</sub>, cytochrome *c*, and cytochrome *aa*<sub>3</sub>, respectively.

TABLE 20.1

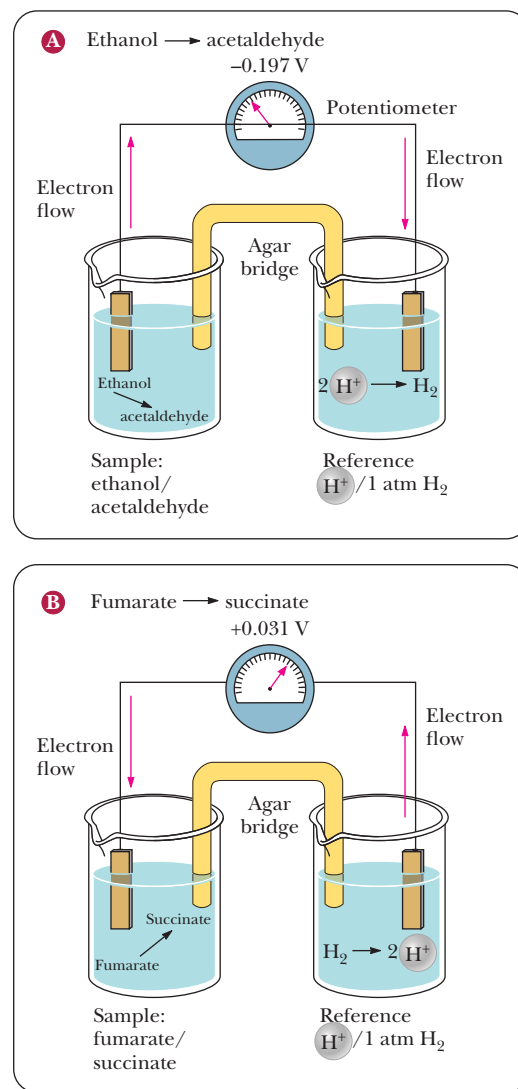
Standard Reduction Potentials for Several Biological Reduction Half Reactions	
Reduction Half Reaction	$E^{\circ}$ (V)
$\frac{1}{2}\text{O}_2 + 2\text{H}^+ + 2e^- \rightarrow \text{H}_2\text{O}$	0.816
$\text{Fe}^{3+} + e^- \rightarrow \text{Fe}^{2+}$	0.771
Cytochrome $a_3(\text{Fe}^{3+}) + e^- \rightarrow$ Cytochrome $a_3(\text{Fe}^{2+})$	0.350
Cytochrome $a(\text{Fe}^{3+}) + e^- \rightarrow$ Cytochrome $a(\text{Fe}^{2+})$	0.290
Cytochrome $c(\text{Fe}^{3+}) + e^- \rightarrow$ Cytochrome $c(\text{Fe}^{2+})$	0.254
Cytochrome $c_1(\text{Fe}^{3+}) + e^- \rightarrow$ Cytochrome $c_1(\text{Fe}^{2+})$	0.220
$\text{CoQH}^{\bullet} + \text{H}^+ + e^- \rightarrow \text{CoQH}_2$ (coenzyme Q)	0.190
$\text{CoQ} + 2\text{H}^+ + 2e^- \rightarrow \text{CoQH}_2$	0.060
Cytochrome $b_{\text{H}}(\text{Fe}^{3+}) + e^- \rightarrow$ Cytochrome $b_{\text{H}}(\text{Fe}^{2+})$	0.050
$\text{Fumarate} + 2\text{H}^+ + 2e^- \rightarrow \text{Succinate}$	0.031
$\text{CoQ} + \text{H}^+ + e^- \rightarrow \text{CoQH}^{\bullet}$	0.030
$[\text{FAD}] + 2\text{H}^+ + 2e^- \rightarrow [\text{FADH}_2]$	0.003–0.091*
Cytochrome $b_{\text{L}}(\text{Fe}^{3+}) + e^- \rightarrow$ Cytochrome $b_{\text{L}}(\text{Fe}^{2+})$	–0.100
$\text{Oxaloacetate} + 2\text{H}^+ + 2e^- \rightarrow \text{Malate}$	–0.166
$\text{Pyruvate} + 2\text{H}^+ + 2e^- \rightarrow \text{Lactate}$	–0.185
$\text{Acetaldehyde} + 2\text{H}^+ + 2e^- \rightarrow \text{Ethanol}$	–0.197
$\text{FMN} + 2\text{H}^+ + 2e^- \rightarrow \text{FMNH}_2$	–0.219
$\text{FAD} + 2\text{H}^+ + 2e^- \rightarrow \text{FADH}_2$	–0.219
$1,3\text{-bisphosphoglycerate} + 2\text{H}^+ + 2e^- \rightarrow$ $\text{Glyceraldehyde-3-phosphate} + \text{P}_i$	–0.290
$\text{NAD}^+ + 2\text{H}^+ + 2e^- \rightarrow \text{NADH} + \text{H}^+$	–0.320
$\text{NADP}^+ + 2\text{H}^+ + 2e^- \rightarrow \text{NADPH} + \text{H}^+$	–0.320
$\alpha\text{-Ketoglutarate} + \text{CO}_2 + 2\text{H}^+ + 2e^- \rightarrow \text{Isocitrate}$	–0.380
$\text{Succinate} + \text{CO}_2 + 2\text{H}^+ + 2e^- \rightarrow \alpha\text{-Ketoglutarate} + \text{H}_2\text{O}$	–0.670

\* Typical values for reduction of bound FAD in flavoproteins such as succinate dehydrogenase.

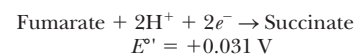
Note that we have shown a number of components of the electron transport chain individually. We are going to see them again as part of complexes. We have also included values for a number of reactions we saw in earlier chapters.

pair has a higher reduction potential than the ethanol/acetaldehyde pair. If we look at Figure 20.3b, we see the opposite. When fumarate and succinate are put into the sample half-cell, the electrons flow in the opposite direction, meaning that fumarate is being reduced to succinate while hydrogen gas is being oxidized to  $\text{H}^+$ . The direction of electron flow and the magnitude of the observed voltage allow us to make a table, as shown in Table 20.1. Because this is a table of standard reduction potentials, all the reactions are shown as reductions. The value being measured is the standard biological voltage of each half reaction  $E^{\circ}$ . This value is calculated based on the compounds in the cells being at 1 M and the pH being 7 at the standard temperature of 25°C.

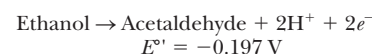
To interpret the data in this table for the purpose of electron transport, we need to look at the reduction potentials of the electron carriers involved. A reaction at the top of the table tends to occur as written if it is paired with a reaction that is lower down on the table. For example, we have already seen that the final step of the electron transport chain is the reduction of oxygen to water. This reaction is at the top of Table 20.1 with a reduction potential of 0.816 V, a very positive number. If this reaction were paired directly with  $\text{NAD}^+/\text{NADH}$ , what would happen? The standard reduction potential for  $\text{NAD}^+$  forming NADH is given near the bottom of the table. Its reduction potential is –0.320 V.



**FIGURE 20.3** Experimental apparatus used to measure the standard reduction potential of the indicated redox couples: (a) the ethanol/acetaldehyde couple, (b) the fumarate/succinate couple. Part (a) shows a sample/reference half-cell pair for measurement of the standard reduction potential of the ethanol/acetaldehyde couple. Because electrons flow toward the reference half-cell and away from the sample half-cell, the standard reduction potential is negative, specifically –0.197 V. In contrast, the fumarate/succinate couple (b) accepts electrons from the reference half-cell; that is, reduction occurs spontaneously in the system, and the reduction potential is thus positive. For each half-cell, a **half-cell reaction** describes the reaction taking place. For the fumarate/succinate half-cell coupled to a  $\text{H}^+/\text{H}_2$  reference half-cell (b), the reaction taking place is indeed the reduction of fumarate.

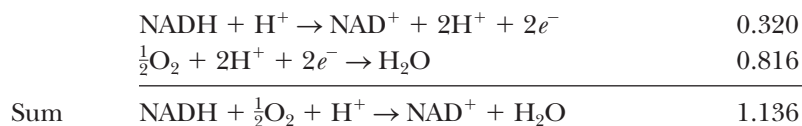


However, the reaction occurring in the ethanol/acetaldehyde half-cell (a) is the oxidation of ethanol, which is the reverse of the reaction listed in Table 20.1.





This means that, if the two half reactions are paired during a redox reaction, the one for the NADH must be reversed. NADH gives up its electrons so that oxygen can be reduced to water:



The overall voltage for this reaction is the sum of the standard reduction potentials—in this case, 0.816 V + 0.320 V, or 1.136 V. Note that we had to change the sign on the standard reduction potential for the NADH because we had to reverse the direction of its reaction.

The  $\Delta G^\circ$  of a redox reaction is calculated using

$$\Delta G^\circ = -nF\Delta E^\circ$$

where  $n$  is the number moles of electrons transferred,  $F$  is Faraday's constant (96.485 kJ V<sup>-1</sup> mol<sup>-1</sup>), and  $\Delta E^\circ$  is the total voltage for the two half reactions. As we can see by this equation,  $\Delta G^\circ$  is negative when  $\Delta E^\circ$  is positive. Therefore, we can always calculate the direction in which a redox reaction will go under standard conditions by combining the two half reactions in the way that gives the largest positive value for  $\Delta E^\circ$ . For this example,  $\Delta G^\circ$  would be calculated as follows:

$$\Delta G^\circ = -(2)(96.485 \text{ kJ V}^{-1} \text{ mol}^{-1})(1.136 \text{ V}) = -219 \text{ kJ mol}^{-1}$$

This would be a very large number if NADH reduced oxygen directly. As we shall see in the next section, NADH passes its electrons along a chain that eventually leads to oxygen, but it does not reduce oxygen directly.

Before moving on, it should be noted that, just as there is a difference between  $\Delta G^\circ$  and  $\Delta G$ , there is a similar difference between  $\Delta E^\circ$  and  $\Delta E$ . Recall from Chapter 15 that we devoted several sections to the question of standard states, including the modified standard state for biochemical reactions. The notations  $\Delta G$  and  $\Delta E$  refer to the free-energy change and the reduction potential under any conditions, respectively. When all components of a reaction are in their standard state (1 atm pressure, 25°C, all solutes at 1 M concentration), we write  $\Delta G^\circ$  and  $\Delta E^\circ$ , respectively, for the standard free-energy change and standard reduction potential. The modified standard state for biochemical reactions takes note of the fact that having all solutes at 1 M concentration includes the hydrogen ion concentration. That implies a pH equal to zero. Consequently, we define a modified standard state for biochemistry that differs from the usual one only in that pH = 7. Under these conditions, we write  $\Delta G^\circ$  and  $\Delta E^\circ$  for the standard free-energy change and the standard reduction potential, respectively. The true direction of electron flow in a redox reaction is also based on the true values of the concentrations for the reactants and products, since the cellular concentrations are never 1 M.

## 20.3 Organization of Electron Transport Complexes

Intact mitochondria isolated from cells can carry out all the reactions of the electron transport chain; the electron transport apparatus can also be resolved into its component parts by a process called fractionation. Four separate *respiratory complexes* can be isolated from the inner mitochondrial membrane. These complexes are multienzyme systems. In the last chapter, we encountered other examples of such multienzyme complexes, such as the pyruvate dehydrogenase complex and the  $\alpha$ -ketoglutarate dehydrogenase complex. Each of the

respiratory complexes can carry out the reactions of a portion of the electron transport chain.

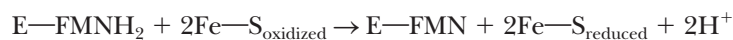
### What reactions take place in the respiratory complexes?

**Complex I** The first complex, **NADH-CoQ oxidoreductase**, catalyzes the first steps of electron transport, namely the transfer of electrons from NADH to **coenzyme Q (CoQ)**. This complex is an integral part of the inner mitochondrial membrane and includes, among other subunits, several proteins that contain an iron–sulfur cluster and the flavoprotein that oxidizes NADH. (The total number of subunits is more than 20. This complex is a subject of active research, which has proven to be a challenging task because of its complexity. It is particularly difficult to generalize about the nature of the iron–sulfur clusters because they vary from species to species.) The flavoprotein has a flavin coenzyme, called flavin mononucleotide, or FMN, which differs from FAD in not having an adenine nucleotide (Figure 20.4).

The reaction occurs in several steps, with successive oxidation and reduction of the flavoprotein and the iron–sulfur moiety. The first step is the transfer of electrons from NADH to the flavin portion of the flavoprotein:



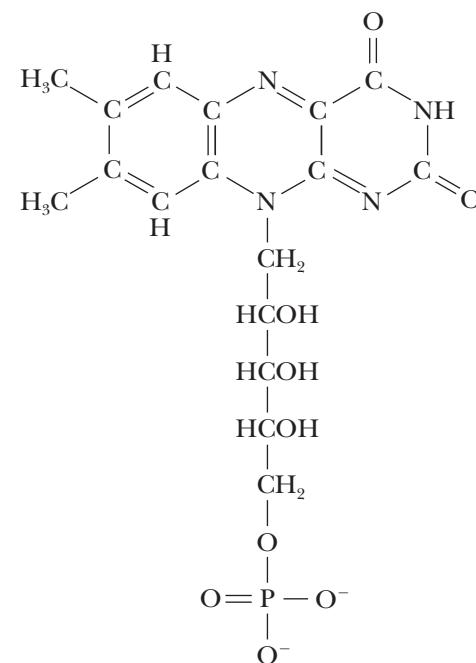
in which the notation E–FMN indicates that the flavin is covalently bonded to the enzyme. In the second step, the reduced flavoprotein is reoxidized, and the oxidized form of the iron–sulfur protein is reduced. The reduced iron–sulfur protein then donates its electrons to coenzyme Q, which becomes reduced to CoQH<sub>2</sub> (Figure 20.5). Coenzyme Q is also called ubiquinone. The equations for the second and third steps are shown here:



The notation Fe–S indicates the iron–sulfur clusters. The overall equation for the reaction is

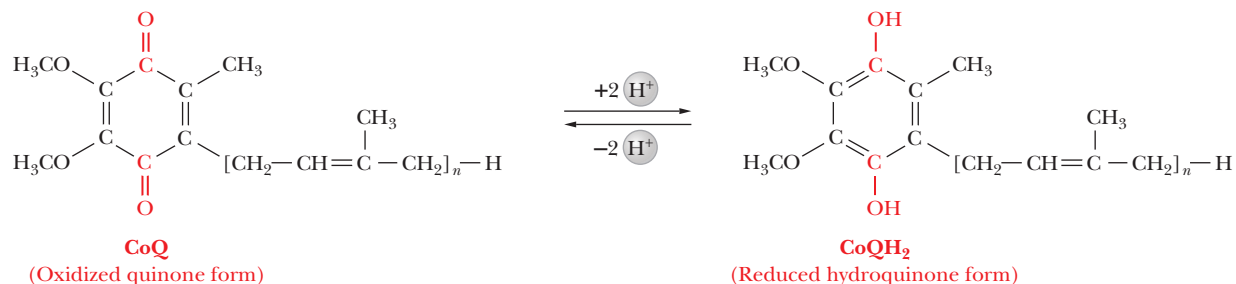


This reaction is one of the three responsible for the proton pumping (Figure 20.6) that creates the pH (proton) gradient. The standard free-energy change ( $\Delta G^\circ = -81 \text{ kJ mol}^{-1} = -19.4 \text{ kcal mol}^{-1}$ ) indicates that the reaction is strongly exergonic, releasing enough energy to drive the phosphorylation of ADP to ATP (Figure 20.7). An important consideration about proton pumping and electron transport is the subtle differences between the electron carriers. Although they can all exist in an oxidized or reduced form, they reduce each

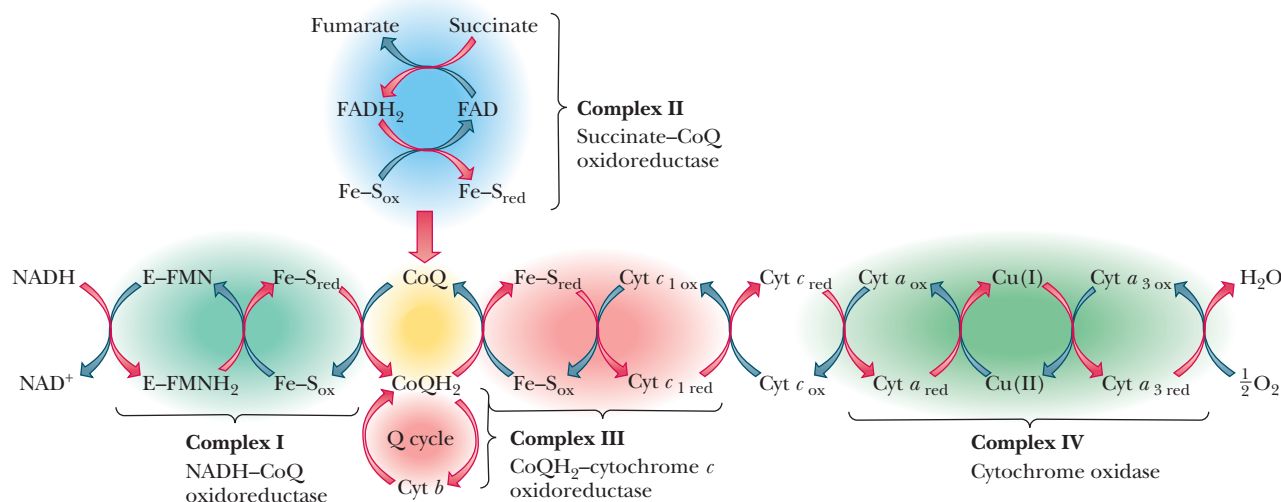


The structure of FMN  
(Flavin mononucleotide)

■ **FIGURE 20.4** The structure of FMN (flavin mononucleotide).



■ **FIGURE 20.5** The oxidized and reduced forms of coenzyme Q. Coenzyme Q is also called ubiquinone.



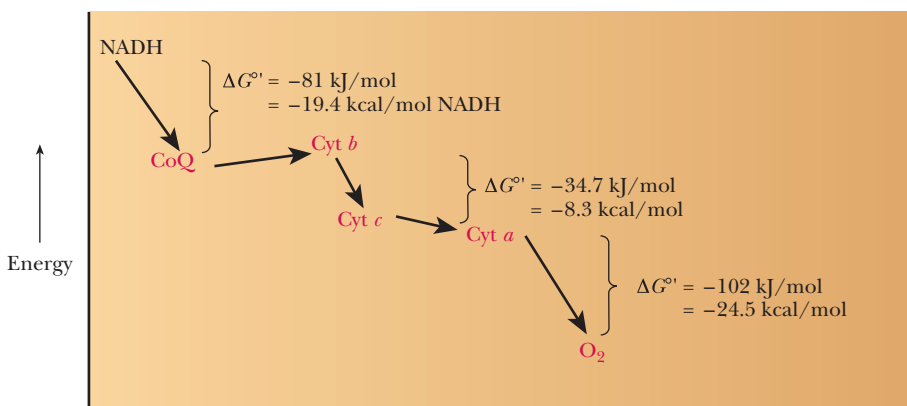
■ **FIGURE 20.6** The electron transport chain, showing the respiratory complexes. In the reduced cytochromes, the iron is in the Fe(II) oxidation state; in the oxidized cytochromes, the oxygen is in the Fe(III) oxidation state.

other in a certain order, as we saw in Section 20.2. In other words, reduced NADH donates its electrons to coenzyme Q, but not the other way around. Thus, there is a direction to the electron flow in the complexes we will study.

The other important subtlety is that some carriers, such as NADH, carry electrons and hydrogens in their reduced forms; others, such as the iron-sulfur protein we just saw, can carry only electrons. This is the basis of the proton pumping that ultimately leads to ATP production. When a carrier such as NADH reduces the iron-sulfur protein, it passes along its electrons, but not its hydrogens. The architecture of the inner mitochondrial membrane and the electron carriers allows the hydrogen ions to pass out on the opposite side of the membrane. We shall look more closely at this in Section 20.5.

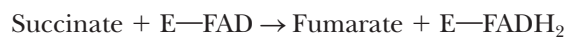
The final electron receptor of complex I, coenzyme Q, is mobile—that is to say, it is free to move in the membrane and to pass the electrons it has gained to the third complex for further transport to oxygen. We shall now see that the second complex also transfers electrons from an oxidizable substrate to coenzyme Q.

**Complex II** The second of the four membrane-bound complexes, **succinate-CoQ oxidoreductase**, also catalyzes the transfer of electrons to coenzyme Q. However, its source of electrons (in other words, the substance being oxidized) differs from the oxidizable substrate (NADH) acted on by NADH-CoQ

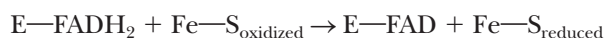


■ **FIGURE 20.7** The energetics of electron transport.

oxidoreductase. In this case the substrate is succinate from the citric acid cycle, which is oxidized to fumarate by a flavin enzyme (see Figure 20.6).



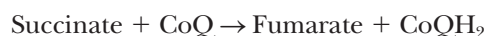
The notation E—FAD indicates that the flavin portion is covalently bonded to the enzyme. The flavin group is reoxidized in the next stage of the reaction as another iron–sulfur protein is reduced:



This reduced iron–sulfur protein then donates its electrons to oxidized coenzyme Q, and coenzyme Q is reduced.



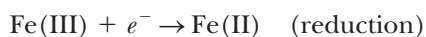
The overall reaction is



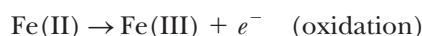
We already saw the first step of this reaction when we discussed the oxidation of succinate to fumarate as part of the citric acid cycle. The enzyme traditionally called succinate dehydrogenase, which catalyzes the oxidation of succinate to fumarate (Section 19.3), has been shown by later work to be a part of this enzyme complex. Recall that the succinate dehydrogenase portion consists of a flavoprotein and an iron–sulfur protein. The other components of Complex II are a *b*-type cytochrome and two iron–sulfur proteins. The whole complex is an integral part of the inner mitochondrial membrane. The standard free-energy change ( $\Delta G^\circ$ ) is  $-13.5 \text{ kJ mol}^{-1} = -3.2 \text{ kcal mol}^{-1}$ . The overall reaction is exergonic, but there is not enough energy from this reaction to drive ATP production, and no hydrogen ions are pumped out of the matrix during this step.

In further steps of the electron transport chain, electrons are passed from coenzyme Q, which is then reoxidized, to the first of a series of very similar proteins called **cytochromes**. Each of these proteins contains a heme group, and in each heme group the iron is successively reduced to Fe(II) and reoxidized to Fe(III). This situation differs from that of the iron in the heme group of hemoglobin, which remains in the reduced form as Fe(II) through the entire process of oxygen transport in the bloodstream. There are also some structural differences between the heme group in hemoglobin and the heme groups in the various types of cytochromes.

The successive oxidation–reduction reactions of the cytochromes

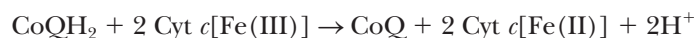


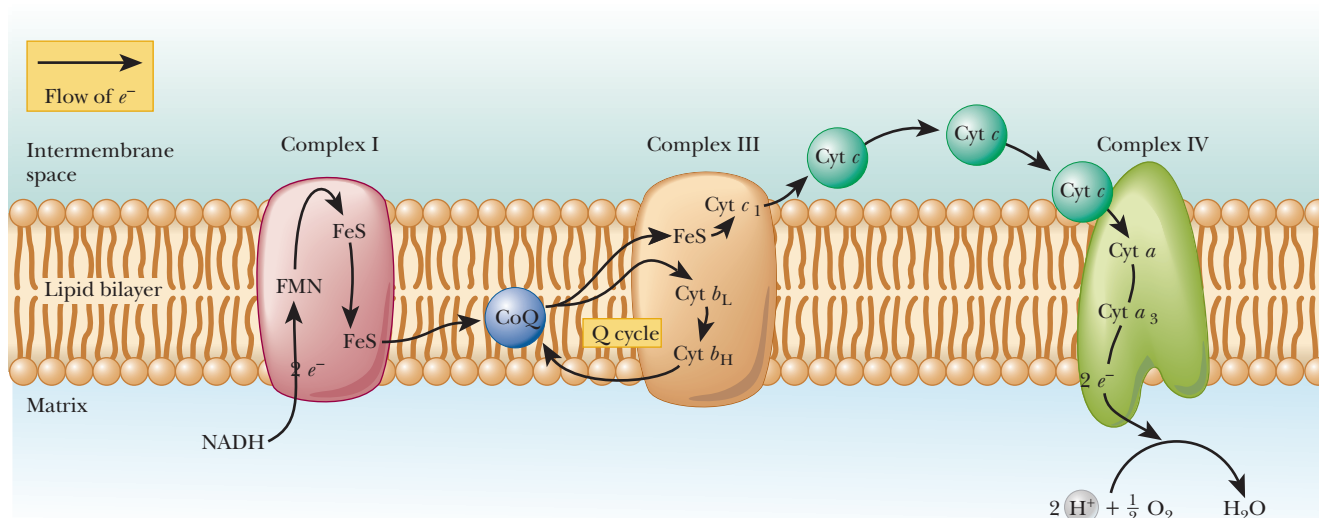
and



differ from one another because the free energy of each reaction,  $\Delta G^\circ$ , differs from the others because of the influences of the various types of hemes and protein structures. Each of the proteins is slightly different in structure, and thus each protein has slightly different properties, including the tendency to participate in oxidation–reduction reactions. The different types of cytochromes are distinguished by lowercase letters (*a*, *b*, *c*); further distinctions are possible with subscripts, as in  $c_1$ .

**Complex III** The third complex, **CoQH<sub>2</sub>-cytochrome *c* oxidoreductase** (also called cytochrome reductase), catalyzes the oxidation of reduced coenzyme Q (CoQH<sub>2</sub>). The electrons produced by this oxidation reaction are passed along to cytochrome *c* in a multistep process. The overall reaction is





■ **FIGURE 20.8** The compositions and locations of respiratory complexes in the inner mitochondrial membrane, showing the flow of electrons from NADH to  $O_2$ . Complex II is not involved and not shown. NADH has accepted electrons from substrates such as pyruvate, isocitrate,  $\alpha$ -ketoglutarate, and malate. Note that the binding site for NADH is on the matrix side of the membrane. Coenzyme Q is soluble in the lipid bilayer. Complex III contains two *b*-type cytochromes, which are involved in the Q cycle. Cytochrome *c* is loosely bound to the membrane, facing the intermembrane space. In Complex IV, the binding site for oxygen lies on the side toward the matrix.

Recall that the oxidation of coenzyme Q involves two electrons, whereas the reduction of Fe(III) to Fe(II) requires only one electron. Therefore, two molecules of cytochrome *c* are required for every molecule of coenzyme Q. The components of this complex include cytochrome *b* (actually two *b*-type cytochromes, cytochrome  $b_H$  and  $b_L$ ), cytochrome  $c_1$ , and several iron–sulfur proteins (Figure 20.6). Cytochromes can carry electrons, but not hydrogens. This is another location where hydrogen ions leave the matrix. When reduced  $CoQH_2$  is oxidized to CoQ, the hydrogen ions pass out on the other side of the membrane.

The third complex is an integral part of the inner mitochondrial membrane. Coenzyme Q is soluble in the lipid component of the mitochondrial membrane. It is separated from the complex in the fractionation process that resolves the electron transport apparatus into its component parts, but the coenzyme is probably close to respiratory complexes in the intact membrane (Figure 20.8). Cytochrome *c* itself is not part of the complex but is loosely bound to the outer surface of the inner mitochondrial membrane, facing the intermembrane space. It is noteworthy that these two important electron carriers, coenzyme Q and cytochrome *c*, are not part of the respiratory complexes but can move freely in the membrane. The respiratory complexes themselves move within the membrane (recall lateral motion within membranes from Section 8.3), and electron transport occurs when one complex encounters the next complex in the respiratory chain as they move.

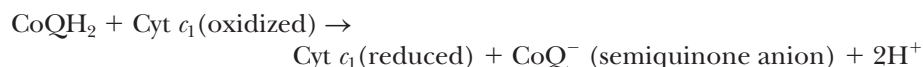
The flow of electrons from reduced coenzyme Q to the other components of the complex does not take a simple, direct path. It is becoming clear that a cyclic flow of electrons involves coenzyme Q twice. This behavior depends on the fact that, as a quinone, coenzyme Q can exist in three forms (Figure 20.9). The semiquinone form, which is intermediate between the oxidized and reduced forms, is of crucial importance here. Because of the crucial involvement of coenzyme Q, this portion of the pathway is called the **Q cycle**.



In part of the Q cycle, *one* electron is passed from reduced coenzyme Q to the iron–sulfur clusters to cytochrome  $c_1$ , leaving coenzyme Q in the semiquinone form.



The notation Fe—S indicates the iron–sulfur clusters. The series of reactions involving coenzyme Q and cytochrome  $c_1$ , but omitting the iron–sulfur proteins, can be written as



The semiquinone, along with the oxidized and reduced forms of coenzyme Q, participates in a cyclic process in which the two  $b$  cytochromes are reduced and oxidized in turn. A second molecule of coenzyme Q is involved, transferring a second electron to cytochrome  $c_1$ , and from there to the mobile carrier cytochrome  $c$ . We are going to omit a number of details of the process in the interest of simplicity. Each of the two molecules of coenzyme Q involved in the Q cycle loses one electron. The net result is the same as if one molecule of CoQ had lost two electrons. It is known that one molecule of CoQH<sub>2</sub> is regenerated, and one is oxidized to CoQ, which is consistent with this picture. Most important, the Q cycle provides a mechanism for electrons to be transferred one at a time from coenzyme Q to cytochrome  $c_1$ .

Proton pumping, to which ATP production is coupled, occurs as a result of the reactions of this complex. The Q cycle is implicated in the process, and the whole topic is under active investigation. The standard free-energy change ( $\Delta G^\circ$ ) is  $-34.2 \text{ kJ} = -8.2 \text{ kcal}$  for each mole of NADH that enters the electron transport chain (see Figure 20.7). The phosphorylation of ADP requires  $30.5 \text{ kJ mol}^{-1} = 7.3 \text{ kcal mol}^{-1}$ , and the reaction catalyzed by the third complex supplies enough energy to drive the production of ATP.

**Complex IV** The fourth complex, **cytochrome  $c$  oxidase**, catalyzes the final steps of electron transport, the transfer of electrons from cytochrome  $c$  to oxygen.

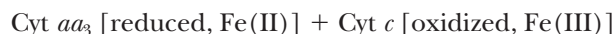
The overall reaction is



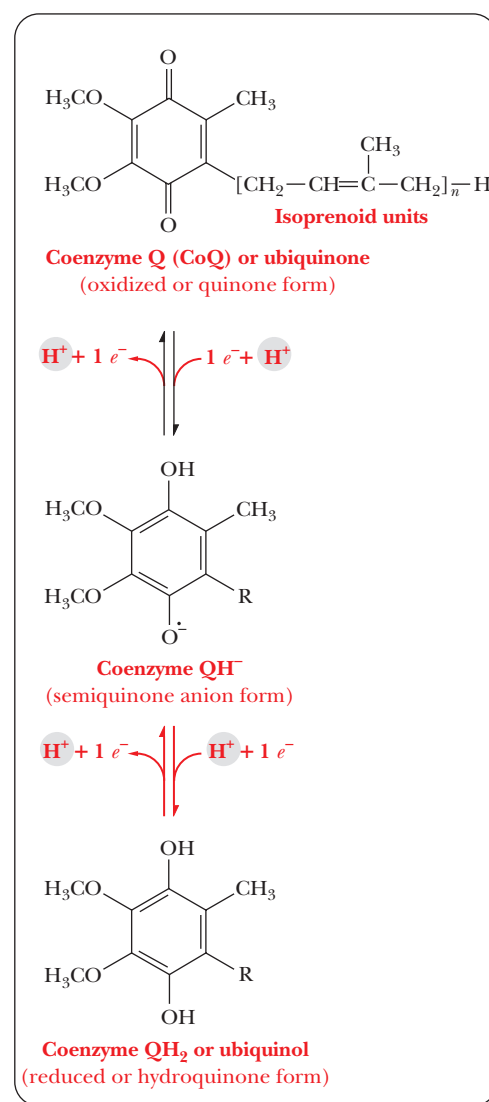
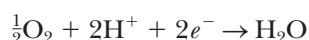
Proton pumping also takes place as a result of this reaction. Like the other respiratory complexes, cytochrome oxidase is an integral part of the inner mitochondrial membrane and contains cytochromes  $a$  and  $a_3$ , as well as two  $\text{Cu}^{2+}$  ions that are involved in the electron transport process. Taken as a whole, this complex contains about 10 subunits. In the flow of electrons, the copper ions are intermediate electron acceptors that lie between the two  $a$ -type cytochromes in the sequence



To show the reactions of the cytochromes more explicitly,



Cytochromes  $a$  and  $a_3$  taken together form the complex known as cytochrome oxidase. The reduced cytochrome oxidase is then oxidized by oxygen, which is itself reduced to water. The half reaction for the reduction of oxygen (oxygen acts as an oxidizing agent) is



■ **FIGURE 20.9** The oxidized and reduced forms of coenzyme Q, showing the intermediate semiquinone anion form involved in the Q cycle.

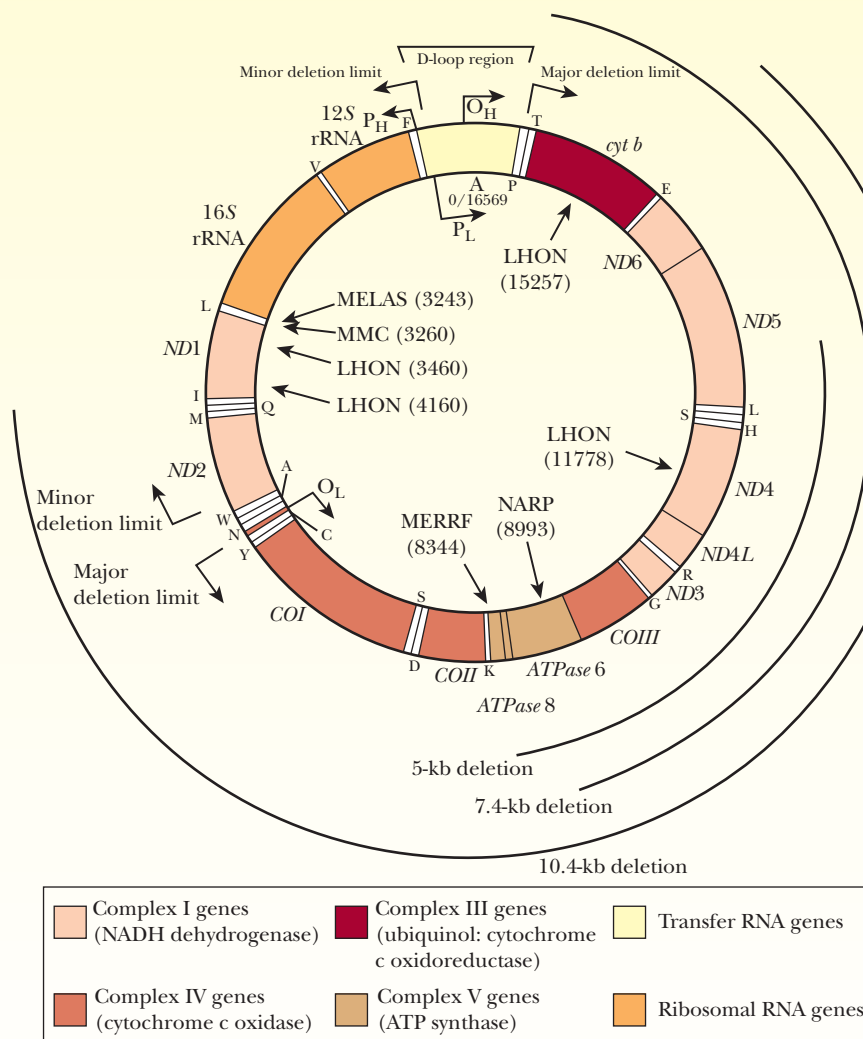
# Biochemical Connections MEDICAL GENETICS

## Mito What . . . ? The Consequences of Mitochondrial Disease

The consequences of mutation in mitochondrial DNA can be wide-ranging because of the central role of mitochondria in metabolism. Some diseases, like mitochondrial myopathy, in which mitochondria fundamentally shut down, are drastic, but fortunately rare. (A fund-raising run for research gave souvenir T-shirts with the slogan “Mito What?” because it was so little known.) Other possible long-term results include heart disease, type 2 diabetes, Parkinson’s disease, Alzheimer’s disease, and many of the effects of aging.

The figure shows the human mitochondrial genome (mt DNA). (Recall that mitochondria not only have their own DNA, but also have ribosomes and carry out protein synthesis.)

Deletion mutations have been correlated with a number of diseases such as hypertrophic cardiomyopathy and ischemic heart disease. Use of animal models has localized the mutations in mt DNA. In July 2008, a study used restriction enzyme methodology to introduce mutations into the cytochrome C oxidase gene in *Drosophila melanogaster*. The mutant flies exhibited a number of defects, including muscular atrophy, growth retardation, neurodegeneration, and shortened life span. This study made it possible to correlate specific changes in DNA with observed disease conditions. The complete proteome of *Drosophila* is known, making this information even more useful as a model for human disease.

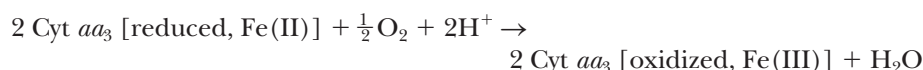


■ **Human mt DNA map, showing the location of major disease mutations.** (From Science, vol. 256, p. 629. Copyright © 1992 by AAAS. Reprinted by permission of the AAAS.)

TABLE 20.2

The Energetics of Electron Transport Reactions		
Reaction	$\Delta G^\circ$	
	$\text{kJ (mol NADH)}^{-1}$	$\text{kcal (mol NADH)}^{-1}$
$\text{NADH} + \text{H}^+ + \text{E-FMN} \rightarrow \text{NAD}^+ + \text{E-FMNH}_2$	-38.6	-9.2
$\text{E-FMNH}_2 + \text{CoQ} \rightarrow \text{E-FMN} + \text{CoQH}_2$	-42.5	-10.2
$\text{CoQH}_2 + 2 \text{ Cyt } b[\text{Fe(III)}] \rightarrow \text{CoQ} + 2\text{H}^+ + 2 \text{ Cyt } b[\text{Fe(II)}]$	+11.6	+2.8
$2 \text{ Cyt } b[\text{Fe(II)}] + 2 \text{ Cyt } c_1[\text{Fe(III)}] \rightarrow 2 \text{ Cyt } c_1[\text{Fe(II)}] + 2 \text{ Cyt } b[\text{Fe(III)}]$	-34.7	-8.3
$2 \text{ Cyt } c_1[\text{Fe(II)}] + 2 \text{ Cyt } c[\text{Fe(III)}] \rightarrow 2 \text{ Cyt } c[\text{Fe(II)}] + 2 \text{ Cyt } c_1[\text{Fe(III)}]$	-5.8	-1.4
$2 \text{ Cyt } c[\text{Fe(II)}] + 2 \text{ Cyt } (aa_3)[\text{Fe(III)}] \rightarrow 2 \text{ Cyt } (aa_3)[\text{Fe(II)}] + 2 \text{ Cyt } c[\text{Fe(III)}]$	-7.7	-1.8
$2 \text{ Cyt } (aa_3)[\text{Fe(II)}] + \frac{1}{2}\text{O}_2 + 2\text{H}^+ \rightarrow 2 \text{ Cyt } (aa_3)[\text{Fe(III)}] + \text{H}_2\text{O}$	-102.3	-24.5
Overall reaction: $\text{NADH} + \text{H}^+ + \frac{1}{2}\text{O}_2 \rightarrow \text{NAD}^+ + \text{H}_2\text{O}$	-220	-52.6

The overall reaction is



*Note that in this final reaction we have finally seen the link to molecular oxygen in aerobic metabolism.*

The standard free-energy change ( $\Delta G^\circ$ ) is  $-110 \text{ kJ} = -26.3 \text{ kcal}$  for each mole of NADH that enters the electron transport chain (see Figure 20.7). We have now seen the three places in the respiratory chain where electron transport is coupled to ATP production by proton pumping. These three places are the NADH dehydrogenase reaction, the oxidation of cytochrome *b*, and the reaction of cytochrome oxidase with oxygen, although the mechanism for proton transfer in cytochrome oxidase remains a mystery. Table 20.2 summarizes the energetics of electron transport reactions.

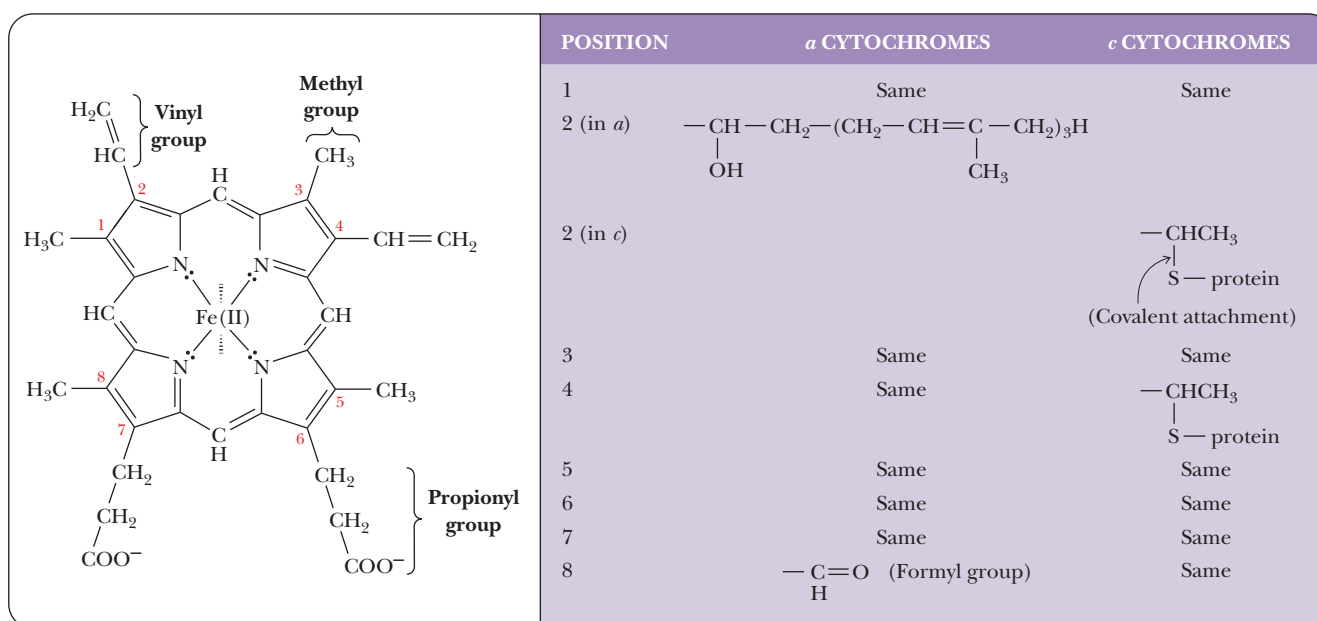
The role of mitochondria in key processes of life makes it imperative that they are intact and functioning properly. The Biochemical Connections box on the preceding page describes some of the consequences of malfunction.

### What is the nature of the iron-containing proteins of electron transport?

In contrast to the electron carriers in the early stages of electron transport, such as NADH, FMN, and CoQ, the cytochromes are macromolecules. These proteins are found in all types of organisms and are typically located in membranes. In eukaryotes, the usual site is the inner mitochondrial membrane, but cytochromes can also occur in the endoplasmic reticulum.

All cytochromes contain the heme group, which is also a part of the structure of hemoglobin and myoglobin (Section 4.5). In the cytochromes, the iron of the heme group does not bind to oxygen; instead, the iron is involved in the series of redox reactions, which we have already seen. There are differences in the side chains of the heme group of the cytochromes involved in the various stages of electron transport (Figure 20.10). These structural differences, combined with the variations in the polypeptide chain and in the way the polypeptide chain is attached to the heme, account for the differences in properties among the cytochromes in the electron transport chain.

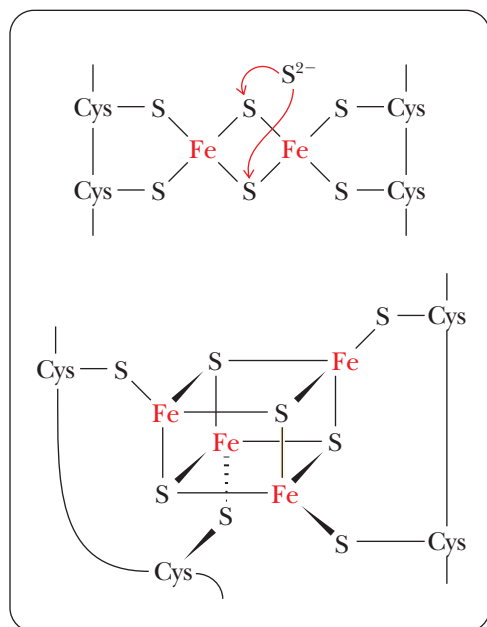
**Nonheme iron proteins** do not contain a heme group, as their name indicates. Many of the most important proteins in this category contain sulfur, as is the case with the iron-sulfur proteins that are components of the respiratory complexes. The iron is usually bound to cysteine or to  $\text{S}^{2-}$  (Figure 20.11). There are still many questions about the location and mode of action of iron-sulfur proteins in mitochondria.



**A** Structures of the heme of all *b* cytochromes and of hemoglobin and myoglobin. The wedge bonds shows the fifth and sixth coordination sites of the iron atom.

**B** A comparison of the side chains of *a* and *c* cytochromes to those of *b* cytochromes.

■ **FIGURE 20.10** The heme group of cytochromes.



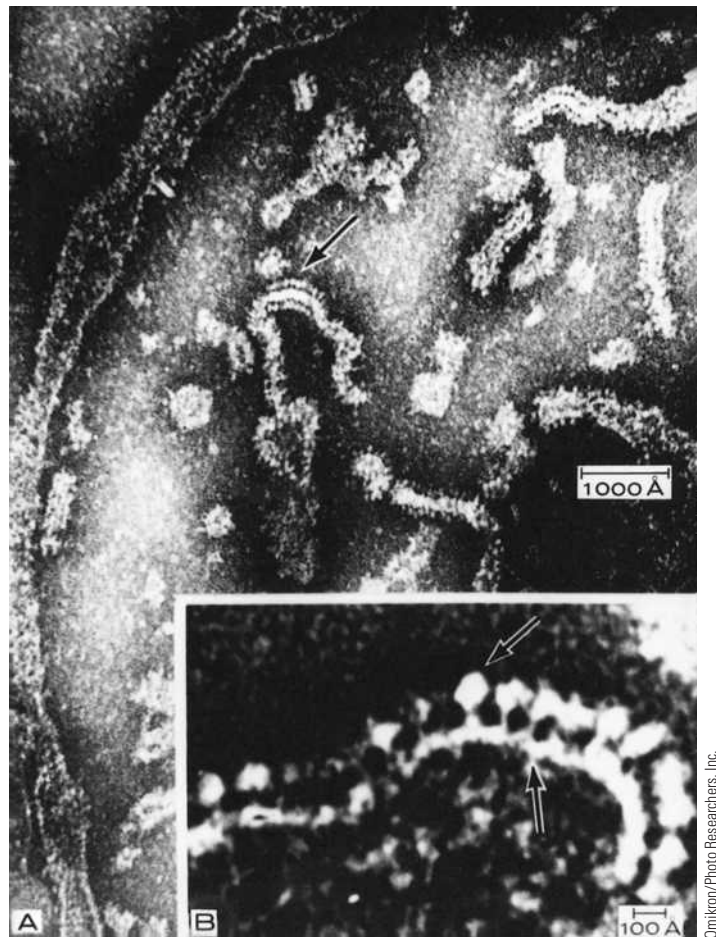
■ **FIGURE 20.11** Iron-sulfur bonding in nonheme iron proteins.

## 20.4 The Connection between Electron Transport and Phosphorylation

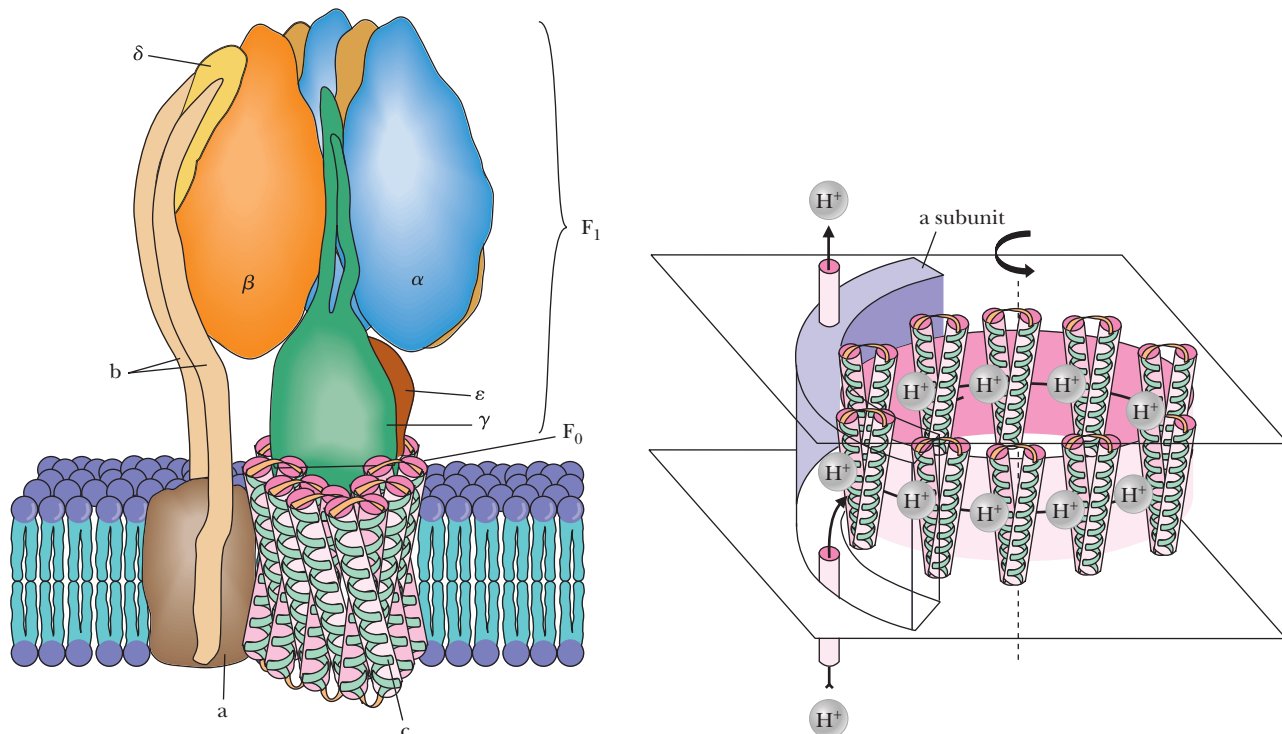
Some of the energy released by the oxidation reactions in the electron transport chain is used to drive the phosphorylation of ADP. The phosphorylation of each mole of ADP requires  $30.5 \text{ kJ} = 7.3 \text{ kcal}$ , and we have seen how each reaction catalyzed by three of the four respiratory complexes provides more than enough energy to drive this reaction, although it is by no means a direct usage of this energy. A common theme in metabolism is that energy to be used by cells is converted to the chemical energy of ATP as needed. The energy-releasing oxidation reactions give rise to proton pumping and thus to the pH gradient across the inner mitochondrial membrane. In addition to the pH gradient, a voltage difference across the membrane is generated by the concentration differences of ions inside and out. The energy of the electrochemical potential (voltage drop) across the membrane is converted to the chemical energy of ATP by the coupling process.

### What is the coupling factor in oxidative phosphorylation?

A coupling factor is needed to link oxidation and phosphorylation. A complex protein oligomer, separate from the electron transport complexes, serves this function; the complete protein spans the inner mitochondrial membrane and projects into the matrix as well. The portion of the protein that spans the membrane is called  $F_0$ . It consists of three different kinds of polypeptide chains (*a*, *b*, and *c*), and research is in progress to characterize it further. The portion that projects into the matrix is called  $F_1$ ; it consists of five different kinds of polypeptide chains in the ratio  $\alpha_3\beta_3\gamma\delta\epsilon$ . Electron micrographs of mitochondria show the projections into the matrix from the inner mitochondrial membrane (Figure 20.12). The schematic organization of the protein can be seen in Figure 20.13. The  $F_1$  sphere is the site of ATP synthesis. The whole protein complex is called **ATP synthase**. It is also known as mitochondrial ATPase because the reverse reaction of ATP hydrolysis, as well as phosphorylation, can be catalyzed

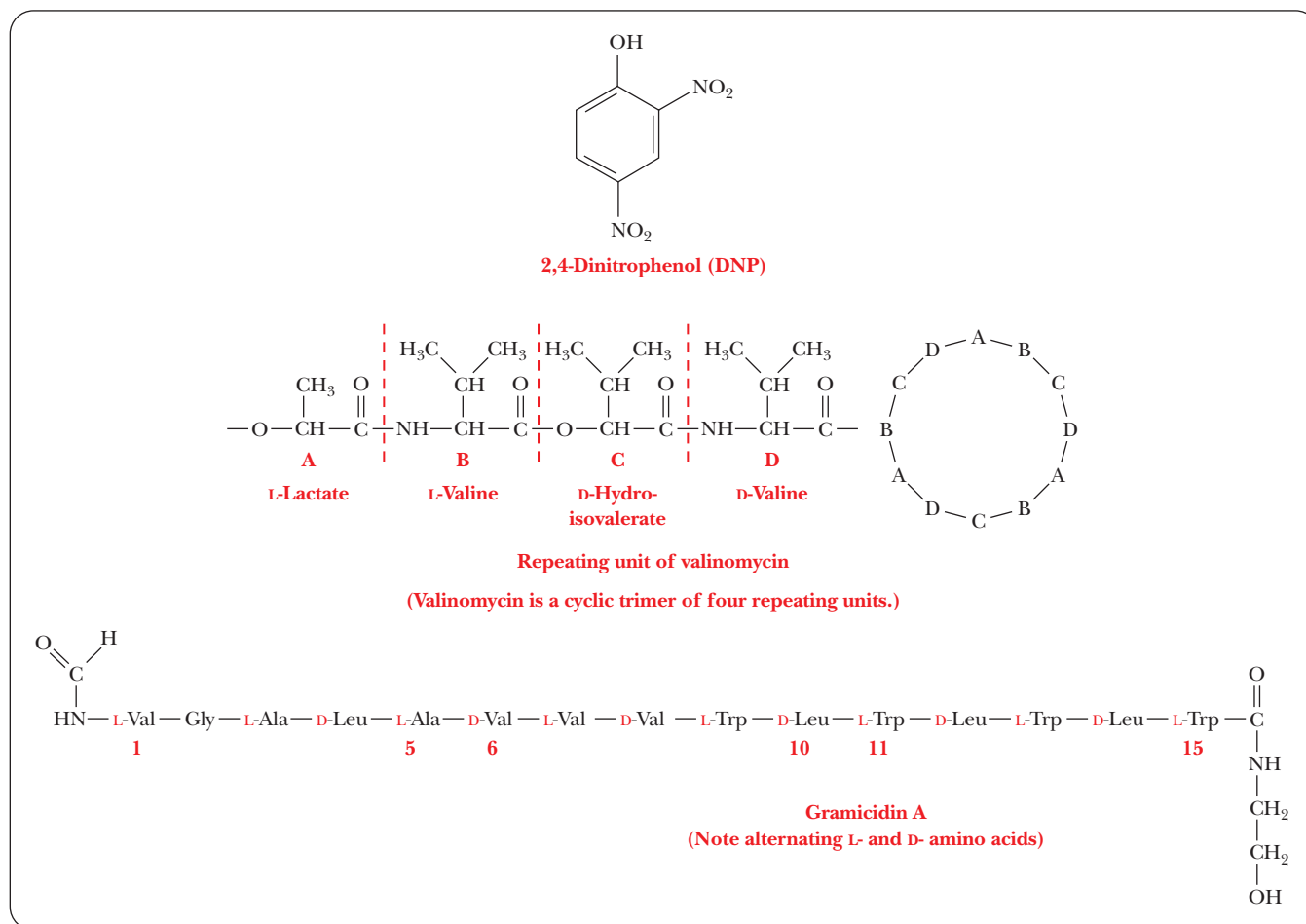


■ **FIGURE 20.12** Electron micrograph of projections into the matrix space of a mitochondrion. Note the difference in scale between part A and part B. The top arrows indicate the matrix side and  $F_1$  subunit. The bottom arrow in part B indicates the intermembrane space.



■ **FIGURE 20.13** A model of the  $F_1$  and  $F_0$  components of the ATP synthase, a rotating molecular motor. The  $a$ ,  $b$ ,  $\alpha$ ,  $\beta$ , and  $\delta$  subunits constitute the stator of the motor, and the  $c$ ,  $\gamma$ , and  $\epsilon$  subunits form the rotor. Flow of protons through the structure turns the rotor and drives the cycle of conformational changes in  $\alpha$  and  $\beta$  that synthesize ATP.





■ **FIGURE 20.14** Some uncouplers of oxidative phosphorylation: 2,4-dinitrophenol, valinomycin, and gramicidin A.

by the enzyme. The hydrolytic reaction was discovered before the reaction of the synthesis of ATP, hence the name. The 1997 Nobel Prize in chemistry was shared by an American scientist, Paul Boyer of UCLA, and a British scientist, John Walker of the Medical Research Council in Cambridge, England, for their elucidation of the structure and mechanism of this enzyme. (The other half of this prize went to a Danish scientist, Jens Skou, for his work on the sodium-potassium pump [Section 8.6], which also functions as an ATPase.)

Compounds known as **uncouplers** inhibit the phosphorylation of ADP without affecting electron transport. A well-known example of an uncoupler is *2,4-dinitrophenol*. Various antibiotics, such as *valinomycin* and *gramicidin A*, are also uncouplers (Figure 20.14). When mitochondrial oxidation processes are operating normally, electron transport from NADH or FADH<sub>2</sub> to oxygen results in the production of ATP. When an uncoupler is present, oxygen is still reduced to H<sub>2</sub>O, but ATP is not produced. If the uncoupler is removed, ATP synthesis linked to electron transport resumes.

A term called the **P/O ratio** is used to indicate the coupling of ATP production to electron transport. The P/O ratio gives the number of moles of P<sub>i</sub> consumed in the reaction  $\text{ADP} + \text{P}_i \rightarrow \text{ATP}$  for each mole of oxygen atoms consumed in the reaction  $\frac{1}{2}\text{O}_2 + 2\text{H}^+ + 2e^- \rightarrow \text{H}_2\text{O}$ . As we have already mentioned, 2.5 moles of ATP are produced when 1 mole of NADH is oxidized to NAD<sup>+</sup>. Recall that oxygen is the ultimate acceptor of the electrons from NADH and that  $\frac{1}{2}$  mole of O<sub>2</sub> molecules (one mole of oxygen atoms) is reduced for each mole of NADH oxidized. The experimentally determined P/O ratio is 2.5

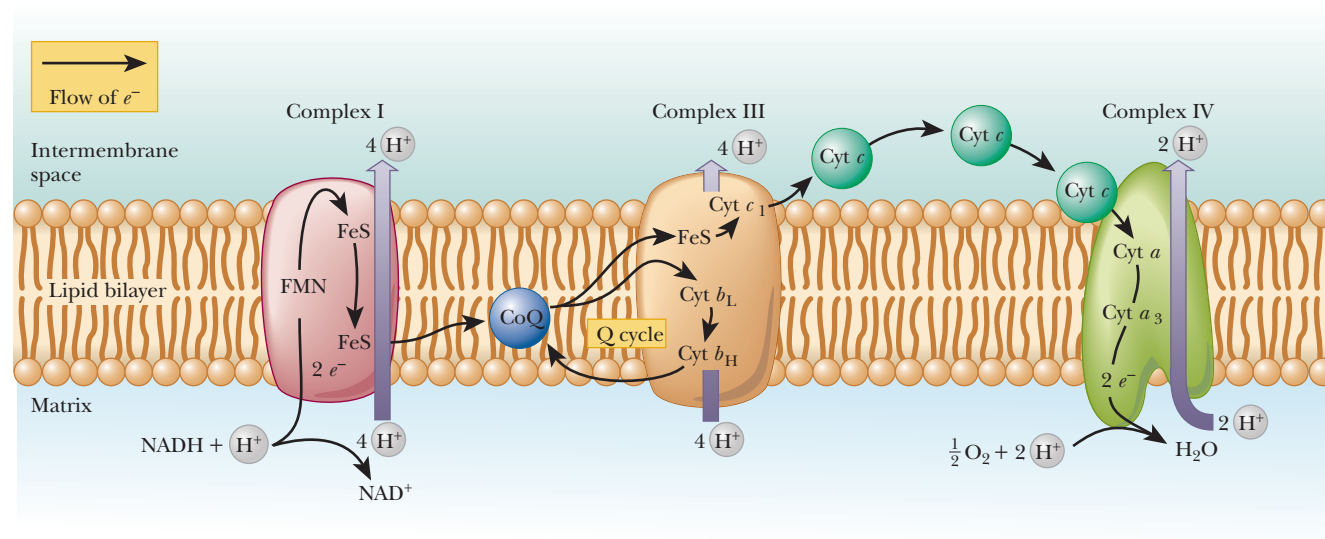
when NADH is the substrate oxidized. The P/O ratio is 1.5 when  $\text{FADH}_2$  is the substrate oxidized (also an experimentally determined value). Until recently, biochemists had used integral values of 3 and 2 for the P/O ratios for reoxidation of NADH and  $\text{FADH}_2$ , respectively. The nonintegral consensus values given here clearly underscore the complexity of electron transport, oxidative phosphorylation, and the manner in which they are coupled.

## 20.5 The Mechanism of Coupling in Oxidative Phosphorylation

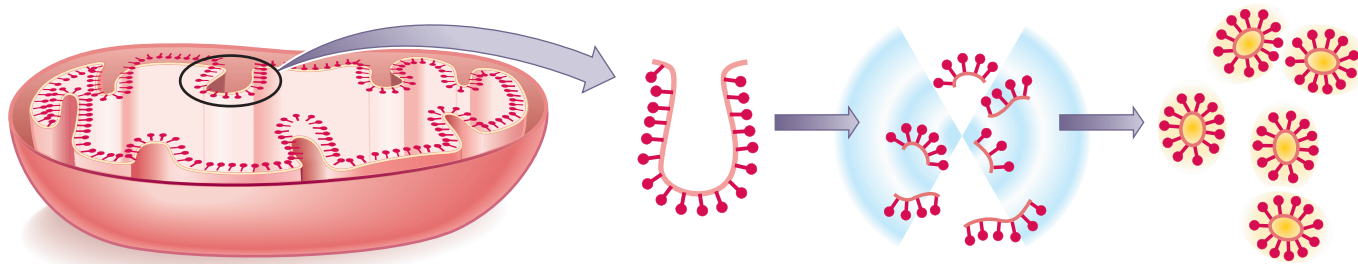
Several mechanisms have been proposed to account for the coupling of electron transport and ATP production. The mechanism that served as the point of departure in all discussions is chemiosmotic coupling, which was later modified to include a consideration of conformational coupling.

### What Is chemiosmotic coupling?

As originally proposed, the **chemiosmotic coupling** mechanism was based entirely on the difference in proton concentration between the intermembrane space and the matrix of an actively respiring mitochondrion. In other words, the proton (hydrogen ion,  $\text{H}^+$ ) gradient across the inner mitochondrial membrane is the crux of the matter. The proton gradient exists because the various proteins that serve as electron carriers in the respiratory chain are not symmetrically oriented with respect to the two sides of the inner mitochondrial membrane, nor do they react in the same way with respect to the matrix and the intermembrane space (Figure 20.15). Note that Figure 20.15 repeats the information found in Figure 20.8, with the addition of the flow of protons. The number of protons transported by respiratory complexes is uncertain and even a matter of some controversy. Figure 20.15 shows a consensus estimate for each complex. In the process of electron transport, the proteins of the respiratory complexes take up protons from the matrix to transfer them in redox reactions; these electron carriers subsequently release protons into the intermembrane space when they are reoxidized, creating the proton gradient. As a result, there



■ **FIGURE 20.15** The creation of a proton gradient in chemiosmotic coupling. The overall effect of the electron transport reaction series is to move protons ( $\text{H}^+$ ) out of the matrix into the intermembrane space, creating a difference in pH across the membrane.

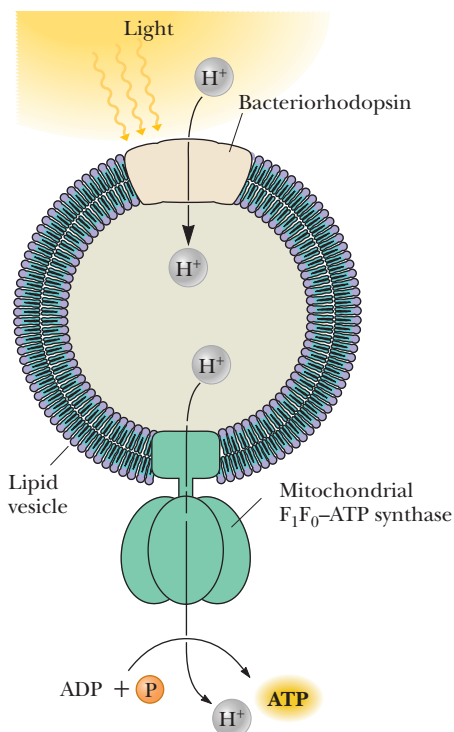


■ **FIGURE 20.16** Closed vesicles prepared from mitochondria can pump protons and produce ATP.

is a higher concentration of protons in the intermembrane space than in the matrix, which is precisely what we mean by a proton gradient. It is known that the intermembrane space has a lower pH than the matrix, which is another way of saying that there is a higher concentration of protons in the intermembrane space than in the matrix. The proton gradient in turn can drive the production of ATP that occurs when the protons flow back into the matrix.

Since chemiosmotic coupling was first suggested by the British scientist Peter Mitchell in 1961, a considerable body of experimental evidence has accumulated to support it.

1. A system with definite inside and outside compartments (closed vesicles) is essential for oxidative phosphorylation. The process does not occur in soluble preparations or in membrane fragments without compartmentalization.
2. Submitochondrial preparations that contain closed vesicles can be produced; such vesicles can carry out oxidative phosphorylation, and the asymmetrical orientation of the respiratory complexes with respect to the membrane can be demonstrated (Figure 20.16).
3. A model system for oxidative phosphorylation can be constructed with proton pumping in the absence of electron transport. The model system consists of reconstituted membrane vesicles, mitochondrial ATP synthase, and a proton pump. The pump is bacteriorhodopsin, a protein found in the membrane of halobacteria. The proton pumping takes place when the protein is illuminated (Figure 20.17).
4. The existence of the pH gradient has been demonstrated and confirmed experimentally.

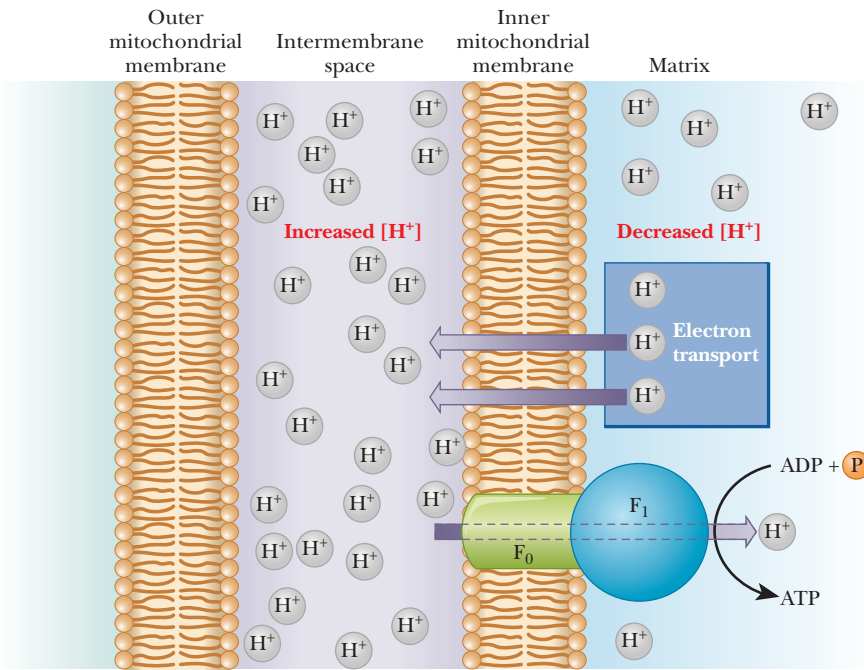


■ **FIGURE 20.17** ATP can be produced by closed vesicles with bacteriorhodopsin as a proton pump.

The way in which the proton gradient leads to the production of ATP depends on ion channels through the inner mitochondrial membrane; these channels are a feature of the structure of ATP synthase. Protons flow back into the matrix through ion channels in the ATP synthase; the  $F_0$  part of the protein is the proton channel. The flow of protons is accompanied by formation of ATP, which takes place in the  $F_1$  unit (Figure 20.18). The unique feature of chemiosmotic coupling is the direct linkage of the proton gradient to the phosphorylation reaction. The details of the way in which phosphorylation takes place as a result of the linkage to the proton gradient are not explicitly specified in this mechanism.

A reasonable mode of action for uncouplers can be proposed in light of the existence of a proton gradient. Dinitrophenol is an acid; its conjugate base, dinitrophenolate anion, is the actual uncoupler because it can react with protons in the intermembrane space, reducing the difference in proton concentration between the two sides of the inner mitochondrial membrane.

The antibiotic uncouplers, such as gramicidin A and valinomycin, are **ionophores**, creating a channel through which ions such as  $H^+$ ,  $K^+$ , and  $Na^+$  can pass through the membrane. The proton gradient is overcome, resulting in the uncoupling of oxidation and phosphorylation. The Biochemical Connections box on page 580 discusses a natural uncoupler.

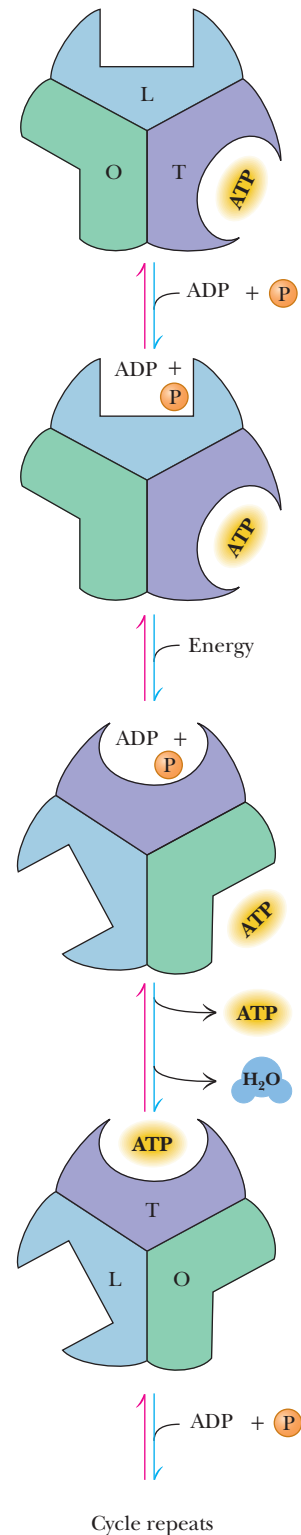


■ **FIGURE 20.18** Formation of ATP accompanies the flow of protons back into the mitochondrial matrix.

### What Is conformational coupling?

In **conformational coupling**, the proton gradient is indirectly related to ATP production. The proton gradient leads to conformational changes in a number of proteins, particularly in the ATP synthase itself. Recent evidence appears to show that the proton gradient is involved in the release of tightly bound ATP from the synthase as a result of the conformational change (Figure 20.19). There are three sites for substrate on the synthase and three possible conformational states: open (O), with low affinity for substrate; loose-binding (L), which is not catalytically active; and tight-binding (T), which is catalytically active. At any given time, each site is in one of three different conformational states. These states interconvert as a result of the proton flux through the synthase. ATP already formed by the synthase is bound at a site in the T conformation, while ADP and  $P_i$  bind at a site in the L conformation. A proton flux converts the site in the T conformation to the O conformation, releasing the ATP. The site at which ADP and  $P_i$  are bound assumes the T conformation, which can then give rise to ATP. More recently, it has been shown that the  $F_1$  portion of ATP synthase acts as a rotary motor. The  $c$ ,  $\gamma$ , and  $\epsilon$  subunits constitute the rotor, turning within a stationary barrel of a domain consisting of the subunit in association with the  $\alpha_3\beta_3$  hexamer and the  $a$  and  $b$  subunits (refer to Figure 20.13 for a detailed picture of the subunits). The  $\gamma$  and  $\epsilon$  subunits constitute the rotating “shaft” that mediates the energy exchange between the proton flow at  $F_0$  and ATP synthesis at  $F_1$ . In essence, the chemical energy of the proton gradient is converted to mechanical energy in the form of the rotating proteins. This mechanical energy is then converted to the chemical energy stored in the high-energy phosphate bonds of ATP.

Electron micrographs have shown that the conformation of the inner mitochondrial membrane and cristae is distinctly different in the resting and active states. It is well established that the shape of mitochondria is not static. This evidence long supported the idea that conformational changes play a role in the coupling of oxidation and phosphorylation.



■ **FIGURE 20.19** The role of conformational change in releasing the ATP from ATP synthase. According to the binding change mechanism, the effect of the binding flux is to cause a conformational change that leads to the release of already formed ATP from ATP synthase.



## Biochemical Connections NUTRITION

### What Does Brown Adipose Tissue Have to Do with Obesity?

When electron transport generates a proton gradient, some of the energy produced takes the form of heat. Dissipation of energy as heat is useful to organisms in two situations: cold-induced nonshivering thermogenesis (production of heat) and diet-induced thermogenesis. Cold-induced nonshivering thermogenesis enables animals to survive in the cold once they have become adapted to such conditions, and diet-induced thermogenesis prevents the development of obesity in spite of prolonged overeating. (Energy is dissipated as heat as food molecules are metabolized instead of being stored as fat.) These two processes may be the same biochemically. It is firmly established that they occur principally, if not exclusively, in brown adipose tissue (BAT), which is rich in mitochondria. (Brown fat takes its color from the large number of mitochondria present in it, unlike the usual white fat cells.) The key to this “inefficient” use of energy in brown adipose tissue appears to be a mitochondrial protein called thermogenin, also referred to as the “uncoupling protein.” When this membrane-bound protein is activated in thermogenesis, it serves as a proton channel through the inner mitochondrial membrane. Like all other uncouplers, it “punches a hole” in the mitochondrial membrane and decreases the effect of the proton gradient. Protons flow back into the matrix through thermogenin, bypassing the ATP synthase complex.

Very little research on the biochemistry or physiology of brown adipose tissue has been done in humans. Most of the work on both obesity and adaptation to cold stress has been done on small mammals, such as rats, mice, and hamsters. What role, if any, brown fat deposits play in the development or prevention of obesity in humans is an open question for researchers. Recently,

researchers have devoted much energy toward identifying the gene that encodes the uncoupling protein involved in obesity. The ultimate goal of this research is to use the protein or drugs that alter its regulation to combat obesity.

Some researchers have also proposed a link between sudden infant death syndrome (SIDS), which is also known as crib death, and metabolism in brown fat tissue. They think that a lack of BAT, or a switch from BAT to normal adipose tissue too early, could lead to body-temperature cooling in a way that could affect the central nervous system.



© Dynamics Graphic Group/IT Stock/Alamy

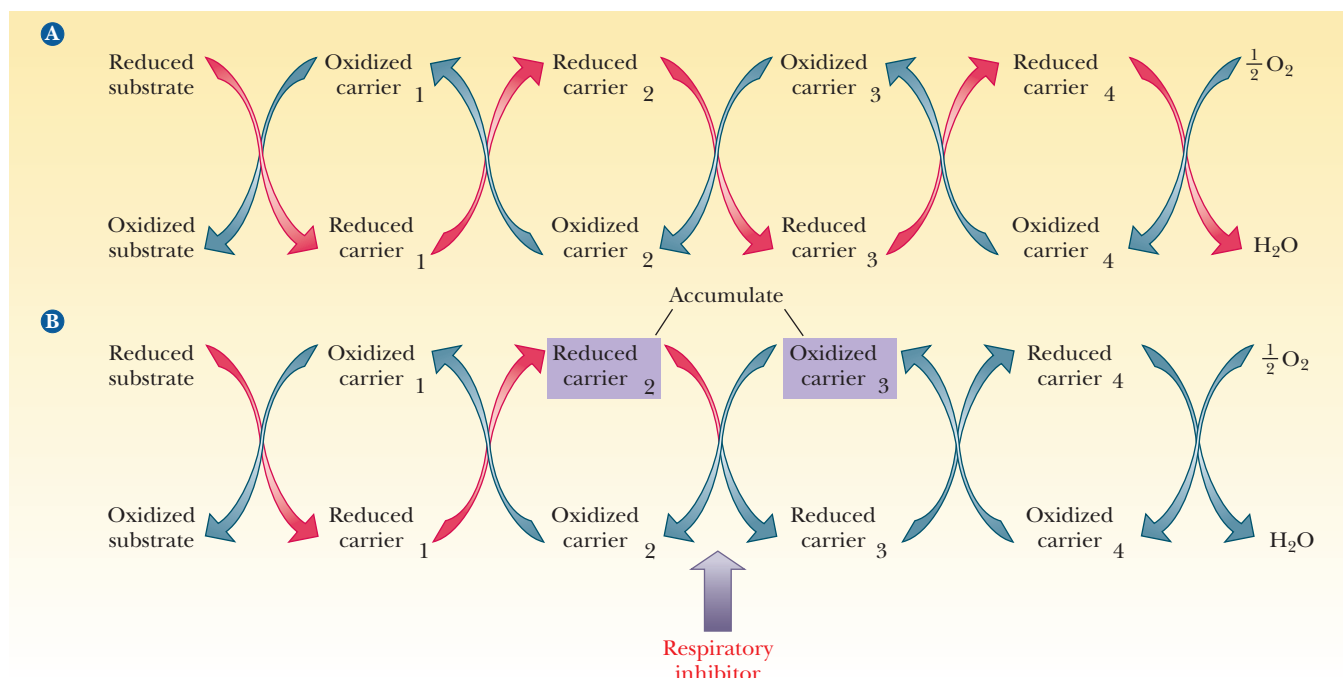
### 20.6 Respiratory Inhibitors Can Be Used to Study Electron Transport

If a pipeline is blocked, a backup occurs. Liquid accumulates upstream of the blockage point, and there is less liquid downstream. In electron transport, the flow of electrons is from one compound to another rather than along a pipe, but the analogy of a blocked pipeline can be useful for understanding the workings of the pathway. When a flow of electrons is blocked in a series of redox reactions, reduced compounds accumulate before the blockage point in the pathway. Recall that reduction is a gain of electrons, and oxidation represents a loss of electrons. The compounds that come after the blockage point will lack electrons and will tend to be found in the oxidized form (Figure 20.20). By using **respiratory inhibitors**, we can gather additional evidence to establish the order of components in the electron transport pathway.

#### *Do respiratory inhibitors have a connection with respiratory complexes?*

The use of respiratory inhibitors to determine the order of the electron transport chain depends on determining the relative amounts of oxidized and reduced forms of the various electron carriers in intact mitochondria. The logic of the experiment can be seen from the analogy of the blocked pipe. In this case, the reduced form of the carrier upstream (reduced carrier 2) accumulates because it cannot pass electrons farther in the chain. Likewise, the oxidized form of the carrier downstream (oxidized carrier 3) also accumulates



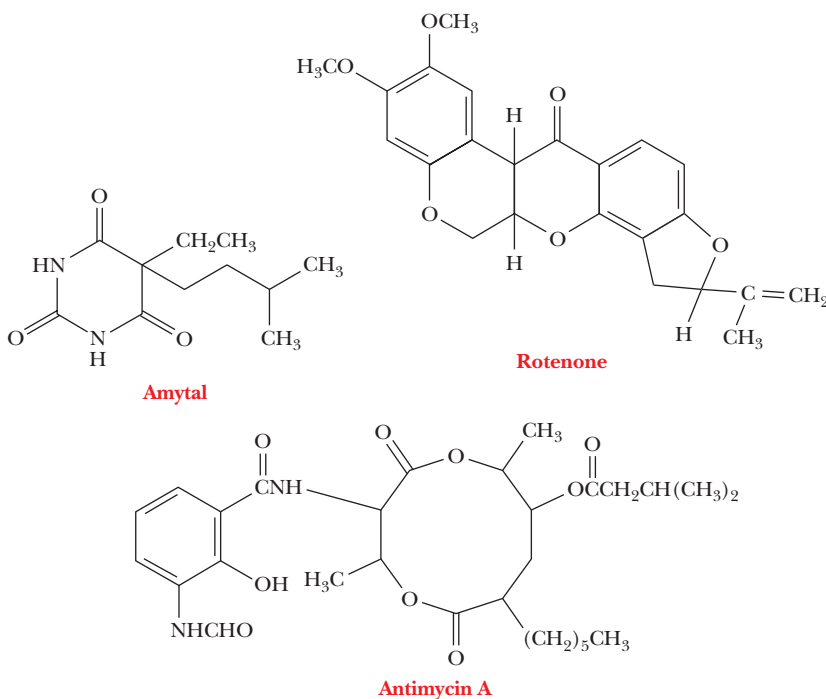


**FIGURE 20.20** The effect of respiratory inhibitors. (a) No inhibitor present. Schematic view of electron transport. The red arrows indicate the flow of electrons. (b) Inhibitor present. The flow of electrons from carrier 2 to carrier 3 is blocked by the respiratory inhibitor. Reduced carrier 2 accumulates, as does oxidized carrier 3, because they cannot react with each other.

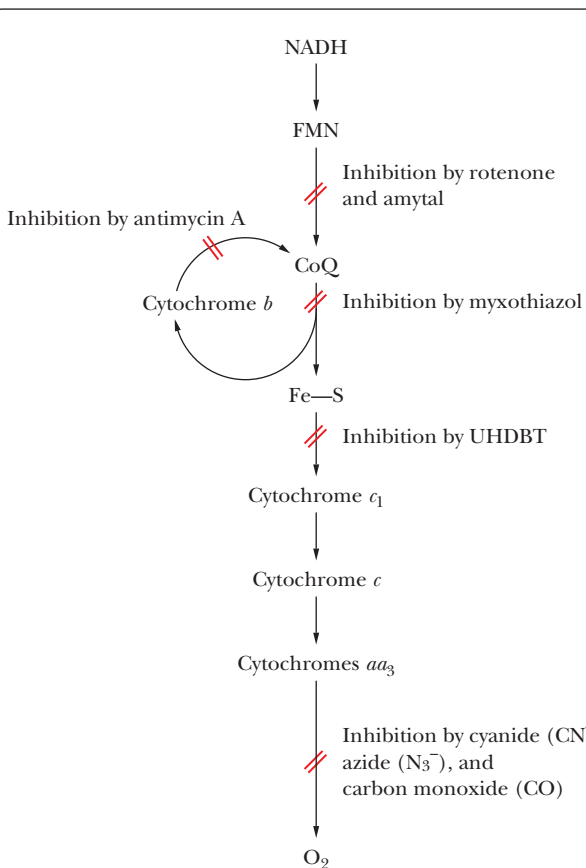
because the supply of electrons that it could accept has been cut off (Figure 20.20). By use of careful techniques, intact mitochondria can be isolated from cells and can carry out electron transport if an oxidizable substrate is available. If electron transport in mitochondria occurs in the presence and absence of a respiratory inhibitor, different relative amounts of oxidized and reduced forms of the electron carriers will be present.

The type of experiment done to determine the relative amounts of oxidized and reduced forms of electron carriers depends on the spectroscopic properties of these substances. The oxidized and reduced forms of cytochromes can be distinguished from one another. Specialized spectroscopic techniques exist to detect the presence of electron carriers in intact mitochondria. The individual types of cytochromes can be identified by the wavelength at which the peak appears, and the relative amounts can be determined from the intensities of the peaks.

Inhibitors have an effect at three sites in the electron transport chain, and we shall look at some classic examples. At the first site, barbiturates (of which amytal is an example) block the transfer of electrons from the flavoprotein NADH reductase to coenzyme Q. Rotenone is another inhibitor that is active at this site. This compound is used as an insecticide; it is highly toxic to fish, but not to humans, and is often used to kill the fish in a lake before introducing fish of a different species. The second site at which blockage can occur is that of electron transfer involving the *b* cytochromes, coenzyme Q, and cytochrome  $c_1$ . The classic inhibitor associated with this blockage is the antibiotic antimycin A (Figure 20.21). More recently developed inhibitors that are active in this part of the electron transport chain include myxothiazol and 5-*n*-undecyl-6-hydroxy-4,7-dioxobenzothiazol (UHDBT). These compounds played a role in establishing the existence of the Q cycle. The third site subject to blockage is the transfer of electrons from the cytochrome  $aa_3$  complex to oxygen. Several potent inhibitors operate at this site (Figure 20.22), such as cyanide ( $\text{CN}^-$ , azide ( $\text{N}_3^-$ ), and carbon monoxide ( $\text{CO}$ ). Note that each of the three



■ **FIGURE 20.21** Structures of some respiratory inhibitors.



■ **FIGURE 20.22** Sites of action of some respiratory inhibitors.

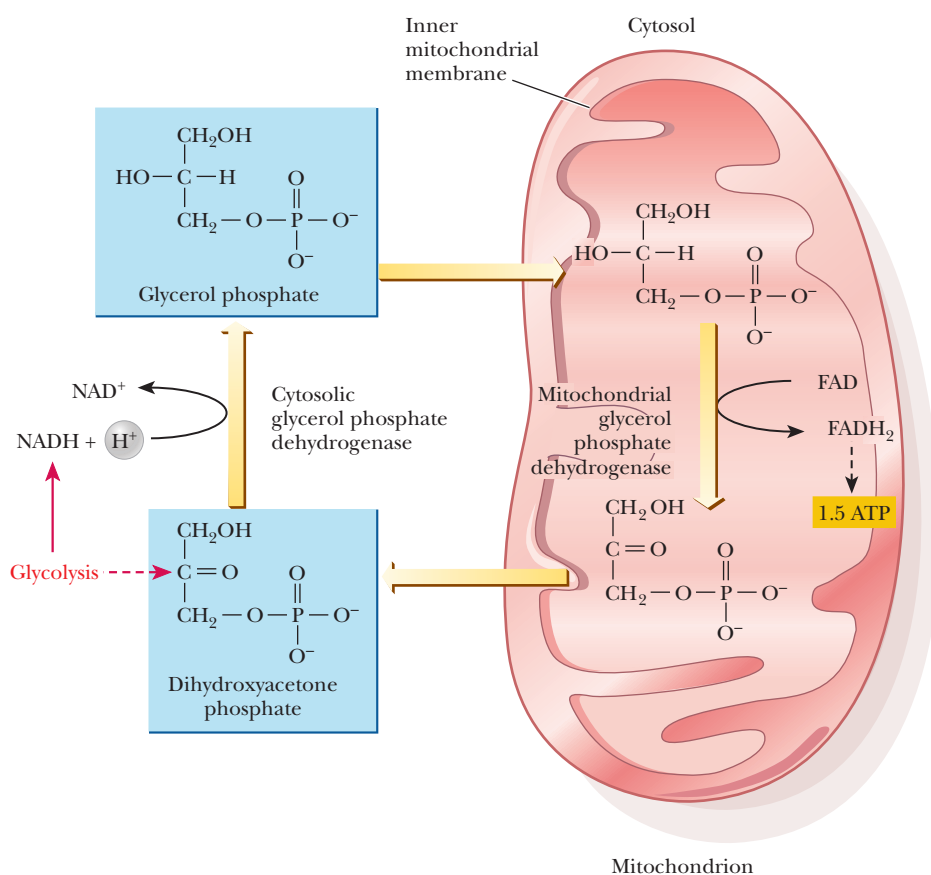
sites of action of respiratory inhibitors corresponds to one of the respiratory complexes. Research is continuing with some of the more recently developed inhibitors; the goal of additional work is to elucidate more of the details of the electron transport process.

## 20.7 Shuttle Mechanisms

NADH is produced by glycolysis, which occurs in the cytosol, but NADH in the cytosol cannot cross the inner mitochondrial membrane to enter the electron transport chain. However, the electrons can be transferred to a carrier that can cross the membrane. The number of ATP molecules generated depends on the nature of the carrier, which varies according to the type of cell in which it occurs.

### How do shuttle mechanisms differ from one another?

One carrier system that has been extensively studied in insect flight muscle is the **glycerol-phosphate shuttle**. This mechanism uses the presence on the outer face of the inner mitochondrial membrane of an FAD-dependent enzyme that oxidizes glycerol phosphate. The glycerol phosphate is produced by the reduction of dihydroxyacetone phosphate; in the course of the reaction, NADH is oxidized to  $\text{NAD}^+$ . In this reaction, the oxidizing agent (which is itself reduced) is FAD, and the product is  $\text{FADH}_2$  (Figure 20.23). The  $\text{FADH}_2$  then passes electrons through the electron transport chain, leading to the production

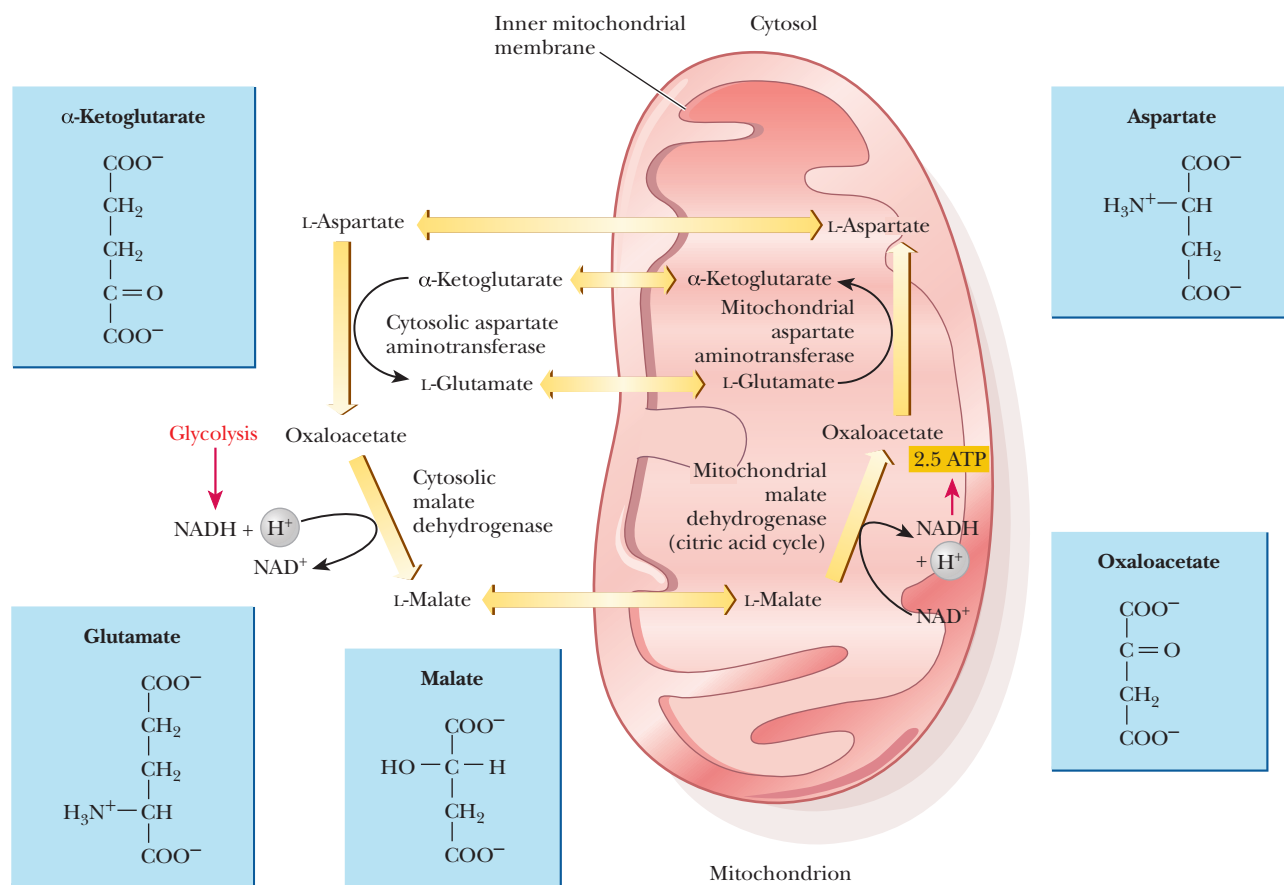


■ **FIGURE 20.23** The glycerol-phosphate shuttle.

of 1.5 moles of ATP for each mole of cytosolic NADH. This mechanism has also been observed in mammalian muscle and brain.

A more complex and more efficient shuttle mechanism is the **malate–aspartate shuttle**, which has been found in mammalian kidney, liver, and heart. This shuttle uses the fact that malate can cross the mitochondrial membrane, while oxaloacetate cannot. The noteworthy point about this shuttle mechanism is that the transfer of electrons from NADH in the cytosol produces NADH in the mitochondrion. In the cytosol, oxaloacetate is reduced to malate by the cytosolic malate dehydrogenase, accompanied by the oxidation of cytosolic NADH to  $\text{NAD}^+$  (Figure 20.24). The malate then crosses the mitochondrial membrane. In the mitochondrion, the conversion of malate back to oxaloacetate is catalyzed by the mitochondrial malate dehydrogenase (one of the enzymes of the citric acid cycle). Oxaloacetate is converted to aspartate, which can also cross the mitochondrial membrane. Aspartate is converted to oxaloacetate in the cytosol, completing the cycle of reactions.

The NADH that is produced in the mitochondrion thus passes electrons to the electron transport chain. With the malate–aspartate shuttle, 2.5 moles of ATP are produced for each mole of cytosolic NADH rather than 1.5 moles of ATP in the glycerol–phosphate shuttle, which uses  $\text{FADH}_2$  as a carrier. The following Biochemical Connections box discusses some practical applications of our understanding of the catabolic pathways.



■ **FIGURE 20.24** The malate–aspartate shuttle.

# Biochemical Connections ALLIED HEALTH

## Sports and Metabolism

Trained athletes, especially at the elite level, are more aware of the results of anaerobic and aerobic metabolism than nonathletes. Genetic endowment and training are important to the success of the athlete, but a keen understanding of physiology and metabolism is equally important. To plan nutrition for performance, a serious athlete must understand the nature of metabolism as it relates to his or her chosen sport. A working muscle has four different sources of energy available after a period of rest:

1. Creatine phosphate, which reacts directly with ADP in substrate-level phosphorylation to produce ATP.
2. Glucose from the glycogen of muscle stores, initially consumed by anaerobic metabolism.
3. Glucose from the liver, from both its glycogen stores and the gluconeogenesis from lactic acid produced in the muscle (the Cori cycle), again initially consumed by anaerobic metabolism.
4. Aerobic metabolism in the muscle mitochondria.

Initially, all four energy sources are available to the muscle. When the creatine phosphate runs out, only the other sources are left. When muscle glycogen runs out, the anaerobic boost it provided slows down, and when the liver glycogen is gone, only aerobic metabolism to carbon dioxide and water is left.

It is difficult to accurately calculate how much each of these nutrients might supply to a rapidly working muscle, but simple calculations are consistent with there being less than a 1-minute supply of creatine phosphate, a figure that can be compared with the length of time for sprint events, which is typically less than a minute. Creatine supplements for athletes are sold in health food stores, and results suggest that for power lifting or short sprints, such as the 100-meter dash, this supplementation is effective. There is about 10 to 30 minutes' worth of glycogen in the muscle cells, with this figure varying dramatically based on the intensity of the exercise. Performance in running events ranging in distance from 1500 meters up to 10 kilometers can be heavily influenced by the muscle glycogen levels at the start of the event. Of course, glycogen loading (Chapter 18) could affect this last figure significantly. One reason for the difficulty in making these calculations is the uncertainty of what proportion of the liver glycogen is metabolized only to lactic acid and how much is metabolized in the liver. It is known that one rate-limiting step for aerobic metabolism is the shuttling of both NADH and pyruvate from the cytoplasm into the mitochondrion.

In this regard, it is interesting to note that well-conditioned and well-trained athletes actually have a higher number of mitochondria in their muscle cells. For long-distance events, such as the marathon or road-cycling events, aerobic metabolism certainly comes into play. "Fat burning" is the term frequently used, and it reflects metabolic fact. Fatty acids are degraded to acetyl-CoA, which then enters the citric acid cycle; marathoners and cyclists are known for their notably lean frames, with a minimal amount of body fat. Running a marathon, which takes between two and three hours for very fit runners, uses more fatty acids, and is done at a lower level of oxygen uptake, than riding in a professional road-cycling event that may take up to seven hours. Clearly, there are differences in metabolism for sports even within the category known as endurance events.

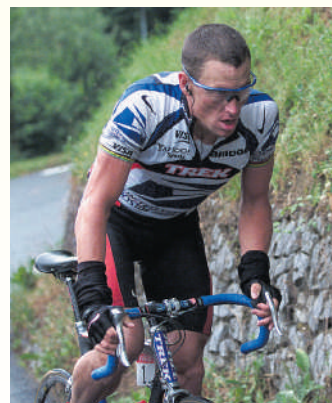
Perhaps the most studied athlete of modern times is cyclist Lance Armstrong. As a young elite rider, he was a world professional road race champion in 1993 and won a few stages of the prestigious Tour de France. He was powerfully built and excelled in time-trial events and single-day road races, but he was never considered a threat in the major stage races because he did not climb the major European mountains very well. After suffering a disappointing

Olympic Games in Atlanta in 1996, he was diagnosed with testicular cancer, which had spread to several organs, including his brain. He was given little chance to live, but after several surgeries and intense chemotherapy, he recovered and resumed his cycling career. The hospitalization and chemotherapy caused him to lose 15 to 20 pounds, and, in 1998, he became not only competitive but also a true challenger again at the World Championships and the Tour of Spain. Cycling fans were surprised at how well he climbed the Spanish Pyrenees, but few suspected that in the next few years he would go on to become the first cyclist to win seven Tour de France titles. Before Armstrong, nobody had ever even won six! After retiring in 2005, he made a comeback in 2009. As much an elder statesman as a contender, he amazingly finished third. At the age of 38, he was set to lead a new team into the 2010 Tour de France.

Lance Armstrong was the second American to win this event. He always was an amazing aerobic machine, and, when his metabolism didn't have to carry as much weight up the mountains, he was able to climb with the best in the world. Of course, his true strength came from his will to win, which he credits to his ordeal with cancer and his upbringing.

Underscoring the importance of the electron transport chain and mitochondria to the athlete is the story of another great cycling champion, Greg LeMond, who was the first American to win the Tour de France and went on to win it a total of three times. Like Lance Armstrong, Greg LeMond also had a tragedy in the middle of his career, a hunting accident. He was shot in the back with buckshot shortly after his first Tour de France victory. Remarkably, he recovered and went on to win two more Tours de France. However, he never really felt well again, and he later commented that, even in his final Tour de France victory in 1990, something was definitely wrong. The next few years were disappointing for LeMond and his fans, and he never made it back into the top places of a race. He seemed to be putting on weight and didn't respond to training. Finally, in desperation, he underwent some painful muscle biopsies and discovered that he had a rare condition called mitochondrial myopathy. When he trained hard, his mitochondria began to disappear. He was essentially an aerobic athlete without the ability to process fuels aerobically. He retired from competition shortly thereafter.

© Reuters/CORBIS



■ Cancer survivor and champion cyclist Lance Armstrong on his way to a Tour de France victory. Lance retired in 2005 after his seventh consecutive tour victory but returned to compete in the tour in 2009, finishing third.

© AP Photo/Lionel Cironneau



■ Winner of the 1986, 1989, and 1990 Tour de France, Greg LeMond was the first American to win the world's most prestigious bicycle race.



**TABLE 20.3**

Yield of ATP from Glucose Oxidation				
Pathway	ATP Yield per Glucose		NADH	FADH <sub>2</sub>
	Glycerol-Phosphate Shuttle	Malate-Aspartate Shuttle		
<b>Glycolysis: glucose to pyruvate (cytosol)</b>				
Phosphorylation of glucose	−1	−1		
Phosphorylation of fructose-6-phosphate	−1	−1		
Dephosphorylation of 2 molecules of 1,3-BPG	+2	+2		
Dephosphorylation of 2 molecules of PEP	+2	+2		
Oxidation of 2 molecules of glyceraldehyde-3-phosphate yields 2 NADH			+2	
<b>Pyruvate conversion to acetyl-CoA (mitochondria)</b>				
2 NADH produced			+2	
<b>Citric acid cycle (mitochondria)</b>				
2 molecules of GTP from 2 molecules of succinyl-CoA	+2	+2		
Oxidation of 2 molecules each of isocitrate, α-ketoglutarate, and malate yields 6 NADH			+6	
Oxidation of 2 molecules of succinate yields 2 FADH <sub>2</sub>				+2
<b>Oxidative phosphorylation (mitochondria)</b>				
2 NADH from glycolysis yield 1.5 ATP each if NADH is oxidized by glycerol-phosphate shuttle; 2.5 ATP by malate-aspartate shuttle	+3	+5	−2	
Oxidative decarboxylation of 2 pyruvate to 2 acetyl-CoA: 2 NADH produce 2.5 ATP each	+5	+5	−2	
2 FADH <sub>2</sub> from each citric acid cycle produce 1.5 ATP each	+3	+3		−2
6 NADH from citric acid cycle produce 2.5 ATP each	+15	+15	−6	
Net Yield	+30	+32	0	0

(Note: These P/O ratios of 2.5 and 1.5 for mitochondrial oxidation of NADH and FADH<sub>2</sub> are “consensus values.” Since they may not reflect actual values and since these ratios may change depending on metabolic conditions, these estimates of ATP yield from glucose oxidation are approximate.)

## 20.8 The ATP Yield from Complete Oxidation of Glucose

In Chapters 17 through 20, we have discussed many aspects of the complete oxidation of glucose to carbon dioxide and water. At this point, it is useful to do some bookkeeping to see how many molecules of ATP are produced for each molecule of glucose oxidized. Recall that some ATP is produced in glycolysis, but that far more ATP is produced by aerobic metabolism. Table 20.3 summarizes ATP production and also follows the reoxidation of NADH and FADH<sub>2</sub>.

## SUMMARY

**What is the importance of mitochondrial structure in ATP production?** In the final stages of aerobic metabolism, electrons are transferred from NADH to oxygen (the ultimate electron acceptor) in a series of oxidation–reduction reactions known as the electron transport chain. In the process, protons are pumped across the inner mitochondrial membrane. This

series of events depends on the presence of oxygen in the final step. This pathway allows for the reoxidation of the reduced electron carriers produced in glycolysis, the citric acid cycle, and several other catabolic pathways, and is the true source of the ATPs produced by catabolism. Phosphorylation depends on the compartmented structure of mitochondria.

**How can reduction potentials be used to predict the direction of electron transport?**

The overall reaction of the electron transport chain shows a very large, negative  $\Delta G^\circ$  due to the large differences in reduction potentials between the reactions involving NADH and those involving oxygen. If NADH were to reduce oxygen directly, the  $\Delta E^\circ$  would be more than 1 V. In reality, many redox reactions are in between, and the correct order of events in the electron transport chain was predicted by comparing the reduction potentials of the individual reactions long before the order was established experimentally.

**What reactions take place in the respiratory complexes?**

Four separate respiratory complexes can be isolated from the inner mitochondrial membrane. Each can carry out the reactions of a portion of the electron transport chain. In addition to the respiratory complexes, two electron carriers, coenzyme Q and cytochrome *c*, are not bound to the complexes but are free to move within and along the membrane, respectively. Complex I accomplishes the reoxidation of NADH and sends electrons to coenzyme Q. Complex II reoxidizes  $\text{FADH}_2$  and also sends electrons to CoQ. Complex III involves the Q cycle and shuttles electrons to cytochrome *c*. Complex IV takes the electrons from cytochrome *c* and passes them to oxygen in the final step of electron transport.

**What is the nature of the iron-containing proteins of electron transport?** A number of iron-containing proteins are part of the electron transport chain. In the cytochrome proteins, the iron is bound to a heme group. In other proteins, the iron is bound to the protein along with sulfur.

**What is the coupling factor in oxidative phosphorylation?**

A complex protein oligomer is the coupling factor that links oxidation and phosphorylation. The complete protein spans the inner mitochondrial membrane and projects into the matrix as well. The portion of the protein that spans the membrane is called  $F_0$ ; it consists of three different kinds of polypeptide chains (a, b, and c).

The portion that projects into the matrix is called  $F_1$ ; it consists of five different kinds of polypeptide chains ( $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\delta$ , and  $\epsilon$ , in the ratio  $\alpha_3\beta_3\gamma\delta\epsilon$ ). The  $F_1$  sphere is the site of ATP synthesis. The whole protein complex is called ATP synthase. It is also known as mitochondrial ATPase. During the process of electron transport, several reactions occur in which reduced carriers that have both electrons and protons to donate are linked to carriers that can only accept electrons. At these points, hydrogen ions are released to the other side of the inner mitochondrial membrane, causing the formation of a pH gradient. The energy inherent in the charge and chemical separation of the hydrogen ions is used to phosphorylate ADP to ATP when the hydrogen ions pass back into the mitochondria through ATP synthase.

**What is chemiosmotic coupling?** Two mechanisms, the chemiosmotic mechanism and the conformational coupling mechanism, have been proposed to explain the coupling of electron transport and ATP production.

Chemiosmotic coupling is the mechanism most widely used to explain the manner in which electron transport and oxidative phosphorylation are coupled to one another. In this mechanism, the proton gradient is directly linked to the phosphorylation process. The way in which the proton gradient leads to the production of ATP depends on ion channels through the inner mitochondrial membrane; these channels are a feature of the structure of ATP synthase. Protons flow back into the matrix through proton channels in the  $F_0$  part of the ATP synthase. The flow of protons is accompanied by formation of ATP, which occurs in the  $F_1$  unit.

**What is conformational coupling?** In the conformational coupling mechanism, the proton gradient is indirectly related to ATP production. Recent evidence appears to show that the effect of the proton gradient is not the formation of ATP but the release of tightly bound ATP from the synthase as a result of the conformational change.

**Do respiratory inhibitors have a connection with respiratory complexes?** Many of the workings of the electron transport chain have been elucidated by experiments using respiratory inhibitors. These inhibitors specifically block the transfer of electrons at specific points in the respiratory complexes. Examples are CO and  $\text{CN}^-$ , both of which block the final step of the electron transport chain, and rotenone, which blocks the transfer of electrons from NADH reductase to coenzyme Q. When such a blockage occurs, it causes electrons to “pile up” behind the block, giving a reduced carrier that cannot be oxidized. By noting which carriers become trapped in a reduced state and which ones are trapped in an oxidized state, we can establish the link between carriers.

**How do shuttle mechanisms differ from one another?** Two shuttle mechanisms—the glycerol–phosphate shuttle and the malate–aspartate shuttle—transfer the electrons, but not the NADH, produced in cytosolic reactions into the mitochondrion. In the first of the two shuttles, which is found in muscle and brain, the electrons are transferred to FAD; in the second, which is found in kidney, liver, and heart, the electrons are transferred to  $\text{NAD}^+$ . With the malate–aspartate shuttle, 2.5 molecules of ATP are produced for each molecule of cytosolic NADH, rather than 1.5 ATP in the glycerol–phosphate shuttle, a point that affects the overall yield of ATP in these tissues.

Approximately 2.5 molecules of ATP are generated for each molecule of NADH that enters the electron transport chain and approximately 1.5 molecules of ATP for each molecule of  $\text{FADH}_2$ . When glucose is metabolized anaerobically, the only net ATPs that are produced are those from the substrate-level phosphorylation steps. This leads to a total of only two ATPs per glucose entering glycolysis. When the pyruvate generated from glycolysis can enter the citric acid cycle, and the resulting NADH and  $\text{FADH}_2$  molecules are re-oxidized through the electron transport chain, a total of 30 or 32 ATPs are produced, with the difference being due to the two possible shuttles.

## REVIEW EXERCISES

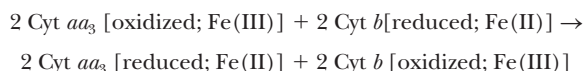
 Interactive versions of these problems are assignable in OWL

## 20.1 The Role of Electron Transport in Metabolism

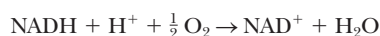
- Recall** Briefly summarize the steps in the electron transport chain from NADH to oxygen.
- Recall** Are electron transport and oxidative phosphorylation the same process? Why or why not?
- Reflect and Apply** List the reactions of electron transport that liberate enough energy to drive the phosphorylation of ADP.
- Reflect and Apply** Show how the reactions of the electron transport chain differ from those in Question 3 when  $\text{FADH}_2$  is the starting point for electron transport. Show how the reactions that liberate enough energy to drive the phosphorylation of ADP differ from the pathway when NADH is the starting point.
- Reflect and Apply** How does mitochondrial structure contribute to aerobic metabolism, particularly to the integration of the citric acid cycle and electron transport?

## 20.2 Reduction Potentials in the Electron Transport Chain

- Recall** Why is it reasonable to compare the electron transport process to a battery?
- Recall** Why are all the reactions in Table 20.1 written as reduction reactions?
- Mathematical** Using the information in Table 20.2, calculate  $\Delta G^\circ$  for the following reaction:



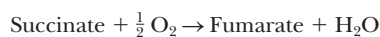
- Mathematical** Calculate  $E^\circ$  for the following reaction:



- Mathematical** Calculate  $E^\circ$  for the following reaction:



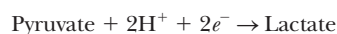
- Mathematical** Calculate  $E^\circ$  for the following reaction:



- Mathematical** For the following reaction, identify the electron donor and the electron acceptor and calculate  $E^\circ$ .



- Mathematical** Which is more favorable energetically, the oxidation of succinate to fumarate by  $\text{NAD}^+$  or by FAD? Give the reason for your answer.
- Reflect and Apply** Comment on the fact that the reduction of pyruvate to lactate, catalyzed by lactate dehydrogenase, is strongly exergonic (recall this from Chapter 15), even though the standard free-energy change for the half reaction



is positive ( $\Delta G^\circ = 36.2 \text{ kJ mol}^{-1} = 8.8 \text{ kcal mol}^{-1}$ ), indicating an endergonic reaction.

## 20.3 Organization of Electron Transport Complexes

- Recall** What do cytochromes have in common with hemoglobin or myoglobin?
- Recall** How do the cytochromes differ from hemoglobin and myoglobin in terms of chemical activity?
- Recall** Which of the following does not play a role in respiratory complexes: cytochromes, flavoproteins, iron-sulfur proteins, or coenzyme Q?

- Recall** Do any of the respiratory complexes play a role in the citric acid cycle? If so, what is that role?
- Recall** Do all the respiratory complexes generate enough energy to phosphorylate ADP to ATP?
- Reflect and Apply** Two biochemistry students are about to use mitochondria isolated from rat liver for an experiment on oxidative phosphorylation. The directions for the experiment specify addition of purified cytochrome *c* from any source to the reaction mixture. Why is the added cytochrome *c* needed? Why does the source not have to be the same as that of the mitochondria?
- Reflect and Apply** Cytochrome oxidase and succinate-CoQ oxidoreductase are isolated from mitochondria and are incubated in the presence of oxygen, along with cytochrome *c*, coenzyme Q, and succinate. What is the overall oxidation-reduction reaction that can be expected to take place?
- Reflect and Apply** What are two advantages of the components of the electron transport chain being embedded in the inner mitochondrial membrane?
- Reflect and Apply** Reflect on the evolutionary implications of the structural similarities and functional differences of cytochromes on the one hand and hemoglobin and myoglobin on the other.
- Reflect and Apply** Experimental evidence strongly suggests that the protein portions of cytochromes have evolved more slowly (as judged by the number of changes in amino acids per million years) than the protein portions of hemoglobin and myoglobin and even more slowly than hydrolytic enzymes. Suggest a reason why.
- Reflect and Apply** What is the advantage of having mobile electron carriers in addition to large membrane-bound complexes of carriers?
- Reflect and Apply** What is the advantage of having a Q cycle in electron transport in spite of its complexity?
- Reflect and Apply** Why do the electron-transfer reactions of the cytochromes differ in standard reduction potential, even though all the reactions involve the same oxidation-reduction reaction of iron?
- Reflect and Apply** Is there a fundamental difference between the one- and two-electron reactions in the electron transport chain?
- Reflect and Apply** What is the underlying reason for the differences in spectroscopic properties among the cytochromes?
- Reflect and Apply** What would be some of the challenges involved in removing respiratory complexes from the inner mitochondrial membrane in order to study their properties?

## 20.4 The Connection between Electron Transport and Phosphorylation

- Recall** Describe the role of the  $\text{F}_1$  portion of ATP synthase in oxidative phosphorylation.
- Recall** Is mitochondrial ATP synthase an integral membrane protein?
- Recall** Define *P/O ratio* and indicate why it is important.
- Recall** In what sense is mitochondrial ATP synthase a motor protein?
- Reflect and Apply** What is the approximate P/O ratio that can be expected if intact mitochondria are incubated in the presence of oxygen, along with added succinate?
- Reflect and Apply** Why is it difficult to determine an exact number for P/O ratios?
- Reflect and Apply** What are some of the difficulties in determining the exact number of protons pumped across the inner mitochondrial membrane by the respiratory complexes?

### 20.5 The Mechanism of Coupling in Oxidative Phosphorylation

38. **Recall** Briefly summarize the main arguments of the chemiosmotic coupling hypothesis.
39. **Recall** Why does ATP production require an intact mitochondrial membrane?
40. **Biochemical Connections** Briefly describe the role of uncouplers in oxidative phosphorylation.
41. **Recall** What role does the proton gradient play in chemiosmotic coupling?
42. **Biochemical Connections** Why was dinitrophenol once used as a diet drug?
43. **Reflect and Apply** Criticize the following statement: “The role of the proton gradient in chemiosmosis is to provide the energy to phosphorylate ADP.”

### 20.6 Respiratory Inhibitors Can Be Used to Study Electron Transport

44. **Recall** What is the effect of each of the following substances on electron transport and production of ATP? Be specific about which reaction is affected.
  - (a) Azide
  - (b) Antimycin A
  - (c) Amytal
  - (d) Rotenone
  - (e) Dinitrophenol
  - (f) Gramicidin A
  - (g) Carbon monoxide

45. **Recall** How can respiratory inhibitors be used to indicate the order of components in the electron transport chain?
46. **Reflect and Apply** What is the fundamental difference between uncouplers and respiratory inhibitors?

### 20.7 Shuttle Mechanisms

47. **Recall** How does the yield of ATP from complete oxidation of one molecule of glucose in muscle and brain differ from that in liver, heart, and kidney? What is the underlying reason for this difference?
48. **Reflect and Apply** The malate–aspartate shuttle yields about 2.5 moles of ATP for each mole of cytosolic NADH. Why does nature use the glycerol–phosphate shuttle, which yields only about 1.5 moles of ATP?

### 20.8 The ATP Yield from Complete Oxidation of Glucose

49. **Mathematical** What yield of ATP can be expected from complete oxidation of each of the following substrates by the reactions of glycolysis, the citric acid cycle, and oxidative phosphorylation?
  - (a) Fructose-1,6-bisphosphate
  - (b) Glucose
  - (c) Phosphoenolpyruvate
  - (d) Glyceraldehyde-3-phosphate
  - (e) NADH
  - (f) Pyruvate
50. **Mathematical** The free-energy change ( $\Delta G^\circ$ ) for the oxidation of the cytochrome  $aa_3$  complex by molecular oxygen is  $-102.3 \text{ kJ} = -24.5 \text{ kcal}$  for each mole of electron pairs transferred. What is the maximum number of moles of ATP that could be produced in the process? How many moles of ATP are actually produced? What is the efficiency of the process, expressed as a percentage?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





# Lipid Metabolism

# 21

Lipid metabolism allows polar bears to thrive in arctic climates and to endure months of hibernation.

© Wayne R. Blanchard/Getty Images

## 21.1 Lipids Are Involved in the Generation and Storage of Energy

In the past few chapters we have seen how energy can be released by the catabolic breakdown of carbohydrates in aerobic and anaerobic processes. In Chapter 16, we saw that there are carbohydrate polymers (such as starch in plants and glycogen in animals) that represent stored energy, in the sense that these carbohydrates can be hydrolyzed to monomers and then oxidized to provide energy in response to the needs of an organism. In this chapter, we shall see how the metabolic oxidation of lipids releases large quantities of energy through production of acetyl-CoA, NADH, and FADH<sub>2</sub> and how lipids represent an even more efficient way of storing chemical energy.

## 21.2 Catabolism of Lipids

The oxidation of fatty acids is the chief source of energy in the catabolism of lipids; in fact, lipids that are sterols (steroids that have a hydroxyl group as part of their structure; Section 8.2) are not catabolized as a source of energy but are excreted. Both triacylglycerols, which are the main storage form of the chemical energy of lipids, and phosphoacylglycerols, which are important components of biological membranes, have fatty acids as part of their covalently bonded structures. In both types of compounds, the bond between the fatty acid and the rest of the molecule can be hydrolyzed (Figure 21.1), with the reaction catalyzed by suitable groups of enzymes—**lipases**, in the case of triacylglycerols (Section 8.2), and **phospholipases**, in the case of phosphoacylglycerols.

Several different phospholipases can be distinguished on the basis of the site at which they hydrolyze phospholipids (Figure 21.2). Phospholipase A<sub>2</sub> is widely distributed in nature; it is also being actively studied by biochemists interested in its structure and mode of action, which involves hydrolysis of phospholipids at the surface of micelles (Section 2.1). Phospholipase D occurs in spider venom and is responsible for the tissue damage that accompanies spider bites. Snake venoms also contain phospholipases; the concentration of phospholipases is particularly high in venoms with comparatively low concentrations of the toxins (usually small peptides) that are characteristic of some kinds of venom. The lipid products of hydrolysis lyse red blood cells, preventing clot formation. Snakebite victims bleed to death in this situation.

The release of fatty acids from triacylglycerols in adipocytes is controlled by hormones. In a scheme that will look familiar from our discussions of carbohydrate metabolism, a hormone binds to a receptor on the plasma membrane of the adipocyte (Figure 21.3). This hormone binding activates adenylate cyclase, which leads to production of active protein kinase A (cAMP-dependent protein kinase). Protein kinase phosphorylates triacylglycerol lipase, which cleaves the fatty acids from the glycerol backbone. The main hormone that has this effect is epinephrine. Caffeine also mimics epinephrine in this regard, which is

## Chapter Outline

### 21.1 Lipids Are Involved in the Generation and Storage of Energy

#### 21.2 Catabolism of Lipids

- How are fatty acids transported to the mitochondrion for oxidation?
- How does oxidation of fatty acids take place?

### 21.3 The Energy Yield from the Oxidation of Fatty Acids

#### 21.4 Catabolism of Unsaturated Fatty Acids and Odd-Carbon Fatty Acids

- How does the oxidation of unsaturated fatty acids differ from that of saturated fatty acids?

#### 21.5 Ketone Bodies

- Do acetone and acetyl-CoA have a connection in lipid metabolism?

#### 21.6 Fatty-Acid Biosynthesis

- How do the first steps of fatty-acid synthesis take place?
- What is the mode of action of fatty-acid synthase?

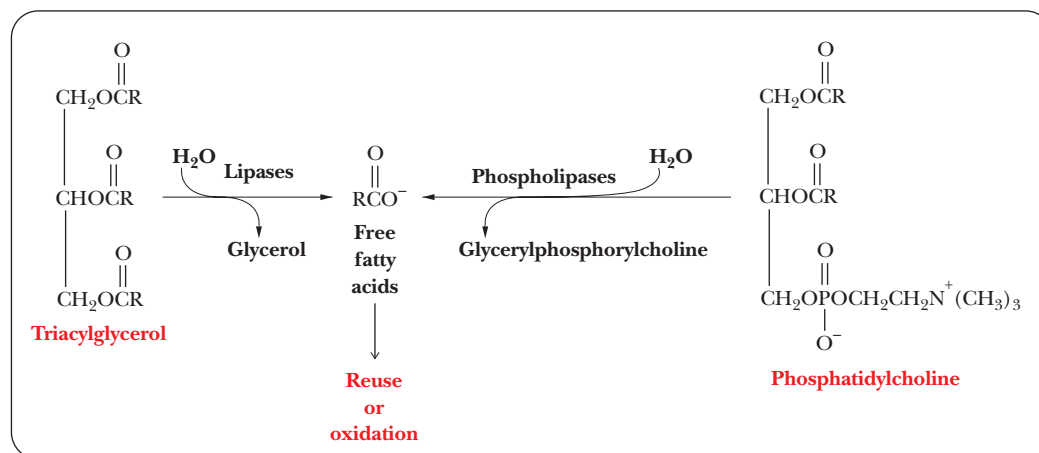
#### 21.7 Synthesis of Acylglycerols and Compound Lipids

- How does the biosynthesis of phosphoacylglycerols take place?
- How does the biosynthesis of sphingolipids take place?

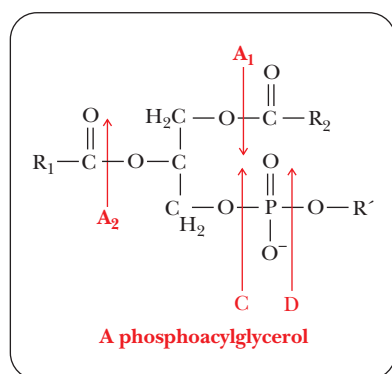
#### 21.8 Cholesterol Biosynthesis

- Why is HMG-CoA so important in cholesterol biosynthesis?
- How does cholesterol serve as a precursor of other steroids?
- What is the role of cholesterol in heart disease?

**OWL** Online homework for this chapter may be assigned in OWL.



■ **FIGURE 21.1** The release of fatty acids for future use. The source of fatty acids can be a triacylglycerol (left) or a phospholipid such as phosphatidylcholine (right).



■ **FIGURE 21.2** Several phospholipases hydrolyze phosphoacylglycerols. They are designated A<sub>1</sub>, A<sub>2</sub>, C, and D. Their sites of action are shown. The site of action of phospholipase A<sub>2</sub> is the B site, and the name phospholipase A<sub>2</sub> is the result of historical accident.

one reason competitive runners often drink caffeine the morning of a race. Distance runners want to burn fat more efficiently to spare their carbohydrate stores for the later stages of the race.

### How are fatty acids transported to the mitochondrion for oxidation?

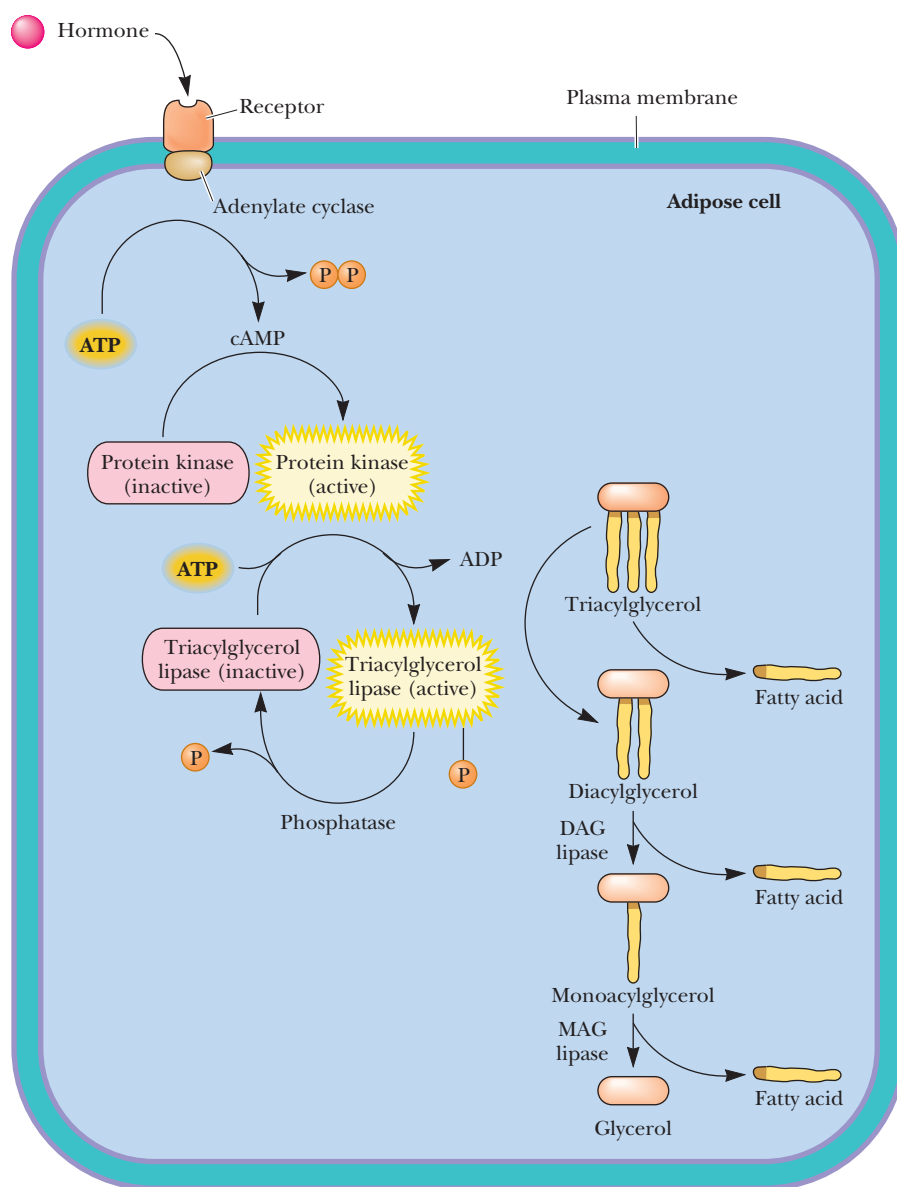
Fatty-acid oxidation begins with **activation** of the molecule. In this reaction, a thioester bond is formed between the carboxyl group of the fatty acid and the thiol group of coenzyme A (CoA-SH). The activated form of the fatty acid is an acyl-CoA, the exact nature of which depends on the nature of the fatty acid itself. Keep in mind throughout this discussion that all acyl-CoA molecules are thioesters, since the fatty acid is esterified to the thiol group of CoA.

The enzyme that catalyzes formation of the ester bond, an *acyl-CoA synthetase*, requires ATP for its action. In the course of the reaction, an acyl adenylate intermediate is formed. The acyl group is then transferred to CoA-SH. ATP is converted to AMP and PP<sub>i</sub>, rather than to ADP and P<sub>i</sub>. The PP<sub>i</sub> is hydrolyzed to two P<sub>i</sub>; the hydrolysis of two high-energy phosphate bonds provides energy for the activation of the fatty acid and is equivalent to the use of two ATP. The formation of the acyl-CoA is endergonic without the energy provided by the hydrolysis of the two high-energy bonds. Note also that the hydrolysis of ATP to AMP and two P<sub>i</sub> represents an increase in entropy (Figure 21.4). There are several enzymes of this type, some specific for longer-chain fatty acids and some for shorter-chain fatty acids. Both saturated and unsaturated fatty acids can serve as substrates for these enzymes. The esterification takes place in the cytosol, but the rest of the reactions of fatty-acid oxidation occur in the mitochondrial matrix. The activated fatty acid must be transported into the mitochondrion so that the rest of the oxidation process can proceed.

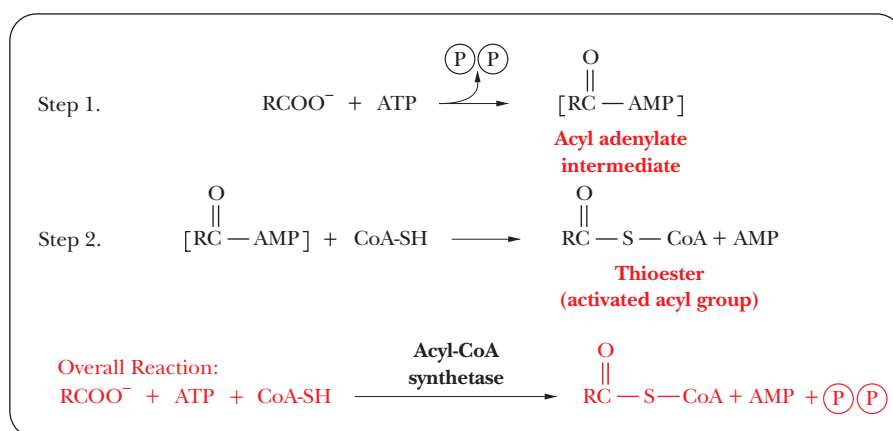
The acyl-CoA can cross the outer mitochondrial membrane but not the inner membrane (Figure 21.5). In the intermembrane space, the acyl group is transferred to **carnitine** by transesterification; this reaction is catalyzed by the enzyme **carnitine acyltransferase**, which is located in the inner membrane. Acyl-carnitine, a compound that can cross the inner mitochondrial membrane, is formed. This enzyme has a specificity for acyl groups between 14 and 18 carbons long and is often called **carnitine palmitoyltransferase (CPT-I)** for this reason. The acyl-carnitine passes through the inner membrane via a specific carnitine/acyl-carnitine transporter called **carnitine translocase**. Once in the matrix, the acyl group is transferred from carnitine to mitochondrial CoA-SH by another transesterification reaction, involving a second carnitine palmitoyltransferase (**CPT-II**) located on the inner face of the membrane.



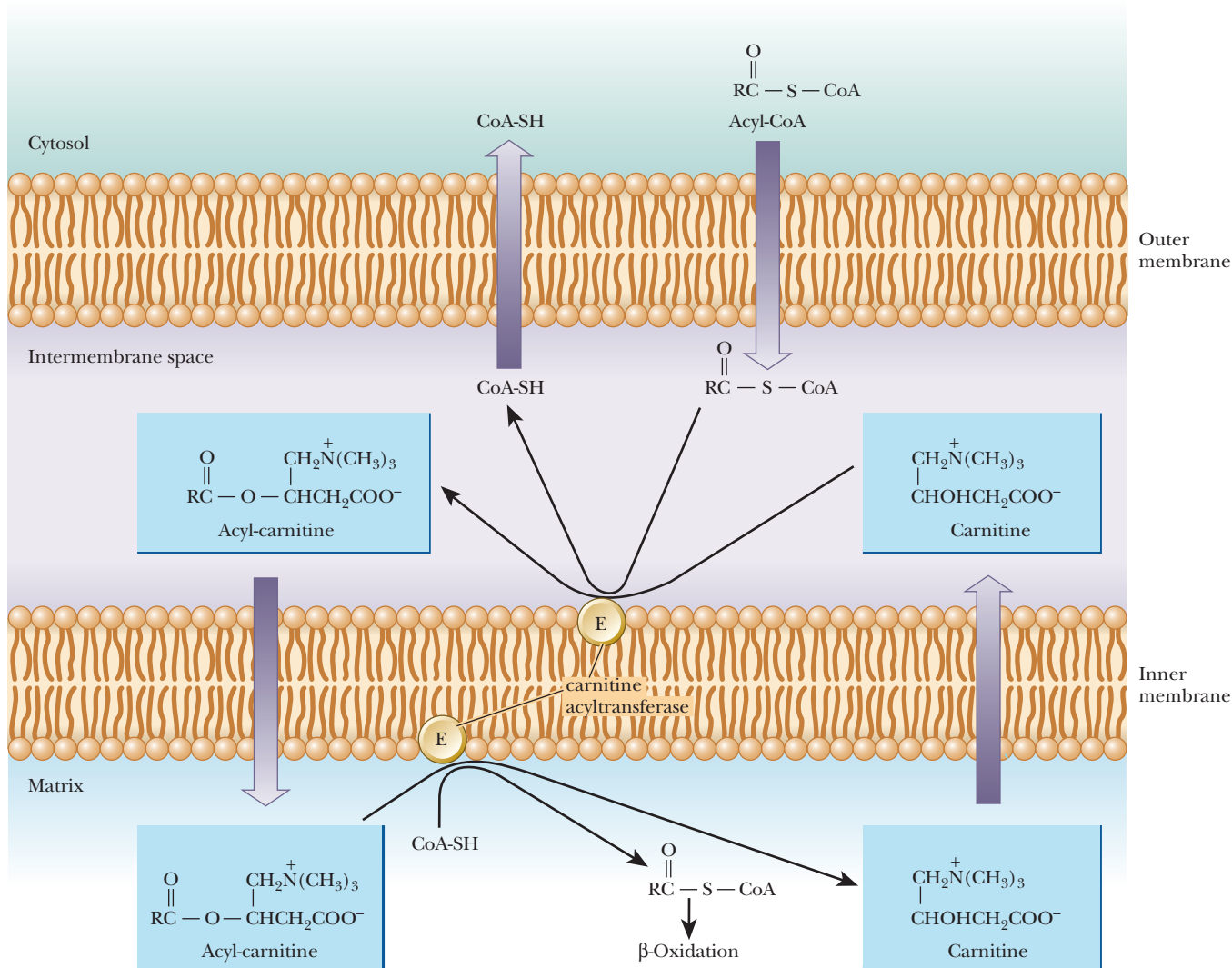
Leonard Lee Rue II/Photo Researchers, Inc.



■ **FIGURE 21.3** Liberation of fatty acids from triacylglycerols in adipose tissue is hormone-dependent.



■ **FIGURE 21.4** The formation of an acyl-CoA.

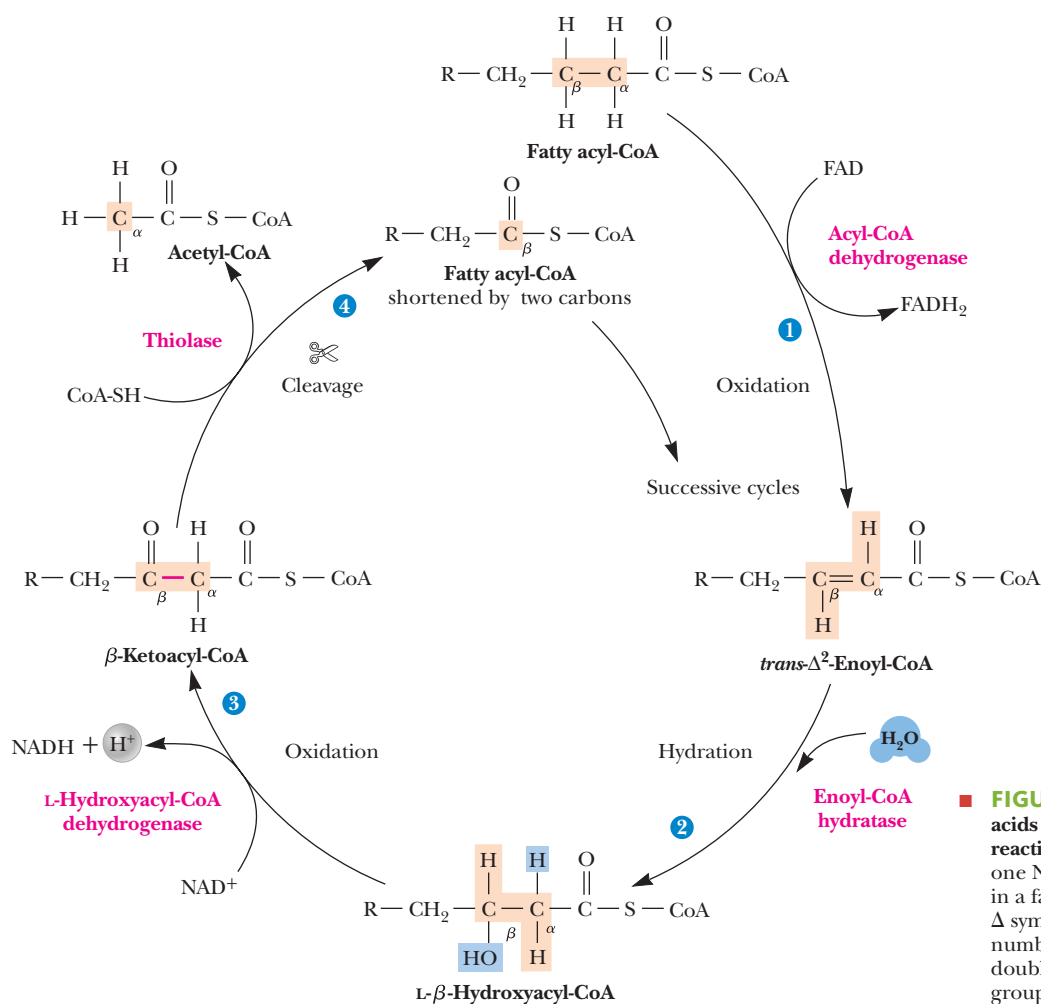


■ **FIGURE 21.5** The role of carnitine in the transfer of acyl groups to the mitochondrial matrix.

### How does oxidation of fatty acids take place?

In the matrix, a repeated sequence of reactions successively cleaves two-carbon units from the fatty acid, starting from the carboxyl end. This process is called **β-oxidation**, since the oxidative cleavage takes place at the β-carbon of the acyl group esterified to CoA. The β-carbon of the original fatty acid becomes the carboxyl carbon in the next stage of degradation. The whole cycle requires four reactions (Figure 21.6).

1. The acyl-CoA is *oxidized* to an α, β unsaturated acyl-CoA (also called a β-enoyl-CoA). The product has the *trans* arrangement at the double bond. This reaction is catalyzed by an FAD-dependent acyl-CoA dehydrogenase.
2. The unsaturated acyl-CoA is *hydrated* to produce a β-hydroxyacyl-CoA. This reaction is catalyzed by the enzyme enoyl-CoA hydratase.
3. A second *oxidation* reaction is catalyzed by β-hydroxyacyl-CoA dehydrogenase, an  $\text{NAD}^+$ -dependent enzyme. The product is a β-ketoacyl-CoA.
4. The enzyme thiolase catalyzes the *cleavage* of the β-ketoacyl-CoA; a molecule of CoA is required for the reaction. The products are acetyl-CoA and an acyl-CoA that is two carbons shorter than the original molecule that entered

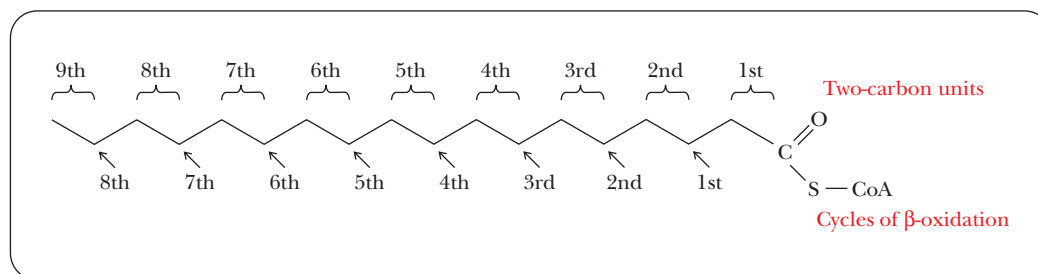


**FIGURE 21.6** The  $\beta$ -oxidation of saturated fatty acids involves a cycle of four enzyme-catalyzed reactions. Each cycle produces one  $\text{FADH}_2$  and one  $\text{NADH}$ , and it liberates acetyl-CoA, resulting in a fatty acid that is two carbons shorter. The  $\Delta$  symbol represents a double bond, and the number associated with it is the location of the double bond (based on counting the carbonyl group as carbon 1).

the  $\beta$ -oxidation cycle. The CoA is needed in this reaction to form the new thioester bond in the smaller acyl-CoA molecule. This smaller molecule then undergoes another round of the  $\beta$ -oxidation cycle.

When a fatty acid with an even number of carbon atoms undergoes successive rounds of the  $\beta$ -oxidation cycle, the product is acetyl-CoA. (Fatty acids with even numbers of carbon atoms are the ones normally found in nature, so acetyl-CoA is the usual product of fatty-acid catabolism.) The number of molecules of acetyl-CoA produced is equal to half the number of carbon atoms in the original fatty acid. For example, stearic acid contains 18 carbon atoms and gives rise to 9 molecules of acetyl-CoA. Note that the conversion of one 18-carbon stearic acid molecule to nine 2-carbon acetyl units requires eight, not nine, cycles of  $\beta$ -oxidation (Figure 21.7). The acetyl-CoA enters the citric acid cycle, with the rest of the oxidation of fatty acids to carbon dioxide and water taking place through the citric acid cycle and electron transport. Recall that most of the enzymes of the citric acid cycle are located in the mitochondrial matrix, and we have just seen that the  $\beta$ -oxidation cycle takes place in the matrix as well. In addition to mitochondria, other sites of  $\beta$ -oxidation are known. Peroxisomes and glyoxysomes (Section 1.6), organelles that carry out oxidation reactions, are also sites of  $\beta$ -oxidation, albeit to a far lesser extent than the mitochondria. Certain drugs, called hypolipidemic drugs, are used in an attempt to control obesity. Some of these work by stimulating  $\beta$ -oxidation in peroxisomes.



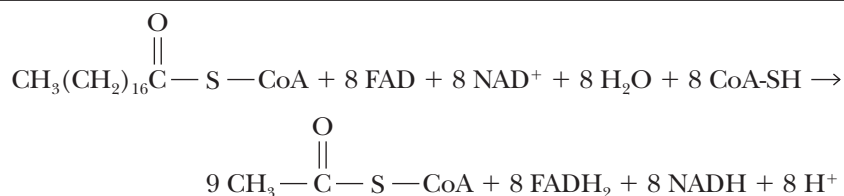


■ **FIGURE 21.7** Stearic acid (18 carbons) gives rise to nine 2-carbon units after eight cycles of  $\beta$ -oxidation. The ninth 2-carbon unit remains esterified to CoA after eight cycles of  $\beta$ -oxidation have removed eight successive 2-carbon units, starting at the carboxyl end on the right. Thus, it takes only eight rounds of  $\beta$ -oxidation to completely process an 18-carbon fatty acid to acetyl-CoA.

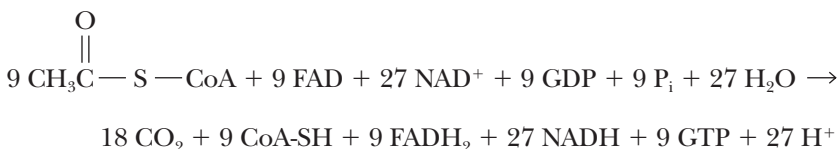
## 21.3 The Energy Yield from the Oxidation of Fatty Acids

In carbohydrate metabolism, the energy released by oxidation reactions is used to drive the production of ATP, with most of the ATP produced in aerobic processes. In the same aerobic processes—namely, the citric acid cycle and oxidative phosphorylation—the energy released by the oxidation of acetyl-CoA formed by  $\beta$ -oxidation of fatty acids can also be used to produce ATP. There are two sources of ATP to keep in mind when calculating the overall yield of ATP. The first source is the reoxidation of the NADH and  $\text{FADH}_2$  produced by the  $\beta$ -oxidation of the fatty acid to acetyl-CoA. The second source is ATP production from the processing of the acetyl-CoA through the citric acid cycle and oxidative phosphorylation. We shall use the oxidation of stearic acid, which contains 18 carbon atoms, as our example.

Eight cycles of  $\beta$ -oxidation are required to convert one mole of stearic acid to nine moles of acetyl-CoA; in the process eight moles of FAD are reduced to  $\text{FADH}_2$ , and eight moles of  $\text{NAD}^+$  are reduced to NADH.

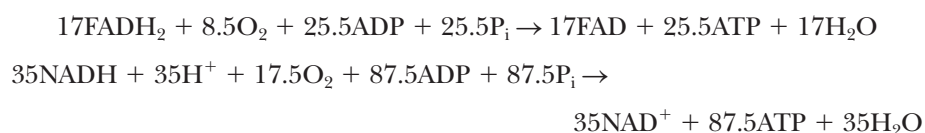


The nine moles of acetyl-CoA produced from each mole of stearic acid enter the citric acid cycle. One mole of  $\text{FADH}_2$  and three moles of NADH are produced for each mole of acetyl-CoA that enters the citric acid cycle. At the same time, one mole of GDP is phosphorylated to produce GTP for each turn of the citric acid cycle.

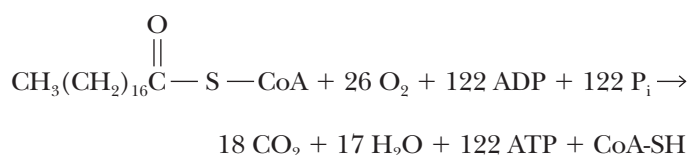


The  $\text{FADH}_2$  and NADH produced by  $\beta$ -oxidation and by the citric acid cycle enter the electron transport chain, and ATP is produced by oxidative phosphorylation. In our example, there are 17 moles of  $\text{FADH}_2$  (8 from  $\beta$ -oxidation

and 9 from the citric acid cycle); there are also 35 moles of NADH (8 from  $\beta$ -oxidation and 27 from the citric acid cycle). Recall that 2.5 moles of ATP are produced for each mole of NADH that enters the electron transport chain, and 1.5 moles of ATP result from each mole of  $\text{FADH}_2$ . Because  $17 \times 1.5 = 25.5$  and  $35 \times 2.5 = 87.5$ , we can write the following equations:



The overall yield of ATP from the oxidation of stearic acid can be obtained by adding the equations for  $\beta$ -oxidation, for the citric acid cycle, and for oxidative phosphorylation. In this calculation, we take GDP as equivalent to ADP and GTP as equivalent to ATP, which means that the equivalent of nine ATP must be added to those produced in the reoxidation of  $\text{FADH}_2$  and NADH. There are 9 ATP equivalent to the 9 GTP from the citric acid cycle, 25.5 ATP from the reoxidation of  $\text{FADH}_2$ , and 87.5 ATP from the reoxidation of NADH, for a grand total of 122 ATP.



The activation step in which stearyl-CoA was formed is not included in this calculation, and we must subtract the ATP that was required for that step. Even though only one ATP was required, two high-energy phosphate bonds are lost because of the production of AMP and  $\text{PP}_i$ . The pyrophosphate must be hydrolyzed to phosphate ( $\text{P}_i$ ) before it can be recycled in metabolic intermediates. As a result, we must subtract the equivalent of two ATP for the activation step. The net yield of ATP becomes 120 moles of ATP for each mole of stearic acid that is completely oxidized. See Table 21.1 for a balance sheet. Keep in mind that these values are theoretical consensus values that not all cells attain.

As a comparison, note that 32 moles of ATP can be obtained from the complete oxidation of one mole of glucose; but glucose contains 6, rather than 18, carbon atoms. Three glucose molecules contain 18 carbon atoms, and a more interesting comparison is the ATP yield from the oxidation of three glucose

**TABLE 21.1**

The Balance Sheet for Oxidation of One Molecule of Stearic Acid			
	NADH Molecules	$\text{FADH}_2$ Molecules	ATP Molecules
Reaction			
1. Stearic acid $\rightarrow$ Stearyl-CoA (activation step)			-2
2. Stearyl-CoA $\rightarrow$ 9 acetyl-CoA (8 cycles of $\beta$ -oxidation)	+8	+8	
3. 9 Acetyl-CoA $\rightarrow$ 18 $\text{CO}_2$ (citric acid cycle); GDP $\rightarrow$ GTP (9 molecules)	+27	+9	+9
4. Reoxidation of NADH from $\beta$ -oxidation cycle	-8		+20
5. Reoxidation of NADH from citric acid cycle	-27		+67.5
6. Reoxidation of $\text{FADH}_2$ from $\beta$ -oxidation cycle		-8	+12
7. Reoxidation of $\text{FADH}_2$ from citric acid cycle	—	-9	+13.5
	0	0	+120

Note that there is no net change in the number of molecules of NADH or  $\text{FADH}_2$ .



George Holony/Photo Researchers, Inc.



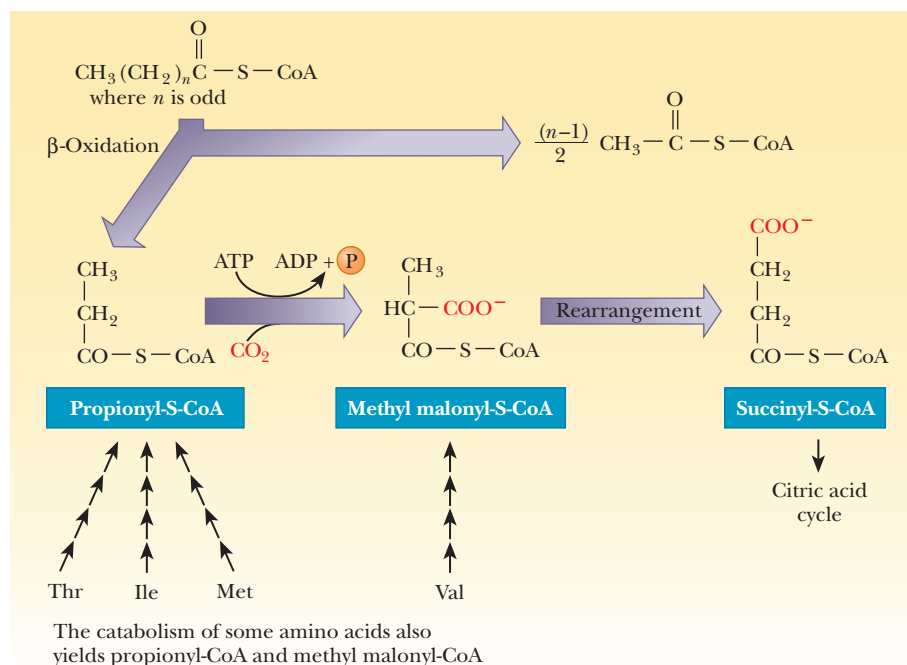
Tom McHugh/Photo Researchers, Inc.

molecules, which is  $3 \times 32 = 96$  ATP for the same number of carbon atoms. The yield of ATP from the oxidation of the lipid is still higher than that from the carbohydrate, even for the same number of carbon atoms. The reason is that a fatty acid is all hydrocarbon except for the carboxyl group; that is, it exists in a highly reduced state. A sugar is already partly oxidized because of the presence of its oxygen-containing groups. Because the oxidation of a fuel leads to the reduced electron carriers used in the electron transport chain, a more reduced fuel, such as a fatty acid, can be oxidized further than a partially oxidized fuel, such as a carbohydrate.

Another point of interest is that water is produced in the oxidation of fatty acids. We have already seen that water is also produced in the complete oxidation of carbohydrates. The production of **metabolic water** is a common feature of aerobic metabolism. This process can be a source of water for organisms that live in desert environments. Camels are a well-known example; the stored lipids in their humps are a source of both energy and water during long trips through the desert. Kangaroo rats provide an even more striking example of adaptation to an arid environment. These animals have been observed to live indefinitely without having to drink water. They live on a diet of seeds, which are rich in lipids but contain little water. The metabolic water that kangaroo rats produce is adequate for all their water needs. This metabolic response to arid conditions is usually accompanied by a reduced output of urine.

## 21.4 Catabolism of Unsaturated Fatty Acids and Odd-Carbon Fatty Acids

Fatty acids with odd numbers of carbon atoms are not as frequently encountered in nature as are the ones with even numbers of carbon atoms. Odd-numbered fatty acids also undergo  $\beta$ -oxidation (Figure 21.8). The last cycle of  $\beta$ -oxidation produces one molecule of propionyl-CoA. An enzymatic



■ **FIGURE 21.8** The oxidation of a fatty acid containing an odd number of carbon atoms.

pathway exists to convert propionyl-CoA to succinyl-CoA, which then enters the citric acid cycle. In this pathway, propionyl-CoA is first carboxylated to methyl malonyl-CoA in a reaction catalyzed by propionyl-CoA carboxylase, which then undergoes rearrangement to form succinyl-CoA. Because propionyl-CoA is also a product of the catabolism of several amino acids, the conversion of propionyl-CoA to succinyl-CoA also plays a role in amino acid metabolism (Section 23.4). The conversion of methyl malonyl-CoA to succinyl-CoA requires vitamin B<sub>12</sub> (cyanocobalamin), which has a cobalt(III) ion in its active state.

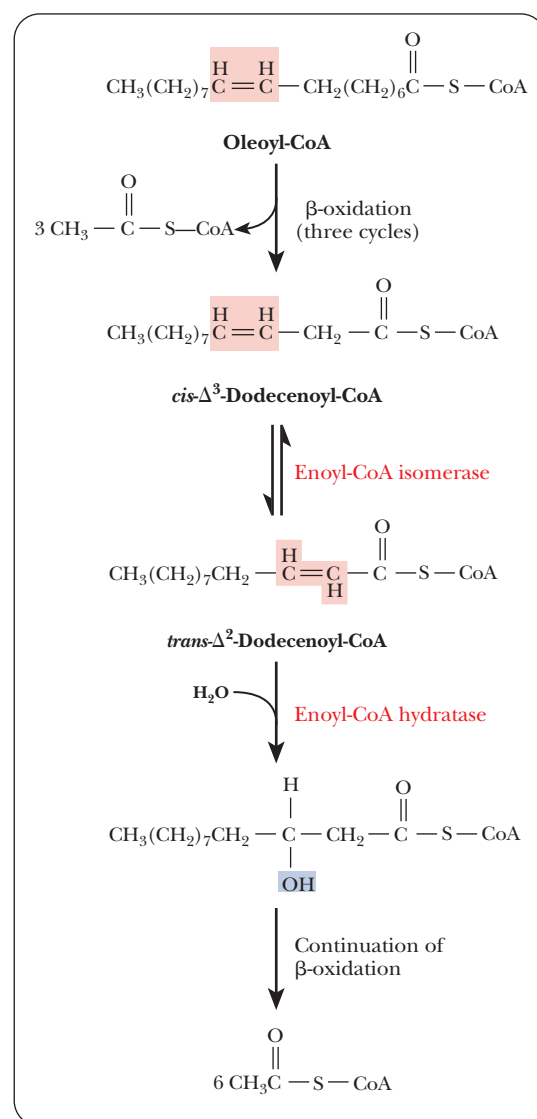
### How does the oxidation of unsaturated fatty acids differ from that of saturated fatty acids?

The conversion of a monounsaturated fatty acid to acetyl-CoA requires a reaction that is not encountered in the oxidation of saturated acids, a *cis*–*trans* isomerization (Figure 21.9). Successive rounds of  $\beta$ -oxidation of oleic acid (18:1) provide an example of these reactions. The process of  $\beta$ -oxidation gives rise to unsaturated fatty acids in which the double bond is in the *trans* arrangement, whereas the double bonds in most naturally occurring fatty acids are in the *cis* arrangement. In the case of oleic acid, there is a *cis* double bond between carbons 9 and 10. Three rounds of  $\beta$ -oxidation produce a 12-carbon unsaturated fatty acid with a *cis* double bond between carbons 3 and 4. The hydratase of the  $\beta$ -oxidation cycle requires a *trans* double bond between carbon atoms 2 and 3 as a substrate. A *cis*–*trans* isomerase produces a *trans* double bond between carbons 2 and 3 from the *cis* double bond between carbons 3 and 4. From this point forward, the fatty acid is metabolized the same as for saturated fatty acids. When oleic acid is  $\beta$ -oxidized, the first step (fatty acyl-CoA dehydrogenase) is skipped, and the isomerase deals with the *cis* double bond, putting it into the proper position and orientation to continue the pathway.

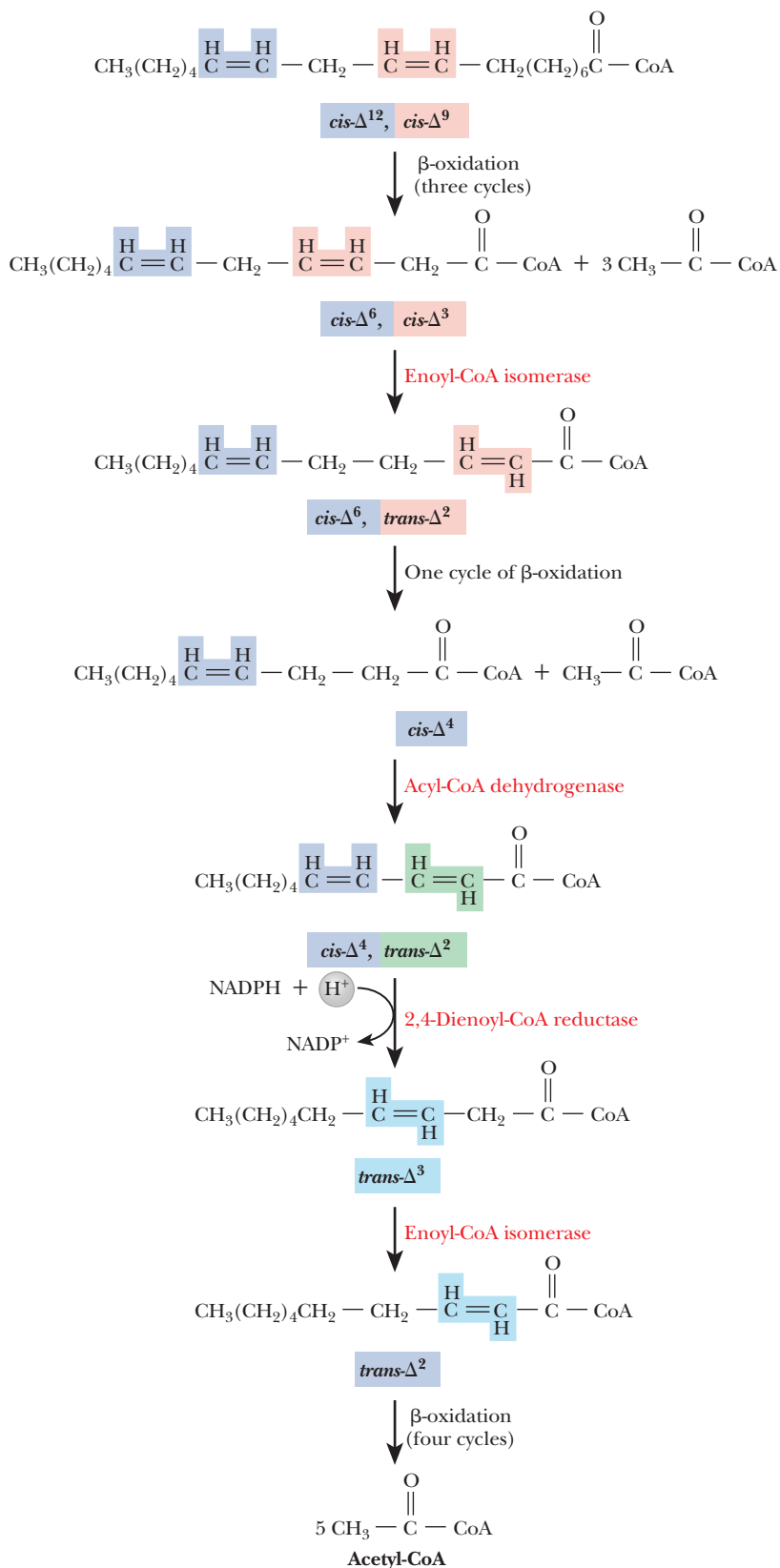
When polyunsaturated fatty acids are  $\beta$ -oxidized, another enzyme is needed to handle the second double bond. Let's consider how linoleic acid (18:2) would be metabolized (Figure 21.10). This fatty acid has *cis* double bonds at positions 9 and 12 as shown in Figure 21.10, which are indicated as *cis*- $\Delta^9$  and *cis*- $\Delta^{12}$ . Three normal cycles of  $\beta$ -oxidation occur, as in our example with oleic acid, before the isomerase must switch the position and orientation of the double bond. The cycle of  $\beta$ -oxidation continues until a 10-carbon fatty acyl-CoA is attained that has one *cis* double bond on its carbon 4 (*cis*- $\Delta^4$ ). Then the first step of  $\beta$ -oxidation occurs, putting in a *trans* double bond between carbons 2 and 3 ( $\alpha$  and  $\beta$ ). Normal  $\beta$ -oxidation cannot continue at this point because the fatty acid with the two double bonds so close together is a poor substrate for the hydratase. Therefore, a second new enzyme, 2,4-dienoyl-CoA reductase, uses NADPH to reduce this intermediate. The result is a fatty acyl-CoA with a *trans* double bond between carbons 3 and 4. The isomerase then switches the *trans* double from carbon 3 to carbon 2, and  $\beta$ -oxidation continues.

A molecule with three double bonds, such as linolenic acid (18:3), would use the same two enzymes to handle the double bonds. The first double bond requires the isomerase. The second one requires the reductase and the isomerase, and the third requires the isomerase. For practice, you can diagram the  $\beta$ -oxidation of an 18-carbon molecule with *cis* double bonds at positions 9, 12, and 15 to see that this is true. Unsaturated fatty acids make up a large enough portion of the fatty acids in storage fat (40% for oleic acid alone) to make the reactions of the *cis*–*trans* isomerase and the epimerase of particular importance.

The oxidation of unsaturated fatty acids does not generate as many ATPs as it would for a saturated fatty acid with the same number of carbons. This is because the presence of a double bond means that the acyl-CoA dehydrogenase step will be skipped. Thus, fewer FADH<sub>2</sub> will be produced.



■ **FIGURE 21.9**  $\beta$ -oxidation of unsaturated fatty acids. In the case of oleoyl-CoA, three  $\beta$ -oxidation cycles produce three molecules of acetyl-CoA and leave *cis*- $\Delta^3$ -dodecenoyl-CoA. Rearrangement of enoyl-CoA isomerase gives the *trans*- $\Delta^2$  species, which then proceeds normally through the  $\beta$ -oxidation pathway.



■ **FIGURE 21.10** The oxidation pathway for polyunsaturated fatty acids, illustrated for linoleic acid. Three cycles of  $\beta$ -oxidation on linoleoyl-CoA yield the *cis*- $\Delta^3$ , *cis*- $\Delta^6$  intermediate, which is converted to a *trans*- $\Delta^2$ , *cis*- $\Delta^6$  intermediate. An additional round of  $\beta$ -oxidation gives *cis*- $\Delta^4$  enoyl-CoA, which is oxidized to the *trans*- $\Delta^2$ , *cis*- $\Delta^4$  species by acyl-CoA dehydrogenase. The subsequent action of 2,4-dienoyl-CoA reductase yields the *trans*- $\Delta^3$  product, which is converted by enoyl-CoA isomerase to the *trans*- $\Delta^2$  form. Normal  $\beta$ -oxidation then produces five molecules of acetyl-CoA.



## 21.5 Ketone Bodies

Substances related to acetone (“**ketone bodies**”) are produced when an excess of acetyl-CoA arises from  $\beta$ -oxidation. This condition occurs when not enough oxaloacetate is available to react with the large amounts of acetyl-CoA that could enter the citric acid cycle. Oxaloacetate in turn arises from glycolysis because it is formed from pyruvate in a reaction catalyzed by pyruvate carboxylase.

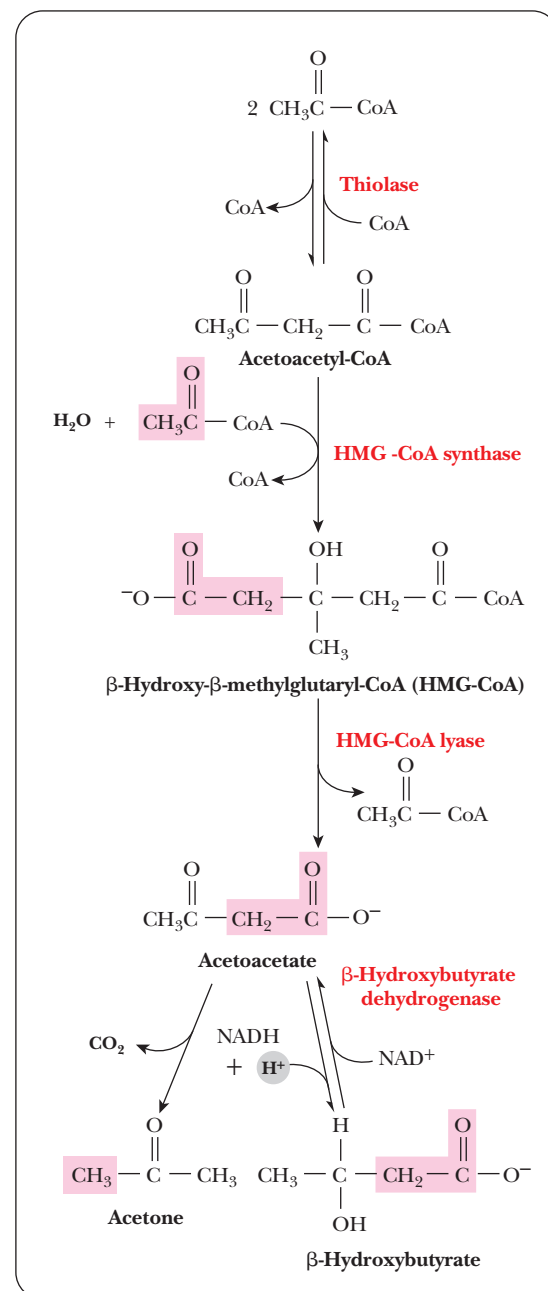
A situation like this can come about when an organism has a high intake of lipids and a low intake of carbohydrates, but there are also other possible causes, such as starvation and diabetes. Starvation conditions cause an organism to break down fats for energy, leading to the production of large amounts of acetyl-CoA by  $\beta$ -oxidation. The amount of acetyl-CoA is excessive by comparison with the amount of oxaloacetate available to react with it. In the case of people with diabetes, the cause of the imbalance is not inadequate intake of carbohydrates but rather the inability to metabolize them.

### Do acetone and acetyl-CoA have a connection in lipid metabolism?

The reactions that result in ketone bodies start with the condensation of two molecules of acetyl-CoA to produce acetoacetyl-CoA. *Acetoacetate* is produced from acetoacetyl-CoA through condensation with another acetyl-CoA to form  $\beta$ -hydroxy- $\beta$ -methylglutaryl-CoA (HMG-CoA), a compound we will see again when we look at cholesterol synthesis (Figure 21.11). HMG-CoA lyase then releases acetyl-CoA to give acetoacetate. Acetoacetate can then have two fates. A reduction reaction can produce  $\beta$ -hydroxybutyrate from acetoacetate. The other possible reaction is the spontaneous decarboxylation of acetoacetate to give *acetone*. The odor of acetone can frequently be detected on the breath of people with diabetes whose disease is not controlled by suitable treatment. The excess of acetoacetate, and consequently of acetone, is a pathological condition known as *ketosis*. Because acetoacetate and  $\beta$ -hydroxybutyrate are acidic, their presence at high concentration overwhelms the buffering capacity of the blood. The body deals with the consequent lowering of blood pH (ketoacidosis) by excreting  $\text{H}^+$  into the urine, accompanied by excretion of  $\text{Na}^+$ ,  $\text{K}^+$ , and water. Severe dehydration can result (excessive thirst is a classic symptom of diabetes); diabetic coma is another possible danger.

The principal site of synthesis of ketone bodies is liver mitochondria, but they are not used there because the liver lacks the enzymes necessary to recover acetyl-CoA from ketone bodies. It is easy to transport ketone bodies in the bloodstream because, unlike fatty acids, they are water-soluble and do not need to be bound to proteins, such as serum albumin. Organs other than the liver can use ketone bodies, particularly acetoacetate. Even though glucose is the usual fuel in most tissues and organs, acetoacetate can be used as a fuel. In heart muscle and the renal cortex, acetoacetate is the preferred source of energy.

Even in organs such as the brain, in which glucose is the preferred fuel, starvation conditions can lead to the use of acetoacetate for energy. In this situation, acetoacetate is converted to two molecules of acetyl-CoA, which can then enter the citric acid cycle. The key point here is that starvation gives rise to long-term, rather than short-term, regulation over a period of hours to days rather than minutes. The decreased level of glucose in the blood over a period of days changes the hormone balance in the body, particularly involving insulin and glucagon (see Section 24.4). (Short-term regulation, such as allosteric interactions or covalent modification, can occur in a matter of minutes.) The rates of protein synthesis and breakdown are subject to change under these conditions. The specific enzymes involved are those involved in fatty-acid oxidation (increase in levels) and those for lipid biosynthesis (decrease in levels).



■ **FIGURE 21.11** The formation of ketone bodies, synthesized primarily in the liver.

## 21.6 Fatty-Acid Biosynthesis

The anabolism of fatty acids is not simply a reversal of the reactions of  $\beta$ -oxidation. Anabolism and catabolism are not, in general, the exact reverse of each other; for instance, gluconeogenesis (Section 18.2) is not simply a reversal of the reactions of glycolysis. A first example of the differences between the degradation and the biosynthesis of fatty acids is that the anabolic reactions take place in the cytosol. We have just seen that the degradative reactions of  $\beta$ -oxidation take place in the mitochondrial matrix. The first step in fatty-acid biosynthesis is transport of acetyl-CoA to the cytosol.

Acetyl-CoA can be formed either by  $\beta$ -oxidation of fatty acids or by decarboxylation of pyruvate. (Degradation of certain amino acids also produces acetyl-CoA; see Section 23.6.) Most of these reactions take place in the mitochondria, requiring a transport mechanism to export acetyl-CoA to the cytosol for fatty-acid biosynthesis. The transport mechanism is based on the fact that citrate can cross the mitochondrial membrane. Acetyl-CoA condenses with oxaloacetate, which cannot cross the mitochondrial membrane, to form citrate (recall that this is the first reaction of the citric acid cycle).

### How do the first steps of fatty-acid synthesis take place?

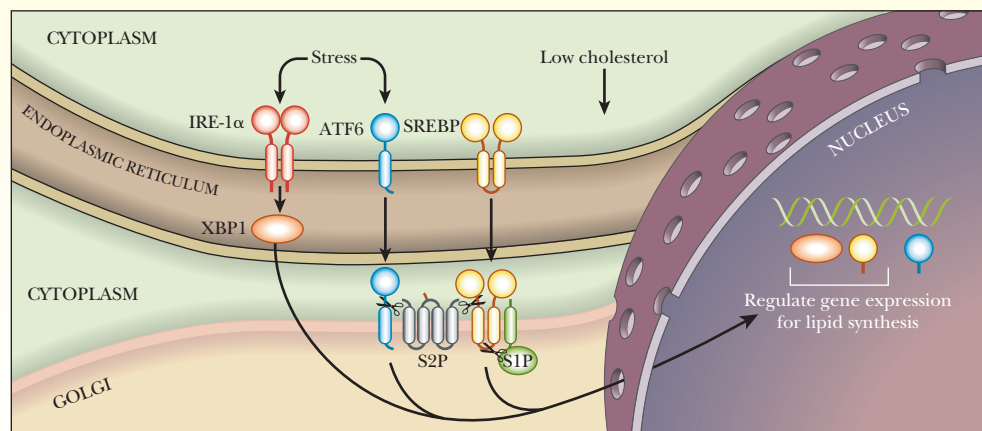
The citrate that is exported to the cytosol can undergo the reverse reaction, producing oxaloacetate and acetyl-CoA (Figure 21.12). Acetyl-CoA enters the pathway for fatty-acid biosynthesis, while oxaloacetate undergoes a series of reactions in which NADPH is substituted for NADH (see the discussion of lipid

## Biochemical Connections GENE EXPRESSION

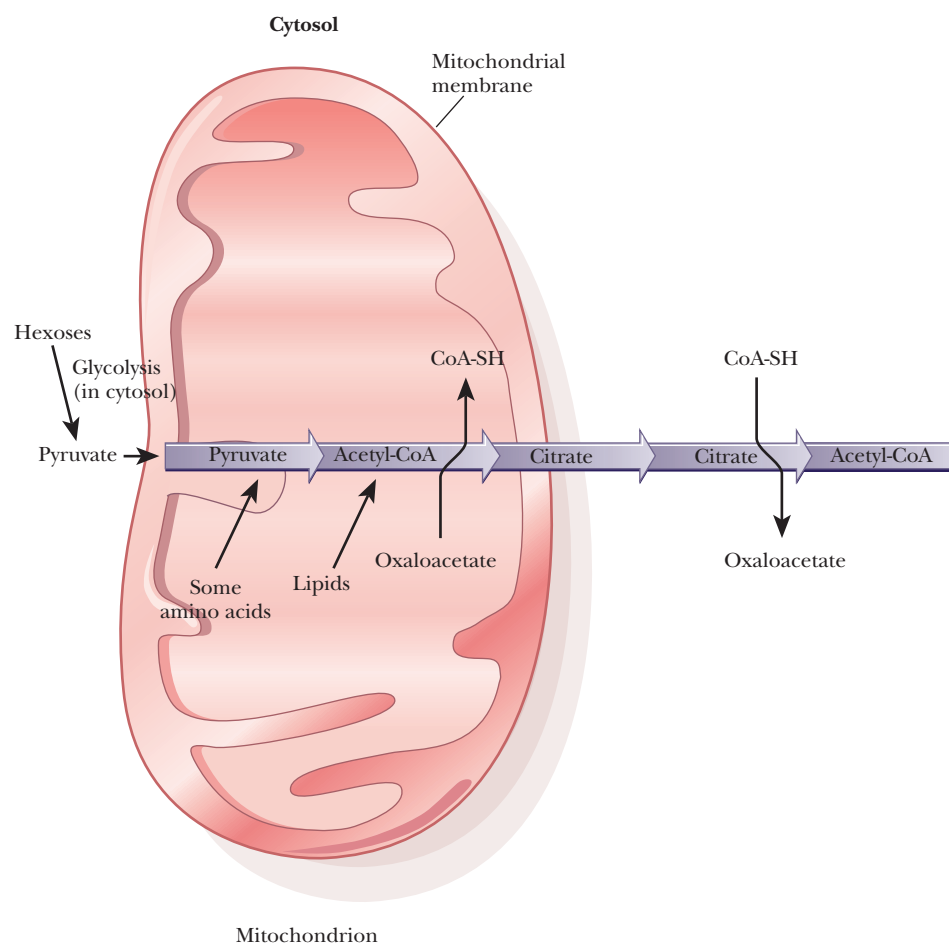
### Transcription Activators in Lipid Biosynthesis

As we saw in Chapter 11, transcription factors can do double duty in cells. An example that directly deals with the material in this chapter is the action of the transcription factor XBP1 in mouse liver. XBP1 regulates genes that deal with improperly folded proteins and genes that control lipid synthesis. It was already well established that other transcription factors, called SREBPs, played a role in regulating gene expression leading to lipid synthesis. The signal for regulation by SREBPs is low cholesterol in the cytoplasm, with the signal being passed into the endoplasmic reticulum and

then into the Golgi apparatus. Processing of the proteins involved takes place in the Golgi before the proteins enter the nucleus. The role of XBP1 in regulating lipid biosynthesis follows a different pathway. As shown in the figure, stress factors trigger a response by binding to IRE-1 $\alpha$  in the endoplasmic reticulum, leading to the release of XBP1 into the Golgi. Another transcription factor, ATF6, enters the Golgi in similar fashion. After processing, all these transcription factors play roles in regulating lipid synthesis. What is new in this picture is the hitherto unknown involvement of XBP1.



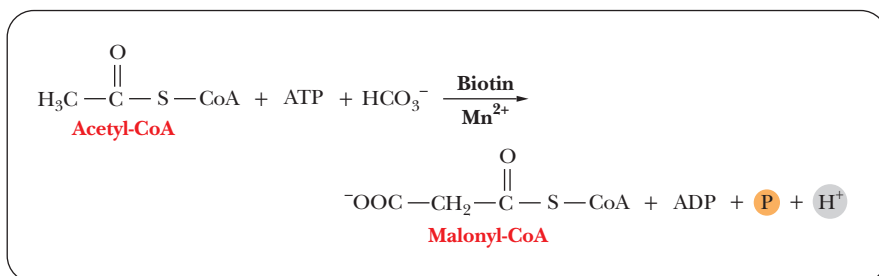
■ **Transcription factors in lipid synthesis.** XBP1, ATF6, and SREBP all undergo processing in the Golgi before they enter the nucleus. ATF6 activates XBP1. In turn, XBP1 and SREBPs regulate the expression of the genes for lipid synthesis. The regulation of SREBPs by cholesterol is a different process from the activation of ATF6 and XBP1 in response to stress. (From Science, vol. 320, p. 1434. Copyright © 2008 by AAAS. Reprinted with permission of the AAAS.)



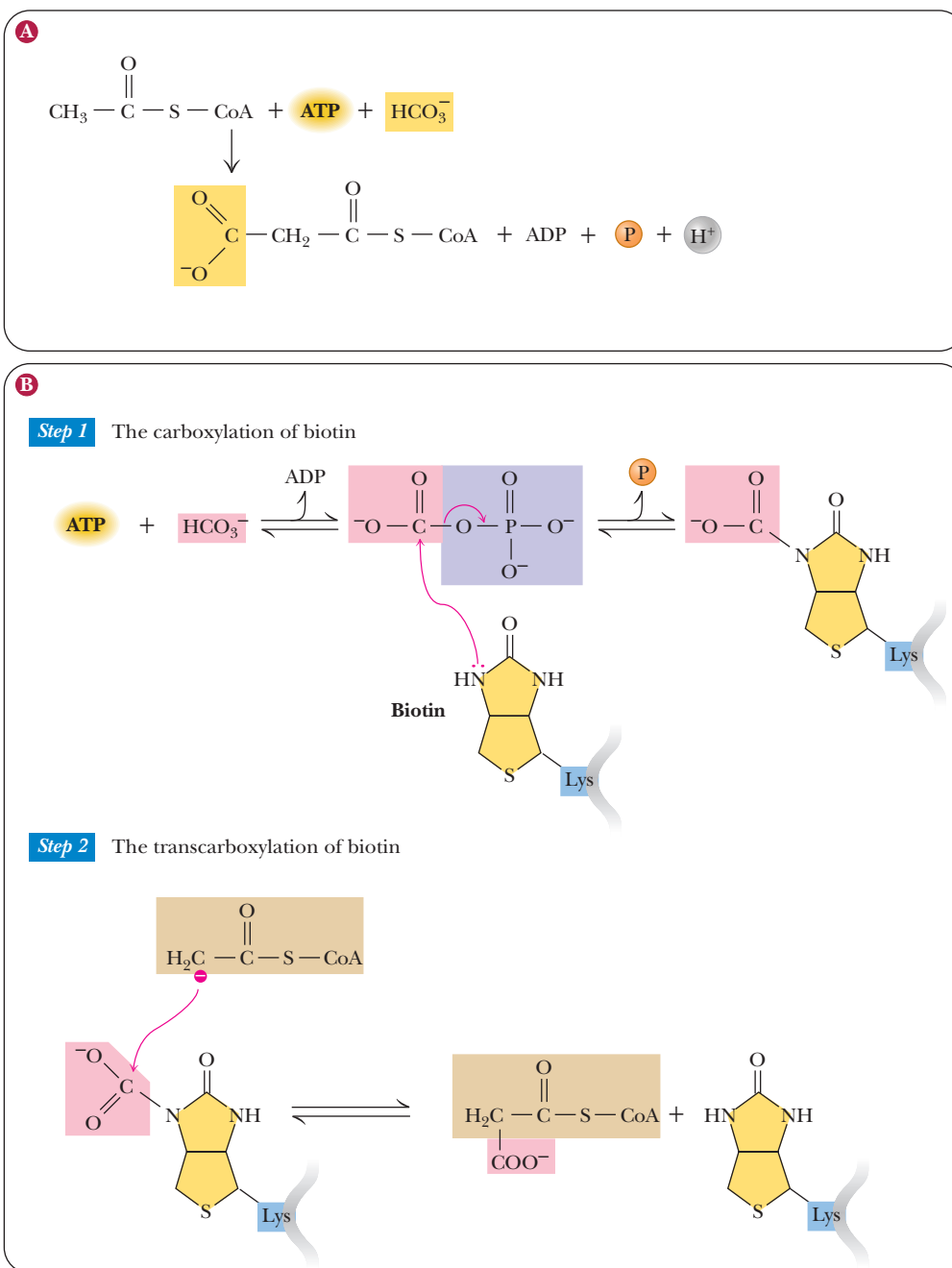
■ **FIGURE 21.12** The transport of acetyl groups from the mitochondrion to the cytosol.

anabolism in Section 19.8). This substitution controls the pathway because NADPH is required for fatty-acid anabolism.

In the cytosol, acetyl-CoA is carboxylated, producing **malonyl-CoA**, a key intermediate in fatty-acid biosynthesis (Figure 21.13). This reaction is catalyzed by the *acetyl-CoA carboxylase* complex, which consists of three enzymes and requires  $\text{Mn}^{2+}$ , biotin, and ATP for activity. We have already seen that enzymes catalyzing reactions that take place in several steps frequently consist of several separate protein molecules, and this enzyme follows that pattern. In this case, acetyl-CoA carboxylase consists of the three proteins *biotin carboxylase*, the *biotin carrier protein*, and *carboxyl transferase*. Biotin carboxylase catalyzes the transfer of the carboxyl group to biotin. The “activated  $\text{CO}_2$ ” (the carboxyl group derived from the bicarbonate ion  $\text{HCO}_3^-$ ) is covalently bound to biotin. Biotin (whether carboxylated or not) is bound to the biotin carrier protein by an amide linkage to the  $\epsilon$ -amino group of a lysine side chain. The amide linkage to the side chain that bonds biotin to the carrier protein is long enough and



■ **FIGURE 21.13** The formation of malonyl-CoA, catalyzed by acetyl-CoA carboxylase.



■ **FIGURE 21.14 The acetyl-CoA carboxylase reaction.** (a) The acetyl-CoA carboxylase reaction produces malonyl-CoA for fatty-acid synthesis. (b) A mechanism for the acetyl-CoA carboxylase reaction. Bicarbonate is activated for carboxylation reactions by formation of *N*-carboxybiotin. ATP drives the reaction forward, with transient formation of a carbonyl-phosphate intermediate (Step 1). In a typical biotin-dependent reaction, nucleophilic attack by the acetyl-CoA carbanion on the carboxyl carbon of *N*-carboxybiotin—a transcarboxylation—yields the carboxylated product (Step 2).

flexible enough to move the carboxylated biotin into position to transfer the carboxyl group to acetyl-CoA in the reaction catalyzed by carboxyl transferase, producing malonyl-CoA (Figure 21.14). In addition to its role as a starting point in fatty-acid synthesis, malonyl-CoA strongly inhibits the carnitine acyl-transferase I on the outer face of the inner mitochondrial membrane. This avoids a futile cycle in which fatty acids are  $\beta$ -oxidized in the mitochondria to make acetyl-CoA just so they can be remade into fatty acids in the cytosol.

### What is the mode of action of fatty-acid synthase?

The biosynthesis of fatty acids involves the successive addition of two-carbon units to the growing chain. Two of the three carbon atoms of the malonyl group of malonyl-CoA are added to the growing fatty-acid chain with each cycle of the biosynthetic reaction. This reaction, like the formation of the malonyl-CoA itself, requires a multienzyme complex located in the cytosol and not attached to any membrane. The complex, made up of the individual enzymes, is called **fatty-acid synthase**.

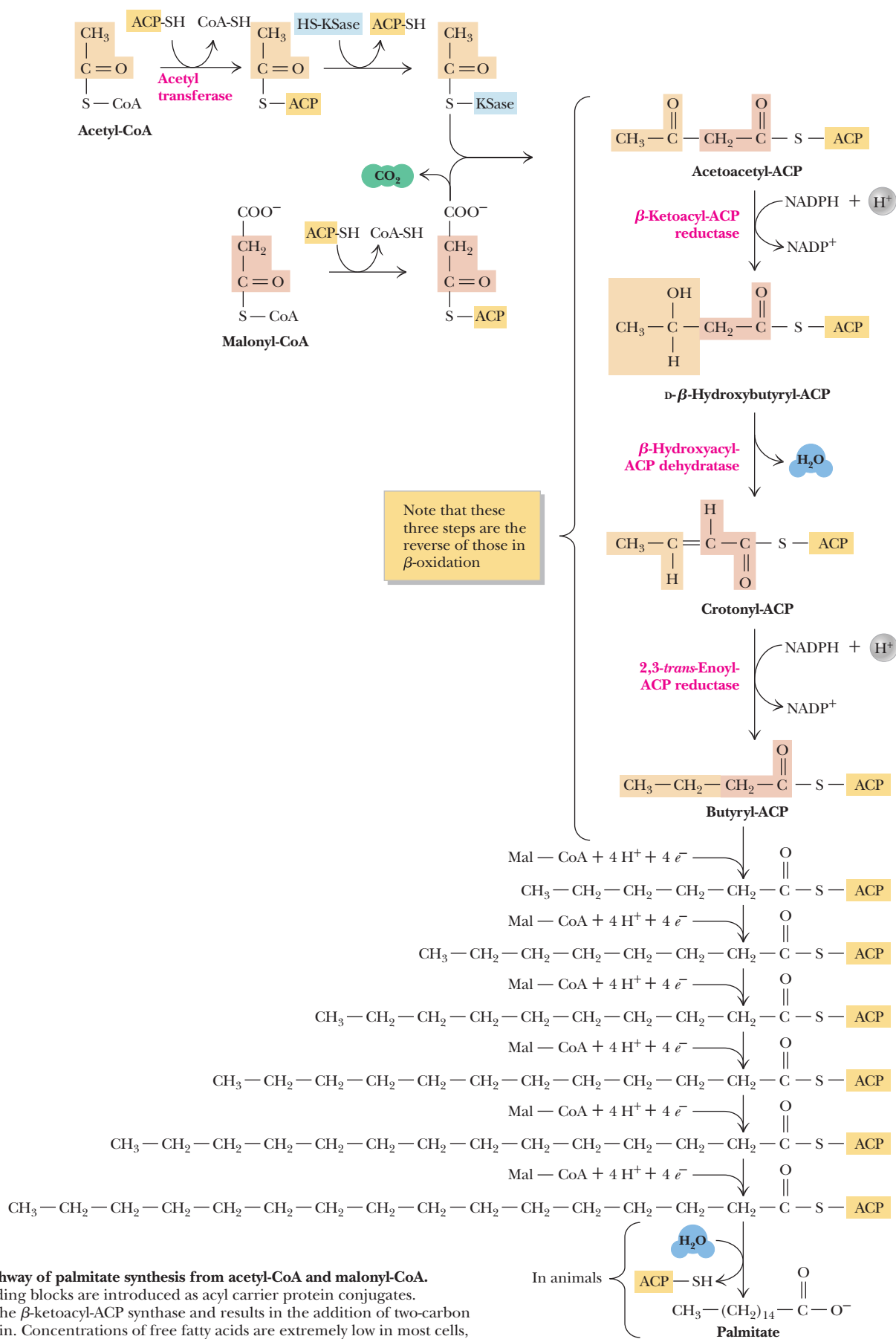
The usual product of fatty-acid anabolism is *palmitate*, the 16-carbon saturated fatty acid. All 16 carbons come from the acetyl group of acetyl-CoA; we have already seen how malonyl-CoA, the immediate precursor, arises from acetyl-CoA. But first there is a priming step in which one molecule of acetyl-CoA is required for each molecule of palmitate produced. In this priming step, the acetyl group from acetyl-CoA is transferred to an **acyl carrier protein (ACP)**, which is considered a part of the fatty-acid synthase complex (Figure 21.15). The acetyl group is bound to the protein as a thioester. The group on the protein to which the acetyl group is bonded is the 4-phosphopantetheine group, which in turn is bonded to a serine side chain; note in Figure 21.16 that this group is structurally similar to CoA-SH itself. The acetyl group is transferred from CoA-SH, to which it is bound by a thioester linkage, to the ACP; the acetyl group is bound to the ACP by a thioester linkage. Although the functional group of ACP is similar to that of CoA-SH, it is noteworthy that fatty-acid synthesis in the cytosol uses only ACP. In essence, the ACP is a label that marks acetyl groups for fatty-acid synthesis.

The acetyl group is transferred in turn from the ACP to another protein, to which it is bound by a thioester linkage to a cysteine-SH; the other protein is  $\beta$ -ketoacyl-S-ACP-synthase (HS-KSase) (Figure 21.15). The first of the successive additions of two of the three malonyl carbons to the fatty acid starts at this point. The malonyl group itself is transferred from a thioester linkage with CoA-SH to another thioester bond with ACP (Figure 21.15). The next step is a condensation reaction that produces acetoacetyl-ACP (Figure 21.15). In other words, the principal product of this reaction is an acetoacetyl group bound to the ACP by a thioester linkage. Two of the four carbons of acetoacetate come from the priming acetyl group, and the other two come from the malonyl group. The carbon atoms that arise from the malonyl group are the one directly bonded to the sulfur and the one in the  $-\text{CH}_2-$  group next to it. The  $\text{CH}_3\text{CO}-$  group comes from the priming acetyl group. The other carbon of the malonyl group is released as  $\text{CO}_2$ ; the  $\text{CO}_2$  that is lost is the original  $\text{CO}_2$  that was used to carboxylate the acetyl-CoA to produce malonyl-CoA. The synthase is no longer involved in a thioester linkage. This is an example of a decarboxylation being used to drive an otherwise unfavorable condensation reaction.

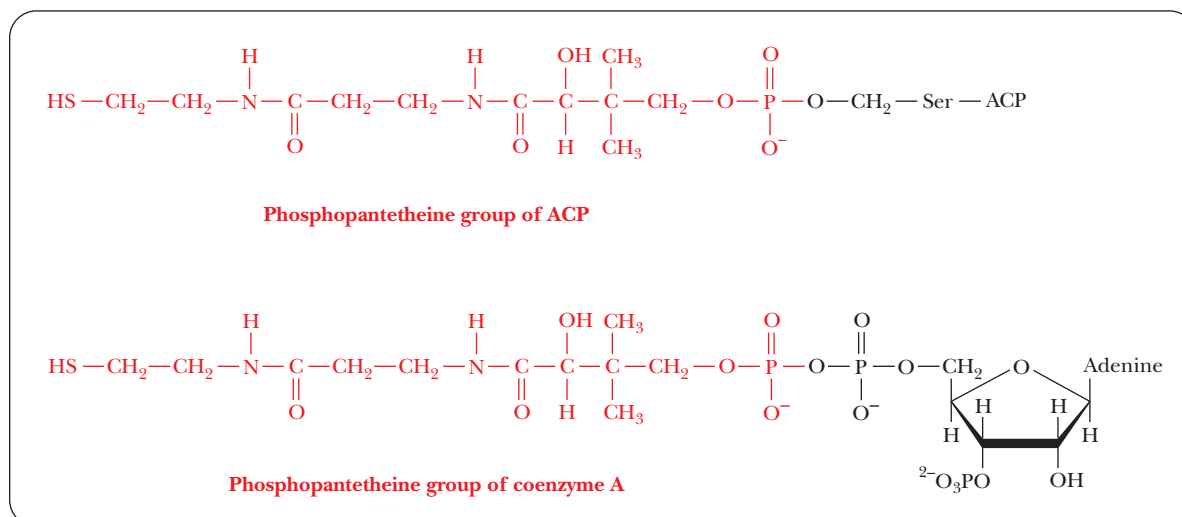
Acetoacetyl-ACP is converted to butyryl-ACP by a series of reactions involving two reductions and a dehydration (Figure 21.15). In the first reduction, the  $\beta$ -keto group is reduced to an alcohol, giving rise to D- $\beta$ -hydroxybutyryl-ACP. In the process, NADPH is oxidized to  $\text{NADP}^+$ ; the enzyme that catalyzes this reaction is  $\beta$ -ketoacyl-ACP reductase (Figure 21.15). The dehydration step, catalyzed by  $\beta$ -hydroxyacyl-ACP dehydratase, produces crotonyl-ACP (Figure 21.15). Note that the double bond is in the *trans* configuration. A second reduction reaction, catalyzed by  $\beta$ -enoyl-ACP reductase, produces butyryl-ACP (Figure 21.15). In this reaction, NADPH is the coenzyme, as it was in the first reduction reaction in this series.

In the second round of fatty-acid biosynthesis, butyryl-ACP plays the same role as acetyl-ACP in the first round. The butyryl group is transferred to the synthase, and a malonyl group is transferred to the ACP. Once again there is a condensation reaction with malonyl-ACP (Figure 21.15). In this second round, the condensation produces a six-carbon  $\beta$ -ketoacyl-ACP. The two added carbon atoms come from the malonyl group, as they did in the first round. The reduction and dehydration reactions take place as before, giving rise to





■ **FIGURE 21.15** The pathway of palmitate synthesis from acetyl-CoA and malonyl-CoA. Acetyl and malonyl building blocks are introduced as acyl carrier protein conjugates. Decarboxylation drives the  $\beta$ -ketoacyl-ACP synthase and results in the addition of two-carbon units to the growing chain. Concentrations of free fatty acids are extremely low in most cells, and newly synthesized fatty acids exist primarily as acyl-CoA esters.

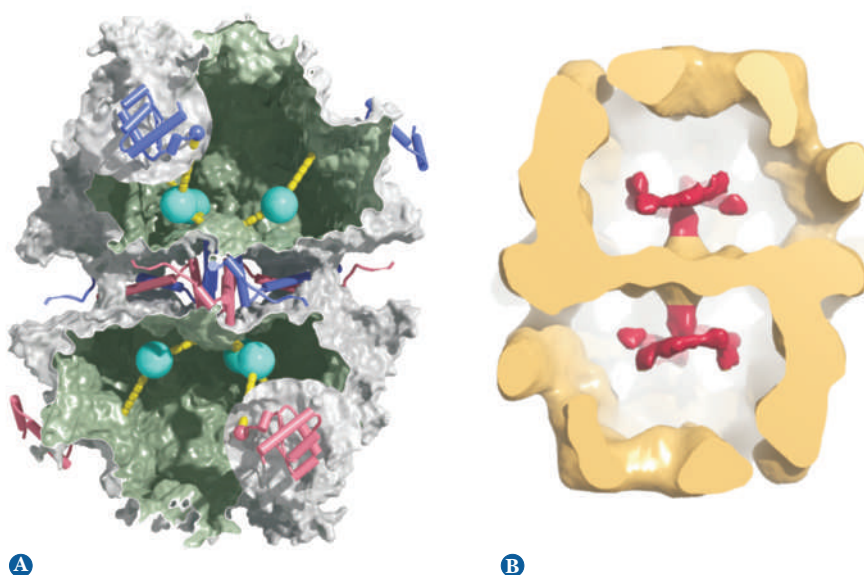


■ **FIGURE 21.16** Structural similarities between coenzyme A and the phosphopantetheine group of ACP.

hexanoyl-ACP. The same series of reactions is repeated until palmitoyl-ACP is produced. In mammalian systems, the process stops at C<sub>16</sub> because the fatty-acid synthase does not produce longer chains. Mammals produce longer-chain fatty acids by modifying the fatty acids formed by the synthase reaction.

Fatty-acid synthases from different types of organisms have markedly different characteristics. In *Escherichia coli*, the multienzyme system consists of an aggregate of separate enzymes, including a separate ACP. The ACP is of primary importance to the complex and is considered to occupy a central position in it. The phosphopantetheine group plays the role of a “swinging arm,” much like that of biotin, which was discussed earlier in this chapter. This bacterial system has been extensively studied and has been considered a typical example of a fatty-acid synthase. In eukaryotes, however, fatty-acid synthesis occurs on a multienzyme complex. In yeast, this complex consists of two different types of subunits, called  $\alpha$  and  $\beta$ , arranged in an  $\alpha_6\beta_6$  complex. In mammals, the fatty-acid synthase contains only one type of subunit, but the active enzyme is a dimer of this single subunit. It has been determined by X-ray crystallography that the mammalian synthase contains two reaction chambers in which the various components are held in proximity to each other as the reaction proceeds. Each of the subunits is a *multifunctional enzyme* that catalyzes reactions requiring several different proteins in the *E. coli* system. The structure of fungal fatty-acid synthase was recently elucidated in even greater detail by X-ray crystallography. The results demonstrated that the multiple active sites for the synthase reactions are arranged in the reaction chamber so that a circular movement of the ACP-bound substrates can deliver the substrate to each specific active site (Figure 21.17). The growing fatty-acid chain swings back and forth between enzyme activities contained on different subunits, using ACP as a “swinging arm.” Like the bacterial system, the eukaryotic system keeps all the components of the reaction in proximity to one another, which shows us another example of the advantages to multienzyme complexes.

Several additional reactions are required for the elongation of fatty-acid chains and the introduction of double bonds. When mammals produce fatty acids with longer chains than that of palmitate, the reaction does not involve cytosolic fatty-acid synthase. There are two sites for the chain-lengthening reactions: the endoplasmic reticulum (ER) and the mitochondrion. In the chain-lengthening reactions in the mitochondrion, the intermediates are of the acyl-CoA type rather than the acyl-ACP type. In other words, the chain-lengthening reactions in the mitochondrion are the reverse of the catabolic reactions of fatty acids, with acetyl-CoA as the source of added carbon atoms; this is a difference between

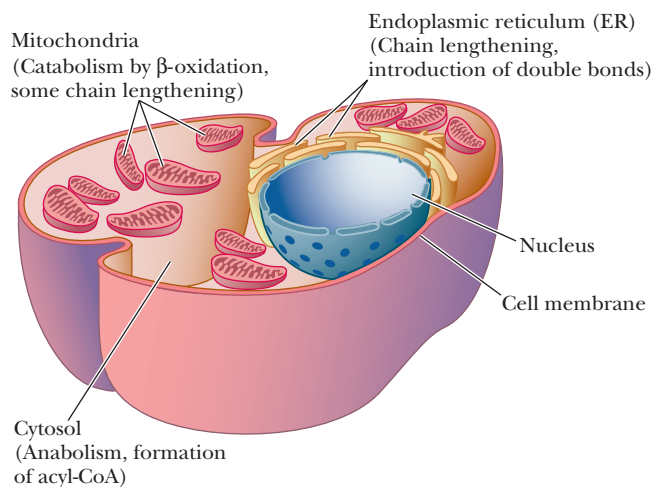


■ **FIGURE 21.17** The architecture of fungal fatty-acid synthase. (a) Side view of two reaction chambers, one on top of the other. The ACP portions are shown as cyan spheres and yellow lines. (b) A comparison structure based on electron microscopy. [Adapted with permission from S. Jenni et al., *Science* 316, 258 (2007).]

the main pathway of fatty-acid biosynthesis and these modification reactions. In the ER, the source of additional carbon atoms is malonyl-CoA. The modification reactions in the ER also differ from the biosynthesis of palmitate in that, like the mitochondrial reaction, no intermediates are bound to ACP.

Reactions in which a double bond is introduced in fatty acids mainly take place on the ER. The insertion of the double bond is catalyzed by a mixed-function oxidase that requires molecular oxygen ( $O_2$ ) and NAD(P)H. During the reaction, both NAD(P)H and the fatty acid are oxidized, while oxygen is reduced to water. Reactions linked to molecular oxygen are comparatively rare (Section 19.9). Mammals cannot introduce a double bond beyond carbon atom 9 (counting from the carboxyl end) of the fatty-acid chain. As a result, linoleate [ $CH_3-(CH_2)_4-CH=CH-CH_2-CH=CH-(CH_2)_7-COO^-$ ], with two double bonds, and linolenate [ $CH_3-(CH_2)_4-CH=CH-CH_2CH=CH-CH_2-CH=CH-(CH_2)_4COO^-$ ], with three double bonds, must be included in the diets of mammals. They are **essential fatty acids** because they are precursors of other lipids, including prostaglandins.

Even though both the anabolism and the catabolism of fatty acids require successive reactions of two-carbon units, the two pathways are not the exact reversal of each other. The differences between the two pathways are summarized in Table 21.2. The sites in the cell in which various anabolic and catabolic reactions take place are shown in Figure 21.18.



■ **FIGURE 21.18** A portion of an animal cell, showing the sites of various aspects of fatty-acid metabolism. The cytosol is the site of fatty-acid anabolism. It is also the site of formation of acyl-CoA, which is transported to the mitochondrion for catabolism by the  $\beta$ -oxidation process. Some chain-lengthening reactions (beyond  $C_{16}$ ) take place in the endoplasmic reticulum (ER), as do reactions that introduce double bonds.

TABLE 21.2

A Comparison of Fatty-Acid Degradation and Biosynthesis	
Degradation	Biosynthesis
1. Product is acetyl-CoA	Precursor is acetyl-CoA
2. Malonyl-CoA is not involved; no requirement for biotin	Malonyl-CoA is source of two-carbon units; biotin required
3. Oxidative process; requires $\text{NAD}^+$ and FAD and produces ATP	Reductive process; requires NADPH and ATP
4. Fatty acids form thioesters with CoA-SH	Fatty acids form thioesters with acyl carrier proteins (ACP-SH)
5. Starts at carboxyl end ( $\text{CH}_3\text{CO}_2^-$ )	Starts at methyl end ( $\text{CH}_3\text{CH}_2^-$ )
6. Occurs in the mitochondrial matrix, with no ordered aggregate of enzymes	Occurs in the cytosol, catalyzed by an ordered multienzyme complex
7. $\beta$ -Hydroxyacyl intermediates have the L configuration	$\beta$ -Hydroxyacyl intermediates have the D configuration

## 21.7 Synthesis of Acylglycerols and Compound Lipids

Other lipids, including triacylglycerols, phosphoacylglycerols, and steroids, are derived from fatty acids and metabolites of fatty acids, such as acetoacetyl-CoA. Free fatty acids do not occur in the cell to any great extent; they are normally found incorporated in triacylglycerols and phosphoacylglycerols. The biosynthesis of these two types of compounds takes place principally on the ER of liver cells or fat cells (adipocytes).

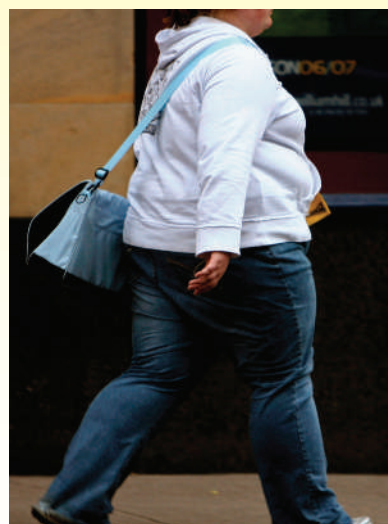
### Triacylglycerols

The glycerol portion of lipids is derived from glycerol-3-phosphate, a compound available from glycolysis. In liver and kidney, another source is glycerol released by degradation of acylglycerols. An acyl group of a fatty acid is transferred from an acyl-CoA. The products of this reaction are CoA-SH and a *lysophosphatidate* (a monoacylglycerol phosphate) (Figure 21.19). The acyl group is shown as esterified at carbon atom 2 (C-2) in this series of equations, but it is equally likely that it is esterified at C-1. A second acylation reaction takes place,

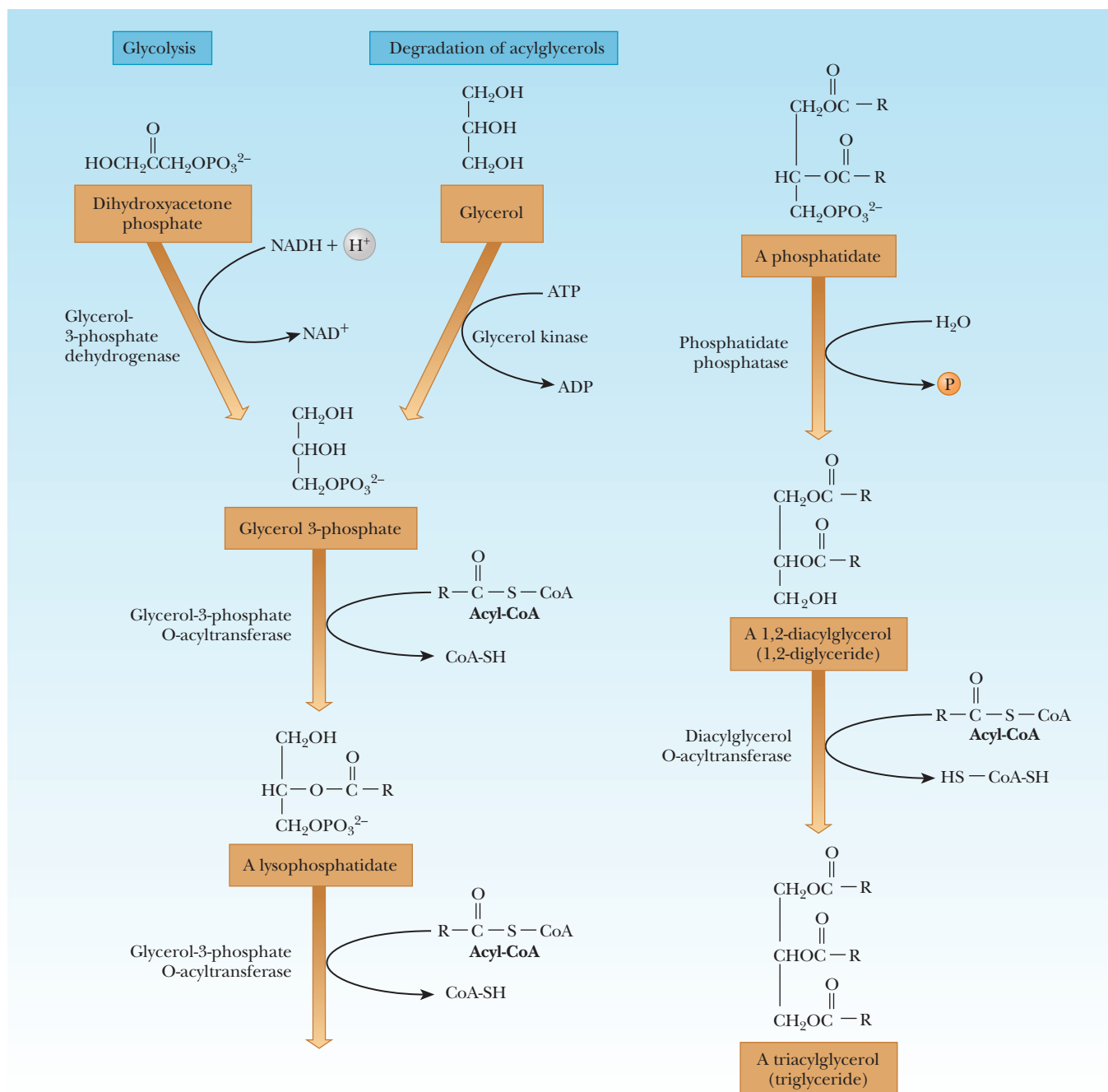
## Biochemical Connections GENETICS

### A Gene for Obesity

Obesity has long been associated with several known disease states, such as diabetes and even cancer, making it a hot topic in modern society. Researchers recently identified the first gene that showed a clear relationship to a tendency toward obesity. The gene has been labeled *FTO*. Interestingly, although this gene is positively correlated with obesity, nobody yet knows what it does. A British team of scientists studied samples from more than 4000 individuals and identified the *FTO* gene, which has been shown to be related to body mass index (BMI). A specific variant of *FTO* with a single nucleotide change was found. Individuals who had two copies of the variant were 1.67 times more likely to be obese than individuals lacking any copies of the variant. The researchers said that even though they currently do not have a function for the gene, its high correlation with obesity will have people racing to understand it.



Getty Images



■ **FIGURE 21.19** Pathways for the biosynthesis of triacylglycerols.

catalyzed by the same enzyme, producing a *phosphatidate* (a diacylglyceryl phosphate). Phosphatidates occur in membranes and are precursors of other phospholipids. The phosphate group of the phosphatidate is removed by hydrolysis, producing a *diacylglycerol*. A third acyl group is added in a reaction in which the source of the acyl group is an acyl-CoA rather than the free fatty acid.

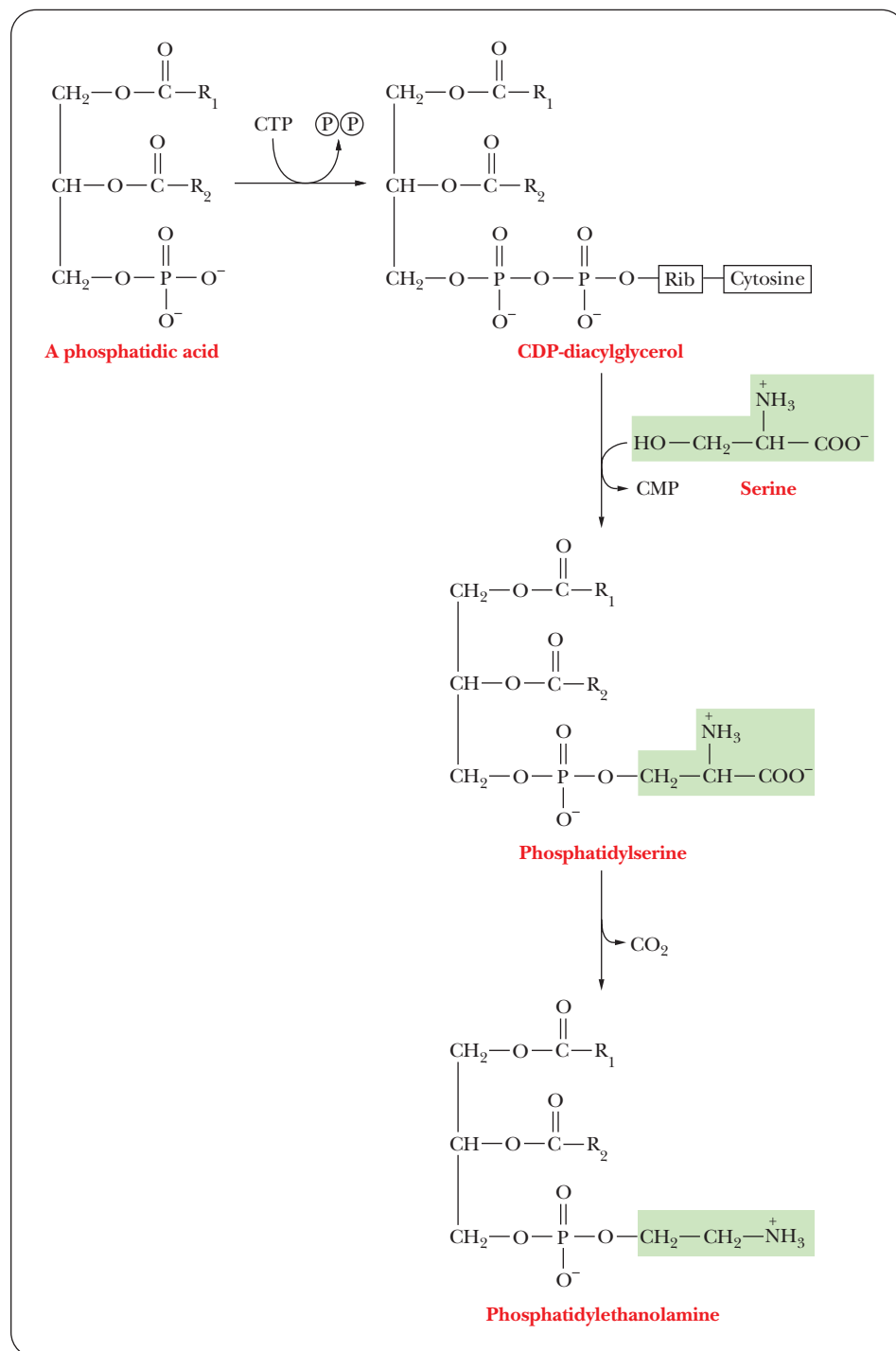
### How does the biosynthesis of phosphoacylglycerols take place?

Phosphoacylglycerols (phosphoglycerides) are based on phosphatidates, with the phosphate group esterified to another alcohol, frequently a nitrogen-containing alcohol such as ethanolamine [see the discussion of phosphoacylglycerols (phosphoglycerides) in Section 8.2]. The conversion of phosphatidates to other phospholipids frequently requires the presence of nucleoside triphosphates, particularly *cytidine triphosphate* (CTP). The role of CTP depends on the type of

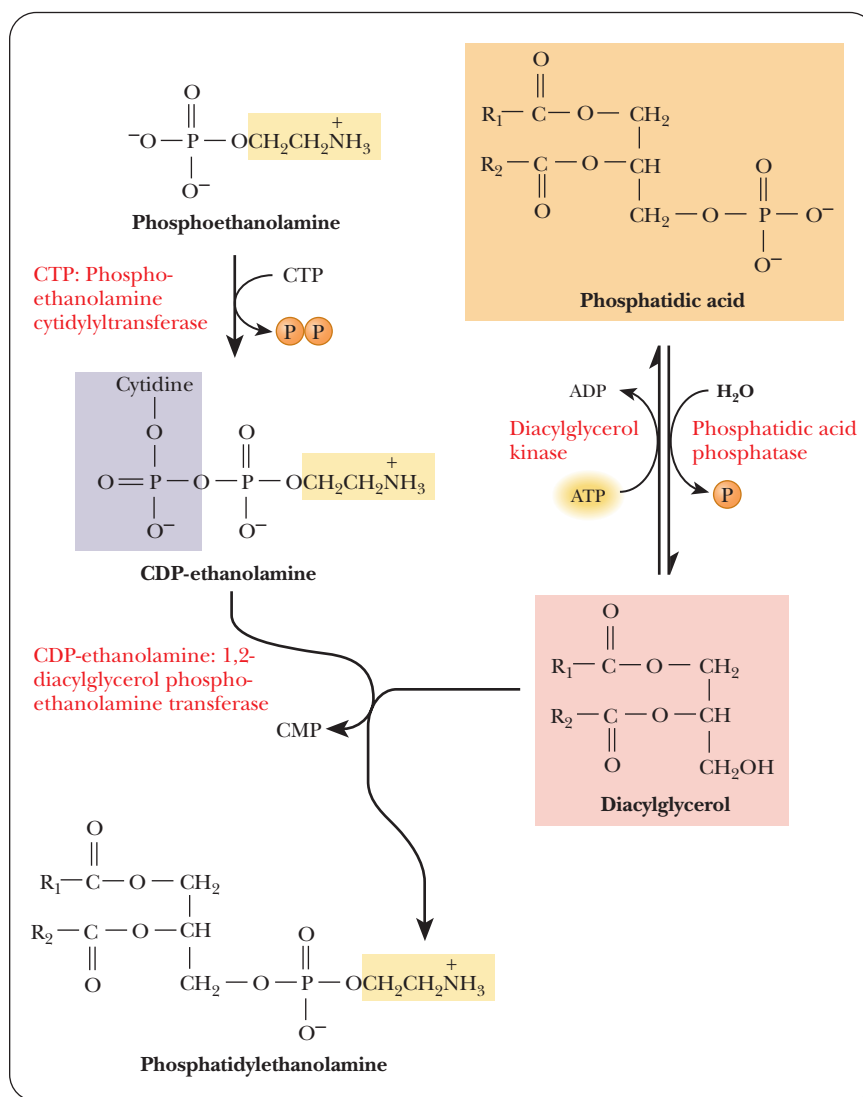


organism, because the details of the biosynthetic pathway are not the same in mammals and bacteria. We shall use a comparison of the synthesis of phosphatidylethanolamine in mammals and in bacteria (Figure 21.20) as a case study of the kinds of reactions commonly encountered in phosphoglyceride biosynthesis.

In bacteria, CTP reacts with phosphatidate to produce cytidine diphosphodiacylglycerol (a CDP diglyceride). The CDP diglyceride reacts with serine to form phosphatidylserine. Phosphatidylserine is then decarboxylated to give *phosphatidylethanolamine*. In eukaryotes, the synthesis of phosphatidylethanolamine requires two preceding steps in which the component parts are processed (Figure 21.21). The first of these two steps is the removal by hydrolysis of the phosphate group of the phosphatidate, producing a diacylglycerol; the second step is the reaction of ethanolamine phosphate with CTP to



■ **FIGURE 21.20** The biosynthesis of phosphatidylethanolamine in bacteria. See text for details about how the pathway differs in mammals.



■ **FIGURE 21.21** Production of phosphatidylethanolamine in eukaryotes.

## Biochemical Connections NUTRITION

### Acetyl-CoA Carboxylase—A New Target in the Fight against Obesity?

Malonyl-CoA has two very important functions in metabolism. First, it is the committed intermediate in fatty-acid synthesis. Second, it strongly inhibits carnitine palmitoyltransferase I and therefore fatty-acid oxidation. The level of malonyl-CoA in the cytosol can determine whether the cell will be oxidizing fats or storing fats. The enzyme that produces malonyl-CoA is acetyl-CoA carboxylase, or ACC. There are two isoforms of this enzyme encoded by separate genes. ACC1 is found in the liver and adipose tissue, while ACC2 is found in cardiac and skeletal muscle. High glucose concentrations and high insulin concentrations lead to stimulation of ACC2. Exercise has the opposite effect. During exercise, an AMP-dependent protein kinase phosphorylates ACC2 and inactivates it.

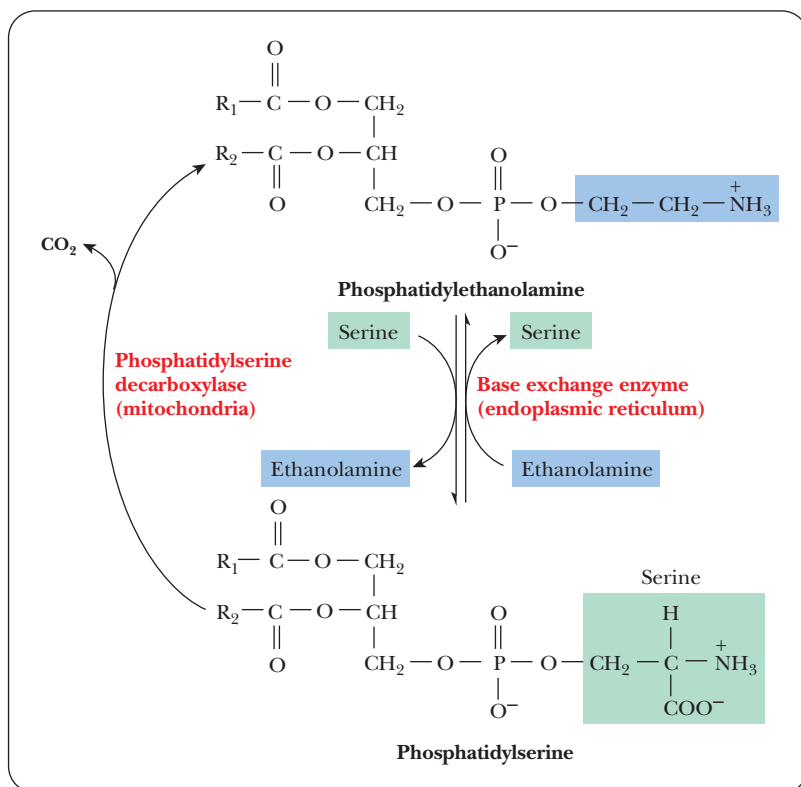
Some recent studies looked at the nature of weight gain and weight loss with respect to ACC2 (see papers by Ruderman and Flier and by Abu-Elheiga et al. cited in the bibliography at the website). The investigators created a strain of mice lacking the gene for ACC2. These mice ate more than their wild-type counterparts but had significantly lower stores of lipids (30%–40% less in skeletal muscle and 10% less in cardiac muscle). Even the

adipose tissue, which still had ACC1, showed a reduction in stored triacylglycerols of up to 50%. The mice showed no other abnormalities. They grew and reproduced normally and had normal life spans. The investigators concluded that reduced pools of malonyl-CoA due to the lack of ACC2 results in increased  $\beta$ -oxidation via removal of the block on carnitine palmitoyltransferase I, and a decrease in fatty-acid synthesis. They speculate that ACC2 would be a good target for drugs used to combat obesity.

■ The amount of white fat under the skin of the mouse on the left, which lacks the gene for ACC2, is less than that for the mouse on the right, which has the gene.



From Continuous Fatty Acid Oxidation and Reduced Fat Storage in Mice Lacking Acetyl-CoA Carboxylase 2 by Lutfi Abu-Elheiga, et al., (30 March 2001). *Science* 291 (5513):2613. Used with permission of AAAS.



■ **FIGURE 21.22** The interconversion of phosphatidylethanolamine and phosphatidylserine in mammals.

produce pyrophosphate ( $\text{PP}_i$ ) and cytidine diphosphate ethanolamine (CDP-ethanolamine). The CDP-ethanolamine and diacylglycerol react to form phosphatidylethanolamine.

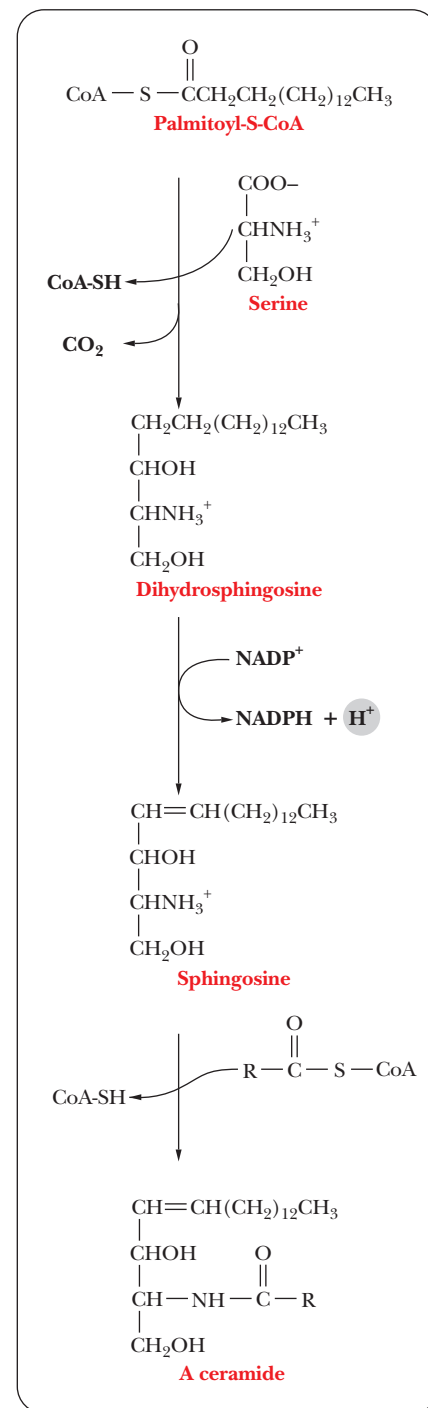
In mammals, phosphatidylethanolamine can be produced another way. Alcohol exchange from serine to ethanolamine allows the interconversion of phosphatidylethanolamine with phosphatidylserine (Figure 21.22).

### How does the biosynthesis of sphingolipids take place?

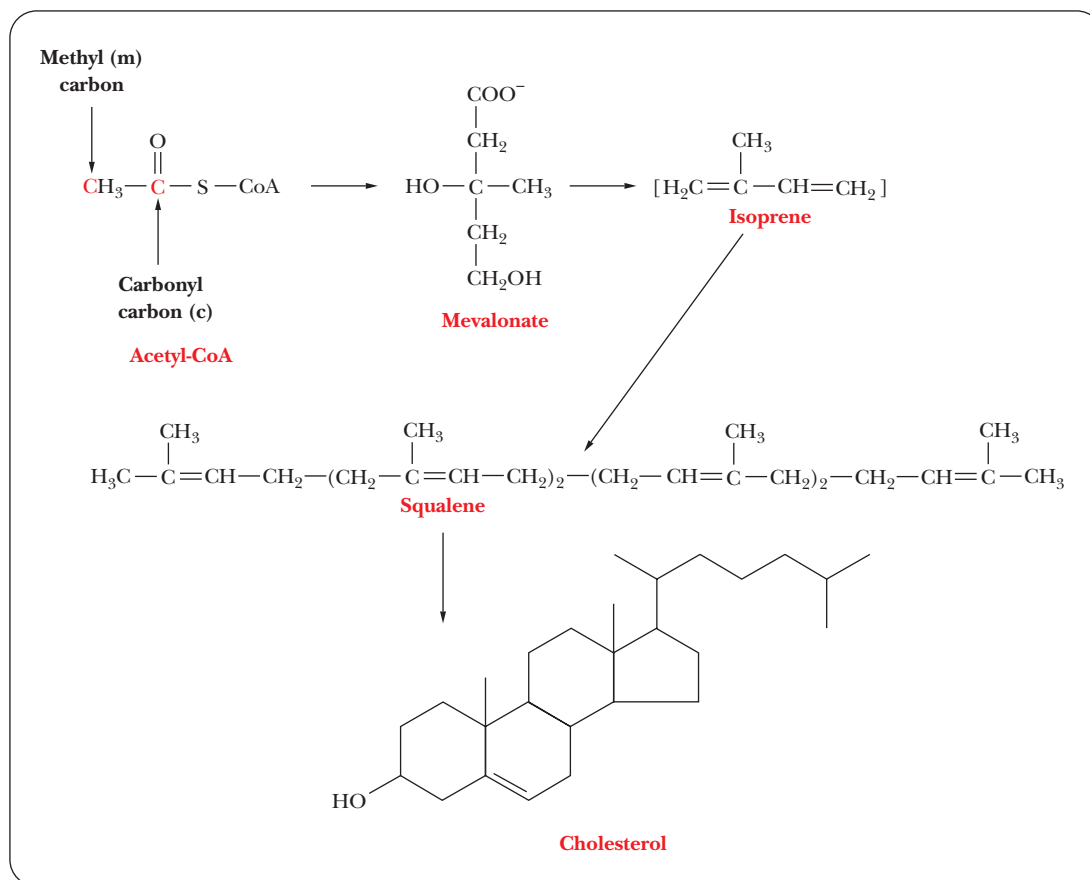
The structural basis of sphingolipids is not glycerol but *sphingosine*, a long-chain amine (see the discussion of sphingolipids in Section 8.2). The precursors of sphingosine are palmitoyl-CoA and the amino acid serine, which react to produce dihydrosphingosine. The carboxyl group of the serine is lost as  $\text{CO}_2$  in the course of this reaction (Figure 21.23). An oxidation reaction introduces a double bond, with sphingosine as the resulting compound. Reaction of the amino group of sphingosine with another acyl-CoA to form an amide bond results in an *N-acylsphingosine*, also called a **ceramide**. Ceramides in turn are the parent compounds of sphingomyelins, cerebroside, and gangliosides. Attachment of phosphorylcholine to the primary alcohol group of a ceramide produces a *sphingomyelin*, whereas attachment of sugars such as glucose at the same site produces *cerebroside*. *Gangliosides* are formed from ceramides by attachment of oligosaccharides that contain a sialic acid residue, also at the primary alcohol group. See the discussion of sphingolipids in Section 8.2 for the structures of these compounds.

## 21.8 Cholesterol Biosynthesis

The ultimate precursor of all the carbon atoms in cholesterol and in the other steroids that are derived from cholesterol is the acetyl group of acetyl-CoA. There are many steps in the biosynthesis of steroids. The condensation of three



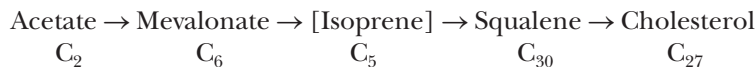
■ **FIGURE 21.23** The biosynthesis of sphingolipids. When ceramides are formed, they can react (a) with choline to yield sphingomyelins, (b) with sugars to yield cerebroside, or (c) with sugars and sialic acid to yield gangliosides.



■ **FIGURE 21.24** Outline of the biosynthesis of cholesterol.

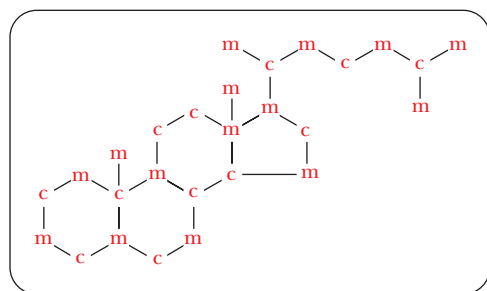
acetyl groups produces mevalonate, which contains six carbons. Decarboxylation of mevalonate produces the five-carbon isoprene unit frequently encountered in the structure of lipids. The involvement of **isoprene units** is a key point in the biosynthesis of steroids and of many other compounds that have the generic name *terpenes*. Vitamins A, E, and K come from reactions involving terpenes that humans cannot carry out. That is why we must consume these vitamins in our diets; vitamin D, the remaining lipid-soluble vitamin, is derived from cholesterol (Section 8.8). Isoprene units are involved in the biosynthesis of ubiquinone (coenzyme Q) and of derivatives of proteins and tRNA with specific five-carbon units attached. Isoprene units are often added to proteins to act as anchors when the protein is attached to a membrane.

Six isoprene units condense to form squalene, which contains 30 carbon atoms. Finally, squalene is converted to **cholesterol**, which contains 27 carbon atoms (Figure 21.24); squalene can also be converted to other sterols.



It is well established that 12 of the carbon atoms of cholesterol arise from the carboxyl carbon of the acetyl group; these are the carbon atoms labeled “c” in Figure 21.25. The other 15 carbon atoms arise from the methyl carbon of the acetyl group; these are the carbon atoms labeled “m.” We shall now look at the individual steps of the process in more detail.

The conversion of three acetyl groups of acetyl-CoA to *mevalonate* takes place in several steps (Figure 21.26). We already saw the first of these steps, the production of acetoacetyl-CoA from two molecules of acetyl-CoA, when we discussed the formation of ketone bodies and the anabolism of fatty acids. A third



■ **FIGURE 21.25** The labeling pattern of cholesterol. Each letter “m” indicates a methyl carbon and each letter “c” indicates a carbonyl carbon, all of which come from acetyl-CoA.

molecule of acetyl-CoA condenses with acetoacetyl-CoA to produce  $\beta$ -hydroxy- $\beta$ -methylglutaryl-CoA (also called HMG-CoA and 3-hydroxy-3-methylglutaryl-CoA).

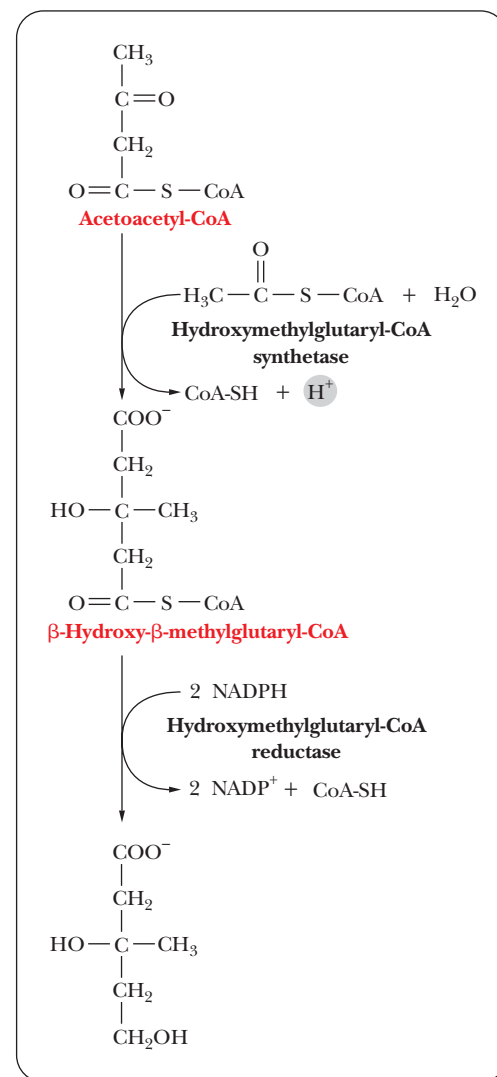
### Why is HMG-CoA so important in cholesterol biosynthesis?

This reaction is catalyzed by the enzyme hydroxymethylglutaryl-CoA synthase; one molecule of CoA-SH is released in the process. In the next reaction, the production of mevalonate from hydroxymethylglutaryl-CoA is catalyzed by the enzyme hydroxymethylglutaryl-CoA reductase (HMG-CoA reductase). A carboxyl group, the one esterified to CoA-SH, is reduced to a hydroxyl group, and the CoA-SH is released. This step is inhibited by high levels of cholesterol and is the major control point of cholesterol synthesis. It is also a target for drugs to lower cholesterol levels in the body. Drugs such as *lovastatin* are inhibitors of hydroxymethyl-CoA reductase and are widely prescribed to lower blood cholesterol levels. The drug is metabolized to mevinolinic acid, which is a transition-state analogue of a tetrahedral intermediate in the reaction catalyzed by HMG-CoA reductase (Figure 21.27).

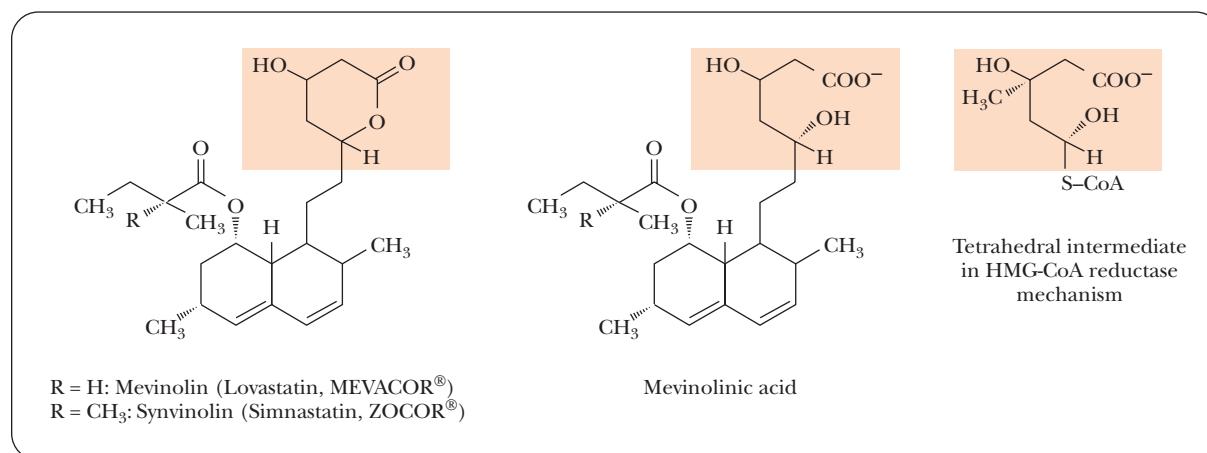
Mevalonate is then converted to an isoprenoid unit by a combination of phosphorylation, decarboxylation, and dephosphorylation reactions (Figure 21.28). Three successive reactions, each of which is catalyzed by an enzyme that requires ATP, give rise to *isopentenyl pyrophosphate*, a five-carbon isoprenoid derivative. Isopentenyl pyrophosphate and *dimethylallyl pyrophosphate*, another isoprenoid derivative, can be interconverted in a rearrangement reaction catalyzed by the enzyme isopentenyl pyrophosphate isomerase.

Condensation of isoprenoid units then leads to the production of squalene and, ultimately, cholesterol. Both of the isoprenoid derivatives we have met so far are required. Two further condensation reactions take place. As a result, *farnesyl pyrophosphate*, a 15-carbon compound, is produced. Two molecules of farnesyl pyrophosphate condense to form *squalene*, a 30-carbon compound. The reaction is catalyzed by squalene synthase, and NADPH is required for the reaction.

Figure 21.29 shows the conversion of squalene to cholesterol. The details of this conversion are far from simple. Squalene is converted to *squalene epoxide* in a reaction that requires both NADPH and molecular oxygen ( $O_2$ ). This reaction is catalyzed by squalene monooxygenase. Squalene epoxide then undergoes a complex cyclization reaction to form *lanosterol*. This remarkable reaction is catalyzed by squalene epoxide cyclase. The mechanism of the reaction is a concerted reaction—that is, one in which each part is essential for any other part to take place. No portion of a concerted reaction can be left out or

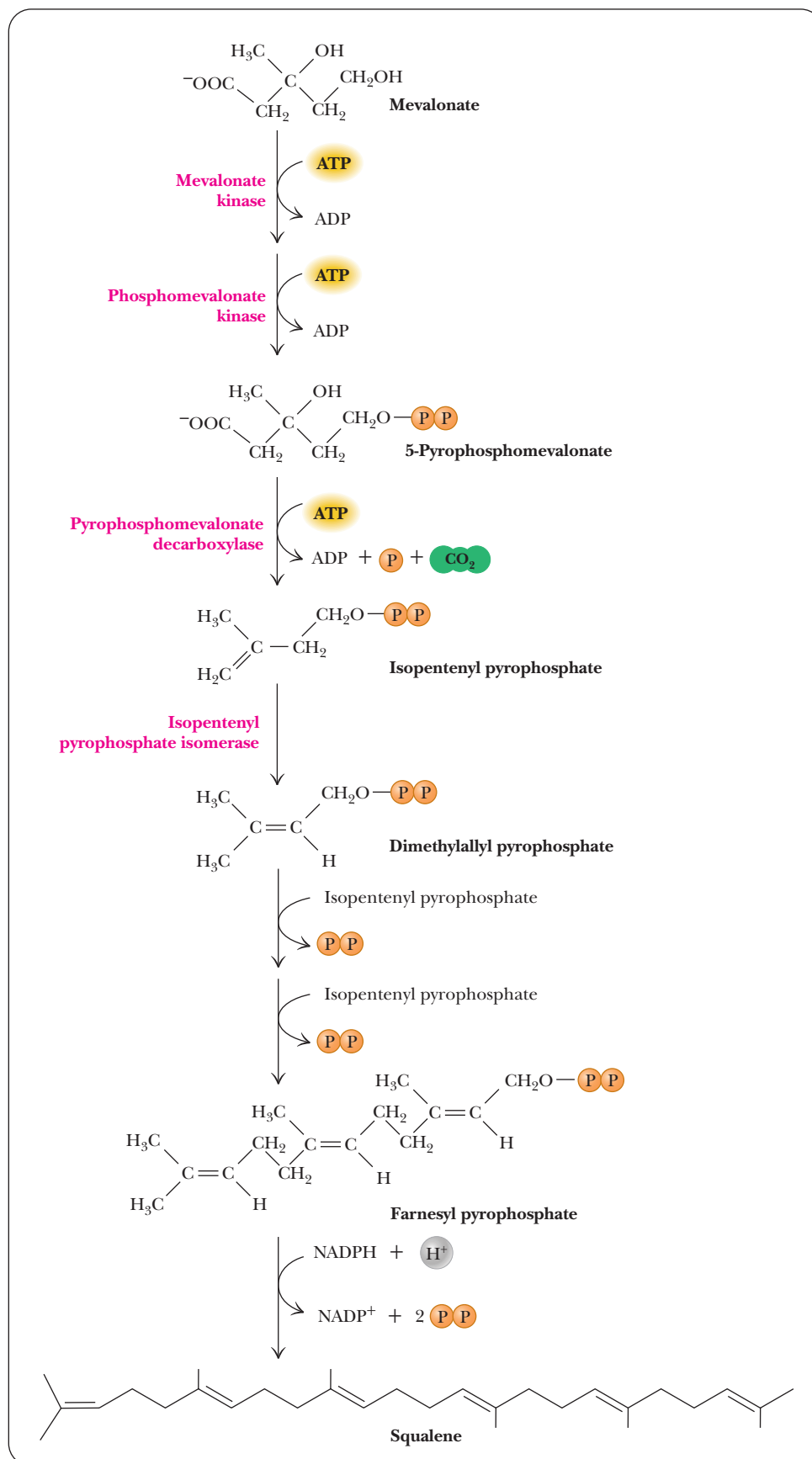


■ **FIGURE 21.26** The biosynthesis of mevalonate.

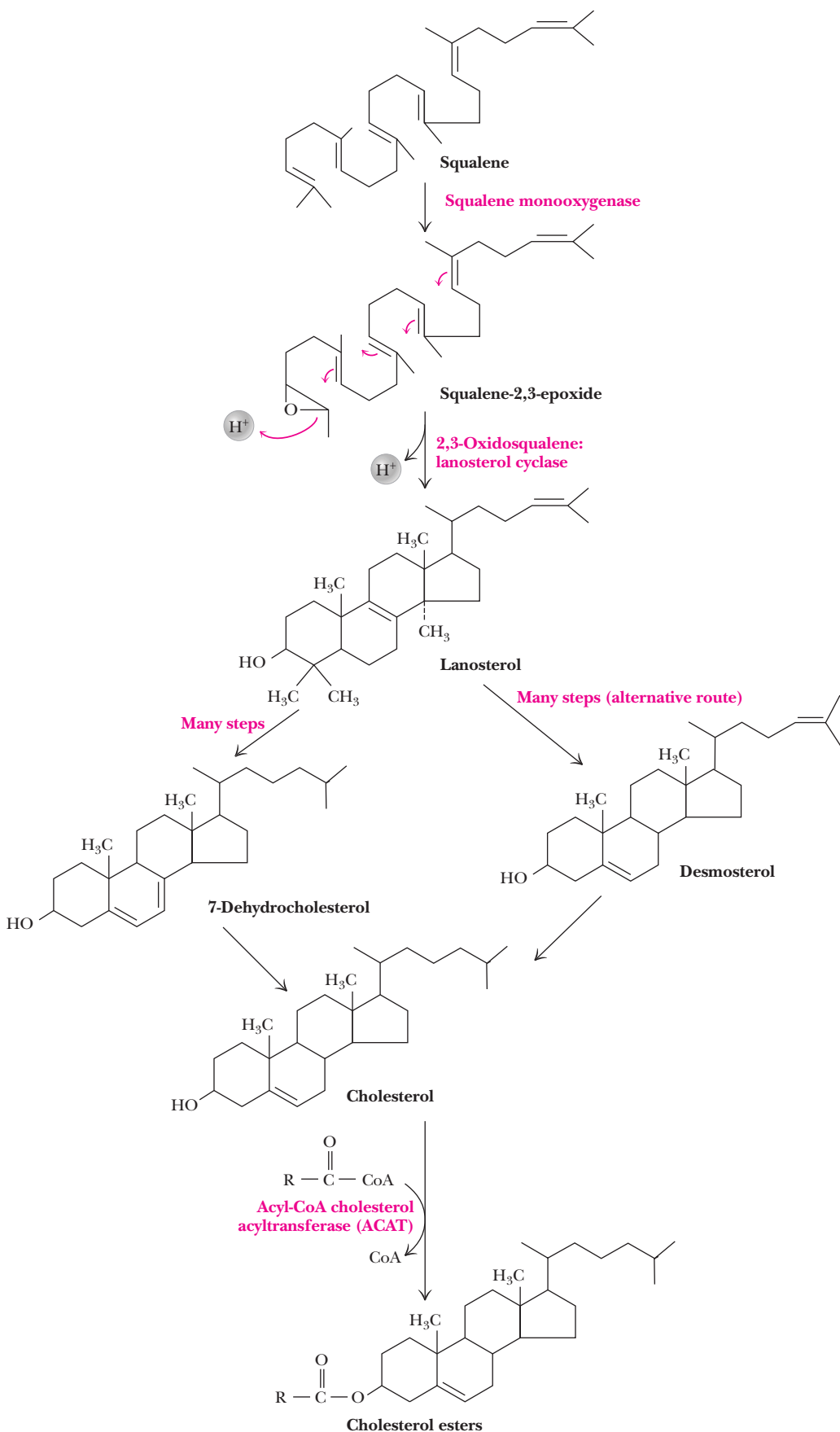


■ **FIGURE 21.27** The structures of (inactive) lovastatin and synvinolin, (active) mevinolinic acid, and the tetrahedral intermediate in the HMG-CoA reductase mechanism.





■ **FIGURE 21.28** The conversion of mevalonate to squalene.



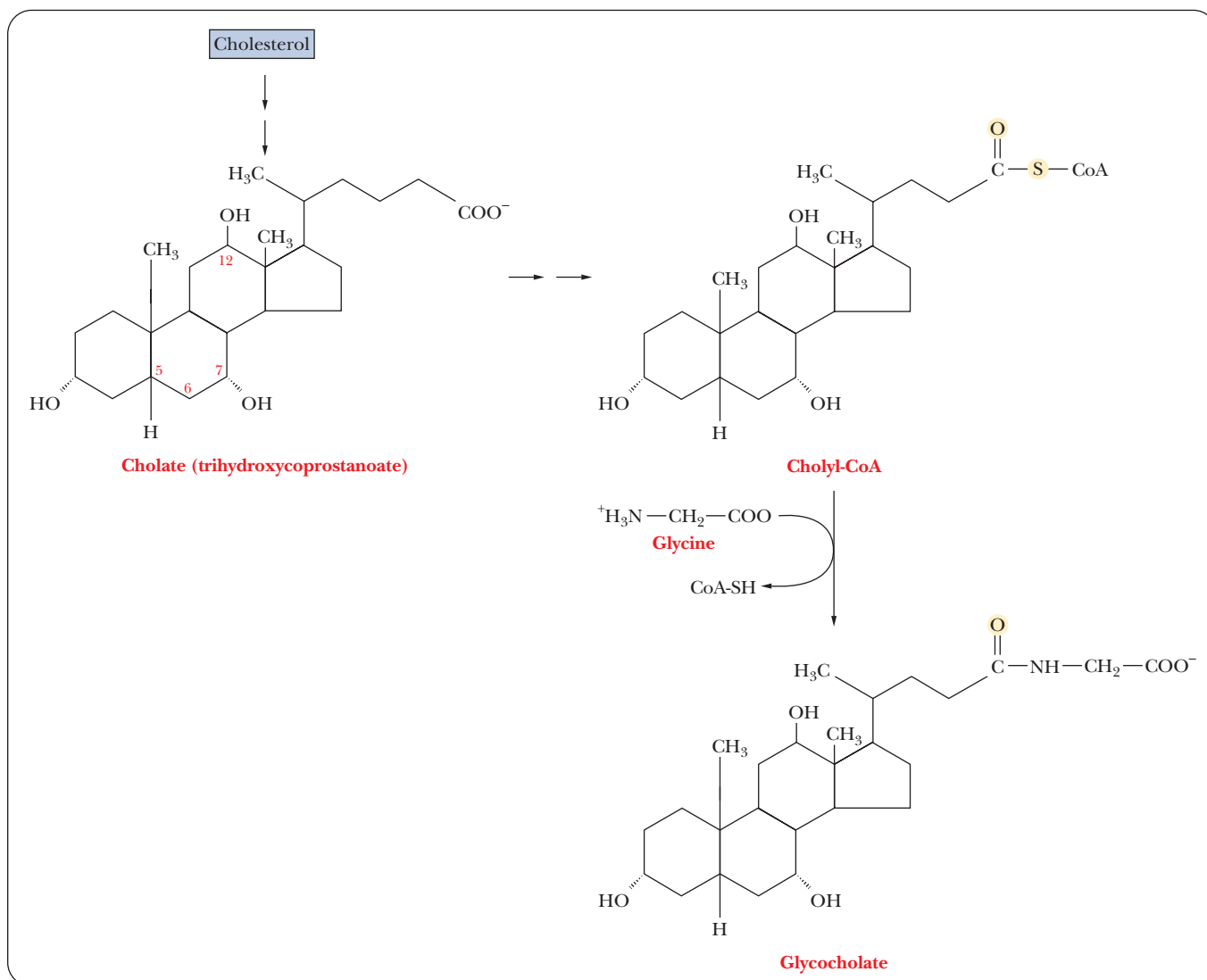
■ **FIGURE 21.29** Cholesterol is synthesized from squalene via lanosterol. The primary route from lanosterol involves 20 steps, the last of which converts 7-dehydrocholesterol to cholesterol. An alternative route produces desmosterol as the penultimate intermediate.

changed because it all takes place simultaneously rather than in a sequence of steps. The conversion of lanosterol to cholesterol is a complex process. It is known that 20 steps are required to remove three methyl groups and to move a double bond, but we shall not discuss the details of the process.

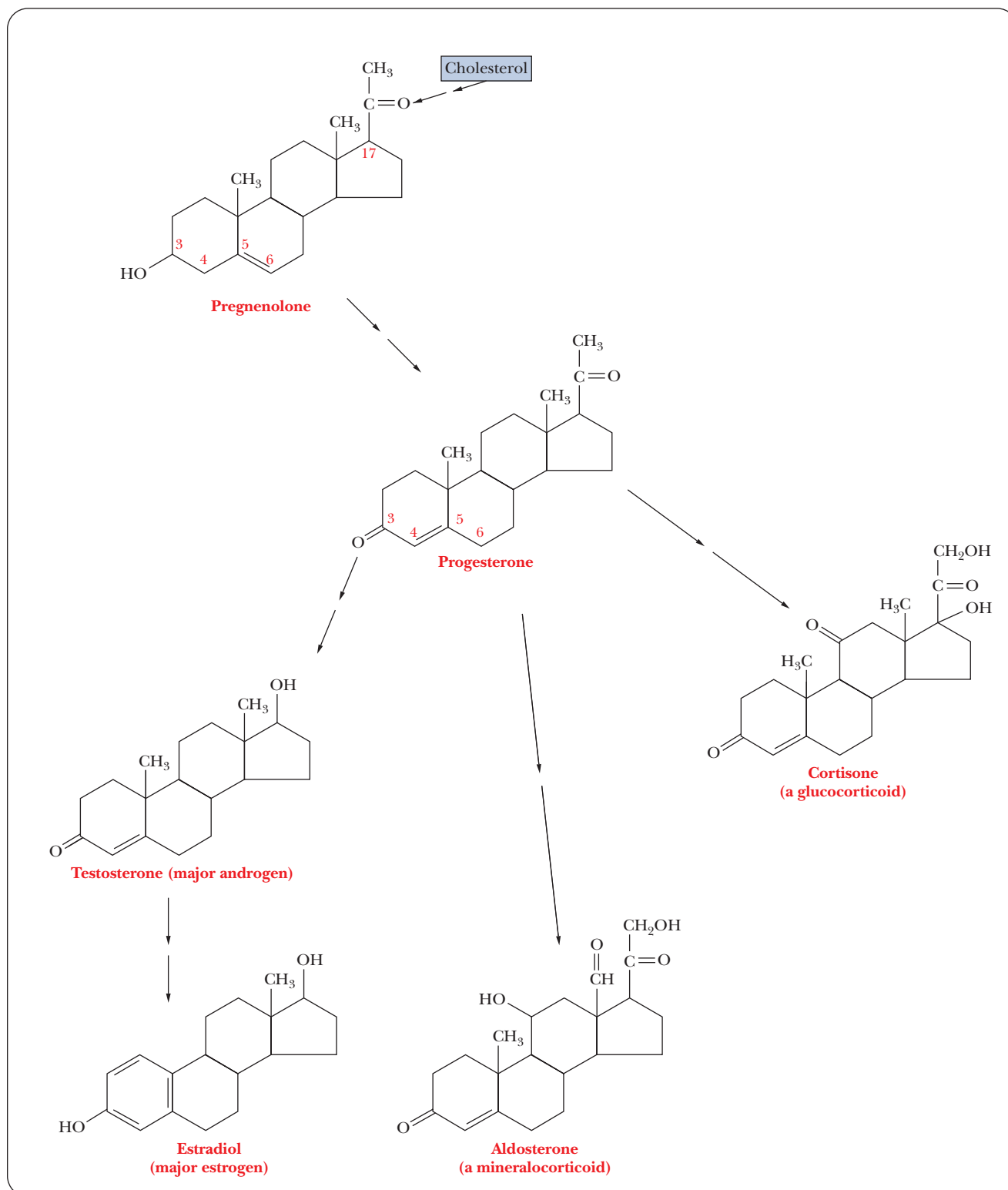
### How does cholesterol serve as a precursor of other steroids?

After cholesterol is formed, it can be converted to other steroids of widely varying physiological function. The smooth ER is an important site for both the synthesis of cholesterol and its conversion to other steroids. Most of the cholesterol formed in the liver, which is the principal site of cholesterol synthesis in mammals, is converted to *bile acids*, such as cholate and glycocholate (Figure 21.30). These compounds aid in the digestion of lipid droplets by emulsifying them and rendering them more accessible to enzymatic attack.

Cholesterol is the precursor of important **steroid hormones** (Figure 21.31), in addition to the bile acids. Like all hormones, whatever their chemical nature (Section 24.3), steroid hormones serve as signals from outside a cell that regulate metabolic processes within a cell. Steroids are best known as sex



■ **FIGURE 21.30** The synthesis of bile acids from cholesterol.



■ **FIGURE 21.31** The synthesis of steroid hormones from cholesterol.

hormones (they are components of birth-control pills), but they play other roles as well. *Pregnenolone* is formed from cholesterol, and *progesterone* is formed from pregnenolone. Progesterone is a sex hormone and is a precursor for other sex hormones, such as *testosterone* and *estradiol* (an estrogen). Other types of steroid hormones also arise from progesterone. The role of sex hormones

in sexual maturation is discussed in Section 24.3. *Cortisone* is an example of *glucocorticoids*, a group of hormones that play a role in carbohydrate metabolism, as the name implies, as well as in the metabolism of proteins and fatty acids. *Mineralocorticoids* constitute another class of hormones that are involved in the metabolism of electrolytes, including metal ions (“minerals”) and water. *Aldosterone* is an example of a mineralocorticoid. In cells in which cholesterol is converted to steroid hormones, an enlarged smooth ER is frequently observed, providing a site for the process to take place.

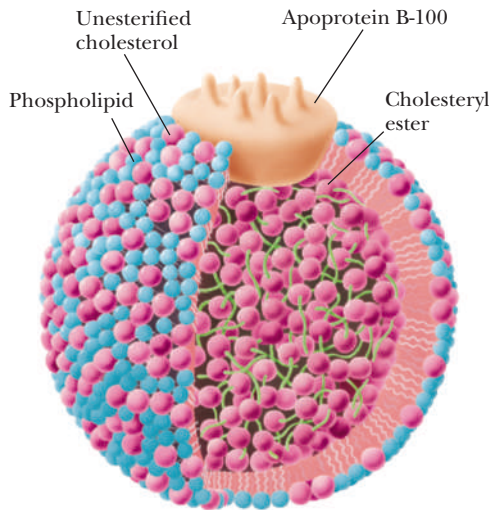
What is the role of cholesterol in heart disease?

Atherosclerosis is a condition in which arteries are blocked to a greater or lesser extent by the deposition of cholesterol plaques, which can lead to heart attacks. The process by which the clogging of arteries occurs is complex. Both diet and genetics are instrumental in the development of atherosclerosis. A diet high in cholesterol and fats, particularly saturated fats, will lead to a high level of cholesterol in the bloodstream. The body also makes its own cholesterol because this steroid is a necessary component of cell membranes. It is possible for more cholesterol to come from endogenous sources (synthesized within the body) than from the diet.

Cholesterol must be packaged for transport in the bloodstream; several classes of lipoproteins (summarized in Table 21.3) are involved in the transport of lipids in blood. These lipoprotein aggregates are usually classified by their densities. Besides chylomicrons, they include very-low-density lipoproteins (**VLDLs**), intermediate-density lipoproteins (**IDLs**), low-density lipoproteins (**LDLs**), and high-density lipoproteins (**HDLs**). The density increases as the protein content increases. LDL and HDL will play the major role in our discussion of heart disease. The protein portions of these aggregates can vary widely. The major lipids are generally cholesterol and its esters, in which the hydroxyl group is esterified to a fatty acid; triacylglycerols are also found in these aggregates. Chylomicrons are involved in the transport of dietary lipids, whereas the other lipoproteins primarily deal with endogenous lipids.

Figure 21.32 shows the architecture of an LDL particle. The interior consists of many molecules of cholesteryl esters (the hydroxyl group of the cholesterol is esterified to an unsaturated fatty acid, such as linoleate). On the surface, protein (apoprotein B-100), phospholipids, and unesterified cholesterol are in contact with the aqueous medium of the plasma. The protein portions of LDL particles bind to receptor sites on the surface of a typical cell. Refer to the discussion of membrane receptors in Section 8.6 for a description of the process by which LDL particles are taken into the cell as one aspect of receptor action. This process is typical of the mechanism of uptake of lipids by cells, and we shall use the processing of LDL as a case study. LDL is the major player in the development of atherosclerosis.

LDL particles are degraded in the cell. LDL particles are taken into the cell by the highly regulated process of endocytosis (Section 8.6), in which a portion



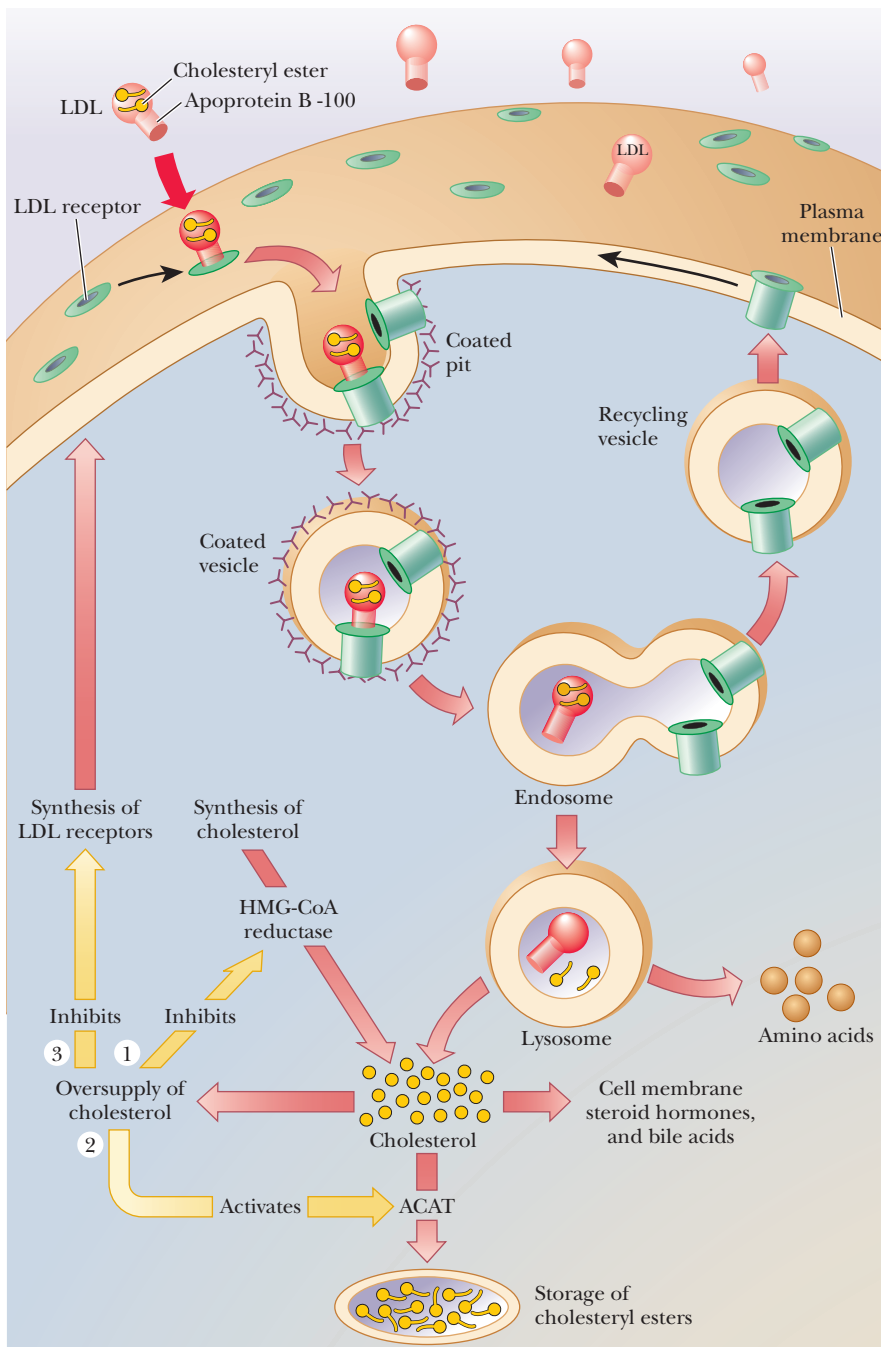
■ **FIGURE 21.32** Schematic diagram of an LDL particle. (From M. S. Brown and J. L. Goldstein, 1984, *How LDL Receptors Influence Cholesterol and Atherosclerosis*, *Sci. Amer.* **251** (5), 58–66.)

TABLE 21.3

Major Classes of Lipoproteins in Human Plasma	
Lipoprotein class	Density (g mL <sup>-1</sup> )
Chylomicrons	<0.95
VLDL	0.95 – 1.006
IDL	1.006–1.019
LDL	1.019–1.063
HDL	1.063–1.210



of the cell membrane containing the LDL particle and its receptor enters the cell. The receptor is returned to the cell surface, while the LDL particles are degraded in the lysosomes (organelles that contain degradative enzymes; see Section 1.6). The protein portion of LDL is hydrolyzed to the component amino acids, while the cholesterol esters are hydrolyzed to cholesterol and fatty acids. Free cholesterol can then be used directly as a component of membranes; the fatty acids can have any of the catabolic or anabolic fates discussed earlier in this chapter (Figure 21.33). Cholesterol not needed for membrane synthesis can be stored as oleate or palmitoleate esters in which the fatty acid is esterified to the hydroxyl group of cholesterol. The production of these esters is catalyzed by acyl-CoA:cholesterol acyltransferase (ACAT), and the presence of free cholesterol increases the enzymatic activity of ACAT. In addition,



**FIGURE 21.33** The fate of cholesterol in the cell (see page 620). ACAT is the enzyme that esterifies cholesterol for storage. (From M. S. Brown and J. L. Goldstein, 1984, *How LDL Receptors Influence Cholesterol and Atherosclerosis*, *Sci. Amer.* **251** (5), 58–66.)

# Biochemical Connections ALLIED HEALTH

## Atherosclerosis

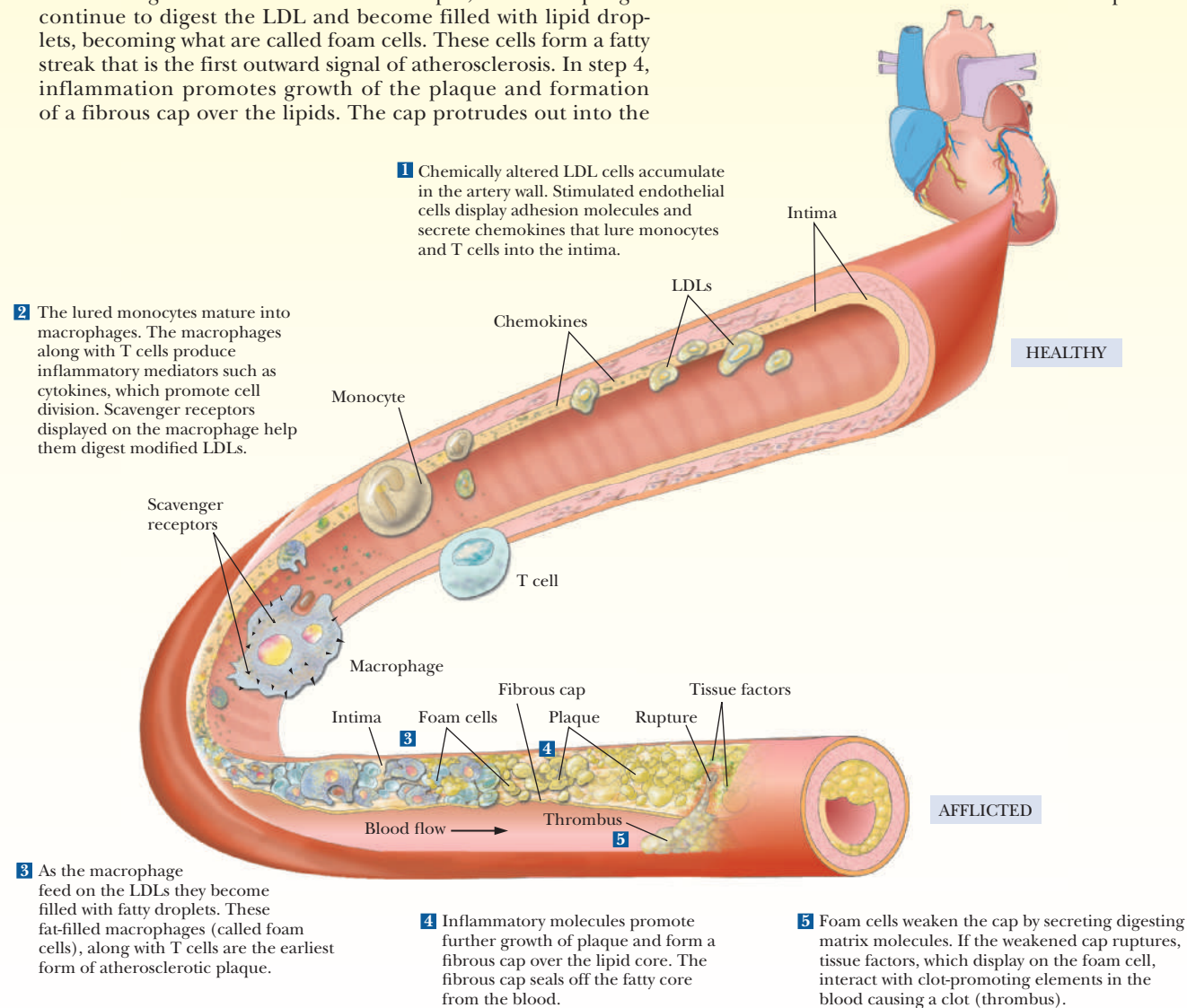
Atherosclerosis causes more deaths every year than cancer. It causes chest pain, heart attack, and stroke. For years it was described with simple models that described how the much-maligned LDL molecules would lay down plaques in blood vessels, leading to blockage. However, more recent evidence shows these old models to be overly simplistic. Lipid metabolism and LDL are certainly involved, but a more important villain in the process is inflammation.

The figure shows a more realistic and complicated picture of the formation of atherosclerosis. Excess LDL is involved in the process, but it is not caused by a simple laying down of the LDL. In step 1, the excess LDLs invade the tissue of the artery and become modified. The modified LDL molecules stimulate the production of adhesion molecules, shown as blue spikes sticking out into the blood stream. These adhesion molecules attract monocytes and T cells. The endothelial cells of the artery wall also secrete chemokines that attract the monocytes into the intima of the artery. In step 2, the monocytes mature into active macrophages and produce many inflammatory molecules. They also scavenge the modified LDLs. In step 3, the macrophages continue to digest the LDL and become filled with lipid droplets, becoming what are called foam cells. These cells form a fatty streak that is the first outward signal of atherosclerosis. In step 4, inflammation promotes growth of the plaque and formation of a fibrous cap over the lipids. The cap protrudes out into the

bloodstream but also protects the blood from the deposits. In the final step, more inflammatory molecules can cause the cap to rupture. The foam cells also release tissue factor, a potent blood-clotting agent. The ruptured plaque leads to the formation of a thrombus, or blood clot in the artery. If the clot is big enough, it can cause a heart attack—the death of cardiac cells.

Recent research has shown a direct connection between inflammation and high lipid levels in the blood. Cytokines play a key role in regulating the levels of enzymes that control lipid metabolism. In mice deficient in LDL receptors (these mice cannot control lipid levels in their blood), inhibition of enzymes produced by cytokines led to a lowering of LDL levels. This result shows clearly that the immune system affects lipid metabolism and provides indicators as to the direction of future research on the subject.

From a practical standpoint, the take-home message may still be the same. Practices that reduce LDL are still considered necessary for a healthy heart. However, when scientists try to come up with medical countermeasures, this full understanding of all the chemical factors involved in atherosclerosis is critical to the development of cures.



■ The formation of atherosclerosis, depicting the growth of an atherosclerotic plaque in a coronary artery. [From P. Libby, *Atherosclerosis: The New View*. Sci. Amer. 286 (5), 50–51 (2002).]

cholesterol inhibits both the synthesis and the activity of the enzyme hydroxymethylglutaryl-CoA reductase (HMG-CoA reductase). This enzyme catalyzes the production of mevalonate, the reaction that is the committed step in cholesterol biosynthesis. This point has important implications. Dietary cholesterol suppresses the synthesis of cholesterol by the body, especially in tissues other than the liver. A third effect of the presence of free cholesterol in the cell is inhibition of synthesis of LDL receptors. As a result of reduction in the number of receptors, cellular uptake of cholesterol is inhibited, and the level of LDL in the blood increases, leading to the deposition of atherosclerotic plaques.

The crucial role of LDL receptors in maintaining the level of cholesterol in the bloodstream is especially clear in the case of *familial hypercholesterolemia*, which results from a defect in the gene that codes for the active receptors. An individual who has one gene that codes for the active receptor and one defective gene is heterozygous for this trait. Heterozygotes have blood cholesterol levels that are above average; therefore, they are at higher risk for heart disease than the general population. An individual with two defective genes, and thus with no active LDL receptor, is homozygous for the trait. Homozygotes have very high blood cholesterol levels from birth, and there are recorded cases of heart attacks in two-year-olds with this condition. Patients who are homozygous for familial hypercholesterolemia usually die before age 20. Another genetic abnormality involved in hypercholesterolemia is the one that gives rise to a faulty apolipoprotein E, a component of IDL and VLDL, which is involved in the uptake of lipids by the cell. The unfortunate result is the same.

Before we leave this discussion, we should mention the “good” cholesterol, HDL. Unlike LDL, which transports cholesterol from the liver to the rest of the body, HDL transports it back to the liver for degradation to bile acids. It is desirable to have low levels of cholesterol and LDL in the bloodstream, but it is also desirable to have as high a proportion of total cholesterol as possible in the form of HDL. It is well known that high levels of LDL and low levels of HDL are correlated with the development of heart disease. Factors that are known to increase HDL levels, such as regular strenuous exercise, decrease the probability of heart disease. Smoking reduces the level of HDL and is highly correlated with heart disease.

## SUMMARY

**How are lipids involved in the generation and storage of energy?** We have already seen how carbohydrates are processed catabolically and anabolically. Lipids are another class of nutrient. The catabolic oxidation of lipids releases large quantities of energy, whereas the anabolic formation of lipids represents an efficient way of storing chemical energy.

**How are fatty acids transported to the mitochondrion for oxidation?** After an initial activation step in the cytosol, with formation of an acyl-CoA corresponding to each fatty acid, each acyl group is transesterified to carnitine for transport across the intermembrane space of the mitochondrion. The acyl group is again transesterified to form an acyl-CoA.

**How does oxidation of fatty acids take place?** The oxidation of fatty acids, which takes place in the mitochondrial matrix, is the chief source of energy in the catabolism of lipids. In this process, two-carbon units are successively removed from the carboxyl end of the fatty acid to produce acetyl-CoA, which subsequently enters the citric acid cycle. The reactions

that liberate the acetyl-CoA units from a fatty acid produce NADH and FADH<sub>2</sub>, which eventually produce ATP via the electron transport chain.

**What is the energy yield from the oxidation of fatty acids?**

There is a net yield of 120 ATP molecules for each molecule of stearic acid (an 18-carbon compound) that is completely oxidized to carbon dioxide and water. The source of these ATP molecules is the production of NADH and FADH<sub>2</sub> in the  $\beta$ -oxidation pathway, as well as the NADH, FADH<sub>2</sub>, and GTP produced when the acetyl-CoA molecules are processed through the electron transport chain.

**How does the oxidation of unsaturated fatty acids differ from that of saturated fatty acids?**

The pathway of catabolism of fatty acids includes reactions in which unsaturated, as well as saturated, fatty acids can be metabolized. Odd-numbered fatty acids can also be metabolized by converting their unique breakdown product, propionyl-CoA, to succinyl-CoA, an intermediate of the citric acid cycle.

**Do acetone and acetyl-CoA have a connection in lipid metabolism?** Ketone bodies are substances related to acetone that are produced when an excess of acetyl-CoA results from  $\beta$ -oxidation. This situation can arise from a large intake of lipids and a low intake of carbohydrates or can occur in diabetes, in which the inability to metabolize carbohydrates causes an imbalance in the breakdown products of carbohydrates and lipids.

**How do the first steps of fatty-acid synthesis take place?** The anabolism of fatty acids proceeds by a different pathway from  $\beta$ -oxidation. Acetyl-CoA is transported to the cytosol, where it is converted to malonyl-CoA. Some of the most important differences between the fatty-acid anabolism and catabolism are the requirement for biotin in anabolism, but not in catabolism, and the requirement for NADPH in anabolism, rather than the  $\text{NAD}^+$  required in catabolism.

**What is the mode of action of fatty-acid synthase?** Fatty-acid biosynthesis occurs in the cytosol, catalyzed by an ordered multienzyme complex called fatty-acid synthase.

**How does the biosynthesis of phosphoacylglycerols take place?** Most compound lipids, such as triacylglycerols, phosphoacylglycerols, and sphingolipids, have fatty acids as precursors. In the case of phosphoacylglycerols, two fatty acids and

phosphoric acid are added to a glycerol backbone. The addition of remaining groups requires nucleoside triphosphates and differs between mammals and bacteria.

**How does the biosynthesis of sphingolipids take place?** Fatty acids are linked to a sphingosine backbone molecule, producing ceramides. Other moieties, including sugars, are added, producing gangliosides and other compounds.

**Why is HMG-CoA so important in cholesterol biosynthesis?** The starting material for steroid biosynthesis is acetyl-CoA. Isoprene units are formed from acetyl-CoA in the early stages of a lengthy process that leads ultimately to cholesterol. HMG-CoA is a key intermediate, and its formation is a target of cholesterol-lowering drugs.

**How does cholesterol serve as a precursor of other steroids?** Cholesterol is converted to other steroids, including bile acids, sex hormones, glucocorticoids, and mineralocorticoids.

**What is the role of cholesterol in heart disease?** Cholesterol must be packaged for transport in the bloodstream; several classes of lipoproteins are involved. One class is LDL (low-density lipoprotein, or “bad cholesterol”) whereas another class is HDL (high-density lipoprotein, or “good cholesterol”). Both dietary cholesterol and genetic factors influence the role of cholesterol in heart disease.

## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

### 21.1 Lipids Are Involved in the Generation and Storage of Energy

1. **Reflect and Apply** (a) The major energy storage compound of animals is fats (except in muscles). Why would this be advantageous? (b) Why don't plants use fats/oils as their *major* energy storage compound?

### 21.2 Catabolism of Lipids

2. **Recall** What is the difference between phospholipase  $A_1$  and  $A_2$ ?
3. **Recall** How are lipases activated hormonally?
4. **Recall** What is the metabolic purpose of linking a fatty acid to coenzyme A?
5. **Recall** Outline the role of carnitine in the transport of acyl-CoA molecules into the mitochondrion. How many enzymes are involved? What are they called?
6. **Recall** What is the difference between the type of oxidation catalyzed by acyl-CoA dehydrogenase and that catalyzed by  $\beta$ -hydroxy-CoA dehydrogenase?
7. **Recall** Draw a six-carbon saturated fatty acid and show where the double bond is created during the first step of  $\beta$ -oxidation. What is the orientation of this bond?
8. **Reflect and Apply** Why does the degradation of palmitic acid (see Question 12) to eight molecules of acetyl-CoA require seven, rather than eight, rounds of the  $\beta$ -oxidation process?
9. **Reflect and Apply** Given the nature of the hormonal activation of lipases, what carbohydrate pathways would be activated or inhibited under the same conditions?

### 21.3 The Energy Yield from the Oxidation of Fatty Acids

10. **Recall** Compare the energy yields from the oxidative metabolism of glucose and of stearic acid. To be fair, calculate it on the basis of ATP equivalents per carbon and also ATP equivalents per gram.
11. **Recall** Which generates more ATP—the processing of the reduced electron equivalents formed during  $\beta$ -oxidation through the electron transport chain, or the processing of the acetyl-CoA generated from  $\beta$ -oxidation through the citric acid cycle and the electron transport chain?
12. **Mathematical** Calculate the ATP yield for the complete oxidation of one molecule of palmitic acid (16 carbons). How does this figure differ from that obtained for stearic acid (18 carbons)? Consider the  $\beta$ -oxidation steps, processing of acetyl-CoA through the citric acid cycle, and electron transport.
13. **Reflect and Apply** It is frequently said that camels store water in their humps for long desert journeys. How would you modify this statement on the basis of information in this chapter?

### 21.4 Catabolism of Unsaturated Fatty Acids and Odd-Carbon Fatty Acids

14. **Recall** Describe briefly how  $\beta$ -oxidation of an odd-chain fatty acid is different from that for an even-chain fatty acid.
15. **Recall** You hear a fellow student say that the oxidation of unsaturated fatty acids requires exactly the same group of enzymes as the oxidation of saturated fatty acids. Is the statement true or false? Why?



16. **Recall** What are the unique enzymes needed to  $\beta$ -oxidize a mono-unsaturated fatty acid?
17. **Recall** What are the unique enzymes needed to  $\beta$ -oxidize a polyunsaturated fatty acid?
18. **Mathematical** Calculate the net ATP yield from the complete processing of a saturated fatty acid containing 17 carbons. Consider the  $\beta$ -oxidation steps, processing of acetyl-CoA through the citric acid cycle, and electron transport.
19. **Mathematical** Calculate the net ATP yield from oleic acid (18:1  $\Delta^9$ ). *Hint:* Remember the step that bypasses acyl-CoA dehydrogenase.
20. **Mathematical** Calculate the net ATP yield from linoleic acid (18:2  $\Delta^{9,12}$ ). For this calculation, assume that the loss of an NADPH is the same as the loss of an NADH.
21. **Reflect and Apply** How many cycles of  $\beta$ -oxidation are required to process a fatty acid with 17 carbons?
22. **Reflect and Apply** It has been stated many times that fatty acids cannot yield a *net* gain in carbohydrates. Why can odd-chain fatty acids be thought to break this rule to a small extent?
34. **Recall** What is the purpose of having ACP as a distinct activating group for fatty-acid synthesis?
35. **Recall** Why are linoleate and linolenate considered essential fatty acids? What step in production of polyunsaturated fatty acids are mammals unable to perform?
36. **Reflect and Apply** Is it possible to convert fatty acids to other lipids without acyl-CoA intermediates?
37. **Reflect and Apply** What is the role of citrate in the transport of acetyl groups from the mitochondrion to the cytosol?
38. **Reflect and Apply** In the mitochondrion is a short-chain carnitine acyltransferase that can take acetyl groups from acetyl-CoA and transfer them to carnitine. How might this be related to lipid biosynthesis?
39. **Reflect and Apply** In fatty-acid synthesis, malonyl-CoA, rather than acetyl-CoA, is used as a “condensing group.” Suggest a reason for this.
40. **Reflect and Apply** (a) Where in an earlier chapter have we encountered something comparable to the action of the acyl carrier protein (ACP) of fatty-acid synthesis? (b) What is a critical feature of the action of the ACP?

### 21.5 Ketone Bodies

23. **Recall** Under what conditions are ketone bodies produced?
24. **Recall** Briefly outline the reactions involved in ketone production.
25. **Reflect and Apply** Why might a doctor smell the breath of a person known to have diabetes who has just passed out?
26. **Reflect and Apply** Why might a person who is an alcoholic have a “fatty liver”?
27. **Reflect and Apply** A friend who is trying to lose weight complains about the odd taste in his mouth in the mornings. He says it seems like a filling has broken loose, and the metallic sensation is bothersome. What would you say?

### 21.6 Fatty Acid Biosynthesis

28. **Recall** Compare and contrast the pathways of fatty-acid breakdown and biosynthesis. What features do these two pathways have in common? How do they differ?
29. **Recall** Outline the steps involved in the production of malonyl-CoA from acetyl-CoA.
30. **Recall** What is the metabolic importance of malonyl-CoA?
31. **Recall** In fatty-acid degradation, we encounter coenzyme A, mitochondrial matrix, *trans* double bonds, L-alcohols,  $\beta$ -oxidation,  $\text{NAD}^+$  and FAD, acetyl-CoA, and separate enzymes. What are the counterparts in fatty-acid synthesis?
32. **Recall** How are the two redox reactions of  $\beta$ -oxidation different from their counterparts in fatty-acid synthesis?
33. **Recall** How is ACP similar to coenzyme A? How is it different?

### 21.7 Synthesis of Acylglycerols and Compound Lipids

41. **Recall** What is the source of the glycerol in triacylglycerol synthesis?
42. **Recall** What is the activating group used in the formation of phosphoacylglycerols?
43. **Recall** What are the differences between synthesis of phosphatidylethanolamine in prokaryotes and eukaryotes?

### 21.8 Cholesterol Biosynthesis

44. **Recall** How are isoprene units important in cholesterol biosynthesis and other biochemical pathways?
45. **Recall** A cholesterol sample is prepared using acetyl-CoA labeled with  $^{14}\text{C}$  at the carboxyl group as precursor. Which carbon atoms of cholesterol are labeled?
46. **Recall** Which molecules have cholesterol as a precursor?
47. **Reflect and Apply** What structural feature do all steroids have in common? What are the biosynthetic implications of this common feature?
48. **Reflect and Apply** In steroid synthesis, squalene is oxidized to squalene epoxide. This reaction is somewhat unusual, in that both a reducing agent (NADPH) and an oxidizing agent ( $\text{O}_2$ ) are required. Why are both needed?
49. **Reflect and Apply** Why must cholesterol be packaged for transport rather than occurring freely in the bloodstream?
50. **Reflect and Apply** A drug that reduces blood cholesterol has the effect of stimulating the production of bile salts. How might this result in lower blood cholesterol? *Hint:* There are two ways.

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





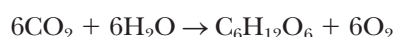
# Photosynthesis

# 22

Lush rain forest vegetation. Photosynthesis linked to oxygen plays an essential role in all plant and animal life.

## 22.1 Chloroplasts Are the Site of Photosynthesis

It is well known that photosynthetic organisms, such as green plants, convert carbon dioxide (CO<sub>2</sub>) and water to carbohydrates such as glucose (written here as C<sub>6</sub>H<sub>12</sub>O<sub>6</sub>) and molecular oxygen (O<sub>2</sub>).



The equation actually represents two processes. One process, the oxidation of water to produce oxygen (the light reactions), requires light energy from the Sun. The light reactions of photosynthesis in prokaryotes and eukaryotes depend on solar energy, which is absorbed by **chlorophyll** to supply the energy needed in the light reactions. The light reactions also generate NADPH, which is the reducing agent needed in the dark reactions. The other process, the fixation of CO<sub>2</sub> to give sugars (the dark reactions), does not use solar energy directly but rather uses it indirectly in the form of the ATP and NADPH produced in the course of the light reactions.

In prokaryotes such as cyanobacteria, photosynthesis takes place in granules bound to the plasma membrane. The site of photosynthesis in eukaryotes such as green plants and green algae is the **chloroplast** (Figure 22.1), a membrane-enclosed organelle that we discussed in Section 1.6. Like the mitochondrion, the chloroplast has inner and outer membranes and an intermembrane space. In addition, within the chloroplast are bodies called **grana**, which consist of stacks of flattened membranes called **thylakoid disks**. The grana are connected by membranes called *intergranal lamellae*. The thylakoid disks are formed by the folding of a third membrane within the chloroplast. The folding of the thylakoid membrane creates two spaces in the chloroplast in addition to the intermembrane space. The **stroma** lies within the inner membrane and outside the thylakoid membrane. In addition to the stroma, there is a **thylakoid space** within the thylakoid disks themselves. The trapping of light and the production of oxygen take place in the thylakoid disks. The dark reactions (also called light-independent reactions), in which CO<sub>2</sub> is fixed to carbohydrates, take place in the stroma (Figure 22.2).

### How does the structure of the chloroplast affect photosynthesis?

It is well established that the primary event in photosynthesis is the absorption of light by chlorophyll. The high energy states (excited states) of chlorophyll are useful in photosynthesis because the light energy can be passed along and converted to chemical energy in the light reaction. There are two principal types of chlorophyll, *chlorophyll a* and *chlorophyll b*. Eukaryotes such as green plants and green algae contain both chlorophyll *a* and chlorophyll *b*.

Prokaryotes such as cyanobacteria (formerly called blue-green algae) contain only chlorophyll *a*. Photosynthetic bacteria other than cyanobacteria have bacteriochlorophylls, with *bacteriochlorophyll a* being the most common. Organisms such as green and purple sulfur bacteria, which contain bacteriochlorophylls, do not use water as the ultimate source of electrons for the redox reactions of

## Chapter Outline

### 22.1 Chloroplasts Are the Site of Photosynthesis

- How does the structure of the chloroplast affect photosynthesis?

### 22.2 Photosystems I and II and the Light Reactions of Photosynthesis

- How does photosystem II split water to produce oxygen?
- How does photosystem I reduce NADP<sup>+</sup>?
- What is known about the structure of photosynthetic reaction centers?

### 22.3 Photosynthesis and ATP Production

- How does ATP production in chloroplasts resemble the process in mitochondria?

### 22.4 Evolutionary Implications of Photosynthesis with and without Oxygen

- Is it possible to have photosynthesis without producing oxygen?

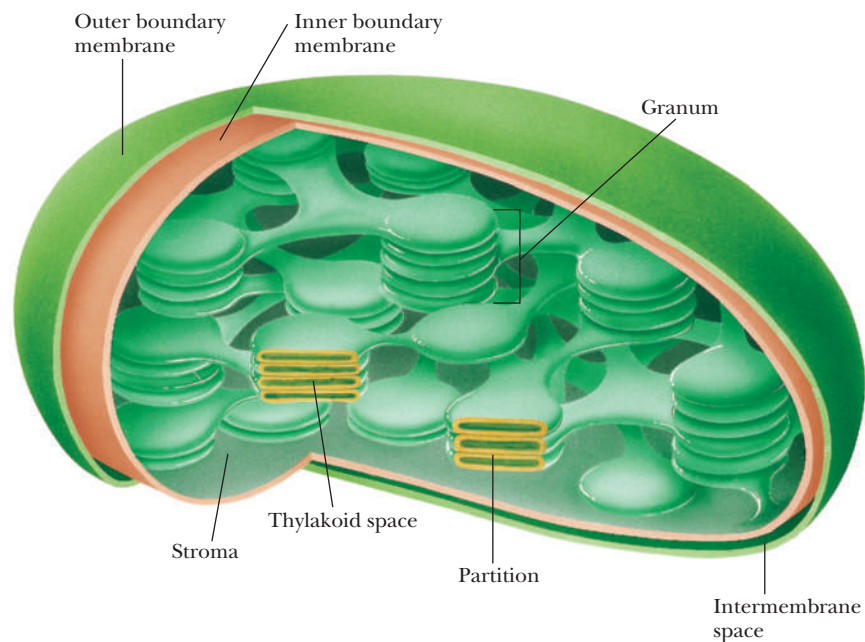
### 22.5 Dark Reactions of Photosynthesis Fix CO<sub>2</sub>

- What is the Calvin cycle?
- How is starting material regenerated in the Calvin cycle?

### 22.6 CO<sub>2</sub> Fixation in Tropical Plants

- What is different about CO<sub>2</sub> fixation in tropical plants?

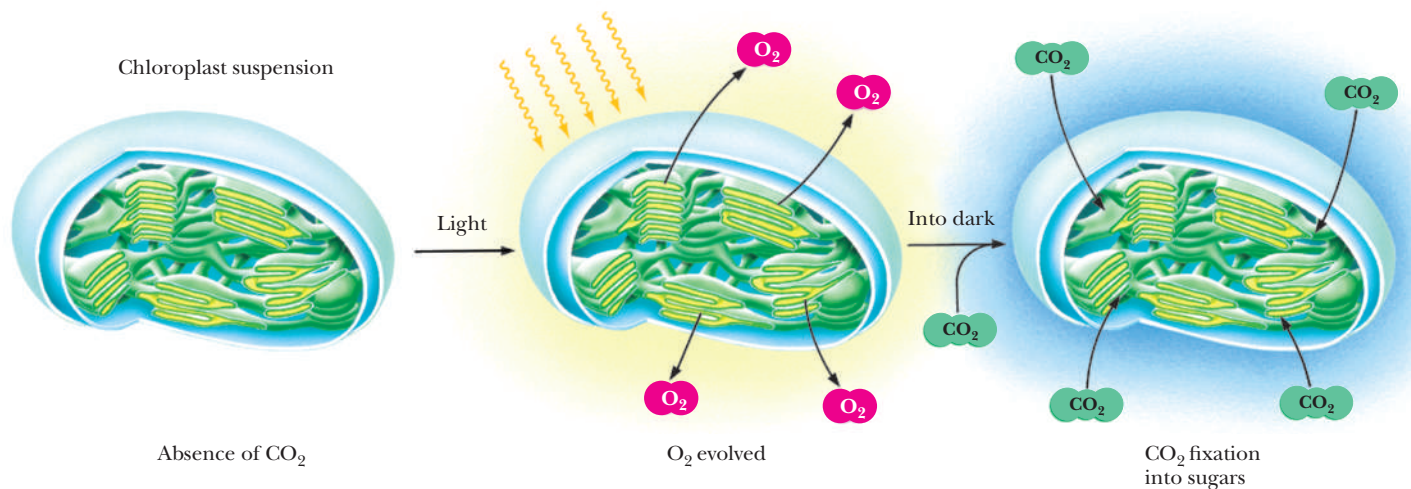
**OWL** Online homework for this chapter may be assigned in OWL.



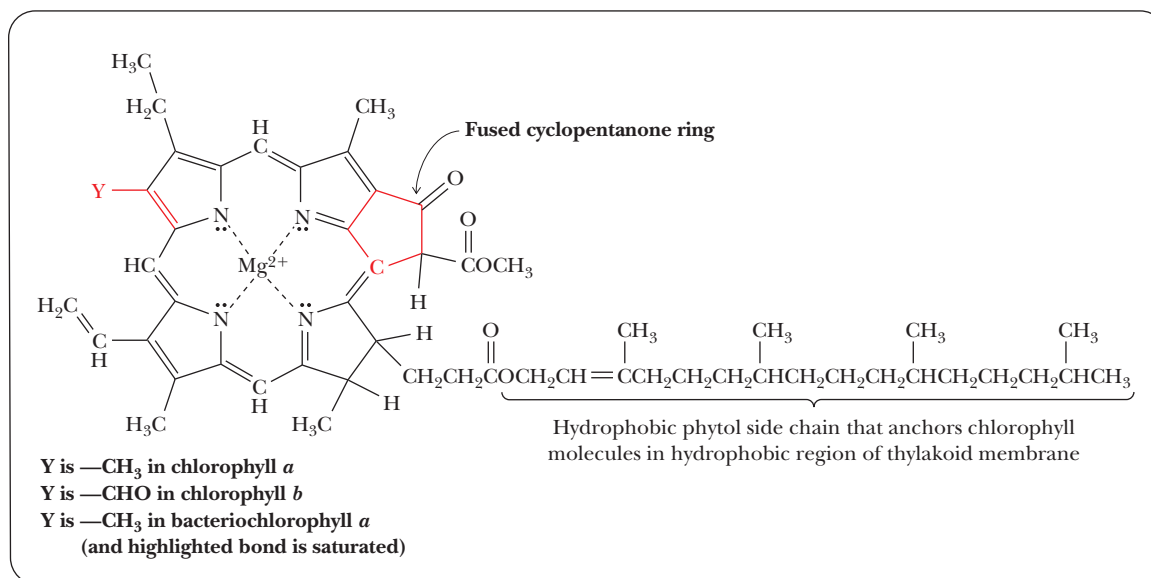
■ **FIGURE 22.1** Membrane structures in chloroplasts.

photosynthesis, nor do they produce oxygen. Instead, they use other electron sources such as  $\text{H}_2\text{S}$ , which produces elemental sulfur instead of oxygen. Organisms that contain bacteriochlorophyll are anaerobic and have only one photosystem, whereas green plants have two different photosystems, as we shall see.

The structure of chlorophyll is similar to that of the heme group of myoglobin, hemoglobin, and the cytochromes in that it is based on the tetrapyrrole ring of porphyrins (Figure 22.3). (See Section 4.5.) The metal ion bound to the tetrapyrrole ring is magnesium,  $\text{Mg(II)}$ , rather than the iron that occurs in heme. Another difference between chlorophyll and heme is the presence of a cyclopentanone ring fused to the tetrapyrrole ring. A long hydrophobic side chain, the phytol group, contains four isoprenoid units (five-carbon units that are basic building blocks in many lipids; Section 21.8) and binds to the thylakoid membrane by hydrophobic interactions. The phytol group is covalently bound to the rest of the chlorophyll molecule by an ester linkage between the alcohol



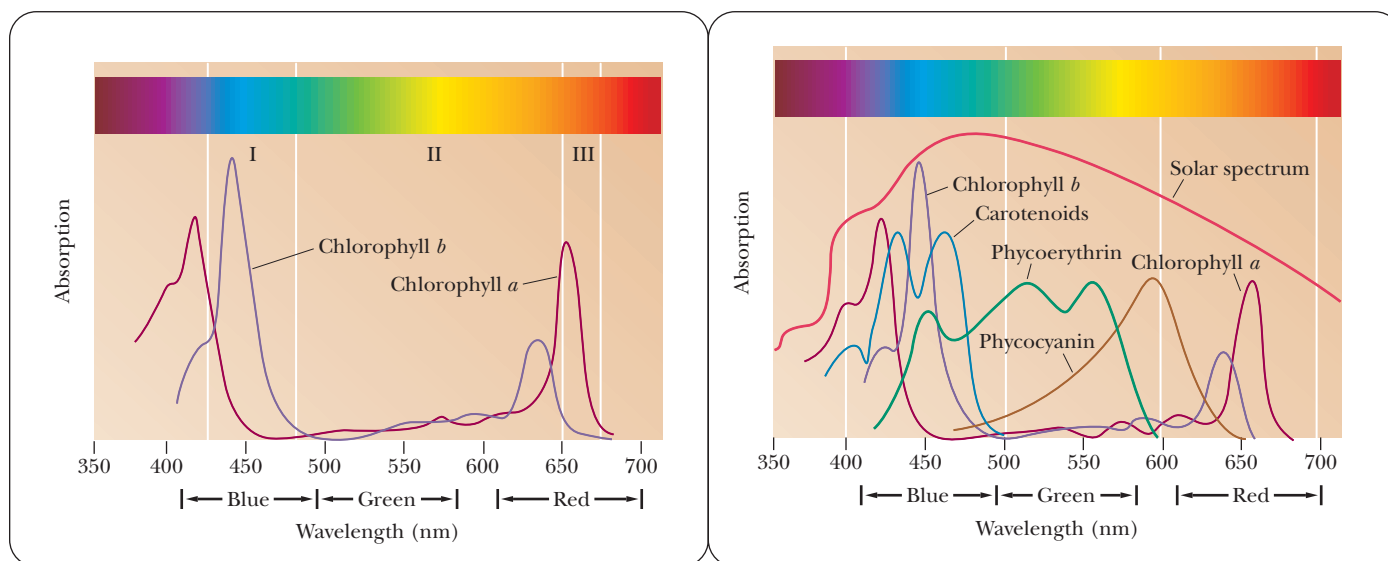
■ **FIGURE 22.2** The light-dependent and light-independent reactions of photosynthesis. Light reactions are associated with the thylakoid membranes, and light-independent reactions are associated with the stroma.



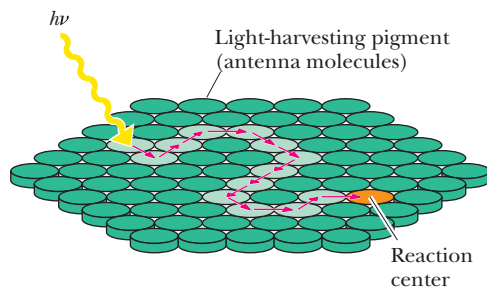
■ **FIGURE 22.3** Molecular structures of chlorophyll *a*, chlorophyll *b*, and bacteriochlorophyll *a*.

group of the phytol and a propionic acid side chain on the porphyrin ring. The difference between chlorophyll *a* and chlorophyll *b* lies in the substitution of an aldehyde group for a methyl group on the porphyrin ring. The difference between bacteriochlorophyll *a* and chlorophyll *a* is that a double bond in the porphyrin ring of chlorophyll *a* is saturated in bacteriochlorophyll *a*. The lack of a conjugated system (alternating double and single bonds) in the porphyrin ring of bacteriochlorophylls causes a significant difference in the absorption of light by bacteriochlorophyll *a* compared with chlorophyll *a* and *b*.

The absorption spectra of chlorophyll *a* and chlorophyll *b* differ slightly (Figure 22.4). Both absorb light in the red and blue portions of the visible



■ **FIGURE 22.4** Visible spectra of chlorophylls.



■ **FIGURE 22.5** Schematic diagram of a photosynthetic unit. The light-harvesting pigments, or antenna molecules (green), absorb and transfer light energy to the specialized chlorophyll dimer that constitutes the reaction center (orange).

spectrum (600 to 700 nm and 400 to 500 nm, respectively), and the presence of both types of chlorophyll guarantees that more wavelengths of the visible spectrum are absorbed than would be the case with either one individually. Recall that chlorophyll *a* is found in all photosynthetic organisms that produce oxygen. Chlorophyll *b* is found in eukaryotes such as green plants and green algae, but it occurs in smaller amounts than chlorophyll *a*. The presence of chlorophyll *b*, however, increases the portion of the visible spectrum that is absorbed and thus enhances the efficiency of photosynthesis in green plants compared with cyanobacteria. In addition to chlorophyll, various **accessory pigments** absorb light and transfer energy to the chlorophylls (Figure 22.4b). Bacteriochlorophylls, the molecular form characteristic of photosynthetic organisms that do not produce oxygen, absorb light at longer wavelengths. The wavelength of maximum absorption of bacteriochlorophyll *a* is 780 nm; other bacteriochlorophylls have absorption maxima at still longer wavelengths, such as 870 nm or 1050 nm. Light of wavelength longer than 800 nm is part of the infrared, rather than the visible, region of the spectrum. The wavelength of light absorbed plays a critical role in the light reaction of photosynthesis because the energy of light is inversely related to wavelength (see the Biochemical Connections box below).

Most of the chlorophyll molecules in a chloroplast simply gather light (antenna chlorophylls). All chlorophylls are bound to proteins, either in antenna complexes or in one of two kinds of **photosystems** (membrane-bound protein complexes that carry out the light reactions). The light-harvesting molecules then pass their excitation energy along to a specialized pair of chlorophyll molecules at a **reaction center** characteristic of each photosystem (Figure 22.5). When the light energy reaches the reaction center, the chemical reactions of photosynthesis begin. The different environments of the antenna chlorophylls

## Biochemical Connections PHYSICS

### The Relationship between Wavelength and Energy of Light

A well-known equation relates the wavelength and energy of light, a point of crucial importance for our purposes. Max Planck established in the early 20th century that the energy of light is directly proportional to its frequency.

$$E = h\nu$$

where  $E$  is energy,  $h$  is a constant (Planck's constant), and  $\nu$  is the frequency of the light. The wavelength of light is related to the frequency.

$$\nu = \frac{c}{\lambda}$$

where  $\lambda$  is wavelength,  $\nu$  is frequency, and  $c$  is the velocity of light. We can rewrite the expression for the energy of light in terms of wavelength rather than frequency.

$$E = h\nu = \frac{hc}{\lambda}$$

Light of shorter wavelength (higher frequency) is higher in energy than light of longer wavelength (lower frequency).



Highest frequency	—————→	Lowest frequency ( $\nu$ )
Highest energy	—————→	Lowest energy ( $E$ )
Shortest wavelength	—————→	Longest wavelength ( $\lambda$ )

■ **The relationships among frequency, energy, and wavelength of light.** In the visible spectrum, violet light has a shorter wavelength ( $\lambda$ ), higher frequency ( $\nu$ ), and higher energy ( $E$ ) than red light. Intermediate values of all these quantities are observed for other colors of the visible spectrum.



and the reaction-center chlorophylls give different properties to the two different kinds of molecules. In a typical chloroplast, several hundred light-harvesting antenna chlorophylls are present for each unique chlorophyll at a reaction center. The precise nature of reaction centers in both prokaryotes and eukaryotes is the subject of active research. X-ray crystallography, for example, is being used to determine the nature of conformational changes in proteins in proximity to the special pair of chlorophylls.

## 22.2 Photosystems I and II and the Light Reactions of Photosynthesis

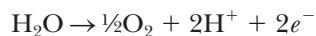
In the light reactions of photosynthesis, water is converted to oxygen by oxidation and  $\text{NADP}^+$  is reduced to NADPH. The series of redox reactions is coupled to the phosphorylation of ADP to ATP in a process called **photophosphorylation**.



The light reactions consist of two parts, accomplished by two distinct but related photosystems. One part of the reaction is the reduction of  $\text{NADP}^+$  to NADPH, carried out by **photosystem I (PSI)**. The second part of the reaction is the oxidation of water to produce oxygen, carried out by **photosystem II (PSII)**. Both photosystems carry out redox (electron transfer) reactions. The two photosystems interact with each other indirectly through an electron transport chain that links the two photosystems. The production of ATP is linked to electron transport in a process similar to that seen in the production of ATP by mitochondrial electron transport.

In the dark reactions, the ATP and NADPH produced in the light reaction provide the energy and reducing power for the fixation of  $\text{CO}_2$ . The dark reactions also constitute a redox process, since the carbon in carbohydrates is in a more reduced state than the highly oxidized carbon in  $\text{CO}_2$ . The light and dark reactions do not take place separately, but they are separated for purposes of discussion only.

The net electron transport reaction of the two photosystems taken together is, except for the substitution of NADPH for NADH, the reverse of mitochondrial electron transport. The half-reaction of reduction is that of  $\text{NADP}^+$  to NADPH, whereas the half-reaction of oxidation is that of water to oxygen.



This is an endergonic reaction with a positive  $\Delta G^\circ = +220 \text{ kJ mol}^{-1} = +52.6 \text{ kcal mol}^{-1}$ . The light energy absorbed by the chlorophylls in both photosystems provides the energy that allows this endergonic reaction to take place. A series of electron carriers embedded in the thylakoid membrane link these reactions. The electron carriers have an organization very similar to that of the carriers in the electron transport chain.

Photosystem I can be excited by light of wavelengths shorter than 700 nm, but photosystem II requires light of wavelengths shorter than 680 nm for excitation. Both photosystems must operate for the chloroplast to produce NADPH, ATP, and  $\text{O}_2$ , because the two photosystems are connected by the electron transport chain. The two systems are, however, structurally distinct in the chloroplast; photosystem I can be released preferentially from the thylakoid membrane by treatment with detergents. The reaction centers of the two photosystems

provide different environments for the unique chlorophylls involved. The unique chlorophyll of photosystem I is referred to as  $P_{700}$ , where P is for pigment and the subscript 700 is for the longest wavelength of absorbed light (700 nm) that initiates the reaction. Similarly, the reaction-center chlorophyll of photosystem II is designated  $P_{680}$  because the longest wavelength of absorbed light that initiates the reaction is 680 nm. Note particularly that the path of electrons starts with the reactions in photosystem II rather than in photosystem I. The reason for the nomenclature is that photosystem I was studied extensively at an earlier date than photosystem II because it is easier to extract photosystem I from the thylakoid membrane than it is to extract photosystem II. There are two places in the reaction scheme of the two photosystems where the absorption of light supplies energy to make endergonic reactions take place (Figure 22.6).

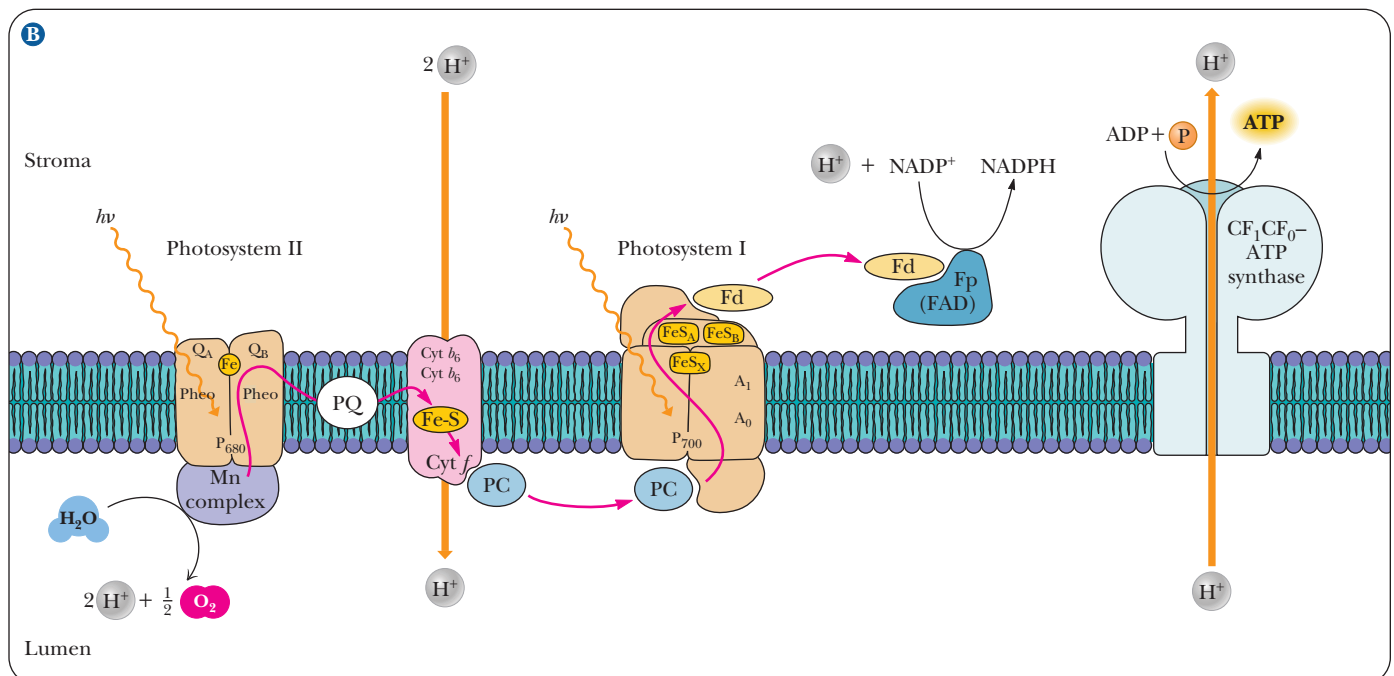
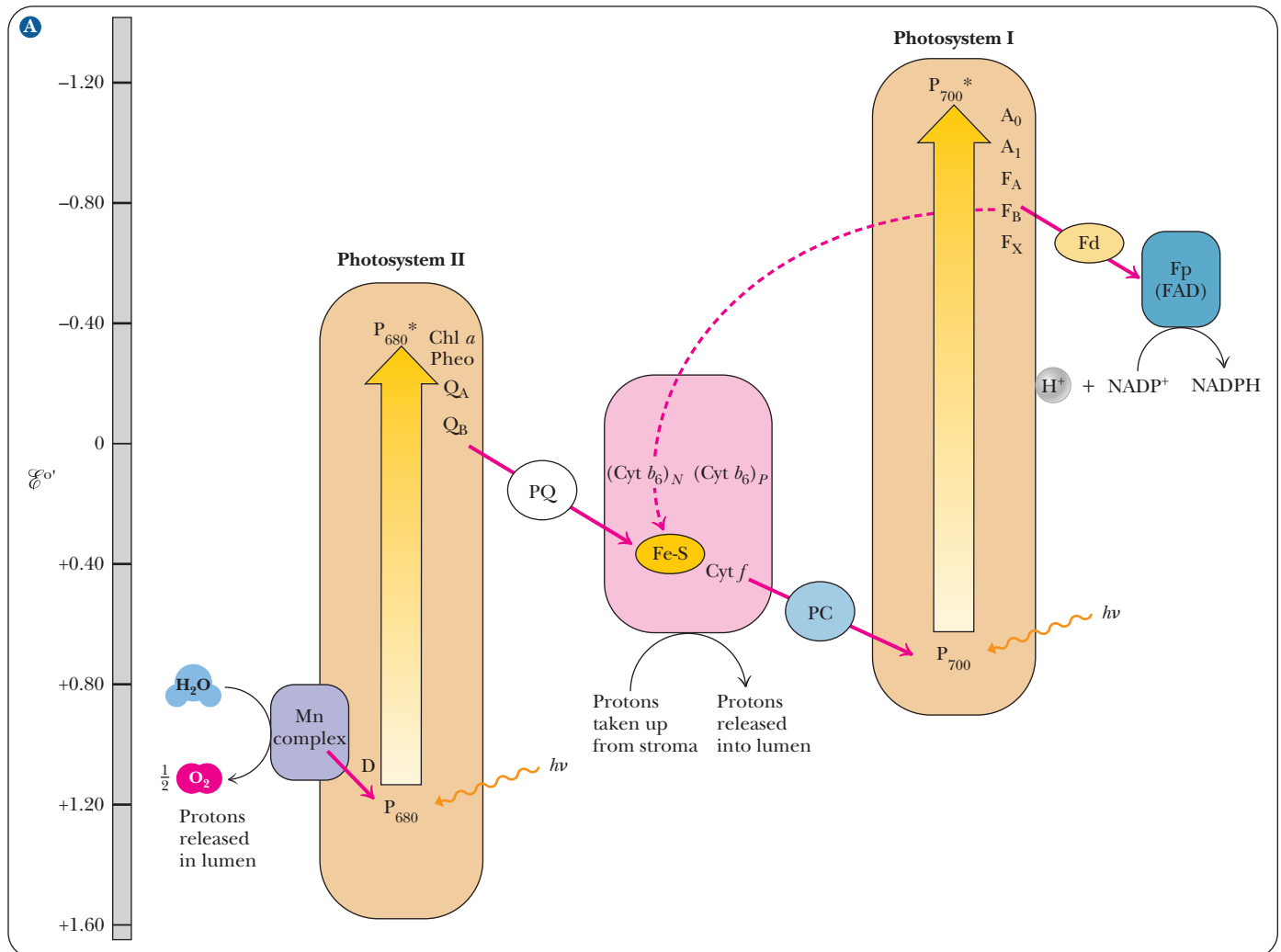
Neither reaction-center chlorophyll is a strong enough reducing agent to pass electrons to the next substance in the reaction sequence, but the absorption of light by the chlorophylls of both photosystems provides enough energy for such reactions to take place. The absorption of light by Chl ( $P_{680}$ ) allows electrons to be passed to the electron transport chain that links photosystem II and photosystem I and generates an oxidizing agent that is strong enough to split water, producing oxygen. When Chl ( $P_{700}$ ) absorbs light, enough energy is provided to allow the ultimate reduction of  $NADP^+$  to take place. (Note that the energy difference is shown on the vertical axis of Figure 22.6. This type of diagram is also called a *Z scheme*. The Z is rather lopsided and lies on its side, but the name is common.) In both photosystems, the result of supplying energy (light) is analogous to pumping water uphill.

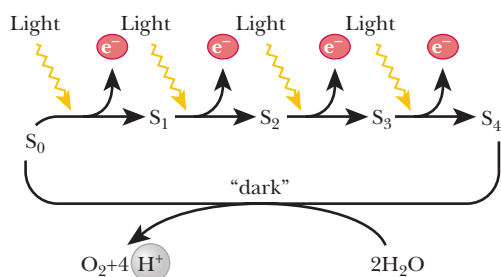
### How does photosystem II split water to produce oxygen?

The oxidation of water by photosystem II to produce oxygen is the ultimate source of electrons in photosynthesis. These electrons are subsequently passed from photosystem II to photosystem I by the electron transport chain. The electrons from water are needed to “fill the hole” that is left when the absorption of one photon of light leads to donation of an electron from photosystem II to the electron transport chain.

The electrons released by the oxidation of water are first passed to  $P_{680}$ , which is reduced. There are intermediate steps in this reaction because four electrons are required for the oxidation of water, and  $P_{680}^*$  can accept only one electron at a time. A manganese-containing protein complex and several other protein components are required. The **oxygen-evolving complex** of photosystem II passes through a series of five oxidation states (designated

■ **FIGURE 22.6 The Z scheme of photosynthesis.** (a) The Z scheme is a diagrammatic representation of photosynthetic electron flow from  $H_2O$  to  $NADP^+$ . The energy relationships can be derived from the  $E^\circ$  scale beside the Z diagram, with lower standard potentials and therefore greater energy as you go from bottom to top. Energy input as light is indicated by two broad arrows, one photon appearing in  $P_{680}$  and the other in  $P_{700}$ .  $P_{680}^*$  and  $P_{700}^*$  represent excited states. Electron loss from  $P_{680}^*$  and  $P_{700}^*$  creates  $P_{680}$  and  $P_{700}$ . The representative components of the three supramolecular complexes (PSI, PSII, and the cytochrome  $b_6-f$  complex) are in shaded boxes enclosed by solid black lines. A number of components are represented by letters of the alphabet—chlorophylls and quinones by A and Q, respectively, and ferredoxins by F—and are further distinguished by subscripts. Proton translocations that establish the proton-motive force driving ATP synthesis are illustrated as well. (b) Figure showing the functional relationships among PSII, the cytochrome  $b_6-f$  complex, PSI, and the photosynthetic  $CF_1CF_0$ —ATP synthase within the thylakoid membrane. Note that  $e^-$  acceptors  $Q_A$  (for PSII) and  $A_1$  (for PSI) are at the stromal side of the thylakoid membrane, whereas the  $e^-$  donors to  $P_{680}$  and  $P_{700}$  are situated at the luminal side of the membrane. The consequence is charge separation (stroma, lumen) across the membrane. Also note that protons are translocated into the thylakoid lumen, giving rise to a chemiosmotic gradient that is the driving force for ATP synthesis by  $CF_1CF_0$ —ATP synthase.





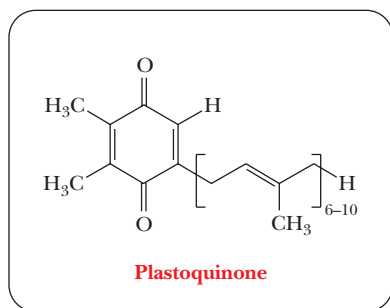
■ **FIGURE 22.7** The PSII reaction center passes through five different oxidation states,  $S_0$  through  $S_4$ , in the course of oxygen evolution.

as  $S_0$  through  $S_4$ ) in the transfer of four electrons in the process of evolving oxygen (Figure 22.7). One electron is passed from water to PSII for each quantum of light. In the process, the components of the reaction center go successively through oxidation states  $S_1$  through  $S_4$ . The  $S_4$  decays spontaneously to the  $S_0$  state and, in the process, oxidizes two water molecules to one oxygen molecule. Note that four protons are released simultaneously. The immediate electron donor to the  $P_{680}$  chlorophyll, shown as D in Figure 22.6, is a tyrosine residue of one of the protein components that does not contain manganese. Several quinones serve as intermediate electron transfer agents to accommodate four electrons donated by one water molecule. Redox reactions of manganese also play a role here. Even this mechanism is an oversimplification. Attempts to observe the direct production of oxygen by the  $S_4$  state imply that some intermediate ( $S_4'$ ) directly produces oxygen after deprotonation of  $S_3$  and loss of an electron by  $S_4$ . The main point is that the oxygen-evolving complex is very complex indeed.

In photosystem II, as in photosystem I, the absorption of light by chlorophyll in the reaction center produces an excited state of chlorophyll. The wavelength of light is 680 nm; the reaction-center chlorophyll of photosystem II is also referred to as  $P_{680}$ . The excited chlorophyll passes an electron to a primary acceptor. In photosystem II, the primary electron acceptor is a molecule of **pheophytin** (Pheo), one of the accessory pigments of the photosynthetic apparatus. The structure of pheophytin differs from that of chlorophyll only in the substitution of two hydrogens for the magnesium. The transfer of electrons is mediated by events that take place at the reaction center. The next electron acceptor is **plastoquinone** (PQ). The structure of plastoquinone (Figure 22.8) is similar to that of coenzyme Q (ubiquinone), a part of the respiratory electron transport chain (Section 20.2), and plastoquinone serves a very similar purpose in the transfer of electrons and hydrogen ions.

The electron transport chain that links the two photosystems consists of pheophytin, plastoquinone, a complex of plant cytochromes (the  $b_6-f$  complex), a copper-containing protein called **plastocyanin** (PC), and the oxidized form of  $P_{700}$  (see Figure 22.6). The  $b_6-f$  complex of plant cytochromes consists of two  $b$ -type cytochromes (cytochrome  $b_6$ ) and a  $f$ -type cytochrome (cytochrome  $f$ ). This complex is similar in structure to the  $b_6$  complex in mitochondria and occupies a similar central position in an electron transport chain. This part of the photosynthetic apparatus is the subject of active research. There is a possibility that a Q cycle (recall this from Section 20.2) may operate here as well, and the object of some of this research is to establish definitely whether this is so. In plastocyanin, the copper ion is the actual electron carrier; the copper ion exists as Cu(II) and Cu(I) in the oxidized and reduced forms, respectively. This electron transport chain has another similarity to that in mitochondria, that of coupling to ATP generation.

When the oxidized chlorophyll of  $P_{700}$  accepts electrons from the electron transport chain, it is reduced and subsequently passes an electron to photosystem I, which absorbs a second photon of light. Absorption of light by photosystem II does not raise the electrons to a high enough energy level to reduce  $NADP^+$ ; the second photon absorbed by photosystem I provides the needed energy. This difference in energy makes the Z of the Z scheme thoroughly lopsided, but the transfer of electrons is complete.

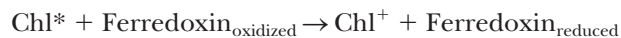


■ **FIGURE 22.8** The structure of plastoquinone. The length of the aliphatic side chain varies in different organisms.

### How does photosystem I reduce $NADP^+$ ?

The absorption of light by  $P_{700}$  then leads to the series of electron transfer reactions of photosystem I. The substance to which the excited-state chlorophyll,  $P_{700}^*$ , gives an electron is apparently a molecule of chlorophyll  $a$ ; this transfer of electrons is mediated by processes that take place in the reaction center.

The next electron acceptor in the series is bound ferredoxin, an iron–sulfur protein occurring in the membrane in photosystem I. The bound ferredoxin passes its electron to a molecule of soluble ferredoxin. Soluble ferredoxin in turn reduces an FAD-containing enzyme called ferredoxin-NADP<sup>+</sup> reductase. The FAD portion of the enzyme reduces NADP<sup>+</sup> to NADPH (Figure 22.6). We can summarize the main features of the process in two equations, in which the notation ferredoxin refers to the soluble form of the protein.



Ferredoxin-NADP  
reductase



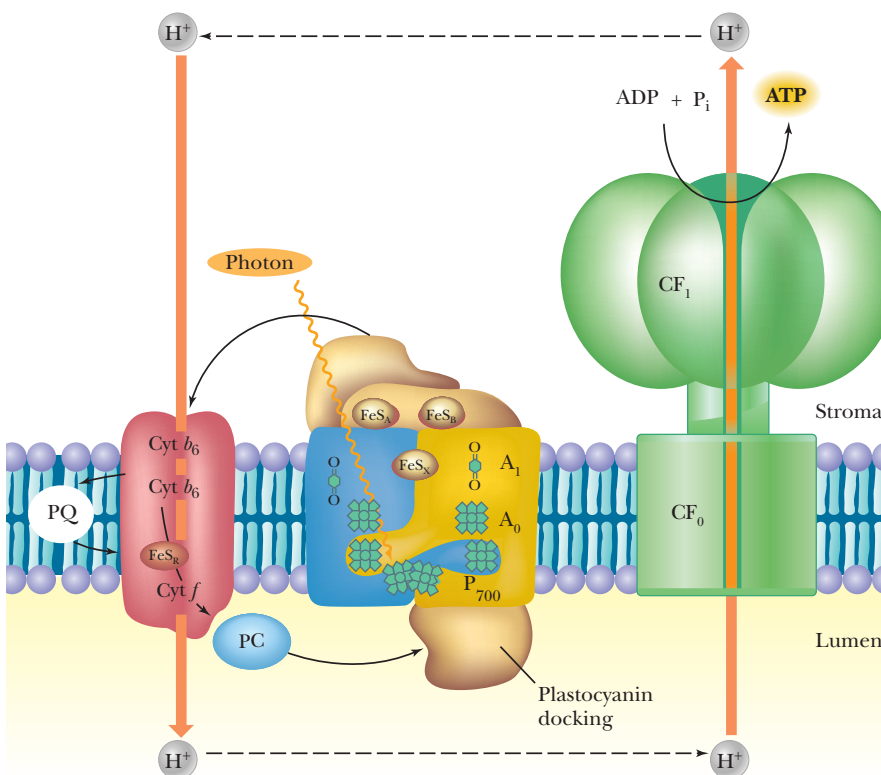
Chl\* donates one electron to ferredoxin, but the electron transfer reactions of FAD and NADP<sup>+</sup> involve two electrons. Thus, an electron from each of two ferredoxins is required for the production of NADPH.

The net reaction for the two photosystems together is the flow of electrons from H<sub>2</sub>O to NADP<sup>+</sup> (see Figure 22.6).



### Cyclic Electron Transport in Photosystem I

In addition to the electron transfer reactions just described, it is possible for cyclic electron transport in photosystem I to be coupled to the production of ATP (Figure 22.9). No NADPH is produced in this process. Photosystem II is not involved, and no O<sub>2</sub> is generated. Cyclic phosphorylation takes place when there is a high NADPH/NADP<sup>+</sup> ratio in the cell: not enough NADP<sup>+</sup> is present in the cell to accept all the electrons generated by the excitation of P<sub>700</sub>.



**FIGURE 22.9** The pathway of cyclic photophosphorylation by PSI. Note that water is not split and that no NADPH is produced. (Adapted from Arnon, D. I., 1984. *The Discovery of Photosynthetic Phosphorylation*. Trends Biochem. Sci. 9, 258–262.)



### What is known about the structure of photosynthetic reaction centers?

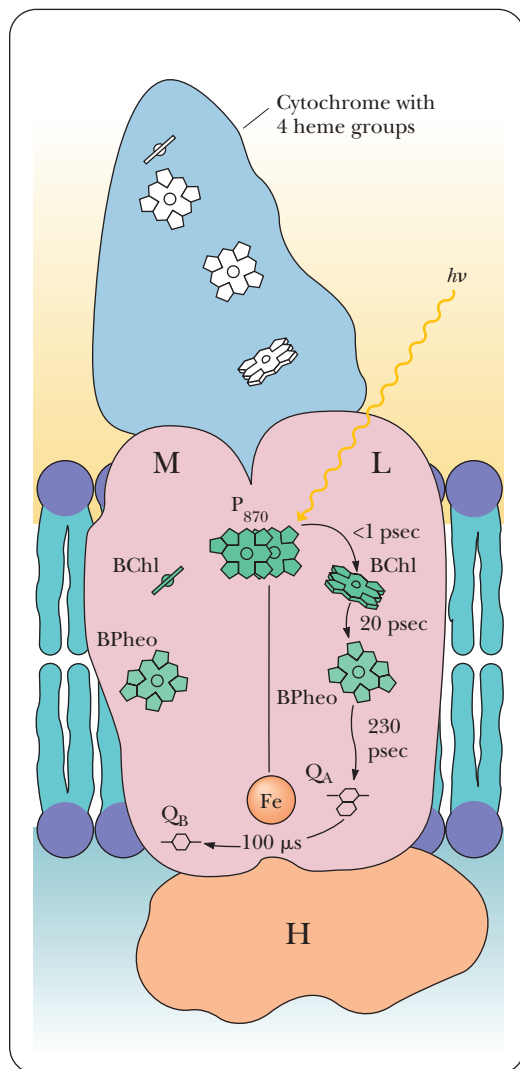
The molecular structure of photosystems is a subject of intense interest to biochemists. The most extensively studied system is that from anaerobic phototrophic bacteria of the genus *Rhodospseudomonas*. These bacteria do not produce molecular oxygen as a result of their photosynthetic activities, but enough similarities exist between the photosynthetic reactions of *Rhodospseudomonas* and photosynthesis linked to oxygen to lead scientists to draw conclusions about the nature of reaction centers in all organisms. Since the structure of this photosystem was elucidated by X-ray crystallography, the structures of PSI and PSII have also been determined and have been shown to be markedly similar. Consequently, the detailed process that goes on at the reaction center of *Rhodospseudomonas* is important enough to warrant further discussion.

It is well established that there is a pair of bacteriochlorophyll molecules (designated  $P_{870}$  from the fact that light of 870 nm is the maximum excitation wavelength) in the reaction center of *Rhodospseudomonas viridis*; the critical pair of chlorophylls is embedded in a protein complex that is in turn an integral part of the photosynthetic membrane. (We shall refer to the bacteriochlorophylls simply as chlorophylls in the interest of simplifying the discussion.)

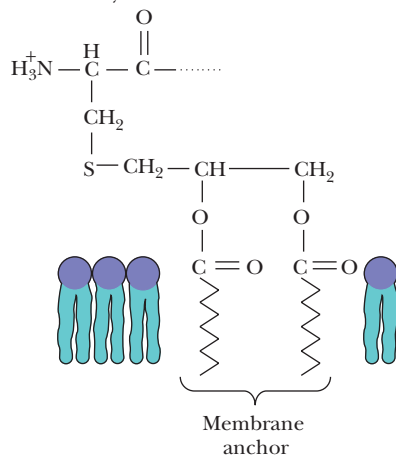
Accessory pigments, which also play a role in the light-trapping process, have specific positions close to the special pair of chlorophylls. The absorption of light by the special pair of chlorophylls raises one of their electrons to a higher energy level (Figure 22.10). This electron is passed to a series of accessory pigments. The first of these accessory pigments is pheophytin, which is structurally similar to chlorophyll, differing only in having two hydrogens in place of the magnesium. The electron is passed along to the pheophytin, raising it in turn to an excited energy level. (Note that the electron travels on only one of two possible paths, to one pheophytin but not the other. Research is in progress to determine why this is so.) The next electron acceptor is menaquinone (QA); it is structurally similar to coenzyme Q, which plays a role in the mitochondrial electron transport chain (Figure 22.11). The final electron acceptor, which is also raised to an excited state, is coenzyme Q itself (ubiquinone, called QB here). The electron that had been passed to QB is replaced by an electron donated by a cytochrome, which acquires a positive charge in the process. The cytochrome is not bound to the membrane and diffuses away, carrying its positive charge with it. The whole process takes place in less than  $10^{-3}$  s. The positive and negative charges have traveled in opposite directions from the chlorophyll pair and are separated from each other. This situation is similar to that of the proton gradient in mitochondria, where the existence of the proton gradient is ultimately responsible for oxidative phosphorylation.

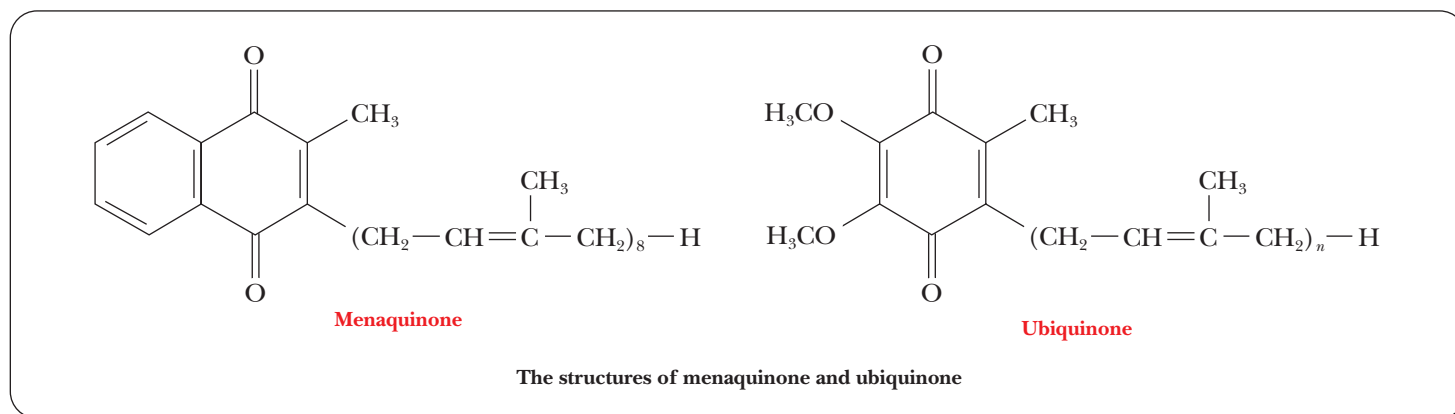
The separation of charge is equivalent to a battery, a form of stored energy. The reaction center has acted as a transducer, converting light energy to a form usable by the cell to carry out the energy-requiring reactions of photosynthesis. The processes that take place in *Rhodospseudomonas* serve as a model for reaction centers in photosynthesis linked to oxygen.

■ **FIGURE 22.10** Model of the structure and activity of the *Rhodospseudomonas viridis* reaction center. Four polypeptides (designated cytochrome, M, L, and H) make up the reaction center, an integral membrane complex. The cytochrome maintains its association with the membrane via a diacylglycerol group linked to its N-terminal Cys residue by a thioether bond. M and L both consist of five membrane-spanning  $\alpha$ -helices; H has a single membrane-spanning  $\alpha$ -helix. The prosthetic groups are spatially situated so that rapid  $e^-$  transfer from  $P_{870}^*$  to QB is facilitated. Photoexcitation of  $P_{870}$  leads in less than 1 picosecond (psec) to reduction of the L-branch BChl only.  $P_{870}$  is re-reduced via an electron provided through the heme groups of the cytochrome.



Note: The cytochrome subunit is membrane associated via a diacylglycerol moiety on its N-terminal Cys residue:





■ **FIGURE 22.11** The structures of menaquinone and ubiquinone.

## 22.3 Photosynthesis and ATP Production

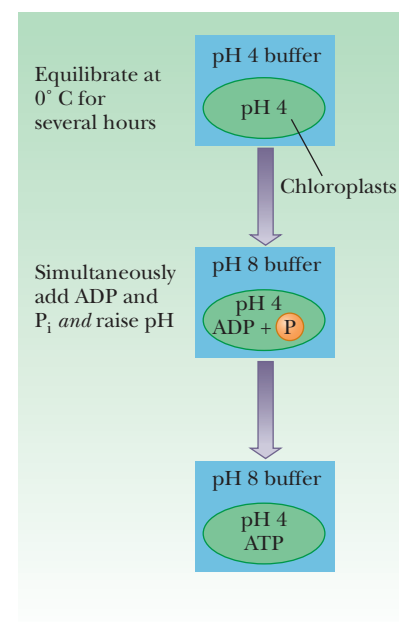
In Chapter 20, we saw that a proton gradient across the inner mitochondrial membrane drives the phosphorylation of ADP in respiration. The mechanism of photophosphorylation is essentially the same as that of the production of ATP in the respiratory electron transport chain. In fact, some of the strongest evidence for the chemiosmotic coupling of phosphorylation to electron transport has been obtained from experiments on chloroplasts rather than mitochondria. Chloroplasts can synthesize ATP from ADP and  $P_i$  *in the dark* if they are provided with a pH gradient.

If isolated chloroplasts are allowed to equilibrate in a pH 4 buffer for several hours, their internal pH will be equal to 4. If the pH of the buffer is raised rapidly to 8 and if ADP and  $P_i$  are added simultaneously, ATP will be produced (Figure 22.12). The production of ATP does not require the presence of light; the proton gradient produced by the pH difference supplies the driving force for phosphorylation. This experiment provides solid evidence for the chemiosmotic coupling mechanism.

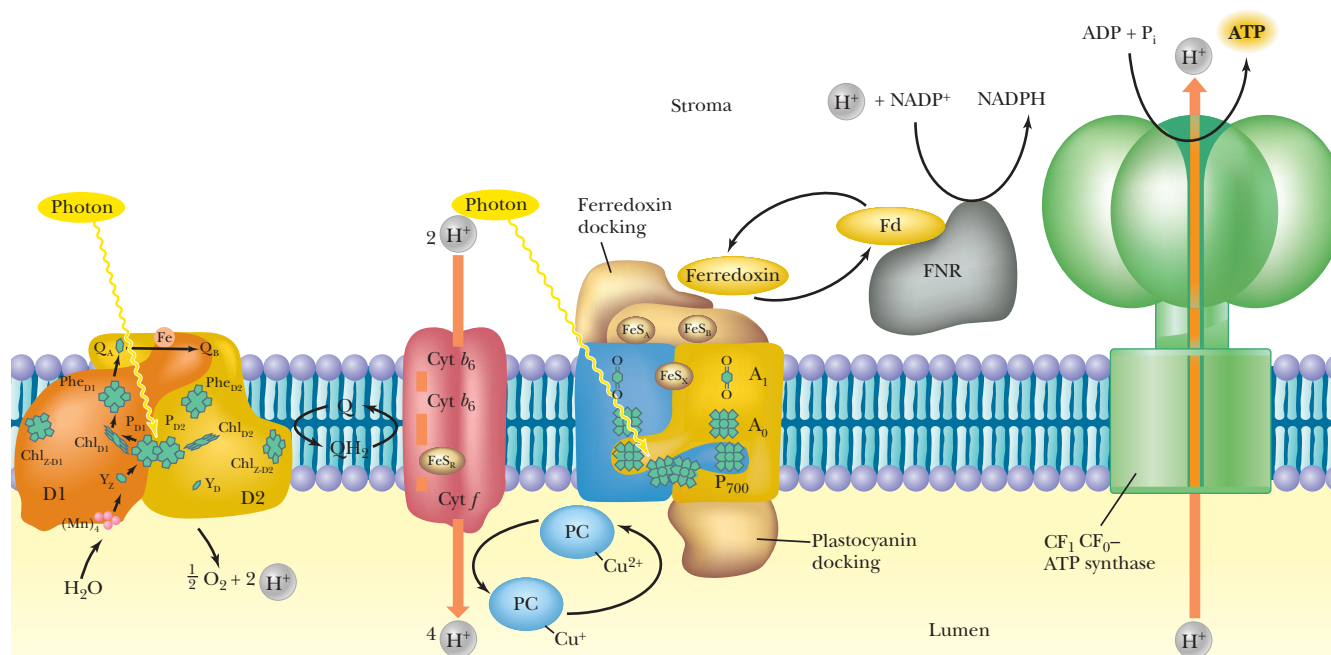
### How does ATP production in chloroplasts resemble the process in mitochondria?

Several reactions contribute to the generation of a proton gradient in chloroplasts in an actively photosynthesizing cell. The oxidation of water releases  $H^+$  into the thylakoid space. Electron transport from photosystem II and photosystem I also helps create the proton gradient by involving plastoquinone and cytochromes in the process. Then photosystem I reduces  $NADP^+$  by using  $H^+$  in the stroma to produce NADPH. As a result, the pH of the thylakoid space is lower than that of the stroma (Figure 22.13). We saw a similar situation in Chapter 20 when we discussed the pumping of protons from the mitochondrial matrix into the intermembrane space. The ATP synthase in chloroplasts is similar to the mitochondrial enzyme; in particular, it consists of two parts,  $CF_1$  and  $CF_0$ , where the C serves to distinguish them from their mitochondrial counterparts,  $F_1$  and  $F_0$ , respectively. Evidence exists that the components of the electron chain in chloroplasts are arranged asymmetrically in the thylakoid membrane, as is the case in mitochondria. An important consequence of this asymmetrical arrangement is the release of the ATP and NADPH produced by the light reaction into the stroma, where they provide energy and reducing power for the dark reaction of photosynthesis.

In mitochondrial electron transport, four respiratory complexes are connected by soluble electron carriers. The electron transport apparatus of the



■ **FIGURE 22.12** ATP is synthesized by chloroplasts in the dark in the presence of a proton gradient, ADP, and  $P_i$ .



■ **FIGURE 22.13** The mechanism of photophosphorylation. Photosynthetic electron transport establishes a proton gradient that is tapped by the  $\text{CF}_1\text{CF}_0$ -ATP synthase to drive ATP synthesis. Critical to this mechanism is the fact that the membrane-bound components of light-induced electron transport and ATP synthesis are asymmetric with respect to the thylakoid membrane so that directional discharge and uptake of  $\text{H}^+$  ensue, generating the proton-motive force. The number of protons pumped through the ATP synthase varies by species and is the subject of active research.

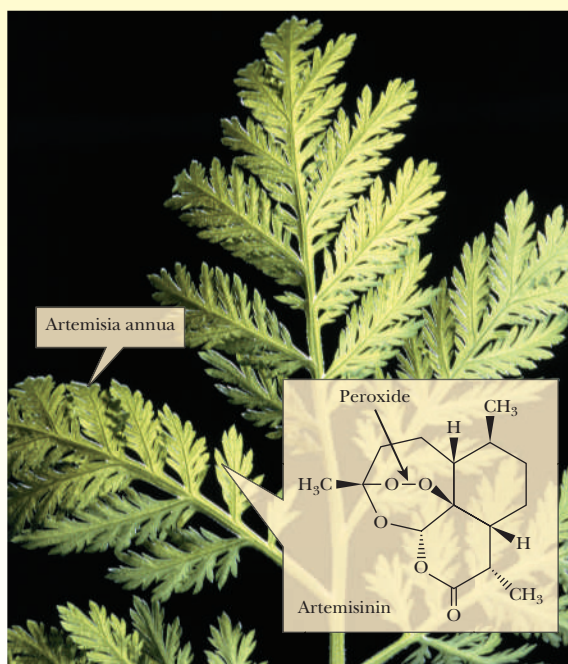
## Biochemical Connections APPLIED GENETICS

### Improving the Yield of Anti-Malarial Plants

A plant called *Artemisia annua* has long been used in Chinese folk medicine. As is the case with many folk remedies, this plant contains a pharmacologically active agent. This compound, which has been named artemisinin, contains a peroxide group as part of a ring system (see figure). It is highly effective in treating malaria, which affects 300 to 500 million people worldwide each year, causing approximately a million deaths. Unfortunately, artemisinin is in short supply.

The genetic map of this plant has recently been determined. In this process, it was possible to identify loci that can be used to improve yields significantly. Efforts have been made to synthesize artemisinin and related compounds in the laboratory. Other attempts have been made to produce it in yeast and bacteria, using recombinant DNA technology. None of these approaches has been particularly successful. It appears that the best way to ensure an adequate supply of this drug is to increase plant yields.

■ Structure of the antimalarial compound artemisinin, shown with the plant *Artemisia annua*, from which it is extracted. The genetic map of the plant is known, including loci that will allow for development of high-yielding plants. (From *The Genetic Map of Artemisia annua L. Identifies Loci Affecting Yield of the Antimalarial Drug Artemisinin* by Ian A. Graham, et al. (15 January 2010) Science 327 (5963), 328. Used with permission of AAAS.)



Scott Bauer/Photo Researchers, Inc.

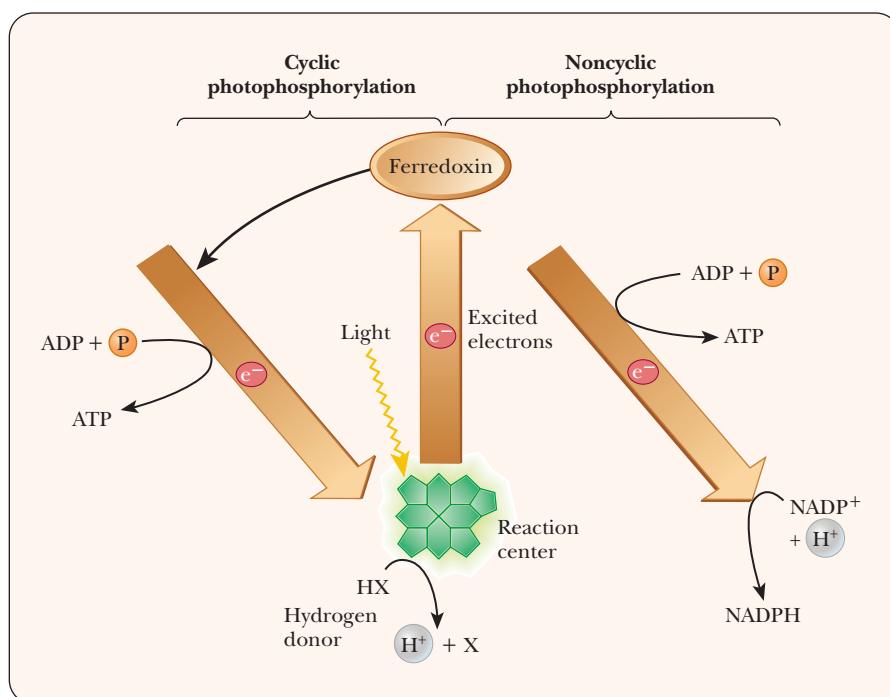
thylakoid membrane is similar in that it consists of several large membrane-bound complexes. They are PSII (the photosystem II complex), the cytochrome  $b_6-f$  complex, and PSI (the photosystem I complex). As in mitochondrial electron transport, several soluble electron carriers form the connection between the protein complexes. In the thylakoid membrane, the soluble carriers are plastoquinone and plastocyanin, which have a role similar to that of coenzyme Q and cytochrome  $c$  in mitochondria (Figure 22.13). The proton gradient created by electron transport drives the synthesis of ATP in chloroplasts, as in mitochondria.

## 22.4 Evolutionary Implications of Photosynthesis with and without Oxygen

Photosynthetic prokaryotes other than cyanobacteria have only one photosystem and do not produce oxygen. The chlorophyll in these organisms is different from that found in photosystems linked to oxygen (Figure 22.14). Anaerobic photosynthesis is not as efficient as photosynthesis linked to oxygen, but the anaerobic version of the process appears to be an evolutionary way station. Anaerobic photosynthesis is a means for organisms to use solar energy to satisfy their needs for food and energy. Although it is efficient in the production of ATP, its efficiency is less than that of aerobic photosynthesis for carbon fixation.

### Is it possible to have photosynthesis without producing oxygen?

A possible scenario for the development of photosynthesis starts with heterotrophic bacteria that contain some form of chlorophyll, probably bacteriochlorophyll. (*Heterotrophs* are organisms that depend on their environment for organic nutrients and for energy.) In such organisms, the light energy absorbed by chlorophyll can be trapped in the forms of ATP and NADPH. The important point about such a series of reactions is that photophosphorylation takes place, ensuring an independent supply of ATP for the organism. In addition, the supply of NADPH facilitates synthesis of biomolecules from simple sources



**FIGURE 22.14** The two possible electron transfer pathways in a photosynthetic anaerobe. Both cyclic and noncyclic forms of photophosphorylation are shown. HX is any compound (such as  $H_2S$ ) that can be a hydrogen donor. (From L. Margulis, 1985. *Early Life*, Science Books International, Boston, p. 45.)



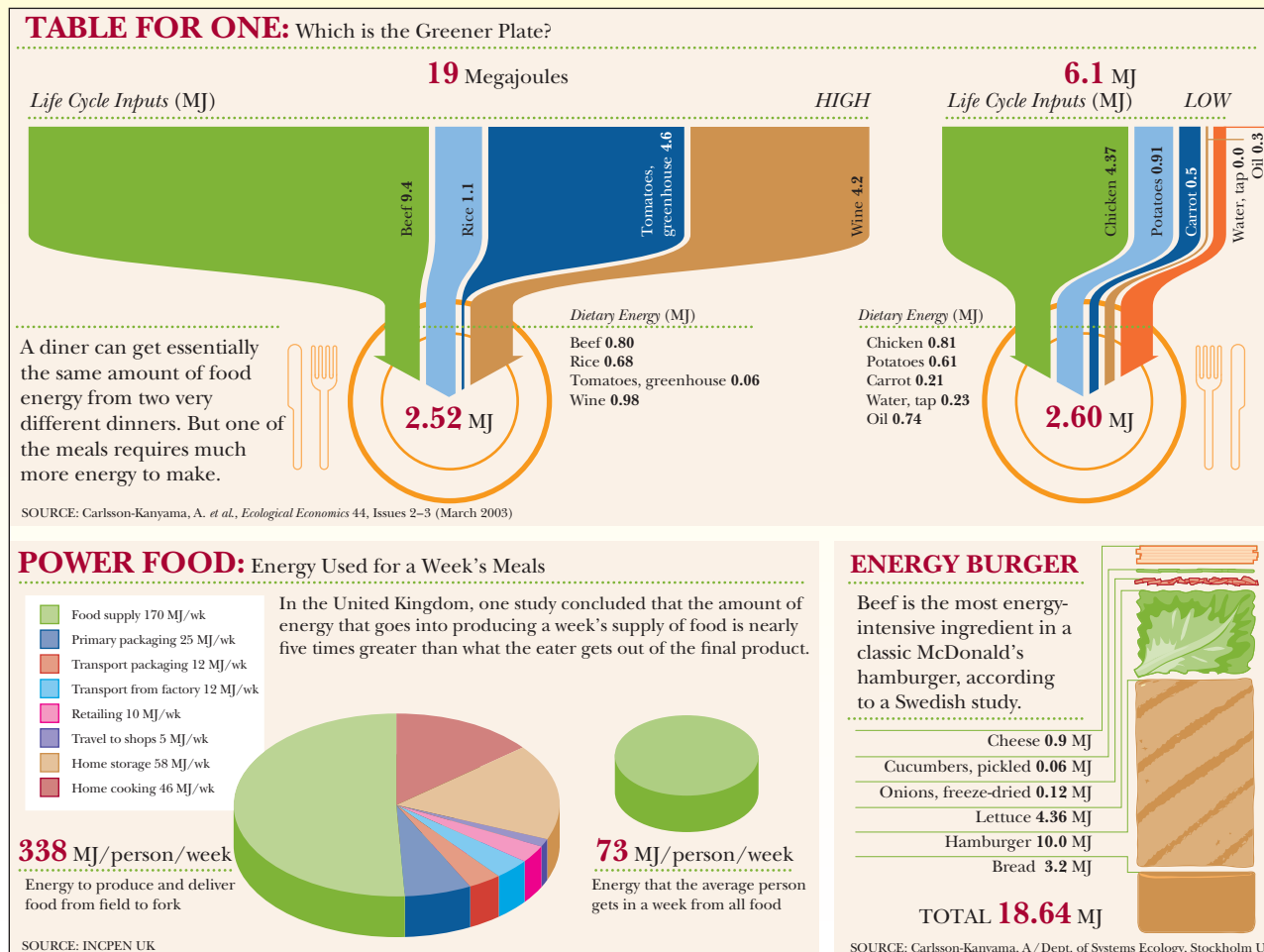
# Biochemical Connections AGRICULTURE

## Plants Feed Animals—Plants Need Energy—Plants Can Produce Energy

All living things on Earth depend on photosynthetic plants for food. This topic is so important that the journal *Science* devoted a special section in its 12 February 2010 issue to the topic of food security. A number of points in this section relate directly to biochemistry. Genetic engineering in particular can address desirable goals. These goals include improving the nutrient content

of seeds and edible plant parts, as well as increasing drought resistance. Strengthening defenses against plant pests also makes use of added genes. If plants can be engineered to carry out nitrogen fixation, that modification can lead to decreased use of fertilizers.

The energy required to grow food crops, especially those used for livestock feed, is another important consideration.



■ The results of a Swedish study on energy requirements for producing food. (From *Science*, vol. 327, p. 809, 2010.)

such as  $\text{CO}_2$ . Under conditions of limited food supply, organisms that can synthesize their own nutrients have a selective advantage.

Organisms of this sort are *autotrophs* (not dependent on an external source of biomolecules) but are also anaerobes. The ultimate electron source that they use is not water but some more easily oxidized substance, such as  $\text{H}_2\text{S}$ , as is the case with present-day green sulfur bacteria (and purple sulfur bacteria), or various organic compounds, as is the case with present-day purple nonsulfur bacteria. These organisms do not possess an oxidizing agent powerful enough to split water, which is a far more abundant electron source than  $\text{H}_2\text{S}$  or organic compounds. The ability to use water as an electron source confers a further evolutionary advantage.



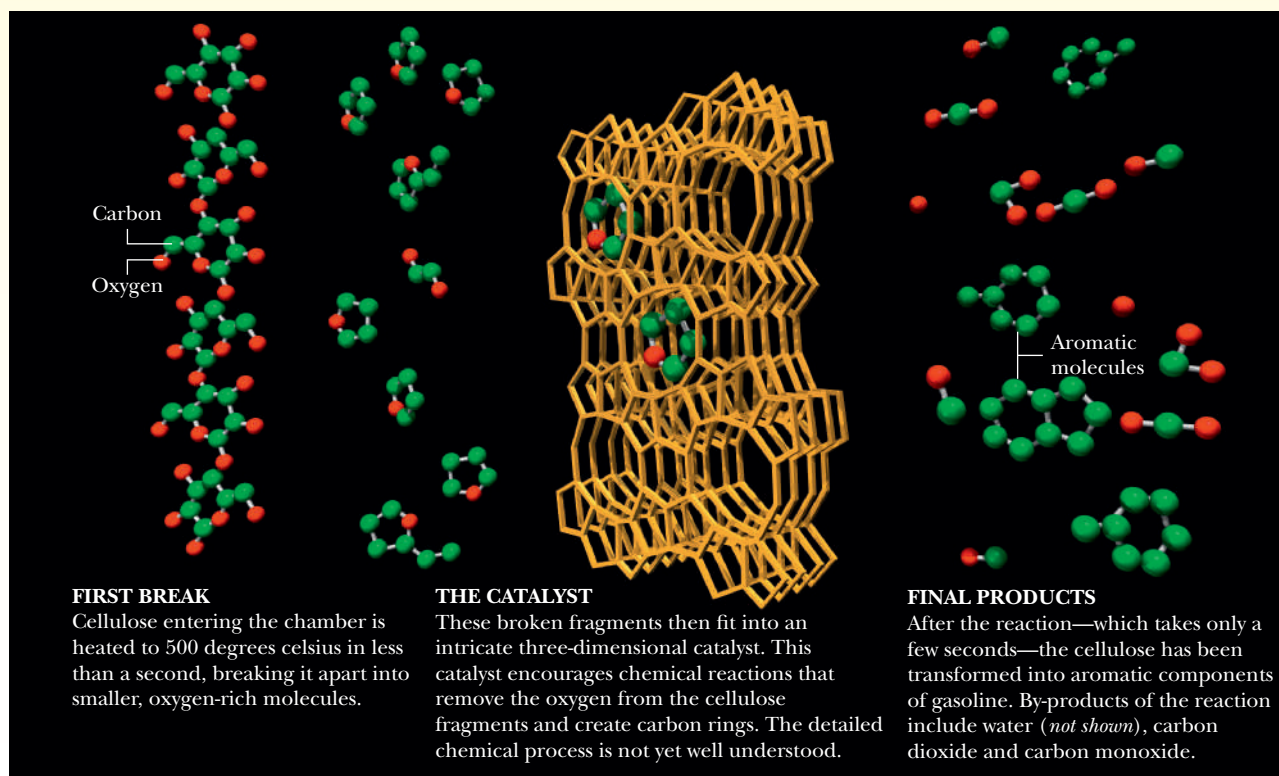
## Biochemical Connections (CONTINUED)

The figure shows the results of a Swedish study on energy requirements for producing food. Energy amounts are given in megajoules. Note the amounts of energy that go into livestock production.

In addition to nutrients, plants produce large quantities of cellulose, which humans cannot digest, but which can be a source of biofuels. Biofuels of plant origin are widely discussed as alternatives to petroleum-based energy sources. Although plants like corn are needed as food by humans and animals, they contain parts that are not. For example, the leftover leaves, stalks, and cobs from corn production can provide an important source of cellulose. Wood chips and wastes from paper mills are another possible source. Plants such as sorghum, which are not an important food source, are yet another possible supply.

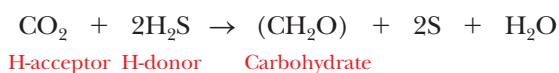
The next step is to turn cellulose into a usable fuel, preferably in liquid form. The process is fundamentally that of breaking up an oxygen-rich polymer to obtain smaller molecules that mostly contain carbon and hydrogen. The goal is achieved by catalytic fast pyrolysis. The method works well and shows promise. A big question in further development is cost. Biofuels need to compete with petroleum-based fuels. Fluctuations in the price of a barrel of oil will play a large role in the future of biofuels.

In catalytic fast pyrolysis, cellulose is introduced into a chamber maintained at 500°C. The cellulose polymer breaks down within a second. The three-dimensional catalyst removes oxygen from the fragments and induces cyclization reactions. The final products are aromatic compounds similar to those found in gasoline.



(From Scientific American, July 2009, vol. 301, issue 1, p. 55.)

As is frequently the case in biological oxidation–reduction reactions, hydrogens as well as electrons are transferred from a donor to an acceptor. In green plants, green algae, and cyanobacteria, the hydrogen donor and acceptor are  $\text{H}_2\text{O}$  and  $\text{CO}_2$ , respectively, with oxygen as a product. Other organisms, such as bacteria, carry out photosynthesis in which there is a hydrogen donor other than water. Some possible donors include  $\text{H}_2\text{S}$ ,  $\text{H}_2\text{S}_2\text{O}_3$ , and succinic acid. As an example, if  $\text{H}_2\text{S}$  is the source of hydrogens and electrons, a schematic equation for photosynthesis can be written with sulfur, rather than oxygen, as a product.



It is also possible for the hydrogen acceptor to be  $\text{NO}_2^-$  or  $\text{NO}_3^-$ , in which case  $\text{NH}_3$  is a product. Photosynthesis linked to oxygen with carbon dioxide as the ultimate hydrogen acceptor is a special case of a far more general process that is widely distributed among many different organisms.

Cyanobacteria were apparently the first organisms that developed the ability to use water as the ultimate reducing agent in photosynthesis. As we have seen, this feat required the development of a second photosystem as well as a new variety of chlorophyll, chlorophyll *a* rather than bacteriochlorophyll in this case. Chlorophyll *b* had not yet appeared on the scene, as it occurs only in eukaryotes. The basic system of aerobic photosynthesis was in place with cyanobacteria. As a result of aerobic photosynthesis by cyanobacteria, the Earth acquired its present atmosphere with its high levels of oxygen. The existence of all other aerobic organisms depended ultimately on the activities of cyanobacteria.

## 22.5 Dark Reactions of Photosynthesis Fix $\text{CO}_2$

The actual storage form of the carbohydrates produced from carbon dioxide by photosynthesis is not glucose but disaccharides (e.g., sucrose in sugarcane and sugar beets) and polysaccharides (starch and cellulose). However, it is customary and convenient to write the carbohydrate product as glucose, and we shall follow this time-honored practice.

Carbon dioxide fixation takes place in the stroma. The equation for the overall reaction, like all equations for photosynthetic processes, is deceptively simple.



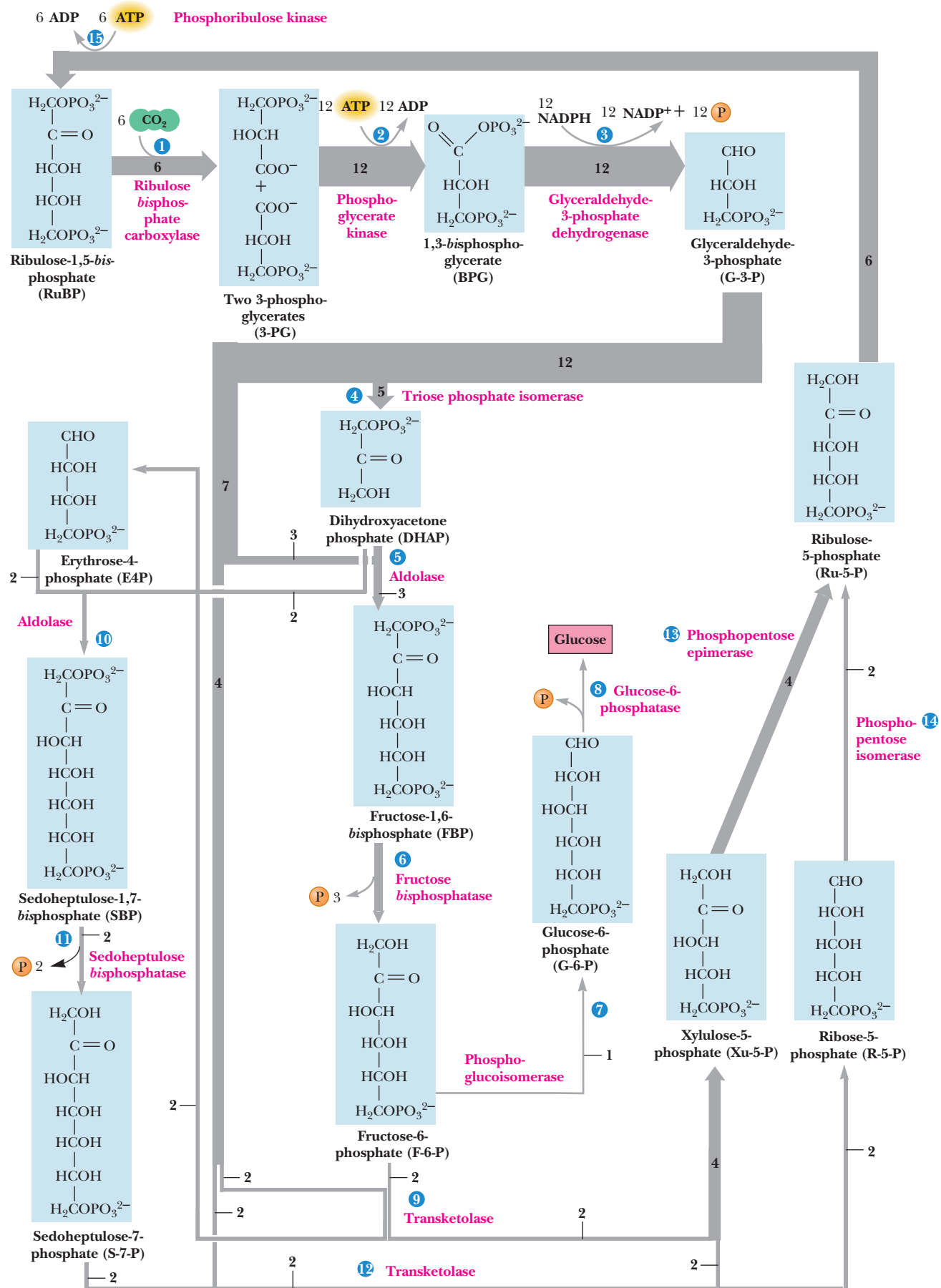
The actual reaction pathway has some features in common with glycolysis and some in common with the pentose phosphate pathway.

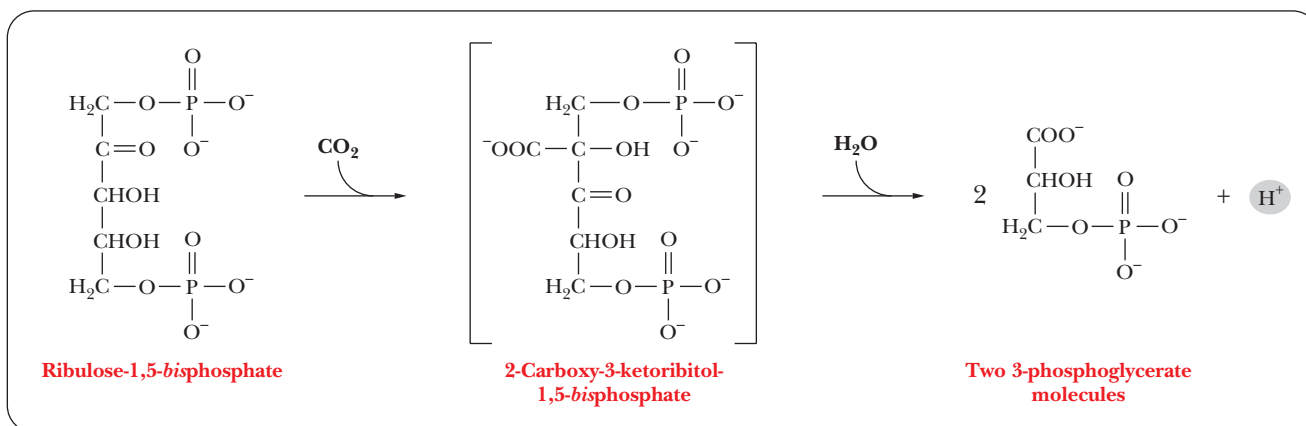
The net reaction of six molecules of carbon dioxide to produce one molecule of glucose requires the carboxylation of 6 molecules of a five-carbon key intermediate, ribulose-1,5-*bis*phosphate, to form 6 molecules of an unstable six-carbon intermediate, which then splits to give 12 molecules of 3-phosphoglycerate. Of these, 2 molecules of 3-phosphoglycerate react in turn, ultimately producing glucose. The remaining 10 molecules of 3-phosphoglycerate are used to regenerate the 6 molecules of ribulose-1,5-*bis*phosphate. The overall reaction pathway is cyclic and is called the *Calvin cycle* (Figure 22.15) after the scientist who first investigated it, Melvin Calvin, winner of the 1961 Nobel Prize in chemistry.

### What is the Calvin cycle?

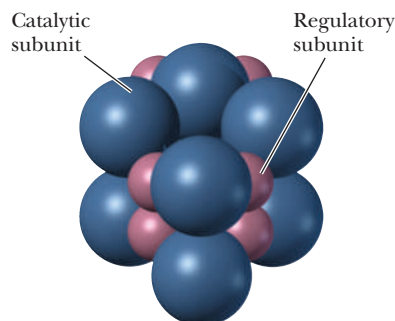
The first reaction of the Calvin cycle is the condensation of ribulose-1,5-*bis*phosphate with carbon dioxide to form a six-carbon intermediate, 2-carboxy-3-ketoribitol-1,5-*bis*phosphate, which quickly hydrolyzes to give two molecules of 3-phosphoglycerate (Figure 22.16). The reaction is catalyzed by the enzyme *ribulose-1,5-bisphosphate carboxylase/oxygenase* (**rubisco**). This enzyme is located on the stromal side of the thylakoid membrane and is probably one of the most abundant proteins in nature, as it accounts for about 15% of the total protein in chloroplasts. The molecular weight of ribulose-1,5-*bis*phosphate carboxylase/oxygenase is about 560,000, and it consists of eight large subunits

■ **FIGURE 22.15** The Calvin cycle of reactions. The number associated with the arrow at each step indicates the number of molecules reacting in a turn of the cycle that produces one molecule of glucose. Reactions are numbered as in Table 22.1.





■ **FIGURE 22.16** The reaction of ribulose-1,5-bisphosphate with  $\text{CO}_2$  ultimately produces two molecules of 3-phosphoglycerate.



■ **FIGURE 22.17** The subunit structure of ribulose-1,5-bisphosphate carboxylase.

(molecular weight 55,000) and eight small subunits (molecular weight 15,000) (Figure 22.17). The sequence of the large subunit is encoded by a chloroplast gene, and that of the small subunit is encoded by a nuclear gene. The endosymbiotic theory for the development of eukaryotes (Section 1.7) is consistent with the idea of independent genetic material in organelles. The large subunit (chloroplast gene) is catalytic, whereas the small subunit (nuclear gene) plays a regulatory role, an observation that is consistent with an endosymbiotic origin for organelles such as chloroplasts.

The incorporation of  $\text{CO}_2$  into 3-phosphoglycerate represents the actual fixation process; the remaining reactions are those of carbohydrates. The next two reactions lead to the reduction of 3-phosphoglycerate to form glyceraldehyde-3-phosphate. The reduction takes place in the same fashion as in gluconeogenesis, except for one unique feature (Figure 22.15): the reactions in chloroplasts require NADPH rather than NADH for the reduction of 1,3-bisphosphoglycerate to glyceraldehyde-3-phosphate. When glyceraldehyde-3-phosphate is formed, it can have two alternative fates: one is the production of six-carbon sugars, and the other is the regeneration of ribulose-1,5-bisphosphate. Table 22.1 summarizes the reactions that take place and indicates their stoichiometry.

The formation of glucose from glyceraldehyde-3-phosphate takes place in the same manner as in gluconeogenesis (Figure 22.15 and reactions 4 through 8 in Table 22.1). The conversion of glyceraldehyde-3-phosphate to dihydroxyacetone phosphate takes place easily (Section 17.2). Dihydroxyacetone phosphate in turn reacts with glyceraldehyde-3-phosphate, in a series of reactions we have already seen, to give rise to fructose-6-phosphate and ultimately to glucose. Because we have already seen these reactions, we shall not discuss them again.

### How is starting material regenerated in the Calvin cycle?

This process is readily divided into four steps: *preparation*, *reshuffling*, *isomerization*, and *phosphorylation*. The preparation begins with conversion of some of the glyceraldehyde-3-phosphate to dihydroxyacetone phosphate (catalyzed by triose phosphate isomerase). This reaction also functions in the production of six-carbon sugars. Portions of both the glyceraldehyde-3-phosphate and the dihydroxyacetone phosphate are then condensed to form fructose-1,6-bisphosphate (catalyzed by aldolase). Fructose-1,6-bisphosphate is hydrolyzed to fructose-6-phosphate (catalyzed by fructose-1,6-bisphosphatase). (See Figure 22.15. Reactions 4 through 6 in Table 22.1 are involved here.) With a supply of glyceraldehyde-3-phosphate, dihydroxyacetone phosphate, and fructose-6-phosphate now available, the reshuffling can begin.

TABLE 22.1

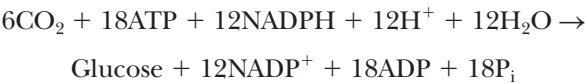
The Calvin Cycle Series of Reactions		
Reactions 1 through 15 constitute the cycle that leads to the formation of one equivalent of glucose. The enzyme catalyzing each step, a concise reaction, and the overall carbon balance are given. Numbers in parentheses show the numbers of carbon atoms in the substrate and product molecules. Prefix numbers indicate in a stoichiometric fashion how many times each step is carried out in order to provide a balanced net reaction.		
1. Ribulose-1,5- <i>bis</i> phosphate carboxylase/oxygenase: 6 CO <sub>2</sub> + 6 H <sub>2</sub> O + 6 RuBP → 12 3-PG		6(1) + 6(5) → 12(3)
2. 3-Phosphoglycerate kinase: 12 3-PG + 12 ATP → 12 1,3-BPG + 12 ADP		12(3) → 12(3)
3. NADP-glyceraldehyde-3-phosphate dehydrogenase: 12 1,3-BPG + 12 NADPH → 12 NADP + 12 G-3-P + 12 P <sub>i</sub>		12(3) → 12(3)
4. Triose phosphate isomerase: 5 G-3-P → 5 DHAP		5(3) → 5(3)
5. Aldolase: 3 G-3-P + 3 DHAP → 3 FBP		3(3) + 3(3) → 3(6)
6. Fructose <i>bis</i> phosphatase: 3 FBP + 3 H <sub>2</sub> O + 3 F-6-P 1 → 3 P <sub>i</sub>		3(6) → 3(6)
7. Phosphoglucosomerase: 1 F-6-P → 1 G-6-P		1(6) → 1(6)
8. Glucose-6-phosphatase: 1 G-6-P + 1 H <sub>2</sub> O → 1 Glucose + 1 P <sub>i</sub>		1(6) → 1(6)
The remainder of the pathway involves regenerating six RuBP acceptors (30 C) from the leftover two F-6-P (12 C), four G-3-P (12 C), and two DHAP (6 C).		
9. Transketolase: 2 F-6-P + 2 G-3-P → 2 Xu-5-P + 2 E4P		2(6) + 2(3) → 2(5) + 2(4)
10. Aldolase: 2 E4P + 2 DHAP → 2 SBP		2(4) + 2(3) → 2(7)
11. Sedoheptulose <i>bis</i> phosphatase: 2 SBP + 2 H <sub>2</sub> O → 2 S-7-P + 2 P <sub>i</sub>		2(7) → 2(7)
12. Transketolase: 2 S-7-P + 2 G-3-P → 2 Xu-5-P + 2 R-5-P		2(7) + 2(3) → 4(5)
13. Phosphopentose epimerase: 4 Xu-5-P → 4 Ru-5-P		4(5) → 4(5)
14. Phosphopentose isomerase: 2 R-5-P → 2 Ru-5-P		2(5) → 2(5)
15. Phosphoribulose kinase: 6 Ru-5-P + 6 ATP → 6 RuBP + 6 ADP		6(5) → 6(5)
Net: 6 CO <sub>2</sub> + 18 ATP + 12 NADPH + 12 H <sup>+</sup> + 12 H <sub>2</sub> O → Glucose + 18 ADP + 18 P <sub>i</sub> + 12 NADP		6(1) → 1(6)

Most of the reactions of the reshuffling process are the same as ones we have already seen as part of the pentose phosphate pathway (Section 18.4). Consequently, we shall concentrate just on the main outline of the process later because the results are summarized in Figure 22.15 and Table 22.1. Reactions catalyzed in turn by *transketolase*, *aldolase*, and *sedoheptulose bisphosphatase* (Reactions 9 through 12 in Table 22.1) are the reactions of rearrangement of carbon skeletons in the reshuffling phase of the Calvin cycle.

The isomerization step (reactions 13 and 14 in Table 22.1) involves the conversion of both ribose-5-phosphate and xylulose-5-phosphate to ribulose-5-phosphate. *Ribose-5-phosphate isomerase* catalyzes the conversion of ribose-5-phosphate to ribulose-5-phosphate, and *xylulose-5-phosphate epimerase* catalyzes the conversion of xylulose-5-phosphate to ribulose-5-phosphate (Figure 22.15). The reverse of both these reactions takes place in the pentose phosphate pathway, catalyzed by the same enzymes.

In the final step (reaction 15 in Table 22.1), ribulose-1,5-*bis*phosphate is regenerated by the phosphorylation of ribulose-5-phosphate. This reaction requires ATP and is catalyzed by the enzyme *phosphoribulose kinase*. The reactions leading to the regeneration of ribulose-1,5-*bis*phosphate summarized in Table 22.1 give a net equation obtained by adding all the reactions.

Taking these points into consideration, we arrive at the *net* equation for the path of carbon in photosynthesis.



The efficiency of energy use in photosynthesis can be calculated fairly easily. The Δ*G*° for the reduction of CO<sub>2</sub> to glucose is +478 kJ (+114 kcal) for each mole of CO<sub>2</sub> (see Question 37), and the energy of light of 600-nm wavelength



is  $1593 \text{ kJ mol}^{-1}$  ( $381 \text{ kcal mol}^{-1}$ ). We shall not explain in detail here how this figure for the energy of the light is obtained, but it comes ultimately from the equation  $E = h\nu$ . Light of wavelength 680 nm or 700 nm has lower energy than light at 600 nm. Thus, the efficiency of photosynthesis is at least  $(478/1593) \times 100$ , or 30%.

## 22.6 CO<sub>2</sub> Fixation in Tropical Plants

In tropical plants, there is a C<sub>4</sub> pathway (Figure 22.18), so named because it involves four-carbon compounds. The operation of this pathway (also called the **Hatch–Slack pathway**) ultimately leads to the C<sub>3</sub> (based on 3-phosphoglycerate) pathway of the Calvin cycle. (There are other C<sub>4</sub> pathways, but this one is most widely studied. Corn [maize] is an important example of a C<sub>4</sub> plant, and it is certainly not confined to the tropics.)

### What is different about CO<sub>2</sub> fixation in tropical plants?

When CO<sub>2</sub> enters the leaf through pores in the outer cells, it reacts first with phosphoenolpyruvate to produce oxaloacetate and P<sub>i</sub> in the mesophyll cells of the leaf. Oxaloacetate is reduced to malate, with the concomitant oxidation of NADPH. Malate is then transported to the bundle-sheath cells (the next layer) through channels that connect the two kinds of cells.

In the bundle-sheath cells, malate is decarboxylated to give pyruvate and CO<sub>2</sub>. In the process, NADP<sup>+</sup> is reduced to NADPH (Figure 22.19). The CO<sub>2</sub> reacts with ribulose-1,5-*bis*phosphate to enter the Calvin cycle. Pyruvate is transported

# Biochemical Connections GENETICS

## Chloroplast Genes

Chloroplasts, like mitochondria, have their own DNA. Scientists speculate that these organelles were originally cyanobacteria that were engulfed by a eukaryotic cell through endosymbiosis (Section 1.8). Some interesting, even elaborate, interactions have evolved from the interplay of both nuclear and chloroplast genes. There are about 3000 chloroplast proteins, but 95% of them are encoded by nuclear genes.

One of the interesting interactions involves rubisco, the principal enzyme of CO<sub>2</sub> fixation. The large subunit of this enzyme is coded by a chloroplast gene, and the small subunit by a nuclear gene. Mechanisms not yet understood must coordinate the two syntheses to ensure equimolar production of both subunits.

The nuclear gene is translated in the cytoplasm, and the protein is then transported to the chloroplast, protected by a chaperonin, via targeting mechanisms (see Section 12.6 and the Biochemical Connections box on page 346). Special targeting sequences are used to direct various nuclear products to the appropriate chloroplast location, using reactions that require ATP hydrolysis. The chaperonin aids in formation of the final, active complex.

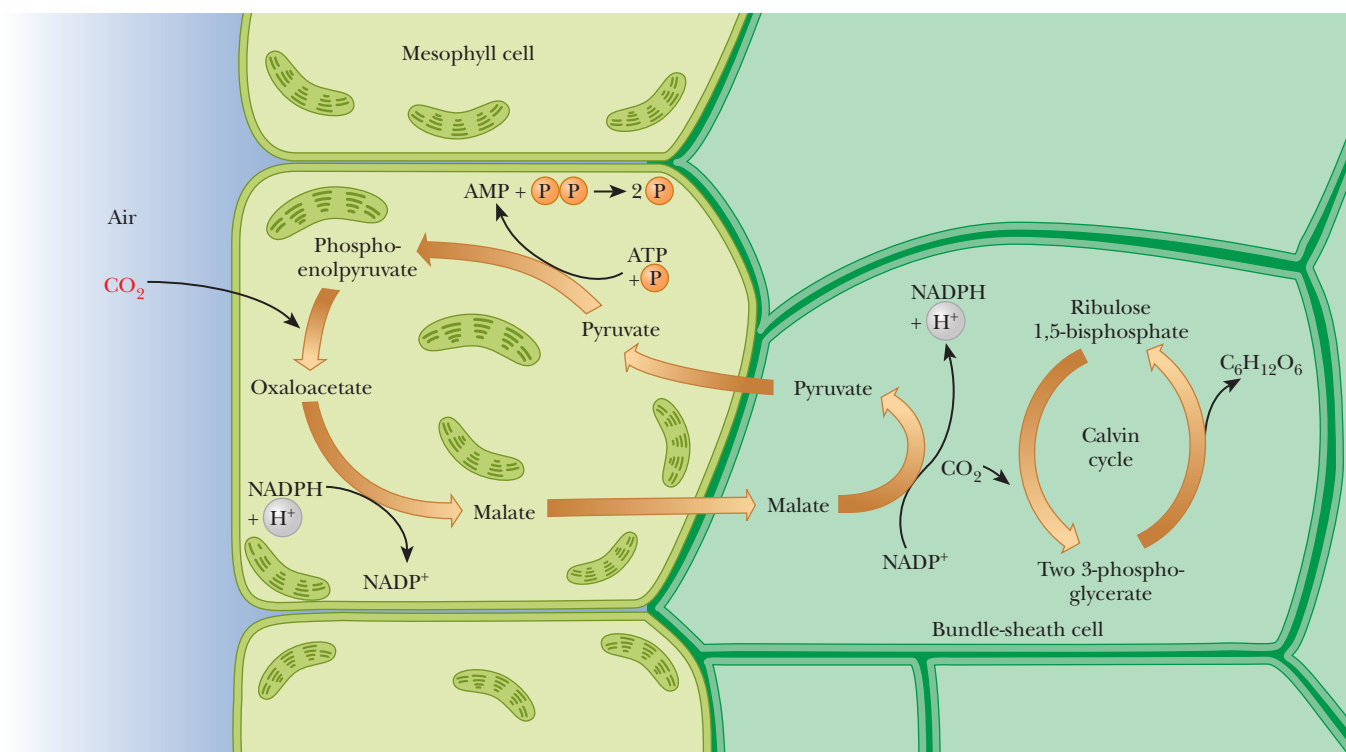
The chloroplast encodes its own RNA polymerase, ribosomal and transfer RNAs, and about one-third of the ribosomal proteins.

The DNA polymerase, aminoacyl synthetases, and the rest of the ribosomal proteins come from the nuclear genes. Different nuclear genes may be used for chloroplasts in different, specialized tissues of the plant. In different classes of plants, different

genes are coded by the nucleus, even though all the same proteins constitute the chloroplast. The sequences of many of these enzymes encoded in the nucleus resemble those found in bacteria more than the sequences of proteins encoded by other nuclear genes. The chloroplast rRNAs likewise resemble bacterial rRNAs. The four subunits of the chloroplast-coded RNA polymerase are homologous to the four subunits of bacterial RNA polymerase. Furthermore, the chloroplast mRNA uses a Shine–Dalgarno sequence to bind to the ribosome, and it does not have a cap or a poly-A tail. These observations are definitely consistent with the idea that the chloroplast (and mitochondrion) originated from bacteria-like organisms, symbiotically taken into early cells, with some subsequent transfer of the organelle genes to the nucleus.

The separation of genetic material between the nucleus and the chloroplast requires a coordination of the transcription of genes in two different locations. How plant cells are able to accomplish this is an active research area. It is now known that there are communications in both directions between the two organelles. A signal proceeding from the nucleus to the chloroplast is considered the more “normal” pathway and is referred to as **anterograde signaling**. When the signal goes from another organelle to the nucleus, it is called **retrograde signaling**. Recently, three different retrograde signaling processes were discovered in algae and higher plants.

The most-studied is signaling by Mg-protoporphyrin IX, a tetrapyrrole generated during chlorophyll biosynthesis. Also studied are signals coming from the expression of chloroplast genes and

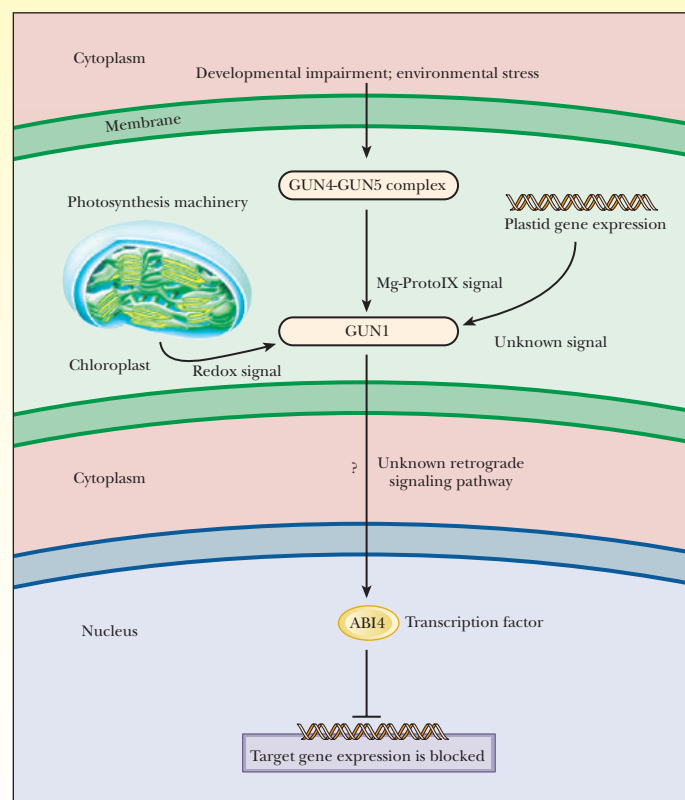


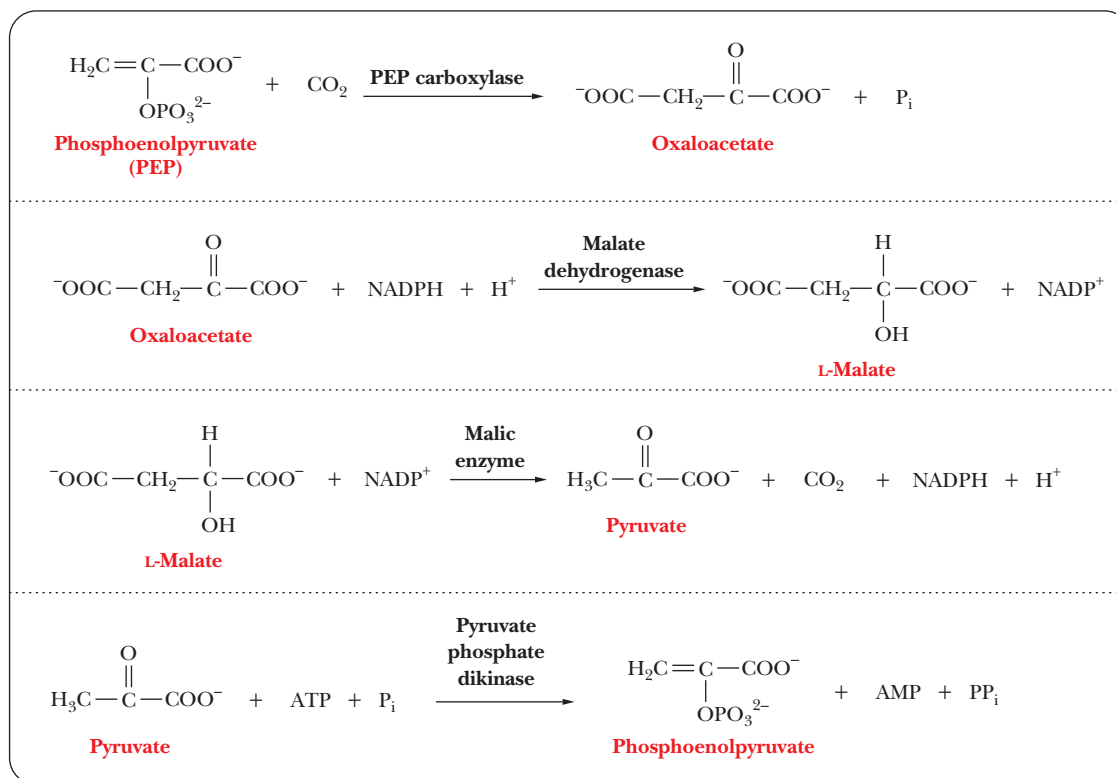
■ **FIGURE 22.18** The C<sub>4</sub> pathway.

## Biochemical Connections (CONTINUED)

those from the photosynthetic electron transport chain (PET). Stress signals and developmental impairment from the cytosol increase the levels of another set of related proteins, GUN4 and GUN5. These three retrograde systems have been shown to act through a GUN1 (genomes uncoupled 1), which is a member of a large family of enigmatic transcription factors in higher plants. It has been shown that all three signal pathways affect the level of GUN1 in the chloroplast (see the figure). It is currently unknown how the GUN1 signal makes its way to the nucleus. However, in the nucleus, the effect is to increase a transcription factor called ABI4 (abscisic acid insensitive 4), which then inhibits transcription of targeted nuclear genes involved in chloroplast metabolism. This common pathway focused to the GUN1 protein allows the smooth coordination of the nuclear part of chloroplast metabolism in response to many environmental signals.

■ **Three different chloroplast signals are focused down to one common signal molecule, GUN1.** By an unknown mechanism, this signal leaves the chloroplast and affects the transcription of nuclear genes by the nuclear transcription factor ABI4. [Redrawn from Zhang, Da-Peng: *Signaling to the Nucleus with a Loaded GUN*. *SCIENCE* 4 may 2007: vol. 316, no. 5825, pp. 700-701. With permission from AAAS].





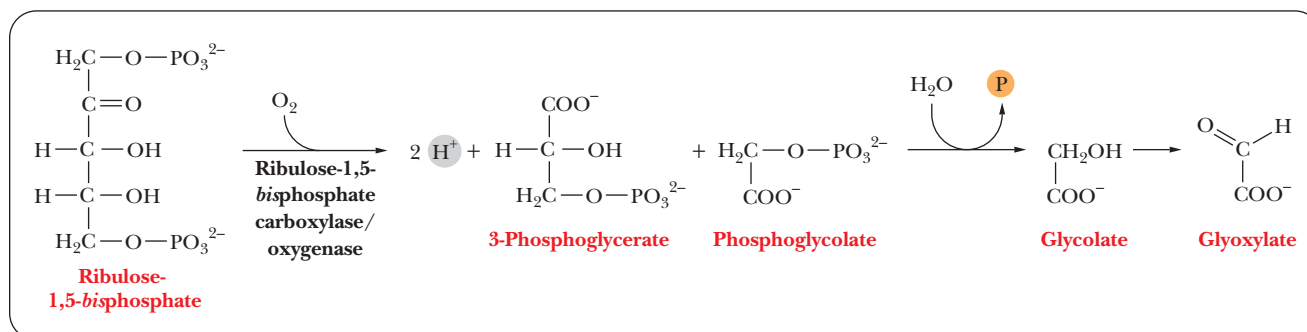
■ **FIGURE 22.19** The characteristic reactions of the C<sub>4</sub> pathway.

back to the mesophyll cells, where it is phosphorylated to phosphoenolpyruvate, which can react with CO<sub>2</sub> to start another round of the C<sub>4</sub> pathway. When pyruvate is phosphorylated, ATP is hydrolyzed to AMP and PP<sub>i</sub>. This situation represents a loss of two high-energy phosphate bonds, equivalent to the use of two ATP. Consequently, the C<sub>4</sub> pathway requires two more ATP equivalents than the Calvin cycle alone for each CO<sub>2</sub> incorporated into glucose. Even though more ATP is required for the C<sub>4</sub> pathway than for the Calvin cycle, there is abundant light to produce the extra ATP by the light reaction of photosynthesis.

Note that the C<sub>4</sub> pathway fixes CO<sub>2</sub> in the mesophyll cells only to unfix it in the bundle-sheath cells, where CO<sub>2</sub> then enters the C<sub>3</sub> pathway. This observation raises the question of the advantage to tropical plants of using the C<sub>4</sub> pathway. The conventional wisdom on the subject focuses on the role of CO<sub>2</sub>, but there is more to the situation than that. According to the conventional view, the point of the C<sub>4</sub> pathway is that it concentrates CO<sub>2</sub> and, as a result, accelerates the process of photosynthesis. Leaves of tropical plants have small pores to minimize water loss, and these small pores decrease CO<sub>2</sub> entry into the plant. Another point to consider is that the *K<sub>M</sub>* for CO<sub>2</sub> of phosphoenolpyruvate carboxylase is lower than that of rubisco, allowing the outer mesophyll cells to fix CO<sub>2</sub> at a lower concentration. This also increases the concentration gradient of CO<sub>2</sub> across the leaf and facilitates the movement of CO<sub>2</sub> into the leaf through the pores. In tropical areas, where there is abundant light, the amount of CO<sub>2</sub> available to plants controls the rate of photosynthesis.

The C<sub>4</sub> pathway deals with the situation, allowing tropical plants to grow more quickly and to produce more biomass per unit of leaf area than plants that use the C<sub>3</sub> pathway. A more comprehensive view of the subject includes a consideration of the role of oxygen and the process of **photorespiration**, in which oxygen is used instead of CO<sub>2</sub> during the reaction catalyzed by rubisco.

Although the actual biological role of photorespiration is not known, several points are well established. The oxygenase activity appears to be an unavoidable,



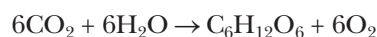
■ **FIGURE 22.20** The characteristic reactions of photorespiration.

wasteful activity of rubisco. Photorespiration is a salvage pathway that saves some of the carbon that would be lost because of the oxygenase activity of rubisco. In fact, the photorespiration is essential to plants even though the plant pays the price in loss of ATP and reducing power; mutations that affect this pathway can be lethal. The principal substrate oxidized in photorespiration is *glycolate* (Figure 22.20). The product of the oxidation reaction, which takes place in peroxisomes of leaf cells (Section 1.5), is *glyoxylate*. (Photorespiration is localized in peroxisomes.) Glycolate arises ultimately from the oxidative breakdown of ribulose-1,5-bisphosphate. The enzyme that catalyzes this reaction is ribulose-1,5-bisphosphate carboxylase/oxygenase, acting as an oxygenase (linked to O<sub>2</sub>) rather than as the carboxylase (linked to CO<sub>2</sub>) that fixes CO<sub>2</sub> into 3-phosphoglycerate.

When levels of O<sub>2</sub> are high compared with those of CO<sub>2</sub>, ribulose-1,5-bisphosphate is oxygenated to produce phosphoglycolate (which gives rise to glycolate) and 3-phosphoglycerate by photorespiration, rather than the two molecules of 3-phosphoglycerate that arise from the carboxylation reaction. This situation occurs in C<sub>3</sub> plants. In C<sub>4</sub> plants, the small pores decrease the entry not only of CO<sub>2</sub> but also of O<sub>2</sub> into the leaves. The ratio of CO<sub>2</sub> to O<sub>2</sub> in the bundle-sheath cells is relatively high as a result of the operation of the C<sub>4</sub> pathway, favoring the carboxylation reaction. C<sub>4</sub> plants have successfully reduced the oxygenase activity by compartmentation and thus have less need of photorespiration. This is an advantage in the hot climates in which C<sub>4</sub> plants are principally found.

## SUMMARY

**How does the structure of the chloroplast affect photosynthesis?** In eukaryotes, the light reactions of photosynthesis take place in the thylakoid membranes of chloroplasts. A series of membrane-bound electron carriers and pigments is able to harness the light energy of the Sun. The equation for photosynthesis



actually represents two processes. One process, the oxidation of water to produce oxygen, requires light energy from the Sun, and the other process, the fixation of CO<sub>2</sub> to give sugars, uses solar energy indirectly. The trapping of light takes place at a reaction center within the chloroplast, and the process requires a pair of chlorophylls in a unique environment.

**How does photosystem II split water to produce oxygen?** In the light reactions, water is oxidized to produce oxygen,

accompanied by the reduction of NADP<sup>+</sup> to NADPH. The light reactions consist of two parts, each carried out by a separate photosystem. When photosystem II transfers electrons from water to an electron-transport chain, oxygen is produced.

**How does photosystem I reduce NADP<sup>+</sup>?** The electrons generated by photosystem II reduce NADP<sup>+</sup> to NADPH in the reactions of photosystem II.

**What is known about the structure of photosynthetic reaction centers?** In the reaction center, specific chlorophylls are associated with membrane-spanning polypeptides. The whole assemblage is arranged to facilitate transfer of electrons to carry out the reactions of photosynthesis.

**How does ATP production in chloroplasts resemble the process in mitochondria?** The two photosystems are linked by an

electron transport chain coupled to the production of ATP. A proton gradient drives the production of ATP in photosynthesis, as it does in mitochondrial respiration.

### Is it possible to have photosynthesis without producing oxygen?

Some forms of bacteria also have photosynthesis, although they often have simpler systems involving only a single photosystem. Early photosynthetic bacteria probably used an electron donor other than water and did not produce oxygen. Bacteria that appeared later, as well as eukaryotes, eventually developed the dual photosystems and the ability to produce oxygen from water, which led to the Earth having an oxygen atmosphere.

**What is the Calvin cycle?** The overall reaction pathway of sugar production is cyclic and is called the Calvin cycle. The dark reactions of photosynthesis involve the net synthesis of 1 molecule of glucose from 6 molecules of  $\text{CO}_2$ . The net reaction of 6 molecules of  $\text{CO}_2$  to produce 1 molecule of glucose requires the carboxylation of 6 molecules of a five-carbon key intermediate, ribulose-1,5-bisphosphate, ultimately forming

12 molecules of 3-phosphoglycerate. Of these, 2 molecules of 3-phosphoglycerate react to give rise to glucose.

**How is starting material regenerated in the Calvin cycle?** The remaining 10 molecules of 3-phosphoglycerate are used to regenerate the 6 molecules of ribulose-1,5-bisphosphate through a series of reactions that include reshuffling of carbon skeletons and isomerizations.

**What is different about  $\text{CO}_2$  fixation in tropical plants?** In addition to the Calvin cycle, there is an alternative pathway for  $\text{CO}_2$  fixation in tropical plants. It is called the  $\text{C}_4$  pathway because it involves four carbon compounds. In this pathway,  $\text{CO}_2$  reacts in the outer (mesophyll) cells with phosphoenolpyruvate to produce oxaloacetate and  $\text{P}_i$ . Oxaloacetate in turn is reduced to malate. Malate is transported from mesophyll cells, where it is produced, to inner (bundle-sheath) cells, where it is ultimately passed to the Calvin cycle. Plants in which the  $\text{C}_4$  pathway operates grow more quickly and produce more biomass per unit of leaf area than  $\text{C}_3$  plants, in which only the Calvin cycle operates.

## REVIEW EXERCISES

 Interactive versions of these problems are assignable in OWL

### 22.1 Chloroplasts Are the Site of Photosynthesis

- Recall** Chlorophyll is green because it absorbs green light less than it absorbs light of other wavelengths. The accessory pigments in the leaves of deciduous trees tend to be red and yellow, but their color is masked by that of the chlorophyll. Suggest a connection between these points and the appearance of fall foliage colors in many sections of the country.
- Recall** The bean sprouts available at the grocery store are white or colorless, not green. Why?
- Recall** What are the principal metal ions used in electron transfer in chloroplasts? Compare them to the ions found in mitochondria.
- Recall** How is the structure of chloroplasts similar to that of mitochondria? How does it differ?
- Recall** List three ways in which the structure of chlorophyll differs from that of heme.
- Reflect and Apply** Suggest a reason why plants contain light-absorbing pigments in addition to chlorophylls *a* and *b*.
- Reflect and Apply** The first amino acid in protein synthesis in the chloroplast is *N*-formyl methionine. What is the significance of this fact?

### 22.2 Photosystems I and II and the Light Reactions of Photosynthesis

- Recall** Is it fair to say that the synthesis of NADPH in chloroplasts is merely the reverse of NADH oxidation in mitochondria? Explain your answer.
- Recall** Outline the events that take place at the photosynthetic reaction center in *Rhodospseudomonas*.
- Recall** What are the two places where light energy is required in the light reaction of photosynthesis? Why must energy be supplied at precisely these points?
- Recall** Do all the chlorophyll molecules in a photosynthetic reaction center play the same roles in the light reactions of photosynthesis?
- Recall** Describe some similarities between the electron transport chains in chloroplasts and in mitochondria.

- Reflect and Apply** Which is likely to have evolved first, the electron transport chain in chloroplasts or in mitochondria? Explain your answer.
- Reflect and Apply** Uncouplers of oxidative phosphorylation in mitochondria also uncouple photoelectron transport and ATP synthesis in chloroplasts. Give an explanation for this observation.
- Reflect and Apply** A larger proton gradient is required to form a single ATP in chloroplasts than in mitochondria. Suggest a reason why. *Hint:* Ions can move across the thylakoid membrane more easily than across the inner mitochondrial membrane.
- Reflect and Apply** Albert Szent-Gyorgi, a pioneer in early photosynthesis research, stated, "What drives life is a little electric current, kept up by the sunshine." What did he mean by this?
- Biochemical Connections** What is implied about the energy requirements of photosystems I and II by the fact that there is a difference in the minimum wavelength of light needed for them to operate (700 nm for photosystem I and 680 nm for photosystem II)?
- Reflect and Apply** Is it reasonable to list standard reduction potentials (see Chapter 20) for the reactions of photosynthesis? Why or why not?
- Reflect and Apply** Why is a photosynthetic reaction center comparable to a battery?
- Reflect and Apply** Antimycin A is an inhibitor of photosynthesis in chloroplasts. Suggest a possible site of action, and indicate the reason for your choice.
- Reflect and Apply** Would you expect  $\text{H}_2\text{O}$  or  $\text{CO}_2$  to be the source of the oxygen produced in photosynthesis? Give the reason for your answer.
- Reflect and Apply** Why do we describe the path of electrons in photosynthesis as starting in photosystem II and ending in photosystem I? In other words, why is the nomenclature "backward"?
- Reflect and Apply** It has taken considerable amounts of research to establish the number of protons pumped across the mitochondrial membrane at the various stages of electron transport. Would you expect to encounter difficulties in determining the number of



protons pumped in electron transport across the thylakoid membrane? Why or why not?

24. **Reflect and Apply** The oxidation of water requires four electrons, but chlorophyll molecules can transfer only one electron at a time. Describe how these two statements can be reconciled.
25. **Reflect and Apply** Why does a loosely bound cytochrome play a unique role in the reaction-center events in *Rhodospseudomonas*?
26. **Reflect and Apply** What are the evolutionary implications of the similarity in structure and function of ATP synthase in chloroplasts and mitochondria?

### 22.3 Photosynthesis and ATP Production

27. **Recall** In cyclic photophosphorylation in photosystem I, ATP is produced, even though water is not split. Explain how the process takes place.
28. **Recall** What are the major similarities and differences between ATP synthesis in chloroplasts, as compared with mitochondria?
29. **Recall** How can a proton gradient be created in cyclic photophosphorylation in photosystem I?
30. **Reflect and Apply** Can ATP production take place in chloroplasts in the absence of light? Give the reason for your answer.
31. **Reflect and Apply** What is the advantage to plants to have the option of both cyclic and noncyclic pathways for photophosphorylation?

### 22.4 Evolutionary Implications of Photosynthesis with and without Oxygen

32. **Recall** Is water the only possible electron donor in photosynthesis? Why or why not?
33. **Reflect and Apply** Suppose that a prokaryotic organism that contains both chlorophyll *a* and chlorophyll *b* has been discovered. Comment on the evolutionary implications of such a discovery.

### 22.5 Dark Reactions of Photosynthesis Fix CO<sub>2</sub>

34. **Recall** Why is rubisco likely to be the most abundant protein in nature?
35. **Biochemical Connections** Is the sequence of amino acids in rubisco encoded by nuclear genes or not? Explain.
36. **Recall** Name some other metabolic pathways that have reactions similar to those of the dark reactions of photosynthesis.

37. **Reflect and Apply** Using information from Sections 15.3 and 15.6, show how the  $\Delta G^\circ$  of 478 kJ (114 kcal) is obtained for each mole of CO<sub>2</sub> fixed in photosynthesis. The reaction in question is  $6\text{CO}_2 + 6\text{H}_2\text{O} \rightarrow \text{Glucose} + 6\text{O}_2$ .
38. **Reflect and Apply** If photosynthesizing plants are grown in the presence of <sup>14</sup>CO<sub>2</sub>, is every carbon atom of the glucose that is produced labeled with the radioactive carbon? Why or why not?
39. **Reflect and Apply** Rubisco has a very low turnover number, about 3CO<sub>2</sub> per second. What might this low number tell us about the evolution of rubisco?
40. **Biochemical Connections** What key aspects of chloroplasts (and mitochondria) are consistent with the theory that they may have once been bacteria? List three specific features.
41. **Reflect and Apply** Suggest a reason why the evolution of the pathway for the regeneration of ribulose-1,5-bisphosphate from glyceraldehyde-3-phosphate was “no big deal.”
42. **Biochemical Connections** Why would nature evolve a key enzyme, rubisco, that is so sensitive to oxygen, resulting in photorespiration?
43. **Reflect and Apply** Does the whole Calvin cycle represent carbon dioxide fixation? Why or why not?
44. **Reflect and Apply** What is the evolutionary advantage to organisms that the Calvin cycle has a number of reactions in common with other pathways?
45. **Reflect and Apply** Why do we refer to the conversion of six molecules of carbon dioxide (six carbon atoms) to one molecule of glucose (also six carbon atoms) as a *net* reaction?

### 22.6 CO<sub>2</sub> Fixation in Tropical Plants

46. **Recall** How does the production of sugars by tropical plants differ from the same reactions in the Calvin cycle?
47. **Recall** How does photosynthesis in C<sub>4</sub> plants differ from the process in C<sub>3</sub> plants?
48. **Recall** What is photorespiration?
49. **Reflect and Apply** Why is it advantageous to tropical plants to use the C<sub>4</sub> rather than the C<sub>3</sub> fixation pathway?
50. **Reflect and Apply** What would be the effect on plants if photorespiration did not exist?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).



# The Metabolism of Nitrogen

# 23

© Robert Pickett/Corbis

Root nodules of leguminous plants play a pivotal role in nitrogen fixation.

## 23.1 Nitrogen Metabolism: An Overview

We have seen the structures of many types of compounds that contain nitrogen, including amino acids, porphyrins, and nucleotides, but we have not discussed their metabolism. The metabolic pathways we have dealt with up to now have mainly involved compounds of carbon, hydrogen, and oxygen, such as sugars and fatty acids. Several important topics can be included in our discussion of the metabolism of nitrogen. The first of these is **nitrogen fixation**, the process by which inorganic molecular nitrogen from the atmosphere ( $\text{N}_2$ ) is incorporated first into ammonia and then into organic compounds that are of use to organisms. Nitrate ion ( $\text{NO}_3^-$ ), another kind of inorganic nitrogen, is the form in which nitrogen is found in the soil, and many fertilizers contain nitrates, frequently potassium nitrate. The process of **nitrification** (nitrate reduction to ammonia) provides another way for organisms to obtain nitrogen. Nitrate ion and nitrite ion ( $\text{NO}_2^-$ ) are also involved in **denitrification** reactions, which return nitrogen to the atmosphere (Figure 23.1).

Ammonia formed by either pathway, nitrogen fixation or nitrification, enters the biosphere. Ammonia is converted to organic nitrogen by plants, and organic nitrogen is passed to animals through food chains. Finally, animal waste products, such as urea, are excreted and degraded to ammonia by microorganisms. The word *ammonia* comes from *sal ammoniac* (ammonium chloride), which was first prepared from the dung of camels at the temple of Jupiter Ammon in North Africa. The process of death and decay releases ammonia in both plants and animals. Denitrifying bacteria reverse the conversion of ammonia to nitrate and then recycle the  $\text{NO}_3^-$  as free  $\text{N}_2$  (Figure 23.1).

The topic of nitrogen metabolism includes the biosynthesis and breakdown of *amino acids*, *purines*, and *pyrimidines*; also, the metabolism of *porphyrins* is related to that of amino acids. Many of these pathways, particularly the anabolic ones, are long and complex. In discussing pathways in which the amount of material is large and highly detailed, we shall concentrate on the most important points. Specifically, we shall concentrate on overall patterns and on interesting reactions of wide applicability. We shall also be interested in health-related aspects of this material. Other reactions will be found at Biochemistry Interactive website for this text. It can be considered a repository of supplementary material for this chapter, and we shall refer to it a number of times.

## 23.2 Nitrogen Fixation

Bacteria are responsible for the reduction of  $\text{N}_2$  to ammonia ( $\text{NH}_3$ ). Typical nitrogen-fixing bacteria are symbiotic organisms that form nodules on the roots of leguminous plants, such as beans and alfalfa. Many free-living microbes and some cyanobacteria also fix nitrogen. Plants and animals cannot carry out nitrogen fixation. This conversion of molecular nitrogen to ammonia is the only source of nitrogen in the biosphere except for that provided by nitrates.

## Chapter Outline

### 23.1 Nitrogen Metabolism: An Overview

### 23.2 Nitrogen Fixation

- How is nitrogen from the atmosphere incorporated into biologically useful compounds?

### 23.3 Feedback Inhibition in Nitrogen Metabolism

### 23.4 Amino Acid Biosynthesis

- What are some common features in amino acid biosynthesis?
- What makes transamination reactions important in amino acid biosynthesis?
- What is the importance of one-carbon transfers?

### 23.5 Essential Amino Acids

### 23.6 Amino Acid Catabolism

- What is the fate of the carbon skeleton in amino acid breakdown?
- What is the role of the urea cycle in amino acid breakdown?

### 23.7 Purine Biosynthesis

- How is inosine monophosphate converted to AMP and GMP?
- What are the energy requirements for production of AMP and GMP?

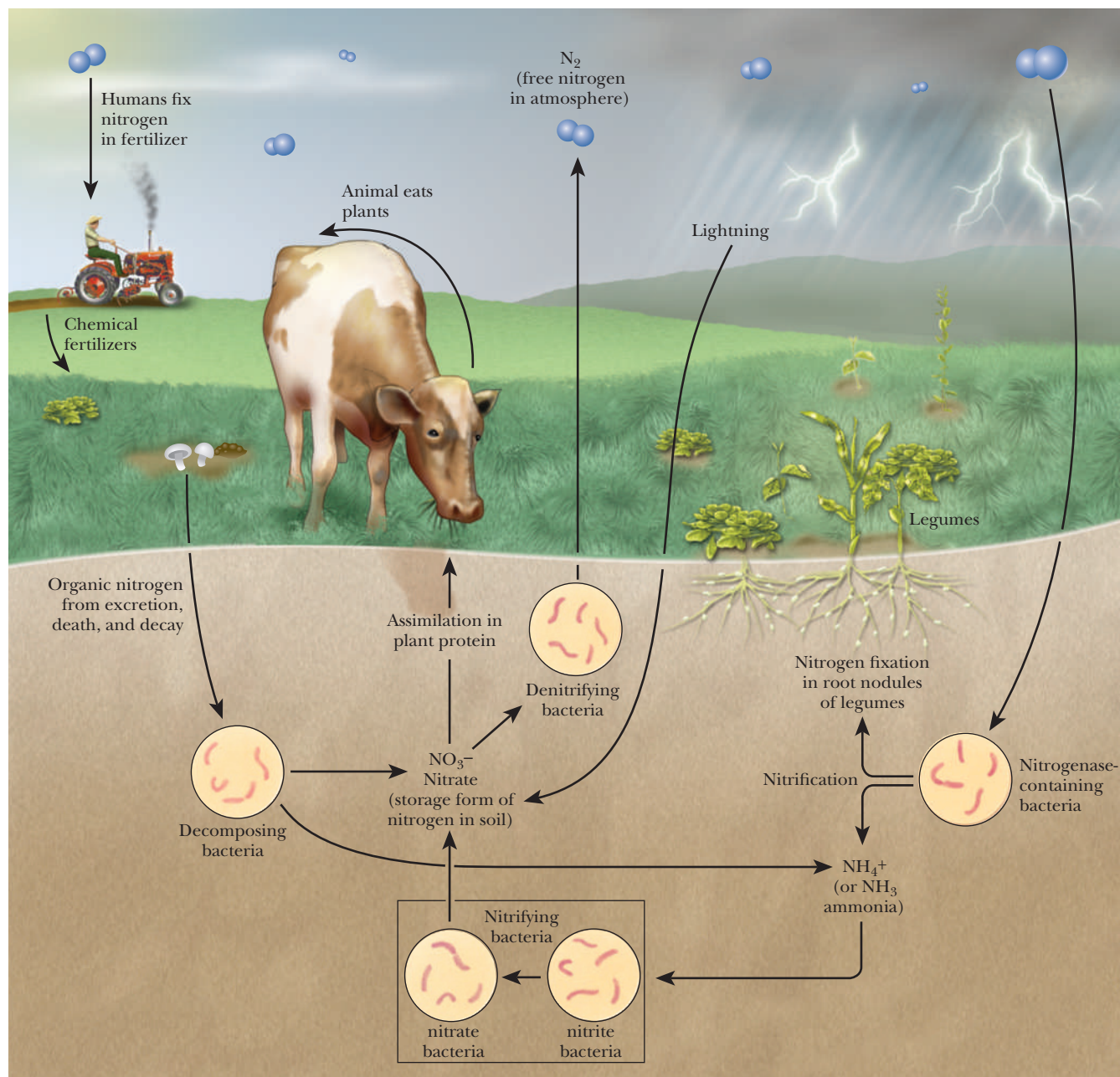
### 23.8 Purine Catabolism

### 23.9 Pyrimidine Biosynthesis and Catabolism

### 23.10 Conversion of Ribonucleotides to Deoxyribonucleotides

### 23.11 Conversion of dUDP to dTTP

**OWL** Online homework for this chapter may be assigned in OWL.

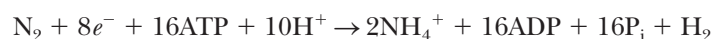


■ **FIGURE 23.1** The flow of nitrogen in the biosphere.

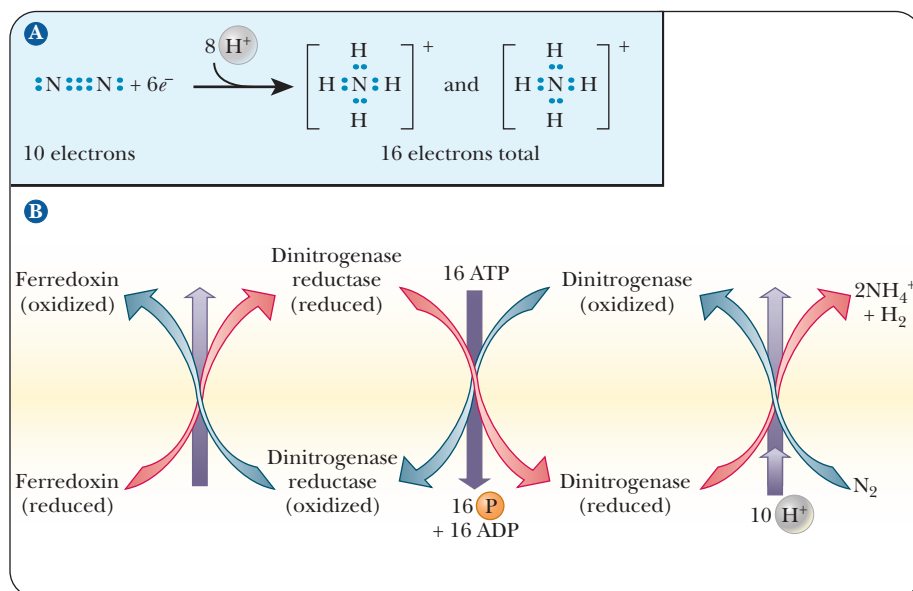
The conjugate acid form of  $NH_3$ , ammonium ion ( $NH_4^+$ ), is the form of nitrogen that is used in the first stages of the synthesis of organic compounds. Parenthetically,  $NH_3$  obtained by chemical synthesis from nitrogen and hydrogen is the starting point for the production of many synthetic fertilizers, which frequently contain nitrates.

### How is nitrogen from the atmosphere incorporated into biologically useful compounds?

The **nitrogenase** enzyme complex found in nitrogen-fixing bacteria catalyzes the production of ammonia from molecular nitrogen. The half-reaction of reduction (Figure 23.2a) is







**FIGURE 23.2** Some aspects of the nitrogenase reaction. (a) The reduction of N<sub>2</sub> to 2NH<sub>4</sub><sup>+</sup>. (b) The path of electrons from ferredoxin to N<sub>2</sub>.

in which six electrons are used to reduce molecular nitrogen to ammonium ion. An additional two electrons are used to reduce hydrogen ion to H<sub>2</sub>. The total reaction catalyzed by nitrogenase is an eight-electron reduction.

The half-reaction of oxidation varies because different organisms vary in terms of the substance oxidized to supply electrons. Several proteins are included in the nitrogenase complex. Ferredoxin is one of them (this protein

## Biochemical Connections PLANT SCIENCE

### Why Is the Nitrogen Content of Fertilizers So Important?

Crop production per acre in the United States is higher than in many areas of the world. In part, this is the result of extensive use of fertilizers, especially those that supply nitrogen in a form that plants can use readily. Both ammonium and nitrate ions are used; even ammonia gas can be pumped into the ground, if enough water is available in the soil to dissolve it.

Ammonia is toxic to animals, so it often comes as a surprise that ammonia gas itself can be used for fertilization. Plants can assimilate ammonia rapidly, but they usually never get the chance to do so because the nitrifying soil bacteria, especially *Nitrosomonas* and *Nitrobacter*, rapidly convert the ammonia first to nitrite and then to nitrate. The final nitrate product is easily converted back to ammonia, but the process requires energy. Ammonia is especially useful as a fertilizer in the early spring and for germinating plants. In the spring, the soil is usually damp enough to dissolve the ammonia so that it can move to the plants. Because light is less available in the early spring, the young plants do not have enough energy to convert the nitrate back to ammonia until their chloroplasts develop fully. Fortunately, because of the condition of the soil, the ammonia goes directly to the plants rather than to the soil bacteria.

The genes for the enzymes for nitrogen fixation have been studied extensively. Much research is going on to determine whether these genes can be incorporated into crop plants, which would reduce the amount of nitrogen fertilizer needed for maximal plant growth and crop production.

Two other sources of nitrogen fixation are often overlooked. The first is the chemical synthesis of ammonia from H<sub>2</sub> and N<sub>2</sub>, called the Haber process, after its discoverer, the German chemist Fritz Haber. This reaction is very important for the formation of

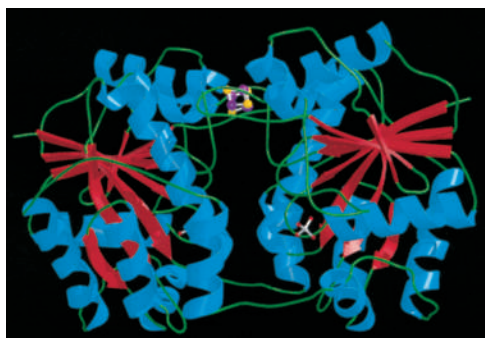
chemical fertilizers, and it is responsible for a great deal of the organic nitrogen currently found in the biosphere. The second source of fixed nitrogen is that produced by lightning.



C. P. Vance/Visuals Unlimited

**Nitrogen-fixing bacteria form nodules on alfalfa roots.**





■ **FIGURE 23.3** The X-ray structure of the *Azotobacter vinelandii* Fe-protein dimer. (From Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. MM Georgiadis, H Komiya, P Chakrabarti, D Woo, JJ Kornuc, and DC Rees. (18 September 1992) *Science* 257 (5077), 1653. Reprinted with permission from AAAS.)

also plays an important role in electron transfer in photosynthesis; Section 22.3). There are also two proteins specific to the nitrogenase reaction. One is an iron-sulfur (Fe-S) protein called *dinitrogenase reductase*. The other is an iron-molybdenum (Fe-Mo) protein, called *dinitrogenase*. The flow of electrons is from ferredoxin to dinitrogenase reductase to dinitrogenase to nitrogen (Figure 23.2b). The nature of the nitrogenase complex is a subject of active research. Significant progress has been made in this work with the determination by X-ray crystallography of the three-dimensional structure of both the Fe protein and the Fe-Mo protein from *Azotobacter vinelandii* (Figure 23.3). The Fe protein is a dimer (“the iron butterfly”), with the iron-sulfur cluster located at the butterfly’s head. The nitrogenase is even more complicated, with several types of subunits arranged into tetramers. Ferredoxin, dinitrogenase reductase, and dinitrogenase combine to perform a series of single-electron transfers, eventually transferring the eight electrons necessary to complete the reduction of  $N_2$  to  $NH_4^+$ . It is worth noting that the reactions of nitrogen fixation consume a great deal of energy. It is estimated that about half of the ATP produced from photosynthesis in legumes is used to fix nitrogen.

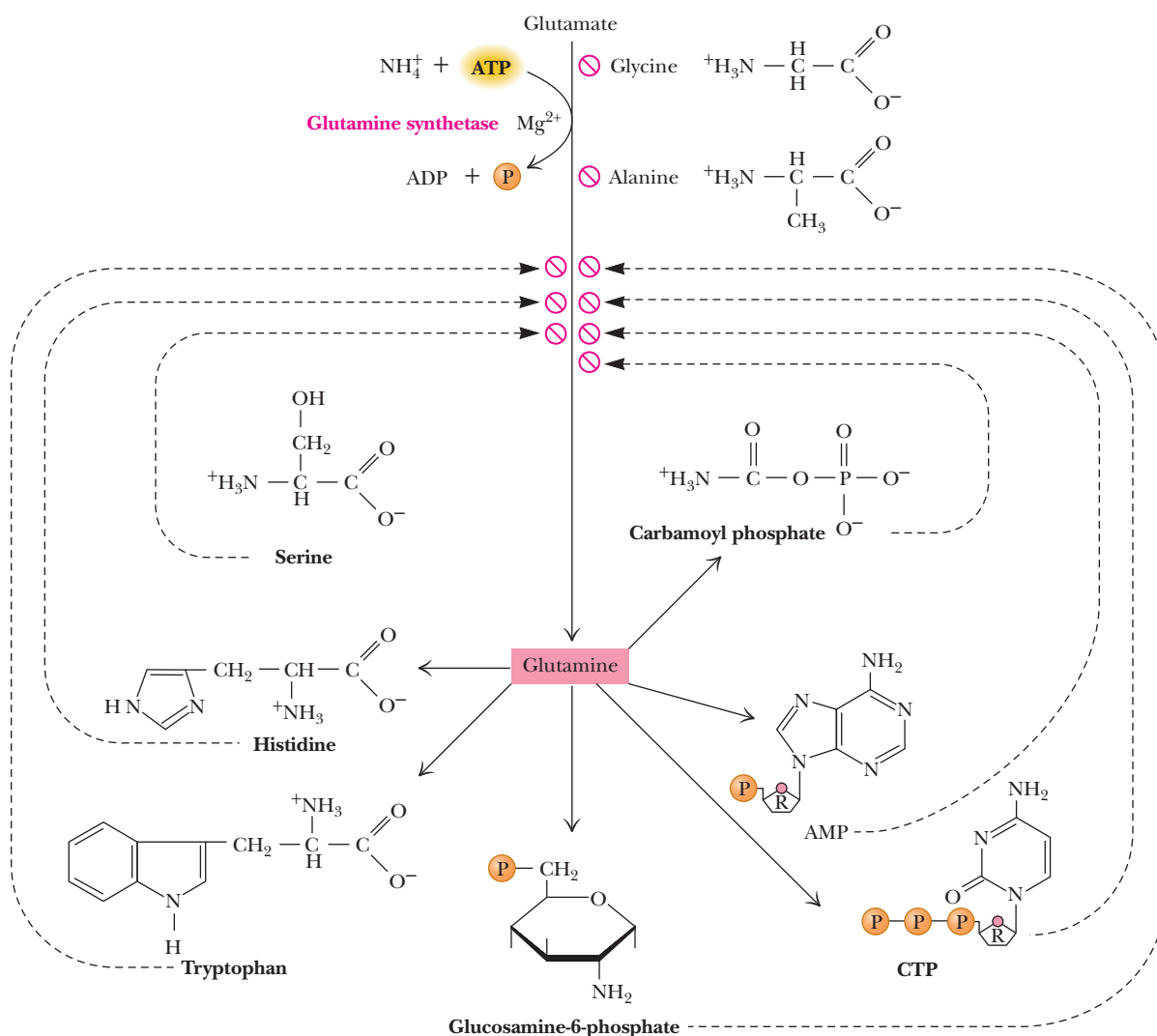
### 23.3 Feedback Inhibition in Nitrogen Metabolism

The biosynthetic pathways that produce amino acids and the bases of nucleotides (purines and pyrimidines) are long and complex, requiring a large investment of energy by the organism. If there is a high level of some end product, such as an amino acid or a nucleotide, the cell saves energy by not making that compound. However, the cell needs a signal to tell it to stop producing more of that particular compound. The signal is frequently part of a **feedback inhibition** mechanism, in which the end product of a metabolic pathway inhibits an enzyme at the beginning of the pathway. We saw an example of such a control mechanism when we discussed the allosteric enzyme aspartate transcarbamoylase in Section 7.1. This enzyme catalyzes one of the early stages of pyrimidine nucleotide biosynthesis, and it is inhibited by the end product of that pathway—namely, cytidine triphosphate (CTP). Feedback inhibition is frequently encountered in the biosynthesis of amino acids and nucleotides. Another prime example of allosteric regulation by feedback inhibition is found in the activity of the enzyme glutamine synthetase, one of the key enzymes in amino acid biosynthesis (Figure 23.4). Nine allosteric inhibitors are involved here (glycine, alanine, serine, histidine, tryptophan, CTP, AMP, carbamoyl phosphate, and glucosamine-6-phosphate).

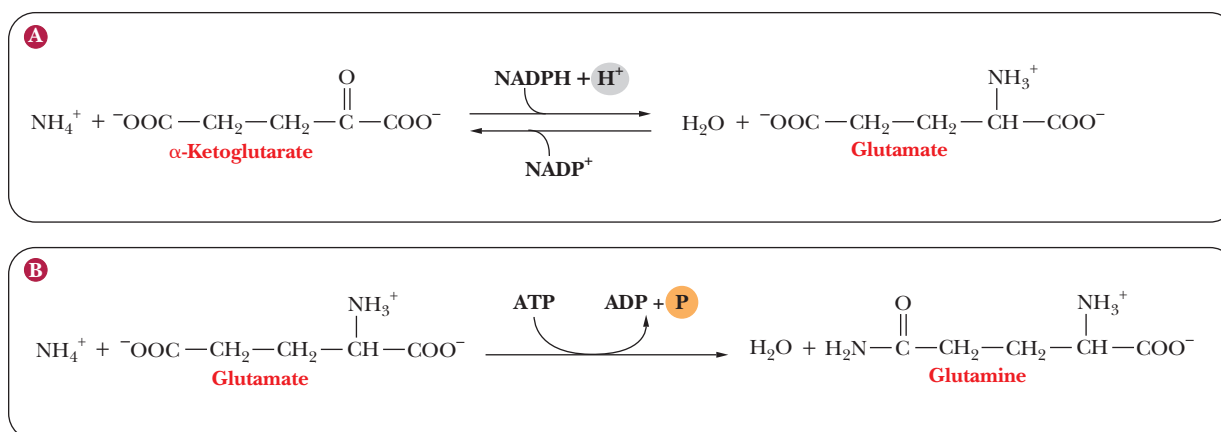
Glycine, alanine, and serine are key indicators of amino acid metabolism in the cell. Each of the other six compounds represents an end product of a biosynthetic pathway that depends on glutamine. Feedback inhibition is very effective because a single product molecule can inhibit an enzyme capable of synthesizing many hundreds or thousands of product molecules.

### 23.4 Amino Acid Biosynthesis

Ammonia is toxic in high concentrations, and so it must be incorporated into biologically useful compounds when it is formed by the reactions of nitrogen fixation discussed earlier in this chapter. The amino acids glutamate and glutamine are of central importance in the process. **Glutamate** arises from  $\alpha$ -ketoglutarate, and **glutamine** is formed from glutamate (Figure 23.5). The production of glutamate is a reductive amination, and the production of glutamine is amidation. In other reactions of amino acid anabolism the  $\alpha$ -amino group of glutamate and the side-chain amino group of glutamine are shifted to other compounds in **transamination** reactions.



■ **FIGURE 23.4** The allosteric regulation of glutamine synthetase activity by feedback inhibition.

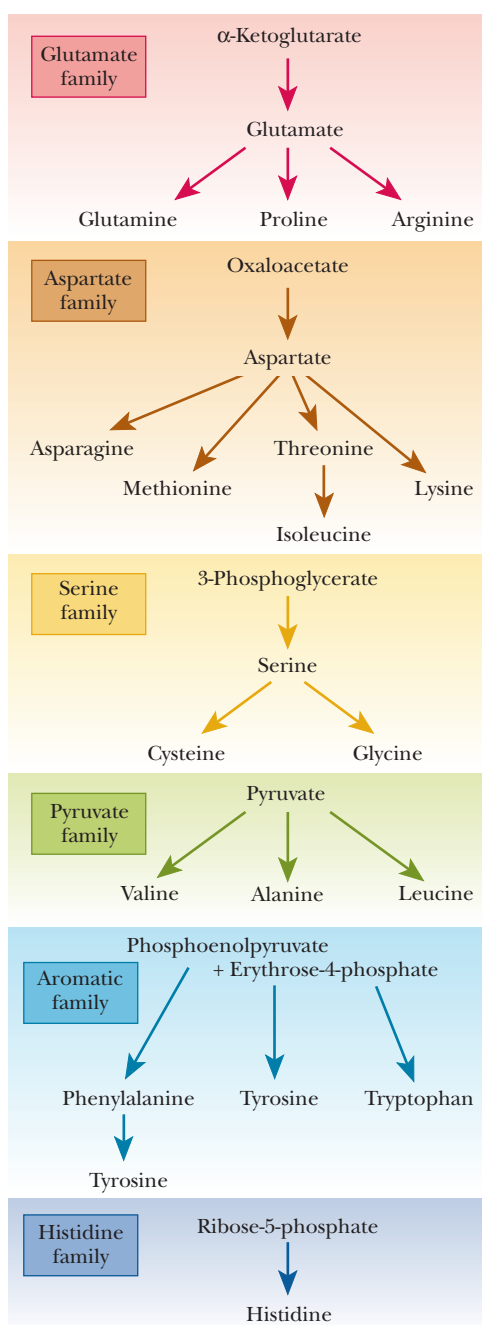


■ **FIGURE 23.5** Biosynthesis of glutamate and glutamine. (a) The production of glutamate from  $\alpha$ -ketoglutarate. (b) The production of glutamine from glutamate.

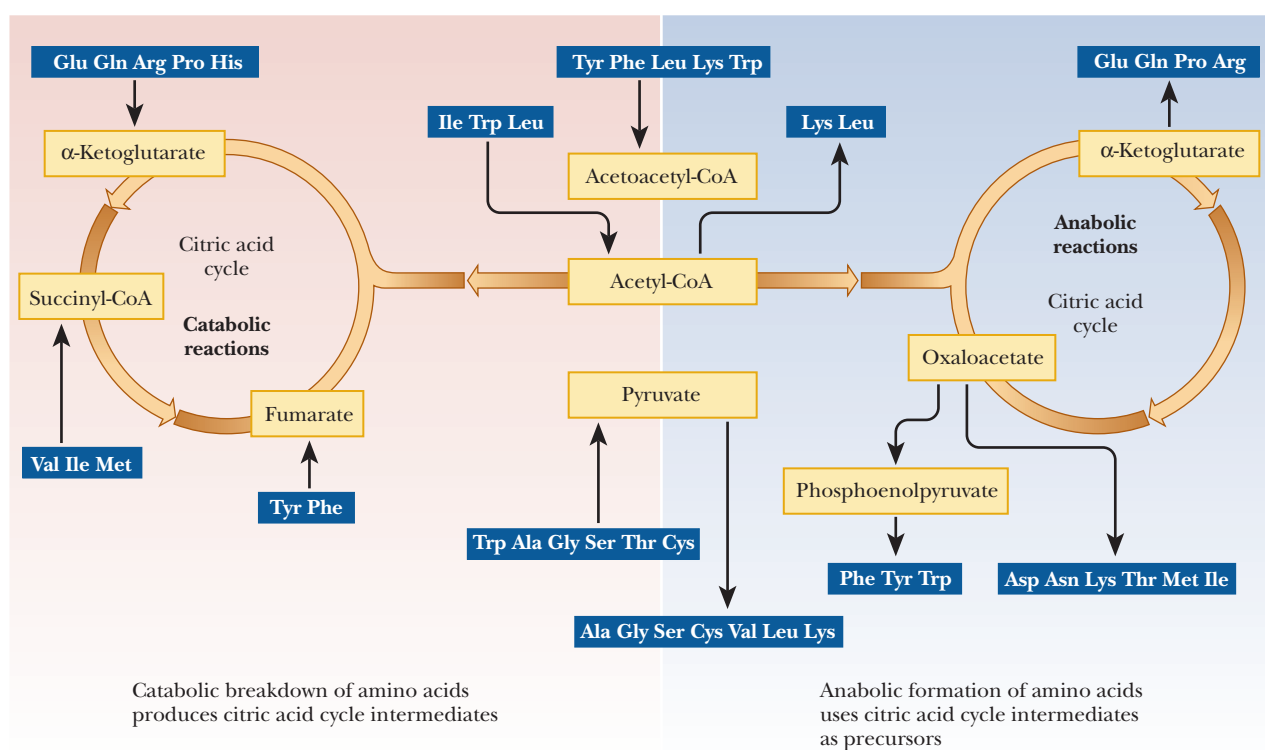
### What are some common features in amino acid biosynthesis?

The biosynthesis of amino acids involves a common set of reactions. In addition to transamination reactions, transfer of one-carbon units, such as formyl or methyl groups, occurs frequently. We are not going to discuss all the details of the reactions that give rise to amino acids. We can, however, organize this material by grouping amino acids into families based on common precursors (Figure 23.6). The reactions of some of the individual families of amino acids provide good examples of reactions that are of general importance, such as transamination and one-carbon transfer.

We can also make some generalizations about amino acid metabolism in terms of the relationship of the carbon skeleton to the citric acid cycle and the related reactions of pyruvate and acetyl-CoA (Figure 23.7). The citric acid cycle is amphibolic; it has a part in both catabolism and anabolism. The anabolic aspect of the citric acid cycle is of interest in amino acid biosynthesis. The catabolic



■ **FIGURE 23.6** Families of amino acids based on biosynthetic pathways. Each family has a common precursor.



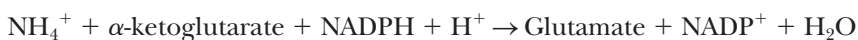
■ **FIGURE 23.7** The relationship between amino acid metabolism and the citric acid cycle.

aspect is apparent in the breakdown of amino acids, leading to their eventual excretion, which takes place in reactions related to the citric acid cycle.

### What makes transamination reactions important in amino acid biosynthesis?

Glutamate is formed from  $\text{NH}_4^+$  and  $\alpha$ -ketoglutarate in a reductive amination that requires NADPH. This reaction is reversible and is catalyzed by *glutamate dehydrogenase* (GDH).

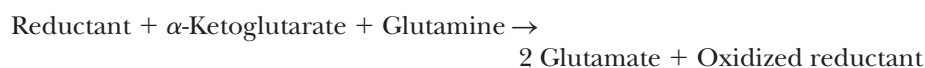
Glutamate is a major donor of amino groups in reactions, and  $\alpha$ -ketoglutarate is a major acceptor of amino groups (see Figure 23.5a). Note the requirement for reducing power.



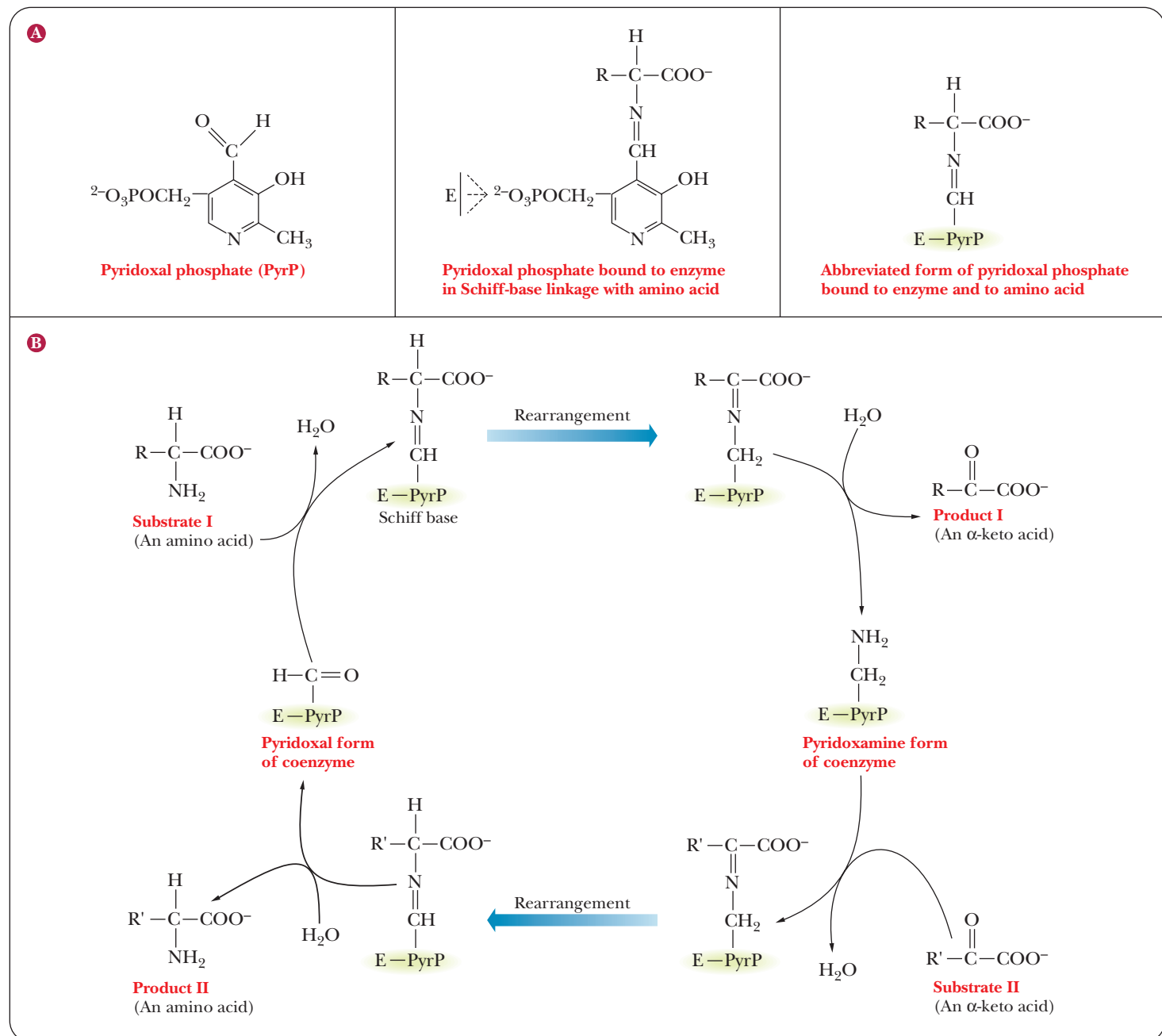
The conversion of glutamate to glutamine is catalyzed by **glutamine synthetase** (GS) in a reaction that requires ATP (see Figure 23.5b).



These reactions fix inorganic nitrogen ( $\text{NH}_3$ ), forming organic (carbon-containing) nitrogen compounds, such as amino acids, but they frequently do not operate in this sequential fashion. In fact, the combination of GDH and GS is responsible for most of the assimilation of ammonia into organic compounds, especially in organisms that are rich in nitrogen sources. However, the  $K_M$  of GS is considerably lower than that of GDH. When nitrogen is limiting, as is frequently the case in plants, the conversion of glutamate to glutamine is the preferred mode of nitrogen assimilation. This means that the supply of glutamate becomes depleted unless there is some way to replenish it. The reductive amination of  $\alpha$ -ketoglutarate with the amide nitrogen of glutamine as the nitrogen source is the way this is done.



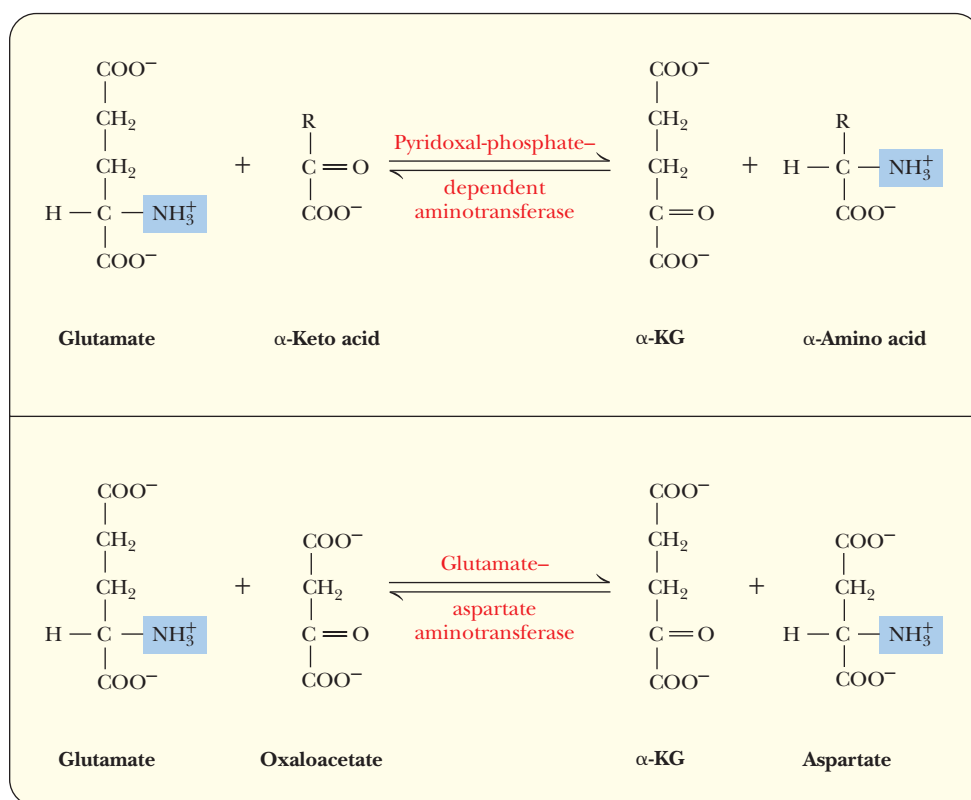
The reductant can be NADH, NADPH (in yeast and bacteria), or reduced ferredoxin (in plants). The enzyme that catalyzes this reaction is glutamate synthase; it is also known as glutamate:oxoglutarate aminotransferase (GOGAT). A GS/GOGAT complex exists in plants and allows them to cope with conditions of limited nitrogen availability. Enzymes that catalyze transamination reactions require pyridoxal phosphate as a coenzyme (Figure 23.8). We discussed this compound in Section 7.8 as a typical example of a coenzyme, and here we can see its mode of action in context.



■ **FIGURE 23.8** The role of pyridoxal phosphate in transamination reactions.

(a) The mode of binding of pyridoxal phosphate (PyrP) to the enzyme (E) and to the substrate amino acid. (b) The reaction itself. The original substrate, an amino acid, is deaminated, while an  $\alpha$ -keto acid is aminated to form an amino acid. The net reaction is one of transamination. Note that the coenzyme is regenerated and that the original substrate and final product are both amino acids.



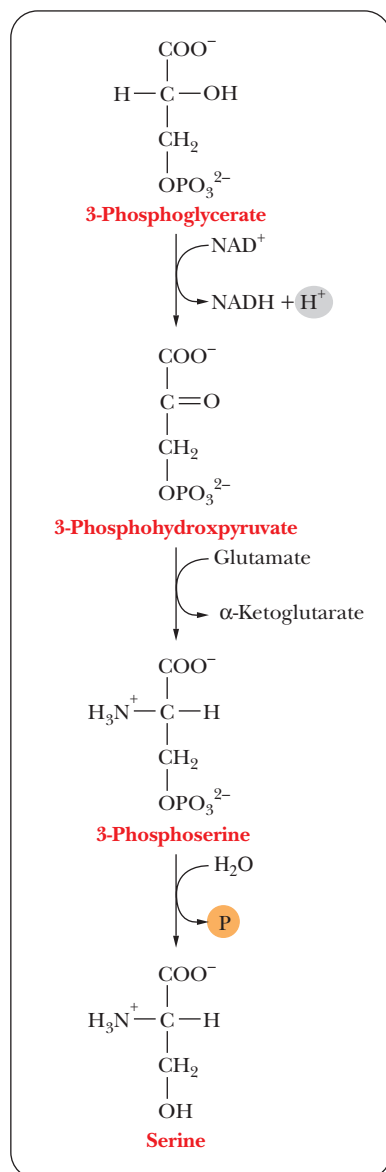


■ **FIGURE 23.9** Transamination reactions switch an amino group from one amino acid to an  $\alpha$ -keto acid. Glutamate and  $\alpha$ -ketoglutarate ( $\alpha$ -KG) are one donor/acceptor pair. Above, a general case. Below, a specific case, in which the other donor/acceptor pair is aspartate and oxaloacetate.

Pyridoxal phosphate (PyrP) forms a Schiff base with the amino group of Substrate I (the amino-group donor). The next stage is a rearrangement followed by hydrolysis, which removes Product I (the  $\alpha$ -keto acid corresponding to Substrate I). The coenzyme now carries the amino group (pyridoxamine). Substrate II (another  $\alpha$ -keto acid) then forms a Schiff base with pyridoxamine. Again there is a rearrangement followed by a hydrolysis, which gives rise to Product II (an amino acid) and regenerates pyridoxal phosphate. The net reaction is that an amino acid (Substrate I) reacts with an  $\alpha$ -keto acid (Substrate II) to form an  $\alpha$ -keto acid (Product I) and an amino acid (Product II). The amino group has been transferred from Substrate I to Substrate II, forming the amino acid, Product II. The overall reaction can be seen for a general case and for a specific case in Figure 23.9. When not involved with one of the substrates, the pyridoxal group is bound in a Schiff-base linkage to an active site  $\epsilon$ -NH<sub>2</sub> group of lysine. Pyridoxal phosphate is a versatile coenzyme that is also involved in other reactions, including decarboxylations, racemizations, and movement of hydroxymethyl groups, as we shall see with the conversion of serine to glycine.

### What is the importance of one-carbon transfers?

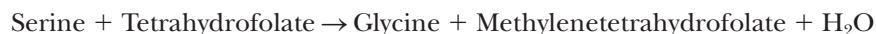
In addition to transamination reactions, one-carbon transfer reactions occur frequently in amino acid biosynthesis. A good example of a one-carbon transfer can be found in the reactions that produce the amino acids of the serine family. This family also includes glycine and cysteine. Serine and glycine themselves are frequently precursors in other biosynthetic pathways. A discussion of the synthesis of cysteine will give us some insight into the metabolism of sulfur, as well as that of nitrogen.



■ **FIGURE 23.10** The biosynthesis of serine.

The ultimate precursor of serine is 3-phosphoglycerate, which is obtainable from the glycolytic pathway. The hydroxyl group on carbon 2 is oxidized to a keto group, giving an  $\alpha$ -keto acid. A transamination reaction in which glutamate is the nitrogen donor produces 3-phosphoserine and  $\alpha$ -ketoglutarate. Hydrolysis of the phosphate group then gives rise to serine (Figure 23.10).

The conversion of serine to glycine involves the transfer of a one-carbon unit from serine to an acceptor. This reaction is catalyzed by *serine hydroxymethylase*, with pyridoxal phosphate as a coenzyme. The acceptor in this reaction is **tetrahydrofolate**, a derivative of folic acid and a frequently encountered carrier of one-carbon units in metabolic pathways. Its structure has three parts: a substituted pteridine ring, *p*-aminobenzoic acid, and glutamic acid (Figure 23.11). Folic acid is a vitamin that has been identified as essential in preventing birth defects; consequently, it is now a recommended supplement for all women of child-bearing age. There is also some evidence that folic acid may prevent heart disease in both men and women over age 50.

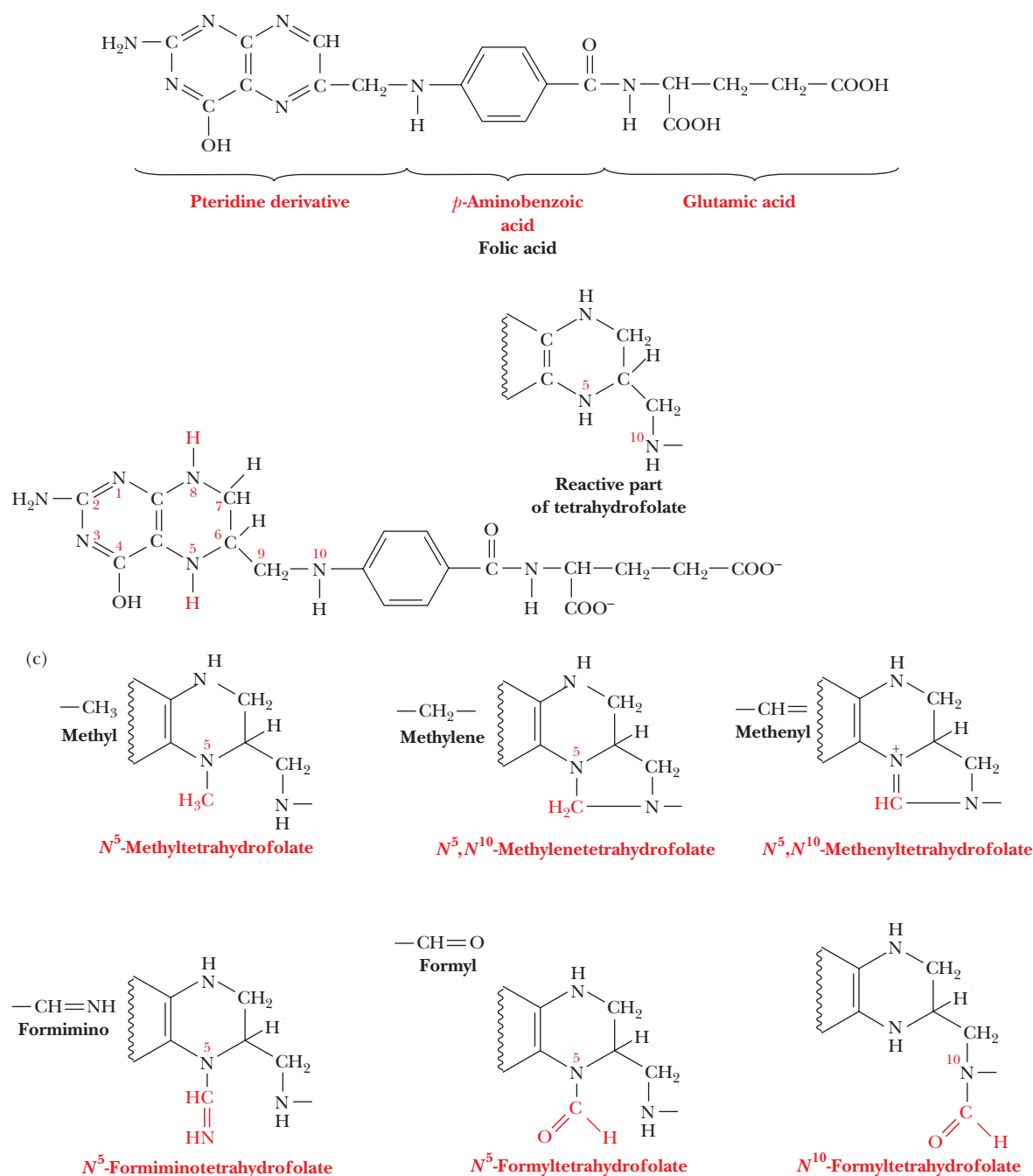


The one-carbon unit transferred in this reaction is bound to tetrahydrofolate, forming  $\text{N}^5, \text{N}^{10}$ -methylenetetrahydrofolate, in which the methylene (one-carbon) unit is bound to two of the nitrogens of the carrier (Figure 23.12). Tetrahydrofolate is not the only carrier of one-carbon units. We have already encountered biotin, a carrier of  $\text{CO}_2$ , and we have discussed the role that biotin plays in gluconeogenesis (Section 18.2) and in the anabolism of fatty acids (Section 21.6).

The conversion of serine to cysteine involves some interesting reactions. The source of the sulfur in animals differs from that in plants and bacteria. In plants and bacteria, serine is acetylated to form *O*-acetylserine. This reaction is catalyzed by *serine acyltransferase*, with acetyl-CoA as the acyl donor (Figure 23.13). Conversion of *O*-acetylserine to cysteine requires production of sulfide by a sulfur donor. The sulfur donor for plants and bacteria is 3'-phospho-5'-adenylyl sulfate. The sulfate group is reduced first to sulfite and then to sulfide (Figure 23.14). The sulfide, in the conjugate acid form  $\text{HS}^-$ , displaces the acetyl group of the *O*-acetylserine to produce cysteine. Animals form cysteine from serine by a different pathway because they do not have the enzymes to carry out the sulfate-to-sulfide conversion that we have just seen. The reaction sequence in animals involves the amino acid methionine.

Methionine, which is produced by reactions of the aspartate family (see the BiochemistryNow Interactive website at [academic.cengage.com/login](http://academic.cengage.com/login)) in bacteria and plants, cannot be produced by animals. It must be obtained from dietary sources. It is an **essential amino acid** because it cannot be synthesized by the body. The ingested methionine reacts with ATP to form **S-adenosylmethionine (SAM)**, which has a highly reactive methyl group (Figure 23.15). This compound is a carrier of methyl groups in many reactions. The methyl group from S-adenosylmethionine can be transferred to any one of a number of acceptors, producing S-adenosylhomocysteine. Hydrolysis of S-adenosylhomocysteine in turn produces homocysteine. Cysteine can be synthesized from serine and homocysteine, and this pathway for cysteine biosynthesis is the only one available to animals (Figure 23.16). Serine and homocysteine react to produce cystathionine, which hydrolyzes to form cysteine,  $\text{NH}_4^+$ , and  $\alpha$ -ketobutyrate.

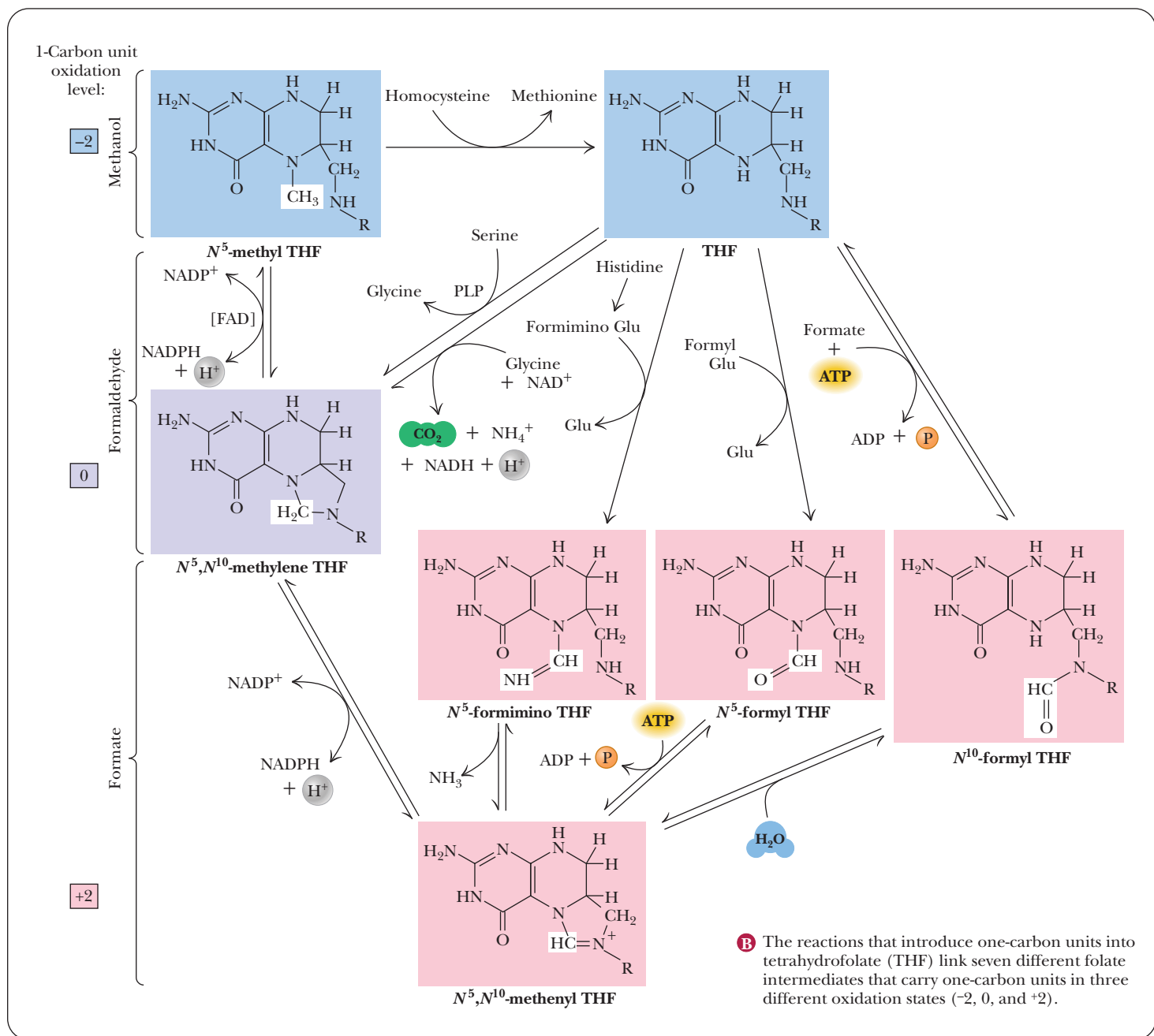
It is worth noting that we have now seen three important carriers of one-carbon units: biotin, a carrier of  $\text{CO}_2$ ; tetrahydrofolate ( $\text{FH}_4$ ), a carrier of methylene and formyl groups; and S-adenosylmethionine, a carrier of methyl groups.



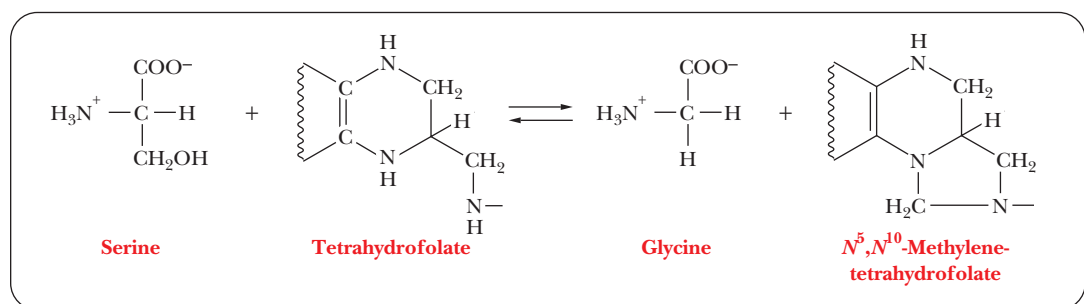
**A** The structure of folic acid, shown in nonionized form. The structures of folic acid derivatives with one-carbon units attached are shown in ionized form.

**FIGURE 23.11 Structure and reactions of folic acid.** (a) The structure of folic acid, shown in non-ionized form. (b) The reactions that introduce one-carbon units into tetrahydrofolate (THF) link seven different folate intermediates that carry one-carbon units in three different oxidation states (−2, 0, and +2). (Adapted from T. Brody *et al.*, in L. J. Machlin. Handbook of Vitamins. New York: Marcel Dekker, 1984.)

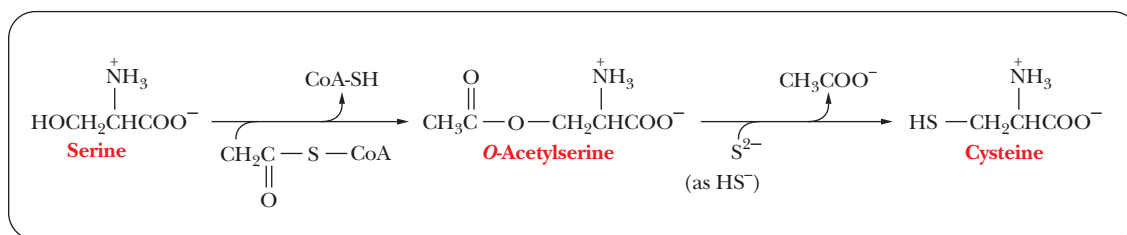
(continued)



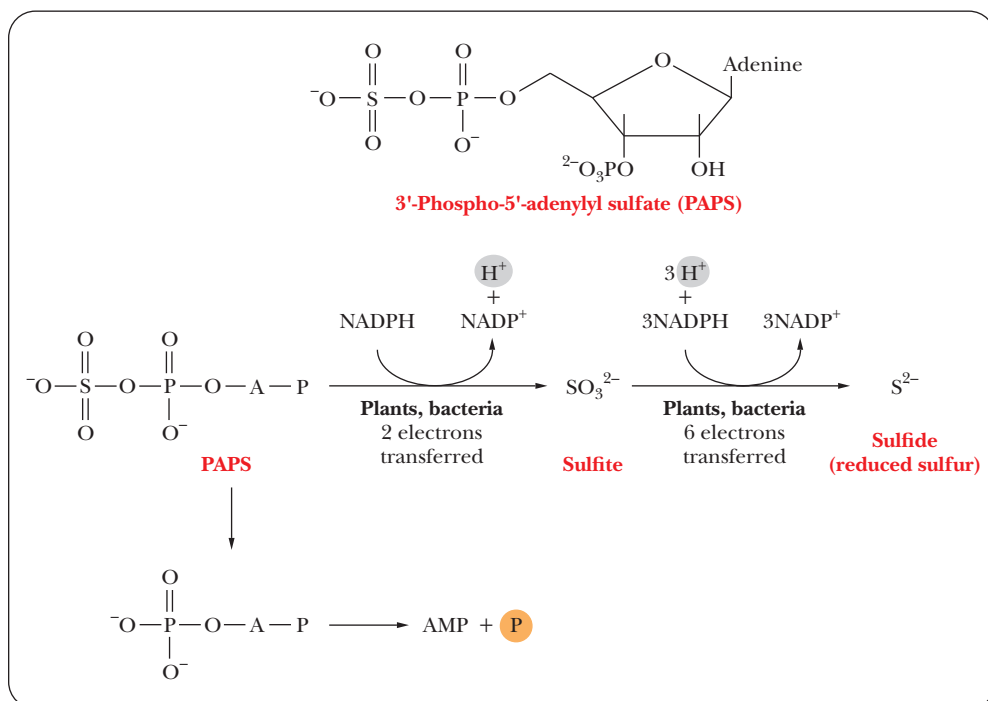
■ FIGURE 23.11—continued



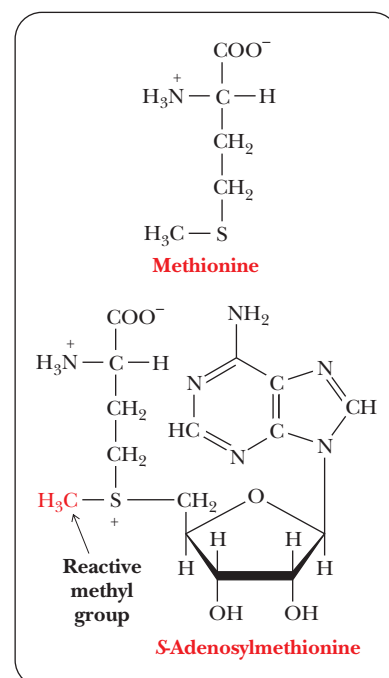
■ FIGURE 23.12 The conversion of serine to glycine, showing the role of tetrahydrofolate.



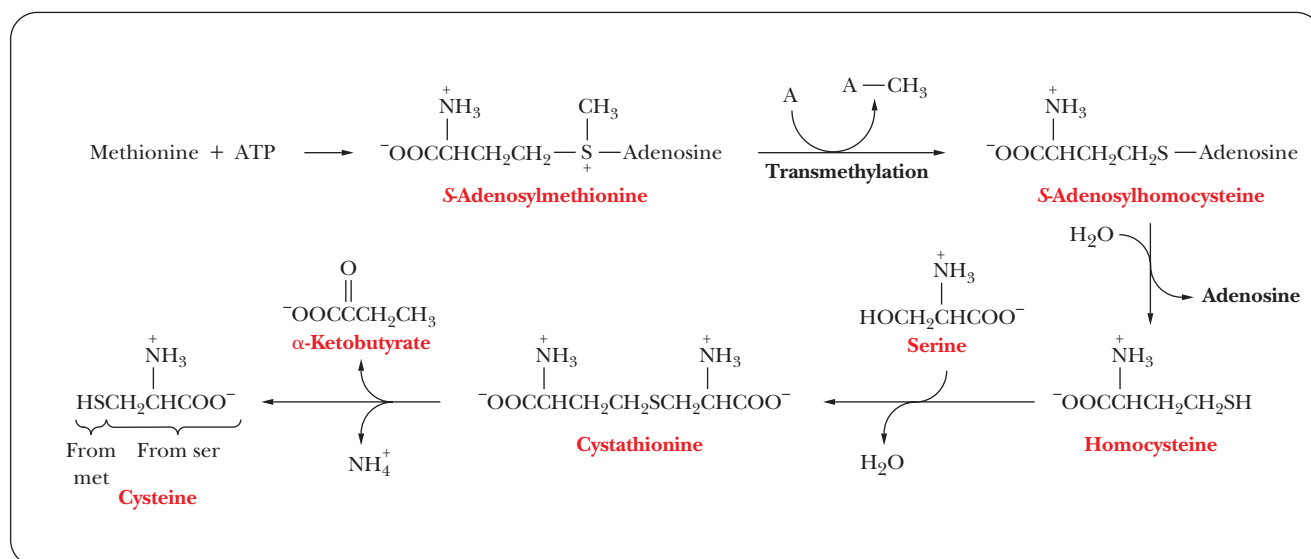
■ **FIGURE 23.13** The biosynthesis of cysteine in plants and bacteria.



■ **FIGURE 23.14** Electron transfer reactions of sulfur in plants and bacteria.



■ **FIGURE 23.15** The structure of S-adenosylmethionine (SAM), with the structure of methionine shown for comparison.



■ **FIGURE 23.16** The biosynthesis of cysteine in animals. (A stands for *acceptor*.)



TABLE 23.1

Amino Acid Requirements in Humans	
Essential	Nonessential
Arginine*	Alanine
Histidine†	Asparagine
Isoleucine	Aspartate
Leucine	Cysteine
Lysine	Glutamate
Methionine	Glutamine
Phenylalanine	Glycine
Threonine	Proline
Tryptophan	Serine
Valine	Tyrosine

\* Mammals synthesize arginine but cleave most of it to urea (Section 23.6).

† Essential for children, but not necessarily for adults.

23.5 Essential Amino Acids

The biosynthesis of proteins requires the presence of all the constituent amino acids. If one of the 20 amino acids is missing or in short supply, protein biosynthesis is inhibited. Some organisms, such as *Escherichia coli*, can synthesize all the amino acids they need. Other species, including humans, must obtain some amino acids from dietary sources. The essential amino acids in human nutrition are listed in Table 23.1. The body can synthesize some of these amino acids, but not in sufficient quantities for its needs, especially in the case of growing children. This last point applies particularly to children’s requirement for arginine and histidine. Amino acids are not stored (except in proteins), and dietary sources of essential amino acids are needed at regular intervals. Protein deficiency, especially a prolonged deficiency in sources that contain essential amino acids, leads to the disease **kwashiorkor**. The problem in this disease, particularly severe in growing children, is not only starvation but the breakdown of the body’s own proteins.

23.6 Amino Acid Catabolism

When we specifically focus on the catabolism of amino acids, the first step we consider is the removal of nitrogen by transamination. Transamination reactions are also important in the anabolism of amino acids, so it is important to remind ourselves that anabolic and catabolic pathways are not the exact reverse of each other, nor do they involve exactly the same group of enzymes. In catabolism, the amino nitrogen of the original amino acid is transferred to  $\alpha$ -ketoglutarate to produce glutamate, leaving behind the carbon skeletons. The fates of the carbon skeleton and of the nitrogen can be considered separately.

What is the fate of the carbon skeleton in amino acid breakdown?

Breakdown of the carbon skeletons of amino acids follows two general pathways, the difference between the two pathways depending on the type of end product. A **glucogenic** amino acid yields pyruvate or oxaloacetate on degradation. Oxaloacetate is the starting point for the production of glucose by gluconeogenesis. A **ketogenic** amino acid breaks down to acetyl-CoA or acetoacetyl-CoA, leading to the formation of ketone bodies (Table 23.2; see also Section 21.5). The carbon skeletons of the amino acids give rise to metabolic intermediates such as pyruvate, acetyl-CoA, acetoacetyl-CoA,  $\alpha$ -ketoglutarate, succinyl-CoA, fumarate, and oxaloacetate (see Figure 23.7). Oxaloacetate is

TABLE 23.2

Glucogenic and Ketogenic Amino Acids		
Glucogenic	Ketogenic	Glucogenic and Ketogenic
Aspartate	Leucine	Isoleucine
Asparagine	Lysine	Phenylalanine
Alanine		Tryptophan
Glycine		Tyrosine
Serine		
Threonine		
Cysteine		
Glutamate		
Glutamine		
Arginine		
Proline		
Histidine		
Valine		
Methionine		

a key intermediate in the breakdown of the carbon skeletons of amino acids because of its dual role in the citric acid cycle and in gluconeogenesis. The amino acids degraded to acetyl-CoA and acetoacetyl-CoA are used in the citric acid cycle, but mammals cannot synthesize glucose from acetyl-CoA. This fact is the source of the distinction between glucogenic and ketogenic amino acids. Glucogenic amino acids can be converted to glucose, with oxaloacetate as an intermediate, but ketogenic amino acids cannot be converted to glucose. Some amino acids have more than one pathway for catabolism, which explains why four of the amino acids are listed as both glucogenic and ketogenic.

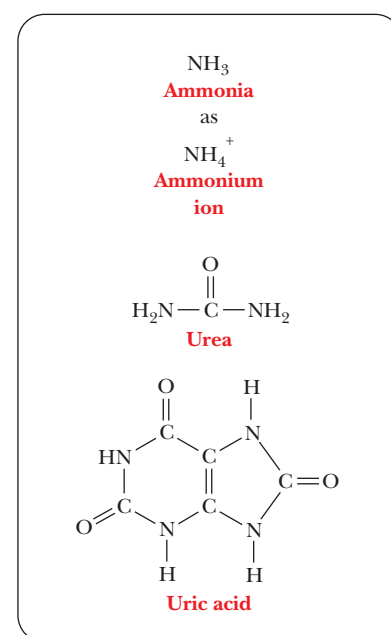
### Excretion of Excess Nitrogen

The nitrogen portion of amino acids is involved in transamination reactions in breakdown as well as in biosynthesis. Excess nitrogen is excreted in one of three forms: *ammonia* (as ammonium ion), *urea*, and *uric acid* (Figure 23.17).

Animals, such as fish, that live in an aquatic environment excrete nitrogen as ammonia; they are protected from the toxic effects of high concentrations of ammonia not only by the removal of ammonia from their bodies but also by rapid dilution of the excreted ammonia by the water in the environment. The principal waste product of nitrogen metabolism in terrestrial animals is urea (a water-soluble compound); its reactions provide some interesting comparisons with the citric acid cycle. Birds excrete nitrogen in the form of uric acid, which is insoluble in water. They do not have to carry the excess weight of water, which could hamper flight, to rid themselves of waste products.

### What is the role of the urea cycle in amino acid breakdown?

A central pathway in nitrogen metabolism is the **urea cycle** (Figure 23.18). The nitrogens that enter the urea cycle come from several sources. One of the nitrogens of urea is added in the mitochondria, and its immediate precursor is glutamate, which releases ammonia via glutamate dehydrogenase. However, the ammonia nitrogens of glutamate have ultimately come from many sources as a result of transamination reactions. Mitochondrial glutaminase also provides free ammonia that can enter the cycle. A condensation reaction between the ammonium ion and carbon dioxide produces **carbamoyl phosphate** in a reaction that requires the hydrolysis of two molecules of ATP for each molecule of carbamoyl phosphate. Carbamoyl phosphate reacts with **ornithine** (Step 1) to form **citrulline**. Citrulline is then transported to the cytosol. A second nitrogen enters the urea cycle when aspartate reacts with citrulline to form



■ **FIGURE 23.17** Nitrogen-containing products of amino acid catabolism.

■ **FIGURE 23.18 The urea cycle series of reactions.** Transfer of the carbamoyl group of carbamoyl-P to ornithine by ornithine transcarbamoylase (OTCase, reaction 1) yields citrulline. The citrulline ureido group is then activated by reaction with ATP to give a citrullyl—AMP intermediate (reaction 2a); AMP is then displaced by aspartate, which is linked to the carbon framework of citrulline via its  $\alpha$ -amino group (reaction 2b). The course of reaction 2 was verified using  $^{18}\text{O}$ -labeled citrulline. The  $^{18}\text{O}$  label (indicated by the asterisk, \*) was recovered in AMP. Citrulline and AMP are joined via the ureido \*O atom. The product of this reaction is argininosuccinate; the enzyme catalyzing the two steps of reaction 2 is argininosuccinate synthetase. The next step (reaction 3) is carried out by argininosuccinase, which catalyzes the nonhydrolytic removal of fumarate from argininosuccinate to give arginine. Hydrolysis of arginine by arginase (reaction 4) yields urea and ornithine, completing the urea cycle.

## Biochemical Connections PHYSIOLOGY

### Water and the Disposal of Nitrogen Wastes

Ammonia gas is toxic to most organisms and must usually be disposed of rapidly. In a certain sense, one can almost guess the mechanism of nitrogen-waste disposal if one knows the amount of water available to the organism in question. For example, bacteria and fish, which live in “infinite” water supplies, usually simply release ammonia into the medium, where organisms that are lower on the evolutionary scale can use it. Fish sometimes produce trimethylamine, another highly water-soluble compound, which is the characteristic “fish odor.” Most terrestrial animals do not have “infinite” water supplies, but mammals, which have bladders, usually live in conditions where adequate water is available. Their mechanism for disposal of most toxins is to prepare a water-soluble compound and then to excrete it through the urine. Thus, urea becomes a major by-product of nitrogen metabolism in mammals.

Reptiles and other desert animals do not usually have much water available, and birds cannot afford to carry the weight of a fluid-filled bladder.

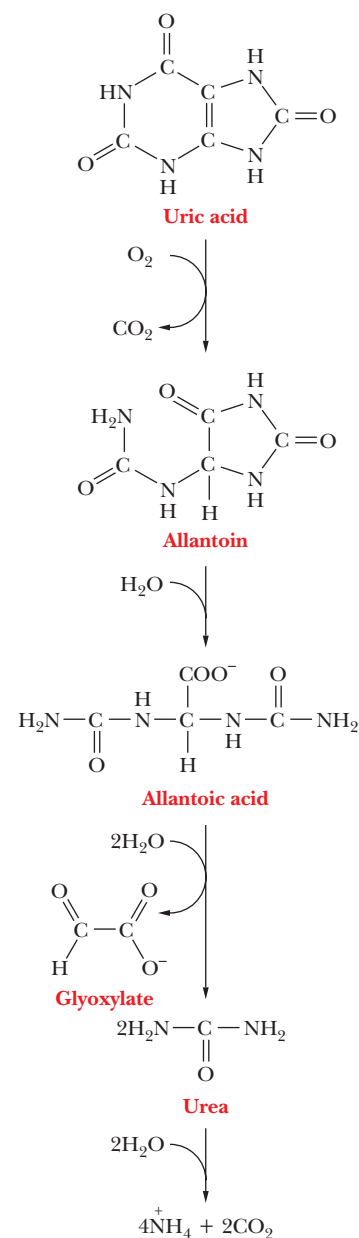
These animals do not make urea; rather, they convert all their waste nitrogen to uric acid (Figure 23.17), the concentrated white solid so familiar in bird droppings. Some desert mammals, such as the kangaroo rat, which never drinks water but rather lives off metabolic water, also convert some of their waste nitrogen to uric acid to conserve the water used in urine.

Uric acid, the typical waste product from purines, can cause problems in primates due to its marginal water solubility. Deposits of uric acid in the joints and extremities cause gout (Section 23.8). Other mammals do not have a problem with uric acid because they convert it to allantoin, which is very water-soluble.

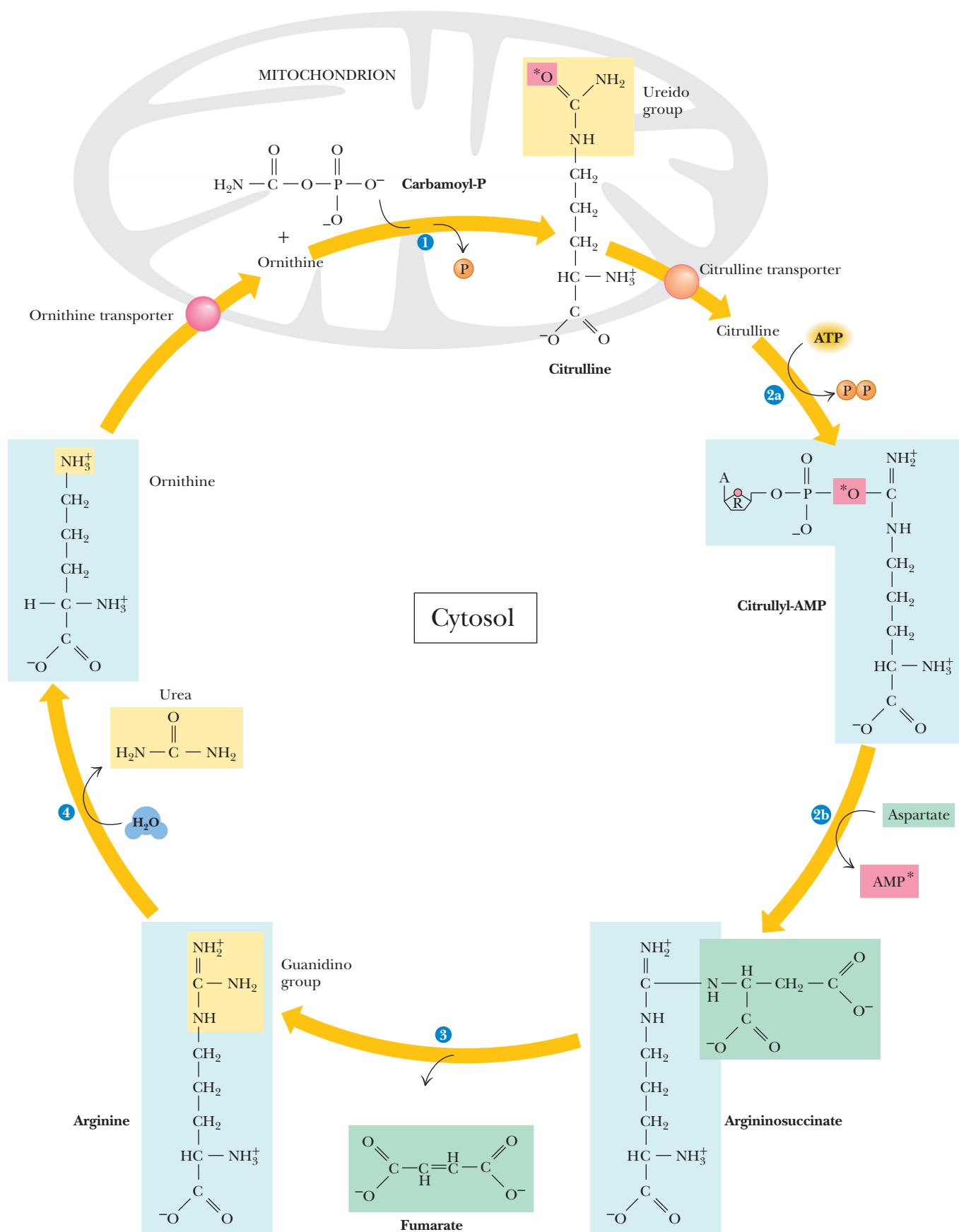


Joe McDonald/Visuals Unlimited

■ A kangaroo rat converts some of its waste nitrogen to uric acid.

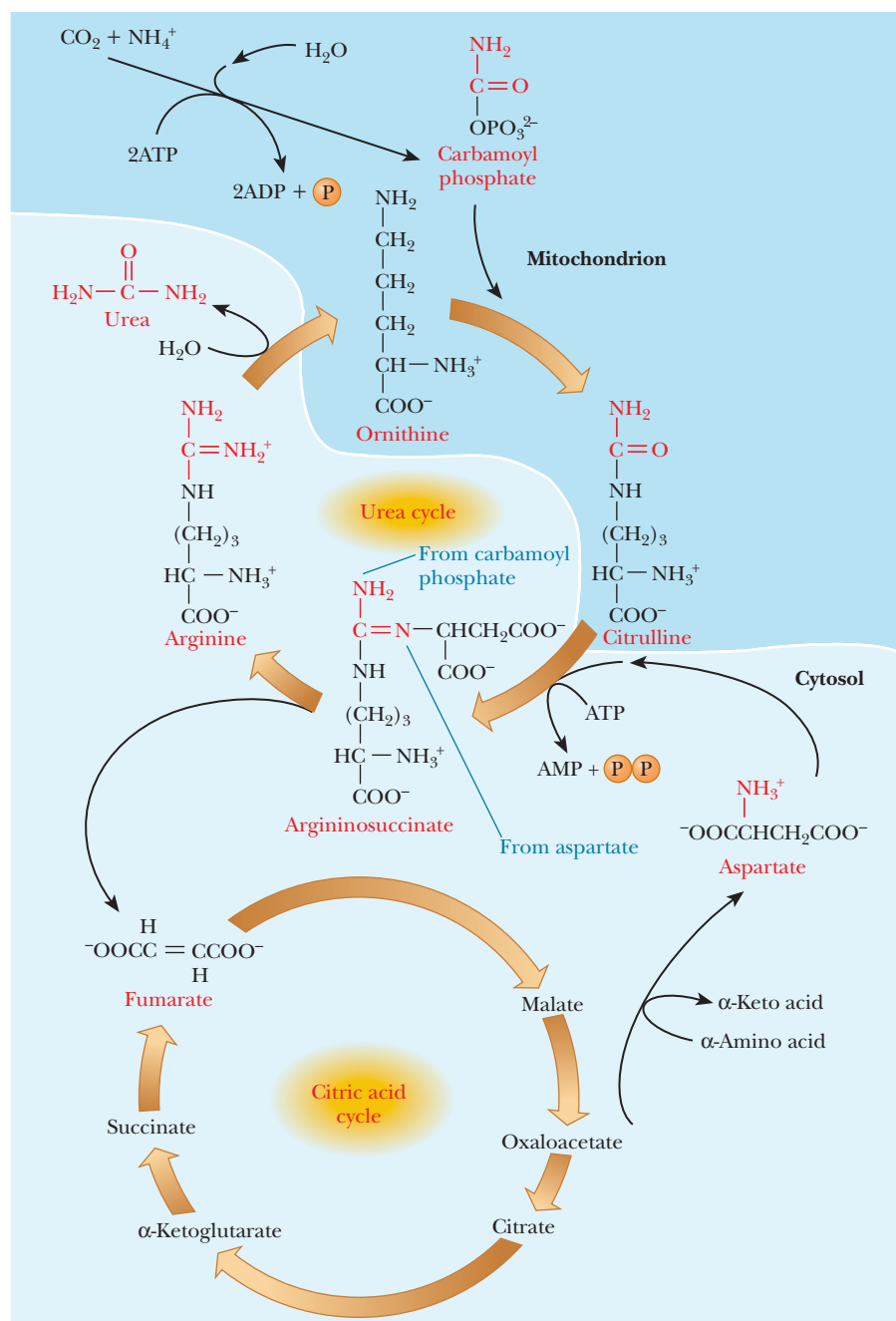


■ Catabolism of uric acid to ammonia and  $\text{CO}_2$ .



**argininosuccinate** in another reaction that requires ATP (AMP and  $\text{PP}_i$  are produced in this reaction; Step 2). The amino group of the aspartate is the source of the second nitrogen in the urea that will be formed in this series of reactions. Argininosuccinate is split to produce **arginine** and **fumarate** (Step 3). Finally, arginine is hydrolyzed to give urea and to regenerate ornithine, which is transported back to the mitochondrion (Step 4). The biosynthesis of arginine from ornithine is discussed on the Biochemistry Interactive website. Another way of looking at the urea cycle is to consider arginine as the immediate precursor of urea and to see it as producing ornithine in the process. According to this point of view, the rest of the cycle is the regeneration of arginine from ornithine.

The synthesis of fumarate is a link between the urea cycle and the citric acid cycle. Fumarate is, of course, an intermediate of the citric acid cycle, and it can be converted to oxaloacetate. A transamination reaction can convert oxaloacetate to aspartate, providing another link between the two cycles (Figure 23.19).



■ **FIGURE 23.19** The urea cycle and some of its links to the citric acid cycle. Part of the cycle takes place in the mitochondrion and part in the cytosol. Fumarate and aspartate are the direct links to the citric acid cycle. Fumarate is a citric acid cycle intermediate. Aspartate comes from transamination of oxaloacetate, which is also a citric acid cycle intermediate.

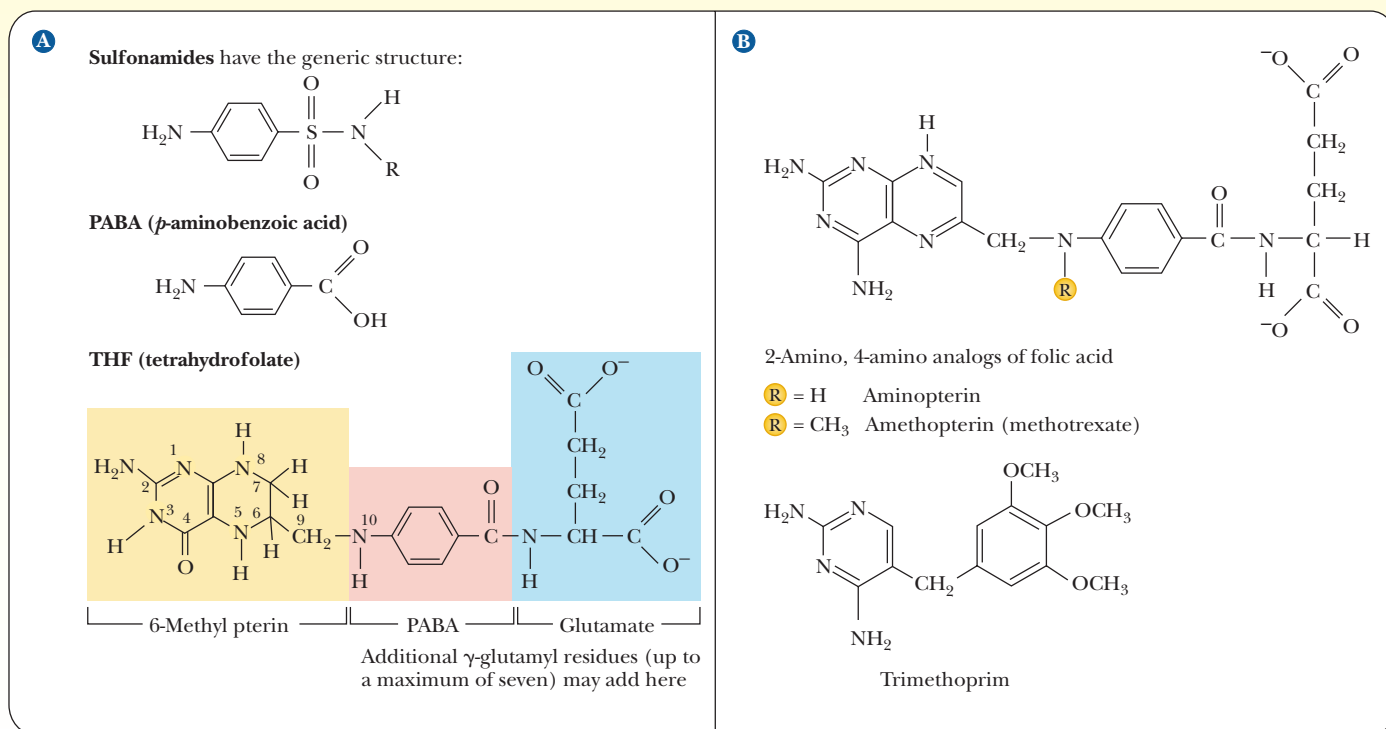


# Biochemical Connections MEDICINE

## Chemotherapy and Antibiotics—Taking Advantage of the Need for Folic Acid

We have already seen the importance of folic acid and its derivative, tetrahydrofolate, in several reactions. This importance has been exploited in human medicine. Bacteria synthesize folic acid from *p*-aminobutyric acid (PABA). A type of antibiotic called a sulfonamide (Figure a) works by competing with PABA in the synthesis of folic acid. Because folic acid is critical to the formation

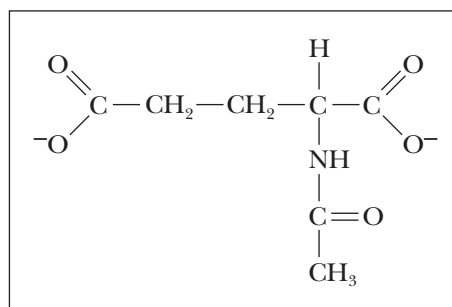
of purines, antagonists of folic acid metabolism are used to inhibit nucleic acid synthesis and cell growth. Rapidly dividing cells, such as those found in cancer and tumors, are more susceptible to these antagonists. Several related compounds, such as methotrexate (Figure b), are used in chemotherapy to inhibit cancer cell growth.



■ (a) Sulfa drugs (sulfonamides) act as antibiotics because of their similarity to *p*-aminobenzoic acid (PABA), a precursor of folic acid synthesis. Sulfa drugs compete with PABA and stop folic acid synthesis in bacteria. (b) Three compounds that are used for chemotherapy because they interfere with folic acid metabolism. They are almost irreversible inhibitors of dihydrofolate reductase, having a 1000-fold greater affinity than dihydrofolate.

In fact, both pathways were discovered by the same person, Hans Krebs. Four high-energy phosphate bonds are required because of the production of pyrophosphate in the conversion of aspartate to argininosuccinate.

In humans, urea synthesis is used to excrete excess nitrogen, such as would be found after consuming a high-protein meal. The pathway is confined to the liver. Note that arginine, the immediate precursor to urea, is the most nitrogen-rich amino acid, but the source of the nitrogen in the arginine varies. The major control point is the mitochondrial enzyme *carbamoyl-phosphate synthetase I* (CPS-I), and the formation of carbamoyl-phosphate is the committed step in the urea cycle. CPS-I is allosterically activated by *N*-acetylglutamate:



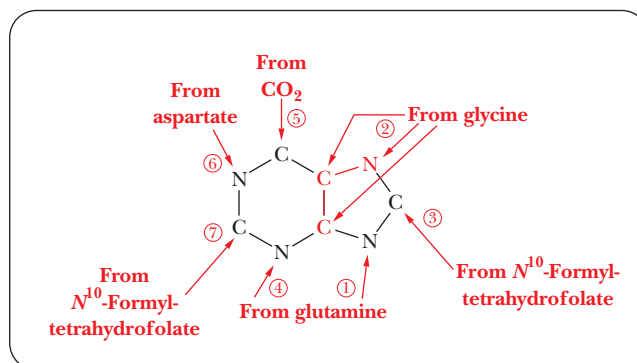
*N*-Acetylglutamate is formed by a reaction between glutamate and acetyl-CoA, which is catalyzed by *N*-acetylglutamate synthase. This enzyme is activated by increased concentrations of arginine. Thus, when amino acid catabolism is high, large amounts of glutamate are present from degradation of glutamine, from synthesis via glutamate dehydrogenase, and from transamination reactions. Increased glutamate levels lead to increased levels of *N*-acetylglutamate followed by increasing the activity of the urea cycle. In addition, any time arginine builds up, either because of protein catabolism or because ornithine is building up because of a low level of CPS-I activity, the arginine stimulates synthesis of *N*-acetylglutamate and therefore increases the CPS-I activity.

## 23.7 Purine Biosynthesis

We have already discussed the formation of ribose-5-phosphate as part of the pentose phosphate pathway (Section 18.4). The biosynthetic pathway for both purine and pyrimidine nucleotides uses preformed ribose-5-phosphate. Purines and pyrimidines are synthesized in different ways, and we shall consider them separately.

### Anabolism of Inosine Monophosphate

In the synthesis of purine nucleotides, the growing ring system is bonded to the ribose phosphate while the purine skeleton is being assembled—first the five-membered ring and then the six-membered ring—eventually producing inosine-5'-monophosphate. All four nitrogen atoms of the purine ring are derived from amino acids: two from glutamine, one from aspartate, and one from glycine. Two of the five carbon atoms (adjacent to the glycine nitrogen) also come from glycine, two more come from tetrahydrofolate derivatives, and the fifth comes from CO<sub>2</sub> (Figure 23.20). The series of reactions producing inosine monophosphate (IMP) is long and complex.



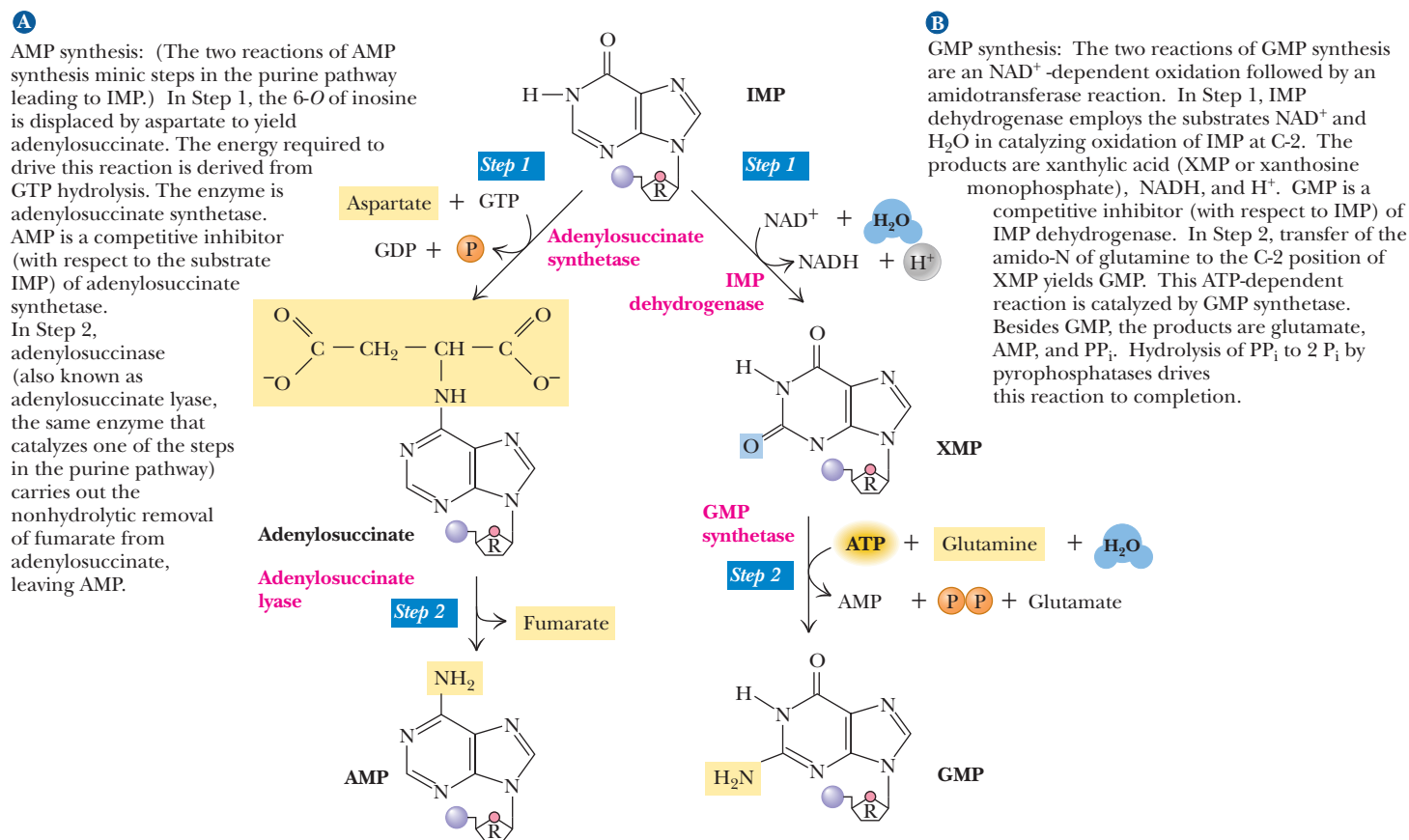
■ **FIGURE 23.20** Sources of the atoms in the purine ring in purine nucleotide biosynthesis. The numbering system indicates the order in which each atom, or group of atoms, is added.

### How is inosine monophosphate converted to AMP and GMP?

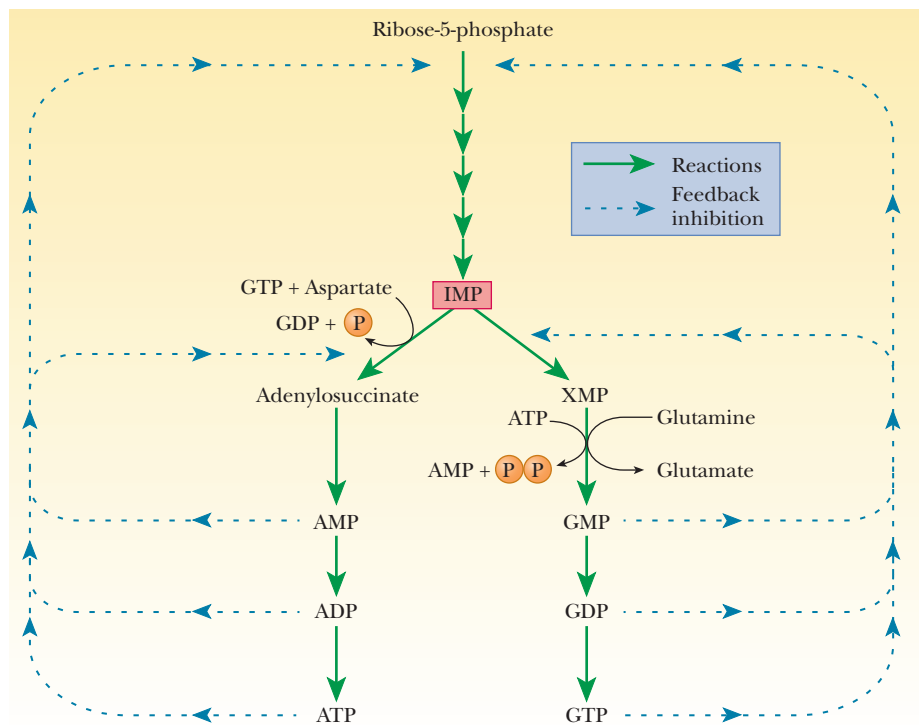
IMP is the precursor of both AMP and GMP. The conversion of IMP to AMP takes place in two stages (Figure 23.21). The first step is the reaction of aspartate with IMP to form adenylosuccinate. This reaction is catalyzed by adenylosuccinate synthetase and requires GTP, not ATP, as an energy source (using ATP would be counterproductive). The cleavage of fumarate from adenylosuccinate to produce AMP is catalyzed by adenylosuccinase. This enzyme also functions in the synthesis of the six-membered ring of IMP.

The conversion of IMP to GMP also takes place in two stages (Figure 23.21). The first of the two steps is an oxidation in which the C—H group at the C-2 position is converted to a keto group. The oxidizing agent in the reaction is  $\text{NAD}^+$ , and the enzyme involved is IMP dehydrogenase. The nucleotide formed by the oxidation reaction is xanthosine-5'-phosphate (XMP). An amino group from the side chain of glutamine replaces the C-2 keto group of XMP to produce GMP. This reaction is catalyzed by GMP synthetase; ATP is hydrolyzed to AMP and  $\text{PP}_i$  in the process. Note that there is some control over the relative levels of purine nucleotides; GTP is needed for the synthesis of adenine nucleotides, whereas ATP is required for the synthesis of guanine nucleotides. Each of the purine nucleotides must occur at a reasonably high level for the other to be synthesized.

Subsequent phosphorylation reactions produce purine nucleoside diphosphates (ADP and GDP) and triphosphates (ATP and GTP). The purine nucleoside monophosphates, diphosphates, and triphosphates are all feedback inhibitors of the first stages of their own biosynthesis. Also, AMP, ADP, and ATP inhibit the conversion of IMP to adenine nucleotides, and GMP, GDP, and GTP inhibit the conversion of IMP to xanthylate and to guanine nucleotides (Figure 23.22).



■ **FIGURE 23.21** The synthesis of AMP and GMP from IMP.



**FIGURE 23.22** The role of feedback inhibition in regulation of purine nucleotide biosynthesis.

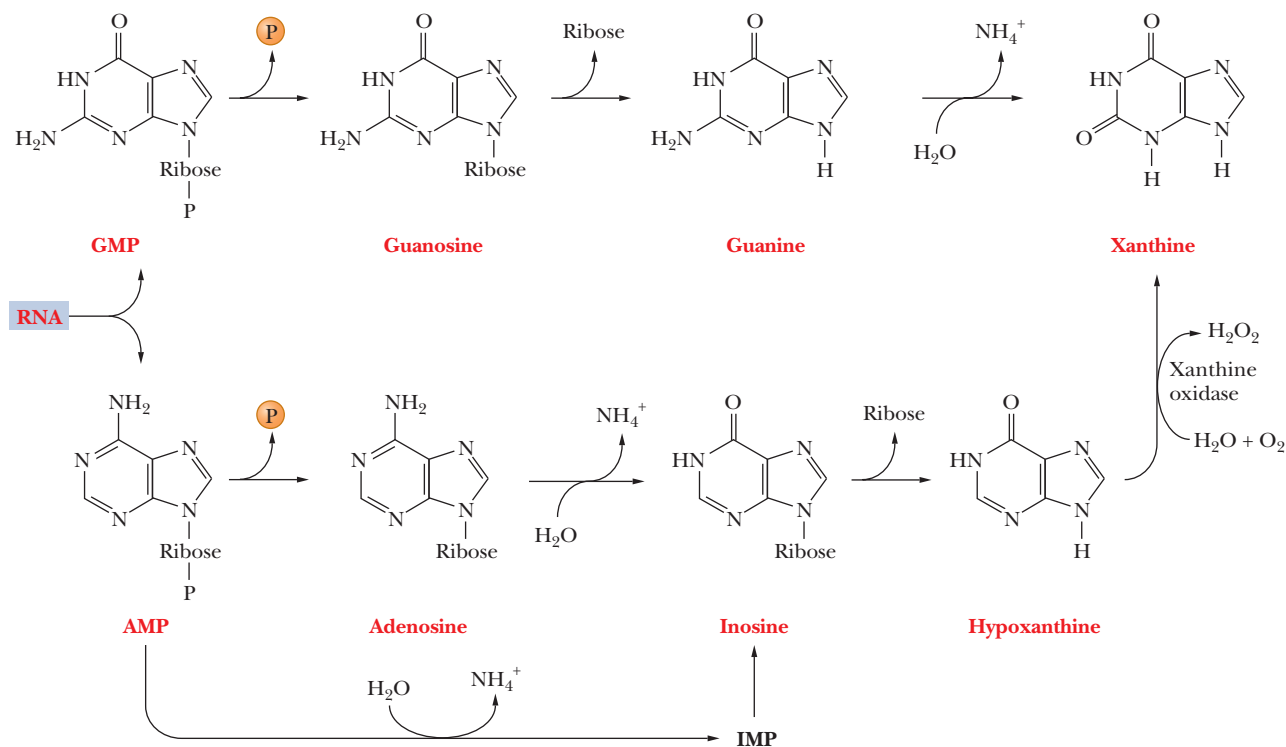
### What are the energy requirements for production of AMP and GMP?

The production of IMP starting with ribose-5-phosphate requires the equivalent of 7 ATP (see the Biochemistry Interactive website). The conversion of IMP to AMP requires hydrolysis of an additional high-energy bond—in this case, that of GTP. In the formation of AMP from ribose-5-phosphate, the equivalent of 8 ATP is needed. The conversion of IMP to GMP requires two high-energy bonds, given that a reaction occurs in which ATP is hydrolyzed to AMP and PP<sub>i</sub>. For the production of GMP from ribose-5-phosphate, the equivalent of 9 ATP is necessary. The anaerobic oxidation of glucose produces only 2 ATP for each molecule of glucose (Section 17.1). Anaerobic organisms require four molecules of glucose (which produce 8 ATP) for each AMP they form, or five molecules of glucose (which produce 10 ATP) for each GMP. The process is more efficient for aerobic organisms. Since 30 or 32 ATP result from each molecule of glucose, depending on the type of tissue, aerobic organisms can optimally produce four AMP (requiring 32 ATP) or three GMP (requiring 36 ATP) for each molecule of glucose oxidized. A mechanism for reuse of purines, rather than complete turnover and new synthesis, saves energy for organisms.

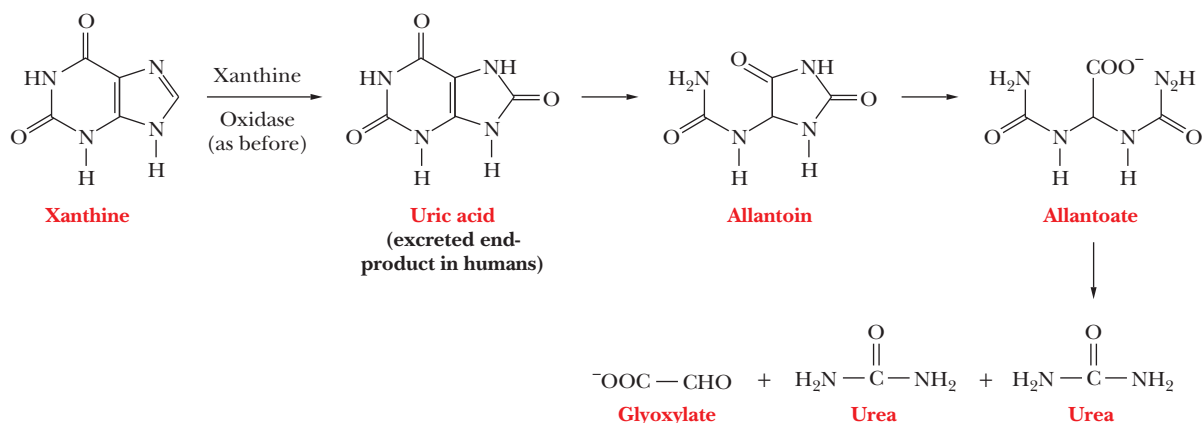
## 23.8 Purine Catabolism

The catabolism of purine nucleotides proceeds by hydrolysis to the nucleoside and subsequently to the free base, which is further degraded. Deamination of guanine produces xanthine, and deamination of adenine produces hypoxanthine, the base corresponding to the nucleoside inosine, which is shown in Figure 23.23a. Hypoxanthine can be oxidized to xanthine, so this base is a common degradation product of both adenine and guanine. Xanthine is oxidized in turn to **uric acid** (Section 23.6). In birds, some reptiles, insects, Dalmatian

**A** Purine nucleotides are converted to the free base and then to xanthine.



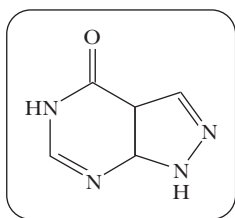
**B** Catabolic reactions of xanthine.



**FIGURE 23.23 The reactions of purine catabolism.** (a) Purine nucleotides are converted to the free base and then to xanthine. (b) Catabolic reactions of xanthine.

dogs, and primates (including humans), uric acid is the end product of purine metabolism and is excreted. In all other terrestrial animals, including all other mammals, allantoin is the product excreted, whereas allantoate is the product in fish. Allantoate is further degraded to glyoxylate and urea by microorganisms and some amphibians, as shown in Figure 23.23b. *Gout* is a disease in humans that is caused by the overproduction of uric acid. Deposits of uric acid (which is barely soluble in water) accumulate in the joints of the hands and feet. Allopurinol is a compound used to treat gout; it inhibits the degradation





■ Allopurinol, a substance used in the treatment of gout.

of hypoxanthine to xanthine and of xanthine to uric acid, preventing the buildup of uric acid deposits. The Biochemical Connections box on next page describes another aspect of gout.

**Salvage reactions** are important in the metabolism of purine nucleotides because of the amount of energy required for the synthesis of the purine bases. A free purine base that has been cleaved from a nucleotide can produce the corresponding nucleotide by reacting with the compound phosphoribosylpyrophosphate (PRPP), formed by a transfer of a pyrophosphate group from ATP to ribose-5-phosphate (Figure 23.24).

Two different enzymes with different specificities with respect to the purine base catalyze salvage reactions. The reaction



is catalyzed by adenine phosphoribosyltransferase. The corresponding reactions of guanine and hypoxanthine

**HGPRT**

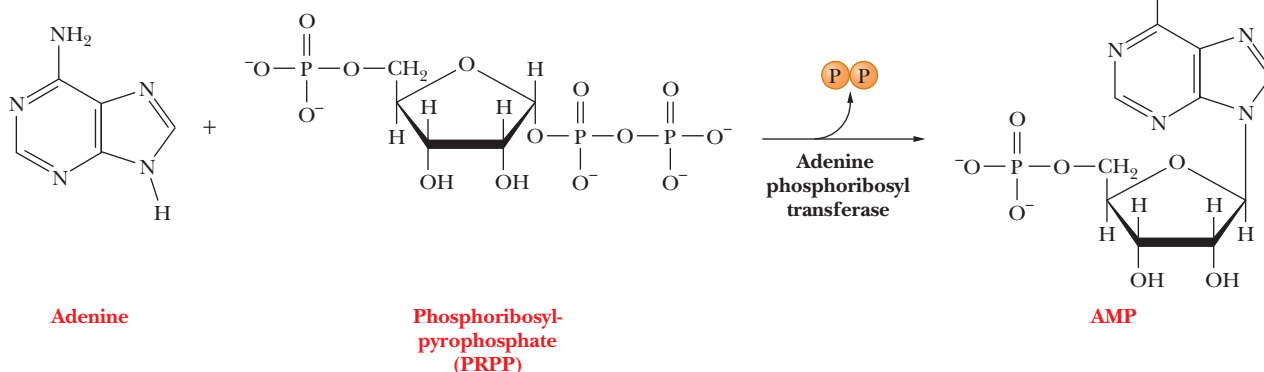


**HGPRT**

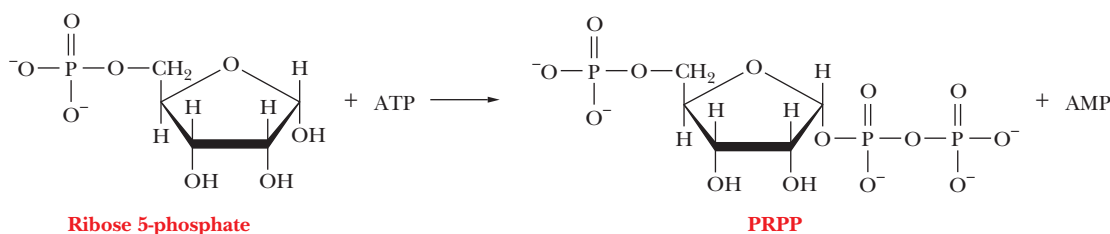


are catalyzed by hypoxanthine-guanine phosphoribosyltransferase (HGPRT) (Figure 23.25).

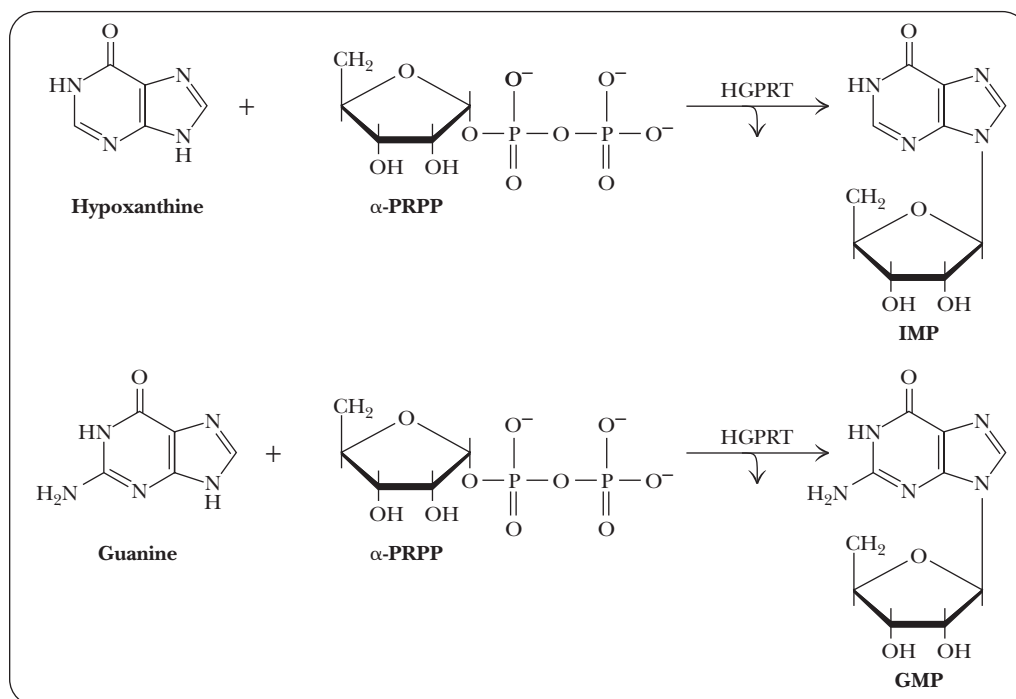
**A** Adenine is the purine in this example. There are analogous reactions for salvage of guanine and hypoxanthine.



**B** The formation of phosphoribosylpyrophosphate (PRPP).



■ **FIGURE 23.24 Purine salvage.** (a) Adenine is the purine in this example. There are analogous reactions for salvage of guanine and hypoxanthine (see page 677). (b) The formation of phosphoribosylpyrophosphate (PRPP).



■ **FIGURE 23.25** Purine salvage by the HGPRT reaction.

## Biochemical Connections CELL BIOLOGY

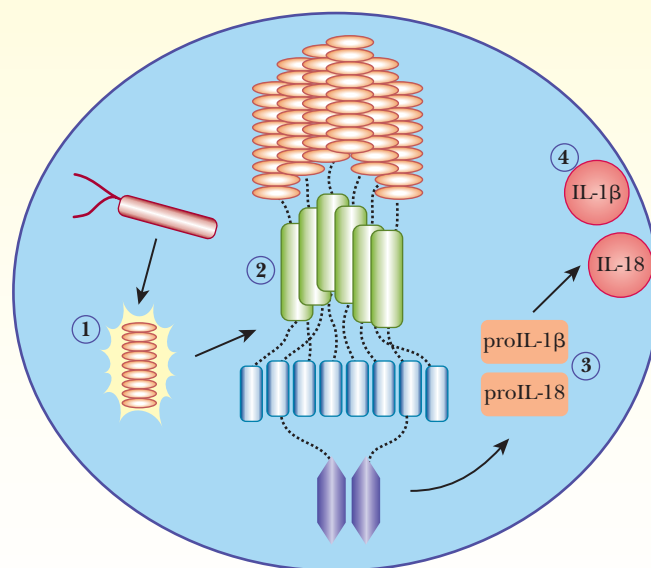
### Cytoplasmic Defenses against Gout

A cell's external defenses against pathogens are part of the innate immune system, which has been studied for some time. Another set of internal defenses, based in the cytoplasm, is now receiving attention. Cytoplasmic sensors respond to a trigger by setting off a chain of events that leads to inflammation.

One example of such a trigger is the presence of crystals of monosodium urate, a salt of uric acid. This insoluble material produces the symptoms of gout. A group of proteins called NLRs play a key role in the defense against this process. A specific member of this group, NLRP3, recognizes a pathogen, such as monosodium urate. This signal leads to the aggregation of more molecules of NLRP3, forming a structure called an inflammasome. The presence of the inflammasome activates the enzyme caspase-1 (shown in purple in the figure), which, in turn, activates the cytokines IL-1 $\beta$  and IL-18. These cytokines promote the development of inflammation.

The exact mechanism by which these proteins operate is not yet known. It is clear that they respond to a number of different

triggers in addition to monosodium urate. Asbestos fibers and portions of bacterial cell walls have also been observed to trigger an inflammatory response. A clearer understanding of the process can lead to new treatments for diseases, including common ones. It is known that nearly half of patients with Crohn's disease carry mutation in an NLR protein. It will be interesting to see what results come from future research along these lines.



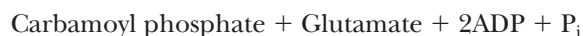
■ **The cytoplasmic alarm system based on the NLRP3 protein.**

The cytoplasmic sensor protein NLRP3 recognizes a pathogen (1). The sensor, which has been activated by the recognition, draws in more NLRP3 molecules and other kinds of proteins to form an inflammasome (2). After the enzyme caspase-1 (purple) is activated, it activates in turn the cytokines IL-1 $\beta$  and IL-18 (3). These cytokines go on to produce inflammation (4). (*From Science, vol. 326, p. 930. Copyright © 2009 by AAAS. Reprinted by permission of the AAAS.*)

## 23.9 Pyrimidine Biosynthesis and Catabolism

### The Anabolism of Pyrimidine Nucleotides

The overall scheme of pyrimidine nucleotide biosynthesis differs from that of purine nucleotides in that the pyrimidine ring is assembled before it is attached to ribose-5-phosphate. The carbon and nitrogen atoms of the pyrimidine ring come from carbamoyl phosphate and aspartate. The production of carbamoyl phosphate for pyrimidine biosynthesis takes place in the cytosol, and the nitrogen donor is glutamine. (We already saw a reaction for the production of carbamoyl phosphate when we discussed the urea cycle in Section 23.6. That reaction differs from this one because it takes place in mitochondria and the nitrogen donor is  $\text{NH}_4^+$ .)



The reaction of carbamoyl phosphate with aspartate to produce *N*-carbamoylaspartate is the committed step in pyrimidine biosynthesis. The compounds involved in reactions up to this point in the pathway can play other roles in metabolism; after this point, *N*-carbamoylaspartate can be used only to produce pyrimidines—thus the term “committed step.” This reaction is catalyzed by aspartate transcarbamoylase, which we discussed in detail in Chapter 7 as a prime example of an allosteric enzyme subject to feedback regulation. The next step, the conversion of *N*-carbamoylaspartate to dihydroorotate, takes place in a reaction that involves an intramolecular dehydration (loss of water) as well as cyclization. This reaction is catalyzed by dihydroorotase. Dihydroorotate is converted to orotate by dihydroorotate dehydrogenase, with the concomitant conversion of  $\text{NAD}^+$  to  $\text{NADH}$ . A pyrimidine nucleotide is now formed by the reaction of orotate with PRPP to give orotidine-5'-monophosphate (OMP), which is a reaction similar to the one that takes place in purine salvage (Section 23.8). Orotate phosphoribosyltransferase catalyzes this reaction. Finally, orotidine-5'-phosphate decarboxylase catalyzes the conversion of OMP to UMP (uridine-5'-monophosphate), which is the precursor of the remaining pyrimidine nucleotides (Figure 23.26).

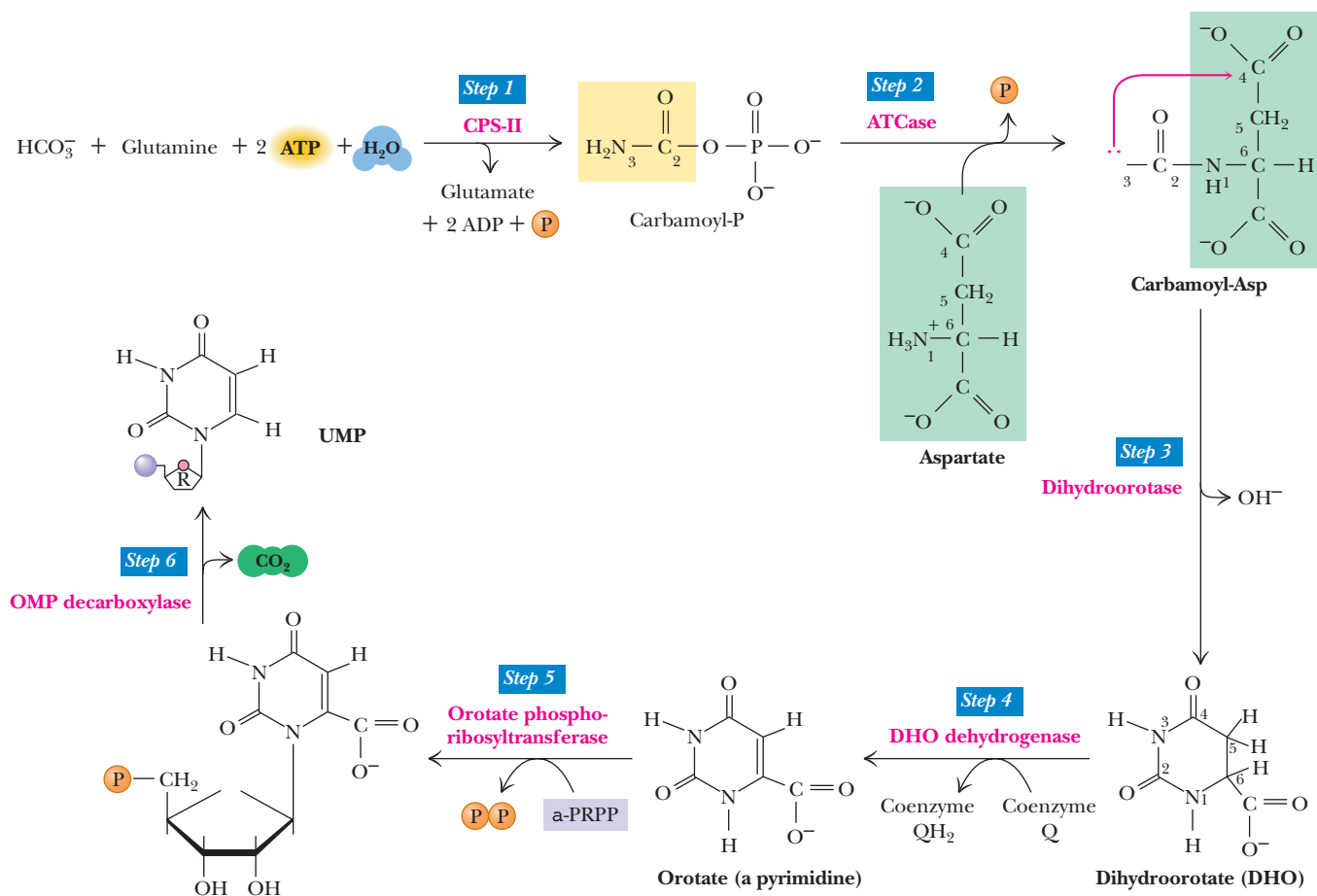
Two successive phosphorylation reactions convert UMP to UTP (Figure 23.27). The conversion of uracil to cytosine takes place in the triphosphate form, catalyzed by CTP synthetase (Figure 23.28). Glutamine is the nitrogen donor, and ATP is required, as we saw earlier in similar reactions.



Feedback inhibition in pyrimidine nucleotide biosynthesis takes place in several ways. CTP is an inhibitor of aspartate transcarbamoylase and of CTP synthetase. UMP is an inhibitor of an even earlier step, the one catalyzed by carbamoyl phosphate synthetase (Figure 23.29).

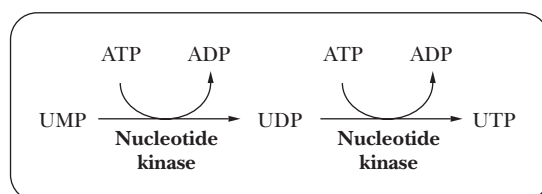
### Pyrimidine Catabolism

Pyrimidine nucleotides are broken down first to the nucleoside and then to the base, as purine nucleotides are. Cytosine can be deaminated to uracil, and the double bond of the uracil ring is reduced to produce dihydrouracil. The ring opens to produce *N*-carbamoylpropionate, which in turn is broken down to  $\text{NH}_4^+$ ,  $\text{CO}_2$ , and  $\beta$ -alanine (Figure 23.30).

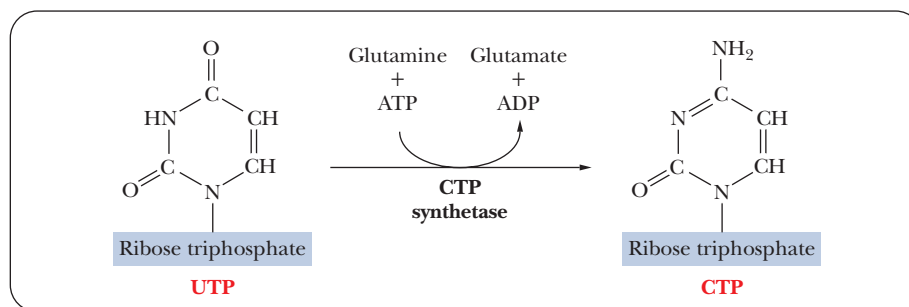


#### Orotidine-5'-Monophosphate (OMP)

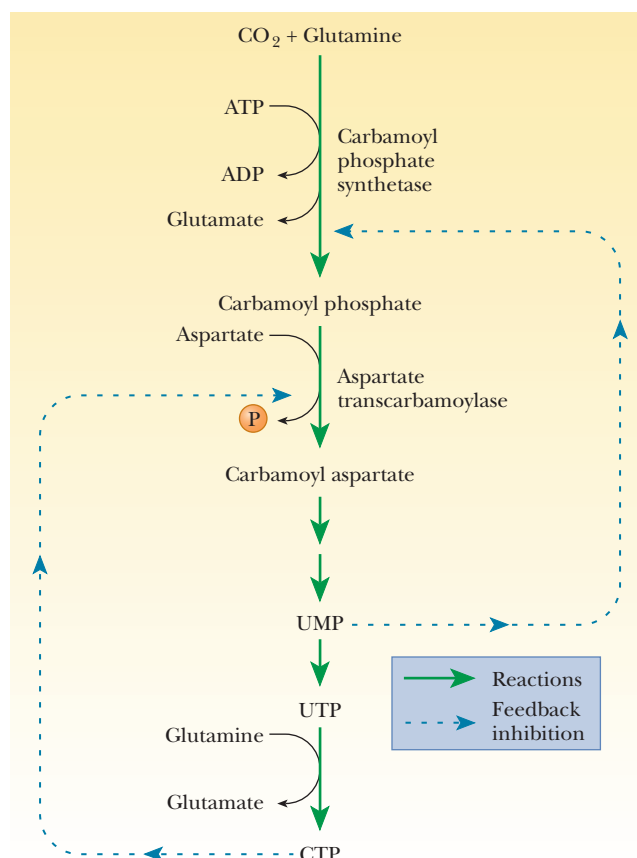
- **FIGURE 23.26** The pyrimidine biosynthetic pathway. Step 1: Carbamoyl-P synthesis. Step 2: Condensation of carbamoyl phosphate and aspartate to yield carbamoyl-aspartate is catalyzed by aspartate transcarbamoylase (ATCase). Step 3: An intramolecular condensation catalyzed by dihydroorotase gives the six-membered heterocyclic ring characteristic of pyrimidines. The product is dihydroorotate (DHO). Step 4: The oxidation of DHO by dihydroorotate dehydrogenase gives orotate. (In bacteria, NAD<sup>+</sup> is the electron acceptor from DHO.) Step 5: PRPP provides the ribose-5-P moiety that transforms orotate into orotidine-5-monophosphate, a pyrimidine nucleotide. Note that orotate phosphoribosyltransferase joins N-1 of the pyrimidine to the ribosyl group in appropriate  $\beta$ -configuration. PP<sub>i</sub> hydrolysis renders this reaction thermodynamically favorable. Step 6: Decarboxylation of OMP by OMP decarboxylase yields UMP.



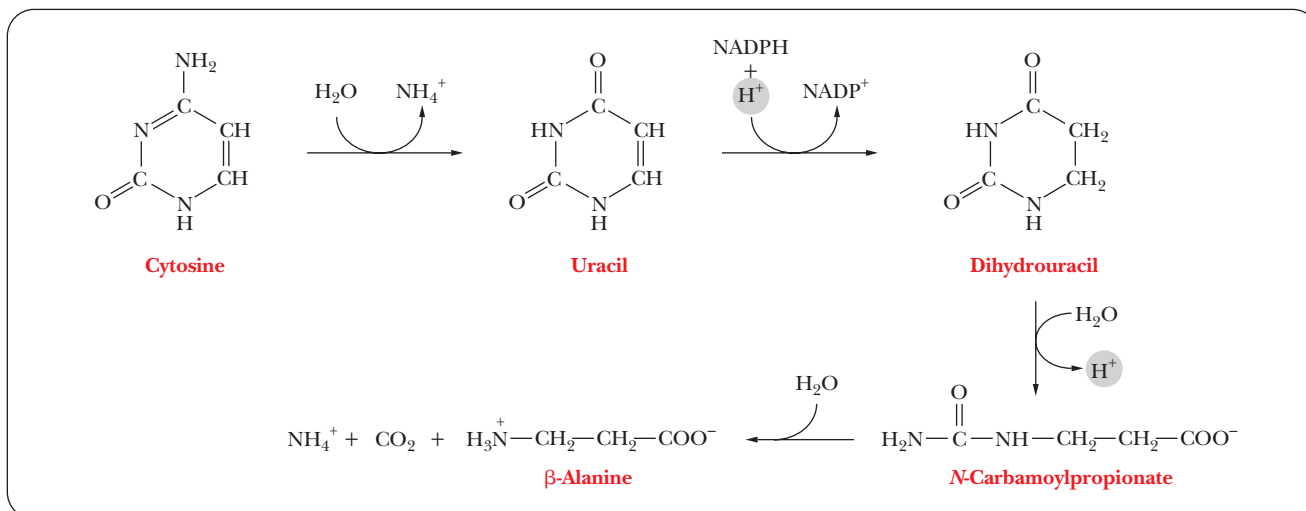
■ **FIGURE 23.27** The conversion of UMP to UTP.



■ **FIGURE 23.28** The conversion of UTP to CTP.



■ **FIGURE 23.29** The role of feedback inhibition in the regulation of pyrimidine nucleotide biosynthesis.



■ **FIGURE 23.30** The catabolism of pyrimidines.

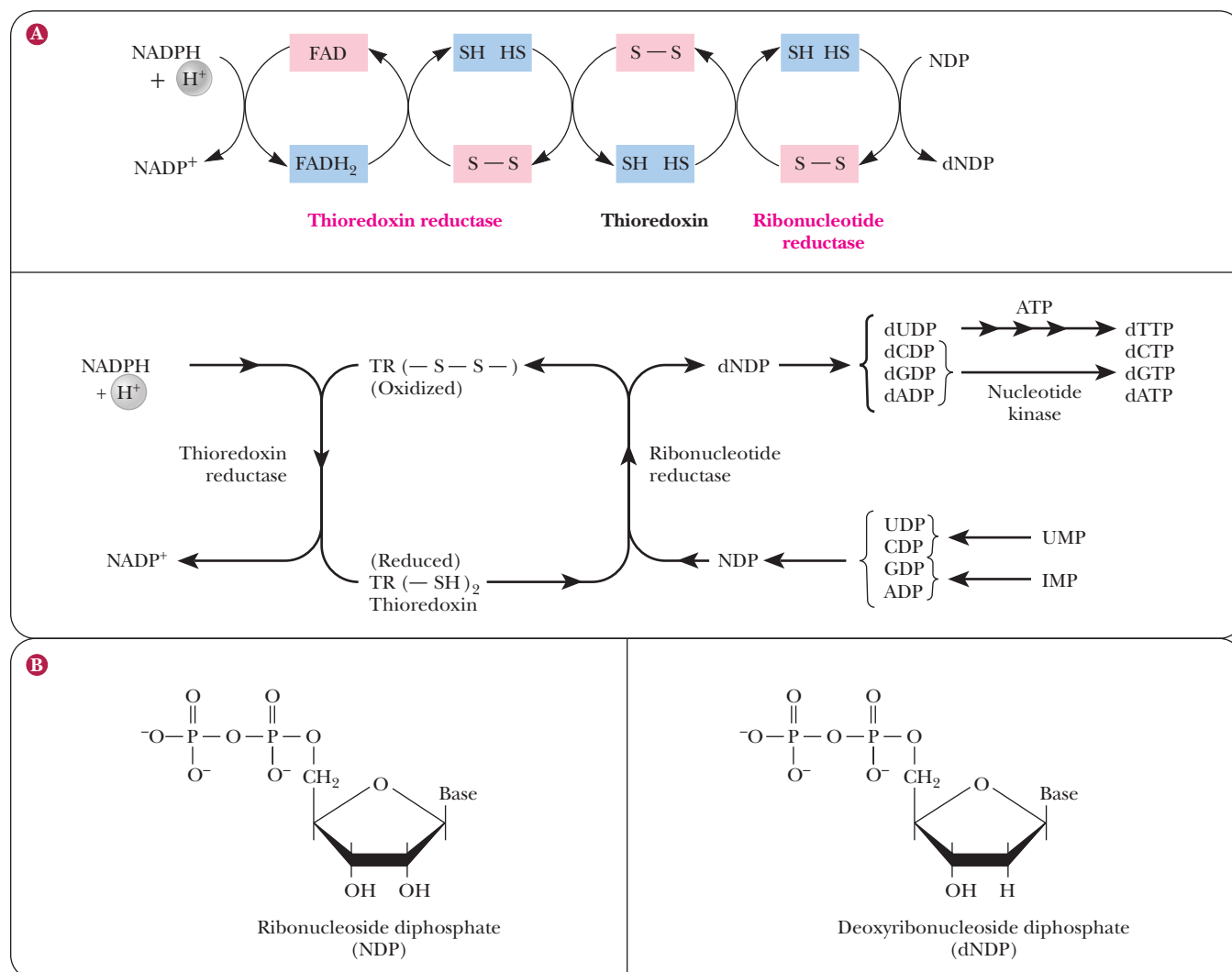


## 23.10 Conversion of Ribonucleotides to Deoxyribonucleotides

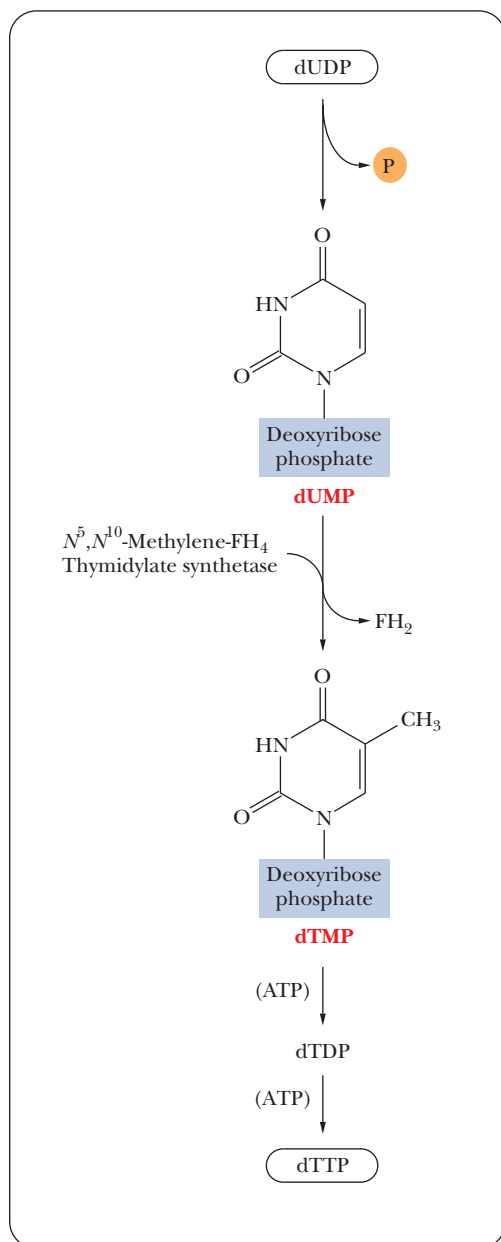
Ribonucleoside diphosphates are reduced to 2'-deoxyribonucleoside diphosphates in all organisms (Figure 23.31a); NADPH is the reducing agent.



The actual process, which is catalyzed by *ribonucleotide reductase*, is more complex than the preceding equation would indicate and involves some intermediate electron carriers. The ribonucleotide reductase system from *E. coli* has been extensively studied, and its mode of action gives some clues to the nature of the process. Two other proteins are required, thioredoxin and thioredoxin reductase. **Thioredoxin** contains a disulfide (S—S) group in its oxidized form and two sulfhydryl (—SH) groups in its reduced form. NADPH reduces thioredoxin in a reaction catalyzed by **thioredoxin reductase**. The reduced thioredoxin in turn reduces a ribonucleoside diphosphate (NDP) to a deoxyribonucleoside diphosphate (dNDP), shown in Figure 23.31b, and this reaction



**FIGURE 23.31** Conversion of ribonucleoside diphosphates to deoxyribonucleoside diphosphates. (a) The (—S—S—)/(—SH HS—) oxidation–reduction cycle involving ribonucleotide reductase, thioredoxin, thioredoxin reductase, and NADPH. (b) The structures of NDP and dNDP.



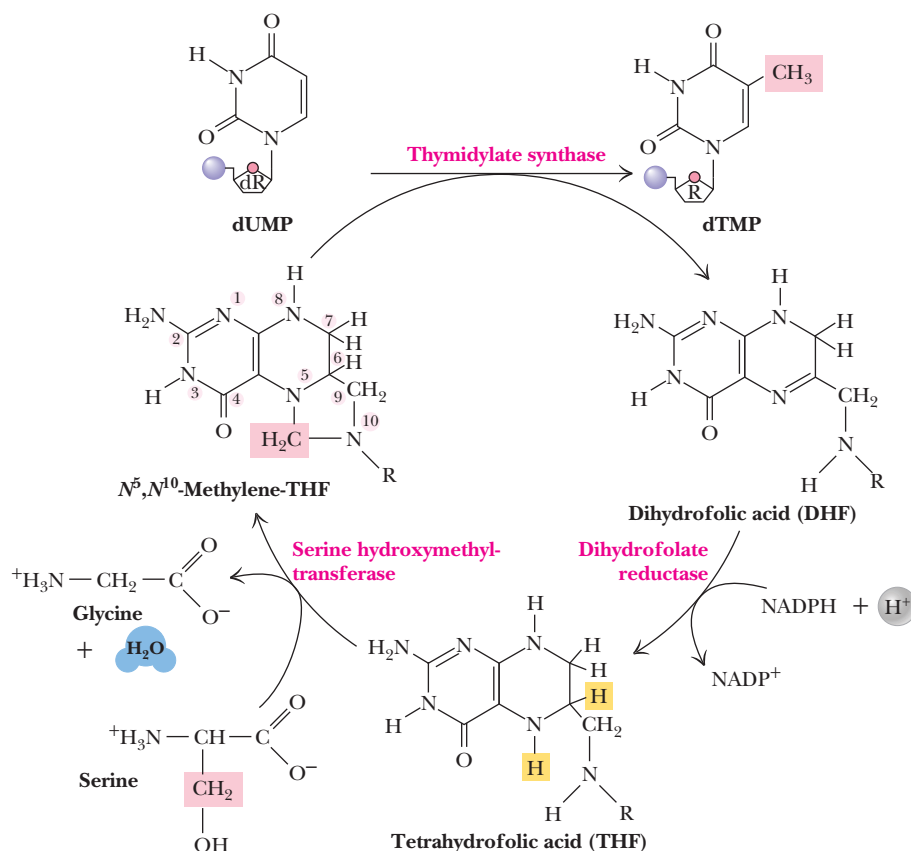
■ **FIGURE 23.32** The conversion of dUDP to dTTP. (FH<sub>4</sub> is tetrahydrofolate; FH<sub>2</sub> is dihydrofolate.)

is actually catalyzed by ribonucleotide reductase. Note that this reaction produces dADP, dGDP, dCDP, and dUDP. The first three are phosphorylated to give the corresponding triphosphates, which are substrates for the synthesis of DNA. Another required substrate for DNA synthesis is dTTP, and we shall now see how dTTP is produced from dUDP.

## 23.11 Conversion of dUDP to dTTP

A one-carbon transfer is required for the conversion of uracil to thymine by attachment of the methyl group. The most important reaction in this conversion is that catalyzed by *thymidylate synthase* (Figure 23.32). The source of the one-carbon unit is  $N^5,N^{10}$ -methylene-tetrahydrofolate, which is converted to dihydrofolate in the process. The metabolically active form of the one-carbon carrier is tetrahydrofolate. Dihydrofolate must be reduced to tetrahydrofolate for this series of reactions to continue, and this process requires NADPH and *dihydrofolate reductase*.

Since a supply of dTTP is necessary for DNA synthesis, inhibition of enzymes that catalyze the production of dTTP inhibits the growth of rapidly dividing cells. Cancer cells, like all fast-growing cells, depend on continued DNA synthesis for growth. Inhibitors of thymidylate synthetase, such as fluorouracil (see Question 50), and inhibitors of dihydrofolate reductase, such as aminopterin and methotrexate (structural analogues of folate), have been used in cancer chemotherapy (Figure 23.33). The intent of such therapy is to inhibit the formation of dTTP and thus of DNA in cancer cells, causing the death of the cancer cells with minimal effect on normal cells, which grow more slowly. Chemotherapy has adverse side effects because of the highly toxic nature of most of the drugs involved; normal cells are affected to some extent, although less than the cancer cells. Enormous amounts of research are focused on finding safe and effective forms of treatment.



■ **FIGURE 23.33** The thymidylate synthase reaction. The 5-CH<sub>3</sub> group is ultimately derived from the β-carbon of serine.

## SUMMARY

**What is nitrogen metabolism?** The metabolism of nitrogen encompasses a number of topics, including the anabolism and catabolism of amino acids, porphyrins, and nucleotides. Atmospheric nitrogen is the ultimate source of this element in biomolecules.

**How is nitrogen from the atmosphere incorporated into biologically useful compounds?** Nitrogen fixation is the process by which molecular nitrogen from the atmosphere is made available to organisms in the form of ammonia. Nitrification reactions convert  $\text{NO}_3^-$  to  $\text{NH}_3$  and provide another source of nitrogen.

**What is feedback inhibition in nitrogen metabolism?** Feedback-inhibition control mechanisms are a unifying factor in biosynthetic pathways involving nitrogen compounds. Most of the nitrogen metabolism pathways are long and complicated and use a great deal of energy. Shutting off these processes when enough of the final product has built up is important to the energy flux of the cell.

**What are some common features in amino acid biosynthesis?** In the anabolism of amino acids, transamination reactions play an important role. Glutamate and glutamine are frequently the amino-group donors. The enzymes that catalyze transamination reactions frequently require pyridoxal phosphate as a coenzyme. One-carbon transfers also operate in the anabolism of amino acids. Carriers are required for the one-carbon groups transferred. Tetrahydrofolate is a carrier of methylene and formyl groups, and S-adenosylmethionine is a carrier of methyl groups.

**What are essential amino acids?** Some species, including humans, cannot synthesize all the amino acids required for protein synthesis and must therefore obtain these essential amino acids from dietary sources. About half of the standard 20 amino acids are essential in humans, including arginine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, threonine, tryptophan, and valine.

**What is the fate of the carbon skeleton in amino acid breakdown?** The catabolism of amino acids has two parts: the fate

of the nitrogen and the fate of the carbon skeleton. The carbon skeleton is converted to pyruvate or oxaloacetate, in the case of glucogenic amino acids, or to acetyl-CoA or acetoacetyl-CoA, in the case of ketogenic amino acids.

**What is the role of the urea cycle in amino acid breakdown?** In the urea cycle, nitrogen released by the catabolism of amino acids is converted to urea. The urea cycle also plays a role in biosynthesis of amino acids.

**How is inosine monophosphate converted to AMP and GMP?** The anabolic pathway of nucleotide synthesis involving purines differs from that involving pyrimidines. Both pathways use preformed ribose-5-phosphate but differ with regard to the point in the pathway at which the sugar phosphate is attached to the base. In the case of purine nucleotides, the growing base is attached to the sugar phosphate during the synthesis, ultimately giving rise to inosine monophosphate. This compound is converted to AMP and to GMP in reactions subject to a high degree of feedback control.

**What are the energy requirements for production of AMP and GMP?** The energy requirements for production of AMP and GMP are high, more than 30 ATP equivalents in both cases.

**What is purine catabolism?** In catabolism, purine bases are frequently salvaged and reattached to sugar phosphates. Otherwise, purines are broken down to uric acid.

**What is pyrimidine biosynthesis and catabolism?** In pyrimidine biosynthesis, the base is first formed and then attached to the sugar phosphate. Pyrimidines are degraded to  $\beta$ -alanine.

**What is conversion of ribonucleotides to deoxyribonucleotides?** Deoxyribonucleotides for DNA synthesis are produced by the reduction of ribonucleoside diphosphates to deoxyribonucleoside diphosphates.

**What is conversion of dUDP to dTTP?** Another reaction specifically needed to produce substrates for DNA synthesis is the conversion of uracil to thymine. This pathway, which requires a tetrahydrofolate derivative as the carrier for one-carbon transfer, is a target for cancer chemotherapy.

## REVIEW EXERCISES

 **OWL** Interactive versions of these problems are assignable in OWL

### 23.1 Nitrogen Metabolism: An Overview

1. **Recall** What kinds of organisms can fix nitrogen? Which ones cannot?

### 23.2 Nitrogen Fixation

2. **Recall** How is nitrogen fixed (converted from  $\text{N}_2$  to  $\text{NH}_4^+$ )? How is it subsequently assimilated into organic compounds?
3. **Biochemical Connections** What is the Haber process?

4. **Recall** Write the overall reaction for the fixation of nitrogen via the nitrogenase complex.
5. **Recall** Describe the nitrogenase complex. How is the enzyme organized? What are its unique components?

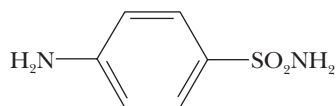
### 23.3 Feedback Inhibition in Nitrogen Metabolism

6. **Recall** How are nitrogen-utilizing pathways controlled by feedback inhibition?

7. **Reflect and Apply** Comment briefly on the usefulness to organisms of feedback control mechanisms in long biosynthetic pathways.
8. **Reflect and Apply** Metabolic cycles are rather common (Calvin cycle, citric acid cycle, urea cycle). Why are cycles so useful to organisms?

### 23.4 Amino Acid Biosynthesis

9. **Recall** What is the relationship between  $\alpha$ -ketoglutarate, glutamate, and glutamine in amino acid anabolism?
10. **Recall** Draw a transamination reaction between  $\alpha$ -ketoglutarate and alanine.
11. **Recall** Diagram the reactions involving glutamate dehydrogenase and glutamine synthetase that produce glutamine from ammonia and  $\alpha$ -ketoglutarate.
12. **Recall** What is the difference between glutamine synthetase and glutaminase?
13. **Recall** Draw the mechanism of transamination with pyridoxal phosphate.
14. **Recall** What cofactors are involved in one-carbon transfer reactions of amino acid anabolism?
15. **Recall** Sketch the structure of folic acid. Also sketch how it serves as a carrier of one-carbon groups.
16. **Recall** Why is there no net gain of methionine if homocysteine is converted to methionine with *S*-adenosylmethionine as the methyl donor?
17. **Recall** Show, by the equation for a typical reaction, why glutamate plays a central role in the biosynthesis of amino acids.
18. **Recall** By means of a structural formula, show how *S*-adenosylmethionine is a carrier of methyl groups.
19. **Reflect and Apply** Sulfanilamide and related sulfa drugs were widely used to treat diseases of bacterial origin before penicillin and more advanced drugs were readily available. The inhibitory effect of sulfanilamide on bacterial growth can be reversed by *p*-aminobenzoate. Suggest a mode of action for sulfanilamide.



Sulfanilamide

20. **Reflect and Apply** Proteins contain methionine but not  $\alpha$ -amino-*n*-hexanoic acid. The only structural difference is the substitution of  $-\text{CH}_2-$  for  $-\text{S}-$ . Both groups are similar in size and hydrophobic character. Why is methionine more advantageous than  $\alpha$ -amino-*n*-hexanoic acid?

### 23.5 Essential Amino Acids

21. **Recall** In general, what categories of amino acids are essential in humans and which are nonessential?
22. **Recall** List the essential amino acids for a phenylketonuric adult and compare them with the requirements for a normal adult.

### 23.6 Amino Acid Catabolism

23. **Recall** How many  $\alpha$ -amino acids participate directly in the urea cycle? Of these, how many can be used for protein synthesis?
24. **Recall** Write an equation for the net reaction of the urea cycle. Show how the urea cycle is linked to the citric acid cycle.
25. **Recall** Describe citrulline and ornithine based on their similarity to one of the 20 standard amino acids.
26. **Recall** Which amino acids in the urea cycle are the links to the citric acid cycle? Show how these links occur.

27. **Recall** How many ATPs are required for one round of the urea cycle? Where do these ATPs get used?
28. **Recall** How is carbamoyl-phosphate synthetase I (CPS-I) controlled?
29. **Recall** What is the logic behind high levels of arginine positively regulating *N*-acetylglutamate synthase?
30. **Recall** How does the level of glutamic acid affect the urea cycle?
31. **Recall** When amino acids are catabolized, what are the end products of the carbon skeletons for glucogenic amino acids? For ketogenic amino acids?
32. **Recall** Will an amino acid be glucogenic or ketogenic if it is catabolized to the following molecules?
  - (a) Phosphoenolpyruvate
  - (b)  $\alpha$ -Ketoglutarate
  - (c) Succinyl-CoA
  - (d) Acetyl-CoA
  - (e) Oxaloacetate
  - (f) Acetoacetate
33. **Biochemical Connections** What species excrete excess nitrogen as ammonia? Which ones excrete it as uric acid?
34. **Biochemical Connections** Would you expect an ostrich to excrete excess nitrogen as uric acid, urea, or ammonia? Make an argument for your answer.
35. **Reflect and Apply** Why is arginine an essential amino acid, when it is made in the urea cycle?
36. **Reflect and Apply** People on high-protein diets are advised to drink lots of water. Why?
37. **Reflect and Apply** Why is it better, when running a marathon, to drink a beverage with sugar for energy rather than one with amino acids?
38. **Reflect and Apply** Argue logically that the urea cycle should not have evolved. Then, logically counter your argument.

### 23.7 Purine Biosynthesis

39. **Biochemical Connections** How is the importance of folic acid related to chemotherapy?
40. **Recall** What are the sources of the carbons and nitrogens in the purine bases?
41. **Recall** What is the structural difference between inosine and adenosine?
42. **Recall** How is tetrahydrofolate important to purine synthesis?
43. **Recall** Does the conversion of IMP to GMP use or produce ATP either directly or indirectly? Justify your answer.
44. **Recall** Discuss the role of feedback inhibition in the anabolism of purine-containing nucleotides.

### 23.8 Purine Catabolism

45. **Recall** How many high-energy phosphate bonds must be hydrolyzed in the pathway that produces GMP from guanine and PRPP by the PRPP salvage reaction, compared with the number of such bonds hydrolyzed in the pathway leading to IMP and then to GMP?
46. **Reflect and Apply** Why do most mammals, other than primates, not suffer from gout?

### 23.9 Pyrimidine Biosynthesis and Catabolism

47. **Recall** What is an important difference between the biosynthesis of purine nucleotides and that of pyrimidine nucleotides?
48. **Recall** Compare the fates of the products of purine and pyrimidine catabolism.

### 23.10 Conversion of Ribonucleotides to Deoxyribonucleotides

49. **Recall** What roles do thioredoxin and thioredoxin reductase play in the metabolism of nucleotides?

**23.11 Conversion of dUDP to dTTP**

50. **Recall** Suggest a mode of action for fluorouracil in cancer chemotherapy.
51. **Reflect and Apply** Chemotherapy patients receiving cytotoxic (cell-killing) agents such as FdUMP (the UMP analogue that contains fluorouracil) and methotrexate temporarily go bald. Why does this take place?

## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





# Integration of Metabolism: Cellular Signaling

# 24

A traffic officer maintains order and stops chaos in the heart of the big city. In the same way, intricate control mechanisms keep the body's biochemical pathways working smoothly together.

## 24.1 Connections between Metabolic Pathways

In the preceding chapters, we learned about a number of individual metabolic pathways. Some metabolites, such as pyruvate, oxaloacetate, and acetyl-CoA, appear in more than one pathway. Furthermore, reactions of metabolism can take place simultaneously, and it is important to consider control mechanisms by which some reactions and pathways are turned on and off.

All metabolism is ultimately linked to photosynthesis and the energy from the Sun (Figure 24.1). The light reactions produce ATP and NADPH, which are then used to make carbohydrates in the dark reactions. These carbohydrates are the source of nutrients for other organisms. ATP and NADPH are the two consistent links between different forms of metabolism. Besides linking the light and dark reactions of photosynthesis, they are the most direct link between catabolism and anabolism (Figure 24.1). Other common molecules, such as sugars, PEP, pyruvate, and acetyl-CoA, also form a bridge between catabolic and anabolic processes. We shall now focus on some of the relationships among pathways by considering some of the physiological responses to biochemical events.

The **citric acid cycle** plays a central role in metabolism. Three main points can be considered in assigning a central role to the citric acid cycle. The first of these is its part in the catabolism of nutrients of the main types: carbohydrates, lipids, and proteins (Section 19.7). The second is the function of the citric acid cycle in the anabolism of sugars, lipids, and amino acids (Section 19.8). The third and final point is the relationship between individual metabolic pathways and the citric acid cycle. When we discuss these broader considerations, we can and should address questions that involve more than individual cells and the reactions that go on in them, such as questions of what goes on in tissues and in whole organs. In this chapter, we shall look at three such topics—nutrition, hormonal control, and the wide-reaching effects of signaling pathways. The following Biochemical Connections box describes the way in which one compound can affect an entire organism.

## 24.2 Biochemistry and Nutrition

The molecules that we process by catabolic reactions ultimately come from outside the body because we are heterotrophic organisms (dependent on external food sources). We shall devote this section to a brief look at how the foods we eat are sources of substrates for catabolic reactions. We should also bear in mind that nutrition is related to physiology as well as to biochemistry. This last point is certainly appropriate in view of the fact that many early biochemists were physiologists by training.

## Chapter Outline

### 24.1 Connections between Metabolic Pathways

#### 24.2 Biochemistry and Nutrition

- What are required nutrients?
- Why do we need vitamins?
- What are minerals?
- Is the old food pyramid still valid?
- What is obesity?

#### 24.3 Hormones and Second Messengers

- What are hormones?
- How do second messengers work?

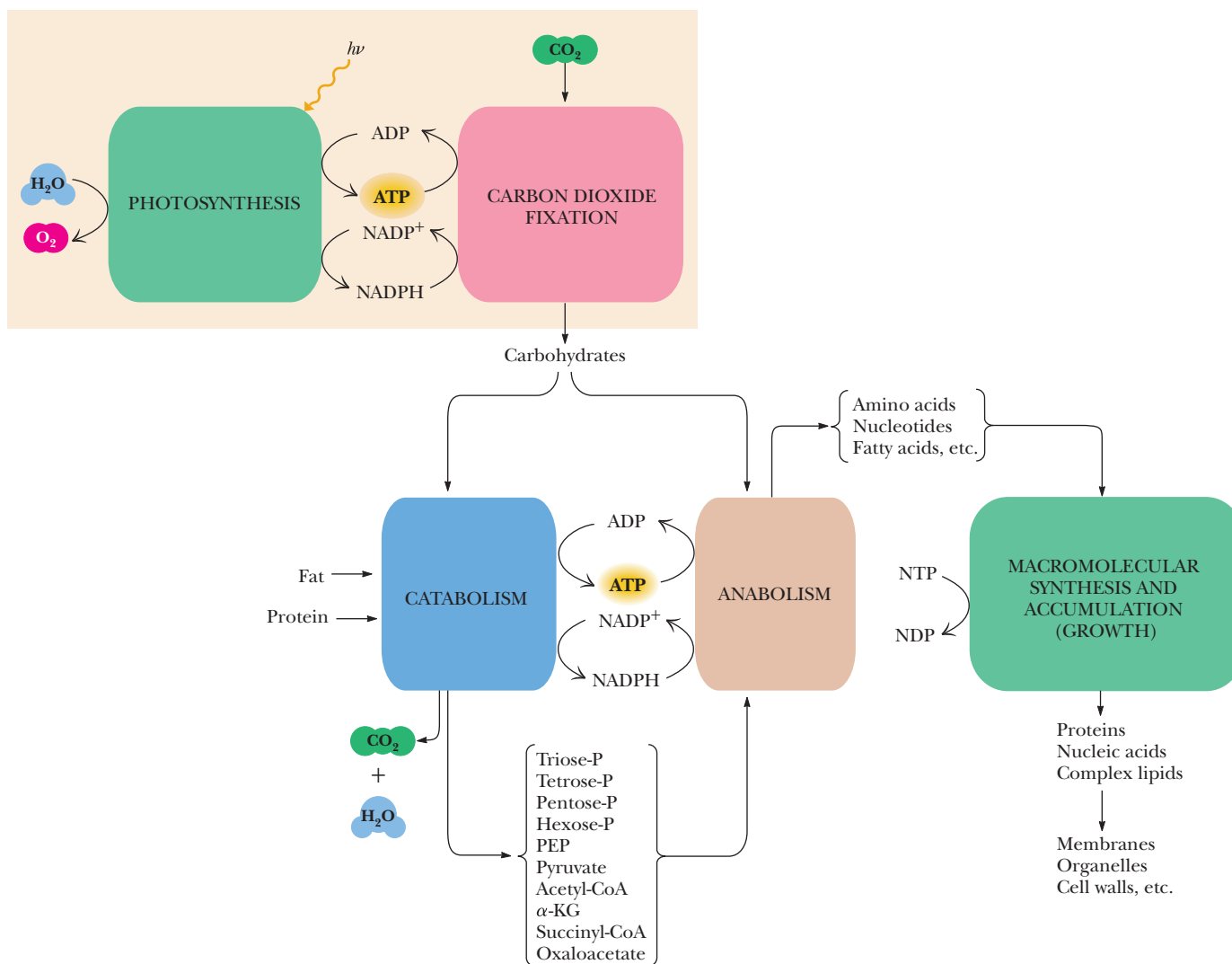
#### 24.4 Hormones and the Control of Metabolism

- What hormones control carbohydrate metabolism?

#### 24.5 Insulin and Its Effects

- What is insulin?
- What does insulin do?

**OWL** Online homework for this chapter may be assigned in OWL.



■ **FIGURE 24.1 All metabolism is connected.** This block diagram of intermediary metabolism shows the relationship between anabolic and catabolic processes and the common metabolites seen in many pathways.

### What are required nutrients?

In humans, the catabolism of **macronutrients** (carbohydrates, fats, and proteins) to supply energy is an important aspect of nutrition. In the United States, most diets provide more than an adequate number of nutritional calories. The typical American diet is high enough in fat that essential fatty acids (Section 21.6) are seldom, if ever, deficient. The only concern is that the diet contains an adequate supply of protein. If the intake of protein is sufficient, the supply of essential amino acids (Section 23.5) is normally also sufficient. Packaging on food items frequently lists the protein content in terms of both the number of grams of protein and the percentage of the daily value (DV) suggested by the Food and Nutrition Board under the auspices of the National Research Council of the National Academy of Sciences (see Table 24.1). Daily values have replaced the recommended daily allowances (RDAs) formerly seen on food packaging.

There are some key biochemical concepts to remember when analyzing a diet for protein content. First, there is no storage form for proteins. This means

# Biochemical Connections ALLIED HEALTH

## Alcohol Consumption and Addiction

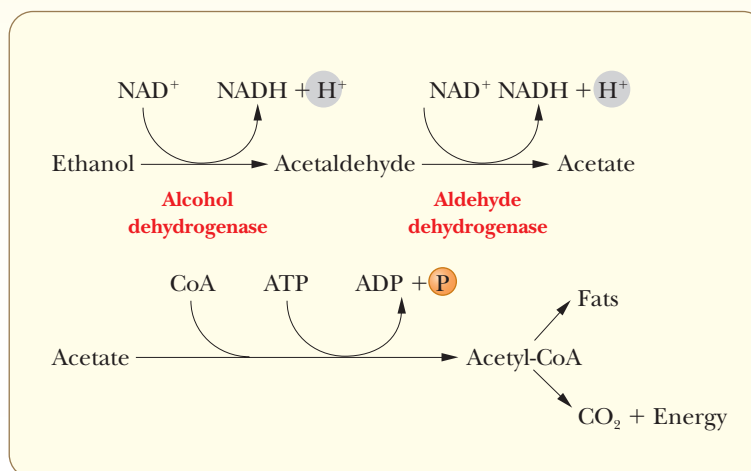
Alcohol is the most abused drug in America, and alcoholism is among the most common diseases. Statistics about deaths due to drunk driving are available, but no one knows how many other accidental deaths may be indirectly caused by alcohol. Many believe that some particular biochemistry must be associated with alcoholism. There is certainly a genetic trait, shown most forcefully in a benchmark study of identical twins raised apart from each other. Attempts to find “the gene for alcoholism” have not, however, met with success. A complex genetic relationship is likely to be involved.

Alcohol dehydrogenase is an inducible enzyme. Its level increases in response to alcohol use. The first reaction occurs very rapidly in alcoholics, so the intoxicating effect of alcohol is actually reduced (i.e., less intoxication per ounce). Alcoholics can tolerate blood alcohol levels that would be lethal to others. For everyone, the second reaction is rate-limiting. Acetaldehyde can cause headaches, nausea, and hangovers. Malnutrition is common among alcoholics because alcohol is a source of “empty calories” without important nutrients, particularly vitamins.

Biochemical, psychological, and nutritional effects of alcohol are not the same for all people. The studies of twins indicate the

possibility of a “born” alcoholic who could be totally hooked on the first drink. Fetal alcohol syndrome (see the Biochemical Connections box on page 503) is of particular concern to women. Ethanol is a teratogen; there is no “safe” level of alcohol during pregnancy. Fetal alcohol syndrome occurs in up to 5 of every 1000 births. Indicators include stunted growth, dysfunction of the central nervous system, and a characteristic facial shape.

Clearly, biochemistry is involved in addiction. Because many of the psychoactive drugs are structural analogs of serotonin and epinephrine (see the Biochemical Connections box on page 68), it is easy to imagine enhancement of their effects or competition with them. There is increasing interest in the effects of drugs (in general) on the production of endorphins and enkephalins (Section 3.5), the short peptides that are the brain’s own opiate painkillers. In people who are not alcoholics, ethanol inhibits enkephalin synthesis because the pleasant effect of alcohol replaces the need for enkephalins. Part of the misery of a hangover is caused by the lack of enkephalins; the hangover usually lasts until the level of these compounds returns to normal.



that proteins eaten in excess do a person no good in terms of satisfying that person’s future protein requirements. All protein consumed in excess of what is needed is turned into carbohydrate or fat, and the nitrogen from the amino group must be eliminated through the urea cycle (Section 23.6). Ingesting too much protein can therefore be stressful on the liver and kidneys because of the overproduction of ammonia that must be eliminated. This is the same risk faced by certain athletes who take creatine to build muscles, because creatine is a highly nitrogenated compound.

Second, the essential amino acids must be consumed daily in order for proteins to be made. It would be difficult to find a protein that did not have at least one residue of each of the common 20 amino acids. Half of these amino acids are essential, and if the diet is lacking or low in even one of these essential amino acids, then protein synthesis is not possible. Not all proteins are created equal (see the Biochemical Connections box on page 85). The protein efficiency ratio (PER) is an indication of how complete a protein is. However,

**TABLE 24.1**

Daily Values for the Average Man and Woman, Aged 19 to 22		
Nutrient	Man	Woman
Protein	56 g	44 g
Lipid-soluble vitamins		
Vitamin A	1 mg RE*	8 mg RE*
Vitamin D	7.5 $\mu$ g <sup>†</sup>	7.5 $\mu$ g <sup>†</sup>
Vitamin E	10 mg $\alpha$ -TE <sup>‡</sup>	8 mg $\alpha$ -TE <sup>‡</sup>
Water-soluble vitamins		
Vitamin C	60 mg	60 mg
Thiamine (vitamin B <sub>1</sub> )	1.5 mg	1.1 mg
Riboflavin (vitamin B <sub>2</sub> )	1.7 mg	1.3 mg
Vitamin B <sub>6</sub>	3 $\mu$ g	3 $\mu$ g
Vitamin B <sub>12</sub>	3 $\mu$ g	3 $\mu$ g
Niacin	3 $\mu$ g	3 $\mu$ g
Folic acid	19 mg	14 mg
Pantothenic acid (estimate)	10 mg	10 mg
Biotin (estimate)	0.3 mg	0.3 mg
Minerals		
Calcium	800 mg	800 mg
Phosphorus	800 mg	800 mg
Magnesium	350 mg	300 mg
Zinc	15 mg	15 mg
Iron	10 mg	18 mg
Copper (estimate)	3 mg	3 mg
Iodine	150 $\mu$ g	150 $\mu$ g

\* RE = retinol equivalent, where 1 retinol equivalent = 1  $\mu$ g retinol or 6  $\mu$ g  $\beta$ -carotene. See Section 8.7.

<sup>†</sup> As cholecalciferol. See Section 8.7.

<sup>‡</sup>  $\alpha$ -TE =  $\alpha$ -tocopherol equivalent, where 1  $\alpha$ -TE = 1  $\mu$ g D- $\alpha$ -tocopherol. See Section 8.7. Data from the Food and Nutrition Board, National Academy of Sciences–National Research Council, Washington, D.C., 1988.

mixing proteins correctly is very important, something that vegetarians know a lot about. A protein that is very low in lysine would have a low PER value. If a second protein had a low PER because it was low in tryptophan, it could be combined with the low-lysine protein to give a combination with a high PER. However, this would work only if the two were eaten together.

Third, proteins are always being degraded (Chapter 12). Because of that, even if a person does not seem to be doing any activities that would tend to require protein replenishment, there is a constant need for quality protein in order to maintain the body’s structures. Athletes are painfully aware of that fact. They must train constantly, and they get out of shape quickly when they stop, the effect being even more pronounced as the athlete reaches middle age.

**Why do we need vitamins?**

**Micronutrients** (vitamins and minerals) are also listed on food packaging. The vitamins we require are compounds that are necessary for metabolic processes; either our bodies cannot synthesize them, or they cannot synthesize them in amounts sufficient for our needs. As a result, we must obtain vitamins from dietary sources. DVs are listed for the fat-soluble vitamins—vitamins A, D, and E (Section 8.7)—but care must be taken to avoid overdoses of these vitamins. Excesses can be toxic when large amounts of fat-soluble vitamins accumulate in adipose tissue. Excess vitamin A is especially toxic. With water-soluble vitamins, turnover is frequent enough that the danger of excess is not normally a problem.



The water-soluble vitamins with listed DVs are vitamin C, which is necessary for the prevention of scurvy (Section 4.3), and the B vitamins—niacin, pantothenic acid, vitamin B<sub>6</sub>, riboflavin, thiamine, folic acid, biotin, and vitamin B<sub>12</sub>. The B vitamins are the precursors of the metabolically important coenzymes listed in Table 7.1, where references to the reactions in which the coenzymes play a role are given. We have seen many pathways in which NADH, NADPH, FAD, TPP, biotin, pyridoxal phosphate, and coenzyme A were found, all of which came from vitamins. A summary of vitamins and their metabolic roles is given in Table 24.2. Frequently, the actual biochemical role is played by a metabolite of the vitamin rather than by the vitamin itself, but this point does not affect the dietary requirement.

### What are minerals?

**Minerals**, in the nutritional sense, are inorganic substances required in the ionic or free-element form for life processes. The macrominerals (those needed in the largest amounts) are sodium, potassium, chloride, magnesium, phosphorus, and calcium. The required amounts of all these minerals, except calcium, can easily be satisfied by a normal diet. Deficiencies of calcium can, and frequently do, occur. Calcium deficiencies can lead to bone fragility, with concomitant risk of fracture, which is a problem especially for elderly women. Calcium supplements are indicated in such cases. Requirements for some microminerals (trace minerals) are not always clear. It is known, for example, from biochemical evidence that chromium is necessary for glucose metabolism (a role that has recently been suggested for chromium picolinate) and that manganese is necessary for bone formation, but no deficiencies of these elements have been recorded. Requirements have been established for iron, copper, zinc, iodide, and fluoride; there are DVs for all of these minerals except fluoride. In the case of copper and zinc, needs are easily met by dietary sources, and overdoses can be toxic. A deficiency of iodide, leading to an enlarged

**TABLE 24.2**

Vitamins: Chemical and Biochemical Facts		
Vitamin	Metabolic Function	Reference
<b>Water-Soluble</b>		
B <sub>1</sub> (thiamine)	Aldehyde transfer, decarboxylation in alcoholic fermentation and citric acid cycle	Sections 17.4, 19.3
B <sub>2</sub> (riboflavin)	Oxidation–reduction reactions, especially in citric acid cycle and electron transport	Sections 19.3, 20.2
B <sub>6</sub> (pyridoxine)	Transamination reactions, especially of amino acids	Section 23.4
Niacin (nicotinic acid)	Oxidation–reduction reactions, found in many metabolic processes	Sections 17.3, 19.3, 20.2
Biotin	Carboxylation reactions in carbohydrate and lipid metabolism	Sections 18.2, 21.6
Pantothenic acid	Acyl transfer in many metabolic processes	Sections 15.7, 21.6
Folic acid	One-carbon group transfer, especially in nitrogen-containing compounds	Sections 23.4, 23.11
C (Ascorbic acid)	Hydroxylates collagen	Biochemical Connections box, p. 461.
Lipoic acid (?)	Acyl transfer, oxidation–reduction (It has been questioned whether lipoic acid is a vitamin.)	Section 19.3
<b>Lipid-Soluble</b>		
A	Isomerization mediates visual process	Section 8.7
D	Regulates calcium and phosphorus metabolism, especially in bone	Section 8.7
E	Antioxidant	Section 8.7
K	Mediates protein modification required for blood clotting	Section 8.7

## Biochemical Connections NUTRITION

### Iron: An Example of a Mineral Requirement

Iron, whether in the form  $\text{Fe(II)}$  or  $\text{Fe(III)}$ , is usually found in the body in association with proteins. Little or no iron can be found “free” in the blood. Because iron-containing proteins are ubiquitous, there is a dietary requirement for this mineral. Severe deficits can lead to iron-deficiency anemia.

Iron usually occurs as the  $\text{Fe(III)}$  form in food. This is also the form released from iron pots when food is cooked in them. However, iron must be in the  $\text{Fe(II)}$  state to be absorbed. Reduction from  $\text{Fe(III)}$  to  $\text{Fe(II)}$  can be accomplished by ascorbate (vitamin C) or by succinate. Factors that affect absorption include the solubility of a given compound of iron, the presence of antacids in the digestive tract, and the source of the iron. To give some examples, iron may form insoluble complexes with phosphate or oxalate, and the presence of antacids in the digestive tract may decrease iron absorption. Iron from meats is more easily absorbed than iron from plant sources.

Requirements for iron vary according to age and gender. Infants and adult men need 10 mg per day; infants are born with a three- to six-month supply. Children and women (ages 16 through 50) need 15 to 18 mg per day. Women lose 20 to 23 mg of iron during each menstrual period. Pregnant and lactating women need more than 18 mg per day. After a blood loss, anyone, regardless of age or gender, needs more than these amounts. Distance runners, particularly marathoners, are also at risk of becoming anemic because of the loss of blood cells in the feet caused by the pounding of the thousands of footfalls that occur during a long run. People with iron deficiencies may experience a craving for nonfood items such as clay, chalk, and ice.

While iron is an essential nutrient, too much iron can also have consequences. Iron normally is bound to proteins, such

as hemoglobin, myoglobin, and transferrin. If a person has too much iron, it will overload the proteins it should bind to. In that case free iron will increase. Free iron is a very reactive ion and causes oxidative damage to cells. Iron toxicity can come from genetic disorders where a person’s ability to manage their iron content is defective, or it can come from taking in far too much iron. As is the case with many compounds, more is not always better, at least not once you surpass the amount you actually need.

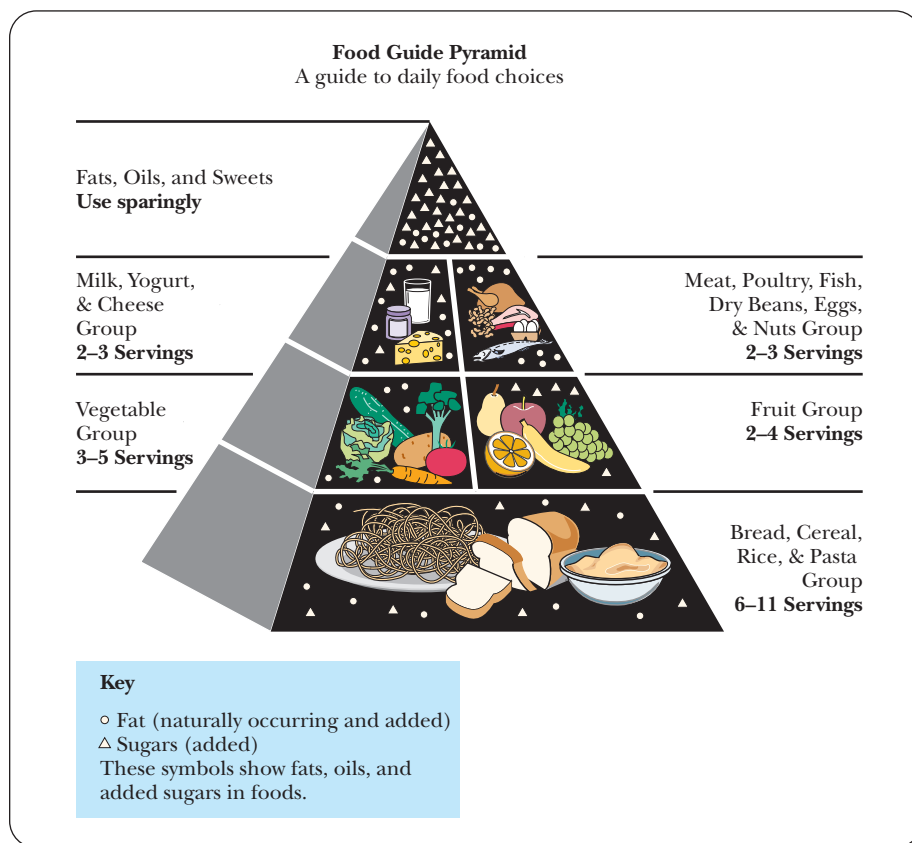


©Cengage Learning/Charles D. Winters

thyroid gland (Section 24.3), has been a problem in some parts of the United States for many years. Iodized salt is used to prevent this deficiency, and it has become unusual to find table salt without an iodine supplement. Fluoride is administered to prevent tooth decay in children and, with that end in mind, has been added to water supplies, sometimes causing considerable controversy. Iron is important because it is part of the structure of the ubiquitous heme proteins. Women of childbearing age are more susceptible to iron deficiencies than are other segments of the population, and in some cases supplements are advised. These recommended levels vary with the age of the individual and are subject to adjustment for level of activity.

### The Food Pyramid

One approach to publicizing healthful food selection was the development of the Food Guide Pyramid, a graphic display that focuses on a diet sufficient in nutrients but without excesses (Figure 24.2). The goal was to use a well-chosen diet to promote good health. To avoid confusion, the development of this scheme had to take into account the fact that many people were familiar with the older recommendations about food groups. The newer recommendations pay particular attention to increasing the amount of fiber and decreasing the amount of fat in the typical diet. Variety and moderation were key concepts of the graphic presentation. From the biochemical point of view, these recommendations translate into a diet based primarily on carbohydrates, with enough protein to meet needs for essential amino acids (Section 23.5). Note



■ **FIGURE 24.2 The Food Guide Pyramid (USDA).** The recommended choices reflect a diet based primarily on carbohydrates. Smaller amounts of proteins and lipids are sufficient to meet the body's needs.

that in Figure 24.2, carbohydrates are the base, with the correct amount suggested to be 6 to 11 servings of foods rich in complex carbohydrates, such as bread, cereal, rice, or pasta. Lipids should not contribute more than 30% of daily calories, but the typical American diet currently is about 45% fat. High-fat diets have been linked to heart disease and to some kinds of cancer, so the recommendation about lipid intake is of considerable importance.

### **Is the old food pyramid still valid?**

Many scientists are questioning some of the details of the food pyramid. Certain types of fat are essential to health and actually reduce the risk of heart disease. Also, there has been little evidence to back up the claim that a high intake of carbohydrates is beneficial. Many people feel that the original food pyramid, which was published in 1992, has serious flaws. It overglorifies carbohydrates while making all fats out to be the bad guys. In addition, meat, fish, poultry, and eggs are all lumped together as if they are equivalent in terms of health. Plenty of evidence links saturated fat with high cholesterol and risk of heart disease, but monounsaturated and polyunsaturated fats have the opposite effect. Although many scientists knew the distinction between the various types of fat, they felt that the average person would not understand them, and so the original pyramid was designed to send the simple message that fat was bad. The implied corollary to fat being bad was that carbohydrates were good. However, after years of study, no evidence can be shown that a diet that has 30% or fewer calories coming from fat is healthier than one with a higher level.

To further complicate matters, we have to recall the effects of the traveling forms of cholesterol—the lipoproteins. Having high levels of cholesterol traveling as high-density lipoproteins (HDL) has been correlated with a healthy heart, while having high levels of cholesterol traveling in the form of low-density lipoproteins (LDL) is related to high risk of heart disease (Chapter 21). When

calories from saturated fat are replaced by carbohydrates, the levels of LDL and total cholesterol decrease, but so does the level of HDL. Because the ratio of LDL to HDL does not decrease significantly, there is little health benefit. However, the increase in carbohydrate has been shown to increase fat synthesis because of increases in insulin production. When calories from unsaturated fat are replaced with calories from carbohydrates, the results are even worse. The LDL levels rise in comparison with the levels of HDL.

Figure 24.3 shows a more modern view of a food pyramid that takes into account the most recent evidence and recommendations from some nutritionists. Note that at the base of the pyramid is the heart and soul of good health—exercise and weight control. There is no replacement for being active and for restricting total calories when it comes to staying healthy. The next level up shows that the good types of carbohydrates and the good forms of fats occupy a prime location. Whole-grain foods are complex carbohydrates that are digested more slowly, so they do not have the effect of raising blood glucose and causing insulin levels to rise to the extent that refined carbohydrates like white rice and pasta do. The healthy fats come from plant oils. Vegetables and fruits still occupy an important place in this pyramid, with nuts and legumes just above them. Next are good sources of protein, such as fish, poultry, and eggs. Note that the recommendation says zero to two servings. This is a change in approach, in that the type of protein is considered important and in the fact that the guide shows that it is not necessary to eat animal protein at all. Dairy products are found high up on the new pyramid. This is because, despite the commercials that suggest “everybody needs milk,” there are some noted health risks in consuming dairy products. Some cultures that consume large quantities of dairy products have the highest incidence of heart disease, probably due to the high concentrations of saturated fatty acids in milk and butter. In addition, many adults are allergic to milk proteins, and many are unable to digest



■ **FIGURE 24.3** A new food pyramid. This version of the recommended amounts of the different food types reflects a distinction between unhealthy and healthy types of carbohydrates and fats. The recommended intake of dairy products has also been reduced compared with the original pyramid. (© Richard Borge. Adapted with permission.)



lactose. At the peak of the pyramid are the items to be eaten only sparingly: red meat and refined carbohydrates, as well as some natural carbohydrate sources, such as potatoes. In April 2006, the U.S. Department of Agriculture launched a nutrition website at [www.mypyramid.gov](http://www.mypyramid.gov). This service allows individuals to find information about foods and the role of physical activity in order to make choices for a healthful lifestyle. The website includes interactive programs to help people assess their lifestyle and nutrition choices, including worksheets to record daily consumption of nutrients.

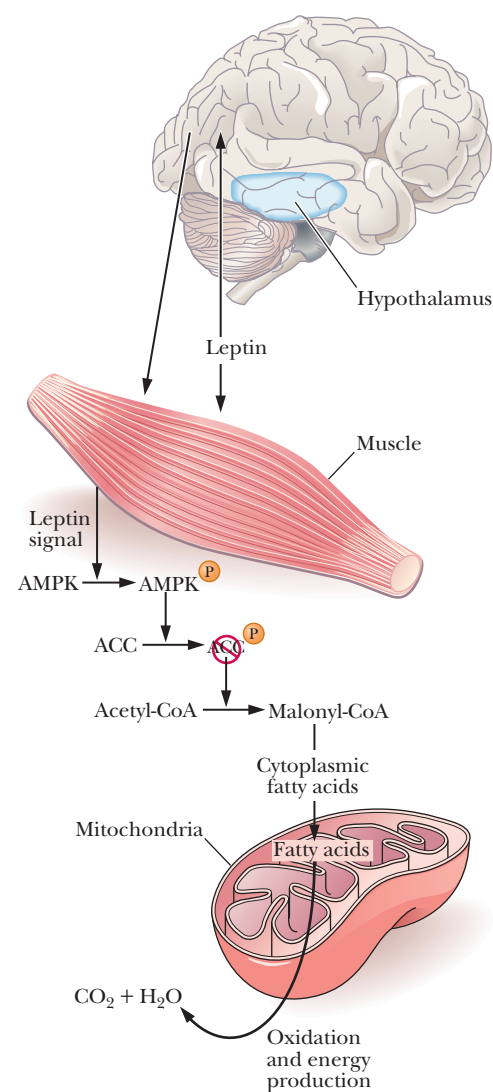
### What is obesity?

Obesity is a major public-health problem in the United States. Recent figures from the National Institutes of Health show that one-third of the population is clinically obese, defined as weighing at least 20% more than their ideal weight. Artificial sweeteners have been introduced, sometimes with great controversy, to help those who wish to control their weight. Fat substitutes have come on the market more recently, again accompanied by controversy. One thing is clear: the topic will continue to be of great interest, with trade-offs between palatability and health concerns providing a driving force for finding new products.

The role of the protein leptin in the control of obesity has been established in mice, and information is just appearing concerning its effect in humans. It is known that in mice leptin is a 16-kDa protein and that it is produced by the *obesity* (*ob*) gene. Mutations in this gene lead to a deficiency of leptin, which in turn leads to increased appetite and decreased activity, ultimately leading to weight gain. Injections of this protein into affected mice lead to decreased appetite and increased activity, with resulting weight loss. Administering leptin to leptin-deficient humans has been reported to reduce obesity; however, in clinically obese subjects, the circulating levels of leptin are often high. Some forms of obesity may be caused by a lack of sensitivity to leptin rather than a lack of leptin itself.

Leptin stimulates the oxidation of fatty acids and the uptake of glucose by muscle cells. It does so by stimulating AMP-activated protein kinase, which phosphorylates an isoform of acetyl-CoA carboxylase (ACC) in muscle cells, rendering it less active (Figure 24.4). Recall from Section 21.6 that ACC plays a pivotal role in fat metabolism. When ACC activity is decreased, malonyl-CoA levels decrease and the mitochondria can take up and oxidize fatty acids. Leptin also inhibits production of the mRNA for hepatic stearoyl-CoA desaturase, an enzyme that adds double bonds to saturated fatty acids, leading to less lipid synthesis.

Leptin also works directly on the nervous system. Both leptin and insulin (Section 24.5) are long-term regulators of appetite. They circulate in the blood at concentrations roughly proportional to body-fat mass. They inhibit appetite by inhibiting specific neurons in the hypothalamus. Several laboratories have shown interest in using this information to develop treatments for human obesity. The Biochemical Connections box on page 708 gives some of the theories on diet, hormones, and obesity.



**FIGURE 24.4 Leptin has multiple effects on metabolism.** It affects the brain, lowering appetite. It also inactivates acetyl-CoA carboxylase (ACC). Reduced activity of ACC leads to a reduction in malonyl-CoA, which stimulates fatty-acid oxidation and reduces fatty-acid synthesis. (From *Nature*, Vol. 415 [January 17, 2002], Fig. 1, p. 268. Copyright © 2002 Nature. Reprinted with permission.)

## 24.3 Hormones and Second Messengers

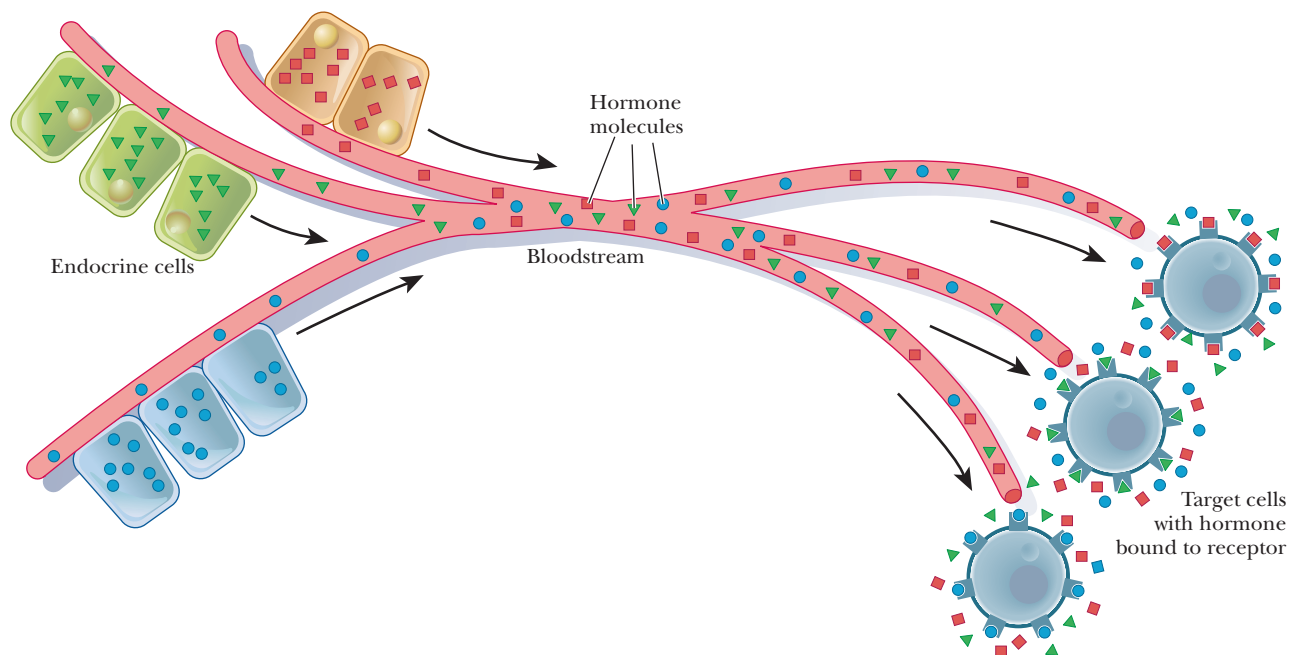
### Hormones

The metabolic processes within a given cell are frequently regulated by signals from outside the cell. A usual means of intercellular communication takes place through the workings of the **endocrine system**, in which the ductless glands produce **hormones** as intercellular messengers.

### What are hormones?

Hormones are transported from the sites of their synthesis to the sites of action by the bloodstream (Figure 24.5). In terms of their chemical structure, some typical





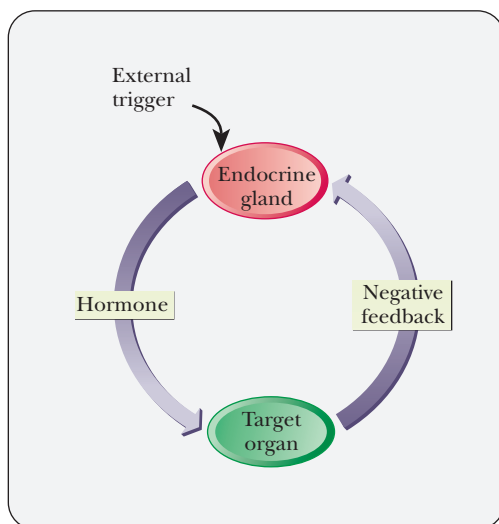
■ **FIGURE 24.5** Blood carries many hormones. Endocrine cells secrete hormones into the bloodstream, which transports them to target cells. Target cells have specific receptors that bind to particular hormones, thereby generating their metabolic effect.

hormones are steroids, such as estrogens, androgens, and mineralocorticoids (Section 21.8); polypeptides, such as insulin and endorphins (Section 3.5); and amino acid derivatives, such as epinephrine and norepinephrine (Table 24.3).

Hormones have several important functions in the body. They help maintain **homeostasis**, the balance of biological activities in the body. The effect of insulin in keeping the blood glucose level within narrow limits is an example of this function. The operation of epinephrine and norepinephrine in the “fight-or-flight” response is an example of the way in which hormones mediate responses to external stimuli. Finally, hormones play roles in growth and development, as seen in the roles of growth hormone and the sex hormones. The methods and insights of biochemistry and physiology alike have helped illuminate the workings of the endocrine system.

The release of hormones exerts control on the cells of target organs; other control mechanisms, however, determine the workings of the endocrine gland that releases the hormone in question. Simple feedback mechanisms, in which the action of the hormone leads to feedback inhibition of the release of hormone, can be postulated (Figure 24.6). The workings of the endocrine system are, in fact, much less simple, with the added complexity allowing for a greater degree of control. To illustrate with a rather restricted example, insulin is released in response to a rapid rise in the level of blood glucose. In the absence of control mechanisms, an excess of insulin can produce **hypoglycemia**, the condition of low blood glucose. In addition to negative feedback control on the release of insulin, the action of the hormone glucagon tends to increase the level of glucose in the bloodstream. The two hormones together regulate blood glucose. This example is far too restricted, as we shall see in the next section. We will look at insulin itself more closely in Section 24.5.

A more sophisticated control system involves the action of the *hypothalamus*, the *pituitary*, and specific *endocrine glands* (Figure 24.7). The central nervous system sends a signal to the hypothalamus. The **hypothalamus** secretes a hormone-releasing factor, which in turn stimulates release of a trophic hormone by the anterior pituitary (Table 24.3). (The action of the hypothalamus on the posterior pituitary is mediated by nerve impulses.) **Trophic hormones**



■ **FIGURE 24.6** A simple feedback control system involving an endocrine gland and a target organ.

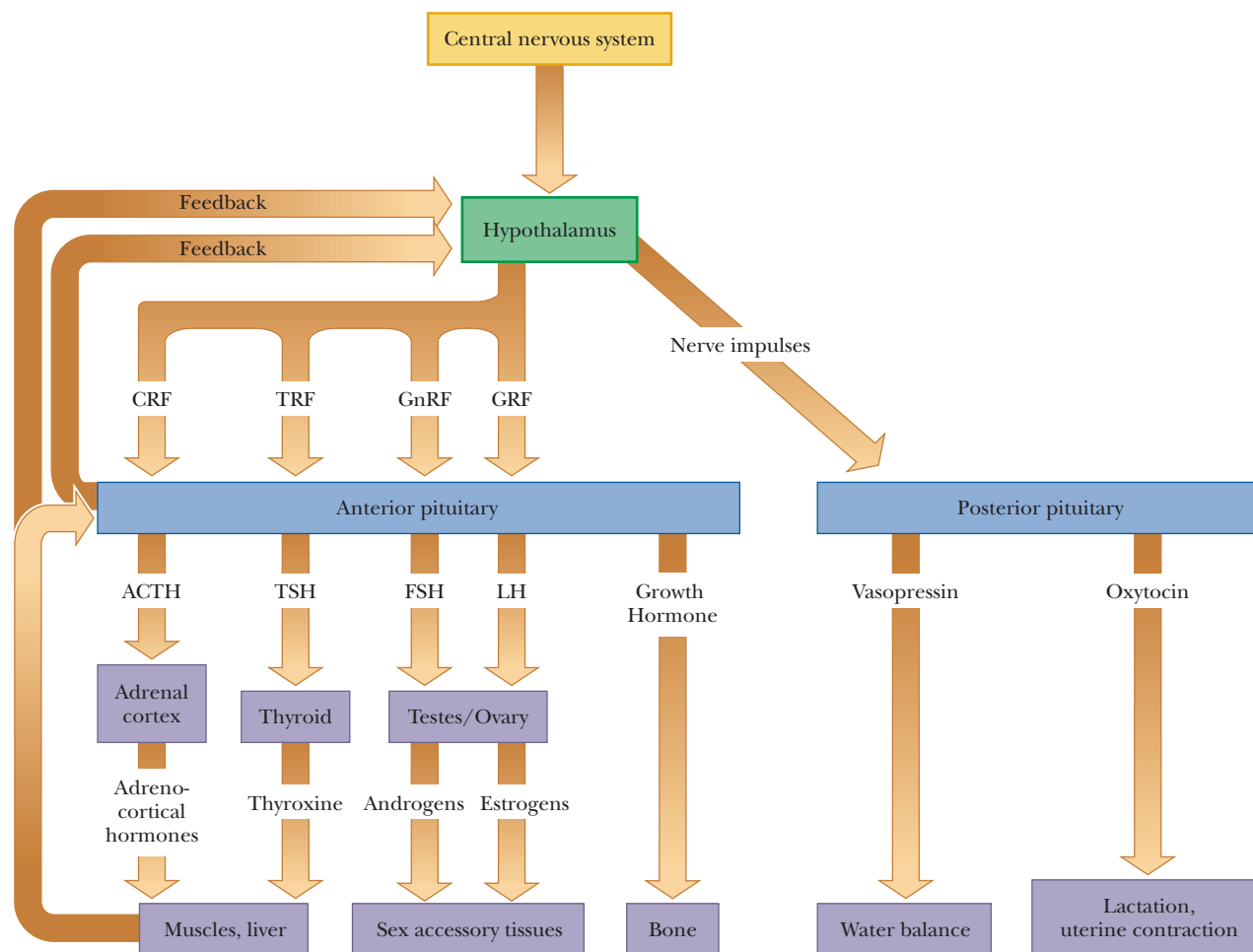
TABLE 24.3

Selected Human Hormones		
Hormone	Source	Major Effects
<b>Polypeptides</b>		
Corticotropin-releasing factor (CRF)	Hypothalamus	Stimulates release of ACTH
Gonadotropin-releasing factor (GnRF)	Hypothalamus	Stimulates release of FSH and LH
Thyrotropin-releasing factor (TRF)	Hypothalamus	Stimulates release of TSH
Growth hormone–releasing factor (GRF)	Hypothalamus	Stimulates release of growth hormone
Adrenocorticotrophic hormone (ACTH)	Anterior pituitary	Stimulates release of adrenocorticosteroids
Thyrotropin (TSH)	Anterior pituitary	Stimulates release of thyroxine
Follicle-stimulating hormone (FSH)	Anterior pituitary	In ovaries, stimulates ovulation and estrogen synthesis; in testes, stimulates spermatogenesis
Luteinizing hormone (LH)	Anterior pituitary	In ovaries, stimulates estrogen and progesterone synthesis; in testes, stimulates androgen synthesis
Met-enkephalin	Anterior pituitary	Has opioid effects on central nervous system
Leu-enkephalin	Anterior pituitary	Has opioid effects on central nervous system
$\beta$ -Endorphin	Anterior pituitary	Has opioid effects on central nervous system
Vasopressin	Posterior pituitary	Stimulates water resorption by kidney and raises blood pressure
Oxytocin	Posterior pituitary	Stimulates uterine contractions and flow of milk
Insulin	Pancreas ( $\beta$ -cells of islets of Langerhans)	Stimulates uptake of glucose from bloodstream
Glucagon	Pancreas ( $\alpha$ -cells of islets of Langerhans)	Stimulates release of glucose to bloodstream
<b>Steroids</b>		
Glucocorticoids	Adrenal cortex	Decrease inflammation, increase resistance to stress
Mineralocorticoids	Adrenal cortex	Maintain salt and water balance
Estrogens	Gonads and adrenal cortex	Stimulate development of secondary sex characteristics, particularly in females
Androgens	Gonads and adrenal cortex	Stimulate development of secondary sex characteristics, particularly in males
<b>Amino acid derivatives</b>		
Epinephrine	Adrenal medulla	Increases heart rate and blood pressure
Norepinephrine	Adrenal medulla	Decreases peripheral circulation, stimulates lipolysis in adipose tissue
Thyroxine	Thyroid	Stimulates metabolism generally

act on specific **endocrine glands**, which release the hormones to be transported to target organs. Note that feedback control is exerted at every stage of the process. Even more fine-tuning is possible with zymogen activation mechanisms (Section 7.4), which exist for many well-known hormones.

The chemical natures of hormones play a predictably important role in their roles in cell signaling. Steroid hormones, for example, can enter the cell directly through the plasma membrane or can bind to plasma membrane receptors. Nonsteroid hormones enter the cell exclusively as a result of binding to plasma membrane receptors (Figure 24.8).

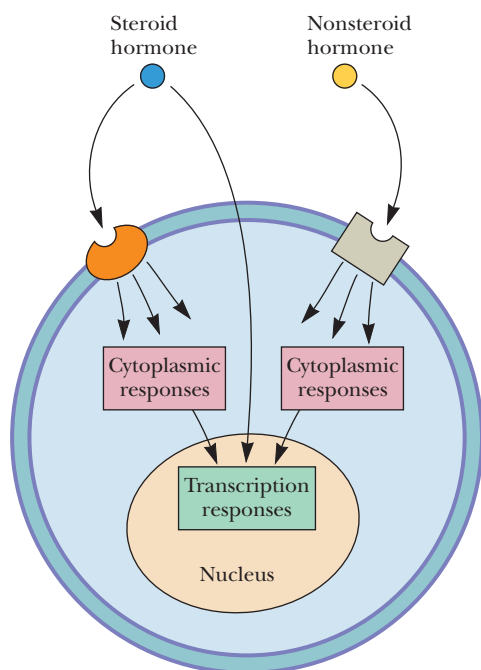
The releasing factors and trophic hormones listed in Table 24.3 tend to be polypeptides, but the chemical natures of the hormones released by specific endocrine glands show greater variation. Thyroxine, for example, produced by the thyroid, is an iodinated derivative of the amino acid tyrosine (Section 3.2). Abnormally low levels of thyroxine lead to **hypothyroidism**, characterized by lethargy and obesity, whereas increased levels produce the opposite effect



■ **FIGURE 24.7 Hormonal control.** This hormonal control system shows the role of the hypothalamus, pituitary, and target tissues. See Table 24.3 for the names of the hormones.

(*hyperthyroidism*). Low levels of iodine in the diet often lead to hypothyroidism and an enlarged thyroid gland (*goiter*). This condition has largely been eliminated by the addition of sodium iodide to commercial table salt (“iodized” salt). (It is virtually impossible to find table salt that is not iodized.)

Steroid hormones (Section 21.8) are produced by the adrenal cortex and the gonads (testes in males, ovaries in females). The **adrenocortical hormones** include **glucocorticoids**, which affect carbohydrate metabolism, modulate inflammatory reactions, and are involved in reactions to stress. The **mineralocorticoids** control the level of excretion of water and salt by the kidneys. If the adrenal cortex does not function adequately, one result is *Addison’s disease*, characterized by hypoglycemia, weakness, and increased susceptibility to stress. This disease is eventually fatal unless it is treated by administration of mineralocorticoids and glucocorticoids to make up for what is missing. The opposite condition, *hyperfunction of the adrenal cortex*, is frequently caused by a tumor of the adrenal cortex or of the pituitary. The characteristic clinical manifestation is *Cushing’s syndrome*, marked by hyperglycemia, water retention, and the easily recognized “moon face.”



■ **FIGURE 24.8 Hormone action.** Nonsteroid hormones bind exclusively to plasma membrane receptors, which mediate the cellular responses to the hormone. Steroid hormones exert their effects either by binding to plasma-membrane receptors or by diffusing to the nucleus, where they modulate transcriptional events.

The adrenal cortex produces some steroid sex hormones, the *androgens* and *estrogens*, but the main site of production is the gonads. Estrogens are required for female sexual maturation and function, but not for embryonic sexual development of female mammals. Animals that are male genetically appear to be females if they are deprived of androgens during embryonic development. As a final example, we shall discuss growth hormone (GH), which is a polypeptide. When overproduction of GH occurs, it is usually because of a pituitary tumor. If this condition occurs while the skeleton is still growing, the result is **gigantism**. If the skeleton has stopped growing before the onset of GH overproduction, the result is **acromegaly**, characterized by enlarged hands, feet, and facial features. Underproduction of GH leads to **dwarfism**, but this condition can be treated by the injection of human GH before the skeleton reaches maturity. Animal GH is ineffective in treating dwarfism in humans. Supplies of human GH were very limited when it could be obtained only from cadavers, but it can now be synthesized by recombinant DNA techniques. (Another discussion of peptide hormones can be found in the Biochemical Connections boxes on pages 77 and 79, which treat oxytocin and vasopressin.) Human growth hormone (HGH) has recently become available to individuals who believe it will help alleviate the effects of aging. It was known that the level of HGH decreases after middle age is reached. Many have assumed that the availability of growth hormone, if one could afford it, would be a virtual fountain of youth. Even though few results are conclusive at this time, HGH is being prescribed, and the medical community has adopted rules for its use. For example, doctors will consider prescribing it only for patients over age 40. The same hormone is also used illegally by endurance athletes.

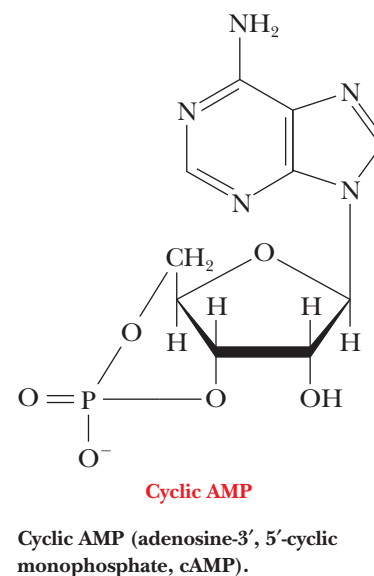
## Second Messengers

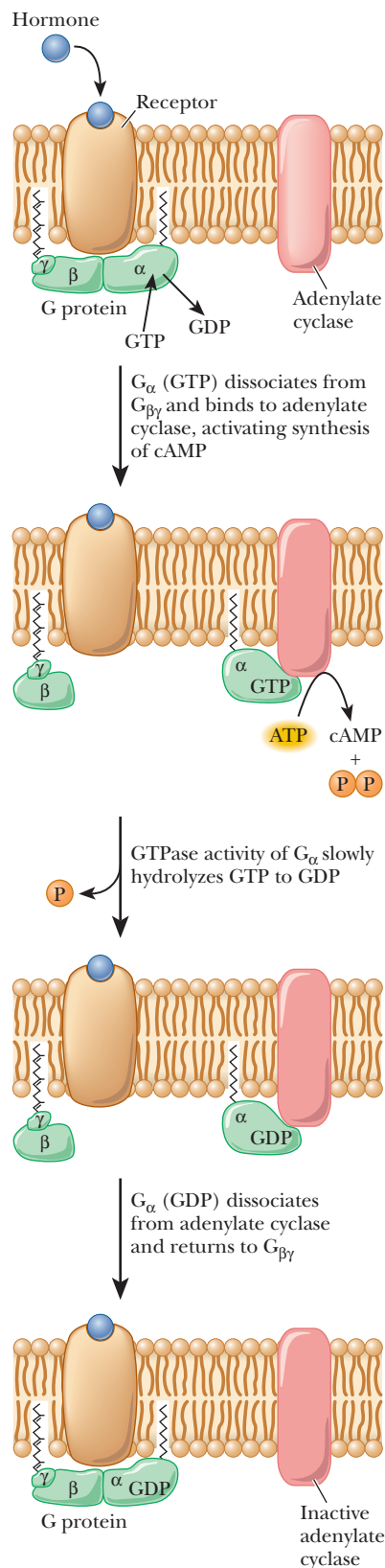
When a hormone binds to its specific receptor on a target cell, it sets off a chain of events in which the actual response within the cell is elicited. Several kinds of receptors are known. The receptors for steroid hormones tend to occur within the cell rather than as part of the membrane (steroids can pass through the plasma membrane); steroid–receptor complexes affect the transcription of specific genes. More frequently, the receptor proteins are a part of the plasma membrane. Binding of hormone to the receptor triggers a change in concentration of a second messenger. The **second messenger** brings about the changes within the cell as a result of a series of reactions.

### How do second messengers work?

#### Cyclic AMP and G Proteins

Cyclic AMP (adenosine-3',5'-monophosphate, cAMP) is one example of a second messenger. The mode of action starts with binding of a hormone to a specific receptor called a  $\beta_1$ - or  $\beta_2$ -adrenergic receptor, which triggers the production of cAMP from ATP, catalyzed by *adenylate cyclase*. This reaction is mediated by a stimulatory G protein, a trimer consisting of three subunits— $\alpha$ ,  $\beta$ , and  $\gamma$ . Binding of the hormone to the receptor activates the G protein; the  $\alpha$ -subunit binds GTP while releasing GDP, giving rise to the name of the protein. The active protein has GTPase activity and slowly hydrolyzes GTP, returning the G protein to the inactive state. GDP remains bound to the  $\alpha$ -subunit and must be exchanged for GTP when the protein is activated the next time (Figure 24.9). The G protein and adenylate cyclase are bound to the plasma membrane, while cAMP is released into the interior of the cell to act as a second messenger. As we have already seen in several pathways, cAMP stimulates protein kinase A, which phosphorylates a host of enzymes and transcription factors. Some examples are known in which the binding of hormone to receptor (an  $\alpha_2$ -receptor) inhibits rather than stimulates adenylate cyclase. A G protein with a different kind of  $\alpha$ -subunit mediates the process. The modified G protein is referred to as an





*inhibitory G protein* to distinguish it from the kind that stimulates response to hormone binding (Figure 24.10).

In eukaryotic cells, the usual mode of action of cAMP is to stimulate a cAMP-dependent protein kinase, a tetramer consisting of two regulatory subunits and two catalytic subunits. When cAMP binds to the dimer of regulatory subunits, the two active catalytic subunits are released. The active kinase catalyzes the phosphorylation of a target enzyme or transcription factor (Figure 24.11). In the scheme shown in Figure 24.11, phosphorylation activates the enzyme. Cases are also known in which phosphorylation inactivates a target enzyme (e.g., glycogen synthase; Section 24.4). The usual site of phosphorylation is the hydroxyl group of a serine or a threonine. ATP is the source of the phosphate group that is transferred to the enzyme. The target enzyme then elicits the cellular response.

G proteins are very important signaling molecules in eukaryotes. They can be activated by combinations of hormones. For example, both epinephrine and glucagon act via a stimulatory G protein in liver cells. The effect can be cumulative, so that if both glucagons and epinephrine have been released, the cellular effect is greater. Besides the effect on cAMP, G proteins are involved in activating many other cellular processes, including stimulating phospholipase C and opening or closing membrane ion channels. They are also involved in vision and smell. There are currently more than 100 known G protein-coupled receptors and more than 20 known G proteins.

A G protein is permanently activated by cholera toxin, leading to excessive stimulation of adenylate cyclase and chronic elevation of cAMP levels. The main danger in *cholera*, caused by the bacterium *Vibrio cholerae*, is severe dehydration as a result of diarrhea. The unregulated activity of adenylate cyclase in epithelial cells leads to the diarrhea because cAMP in epithelial cells stimulates active transport of  $\text{Na}^+$ . Excessive cAMP in the epithelial cells of the small intestine produces a large flow of  $\text{Na}^+$  and water from the mucosal surface of the epithelial cells into the lumen of the intestine. If the lost fluid and salts can be replaced in cholera victims, the immune system can eliminate the actual infection within a few days.

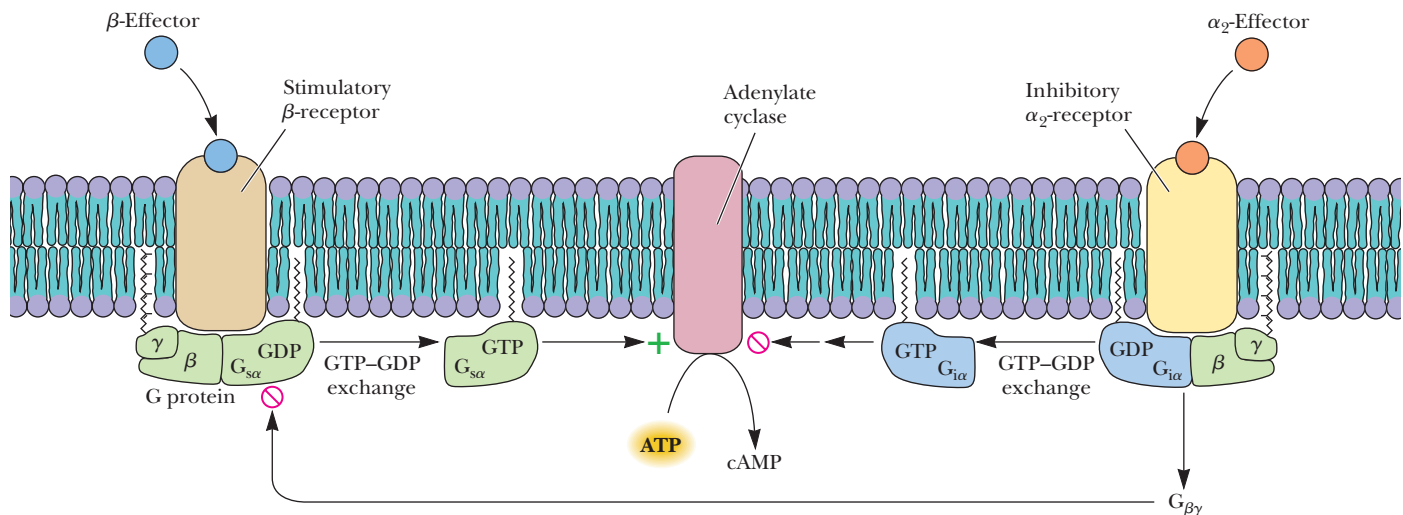
### Calcium Ion as a Second Messenger

Calcium ion ( $\text{Ca}^{2+}$ ) is involved in another ubiquitous second-messenger scheme. Much of the calcium-mediated response depends on release of  $\text{Ca}^{2+}$  from intracellular reservoirs, similar to the release of  $\text{Ca}^{2+}$  from the sarcoplasmic reticulum in the action of the neuromuscular junction. A component of the inner layer of the phospholipid bilayer, *phosphatidylinositol 4,5-bisphosphate* ( $\text{PIP}_2$ ), is also required in this scheme (Figure 24.12).

When the external trigger binds to its receptor on the cell membrane, it activates *phospholipase C* (Section 21.2), which hydrolyzes  $\text{PIP}_2$  to *inositol 1,4,5-triphosphate* ( $\text{IP}_3$ ) and a *diacylglycerol* (DAG), in a process mediated by a different member of the family of G proteins. The  $\text{IP}_3$  is the actual second messenger. It diffuses through the cytosol to the endoplasmic reticulum (ER), where it stimulates the release of  $\text{Ca}^{2+}$ . A complex is formed between the calcium-binding protein calmodulin and  $\text{Ca}^{2+}$ . This calcium-calmodulin complex activates a cytosolic protein kinase, which phosphorylates target enzymes in the same fashion as in the cAMP second-messenger scheme. DAG also plays a

**FIGURE 24.9** Activation of adenylate cyclase by heterotrimeric G proteins. Binding of hormone to its receptor causes a conformational change that induces the receptor to catalyze a replacement of GDP by GTP on  $G_{\alpha}$ . The  $G_{\alpha}$  (GTP) complex dissociates from  $G_{\beta\gamma}$  and binds to adenylate cyclase, stimulating synthesis of cAMP. Bound GTP is slowly hydrolyzed to GDP by the intrinsic GTPase activity of  $G_{\alpha}$ .  $G_{\alpha}$  (GDP) dissociates from adenylate cyclase and reassociates with  $G_{\beta\gamma}$ .  $G_{\alpha}$  and  $G_{\gamma}$  are lipid-anchored proteins. Adenylate cyclase is an integral membrane protein consisting of 12 transmembrane  $\alpha$ -helical segments.



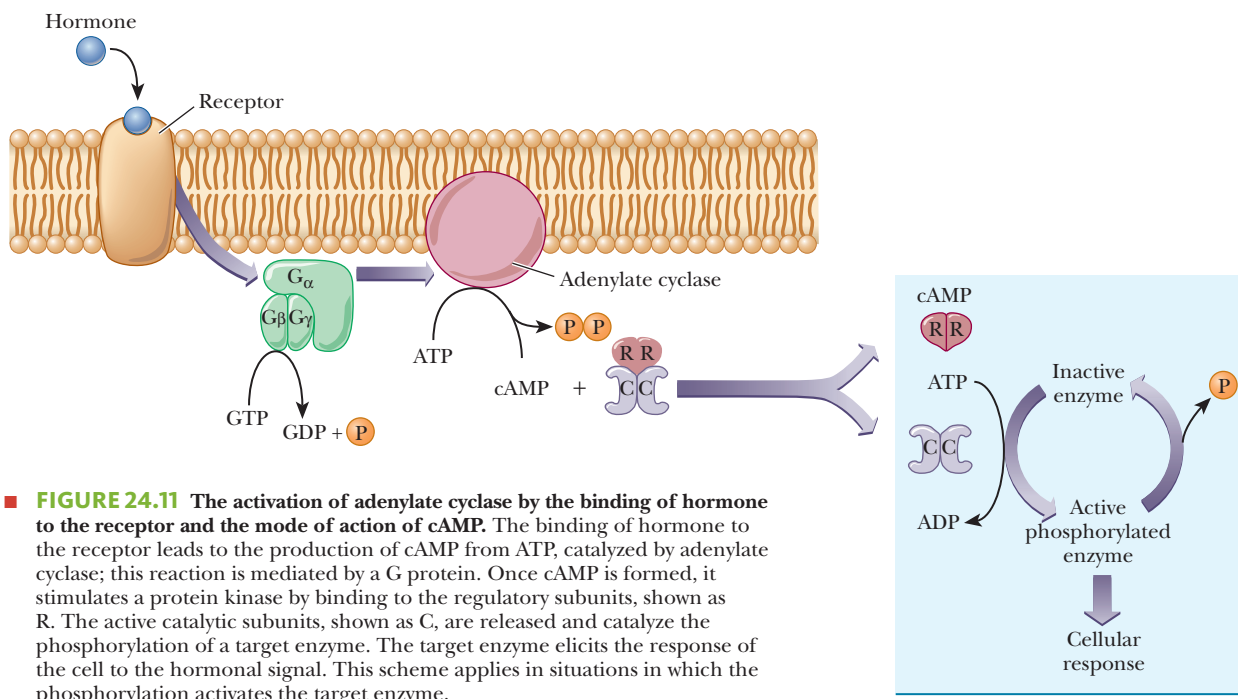


**FIGURE 24.10 Control of adenylate cyclase.** Adenylate cyclase activity is modulated by the interplay of stimulatory ( $G_s$ ) and inhibitory ( $G_i$ ) G proteins. Binding of hormones to  $\beta$ -receptors activates adenylate cyclases via  $G_s$ , whereas hormone binding to  $\alpha_2$  receptors leads to the inhibition of adenylate cyclase. Inhibition may occur by direct inhibition of cyclase activity by  $G_{i\alpha}$  or by binding of  $G_{i\beta\gamma}$  to  $G_{s\alpha}$ .

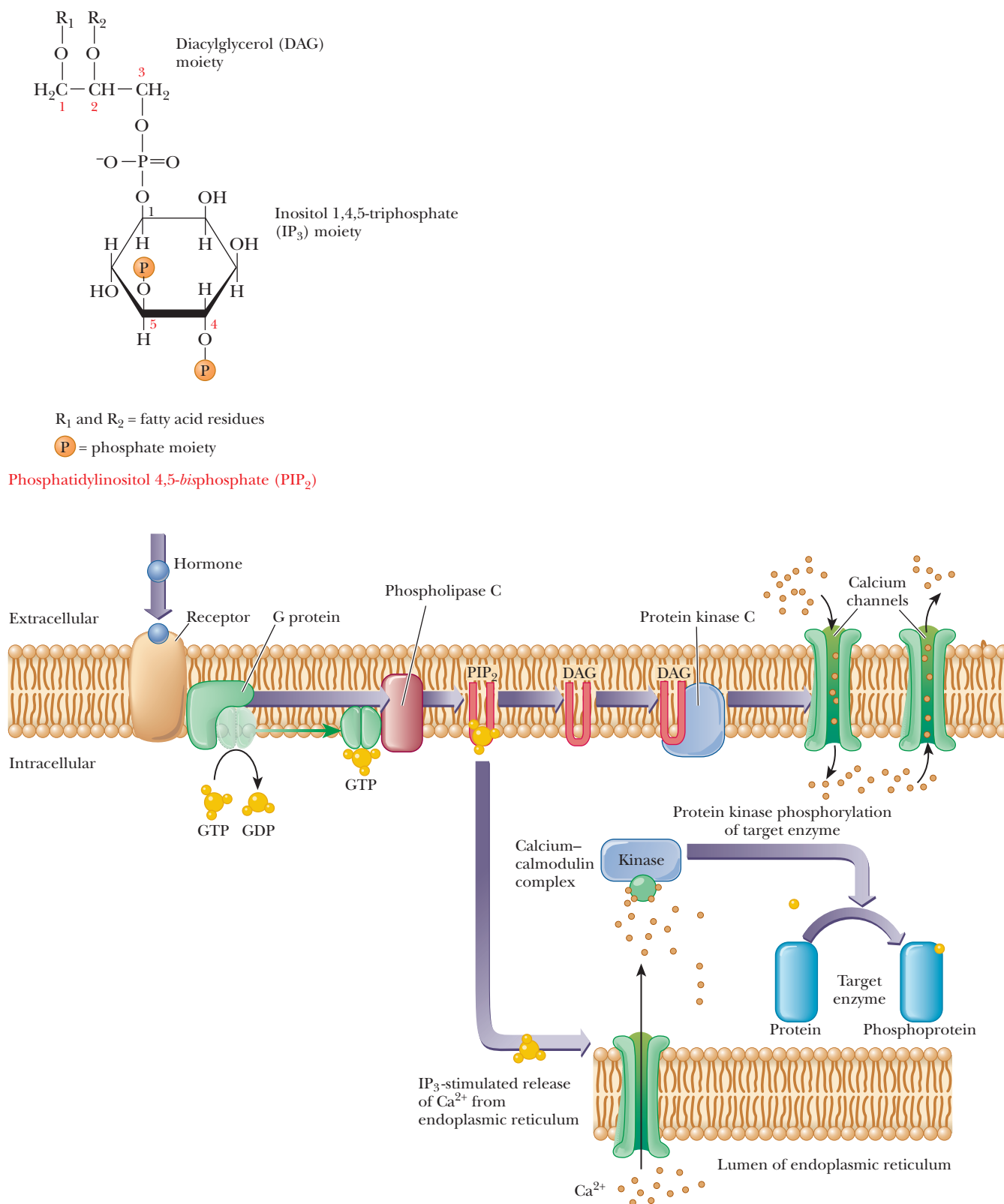
role in this scheme; it is nonpolar and diffuses through the plasma membrane. When DAG encounters the membrane-bound protein kinase C, it too acts as a second messenger by activating this enzyme (actually a family of enzymes). **Protein kinase C** also phosphorylates target enzymes, including channel proteins that control the flow of  $Ca^{2+}$  into and out of the cell. By controlling the flow of  $Ca^{2+}$ , this second-messenger system can produce sustained responses even when the supply of  $Ca^{2+}$  in the intracellular reservoirs becomes exhausted.

### Receptor Tyrosine Kinases

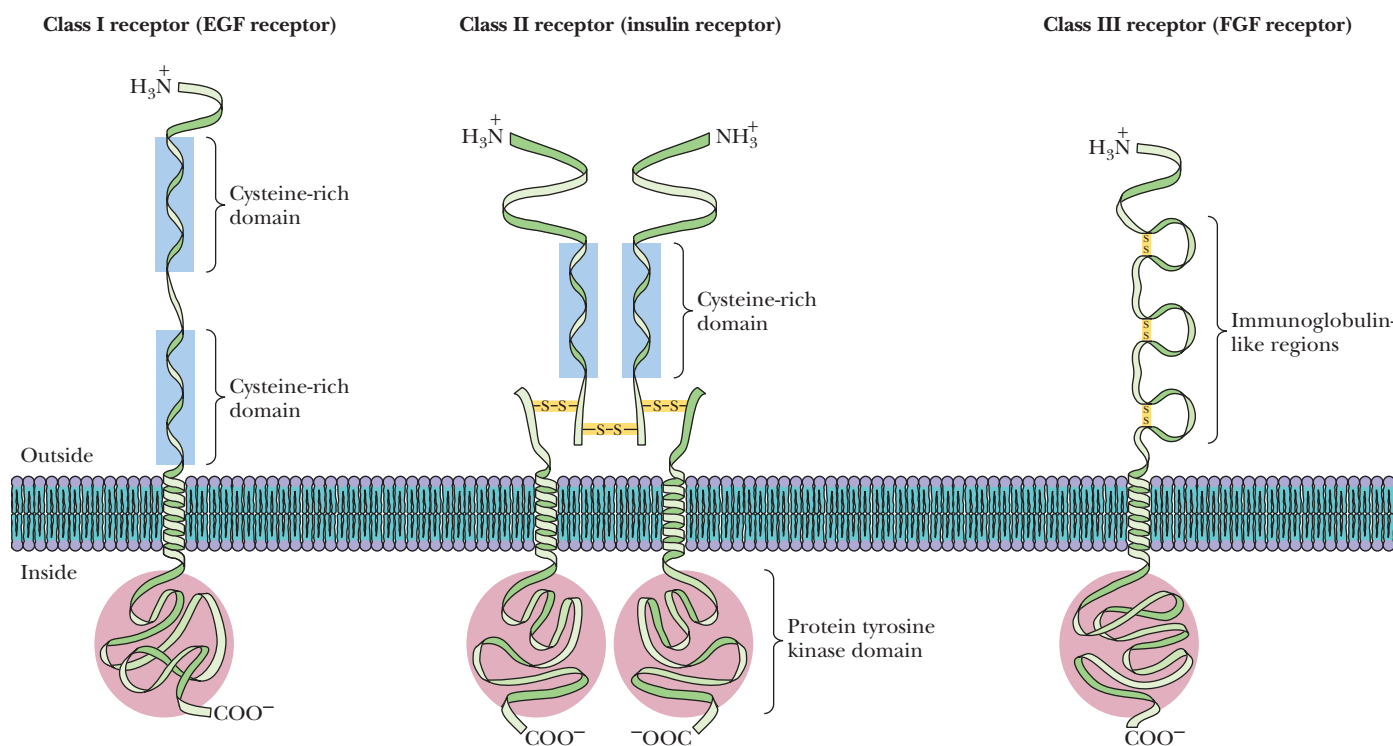
Another important type of second-messenger system involves a receptor type called a **receptor tyrosine kinase**. These receptors span the membrane of the



**FIGURE 24.11 The activation of adenylate cyclase by the binding of hormone to the receptor and the mode of action of cAMP.** The binding of hormone to the receptor leads to the production of cAMP from ATP, catalyzed by adenylate cyclase; this reaction is mediated by a G protein. Once cAMP is formed, it stimulates a protein kinase by binding to the regulatory subunits, shown as R. The active catalytic subunits, shown as C, are released and catalyze the phosphorylation of a target enzyme. The target enzyme elicits the response of the cell to the hormonal signal. This scheme applies in situations in which the phosphorylation activates the target enzyme.



■ **FIGURE 24.12 The PIP<sub>2</sub> second-messenger scheme.** When a hormone binds to a receptor, it activates phospholipase C, in a process mediated by a G protein. Phospholipase C hydrolyzes PIP<sub>2</sub> to IP<sub>3</sub> and DAG. IP<sub>3</sub> stimulates the release of Ca<sup>2+</sup> from intracellular reservoirs in the ER. A complex formed between Ca<sup>2+</sup> and the calcium-binding protein calmodulin activates a cytosolic protein kinase for phosphorylation of a target enzyme. DAG remains bound to the plasma membrane, where it activates the membrane-bound protein kinase C (PKC). PKC is involved in the phosphorylation-channel proteins that control the flow of Ca<sup>2+</sup> in and out of the cell. Ca<sup>2+</sup> from extracellular sources can produce sustained responses even when the supply of Ca<sup>2+</sup> in intracellular reservoirs is exhausted.



■ **FIGURE 24.13** The three classes of receptor tyrosine kinases. Class I receptors are monomeric and contain a pair of Cys-rich repeat sequences. The insulin receptor, a typical class II receptor, is a glycoprotein composed of two kinds of subunits in a  $\alpha_2\beta_2$  tetramer. The  $\alpha$ - and  $\beta$ -subunits are synthesized as a single peptide chain, together with an N-terminal signal sequence. Subsequent proteolytic processing yields the separate  $\alpha$ - and  $\beta$ -subunits. The  $\beta$ -subunits, of 620 residues each, are integral transmembrane proteins, with only a single transmembrane  $\alpha$ -helix and with the amino terminus outside the cell and the carboxyl terminus inside. The  $\alpha$ -subunits, of 735 residues each, are extracellular proteins that are linked to the  $\beta$ -subunits and to each other by disulfide bonds. The insulin-binding domain is located in a cysteine-rich region on the  $\alpha$ -subunits. Class III receptors contain multiple immunoglobulin-like domains. Shown here is fibroblast growth factor (FGF) receptor, which has three immunoglobulin-like domains. (Adapted from A. Ullrich and J. Schlessinger, 1990. *Signal Transduction by Receptors with Tyrosine Kinase Activity*. Cell, **61**, 203–212 [1990].)

cell and have a hormone receptor on the outside and a tyrosine kinase portion on the inside. There are several subclasses of these receptor kinases, as shown in Figure 24.13. The best known of these is class II, which includes the insulin receptor (which we will look at in more detail in Section 24.5).

These kinases are allosteric enzymes. When the hormone binds to the binding region on the outside of the cell, it induces a conformational change in the tyrosine kinase domain that activates the kinase activity. The activated tyrosine kinases phosphorylate tyrosines on a variety of target proteins, causing alterations in membrane transport of ions and amino acids and the transcription of certain genes. Phospholipase C (seen in Figure 24.12) is one of the targets of tyrosine kinases. Another is an **insulin-sensitive protein kinase**, which phosphorylates and activates protein phosphatase 1.

## 24.4 Hormones and the Control of Metabolism

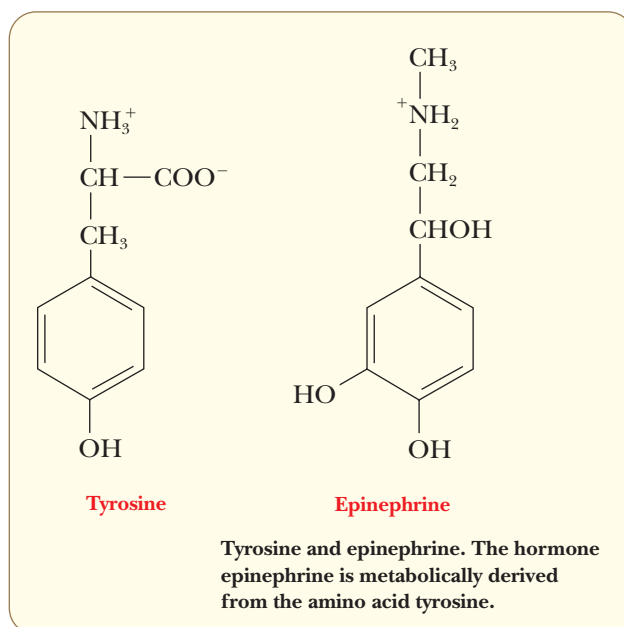
Now that we know something about the effects of hormones in triggering responses within the cell, we can return to and expand on some earlier points about metabolic control. In Section 18.3, we discussed some points about

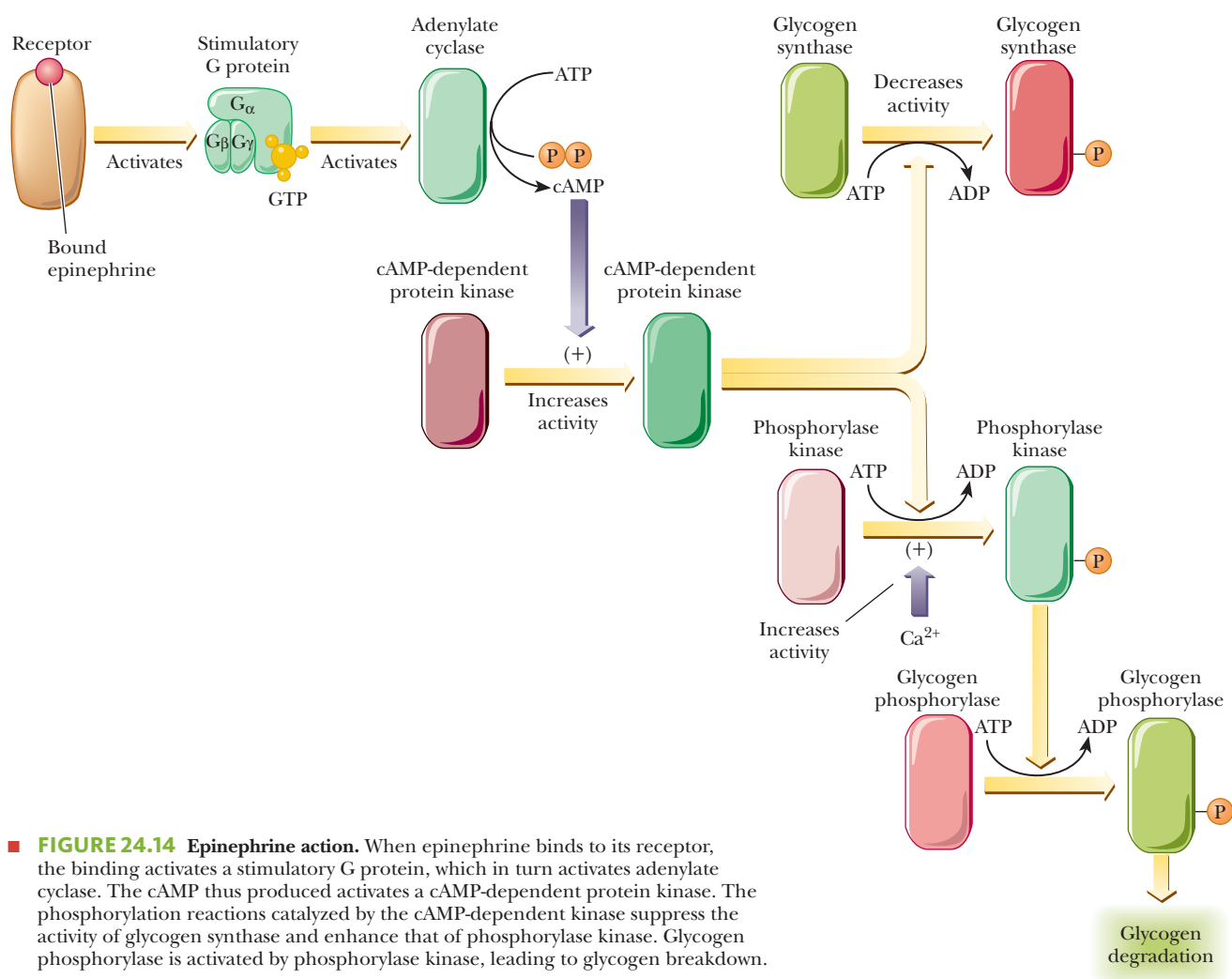
control mechanisms in carbohydrate metabolism. We saw at that time how glycolysis and gluconeogenesis can be regulated and how glycogen synthesis and breakdown can respond to the body's needs. Phosphorylation and dephosphorylation of the appropriate enzymes played a large role there, and that whole scheme is subject to hormonal action.

### What hormones control carbohydrate metabolism?

Three hormones play a part in the regulation of carbohydrate metabolism: epinephrine, glucagon, and insulin. Epinephrine acts on muscle tissue to raise levels of glucose on demand, while glucagon acts on the liver, also to increase the availability of glucose. Feedback control plays a role in the process and ensures that the amount of glucose made available does not reach an excessive level (Section 24.3). The role of insulin is to trigger the feedback response that achieves this further control.

Epinephrine (also called *adrenalin*) is structurally related to the amino acid tyrosine. Epinephrine is released from the adrenal glands in response to stress (the “fight or flight” response). When it binds to specific receptors, it sets off a chain of events that leads to increased levels of glucose in the blood, increased glycolysis in muscle cells, and increased breakdown of fatty acids for energy. Glucagon (a peptide that contains 29 amino acid residues) is released by the  $\alpha$ -cells of the islets of Langerhans in the pancreas, and it too binds to specific receptors to set off a chain of events to make glucose available to the organism. Each time a single hormone molecule, whether epinephrine or glucagon, binds to its specific receptor, it activates a number of stimulatory G proteins. This effect starts an amplification of the hormonal signal. Each active G protein in turn stimulates adenylate cyclase several times before the G protein is inactivated by its own GTPase activity, leading to still more amplification. The cAMP produced by the increased adenylate cyclase activity allows for increased activity of the cAMP-dependent protein kinase, phosphorylating target enzymes that lead to increased glucose levels. In particular, this means an increase in the activity of the enzymes involved in gluconeogenesis and glycogen breakdown as well as a decrease in the activity of enzymes involved in glycolysis and glycogen synthesis. The series of amplifying steps is called a **cascade**, and the cumulative effect is the underlying reason why small amounts of hormones can have such marked effects.



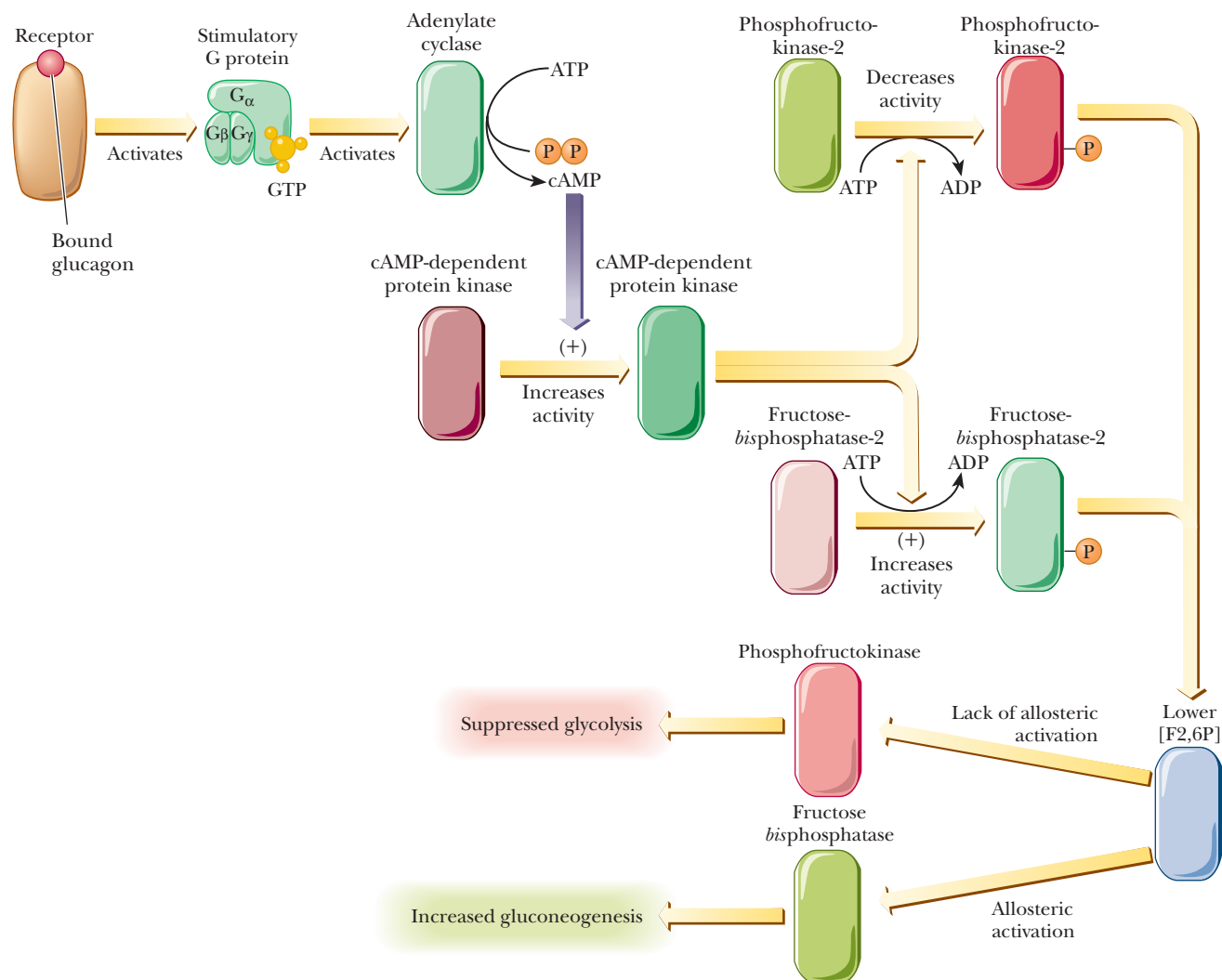


■ **FIGURE 24.14 Epinephrine action.** When epinephrine binds to its receptor, the binding activates a stimulatory G protein, which in turn activates adenylate cyclase. The cAMP thus produced activates a cAMP-dependent protein kinase. The phosphorylation reactions catalyzed by the cAMP-dependent kinase suppress the activity of glycogen synthase and enhance that of phosphorylase kinase. Glycogen phosphorylase is activated by phosphorylase kinase, leading to glycogen breakdown.

Figure 24.14 shows how the binding of epinephrine to specific receptors leads to increased glycogen breakdown in muscle and suppression of glycogen synthesis. The hormonal stimulation leads to activation of adenylate cyclase, which in turn activates the cAMP-dependent protein kinase responsible for activating glycogen phosphorylase and inactivating glycogen synthase.

The effect of glucagon binding to receptors in stimulating gluconeogenesis in the liver and suppressing glycolysis depends on changes in the concentration of the key allosteric effector, fructose-2,6-bisphosphate (F2,6P). Recall from Section 18.3 that this compound is an important allosteric activator of phosphofructokinase, the key enzyme of glycolysis; it is also an inhibitor of fructose-bisphosphate phosphatase, which plays a role in gluconeogenesis. A high concentration of F2,6P stimulates glycolysis, whereas a low concentration stimulates gluconeogenesis. The concentration of F2,6P in a cell depends on the balance between its synthesis [catalyzed by phosphofructokinase-2 (PFK-2)] and its breakdown [catalyzed by fructose-bisphosphatase-2 (FBPase-2)]. The enzyme activities (on a single multifunctional protein) that control the formation and breakdown of F2,6P are themselves controlled by a phosphorylation/dephosphorylation mechanism, which in turn is subject to the same sort of hormonal control we just discussed for the enzymes of glycogen metabolism. Figure 24.15 summarizes the chain of events that leads to increased gluconeogenesis in the liver as a result of the binding of glucagon to its specific receptor. The following Biochemical Connections box discusses the role of insulin in overall metabolism and some current dieting trends.





■ **FIGURE 24.15 Glucagon action.** Binding of glucagon to its receptor sets off the chain of events that leads to the activation of a cAMP-dependent protein kinase. The enzymes phosphorylated in this case are phosphofructokinase-2, which is inactivated, and fructose-bisphosphatase-2, which is activated. The combined result of phosphorylating these two enzymes is to lower the concentration of fructose-2,6-bisphosphate (F2,6P). A lower concentration of F2,6P leads to allosteric activation of the enzyme fructose-bisphosphatase, thus enhancing gluconeogenesis. At the same time, the lower concentration of F2,6P implies that phosphofructokinase is lacking a potent allosteric activator, with the result that glycolysis is suppressed.

## 24.5 Insulin and Its Effects

To the average person, insulin is best known as the hormone that is deficient in people with diabetes, and it certainly was this relationship that spurred on the study of this fascinating hormone. We are just now realizing that insulin is involved in many cellular processes and in many different ways from what was previously thought.

### What is insulin?

Insulin is a peptide hormone secreted from the pancreas. In its active form, it is a 51-amino-acid peptide with two different chains, the A chain and the B chain, held together by disulfide bonds. As described in Chapter 13, insulin was one of the first proteins to be cloned and expressed for human need. Insulin is created as an 86-residue precursor, called *proinsulin*. Residues 31 to 65 of proinsulin



# Biochemical Connections NUTRITION

## Insulin and Low-Carbohydrate Diets

In the 1970s, diets very high in carbohydrates became popular, with both athletes and the population at large. It was felt that a diet consisting of 60%–70% carbohydrates and 15%–20% each of fat and protein would be the healthiest (because of the high carbohydrate-to-fat ratio) and the best for athletes (because of the high levels of carbohydrates for replenishing glycogen). In the 1990s, newer diets that were based on a lower carbohydrate level became fashionable. Instead of a 70/15/15 ratio of carbohydrate/protein/fat, these diets recommended a 60/20/20 ratio or even a 50/25/25 ratio. The most notable of these diets was the one called the Zone Diet, promoted by Dr. Barry Sears. The idea behind such diets is that a calorie is not always a calorie. In other words, the source of the calorie does matter. People had maintained high-carbohydrate diets in the belief that carbohydrates were healthier than fats, because of the myriad health problems that were attributed to too much dietary fat. Even though this seemed logical and few would argue in favor of the benefits of excessive fat, there is a possible downside to too many carbohydrates. For one thing, excess carbohydrates become fat. This may be a big consideration to nonathletes, who do not need to replenish muscle and liver glycogen as quickly and as often as an endurance athlete would. Also, a high-carbohydrate meal stimulates the production of insulin. Insulin inhibits the body's ability to use fat for energy and stimulates the uptake of fat and its storage as triacylglycerol. A high-carbohydrate meal also has the potential of causing **reactive hypoglycemia**, which occurs when high blood glucose stimulates a large insulin release, which then proceeds to clear too much glucose from the blood, causing a blood sugar crash shortly

thereafter. Many people find themselves weak, shaky, or sleepy by 10 A.M. after a high-carbohydrate breakfast. As we saw in Section 24.2, replacing fat with carbohydrates does nothing to increase the HDL/LDL ratio, either. The Zone Diet was designed to avoid reactive hypoglycemia and the effects insulin has on fat storage. Because of the differences between fats and carbohydrates, many people may also tend to eat many more calories in the form of carbohydrates than they would in the form of fats, because carbohydrates do not give the same “filling” sensation as fats. Diet is a very personal thing. Many people have found that they feel better and actually lose weight while on a lower-carbohydrate diet. Others find just the opposite. There has been little evidence to suggest that a lower-carbohydrate diet is effective for athletes, however.



©Lightworks/Alamy



©Jupiter Images/Comstock/Alamy

## Insulin Affects Many Enzymes

Insulin affects the activity of many enzymes, most of which are involved in getting rid of glucose. However, fat metabolism is also affected. Glucokinase is a liver enzyme that phosphorylates glucose to glucose-6-phosphate (see the Biochemical Connections box on page 157). It is induced by insulin, so that when insulin is present, glucose in the liver is sent toward catabolic pathways, such as pentose phosphate or glycolysis. Insulin also activates liver glycogen synthase and deactivates glycogen phosphorylase, causing glucose to be put into a polymeric form. In addition, insulin stimulates glycolysis through activation of phosphofructokinase and pyruvate dehydrogenase. Insulin also has a large effect on fatty-acid metabolism. It increases fatty-acid synthesis via stimulation of acetyl-CoA carboxylase (ACC), and it increases triacylglycerol synthesis in the liver through activation of lipoprotein lipase. It also increases cholesterol synthesis via activation of hydroxymethylglutaryl-CoA reductase (Section 21.8). Table 24.4 summarizes the effects of insulin on metabolism.

## Diabetes

Much has been learned about insulin because of its relationship with diabetes. In classical, type I diabetes (or insulin-dependent diabetes), the individual does not make insulin, or at least not enough of it. This is usually caused by destruction of the beta cells of the islets of Langerhans in the pancreas resulting from a type of autoimmune disease. The only remedy for type I diabetes is regular insulin injections, and insulin is produced for this purpose by recombinant DNA technology (Chapter 13).

The medical community is also concerned about the large increases in type II diabetes (or non-insulin-dependent diabetes), which is characterized by cells

TABLE 24.4

Effect of Insulin on Metabolism			
Metabolic Process	Location	Effect	Target
Glucose uptake	Muscle	Increases	GLUT4 transporter
Glucose breakdown	Liver	Increases	Glucokinase
Glycolysis	Muscle and liver	Increases	PFK-1
Acetyl-CoA production	Muscle and liver	Increases	Pyruvate dehydrogenase
Glycogen synthesis	Muscle and liver	Increases	Glycogen synthase
Glycogen breakdown	Muscle and liver	Decreases	Glycogen phosphorylase
Fatty-acid synthesis	Liver and muscle	Increases	Acetyl-CoA carboxylase
Triacylglycerol synthesis	Adipocytes	Increases	Lipoprotein lipase

not responding correctly to insulin. In these cases, the person may make a normal amount of the hormone, but it does not have sufficient effect, either because it does not bind correctly to the receptor, or because the receptor does not correctly transmit the second messenger. This disease often begins later in life and is then called **adult-onset diabetes**. Whereas people with type I diabetes are often thin, people with type II diabetes are often obese. Evidence suggests that type II diabetes in the elderly is related to dysfunction of muscle mitochondria.

One of the most recent discoveries is that people with type II diabetes also have an increased risk of Alzheimer's disease. In this type of diabetes, insulin is increased because it takes more insulin to accomplish the same clearance of glucose into the cells. Insulin appears to increase levels of the  $\beta$ -amyloid protein that forms plaques in the brain. A brain protein called **insulin-degrading enzyme (IDE)** is involved in binding to and degrading insulin. This enzyme also binds to  $\beta$ -amyloid protein and clears it from the brain. When insulin levels are high, IDE spends more time tied up with insulin and less time clearing the  $\beta$ -amyloid protein. Because insulin is produced in high quantities when a

## Biochemical Connections ALLIED HEALTH

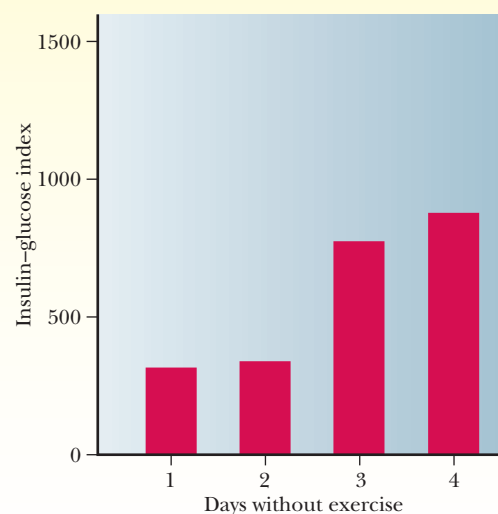
### A Workout a Day Keeps Diabetes Away?

There seems to be a link between obesity and type II diabetes, although it is not clear whether diabetes leads to obesity or vice versa. The GLUT4 transporter is one of many glucose transporters, and it is the one most affected by insulin levels. It is also a protein whose levels can be affected by physical training. Studies have shown that one of the major changes associated with physical activity is an increase in the amount of GLUT4 in the muscle. In the trained state, a person transports more glucose into the cell than when untrained. Studies showed that only one week of moderate exercise (1–2 hours a day at 70% of maximal oxygen uptake) would double the GLUT4 protein content of the muscles of sedentary people.

By definition, loss of function of glucose transport is type II diabetes. The training effect is such that it takes only a few days without training for the activity of GLUT4 to decrease to only half its normal level. Fortunately, the intensity of the training has less

to do with the effect, at least in young to middle-aged people. With the apparent link between type II diabetes and obesity, one method of maintaining proper glucose transport appears to be staying light and fit.

■ **Insulin–glucose index versus detraining time.** Moderately trained middle-aged men were tested for the effect of detraining on their muscles' ability to clear glucose out of the blood (measured as the insulin–glucose index—the amount of insulin it takes to clear glucose from the blood). This shows that on the third day without training, there is a pronounced increase in the amount of insulin required to clear glucose. (Adapted from “Metabolic basis of the health benefits of exercise” by Adrienne Hardman, *The Biochemist* 20 [3], pp. 18–22 [1998].)





person eats a high-carbohydrate meal, one could easily theorize that the increasing number of people getting Alzheimer's disease could be tied to our fast-food, high-sugar diet and lifestyle.

### Insulin and Sports

Athletes must be able to control their diets for maximum performance. Insulin plays a large role in the choice of a pre-race breakfast for aerobic athletes. If an athlete eats a large, carbohydrate-filled breakfast in the morning, there is a rise in blood glucose, followed by a rise in insulin. This often leads to a fall in blood glucose below the baseline level. It may take hours for the blood glucose levels to come back up. In addition, the high insulin level would cause activation of

## Biochemical Connections ALLIED HEALTH

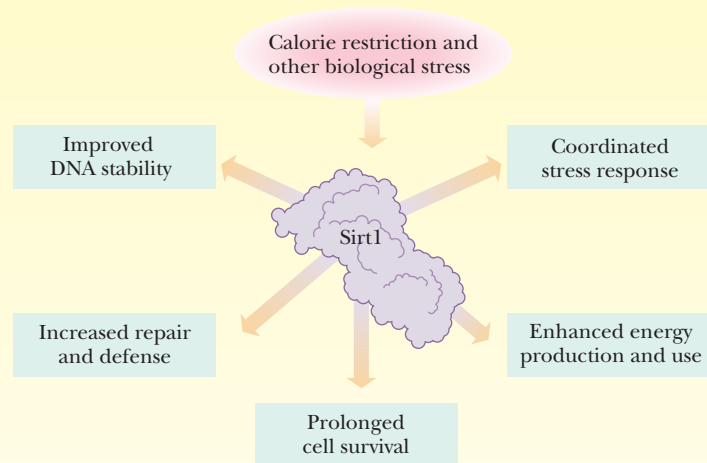
### Aging and the Search for Longevity

Soon after the dawn of humanity, people discovered that if they lived long enough, they experienced a gradual deterioration in health as they got old. Immediately after that, they realized they did not like it. From that moment forward, humankind has been obsessed with finding ways to turn back the clock, or at least to stop it from advancing. In this book we have seen many references to practical applications of biochemistry that can lead to a higher quality of life and perhaps even a longer one. Most of these have been intuitive, such as maintaining a healthy lifestyle through diet and exercise. However, we are never satisfied. This is the age of instant gratification, and we would immediately invest in a company that could develop an anti-aging pill—a fountain of youth in a bottle.

Although the exact causes of aging are still not clear, we believe it to be the gradual wearing out over time of the body's natural ability to maintain itself and repair damages. Logic dictates that natural selection cannot help our longevity, because the difference between living to 70 and living to 130 happens after the reproductive years, so there is no selective pressure for longevity. In other words, if there is a gene for longevity, it can only be selectively passed on if it affects reproductive success.

We discovered more than 70 years ago that calorie restriction (CR) is associated with increased longevity in lifeforms as varied as yeast and rodents. Just last year, primates were added to that list. In some species, restricting caloric intake by 30% compared to normal levels was shown to increase life span by 30% or more. This technique is still the only absolutely proven method of extending life span, other than not smoking and avoiding the more obvious dangerous behaviors. In addition to the life span extension, CR leads to a higher quality of life and forestalls many diseases, such as cancer, diabetes, inflammation, and even neurodegenerative diseases. Many mechanisms for this longevity increase have been suggested, including general health benefits of weight reduction and specific improvements in DNA management due to lower levels of oxidative compounds that are created as by-products of metabolism. However, about 15 years ago researchers began to pinpoint a family of genes in the yeast *Saccharomyces cerevisiae* that seemed to be at the center of these increases in longevity due to CR. The best characterized of these genes is *SIR2* in yeast. *SIR2* is a member of a family of genes called the *sirtuins*, and evidence indicates that they are key regulators of the longevity mechanism. In yeast and in roundworms, genetic manipulations that doubled the number of *SIR2* genes increased life span by 50%!

The protein product of the mammalian version of the *SIR2* gene is a protein called Sirt1. It is a stress-induced protein deacetylase that is dependent on  $\text{NAD}^+$ . It regulates cell survival, replicative senescence, inflammation, and metabolism via the deacetylation of histones (Chapter 10). Calorie restriction is



■ **Sirt1 and its putative relationship to health and longevity.** The Sirt1 enzyme appears to be responsible for the health and longevity-enhancing effects of calorie restriction in mammals. Food scarcity and other biological stressors trigger increased activity by Sirt1, which in turn alters activities within cells. By boosting manufacture of certain signaling molecules, such as insulin, Sirt1 may also coordinate the stress response throughout the body. (Reprinted by permission of Scientific American, "Unlocking the Secrets of Longevity Genes" by David A. Sinclair and Lenny Guarente, March 2006.)

a biological stressor like natural food scarcity. Sirt1 seems to be at the center of a generalized response to stress that primes the organism for survival. As the figure below shows, Sirt1 in mammals occupies a pivotal role in longevity through improved DNA stability, increased repair and defense, prolonged cell survival, enhanced energy production and use, and other coordinated stress responses. Mice that have been engineered to lack SIRT1 do not show the longevity increase associated with calorie restriction. Furthermore, doubling the number of SIRT1 genes in an organism renders it unresponsive to calorie restriction. Thus, it is now generally accepted that calorie restriction promotes longevity by activation of sirtuins in general and Sirt1 in particular.

Of course, humans prefer not to live a life of deprivation in order to reap the benefits of life span extension; and thus, the search for a stimulator of SIRT1 was on. One of the first compounds found that is a natural activator of sirtuins is a small molecule called resveratrol, a polyphenol that is present in red wine and made by many plants when stressed. Feeding resveratrol to yeast, worms, or flies, or placing them on a calorie-restricted diet, extends their life spans about 30%, but only if they possess the *SIR2* gene. Increased levels of Sirt1



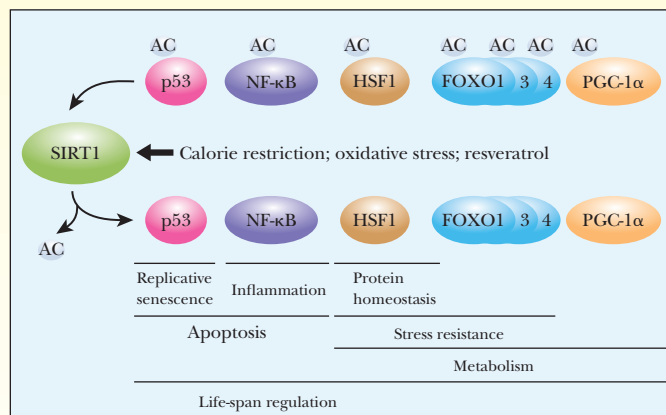
fat and glycogen synthesis and would inhibit glycogen breakdown. Thus, for a period of time after the high-carbohydrate breakfast, the athlete would be essentially running on empty. For this reason, many runners do not eat before a morning event, or, if they do, they eat little and avoid large amounts of high-glycemic-index foods. Athletes often drink coffee or tea in the morning. Besides the general stimulation of the central nervous system, caffeine also inhibits insulin production and stimulates fat mobilization.

In this chapter, and indeed in this entire book, we have explored many topics, most of which are interrelated. In the following Biochemical Connections box, we put many of them together to discuss a topic that should be of interest to everybody, aging.

## Biochemical Connections (CONTINUED)

in mice and rats allow some of the animals' cells to survive in the face of stress that would normally trigger their programmed suicide. It does this by regulating several other key cellular proteins, such as p53 (Chapter 14), NF- $\kappa$ B, HSF-1, FOXO1, 3, and 4, and PGC-1 $\alpha$ . In addition, Sirt1 is stimulated by increased ratios of NAD<sup>+</sup>/NADH, a situation that arises when respiration is increased, as happens with fasting. Thus, Sirt1 is believed to act as both a sensor of nutrient availability and a regulator of the liver's response. Sirt1 has been linked to regulation of insulin and insulin-like growth factor. As we saw in this section, insulin is known to play an important role in the general metabolic state of the organism.

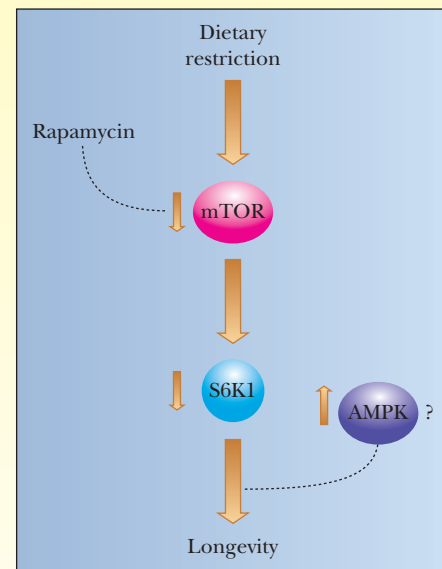
The discovery of the sirtuins and of the effect of calorie restriction and resveratrol (which was much appreciated by wine



(From Saunders, L. R. & Verdin, E. *Stress Response to Aging*. Science, vol. 323, p. 1021. Copyright © 2009 by AAAS. Reprinted with permission of the AAAS.)

drinkers) led to further research into aging and longevity. Several important signaling pathways have now been found to play a role. A drug called rapamycin was found to increase life span in mice. Its direct target is a protein that was given the name **mammalian target of rapamycin (mTOR)**. Both calorie restriction and rapamycin lower the activity of the mTOR enzyme, as shown in the figure to the right.

The mTOR enzyme activates a **ribosomal S6 protein kinase (RSK)**, called **S6K1**, which phosphorylates S6 ribosomal proteins. The RSKs modulate mRNA translation and protein synthesis in response to mTOR signaling. It has been shown that longevity is increased by inhibiting the mTOR enzyme, which in turn inhibits the S6K1 enzyme. Another protein kinase, AMPK, appears to be stimulated by the process.



(From Kaeblerlein, M & Kapahi, P. *Aging is a RSKy Business*. Science, vol. 326, p. 55. Copyright © 2009 by AAAS. Reproduced by permission of the AAAS.)

Although we are decades away from seeing a true longevity pill, the studies referenced here indicate promise that such a compound can be found. As is often the case, it should be much easier to find the treasure when we are sure the treasure exists. Both mTOR and S6K1 can be modified by small molecules, as we have seen in the case of rapamycin. Rapamycin has been shown to reduce adiposity in mice, at least in the short term. Why then have we not seen rapamycin on the shelves at our local pharmacy? The reason is that we have a long way to go before we truly understand this process. For one thing, scientists are concerned about side effects. A known side effect of rapamycin when used long-term is immune suppression. Furthermore, evidence exists that attempts to prolong life also often have the consequence of stimulating cancers. Several studies of cancer cell lines have shown they have significantly greater levels of sirtuins than regular cells. Thus stimulating the longevity of cells doesn't work if we stimulate the wrong kind of cells.

However, the knowledge gained from the studies on sirtuins and mTOR has been the first indication that we may yet be able to take control of our own longevity destiny, albeit sometime in the future.

## SUMMARY

**What are required nutrients?** Required nutrients are substances that we must include in the diet and those that we keep track of. Sometimes this varies by area. For example, in the United States, the only macronutrient we typically monitor is protein, as Americans get plenty of carbohydrates and fats. For micronutrients, we keep track of many vitamins and minerals, all of which are required for optimum health.

**Why do we need vitamins?** Vitamins are small substances required for metabolism that our bodies cannot synthesize. These include the fat-soluble vitamins A, D, and E and the water-soluble C and B vitamins. The B vitamins are precursors to important coenzymes such as NAD, FAD, and TPP.

**What are minerals?** In a nutritional sense, minerals refer to inorganic substances, such as sodium, potassium, chloride, magnesium, phosphorus, and calcium. Minimum daily values for many minerals have been established, and food labels give the quantities of these substances.

**Is the old food pyramid still valid?** Although you can still find the older food pyramid in publications, many scientists recommend the use of the newer pyramid as it better reflects the nature of some of the components. For example, not all fats are bad, and not all carbohydrates are good. The new pyramid distinguishes between types of compounds within a given category.

**What is obesity?** Obesity is a major public-health problem in the United States. The clinical definition of obesity is body weight greater than 20% more than an ideal body weight based on a height-to-weight index.

**What are hormones?** Hormones are chemicals that are produced in one part of the body that affect cells in another part of the body. This takes place through the workings of the endocrine system. Glands produce hormones that circulate in

the blood until they find cells that have the correct receptors. Once the hormone binds to the cell receptor, it effects a change in the metabolism of the cell. Chemically hormones can be peptides, steroids, or amino acid derivatives.

**How do second messengers work?** A second messenger is a molecule that acts as a bridge between the binding of a hormone to a cell receptor and the metabolic effect the hormone has. Common second messengers include cyclic AMP and  $\text{Ca}^{2+}$ . When cAMP is the second messenger, a hormone binds to a receptor and activates a G protein. The G protein stimulates the enzyme adenylate cyclase, which produces cAMP. The cAMP then affects target enzymes that lead to the metabolic effect.

**What hormones control carbohydrate metabolism?** The principal hormones that control carbohydrate metabolism are epinephrine, glucagon, and insulin. Epinephrine promotes the usage of glucose for energy, stimulating catabolic metabolism of carbohydrates in muscle. Glucagon is triggered when the body needs blood sugar. It stimulates gluconeogenesis in the liver.

**What is insulin?** Insulin is a peptide hormone produced in the pancreas. In its active form it is a 51-amino acid peptide with one A chain and one B chain held together by disulfide bonds.

**What does insulin do?** Insulin's primary role is to stimulate cells to take up glucose out of the blood. This is the role that it is best known for, primarily through studying the disease diabetes. A person who does not produce insulin or does not produce enough insulin has type I diabetes. A person who produces insulin but whose cells do not respond correctly to it has type II diabetes.

Insulin also affects many enzymes. In general insulin puts the body into an anabolic state, stimulating the production of glycogen and fat and inhibiting the use of glycogen and breakdown of fat.

## REVIEW EXERCISES

 Interactive versions of these problems are assignable in OWL

### 24.1 Connections between Metabolic Pathways

- Recall** What are the two primary molecules that link anabolic and catabolic reactions?
- Recall** Name some of the key metabolic intermediates that are seen in more than one pathway.
- Recall** Many components of the glycolytic pathway and the citric acid cycle are direct exit or entry points to metabolic pathways of other substances. Indicate another pathway available to the following compounds.
  - Fructose-6-phosphate
  - Oxaloacetate
  - Glucose-6-phosphate
  - Acetyl-CoA
  - Glyceraldehyde-3-phosphate
  - $\alpha$ -Ketoglutarate
  - Dihydroxyacetone phosphate
  - Succinyl-CoA
  - 3-Phosphoglycerate
  - Fumarate
  - Phosphoenolpyruvate
  - Citrate
  - Pyruvate
- Reflect and Apply** People who begin to lose weight often have a rapid weight loss in the first few days. Common knowledge says that this is “just” because of a loss of water from the body. Why might this be true?
- Reflect and Apply** The functioning of a particular pathway often depends not only on control enzymes in that pathway but also on

control enzymes of other pathways. What happens in the following pathways under the indicated conditions? Suggest what other pathway or pathways might be influenced.

- (a) High ATP or NADH concentration and the citric acid cycle.
  - (b) High ATP concentration and glycolysis.
  - (c) High NADPH concentration and the pentose phosphate pathway.
  - (d) High fructose-2,6-bisphosphate concentration and gluconeogenesis.
6. **Reflect and Apply** Why is it somewhat misleading to study biochemical pathways separately?
  7. **Reflect and Apply** To what extent can metabolic pathways be considered reversible? Why?
  8. **Reflect and Apply** In eukaryotic cells, metabolic pathways occur in specific locations, such as the mitochondrion or the cytosol. What sort of transport processes are required as a result?
  9. **Reflect and Apply** Why is it advantageous for a metabolic pathway to have a large number of steps?
  10. **Conjectural Question** If you had your choice of doing research on any topic in this book, which would you choose? Why do you consider that topic to be interesting and important?

## 24.2 Biochemistry and Nutrition

11. **Recall** What is the difference between the old food pyramid and the new version?
12. **Recall** What do we mean when we say that there is no storage form for protein? How is this different from fats and carbohydrates?
13. **Recall** What is the relationship between saturated fatty acids and LDL?
14. **Recall** What is leptin, and how does it work?
15. **Recall** Many have suggested that vitamin D could be more appropriately called a hormone than a vitamin. Is this correct?
16. **Reflect and Apply** Recent recommendations on diet suggest that the sources of calories should be distributed as follows: 50%–55% carbohydrate, 25%–30% fats, and 20% protein. Suggest some reasons for these recommendations.
17. **Biochemical Connections** People who are both alcoholic and exposed to halogen compounds often die of liver failure. Why might this be a logical ultimate result?
18. **Reflect and Apply** It has been suggested that limits be put on the dose in vitamin A supplements sold in stores. What is a possible reason for this limitation?
19. **Recall** In the early 20th century, goiter was relatively common in the Midwest. Why was this so? How has it been eliminated?
20. **Reflect and Apply** A cat named Lucullus is so spoiled that he will eat nothing but freshly opened canned tuna. Another cat, Griselda, is given only dry cat food by her far less indulgent owner. Canned tuna is essentially all protein, whereas dry cat food can be considered 70% carbohydrate and 30% protein. Assuming that these animals have no other sources of food, what can you say about the differences and similarities in their catabolic activities? (The pun is intended.)
21. **Reflect and Apply** Immature rats are fed all the essential amino acids but one. Six hours later they are fed the missing amino acid. The rats fail to grow. Explain this observation.
22. **Reflect and Apply** Kwashiorkor is a protein-deficiency disease that occurs most commonly in small children, who characteristically have thin arms and legs and bloated, distended abdomens due to fluid imbalance. When such children are placed on adequate diets, they tend to lose weight at first. Explain this observation.
23. **Reflect and Apply** Why are amino acids such as arginine and histidine required in relatively large amounts by children but in smaller amounts by adults? The adult human is not able to make these amino acids.
24. **Biochemical Connections** During colonial times, iron-deficiency anemia was almost unknown in America. Why? *Hint:* The answer has nothing to do with what foods they ate.

25. **Reflect and Apply** People on high-fiber diets often have less cancer (especially of the colon) and lower blood-cholesterol levels than people on low-fiber diets, even though fiber is not digestible. Suggest reasons for the benefits of fiber.
26. **Reflect and Apply** Most calcium supplements have calcium carbonate as the main ingredient. Other supplements that have calcium citrate as the main ingredient are advertised as being more easily absorbed. Do you consider this a valid claim? Why?
27. **Biochemical Connections** Alcoholics tend to be malnourished, with thiamine deficiency being a particularly severe problem. Suggest a reason why this is so.
28. **Reflect and Apply** Biologically and nutritionally important trace elements tend to be metals. What is their likely biochemical function?
29. **Reflect and Apply** An athletic friend is preparing to run a marathon and intends to load glycogen before the race. Someone told your friend that the way to load more glycogen is by exercising excessively for two days to deplete the glycogen stores completely, and that is what your friend intends to do. What do you say about this regimen?
30. **Reflect and Apply** Over a period of several decades, an adult human consumes tons of nutrients and more than 20,000 L of water without significant weight gain. How is this possible? Is it an example of chemical equilibrium?

## 24.3 Hormones and Second Messengers

31. **Recall** Are all hormones closely related in their chemical structure?
32. **Recall** How is hormone production affected by damage to the pituitary gland? To the adrenal cortex?
33. **Recall** How do the actions of the hypothalamus and the pituitary gland affect the workings of endocrine glands?
34. **Recall** The hormone thyroxine is given as an oral dose, but insulin needs to be injected into the body. Why?
35. **Recall** What is the difference between a G protein and a receptor tyrosine kinase? Give an example of a hormone that uses each.
36. **Recall** Give three examples of second messengers.
37. **Reflect and Apply** The average male with a computer hooked up to the Internet receives thousands of spam e-mails for all kinds of things, such as Viagra, weight-loss pills, and others not fit to print. Among the spam, one can find offers for human growth hormone pills guaranteed to make one young again. Why is this unlikely?

## 24.4 Hormones and the Control of Metabolism

38. **Recall** List two hormones that work through the cAMP second messenger.
39. **Recall** How does glucagon affect the following enzymes:
  - (a) Glycogen phosphorylase
  - (b) Glycogen synthase
  - (c) Phosphofructokinase I
40. **Recall** How does epinephrine affect the enzymes listed in Question 39?
41. **Recall** What stops a hormone response when a G protein is involved?
42. **Reflect and Apply** When  $\text{PIP}_2$  is hydrolyzed, why does  $\text{IP}_3$  diffuse into the cytosol while DAG remains in the membrane?
43. **Reflect and Apply** Briefly describe the series of events that takes place when cAMP acts as a second messenger.
44. **Reflect and Apply** For each of three hormones discussed in this chapter, give its source and chemical nature; also discuss the mode of action of each hormone.
45. **Reflect and Apply** Is it likely that any metabolic pathway can exist without control mechanisms?
46. **Reflect and Apply** Cholera harms the body by its effect on a second messenger. Describe how this takes place.
47. **Reflect and Apply** The epinephrine-mediated “amplification cascade” of Figure 24.14 has six steps, all of which are catalytic with

one exception. This cascade leads to the activation of glycogen phosphorylase. This enzyme acts in turn on glycogen to yield glucose-1-phosphate (G-1-P).

- (a) Which step is not catalytic?
  - (b) If each catalytic step had a turnover (molecules of substrate acted on per molecule of enzyme) of 10, how many molecules of G-1-P would result from one molecule of epinephrine?
  - (c) What is the biochemical advantage of such a cascade?
48. **Reflect and Apply** How is the amplification cascade of Question 47 reversed?
49. **Biochemical Connections** Explain what insulin and low-carbohydrate diets have to do with one another.

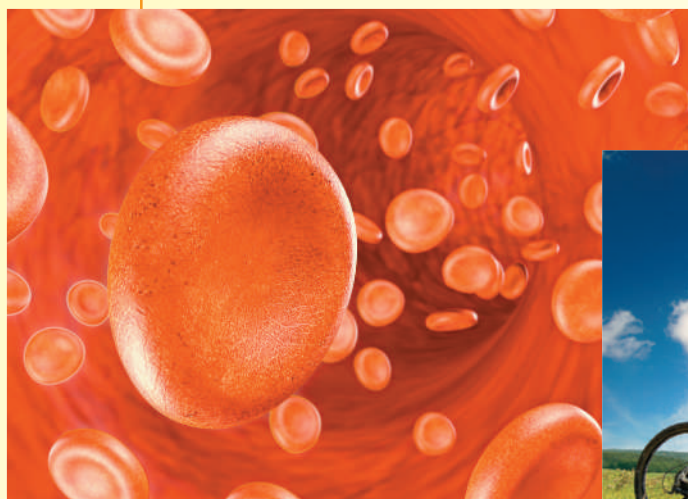
### 24.5 Insulin and Its Effects

50. **Recall** What is the primary function of insulin?
51. **Recall** What is the second messenger for the insulin response?
52. **Recall** What is the link between insulin binding to the receptor and the eventual second messenger?
53. **Recall** What is insulin's effect on the following?
  - (a) Glycogen breakdown
  - (b) Glycogen synthesis
  - (c) Glycolysis
  - (d) Fatty-acid synthesis
  - (e) Fatty-acid storage
54. **Reflect and Apply** How is it possible that both insulin and epinephrine stimulate muscle glycolysis?
55. **Reflect and Apply** Why would a runner who has a 5-km race to run at 9 AM be concerned about insulin?
56. **Biochemical Connections** Why do some people call GLUT4 the training glucose transporter?
57. **Reflect and Apply** How are insulin, GLUT4, obesity, and type II diabetes related?
58. **Biochemical Connections** Explain the relationship of Sirt1 to longevity.
59. **Biochemical Connections** How is rapamycin involved in studies of longevity?
60. **Biochemical Connections** What are the targets of SIRT1?

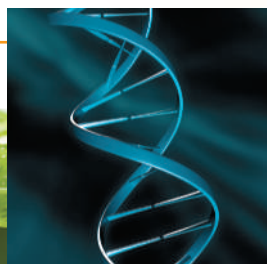
## ANNOTATED BIBLIOGRAPHY

Explore the annotated bibliography for this chapter online at [www.cengage.com/chemistry/campbell](http://www.cengage.com/chemistry/campbell).





# BIOCHEMISTRY



## HOT TOPICS

### Green Fluorescent Protein:

Jellyfish and Green Monkeys **p. HT3**

### DNA and Family Trees:

Who Is a Relative? **p. HT6**

### Diabetes:

An Epidemic for Modern Times **p. HT10**

### Just One Word:

Nanotechnology **p. HT13**

### Small, Smaller, Smallest:

Beyond the Electron Microscope to Single Molecules **p. HT17**

### HPV Vaccines:

Waging the War on Cervical Cancer **p. HT20**

### Stem Cells:

Hope or Hype? **p. HT23**

### Doping in Sports:

Good Science Gone Bad **p. HT28**





# Green Fluorescent Protein: Jellyfish and Green Monkeys

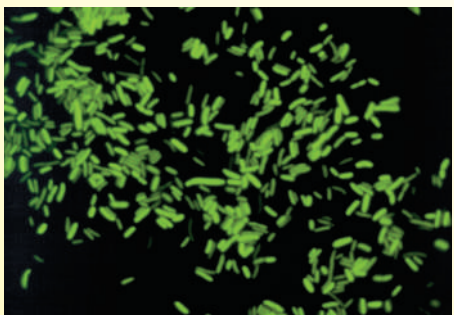


David Woelbel

■ The luminescent jellyfish *Aequorea victoria*.

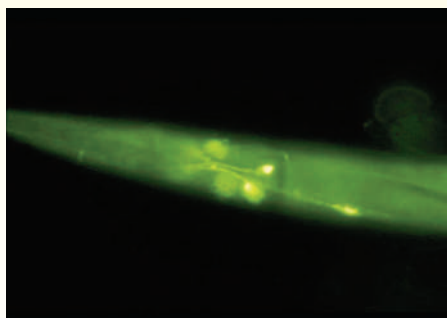
In molecular biology, it is of paramount importance to determine whether or not a gene is expressed. One of the most useful ways of doing so came to be as the result of research on glowing jellyfish. The jellyfish in question, *Aequorea victoria*, emits green light.

Two luminescent proteins were found in this organism. (The terms fluorescence and luminescence both refer to emission of light after absorption of light of shorter wavelength.) One protein, called aequorin, absorbs ultraviolet light and emits blue light, and the other, the green fluorescent protein (GFP) itself, absorbs ultraviolet light and emits green light. Most luminescent proteins, including aequorin, require some other protein or cofactor to emit light. GFP does not



Dennis Kunkel Microscopy, Inc./Phototake

■ Green fluorescent protein in *E. coli*.



Dennis Slade

■ Green fluorescent protein in *C. elegans*.

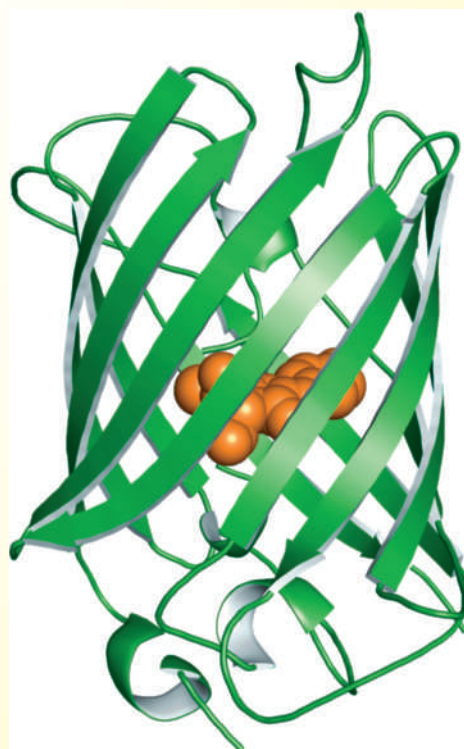
have this requirement. It was clear that, if the gene for this protein could be introduced into other organisms, the protein could be used as a reporter for gene expression. In time, the gene was indeed sequenced and cloned, with a view to introducing it into other organisms. In 1994, the gene was successfully introduced into a prokaryote (*Escherichia coli*) and a eukaryote (*Caenorhabditis elegans*). This breakthrough made it possible to follow gene expression in organisms of all sorts.

Fluorescence is inherently a highly sensitive method, and it has been used as a visualization aid in microscopy for some time. The availability of GFP has revolutionized the method, because it does not require toxic compounds to make samples fluorescent, which was frequently the case beforehand. It becomes possible to monitor the growth of any kind of cell into which the GFP is introduced. It even means that the spread of cancer cells can be tracked with GFP as the tag. By the time the co-discoverers of GFP received the Nobel Prize for chemistry in 2008, more than 12,000 papers about GFP and similar fluorescent proteins had appeared in the scientific literature.

The main structural motif of GFP is that of a  $\beta$ -barrel. (A sculpture based on

this structure is on display at the Friday Harbor Laboratories, where GFP was discovered.) The **chromophore**, shown in orange, is the part of the molecule responsible for the fluorescence. It forms from adjacent serine, tyrosine, and glycine residues in the GFP sequence. This reaction is *autocatalytic*, which means that it takes place without need of enzymatic intervention.

Since the discovery of GFP, many variant forms with different uses have been devised. Mutants have been engineered to have specific properties, such as increased fluorescence. Other variants emit different wavelengths of light, including the blue, red, and yellow regions of the visible spectrum. The variations in the wavelength of emitted light and



■ The structure of green fluorescent protein (see text).

other properties are achieved by varying the amino acid sequence of GFP using standard techniques of site-directed mutagenesis. Another variation depends on emission of infrared radiation. Visible light does not penetrate living tissue deeply, but infrared radiation can penetrate to some extent. A protein that emits infrared radiation was discovered in the bacterium *Deinococcus radiodurans*. When the gene for this protein was cloned into mice, it made it possible to visualize a tumor in a living mouse. In addition, it is possible to use more than one fluorescent protein in a single experiment. Two fluorescent proteins, emitting green and red light, respectively, were used to study fertilization in *Drosophila melanogaster*. The different proteins identified the sperm of two different male flies to see which one actually fertilized a specific egg.

Zebrafish are ideal organisms in which to use GFP to follow development. This fish, *Danio rerio*, is a favorite of developmental biologists for several reasons. Not only do embryos grow outside the mother's body, they are transparent enough to follow all stages of organ growth and to observe how changes in conditions, including removal of cells, affect the growth process. These fish are easily obtained, to the point that they are commonly available in pet shops.



■ Zebrafish are widely used in biological research.

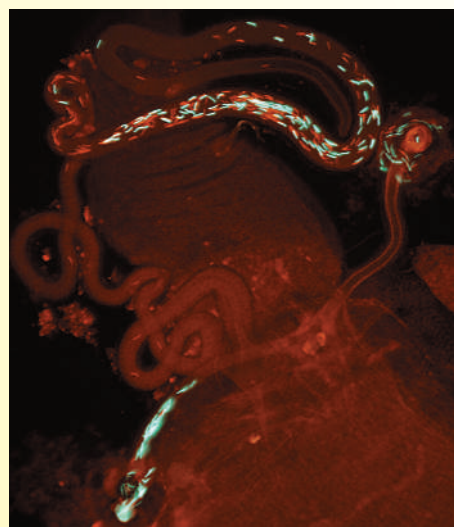
Zebrafish embryos that are transgenic for GFP (fish that have had the GFP gene introduced into their genome) can be observed in real time during their development. This method was used to follow the growth of the heart, blood vessels, and blood. It was a big step forward in the basic understanding of organ growth in vertebrate embryos. Using standard cloning methods, the GFP gene and a suitable promoter were cloned into T cells, which were introduced into the fish. Extending the method still further, a mouse gene for T cell leukemia can be introduced into zebrafish along with the GFP gene. The leukemia cells contained the mouse leukemia gene and the GFP gene under the control of a zebrafish promoter that targets gene expression specifically to lymphoid cells. It is possible to follow the

spread of the cancer from one organ to another by direct observation. It was also possible to screen for drug effectiveness in this experimental system.

In zebrafish hearts, mature heart cells, rather than stem cells, are responsible for repair after injury. The mature cells, and only the mature cells, are labeled with GFP. Images from the stages of the repair process clearly show the role of the mature cells.

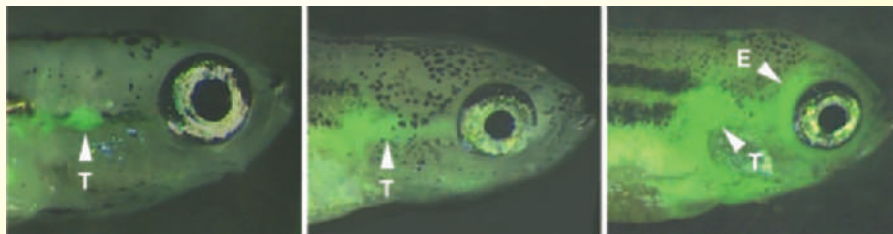
A number of other animals, particularly mammals, have been transformed with the GFP gene. In Asia, researchers at National Taiwan University bred green fluorescent pigs. Images of glowing green mice are easily found on the Internet. Both mice and zebrafish with the GFP gene are sold as pets. The next question is "What about primates?" Apes and monkeys are frequently used as models for human disease because of the close biological relationship, so GFP is an especially valuable reporter here. A number of reports have described the introduction of the GFP gene into primates.

In 2001, researchers at the Oregon Health Sciences University introduced to the world a rhesus monkey named ANDi ("inserted DNA" spelled in reverse). Using a viral vector, the researchers introduced the GFP gene into 224 egg cells, which were then fertilized with monkey sperm. The most promising embryos were implanted into surrogate mothers. Three healthy monkeys were born, but ANDi was the only one who received the GFP gene. Even though ANDi received the gene, he did not produce the protein at detectable levels. It is possible that he was not able to express the gene, even though it was present in his genome. Still, ANDi's existence proved that it is possible in principle to modify primates with a foreign gene.



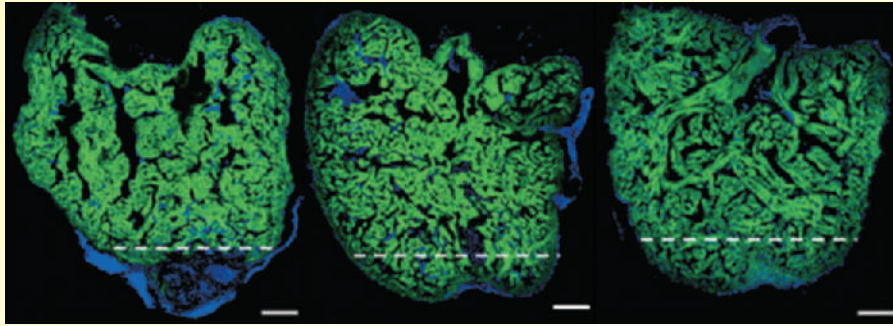
Mollie Menier and Scott Pitnick

■ The reproductive tract of a female fruit fly (*Drosophila melanogaster*), showing relative abundances and locations of first-male (green fluorescent protein) and second-male (red fluorescent protein) sperm.



■ Progression of leukemia in zebrafish after injection of leukemia-containing cells. The successive panels show the spread over a 14-day period starting with the thymic mass (T) and spreading to the eye (E). (From *Myc-Induced T Cell Leukemia in Transgenic Zebrafish* by David M. Langenau, et al., (7 February 2003) *Science* 299 (5608), 887. Reprinted with permission from AAAS.)

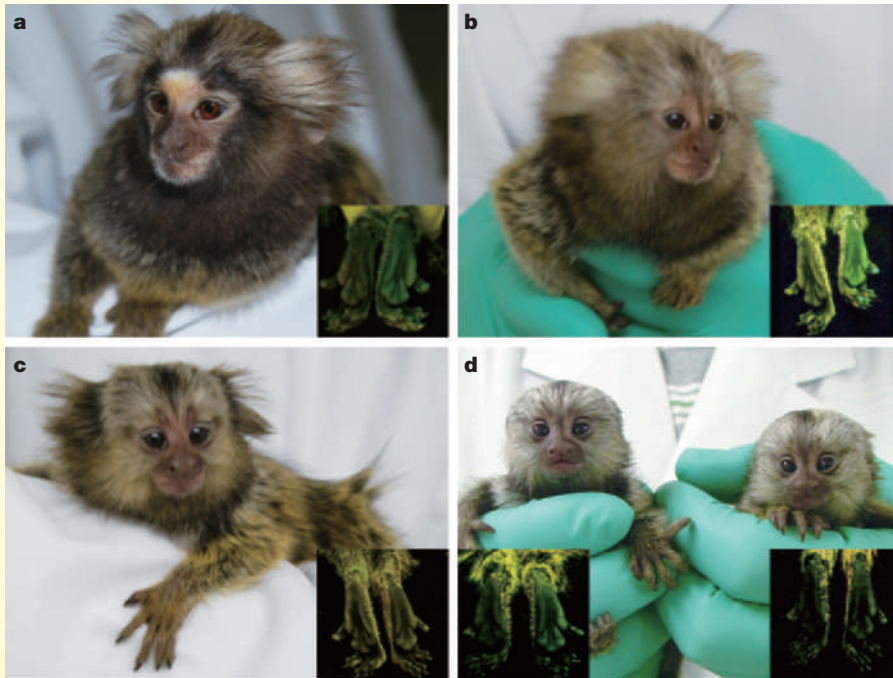




■ Mature muscle cells are responsible for the repair of lost tissue in zebrafish. (Reprinted by permission from *Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation* by Chris Jopling, et al., *Nature* 464, 606–609 [2010].)

In 2009, scientists in Japan succeeded in introducing the GFP gene into marmosets (a species of New World monkey). These animals not only expressed the gene and glowed green under ultraviolet light, they were able to pass the gene to the next generation. Five

second-generation marmosets carried the GFP gene, including the twins Kei and Kou (“keikou” is the Japanese word for “fluorescence”). The next question is the extent to which such animals may be useful models for human disease. What the future will bring remains to be seen.



■ The soles of these marmosets' feet glow on irradiation by ultraviolet light. They received the GFP gene from their transgenic father. (Reprinted by permission from *Generation of transgenic non-human primates with germline transmission* by Erika Sasaki, et al., *Nature* 459, 523–527 [2009].)

## Annotated Bibliography

Chalfie, M., Y. Tu, G. Euskirchen, D. Ward, and D. Prasher. Green Fluorescent Protein as a Marker for Gene Expression. *Science* **263**, 802–805 (1994). [The original article describing the cloning and expression of the GFP gene in a prokaryote and a eukaryote.]

Jopling, C., E. Sleep, M. Raya, M. Marti, E. Raya, and J. Belmonte. Zebrafish Heart Regeneration Occurs by Cardiomyocyte Differentiation and Proliferation. *Nature* **464**, 606–609 (2010). [GFP labeling shows that mature cells are responsible for cardiac regeneration.]

Langenau, D., D. Traver, A. Ferrando, J. Kutok, J. Aster, J. Kanki, S. Lin, E. Prochownik, N. Trede, L. Zon, and A. Look. Myc-Induced T Cell Leukemia in Transgenic Zebrafish. *Science* **299**, 887–890 (2003). [Use of GFP to follow the spread of cancer in zebrafish.]

Langenau, D., A. Ferrando, D. Traver, J. Kutok, J. Hezel, J. Kanki, L. Zon, A. Look, and N. Trede. In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc. Nat. Acad. Sci. (U.S.)* **101**, 7369–7374 (2004). [The role of GFP in tracking T cell development in zebrafish.]

Manier, M., J. Belote, K. Berben, D. Novikov, W. Stuart, and S. Pitnick. Resolving Mechanisms of Comparative Fertilization Success in *Drosophila melanogaster*. *Science* **327**, 18 March 2010 DOI 10.1126/science.1187096. [The use of red and green fluorescent proteins to study fertilization.]

Nogrady, B. Deep into the Red. *Sci. Amer.* **301** (1) 26–27 (2009). [A report about a protein that emits infrared radiation as a visualization tool in living organisms.]

Sasaki, E., et al. Generation of Transgenic Non-Human Primates with Germline Transmission. *Nature* **459**, 523–527 (2009). [Marmosets that have received the GFP gene can pass it to their progeny.]

Service, R. Three Scientists Bask in Prize's Fluorescent Glow. *Science* **322**, 361 (2008). [A single-page news report about the award of the Nobel Prize for the discovery of green fluorescent protein.]

Shu, X., A. Royant, M. Lin, T. Aguilera, V. Lev-Ram, P. Steinbach, and R. Tsien. Mammalian Expression of Infrared Fluorescent Proteins Engineered from a Bacterial Phytochrome. *Science* **324**, 804–807 (2009). [Describes cloning of a fluorescent protein from a bacterial source in mice.]

Thisse, C., and L. Zon. Organogenesis—Heart and Blood Formation from the Zebrafish Point of View. *Science* **295**, 457–462 (2002). [Using GFP to track development in zebrafish embryos.]

# DNA and Family Trees: Who Is a Relative?

It is now common knowledge that individuals can be identified by their DNA, and that this information has wide-reaching forensic applications. Guilt or innocence has been established by DNA evidence, and DNA tests have also been used to establish paternity. Human remains, even ones that are otherwise unrecognizable, can be identified by their DNA.

The identification of the remains of the last tsar of Russia, Nicholas II, and his family has a long, and rather convoluted, story. In 1918, the tsar, the tsarina, their four daughters, and their son were executed by firing squad in Ekaterinburg, Siberia, along with their personal physician and three servants. In 1991, nine skeletons were exhumed from a grave in Siberia near the site of the execution. The initial analysis of the bones from those skeletons indicated that the group consisted of five related persons—a father, a mother, and three daughters—and four unrelated ones. The skeletons of the four unrelated persons were presumably those of the physician and three servants who were executed with the Romanov family. Mitochondrial DNA tests were used to confirm the identity of the family. Prince Philip of England was in a position to contribute a DNA sample, and he did so. Recall that mitochondrial DNA is inherited from the mother only. Prince Philip's maternal grandmother was the tsarina's sister. The match confirmed the identity of the mother and daughters as the tsarina and three of her four daughters. The tsar's identity was established by comparison of mitochondrial DNA from relatives of his mother. Since that time, more information has come to light, strongly confirming the identification.

Another grave was found nearby in 2007, containing 44 badly damaged bone fragments. It was possible to recover enough DNA of both mitochondrial and nuclear origin to establish the fragments as having come from the other two children, the fourth daughter and the son.



■ Remains of the Romanov family were identified by DNA analysis.

The process was helped by the discovery in a museum of a shirt stained with the tsar's blood during a failed assassination attempt. An authentic sample of the tsar's DNA was extracted and used for identification in the paternal line. Even though the remains were in bad condition, the DNA tests made positive identification of the whole family possible. All the conjectures about any survivors of the execution were finally laid to rest.

Tracing inheritance in the male line can be done through the Y chromosome, which passes from father to son. Such inheritance can be traced for many generations. For example, in 2003 genetics researchers tested the Y chromosome of 2123 men living in Central Asia. In 8% of the group the DNA sequence of the Y chromosome was virtually identical. The few differences, arising from mutations, indicate common ancestry about 800 to 1000 years ago. This common ancestor is usually conjectured to have been Genghis Khan (1162–1227). Based on the sample, this result suggests that about 16 million men who live between northeastern China and the Caspian Sea in Central Asia are direct descendants of Genghis Khan in the male line. This finding has captured the popular imagination to the extent that a number of men have



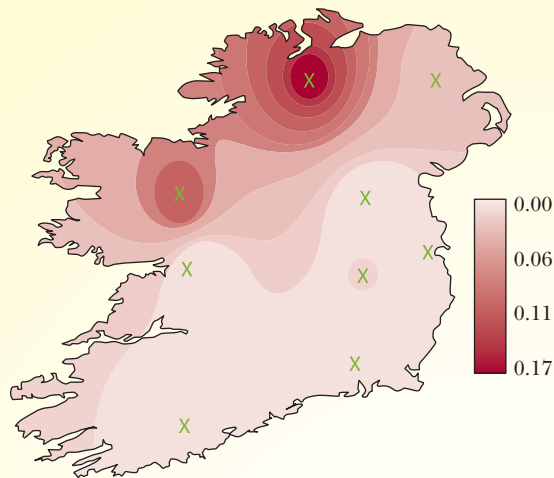
■ A widely distributed Y chromosome is thought to be inherited from Genghis Khan.

asked to be tested to see whether they have this Y chromosome.

A few years later, a group of researchers at Trinity College in Dublin discovered a similar widely distributed Y chromosome in Irish men. In a sample of 796 Irish men, the Y chromosome analysis indicated that roughly a tenth of Irish men, many in northwestern Ireland, had a common male ancestor about 1500 to 1700 years ago. In some parts of northwestern Ireland, the incidence is closer to one in five. This common ancestor was tentatively identified as the chieftain Naill of the Nine Hostages, who established a powerful and long-lasting dynasty about 1500 years ago. This dynasty was called the Uí Néill (descendants of Naill). Linguistic evidence supports this claim. The surname O'Neill remains a common one in Ireland and in many parts of the world among a number of people with Irish ancestry.

Y chromosome markers have an even wider application in their use to track human migration from the African origin of our species. Genetic markers in the Y chromosome (specific mutations) can be tracked, and approximate dates can be assigned to them. The figure shows the pattern of human expansion indicated by Y chromosome analysis. The designations such as M3 and M173 identify a lineage and where it originated. The color coding indicates the date.

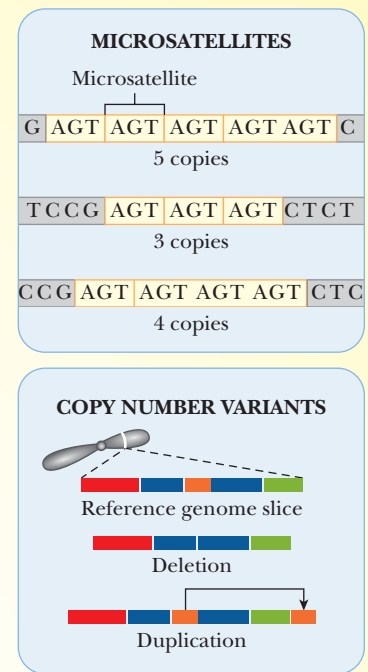




■ The color intensity indicates the distribution of a specific Y chromosome in Ireland. (From Moore, L. T., et al. *A Y-Chromosome Signature of Hegemony in Gaelic Ireland*. American Journal of Human Genetics, vol. 78, p. 5. Copyright © 2005. Reprinted by permission of Elsevier.)

As sequencing techniques improve, it is becoming possible to analyze more of the overall genome, not just mitochondrial DNA and the Y chromosome, to establish relationships among all humans. Two key points emerge in analysis of the whole genome. One is the existence of polymorphisms (differences in individual nucleotides at specific places in the DNA sequence) that distinguish one person from another. The other is the variations in microsatellites, which are short repetitive sequences encountered in all chromosomes. Copy number variants are key to working with microsatellites. The deletions or duplications of the repeated sequences constitute a marker.

The use of copy number variations has revolutionized whole-genome analysis. In 2007, *Science* magazine designated it the Breakthrough of the Year. We have 99.9% of our DNA in common with the other members of the human race. It is the remaining 0.1% that makes us unique individuals. A number of 15 million is a reasonable estimate for the number of places at which the DNA of one person can differ from that of others. When a single base differs from others, the marker is called a single-nucleotide polymorphism (SNP). The availability of millions of markers makes more detailed genetic analysis possible than could ever have been envisioned in the past. It even makes it

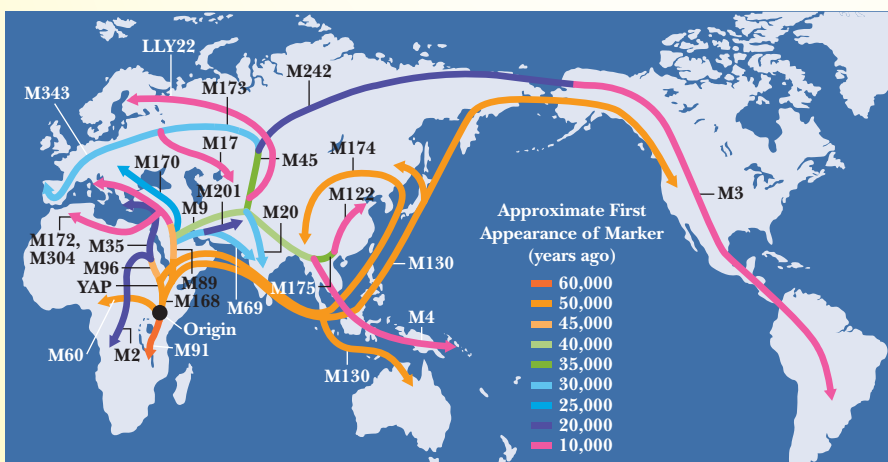


■ Microsatellites and copy number variation. Top: Microsatellites are short repetitive sequences. All chromosomes have microsatellites. Bottom: Copy number variants are markers that arise when repeated sequences are duplicated or deleted. (From Heath, et al. *Traces of a Distant Past*. Scientific American (2008). Copyright © 2008 Scientific American, a division of Nature America, Inc. All rights reserved. Reprinted by permission.)

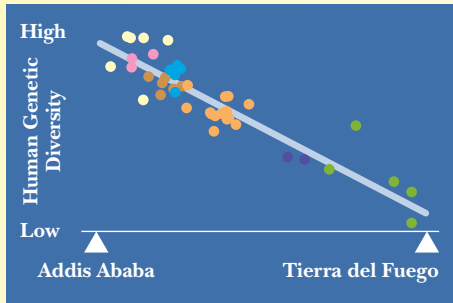
possible to look into relationships among groups of humans.

In 2008 a detailed DNA study appeared, tracing the spread of humans across the globe. This report also included a “family tree” of human populations around the world. The researchers compared haplotypes from a number of individuals around the world. The word “haplotype” refers to a haploid genotype. It is a stretch of one of the two copies of the chromosome (haploid, rather than diploid, which refers to both chromosomes) in which genetic markers are transferred with each other, implying that they are physically close to one another. The results (see figure) indicate that genetic diversity decreases with distance from Africa. This result is consistent with the out-of-Africa model of human development. The original gene pool had a full spectrum of variations, with genetic diversity decreasing as individual populations split off from their roots.

A more detailed analysis allowed construction of a “family tree” for the human race based on geographic distribution.



■ Patterns of human migration as traced by Y chromosome variation. The designations M3, M4, etc. indicate genetic markers and the location of their origin. The color coding indicates the approximate date at which the marker originated. (Reprinted with permission. Copyright © 2008 Scientific American, a division of Nature America, Inc. All rights reserved.)

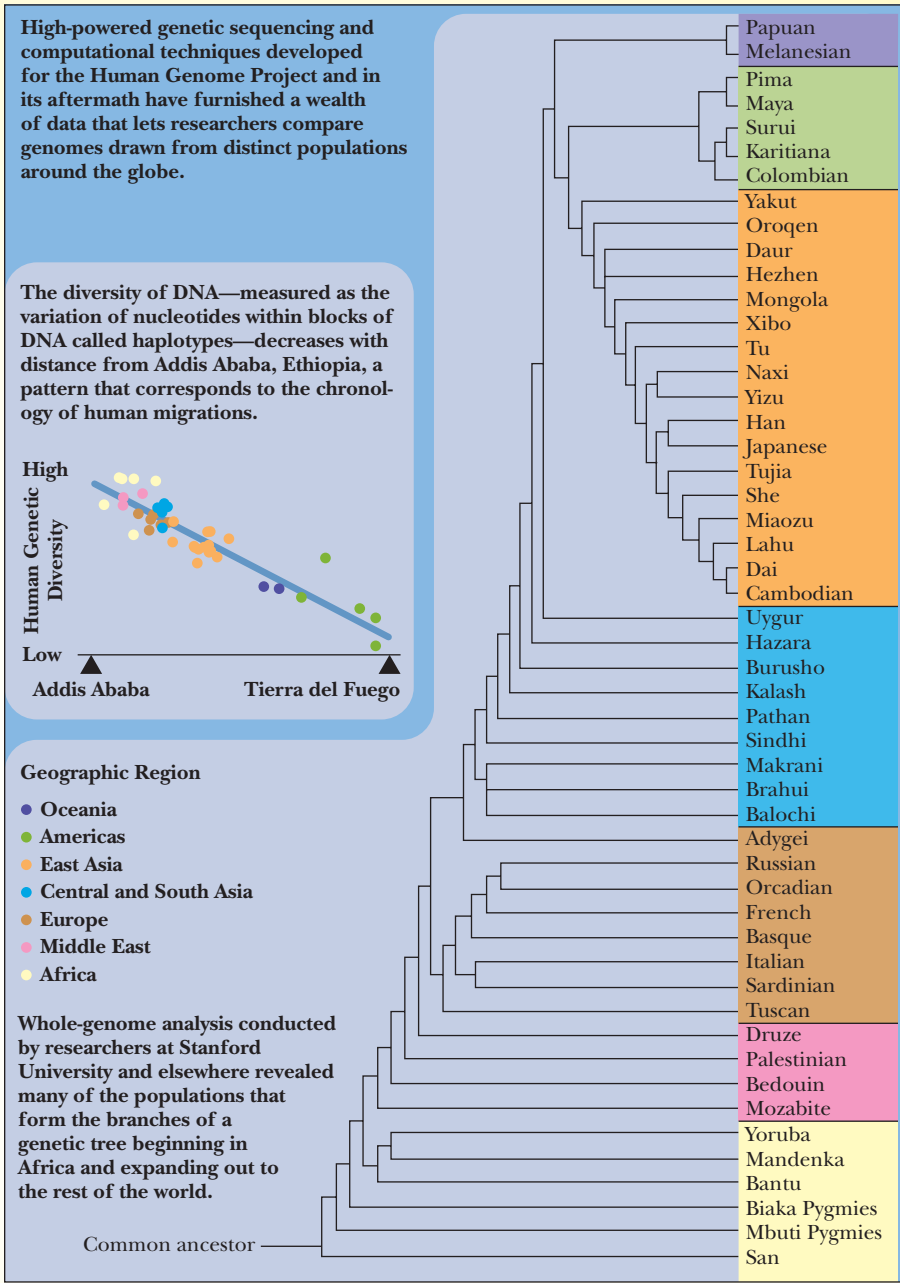


■ Human genetic diversity decreases with distance from African origins. Comparison of haplotypes from around the world shows a greater number of variations in Africa than in other parts of the world. (See next figure for color coding.) (From Li, J. Z., et al. *World Human Relationships Inferred from Genome-Wide Patterns of Variation*. Science, vol. 319, p. 1102. Copyright © 2008. Reprinted by permission of AAAS.)

Related groups, as shown in the figure, are clustered in given regions.

Until very recently, the conventional wisdom was that the spread of *Homo sapiens*, our own species, across the globe took place without significant interbreeding with other hominid species. That view changed drastically in May 2010, when a draft sequence of Neanderthal DNA appeared in print. An international team sequenced DNA from bones of three Neanderthals and compared the results with five individual human genomes from around the world. The five modern genomes were those of a San from southern Africa, a Yoruba from western Africa, a Papua New Guinean, a Han Chinese, and a Frenchman. The results indicated that there were no Neanderthal genes in the present-day African genomes, but that they occurred in 1 to 4 % of the European and Asian genomes. Note that it is possible to obtain these results because of efficient methods for sequencing individual genomes. Neanderthals lived in Europe and some parts of western Asia before they died out about 30,000 years ago. The presence of Neanderthal genes in China and Papua New Guinea, far distant from the places where Neanderthals lived, led the researchers to conclude that the interbreeding took place soon after the first group of humans left Africa 50,000 to 80,000 years ago and then spread through Europe, Asia, and beyond.

The real importance of this new discovery of our DNA is what it tells us about modern humans. Dr. Svante Paabo,

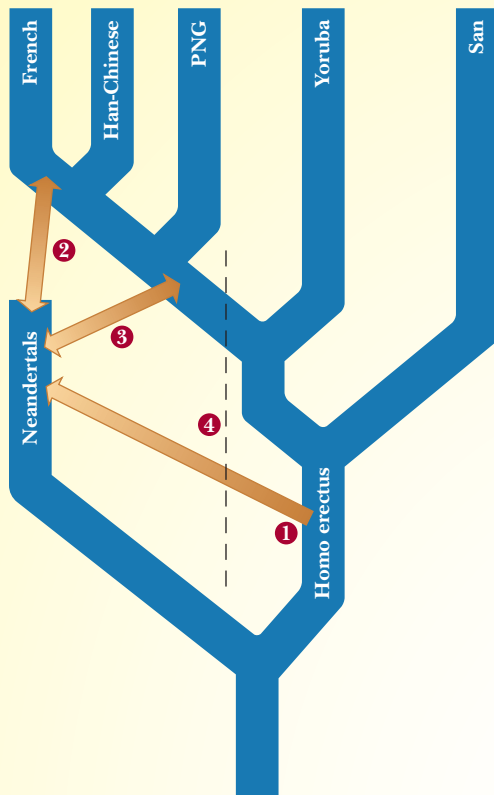


■ The genetic tree of human populations, branching out from African origins. (From Li, J. Z., et al. *World Human Relationships Inferred from Genome-Wide Patterns of Variation*. Science, vol. 319, p. 1102. Copyright © 2008. Reprinted by permission of AAAS.)

the leader of the research team, said the following about Neanderthals: “They are not totally extinct. In some of us they live on, a little bit.” That piece of information tells us something important about ourselves.

The most important point for the human race in all of this study of the DNA family tree is that we are now approaching the point at which we can determine individual genomes relatively easily and relatively cheaply. Ease and

cost-effectiveness are constantly improving. We will soon be able to understand the distribution of diseases in various populations more than ever before. More to the point, we will be able to provide genetically based medicine tailored to the needs of each person to a degree that would have been impossible to envision even a few years ago. It may not be possible for most of us to find a famous ancestor, but we will all be able to profit from the improvements in medicine.



■ A possible modification of the DNA family tree to include the introduction of Neanderthal genes into non-African DNA.



Frank Vrinken, Courtesy of the Max Planck Institute for Evolutionary Anthropology

■ Dr. Svante Paabo with a Neanderthal skull.

## Annotated Bibliography

Gill, P., P. Ivanov, R. Piercy, N. Benson, G. Tully, I. Evett, E. Hagelberg, and G. Sullivan. Identification of the Remains of the Romanov Family by DNA Analysis. *Nature Genetics* **6**, 130–135 (1994). [The identification of the remains of five of the seven members of the family.]

Green, R., 55 other co-authors, and S. Paabo. A Draft Sequence of the Neandertal Genome. *Science* **328**, 710–722 (2010). [An international team reports on the result of its analysis.]

Li, J., et al. Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science* **319**, 1100–1104 (2008). [Whole-genome analysis showing the decrease in genetic diversity with increasing distance from Africa.]

Moore, L., B. McEvoy, E. Cape, K. Simms, and D. Bradley. A Y Chromosome Signature of Hegemony in Celtic Ireland. *Am. J. Hum. Genet.* **78**, 334–338 (2006). [A Y chromosome that may derive from Naill of the Nine Hostages is widely distributed in Ireland.]

Rogaev, E., A. Grigorenko, Y. Moliaka, G. Faskhutdinova, A. Goltsov, A. Lahti, C. Hildebrandt, E. Kittler, and I. Morozova. Genomic Identification in the Historical Case of the Nicholas II Royal Family. *Proc. Nat. Acad. Sci.* **106**, 5258–5263 (2009). [The unequivocal identification of the last two members of the Russian royal family.]

Stix, G. Traces of a Distant Past. *Sci. Amer.* **299** (1) 56–63 (2008). [Using DNA analysis to trace human migration from its origins to the present.]

Zerjal, T., et al. The Genetic Legacy of the Mongols. *Am. J. Hum. Genet.* **72**, 717–721 (2003). [Article describing the discovery of the widespread Y chromosome in Central Asia, likely that of Genghis Khan.]



# Diabetes: An Epidemic for Modern Times

As we have seen throughout this book, maintenance of proper blood sugar levels is critical to mammalian life, especially humans, due to the reliance of the brain on glucose for fuel. There are a number of diseases that affect the metabolism of glucose, but none is as well known, while simultaneously being misunderstood, as diabetes. Diabetes is a disease based on flaws in the pathways involving the hormone insulin. In Type 1 diabetes, the  $\beta$  cells of the pancreas do not make enough or any insulin. For people with this type of diabetes, receiving insulin injections is the only treatment. In the last few decades, however, another type of diabetes, Type 2, has taken the spotlight. This type of diabetes is often referred to as a “lifestyle” disease, as it is intimately related to weight, diet, and exercise. It is this type of diabetes that is becoming an epidemic.

## Insulin and Its Functions

Insulin is a peptide hormone secreted from the pancreas in response to blood glucose levels. It is a primary regulator of carbohydrate, fat, and protein metabolism. It regulates the synthesis of glycogen, the body's storage form for glucose in the muscle and liver. It stimulates the synthesis and storage of fats in fat deposits and the liver. It also inhibits the breakdown of fat. Insulin stimulates the synthesis of proteins. It also serves as a signaling molecule that tells the brain about the availability of fuels. In general, insulin primes the body to store fuels for a rainy day. After a meal, blood glucose rises, which causes a rise in insulin levels, which in turn stimulates storage of fuel and inhibits its oxidation. It also stimulates the muscles to use glucose as a primary fuel source. Another way to think about insulin's role is that it is formed when glucose is plentiful, so it tells the body to use it up in every way it can, whether by burning it as fuel or storing it as glycogen.

HT10

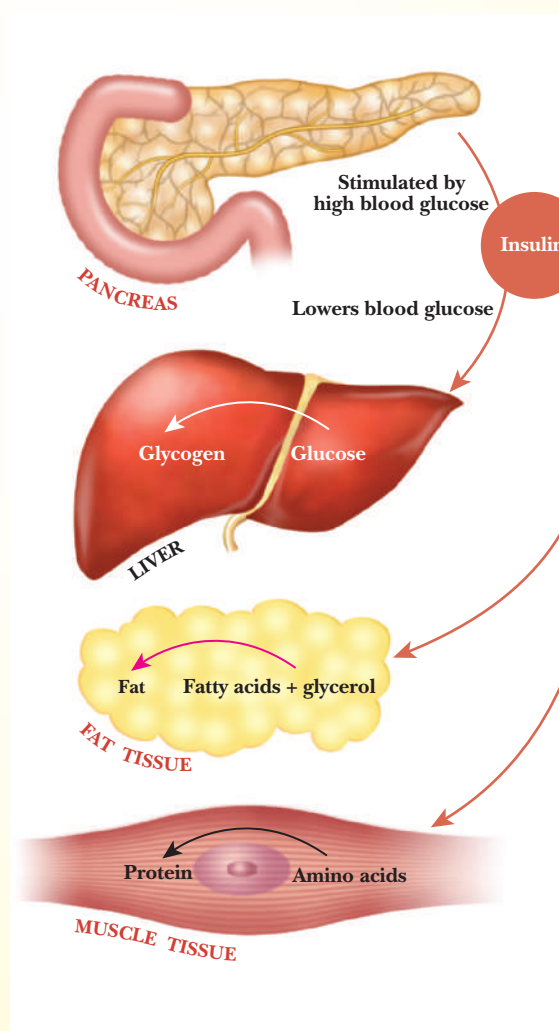
As time since the last meal passes, blood glucose levels drop, followed by a decrease in insulin production. This change helps shift the body away from the metabolism of glucose and toward the metabolism of fat instead, as well as its use by the mitochondria. As such, insulin helps partition the usage of fuels between the main organs as shown in Figure 1.

When the body does not respond properly to insulin, called **insulin resistance**, the delicate balance between fuel

usage and storage breaks down and the symptoms of diabetes become apparent and eventually pathological.

## The Insulin Resistance Epidemic

While virtually unknown a half century ago, insulin resistance is now the common feature of many modern afflictions, such as obesity, heart disease, and Type 2 diabetes. All are linked to diet and other



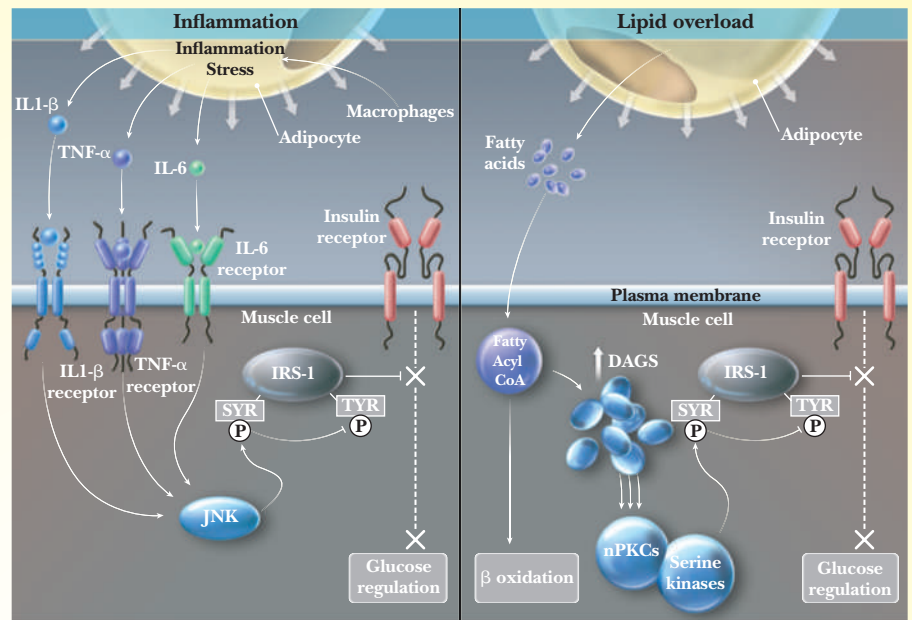
■ **FIGURE 1 Fuel economy.** Among insulin's many functions is as partitioner of metabolic fuels—carbohydrates, fats, and protein—for use and storage in tissues. (From Science, vol. 325, p. 257 (2009). Copyright © 2009 by AAAS. Reprinted by permission of the AAAS.)

lifestyle concerns, and all are linked causally to a diminished response to insulin. Insulin resistance is the primary defect in Type 2 diabetes, which afflicts 6% of adult Americans. This is a 100% increase since 1970. Most people with Type 2 are obese. Obesity is so closely correlated with insulin resistance that most people assume it is the cause. Obesity has increased 150% since the 1970s. Right now 34% of Americans are clinically obese. Another insulin resistance condition is called **metabolic syndrome**, and it involves a number of abnormalities, such as abdominal obesity, hypertension, and high blood sugar. Other diseases, such as asthma, some cancers, stroke, and Alzheimer's disease are also associated with insulin resistance.

Insulin resistance is part of a deadly cycle. A tissue becomes unresponsive to insulin, so the pancreas compensates by producing more insulin. The tissues then respond by becoming even more resistant. It is one of the main goals of diabetes researchers to find and break the causes of this destructive cycle, but studying the problem is not easy, especially in humans. For example, even with healthy individuals, there is a variation of 600–800% in insulin-stimulated glucose uptake, insulin sensitivity, and insulin resistance. Some of this variation is explained by variations in physical fitness. Some is due to weight. There are currently two major theories on the causes of insulin resistance, although some think the truth would be a combination of the two. One of the two theories is based on **lipotoxicity**, the idea that cells get poisoned by fat. The other is based on **inflammation** being the cause.

## Too Much Fat Leads to Insulin Resistance

Since the 1990s, researchers have noted that insulin resistance is found alongside elevated levels of fatty acids in the bloodstream. This has led them to focus on high lipid levels as the putative cause of the problem. Observations showed that muscle cells that contained excess fat inside them could be correlated to being more insulin-resistant. Gerald Shulman of Yale University established that fat accumulation inside muscle cells blocks intracellular events that would normally lead to glucose transport into the cell.



■ **FIGURE 2 Main contenders.** Two explanations for the mechanism of insulin resistance have emerged: inflammation (left) and lipid overload (right). In the inflammation hypothesis, enlarged fat cells attract macrophages and excrete inflammatory signals that work in the muscle cell, via the kinase JNK, to block an insulin receptor substrate (IRS-1) and shut down the insulin-signaling pathway. In the lipid-overload hypothesis, enlarged fat cells leak fatty acids, causing diacylglycerols (DAGs) to accumulate in muscle cells. These inhibit signaling through nPKCs and then block the insulin receptor substrate IRS-1. (From Science, vol. 325, p. 259 (2009). Copyright © 2009 by AAAS. Reprinted by permission of the AAAS.)

He believes the causative agent of this blockage is an increase in diacylglycerols (DAGs), which are intermediates in the production of triacylglycerols, the storage form of fat (see Figure 2).

When the DAGs build up inside liver or muscle cells, they turn off the entire insulin signaling pathway. They inhibit movement of the GLUT4 transporter (Section 24.5), which we saw previously is a glucose transporter that can also be influenced by physical exercise, to the cell membrane. The cell's response to insulin is no longer effective. Through a variety of studies, Shulman showed that in many different tissues, if the level of DAG was increased, the tissue became insulin-resistant. The DAGs are believed to function by modulating novel protein kinase C (nPKC) and serine kinases, which block the insulin receptor substrate, IRS-1.

## Inflammation Is Also Involved

Other researchers believe inflammation is the real culprit in insulin resistance. Gokhan Hotamisligil from Harvard found that the inflammatory cytokine, TNF- $\alpha$ , is

overexpressed in other obese animals. He and other researchers also were able to induce insulin resistance by exposing fat cells to TNF- $\alpha$ . Cells under stress undergo an inflammation response that involves many different inflammatory cytokines (see Figure 2). These cytokines react with cell surface receptors, which then activate another kinase, JNK, which blocks the IRS-1. It is thought that overstuffing cells with fat causes them to attract macrophages and to release the inflammatory cytokines. How the macrophages get recruited is not known, but studies showed that adipose tissue from lean mammals contains about 5% macrophages, whereas in obese animals this number climbs to 50%.

## Causes and Effects

Many researchers believe it is a combination of fat overload and inflammation that causes insulin resistance, but many questions remain. For example, what causes the inflammation? What causes DAGs to accumulate? Nobody is sure. The answer might be something as simple as overeating. If a person eats more than they burn off with daily metabolism, fatty acids



accumulate and overwhelm fat cells. This causes the fat cells to expand and secrete the inflammatory cytokines, which spill out of the cells and accumulate where they should not be. Unfortunately for researchers wanting to have a definitive answer, all signaling pathways are complicated and interrelated. In Chapter 24 we saw how several hormones affect metabolism. For example, epinephrine binds to alpha adrenergic receptors in a variety of pathways. An article in *Science* in January 2010 demonstrated that one of the contributing factors to Type 2 diabetes is an overexpression of Alpha2A-Adrenergic receptors. Epinephrine binding to adrenergic receptors is known to reduce insulin secretion, so having too many receptors can cause the same thing.

## The Brain Connection

While most people think of diabetes and insulin as factors related to the pancreas, liver, muscle, and fat cells, it is also very important to the correct functioning of

the brain. Insulin problems are known to be associated with several neurological disorders, such as Alzheimer's disease, Parkinson's disease, and Huntington's disease. Insulin plays a role in memory and learning, to the point that people that inject insulin are better at remembering things almost immediately. The relationship goes both ways, as has been shown by the fact that learning raises insulin levels in rats. Suzanne de la Monte from Brown University studied the levels of insulin and insulin receptors. She found that insulin levels in the part of the brain associated with learning were four times higher in healthy brains than in Alzheimer's brains. They also had ten times as many receptors. She dubbed Alzheimer's disease "Type 3 Diabetes" due to the connection.

While nobody is sure about the "hows" of the connection, studies have shown that insulin modulates the production and degradation of amyloid beta, the protein that forms the plaques in the brains of Alzheimer's patients. Treatments that

improve insulin response also have been shown to improve cognitive processes in Alzheimer's patients.

The insulin-brain connection is just another reason to pay careful attention to our lifestyles. Eating too much and exercising too little have their own set of physical problems that are well documented. Now evidence indicates that the troubles can extend to neurodegenerative diseases as well. These diseases are difficult to recover from. A person might think they can always lose weight or start exercising more, but if the brain starts to go, it is difficult, if not impossible, to get it back.

## Annotated Bibliography

Taubes, G. Prosperity's Plague, *Science* **325**, 256–260 (2009).

Rosengren, A.H. et. al. Overexpression of Alpha2A-Adrenergic Receptors Contributes to Type 2 Diabetes. *Science* **327**, 217–220 (2010).

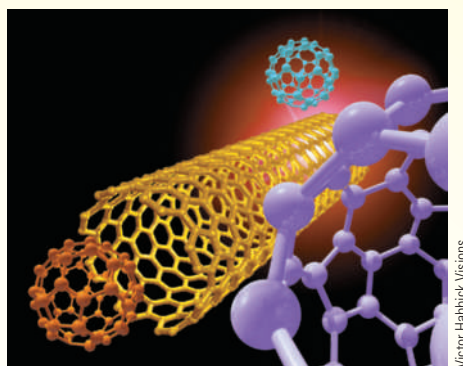
Wenner, M. Your Brain on Diabetes. *Scientific American*, June 2008, 26–28.

# Just One Word: Nanotechnology

In the classic 1967 film *The Graduate*, a young man hears the following career advice. “I want to say one word. Just one word . . . plastics.” If that film were made today, the single word might well be *nanotechnology*. The term covers a wide range of topics, and a number of journals have appeared over the past two decades with the word in their titles. Given the size of this field, we are going to have to concentrate on aspects of it that are related to the material in this book. Most, but not all, of our discussion will focus on the role of DNA in nanotechnology.

First of all, what do we mean when we refer to nanotechnology? The prefix “nano” means “one-billionth.” A nanometer is one billionth ( $1 \times 10^{-9}$ ) of a meter. A nanometer is a convenient yardstick for molecular dimensions. In its most basic terms, nanotechnology is the manipulation of matter on the scale of individual molecules. If an object is smaller than 100 nanometers in one direction, it can be a part of the nanotechnology world. Nanomaterials originally appeared in the public eye when the fullerenes came on the scene. These molecules, discovered in the mid-1980s, consist entirely of carbon atoms. The first to be discovered, buckminsterfullerene, was named in honor of the architect Richard Buckminster Fuller because of its resemblance to his geodesic domes. Other fullerenes were discovered in due course, including carbon nanotubes. They immediately captured the imagination, and their relevance to the field of materials science, especially for applications in electronics, was obvious and readily popularized.

The connection between nanomaterials and molecular self-assembly soon became a topic of importance to researchers in the field. We have already seen numerous examples of self-assembly in this book, such as ribosomes and virus particles. Many examples were discovered, a number of them not resembling biomolecules, but the connection to self-

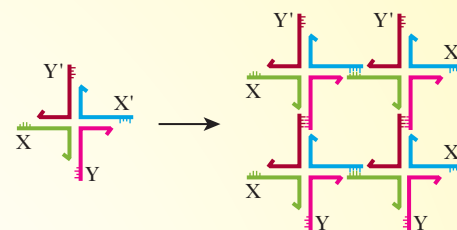


■ Buckminsterfullerene. Carbon nanotubes.

assembly in living systems was always apparent.

The role of DNA in nanotechnology dates to the 1980s. The specificity of base pairing and the action of sticky ends are prime examples of approaches to self-assembly. The applicability of DNA as a tool for nanotechnology arises from these properties. The diameter of the DNA double helix is about 2 nanometers (nm), and the repeat distance for one turn of the helix (its pitch) is about 3.5 nm. Using sticky ends, it is possible to construct long stretches of DNA. These are one-dimensional arrays, and the question immediately arises as to whether we can construct two-dimensional or three-dimensional arrays. The answer is that the properties of DNA allow us to do so readily. Two-dimensional arrays of four-stranded DNA with branch points occur in nature during genetic recombination.

As Figure 1 shows, a branch point forms when four strands of DNA combine. Two of the strands have sticky ends labeled X and X', respectively. The sticky ends of the other two strands are labeled Y and Y', respectively. The sticky ends align, forming a square with more sticky ends on the periphery. These sticky ends on the outside can base-pair with other branched units, eventually giving rise to a two-dimensional array. It is possible to make use of crossing over and exchanges between



■ **FIGURE 1** Sticky ends at branch points can give rise to two-dimensional arrays.

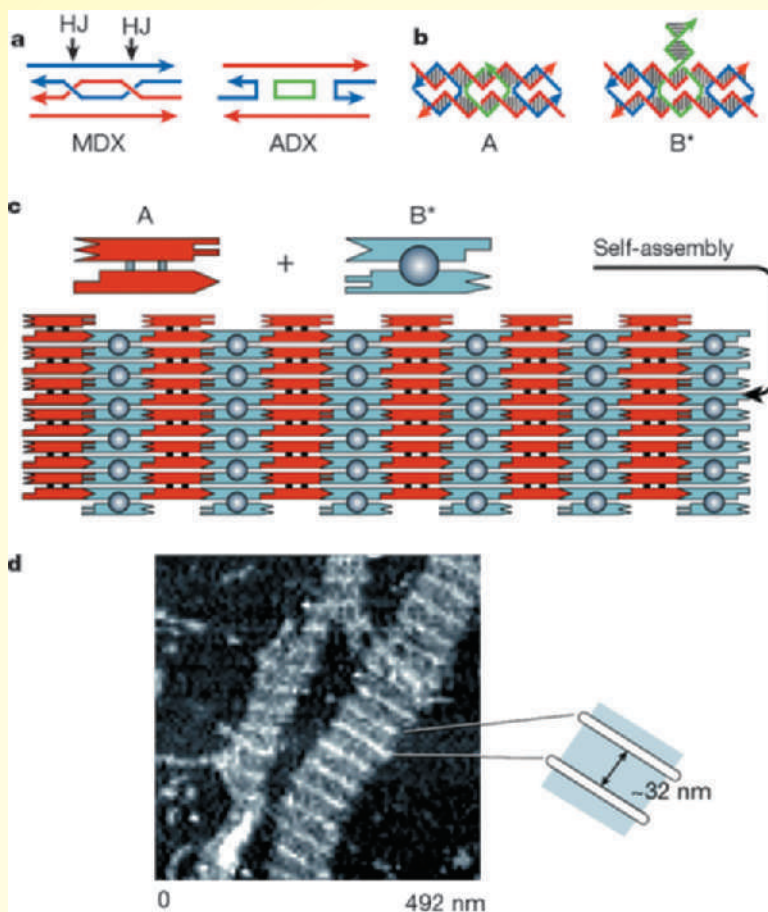
(Reprinted by permission from MacMillan Publishers, Ltd. From Seeman, N. C. DNA in a Material World. *Nature*, vol. 421, pp. 427–431. Copyright © 2003.)

strands to construct two-dimensional motifs that can serve as building blocks for still larger arrays.

Figure 2 shows how two-dimensional DNA arrays can be constructed by using repeated motifs similar to tiles on a floor. In **a**, two kinds of double crossover (DX) units are shown. The one labeled MDX is a meiotic DX recombination intermediate. Each line consists of double-helical DNA. At the points marked HJ (Holliday junctions), DNA strands cross over and exchange genetic information. The other kind of double crossover unit, the analogue unit (ADX), is the tiling unit used to construct arrays. It consists of two red strands, two blue strands, and a central green crossover strand. In **b**, we see the ADX unit shown schematically, along with a variant B\*, which has an extra DNA domain protruding perpendicular to the main strand. In **c**, the A and B\* units and how they fit together are shown. The protrusion on the B\* unit is shown as a circle. The dimensions of the individual tiles are about  $4 \times 16$  nm. In **d**, we see an atomic force micrograph of such an array. The stripes are the B\* protrusions, which are about 32 nm apart.

When these tiles are formed, they can be used as components of DNA-based computers. The array of tiles is programmable in the same way as the components of a computer. A two-dimensional



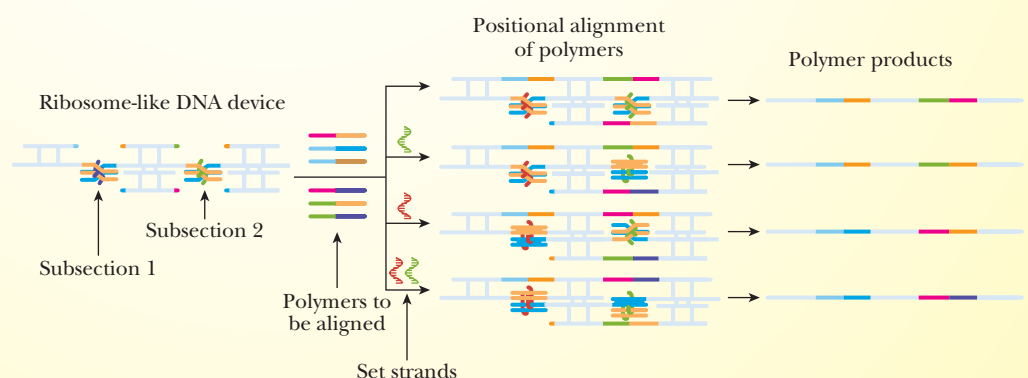


■ **FIGURE 2** Two-dimensional arrays constructed by using repeated motifs like tiles on a floor.  
(Reprinted by permission from *DNA in a Material World* by N. C. Seeman, *Nature* 42, 427–431 [2004].)

template layer can then direct the self-assembly of other layers. Needless to say, these small arrays do not have the power of the silicon-based computers that are so ubiquitous in the world. However, the fact that the process can take place at all is remarkable. It also recalls the theories of the origin of life with layers of ions in clay as the first coding system.

The next step is to use DNA for the construction of nanomachines. The transition between the B and Z forms of DNA is the basis of the first device of this sort. The first nanomachine ever devised used the transition to do rotations. It was not long after that that a DNA machine was devised to do translational motion, namely aligning polymers for subsequent linking in a manner that mimics the translational motion of protein synthesis on a ribosome. The machine has two subsections, each of which has two structural states. Thus, the device can be put into any one of four states by adding

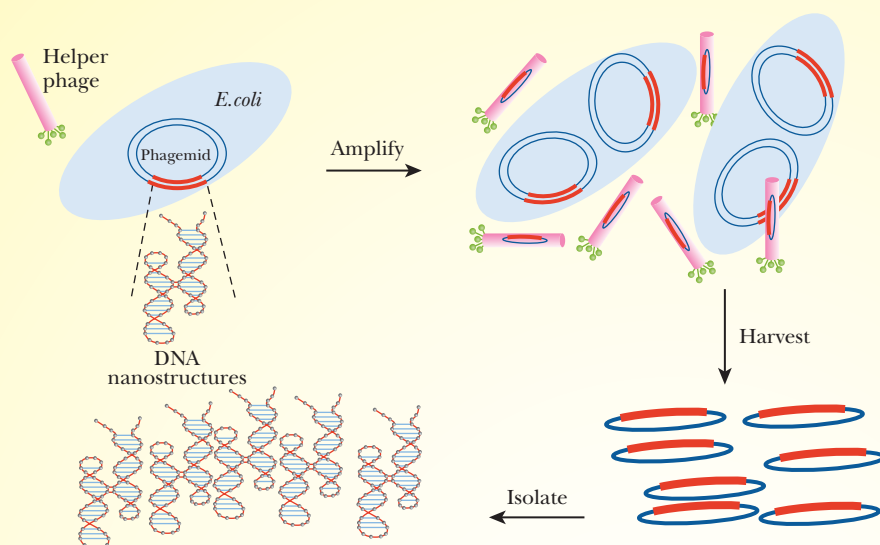
or removing “set strands,” which are themselves DNA strands. In each of the four states, a specific pair of DNA motifs (subsections shown on the left) is selected from a pool. The pairs are linked to the polymers, which in turn can be linked in a specific order.



■ A DNA device to control polymer assembly. One of the four possible states is shown on the left of the drawing. It is the uppermost one in the centre panel. The other three possibilities are shown as well. (From *Science*, vol. 306, p. 2049 (2004). Copyright © 2004 by AAAS. Reprinted by permission of the AAAS.)

It is an expensive and time-consuming process to make these DNA nanostructures. It would make life much easier if we could clone them, and that is precisely what has happened. Researchers used small pieces of phage DNA modified to act as plasmids in *E. coli*. These synthetic cloning vectors are called phagemids. As shown in the figure, the DNA nanostructure is inserted into the phagemid. Amplification takes place in the usual way in cloning experiments, with one change, which is the presence of helper phage. The phagemids are harvested and the DNA nanostructures are harvested from them. So far this technique is applicable only in cases where the DNA nanostructure is topologically simple, but that is bound to change as time goes on.

DNA nanostructures make excellent scaffolds for all sorts of materials because of the way they self-assemble. The classic examples of self-assembly in the natural world, such as ribosomes and virus coats, include proteins as well as nucleic acids. It is reasonable to expect a role for proteins in nanotechnology, and that expectation is definitely fulfilled. Nanoparticles derived from viruses, especially from their protein coats, have a number of advantages. Their size is suitable, and they can self-assemble into symmetrical shapes with a number of binding sites. In addition, they can be produced in large quantities. The protein coats without the genetic material of the virus can be used as nanoscale reaction vessels or can be used to transport therapeutic agents. The cowpea mosaic virus has been used for applications in medicine



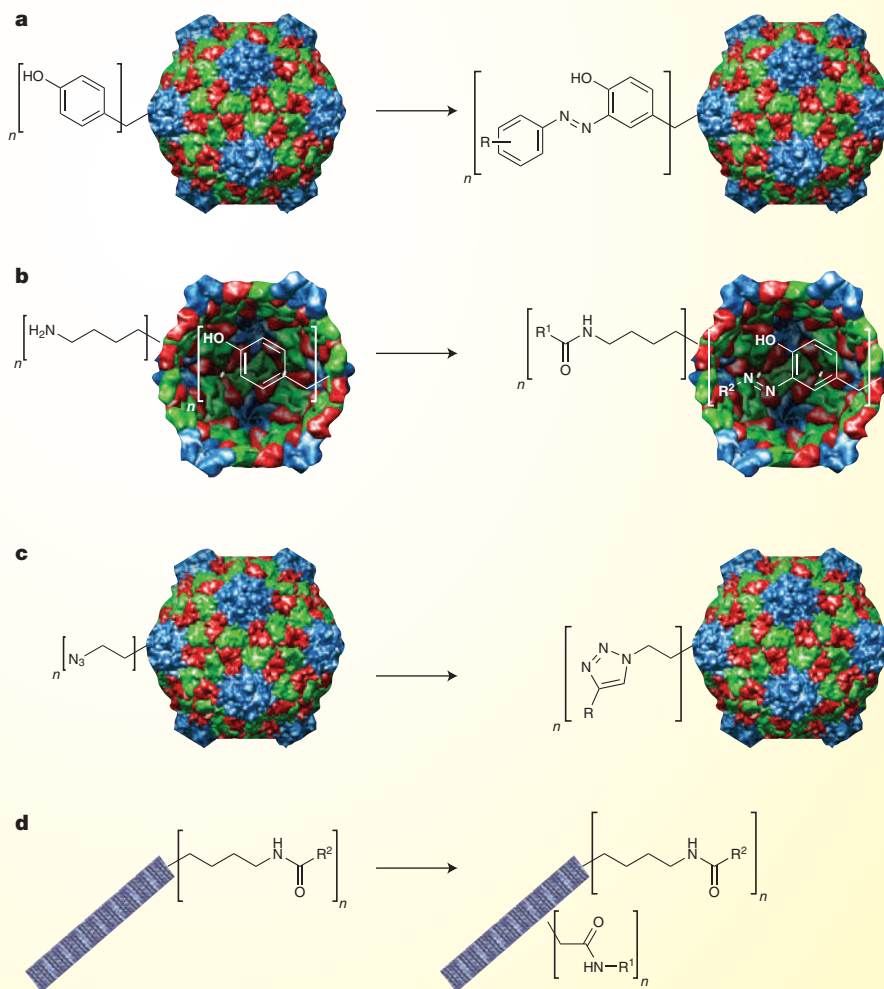
■ Cloning of DNA nanostructures. Phage DNA modified to act as a plasmid in *E. coli* (a phagemid) is the vector. (Reprinted by permission from MacMillan Publishers, Ltd. From Zhang, C. & Mao, C. DNA Nanotechnology: Bacteria as factories. *Nature Nanotechnology* vol. 3 (12), pp. 707–708. Copyright © 2008.)

and electronics. The rod-like shape of the tobacco mosaic virus makes it useful to develop nanowires. Another virus is potentially even more versatile. This virus, *Sulfolobus islandicus* virus 2 (SIRV2), was discovered in an acidic hot spring with a temperature of 88°C and a pH of 2.5. It is stable at high concentrations of ethanol and dimethyl sulfoxide (DMSO), which means that it can be modified under the conditions used in the organic chemistry laboratory. As shown in the figure, the hydroxyl group of tyrosine residues can be modified. The top equation in **a** shows modification of an exterior tyrosine, and the second, in **b**, shows modification of an exterior lysine as well as an interior tyrosine. Unnatural amino acids can be incorporated into the virus coat protein. In **c**, we see how an azide group that can react with an alkyne to give a new functional group. In **d**, rod-shaped viruses can have functional groups added to the ends or the shaft to produce modified nanoparticles, which have the advantage of being stable in organic solvents.

In another use of proteins in nanotechnology, two kinds of proteins can be combined to form a molecular switch. The proteins in question are ion channels and G-protein coupled receptors. When a molecule, such as a possible drug, binds to the receptor portion of the combination, the ion channel portion changes the voltage gradient across cell membranes

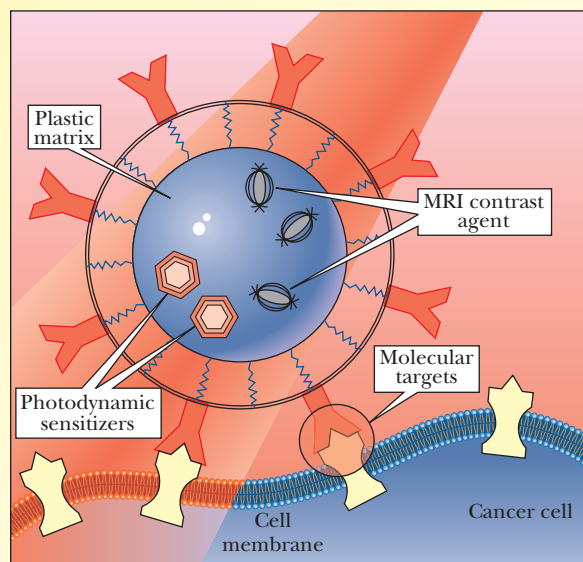
and gives rise to a measurable signal. A number of receptors can be coupled to a single ion channel, making it possible to screen a number of receptor-drug interactions simultaneously, using the kinds of chips we saw in Chapter 13. Other applications are constantly being developed, so that new uses for nucleic acids, proteins, and their derivatives in nanotechnology appear frequently.

So much of medical practice and research has biochemical connections nowadays that it would be profitable for us to look at medical applications of nanotechnology before we leave the subject. They can be based on metal or plastic, but they interact with biological material of necessity. Everything that is known about the nanotechnology of proteins, nucleic acids, or any sort of material is of use here.



■ Proteins in nanotechnology. Modification of virus protein coats can produce nanoparticles that are stable in organic solvents. (Reprinted by permission from MacMillan Publishers, Ltd. From Carrico, I. S. & Kirshenbaum, K. Nanoparticles: Designer labels for virus coats. *Nature Nanotechnology* vol. 4 (1), pp. 14–15. Copyright © 2009.)





■ The multiple uses of nanoparticles in combating cancer. A single particle can target, image, and treat cancer. (From *Science*, vol. 310, p. 1133 (2005). Copyright © 2005 by AAAS. Reprinted by permission of the AAAS.)

Nanoparticles of all sorts are being used in cancer research. The goal is to produce multifunctional agents that can be used both to detect and to destroy cancer cells. As shown in the figure, a single particle can be designed to have antibodies specific for cancer cells on its surface. The interior of the particle can contain imaging agents, such as contrast agents for MRI, and can also contain the toxin to be targeted to the cancer cell. Since the toxin can be specifically targeted to the cancer cell, it will be possible to use higher doses and to avoid the side effects of the usual kinds of chemotherapy.

Many challenges face the application of nanotechnology in medicine.

One of the most important is the task of following the distribution of nanoparticles in the body and imaging them. The substances delivered by nanoparticles have to cross membrane boundaries, and questions of mass transport across these boundaries have to be addressed. It will be highly useful to develop computer models that predict risk and benefit factors for the major classes of nanoparticles and to establish consistent testing standards. A lot of work is necessary to bring nanotechnology from the laboratory to your doctor's office. The idea of introducing "nanobots" (substances derived from nanotechnology) into our bodies to cure our ills is still in the realm

of science fiction. It will be interesting to see what will happen to this idea in the future.

## Annotated Bibliography

Carbone, A., and N. Seeman. Circuits and Programmable Self-Assembling DNA Structures. *Proc. Nat. Acad. Sci.* **99**, 12577–12582 (2002). [The use of DNA nanotechnology to construct programmable devices.]

Carrico, I., and K. Kirshenbaum. Designer Labels for Virus Coats. *Nature Nanotechnology* **4**, 14–15 (2009). [A description of chemical modifications to the coat of a rod-shaped thermophilic virus and possible applications resulting from the modification.]

Sanhai, W., J. Sakamoto, R. Canady, and M. Ferrari. Seven Challenges for Nanomedicine. *Nature Nanotechnology* **3**, 242–244 (2008). [A report on how to bring the results of laboratory research to clinical practice.]

Seeman, N. Biochemistry and Structural DNA Nanotechnology: An Evolving Symbiotic Relationship. *Biochemistry* **42**, 7259–7269 (2003). [A review of the subject by one of the leaders in the field.]

Seeman, N. DNA in a Material World. *Nature* **421**, 427–431 (2003). [A feature article outlining the connections of DNA nanotechnology to other fields.]

Service, R. Nanotechnology Takes Aim at Cancer. *Science* **310**, 1132–1134 (2004). [A review of possible uses for nanotechnology in the detection and treatment of cancer.]

Yan, H. Nucleic Acid Nanotechnology. *Science* **306**, 2048–2049 (2004). [A report on nanomachines based on both RNA and DNA.]

Zhang, C., and C. Mao. Bacteria as Factories. *Nature Nanotechnology* **3**, 707–708 (2008). [A news article about cloning of DNA nanostructures in *E. coli*.]

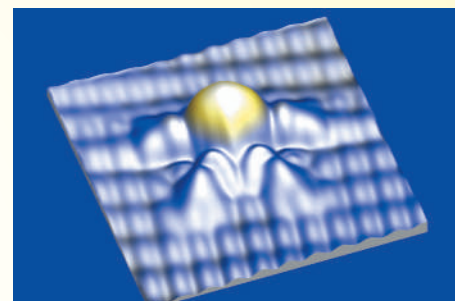
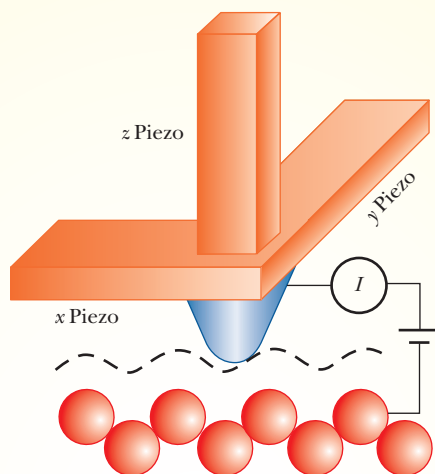


# Small, Smaller, Smallest: Beyond the Electron Microscope to Single Molecules

In the 17th century, the view of the natural world changed drastically as a result of the invention of two pieces of scientific equipment: the telescope and the microscope. Both depend on lenses to focus visible light. The telescope expanded our view first of the solar system and then of the universe as a whole. The microscope opened the world of bacteria and other microorganisms to our view and then allowed us to see cellular structure. Because of the nature of this text, we are going to focus on microscopy. The method has changed greatly since the middle of the 20th century.

The electron microscope is a 20th-century invention. It depends on the discovery, ultimately based on quantum mechanics, that beams of electrons can be focused much like visible light, using electrostatic "lenses." Other, more recent, methods of visualization depend on other physical principles. In the 1980s, two highly sensitive methods, scanning tunneling microscopy (STM) and atomic force microscopy (AFM), were developed. Both are capable of visualizing samples at atomic resolution. Both depend on piezoelectricity, the electric field generated when the material in the microscope probe is subjected to mechanical stress. STM, which was developed before AFM, requires a sample that conducts electricity. AFM does not.

The first figure shows the fundamentals of operation of an STM. The tip, which ends with a single atom, is mounted on a piezoelectric scanner that moves in the x, y, and z directions. A voltage is applied between the tip and the sample, which are not in contact. Electrons do, however, cross the short distance of vacuum between the two by the quantum mechanical tunnel effect, giving rise to the name of the technique. The motion of electrons means that the current flows, and the current can be monitored to give rise to an image of the surface on the atomic level.



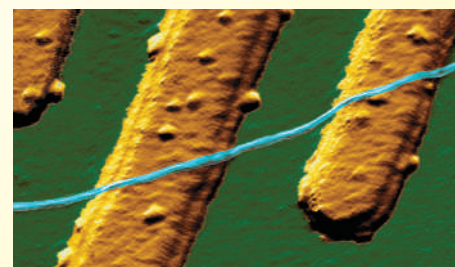
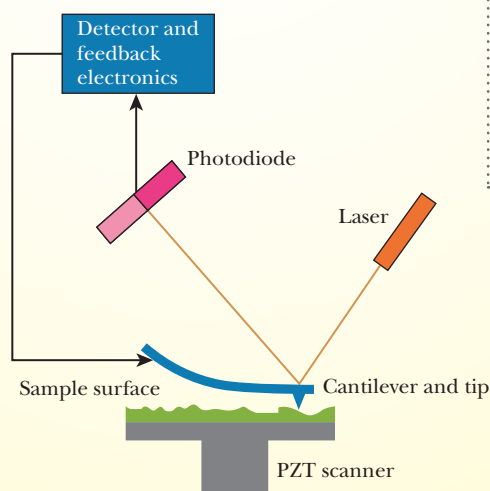
Drs. Ali Yazdani & Daniel J. Hor

**FIGURE 1 Scanning Tunneling Microscopy.** Part A. The mode of operation of an STM. The tip, shown as a rounded cone, is mounted on a piezoelectric scanner. A voltage is applied between the scanner and the sample, causing an electric current to flow. Part B. STM image of a manganese atom (yellow sphere) in a gallium arsenide semiconductor. Single atoms are visible in this image. (From *Scanning tunneling microscopy and atomic force microscopy: Application to biology and technology* by P. K. Hansma, V. B. Elings, O. Marti, and C. E. Bracker (14 October 1988) *Science* 242 (4876), 209. Reprinted with permission from AAAS.)

In the case of AFM, the tip touches the sample, and the sample does not have to be a conducting material. The tip is attached to a cantilever, which moves along the surface. The force exerted by

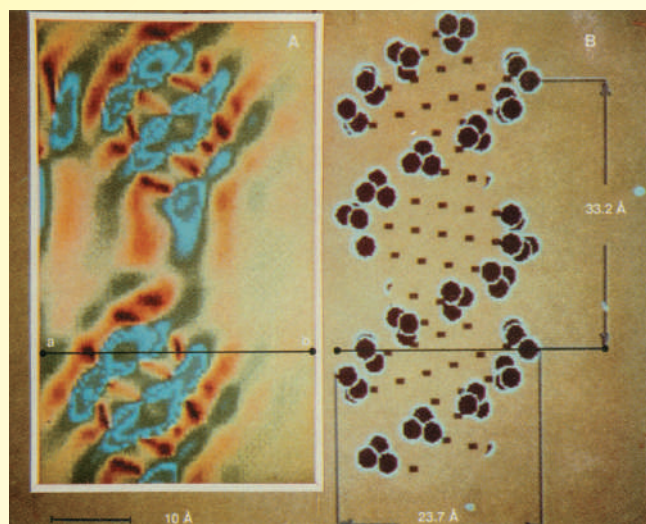
the tip on the sample is measured indirectly. The motion of the tip is tracked by reflection onto a photodiode array of the laser light beam shining onto the moving tip. The output of the array is processed by a detector and feedback electronics. The experimental setup and a typical scan are shown in Figure 2.

A scan of DNA is an obvious goal for STM. In 1989, DNA samples were prepared for scanning by chemical modification with a marker, allowing for visualization with a



Delft University of Technology

**FIGURE 2 Atomic Force Microscopy.** Part A. The experimental setup for AFM. Part B. AFM image of a carbon nanotube (blue line) on platinum electrodes (yellow). The nanotube is 1.5 nanometers (10 atoms) wide.



**FIGURE 3** Application of STM to DNA Structure. Part A The experimentally obtained image of the DNA sample. Part B Comparison with standard model of DNA. (From *Native Escherichia coli OmpF porin surfaces probed by atomic force microscopy* by F. A. Schabert, C. Henn, and A. Engel. (7 April 1995) *Science* 268 (5207), 92. Reprinted with permission from AAAS.)

specially constructed tungsten tip. The results are shown in Figure 3, with a standard model of DNA for comparison. The helical structure of DNA is clearly visible. Note that DNA structure was already well known at the time that this work was done. The results show what the technique can do.

The structure of the membrane protein OmpF (one of the class called porins) was already known from x-ray crystallography when it was the subject of an AFM study. This work was done in 1995, and the quality of the results shows how the technique had advanced since its invention in the 1980s. There was enough resolution to generate a detailed molecular model of the porin in the membrane bilayer.

The development of STM and AFM has not meant the end of techniques that use visible light. Some of the most important improvements have depended on the use of fluorescence. By definition,

fluorescence is the emission of light by a molecule after it has been irradiated with light of shorter wavelength. The shorter-wavelength exciting light is usually in the ultraviolet, with emission of visible light of various colors. The technique is inherently very sensitive. The use of lasers to provide the exciting light to irradiate the sample, combined with sophisticated electronic data-handling systems, helps to increase the sensitivity. With these modifications, it was possible by 1994 to probe individual molecules using confocal fluorescence microscopy. It has since become possible to use fluorescence

of individual molecules to achieve a truly remarkable feat, the sequencing of a single molecule of viral DNA.

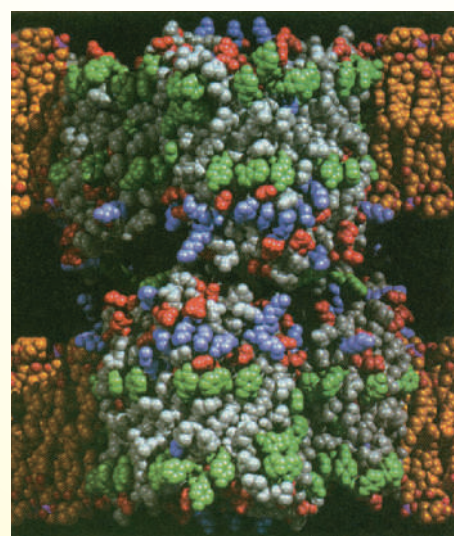
A review of DNA sequencing in general will be useful background for discussion of sequencing of a single molecule. The Sanger method, which we saw in Chapter 13, is the point of departure for newer sequencing methods. The Sanger method is a form of *sequencing by separation*. It uses electrophoresis to isolate the individual sites of interrupted synthesis in a collection of molecules that represent all possible sites at which addition of nucleotides is interrupted. The drawback of the Sanger method is that it requires a lot of time and money. Realizing the full potential of the Human Genome Project requires the individual genome sequences from a number of different people, not just the consensus sequence originally obtained.

In order to increase speed and lessen costs, it is useful to bypass the separation steps, and to sequence millions of small fragments simultaneously. This goal can be achieved by *sequencing by synthesis*. A single stranded DNA template is anchored to a support, after which a primer base-pairs to one end. Nucleotides with fluorescent labels are added, as is DNA polymerase. The correct nucleotide is added to the primer. The excess nucleotides and the polymerase are removed. The identity of the added nucleotide is determined by fluorescence excited by laser light. The fluorescent tag is then removed. The process repeats until the sequence of the whole fragment is determined.

With this information in mind, we can discuss the sequencing of a single molecule of DNA from the virus M13. The researchers used a variation on sequencing by synthesis called a two-pass approach. A primer is covalently attached to the support, and the DNA strand to be sequenced hybridizes to it by base pairing. A high-fidelity polymerase is used to copy the DNA strand with the sequence complementary to the far end of the template (marked “P2 comp” in the figure). Note that the copy is covalently attached to the support. The original template strand is melted off, using hot water. The sequence at the far end is complementary to the primer to be hybridized to the copy strand. Sequencing by synthesis is done, recreating the original template

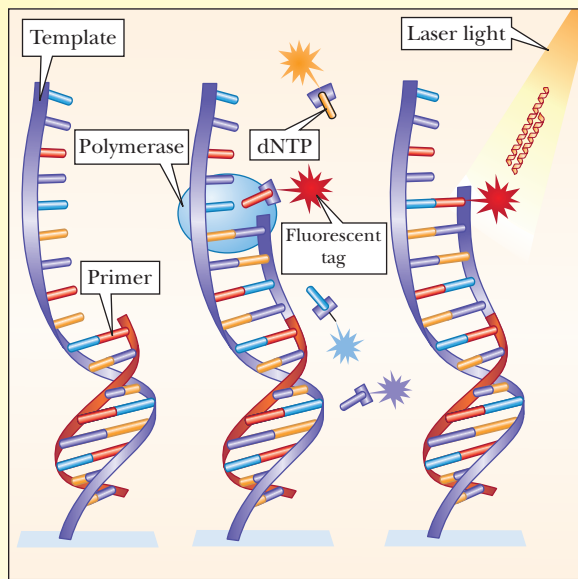


**FIGURE 4A** The AFM scan of the OmpF porin crystal in a membrane model. (From *Native Escherichia coli OmpF porin surfaces probed by atomic force microscopy*. F. A. Schabert, C. Henn, and A. Engel. (7 April 1995) *Science* 268 (5207), 92. Reprinted with permission from AAAS.)

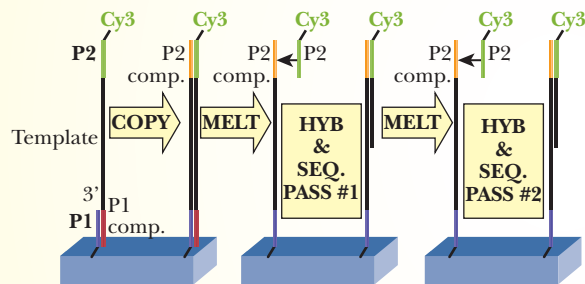


**FIGURE 4B** Molecular model of the OmpF porin in a bilayer. (From *Native Escherichia coli OmpF porin surfaces probed by atomic force microscopy*. F. A. Schabert, C. Henn, and A. Engel. (7 April 1995) *Science* 268 (5207), 92. Reprinted with permission from AAAS.)





■ **FIGURE 5** Sequencing by synthesis. (a) The single-stranded template DNA is anchored to a support, and the primer base-pairs to one end. (b) Nucleotides (dNTPs) that have fluorescent tags are added, along with polymerase. A nucleotide complementary to the template is added to the primer. (c) After removal of polymerase and excess nucleotides, laser excitation identifies the added nucleotide. The process repeats until the whole sequence is determined. (Copyright © 2003 Terese Winslow)



■ **FIGURE 6** The method for two-pass sequencing. A primer (P1) is covalently anchored to a support, and the DNA sample hybridizes with it by base pairing. A copy of the DNA is made, giving rise to a copy of the template anchored to the surface. The template strand is melted off. A primer complementary to the far end of the copied strand is hybridized to it, and sequencing by synthesis is carried out. The process of melting and sequencing is repeated, minimizing errors. (From *Science*, vol. 320, p. 108 (2008). Copyright © 2008 by AAAS. Reprinted by permission of the AAAS.)

strand. The new strand is melted off, and the process is repeated. The point of the repetition is that it drastically reduces errors. The 6407-nucleotide sequence was done in this fashion. The result points the way to sequencing of individual human genomes both quickly and relatively cheaply. Having that information readily available will help achieve the full potential of the Human Genome Project.

## Annotated Bibliography

Cricenti, E., S. Selci, A. Felici, R. Generosi, E. Gori, W. Dzaczenko, and G. Chiarotti. Molecular Structure of DNA by Scanning Tunneling Microscopy. *Science* **245**, 1226–1227 (1989). [An early application of STM to DNA structure.]

Church, G. Genomes for All. *Sci. Amer.* **294** (1) 46–54 (2006). [A review of DNA sequencing methods with emphasis on progress toward obtaining inexpensive and complete genomes for a number of individuals.]

Hansma, P., V. Elings, O. Marti, and C. Bracker. Scanning Tunneling Microscopy and Atomic Force Microscopy: Applications to Biology and Technology. *Science* **242**, 209–216 (1988). [A classic description of the two methods at the time when they were first applied to biological samples.]

Harris, T., et al. Single-Molecule Sequencing of a Viral DNA Genome. *Science* **320**, 106–109 (2008). [The title says it all.]

Nie, S., D. Chu, and R. Zare. Probing Individual Molecules with Confocal Fluorescence Microscopy. *Science* **266**, 1018–1021 (1994). [Real-time observation of fluorescence from single molecules.]

Schaeber, F., C. Henn, and A. Engel. Native *Escherichia coli* OmpF Porin Surfaces Probed by Atomic Force Microscopy. *Science* **286**, 92–94 (1995). [Visualization of a complex protein array by AFM.]

# HPV Vaccines: Waging the War on Cervical Cancer

In the United States, 3,700 women die of cervical cancer every year, and this figure is much higher in the undeveloped world. According to the World Health Organization, the rate in the United States, Canada, Australia, and China is 9.4 per 100,000. In Mexico, the rate is about 33 per 100,000. The highest rates are seen in Colombia, Venezuela, and several African countries at 87 per 100,000, as shown in Figure 1. In 1975, virologist Harald zur Hausen demonstrated that human papillomavirus (HPV) could cause cervical cancer. This finding was confirmed in 1999 when studies found HPV DNA in 99.7% of cervical cancers studied.

HPV is a very prevalent sexually transmitted disease (STD), with estimates indicating that half of adult Americans who are sexually active will become infected at some point in their lives. In one study, 60% of the women studied were infected during a five-year period beginning when they entered the university, as shown in Figure 2. Fortunately, in the United States the number of deaths due to cervical cancer is dropping. In 2002, there were 5,000 deaths, which represents a 75% drop from 1950. The decrease is largely due to

preventive medicine, education, and the standard use of the Pap smear. However, the virus is still a major concern. Scientists have identified more than 100 different types of HPV, only 40 of which infect the genital tract. Some strains cause genital warts. Some cause rectal warts. Some are linked to cervical cancer, and others don't seem to cause any symptoms at all. Of all 100 strains, only about 15 put women into a category of high risk for cervical cancer. In most cases, the immune system is able to defeat the infection and clear it from the system. Two of the HPV types, 16 and 18, occur in more than 70% of the cervical cancers studied. Another 6% contain type 45, and another 4% show type 31, as shown in Figure 3.

In the cervix, HPV infects epithelial cells that lie under the mucosal membrane (see Figure HPV-4). The types most responsible for cervical cancer, HPV 16 and 18, make proteins that bind to two tumor suppressors, one of which is p53, which we saw in Chapter 24. This allows the endothelial cells to divide abnormally. Cancer occurs for reasons still not well understood when the abnormal endothelial cells contact the columnar cells.

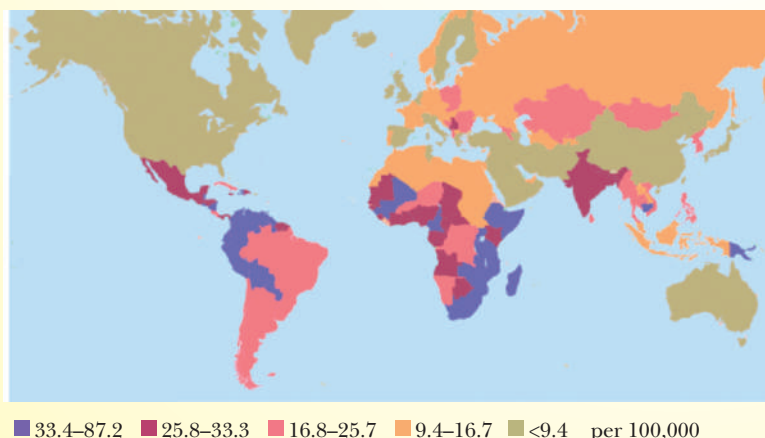
For several years, two pharmaceutical companies have been racing to produce a vaccine against HPV: Merck and Co. (Merck) and Glaxo-Smith-Kline (GSK). The latter has focused on getting approval for use in Europe. Merck has been undergoing efficacy trials in the United States. In June 2006, the U.S. Food and Drug Administration (FDA) approved the use of the Merck vaccine that is effective against HPV 16, 18, 6, and 11. The Merck vaccine is called Gardasil. It is very effective, but it is not cheap, costing \$360 for the complete regimen of immunizations. As the immunizations are much less effective once the infection is established, the recommended target group is young girls before they become sexually active. In the June 29 announcement, the FDA recommended that girls and women ages 12 to 26 should receive the vaccine and that girls age 9 and older could receive it on the advice of a physician. These age recommendations have been the main points of controversy, as will be discussed later.

The vaccines from both companies are based on identical science. A specific protein called L1 makes up the bulk of the viral envelope. By splicing the L1 sequence into a different virus or a yeast particle, scientists could produce a pseudovirus outer particle that had no viral DNA. This protein was then injected into host animals and the antibodies collected (Chapter 24). The antibodies then can attack the true HPV viruses in the body, keeping them from being able to infect as shown in Figure 4.

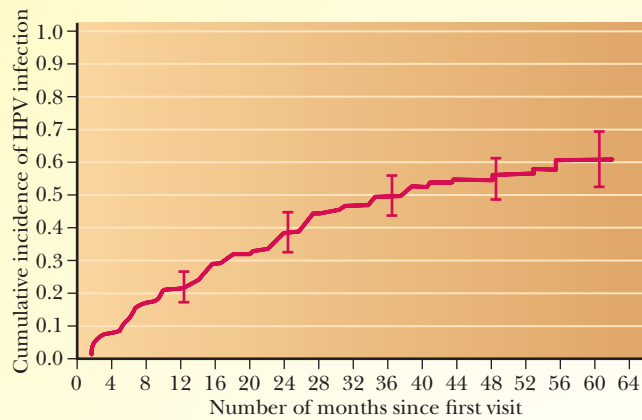
In clinical trials, both Gardasil and the GSK vaccine showed stunning results. The vaccines prevented persistent infection in 100% of the vaccinated women and reduced cervical cell abnormalities by more than 90%. Initial vaccines were based on the HPV 16 and 18 strains. Merck has since produced a quadravalent vaccine that includes two other strains, 6 and 11.

Although the clinical results were surprisingly good, the reality of the vaccine has left as many questions as answers. One important question is, who is going

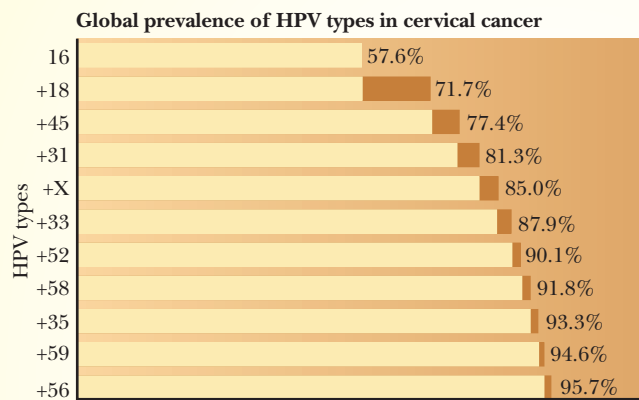
Cervical cancer rates worldwide



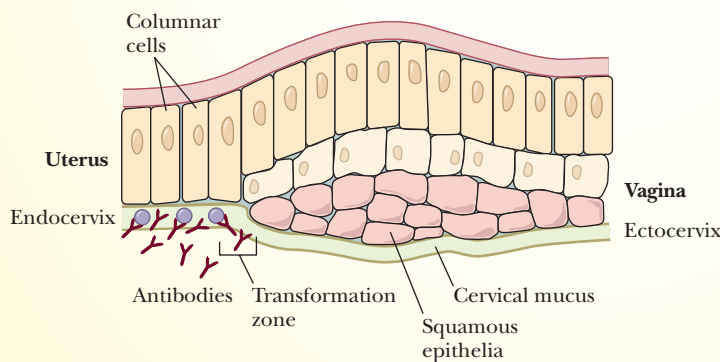
■ **FIGURE 1 Disproportionate impact.** As the Pap smear has become common in wealthy countries, cervical cancer cases and deaths have become increasingly concentrated in the poorer areas of the world. (From *Science*, vol. 308, p. 618 (2005). Copyright © 2005 by AAAS. Reprinted by permission of the AAAS.)



**FIGURE 2 Big virus on campus.** A University of Washington study found that more than 60% of college women studied became infected with HPV over five years. (Adapted from R. L. Winer, et al. American Journal of Epidemiology, vol. 157, p. 218 (2003). Copyright © 2003 Oxford University Press. Reprinted with permission.)



**FIGURE 3 Typical types.** An international ranking of HPV types that put women at high risk of cervical cancer shows that the six most common ones account for nearly 90% of the cases. The Merck and GSK vaccines, now in efficacy trials, both contain HPV 16 and 18, the two most responsible for causing cervical cancer. (From Science, vol. 308, p. 621 (2005). Copyright © 2005 by AAAS. Reprinted by permission of the AAAS.)



**FIGURE 4 Blocking entry.** HPV infects the endothelial cells in the cervix. Antibodies triggered by the vaccine bind the L1 protein and prevent HPV infection. (Adapted from and reproduced with permission from MacMillan Publishers, Ltd. From J. T. Schiller and P. Davies. Nature Reviews Microbiology, vol. 2, p. 343 (April 2004). Copyright © 2004.)

to pay for the vaccine? It is very costly, and the countries that could most benefit from it will be hard pressed to afford it. Although doctors and scientists consider the general well-being of humankind, pharmaceutical companies are

not nonprofit, and they have already invested millions of dollars in production and testing. Therefore, they cannot give the vaccine away. In the United States, it is less clear how important this vaccine can be. Rates of cervical cancer are

already declining because of other medical practices.

Another important question is, who should receive it? Studies indicate that the best target group would be young girls, but this has opened up a lot of controversy. Religious groups oppose the vaccination of young girls on the grounds that it might promote a belief that promiscuity is acceptable. They would prefer that abstinence, rather than a vaccine, be used as a tool against STDs. Some doctors agree that teens and young women should be the first recipients just because there is a ready market for this demographic, while few mothers of 9-year-old girls would currently be thinking about vaccinating their daughters against STDs. Another important consideration will be how long the immunity lasts, information that is still being gathered. It would not make much sense to vaccinate a 9-year-old girl if the protection would end before she became sexually active.

Another question is whether men should receive the vaccine, as they also experience symptoms of HPV infection. Given that genital warts are visible, men might be very motivated to receive such a vaccination. As one of the most ubiquitous of STDs, HPV is unlike any of the others. It can be spread even with condom use and sweeps through a population quickly, all of which should favor the use of a vaccine. Some are considering whether the vaccine should become mandatory. Although this would be very beneficial to the companies that produce it, the question of who would pay for it would then be multiplied, as ultimately the government and thus the taxpayers would foot the bill.

The most important question still remains: how do we get past the international barriers to delivering the vaccine to the less-developed countries that need it the most? Companies find themselves in a moral dilemma in which they know that the poorer countries that need it cannot afford it. This has become a double-edged sword in some cases. Although the companies have patents that protect their rights to produce and sell the product, some countries do not honor the patents, producing the compounds in their own labs. In the case of HIV-fighting drugs, for example, some U.S. companies have made special deals to sell low-cost drugs



to developing nations in return for these countries' honoring the patents and not producing bootleg versions of the drugs.

The conception and production of the vaccines against HPV were a stunning victory for scientists, but as we are seeing, producing the vaccine was the easy part. Now the hard work of blending

international law, public perception, morality, and business finance and accounting has begun.

### Annotated Bibliography

Cohen, J. High Hopes and Dilemmas for a Cervical Cancer Vaccine. *Science* 308, 618–621 (2005)

[An overview of the issues surrounding a cervical cancer vaccine].

Cohen, J. HPV's Peculiarities, from Infection to Disease. *Science* 308, 619 (2005) [A more in-depth article about the subtleties of HPV compared to other viruses].

Kennedy, D. News on Women's Health. *Science* 313, 273 (2006) [A brief synopsis of the status of approval for the HPV vaccine as of 2006].

# Stem Cells: Hope or Hype?

Stem cells are the precursors of all other cell types. They are undifferentiated cells that have the ability to form any cell type as well as to replicate into more stem cells. Stem cells are often called **progenitor cells** because of their ability to differentiate into many cell types. A **pluripotent** stem cell is one that can give rise to all cell types in an embryo or in an adult. Some cells are called **multipotent** because they can differentiate into more than one cell type, but not into all cell types. The further from a zygote a cell is in the course of development, the less the potency of the cell type. The use of stem cells, especially **embryonic stem (ES) cells**, has been an exciting field of research that has really taken off since the late 1990s.

## History of Stem Cell Research

Stem cell research began in the 1970s with studies on teratocarcinoma cells, which are found in testicular cancers. These cells are bizarre blends of differentiated and undifferentiated cells. They were referred to as **embryonal carcinoma (EC) cells**. They were found to be pluripotent, which led to the idea of using them for therapy. However, such research was suspended because the cells had come from tumors, which made their use dangerous, and because they were **aneuploid**, which means they had the wrong number of chromosomes. As we shall see, the possibility of cells becoming cancerous is one of the major hurdles to overcome when we consider using stem cells for tissue therapy.

Early work with embryonic stem (ES) cells came from cells that were grown in culture after being taken from embryos. Researchers found that these stem cells could be grown in culture and maintained for long periods. Most differentiated cells, on the other hand, will not grow for

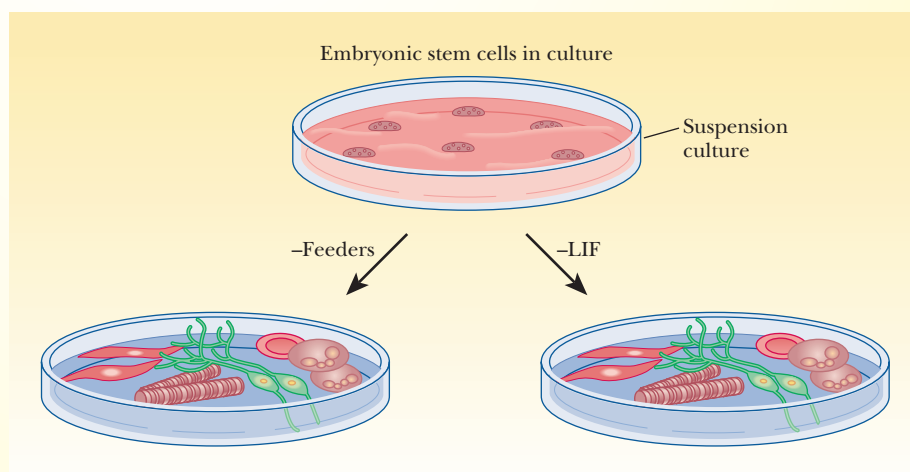
extended periods in culture. Stem cells are maintained in culture by the addition of certain factors, such as leukemia inhibitory factor or feeder cells (nonmitotic cells such as fibroblasts). Once released from these controls, ES cells differentiate into all kinds of cells, as shown in Figure 1.

## Stem Cells Offer Hope

Stem cells placed into a particular tissue, such as blood, differentiate and grow into blood cells. Others placed into brain tissue grow into brain cells. This is a very exciting discovery. Previously it had been believed that there was little hope for patients with spinal cord and other nerve damage, because nerve cells do not normally regenerate. In theory, neurons could be produced to treat neurodegenerative diseases, such as Alzheimer's disease or Parkinson's disease. Muscle cells could be produced to treat muscular dystrophies and heart disease. In one study, mouse stem cells were injected into a mouse heart that had undergone a myocardial infarction. The cells spread from

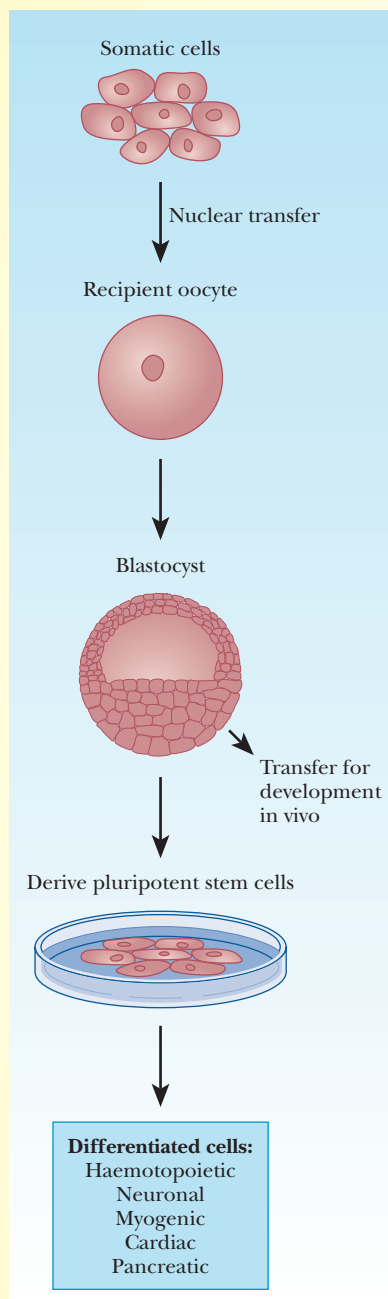
an unaffected region into the infarcted zone and began to grow new heart tissue. Human pluripotent stem cells have been used to regenerate nerve tissue in rats with nerve injuries and have been shown to improve motor and cognitive ability in rats that suffered strokes. (See the articles by Sussman, by Aldhous, and by Donovan cited in the annotated bibliography.) Results such as these have led some scientists to claim that stem cell technology will be the most important advancement since cloning.

Truly pluripotent stem cells have been harvested primarily from embryonic tissue, and these cells show the greatest ability to differentiate into various tissues and to reproduce in cell culture. Stem cells have also been taken from adult tissues, because organisms always contain some stem cells even at the adult stage. These cells are usually multipotent, as they can form several different cell types, but they are not as versatile as embryonic stem (ES) cells. Many scientists believe that for this reason the ES cells represent a better source for tissue therapy



■ **FIGURE 1** Pluripotent embryonic stem cells can be grown in cell culture. They can be maintained in an undifferentiated state by growing them on certain feeder cells, such as fibroblasts, or by using leukemia inhibitory factor (LIF). When removed from the feeder cells or when the LIF is removed, they begin to differentiate in a wide variety of tissue types, which could then be harvested and grown for tissue therapy. (Taken from Donovan, P. J., and Gearhart, J. *Nature*, 414, 92–97 [2001].)

than adult stem cells. The acquisition and use of stem cells can also be related to a technique called **cell reprogramming**, which is a necessary component of whole-mammal cloning, such as the cloning that produced the world's most



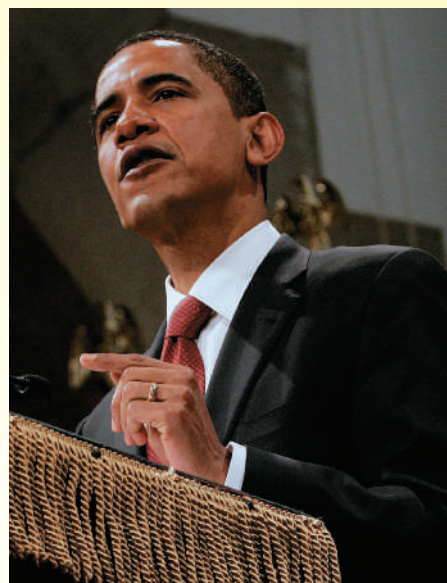
■ **FIGURE 2** Reprogramming a somatic nucleus. When transplanted into an oocyte, a somatic nucleus may respond to the cytoplasmic factors and be reprogrammed back to totipotency. These cytoplasmic factors erase the molecular memory of the somatic cells. Such cells can then be used to harvest pluripotent stem cells or to transfer a blastocyst into a carrier and develop an organism in vivo. (Taken from Surani, M. A. *Nature*, 414, 122–127 [2001].)

famous sheep, Dolly. Most somatic cells in an organism contain the same genes, but the cells develop as different tissues with extremely different patterns of gene expression. A mechanism that alters expression of genes without changing the actual DNA sequence is called an **epigenetic** mechanism. An epigenetic state of the DNA in a cell is a heritable trait that allows a “molecular memory” to exist in the cells. In essence, a liver cell remembers where it came from and continues to divide and to remain a liver cell. These epigenetic states involve methylation of cytosine–guanine dinucleotides and interactions with proteins of chromatin. Mammalian genes have an additional level of epigenetic information called **imprinting**, which allows the DNA to retain a molecular memory of its germ-line origin. The paternal DNA is imprinted differently from the maternal DNA. In normal development, only DNA that came from both parents would be able to combine and to lead to a viable offspring.

Normally, the epigenetic states of somatic cells are locked in such a way that the differentiated tissues remain stable. The key to whole-organism cloning was the ability to erase the epigenetic state and to return to the state of a fertilized egg, which has the potential to produce all cell types. It has been shown that if the nucleus of a somatic cell is injected into a recipient oocyte (see Figure 2), the epigenetic state of the DNA can be reprogrammed, or at least partially reprogrammed. The molecular memory is erased, and the cell begins to behave like a true zygote. This can be used to derive pluripotent stem cells or to transfer a blastocyst into a mother-carrier for growth and development. In November 2001, the first cloned human blastocyst was created in this way, with the aim of growing enough cells to harvest pluripotent stem cells for research.

## The Stem Cell Controversies and Politics

A controversy is currently raging worldwide over the use of embryonic stem cells. The issue is one of ethics and the definition of life. Embryonic stem cells



Reuters/Shannon Stapleton

■ “As president, I will lift the current administration’s ban on federal funding of research on embryonic stem cell lines created after August 9, 2001 through executive order.”

come from many sources, including aborted fetuses, umbilical cords, and embryos from in vitro fertilization clinics. The report about the cloned human embryonic cells added to the controversy. Under the Bush administration, in 2001 the U.S. government banned government funding for stem cell research, but it allowed research to continue on 21 existing embryonic cell lines. The questions driving the controversy include the following: Do a few cells created by therapeutic cloning of your own somatic cells constitute life? If these cells do constitute life, do they have the same rights as a human being conceived naturally? If it were possible, should someone be allowed to grow his or her own therapeutic clone into an adult?

In March of 2009, President Obama announced that he was overturning the Bush administration’s stem cell policy, giving new hope to stem cell researchers, although there is still plenty of red tape regarding which cell lines can be used and the ethics behind them. In December of 2009, the National Institute of Health (NIH) named the first 13 new lines approved for federal funding. Over a hundred new lines are expected to be approved by the end of 2010, and there are hundreds of lines under consideration. In April 2009, the NIH released a draft



of their guidelines, which were seen as a huge improvement over the 21 legal cell lines approved before. Some restrictions are based on when the cell lines were derived and the ethics behind whether donors gave consent. The new cell lines must be derived from surplus embryos donated by couples receiving fertility treatment. Stem cell lines derived from research cloning or somatic cell nuclear transfer will not be eligible. The government also will not fund the development of new cell lines, but will fund research using the lines once developed.

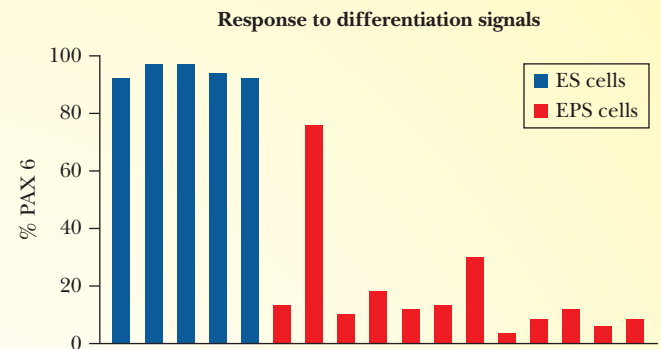
## The Search for Less Controversial Stem Cells

Although the future of stem cell research is much brighter now than it was in 2001, researchers did not stand still in the eight years between. Rather, they began attempting to find ways to create stem cells that are not seen as controversial. Researchers at Columbia University proposed that embryos that had stopped dividing could provide a good source of stem cells. Fertility clinics that do in vitro fertilization have many more failed embryos than successful ones. At the point at which they realize the embryos are not viable and decide not to implant them, they would be able to harvest stem cells from them. Other attempts to generate stem cells involve using other tissues, such as cells from amniotic fluid. These have many of the properties of true stem cells, although they are not truly pluripotent.

Catherine Verfaillie of the University of Minnesota has purified cells from bone marrow called multipotent adult progenitor (MAP) cells. These cells are less versatile than embryonic stem cells, but her research suggested that they can be very useful for therapeutic techniques. In her research she used MAP cells to repopulate the blood cells of mice whose cells had been destroyed by radiation. (See the articles by Holden cited in the bibliography for more on methods for producing these less-controversial stem cells.) Several other labs began looking at other ways to produce cells with pluripotent properties without having to use embryonic tissue.

One of the biggest breakthroughs was made by Japanese researcher Shinya Yamanaka. While many other researchers

■ For more on the initial production of iPS cells, see the articles by Hornyak and by Stadtfeld in the bibliography. (From *Reprogrammed Cells Come Up Short, For Now* by G. Vogel (5 March 2010) *Science* 327 (5970), 1191. Reprinted with permission from AAAS.)



were taking the new embryonic stem cells and working with controlling how to correctly differentiate them into target tissues, he took an opposite approach and began looking for ways of creating stem cells from regular somatic cells. He hypothesized that specific proteins would be found in embryonic cells but not in differentiated cells. He thought that if he could introduce the genes for these proteins into differentiated cells, he might be able to convert them back to a pluripotent state. After four years of research, he had uncovered 24 factors that would transform skin fibroblast cells into pluripotent cells in mice, and found that these cells were almost identical to stem cells. He then found that by introducing four specific genes into the fibroblast cells, he could accomplish the same thing. In 2006 he published his landmark article in which he identified these four critical genes as *Oct3/4*, *Sox2*, *c-Myc*, and *Klf4*. The cells derived from this process are called **induced pluripotent stem cells (iPS cells)**. Yamanaka and other researchers have derived iPS cells from many tissues, including liver, stomach, and brain. These cells show some of the same abilities as true stem cells and have been turned into skin, muscle, cartilage, and nerve cells. In 2007, researchers in the United States extended the technique to create human iPS cells in work recognized as the first runner-up in *Science* magazine's *Breakthrough of the Year*.

There are two concerns with the original research that led to iPS cell production. First was the use of the gene *c-Myc*, which is a powerful cancer gene. In essence, making pluripotent cells can be looked at as being very similar to making cancer cells, and the two have many of the same properties. Further research has shown that in mice, the *c-Myc* can be

avoided and iPS cells can still be produced. The other risk is that the four genes are delivered using retroviruses, which, as we saw in chapter 13, is a procedure that carries its own risks. This second danger was addressed in November 2008, when Matthias Stadtfeld and coworkers generated iPS cells without viral integration by using the common cold adenovirus. The adenovirus was able to produce the required cofactors in the cells without integrating into the host cells' DNA. Their iPS cells showed all of the ability of the iPS cells generated with the retroviruses. This is thought to be a safer alternative.

Scientists know more about how to create iPS cells than they do about why their techniques work, but they continue to search for the reasons. In 2009, Yamanaka, as well as several other researchers, simultaneously published evidence that a critical part of iPS production is the suppression of *p53*, a tumor suppressor gene we saw in Chapter 14. This leads to a tricky situation, however. To make the iPS cells, the *p53* must be suppressed, but for the cells to be stable and useful once created, this transcription factor must be reactivated. The creation of iPS cells was quite a breakthrough, but the new kid on the block has still not convinced many people. The most recent comparisons of iPS cells and traditional ES cells indicates that ES cells are easier to differentiate into target tissues than their iPS counterparts, as shown in Figure 3 below:

## Cutting Out the Middleman

Whether using ES cells or iPS cells, the original research that headed toward tissue therapy involved several changes to cell type. First a cell had to be turned into a pluripotent cell. Then the pluripotent

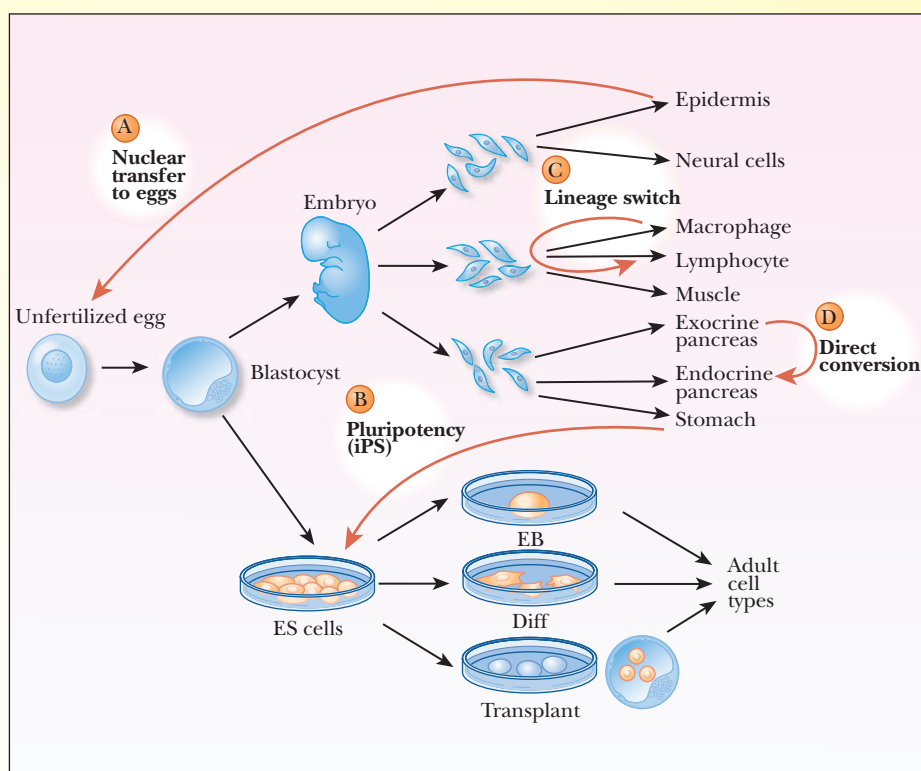
cell had to be converted to the desired tissue type. The newest process involves avoiding this transition and changing one cell type directly into another one. In late 2008, researchers at Harvard University used cell reprogramming techniques to change one type of pancreatic cell in mice into the beta cells that produce insulin. This technique allows the **direct programming** of one cell type into another. After sifting through 1000 transcription factors, the researchers found that by turning on only three genes in the exocrine cells of the mouse pancreas, they could change exocrine cells to insulin-producing beta cells. A similar technique, called **lineage switching**, involves another way to change cell types, in this case by moving the cell back toward a less differentiated cell until a branch point and then re-differentiating it into something else. Such feats have obvious implications for the treatment of diabetes and many other diseases.

## Many Paths to Reprogramming

Researchers interested in studying paths to cell reprogramming have a variety of paths to choose from these days, as shown in Figure 4. Path A shows nuclear transfer to an unfertilized egg, such as the process that led to the cloned sheep, Dolly. Path B shows how a differentiated cell from the stomach can be used to regress to an undifferentiated state in the form of an induced pluripotent cell. Path C shows a lineage switch where a macrophage becomes a lymphocyte. Path D shows the direct conversion of an exocrine cell to an endocrine beta cell.

## Are Stem Cells the New Snake Oil?

No ethical researcher would tell you that we will be healing severed spinal cords in the next year or so, despite the many promising experiments done on animal models. Given the difficulty with repairing nerve tissue, the fact that any improvement can be seen is remarkable, but it will be many years or decades before treatments are readily accessible. There is still a lot to learn about all forms of



■ **Many Paths to Reprogramming.** See text for details. (From *Nuclear Reprogramming in Cells* by J. B. Gurdon and D. A. Melton (19 December 2008) *Science* 322 (5909), 1811. Reprinted with permission from AAAS.)

cell reprogramming. For example, even though researchers were able to transform mouse pancreatic cells from one type to another, approval for human trials would depend on a better understanding

of exactly why the process worked, something that is still a mystery.

Unfortunately, the hope offered by stem cells has spawned many clinics that promise results to desperate clients.



■ **FIGURE 3 Celebrity activists.** Michael J. Fox has been interviewed many times regarding the U.S. policy on stem cell research. He has also campaigned for politicians who support it.



A dozen or more companies produce stem cells for clinical use, and clinics exist in several countries, including Turkey, Azerbaijan, the Dominican Republic, the Netherlands, and China. Advertising campaigns from such companies have attracted many patients willing to spend more than \$20,000 for stem cell therapy, but there is little proof yet that any of the claims of these clinics are valid. Many reputable scientists in the stem cell research world are currently investigating some of the claims and facilities of these companies. (See the article by Enserink cited in the bibliography for a complete description of the state of human stem cell therapy clinics.) The political fight over stem cells continues in the United States. Even celebrities have come forward to endorse stem cell research and politicians that support it. Actor Michael J. Fox, himself a victim of Parkinson's disease, is a most vocal proponent for stem cell research (see Figure 5).

## Annotated Bibliography

- Aldhous, P. Can They Rebuild Us? *Nature* **410**, 622–625 (2001). [An article about how stem cells might lead to therapies to rebuild tissues.]
- Donovan, P. J., and G. Gearhart. The End of the Beginning for Pluripotent Stem Cells. *Nature* **414**, 92–97 (2001). [A review of the status of stem-cell research.]
- Enserink, M. Selling the Stem Cell Dream. *Science* **313**, 160–163 (2006). [An article about stem cell therapy clinics that considers the hype surrounding stem cell therapies.]
- Gurdon, J. B., and Melton, D.A. Nuclear Reprogramming in Cells. *Science* **322**, 1811–1815 (2008)
- Holden, C. Biologists Change One Cell Type Directly into Another. *Science* **321**, 1143 (2008).
- . Controversial Marrow Cells Coming into Their Own? *Science* **315**, 760–761 (2007). [An article about multipotent progenitor cells from bone marrow.]
- . A Fresh Start for Embryonic Stem Cells. *Science* **322**, 1619 (2008) [One of the first announcements about Obama lifting the stem cell ban.]
- . NIH Approves First New Lines; Many More on the Way. *Science* **326**, 1467 (2009). [An

article about new stem cell lines approved under the Obama administration.]

- . Scientists Create Human Stem Cell Line from “Dead” Embryos. *Science* **313**, 1869 (2006). [An article about a possible less-controversial source of stem cells.]
- . Versatile Stem Cells without the Ethical Baggage? *Science* **315**, 170 (2007). [An article about amniotic fluid cells and their use as stem cells substitutes.]
- Holden, C. and Kaiser, J. Draft Stem Cell Guidelines Please Many, Disappoint Some. *Science* **324**, 446 (2009). [An article about reactions to the new stem cell guidelines.]
- Hornyak, T. Turning Back the Cellular Clock. *Scientific American*, 112–115 (December 2008). [A review of the initial research that led to iPS cells.]
- Normile, D. Recipe for Induced Pluripotent Stem Cells Just Got Clearer. *Science* **325**, 803 (2009).
- Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., and Hochedlinger, K. Induced Pluripotent Stem Cells Generated Without Viral Integration. *Science* **322**, 945–949 (2008). [The original work showing that iPS cells can be produced using adenovirus.]
- Sussman, M. Cardiovascular Biology: Hearts and Bones. *Nature* **410**, 640–641 (2001). [An article about how stem cells can be used to regenerate tissue.]

# Doping in Sports: Good Science Gone Bad

Since sports were invented, humankind has always sought a way to gain a competitive advantage, and there are many ways, both ethical and unethical, to do so. Ethical ways of improving performance include superior training, improved diet, nutritional supplementation, better rest, physical therapy, massage therapy, stretching, and a host of sports psychology techniques. Unfortunately, since humankind has roamed the planet, there have always been those willing to cheat to get a competitive advantage, although sometimes the line between cheating and fair play is rather slim.

The last decade has been marked by many sports scandals involving doping. By definition, *doping* is the use of any prohibited compound or method in order to gain a competitive advantage. The type of compound chosen depends on the type of activity. Athletes in power sports use different methods from those in endurance sports. The International Olympic Committee has a comprehensive list of all agents and practices that are considered “illegal” from a sporting point of view for particular sports. It should be noted that “unethical” and “illegal” are not always the same thing. You and I can take a cold medicine that contains ephedrine in it if we have a cold. That is neither illegal nor unethical. A professional cyclist cannot take the same cold medicine, as ephedrine is on the banned list. It would not be illegal, but it would be considered an infraction of the rules, and the rider would be penalized. In the remainder of this article, we will look at some of the more common doping practices.

Doping in sports came under intense scrutiny beginning on July 13, 1967, when professional cyclist Tom Simpson died during the Tour de France. Simpson, like many riders of the time, was using artificial stimulants, such as amphetamines, to allow him to ride longer, higher, and faster. Unfortunately, amphetamines allowed humans to do something that horses



■ British cyclist Tom Simpson died on the slopes of Mt. Ventoux during the 1967 Tour de France. His sacrifice brought the use of drugs in sports to light and was the beginning of the modern era of antidoping controls.

can do—run themselves to death. Natural controls keep humans from doing so, but under the influence of amphetamines, humans can create an oxygen debt that they cannot repay. This is what happened to Simpson on the slopes of the brutal volcano Mt. Ventoux on that tragic day. He collapsed and died a few minutes later despite being given oxygen. In response to the Simpson tragedy, the worldwide governing body of cycling, the Union Cycliste Internationale (UCI), began formal drug testing at events. Cycling has always led the way in the fight against doping. Unfortunately, this has had the unintended effect of casting a large shadow over the sport. Cycling does 10 times the number of tests that other sports do. For this reason it also discovers a larger number of positive cases of drug use. This has reached epidemic proportions in the last few years, especially with cycling’s signature event, the Tour de France.

Throughout the late 1960s and into the present, the use of amphetamines waned as improved testing procedures and education into the dangers caused athletes and coaches to abandon their use. The 1970s and 1980s marked the era of the use of anabolic steroids. These were principally seen in power sports, such as power lifting and sprinting. The use of

steroids was nearly accepted in several sports, such as professional baseball and football, until recently. One of the most infamous cases of steroid use led to the disqualification of Canadian sprinter Ben Johnson after the 1988 Olympic Games. More recently, several professional baseball players have lived in public scrutiny over their alleged use of anabolic steroids. In the summer of 2007, one of these athletes, Barry Bonds, broke the all-time home run record previously held by Hank Aaron since 1976. Fans of Hank Aaron, and indeed of previous record holder Babe Ruth, believe that Bonds’s new record is meaningless by comparison.

Although power athletes were tempted by steroids in the 1980s for the purpose of building muscle strength and power, endurance athletes were experimenting with two drugs. One was also a steroid, testosterone or its precursors. In the case of an endurance athlete, testosterone is used to help recovery rather than build muscle. Endurance sports require long hours of training, and the ability to recover quickly between workouts is critical



■ Canadian sprinter Ben Johnson was disqualified for steroid use after winning the 1988 Olympic Games 100-meter dash.



to success. Extra testosterone helps such athletes be ready to train hard again the next day. Therefore, use of this drug allows superior training for those that take it. As testosterone is a natural hormone, detecting its use is difficult.

At the same time, endurance athletes were trying to find a way to get more oxygen-carrying capacity out of their cardiovascular systems. Training at altitude had been a favorite technique, and athletes who trained at altitude would have more red blood cells than those who trained at sea level. The downside was that a human body cannot train as effectively at altitude. The perfect combination would be to train at sea level but still get the benefit of high-altitude training. Thus, the early to mid-1980s marked the beginning of the blood-doping era. *Blood doping* involves removing some of the athlete's blood and storing the red blood cells. The athlete then has time to regenerate the lost blood volume and red blood cells. Then, just before the competition, the athlete has his or her own stored red blood cells reinjected. In this way the athlete gets to have extra blood cells without having to suffer the poorer training at altitude. This practice is not without risks, however, as mistakes that lead to blood too thick in red blood cells can have fatal consequences. Athletes have died during competitions from heart failures caused by such thick blood.

In the 1984 Olympic Games in Los Angeles, the technique of blood doping was



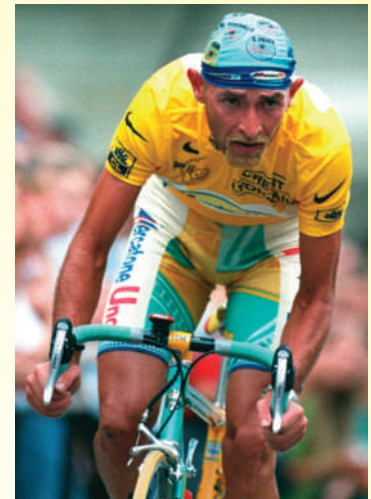
© Lucy Nicholson/Reuters/Corbis

■ In 2007, Barry Bonds broke the 30-year-old home run record previously held by Hank Aaron. The last several years of Bonds's remarkable career were clouded by constant allegations of steroid use.

in full swing, but the practice was not officially banned by the IOC at this time. It was also impossible to detect because athletes had received their own blood, and there was no unnatural substance to detect. However, it was also being realized that blood doping did carry a significant downside. In most sports in which athletes benefit from practice, they train hard all year. There is never a time when they can afford the poor training accomplished when they are trying to replenish their blood supply. Also, the blood cells do not last long in storage, so the blood must be drawn during the peak competition season. This makes the practice of questionable use these days for sports in which the athlete must be competitive for months at a time. Shortly after the 1984 Games, the IOC and UCI did ban the practice, although they recognized the near impossibility of detecting it. In the late 1980s, another technique was developed to make the removal of blood unnecessary—the use of recombinant erythropoietin (EPO) (Chapter 12). EPO is a natural hormone that stimulates production of red blood cells. It was originally purified for medical use to help surgery patients and kidney dialysis patients replenish lost blood cells. By using EPO, an endurance athlete would have all the benefits of blood cell enrichment without having to train at high altitude or remove blood cells.

In the late 1980s and early 1990s, EPO use was rampant but also had tragic consequences. More than 20 athletes died while using EPO in a four-year period, most likely because they took too much of the hormone and reached a blood cell level that was too high for safety. Blood cell concentration is measured by a simple technique called a *hematocrit*, which compares the red blood cell volume to plasma volume. In the 1990s, cyclists using EPO had hematocrit levels in the 60s. Normal levels would be in the high 30s to mid-40s by comparison. Unfortunately, there was still no reliable test for EPO use because the difference between the natural compound and the recombinant version is very subtle.

The EPO era blew up in 1998 when the entire Festina professional cycling team was ejected from the Tour de France when their team doctor was caught



© Fabrice Coffrini/AP

■ Italian cycling star Marco Pantani was a former winner of the Tour de France. He was leading the Giro d'Italia in 1999 when he was expelled for having a hematocrit that was too high. His career and life in ruins, he committed suicide in 2006.

crossing the border into France with hundreds of vials of EPO and steroids. Professional cycling, as well as other endurance sports, was truly in a crisis. At the end of the season, the professional teams, riders, coaches, and the UCI met to discuss what had to be done. The athletes' position was that nobody wanted to take drugs for their sport, but that they were professionals, and because the others were doing it, they had to do it in order to survive and feed their families. The top teams agreed that they would all stop if everyone else would. Conceding the human nature to cheat, they all agreed to undergo frequent blood tests to check their hematocrits. Cycling's Commission for Safety in Sport decided that athletes with a hematocrit over 50 would be declared unfit to continue in the event. This would have the same effect as a disqualification for a doping offense, but without the same stigma, as there was still no proof that an illegal substance had been taken. Athletes often claimed that they had naturally high hematocrits, had recently returned from altitude, or were severely dehydrated, which raises the blood cell-to-plasma ratio. It did not take long for the first real test case. In the 1999 Tour of Italy (Giro d'Italia), former Tour de France winner Marco Pantani was ejected late in the event for having a hematocrit well over 50. Pantani was the race leader





© Greg Breiber/epa/Corbis

■ The 2007 Tour de France was marred by several scandals, including the expulsion of Kazakhstan cycling star Alexandre Vinokourov, who was accused of having taken autologous blood transfusions to boost his red cell count.

at the time. When asked to comment, a well-known Italian coach replied, “After a three-week stage race, all of my riders’ hematocrits are down five points. Pantani’s is up five points. You can draw your own conclusions.” After serving a prolonged suspension, Pantani eventually returned to the sport, but never attained his former stature or reputation. He was found dead in his apartment in 2006, supposedly by his own hands.

Blood testing in sports reached a new level by 2004. Top athletes are tested frequently, both in competition and out. They must make their whereabouts known 24 hours a day, and the antidoping inspectors can show up at their homes or their hotels unannounced at any time. Sporting antidoping is now governed by the World Antidoping Agency (WADA). Each country has an affiliate. In the United States, that affiliate is the U.S. Antidoping Agency (USADA). In the early 2000s, researchers came up with a reliable test for recombinant EPO, and many athletes have been caught using this drug, which carries an immediate suspension of up to two years with the possibility of a lifetime ban for a subsequent offense. At the same time, researchers improved blood testing in a way that allows the detection of multiple types of blood cells. Because EPO was easier to find and because standard blood doping was problematical for

reasons discussed earlier, some athletes began having blood transfusions from relatives. An athlete could have a parent or sibling donate blood and then have the red blood cells injected. However, similar is not the same as identical. Red blood cells carry very specific markers (Chapter 16) that can be detected by fluorescence. A person’s blood should only have one type of blood cell in it, but a person who has received blood from someone else, other than an identical twin, will have multiple cell types and multiple markers. This technique led to the ejection of Alexandre Vinokourov after a difficult and occasionally brilliant 2007 Tour de France. He was suspended until 2009, returning to competition in 2010 as a member of the team of 2010 Tour de France winner, Alberto Contador.

Finally, researchers have perfected techniques of using carbon isotope ratios to identify unnatural molecules of otherwise natural compounds. This ability has led to the detection of synthetic testosterone in several athletes. In 2006, American Floyd Landis took over where Lance Armstrong left off, winning the Tour de France after the former champion’s retirement. Unfortunately, after stage 17, Landis tested positive for an elevated ratio of testosterone to its precursor

epitestosterone. Subsequent tests of carbon isotope ratios showed that the samples contained synthetic testosterone. Landis fought the finding in the American Arbitration Association (AAA) on the grounds of several alleged irregularities in the testing procedures. A hearing was held in May 2007. In September 2007, the AAA ruled that while the testing procedures used in the French lab were far from ideal, the results were enough to conclude that Landis had committed a doping infraction. He was stripped of his 2006 Tour de France title and served a suspension for two years.

Sports in general and cycling in particular are in a state of emergency. In the last several years many athletes have been found guilty of doping violations, and many more former athletes have come forward admitting that they too were using prohibited substances during their careers. The reputations of the sports and the athletes are suffering. Athletes and teams are finding it harder to find sponsors, and the public does not know what to think about their sporting heroes. There are two extreme views about how to solve the problem. One is to just let professional athletes do what they want, assuming that the ultimate health risks will be self-governing. The other is that



© Reed Saxon/AP

■ Floyd Landis carried on where Lance Armstrong left off, winning the 2006 Tour de France. Unfortunately, on stage 17 Landis tested positive for an elevated level of testosterone. Samples were later shown to include synthetic testosterone. In September of 2007, the American Arbitration Association ruled that the evidence presented in May of 2007 was conclusive that Landis’s sample taken after stage 17 of the 2006 Tour de France indicated he had committed a doping infraction. Floyd Landis was stripped of his 2006 Tour de France title and served two-year suspension.

athletes should be tested daily. Both extremes would be difficult. The first might force athletes to take drugs to be competitive, although it is certainly debatable whether this is true (see the article by Fitzgerald cited in the bibliography for a debate on this topic). The second would be prohibitively expensive. For reference, each lab test for a single sample costs between \$200 and \$400. At the Tour de France, the race leader is tested every day for 21 days. With his seven Tour de France victories, Lance Armstrong must clearly be the most tested athlete in the history of sport, not to mention the hundreds of out-of-competition tests he underwent in his long career (all without a single positive test, it should be noted).

The last few years of doping scandals have put a microscope on the lives and careers of professional athletes, and many suspect the problem does not stop there, extending into the realm of recreational athletes and “masters” age categories as well. The problems at the Tour de France in

2006 and 2007 are evidence that we have a long way to go before we can have confidence that athletes are all competing clean and fair. Some hopefuls have thought that the constant efforts were paying off and that the old riders who were doping were gradually being driven out of the sport, while the newer, younger riders would pave the way for a new era. Unfortunately, that belief was dealt a setback in February 2010, when the first- and second-place riders in the Under 23 category at the World Cyclocross Championships were found positive for EPO usage. As part of the latest fight against doping, the UCI has instituted a system called the “biological passport.” Professional riders are tested often, and at least quarterly, and the profiles of their blood are compared. In this way they can establish a baseline for how their blood should look. Any anomaly can indicate a potential issue, and then the rider can be checked more thoroughly. It is safe to say that the next few years will see a major reorganization in the sporting world

regarding this topic. In the meantime, scientists are working to create reliable tests of other current doping methods, such as the use of human growth hormone. WADA declared a victory in 2010 with a new test for HGH that caught a British Rugby star using this banned substance. They are also worried that the next frontier for doping will involve gene doping, using gene therapy techniques (Chapter 13) to deliver critical genes for muscle growth and endurance. For more information on this possibility, see the article by Sweeney cited in the bibliography.

## Annotated Bibliography

Fitzgerald, M. No Need to Cheat. *Running Times* (January/February), 35–40 (2006). [A point/counterpoint debate between two famous runners over whether it is possible to compete clean at the world level.]

Sweeney, H. L. Gene Doping. *Sci. Amer.* (July), 63–69 (2004). [An article describing what might be the next wave in doping in sports – doping at the genetic level.]





# Glossary

**abzyme** an antibody that is produced against a transition-state analog and that has catalytic activity similar to that of a naturally occurring enzyme (7.7)

**accessory pigments** plant pigments other than chlorophyll that play a role in photosynthesis (22.1)

**acid dissociation constant** a number that characterizes the strength of an acid (2.3)

**acidic domain** a common motif in the transcription-activation domain portion of a transcription factor (11.6)

**acid strength** the tendency of an acid to dissociate to a hydrogen ion and its conjugate base (2.3)

**acromegaly** a disease caused by an excess of growth hormone after the skeleton has stopped growing and characterized by enlarged hands, feet, and facial features (24.3)

**activation energy** the energy required to start a reaction (6.2)

**activation step** the beginning of a multistep process in which a substrate is converted to a more reactive compound (12.1)

**active site** the part of an enzyme to which the substrate binds and at which the reaction takes place (6.4)

**active transport** the energy-requiring process of moving substances into a cell against a concentration gradient (8.6)

**acyl carrier protein** a protein that functions in fatty-acid synthesis to carry activated carbon groups (21.6)

**acyl-CoA synthetase** the enzyme that catalyzes the activation step in lipid catabolism (21.6)

**adenine** one of the purine bases found in nucleic acids (9.2)

**S-adenosylmethionine** an important carrier molecule of methyl groups in amino acid metabolism (23.4)

**adenylate cyclase** the enzyme that catalyzes the production of cyclic AMP (24.3)

**A-DNA** a form of a DNA double helix characterized by having fewer residues per turn and major and minor grooves with dimensions that are more similar to each other than those of B-DNA (9.3)

**ADP** (adenosine diphosphate) a compound that can serve as an energy carrier when it is phosphorylated to form ATP (15.6)

**adrenocortical hormones** steroid hormones secreted by the adrenal cortex that have an effect on inflammation and salt and water balance (24.3)

**affinity chromatography** a powerful column separation procedure based on specific binding of molecules to a ligand (5.2)

**agarose** a complex polysaccharide used to make up resins for use in electrophoresis and in column chromatography (5.2)

**alcoholic fermentation** the anaerobic pathway that converts glucose to ethanol (17.1)

**aldolase** in glycolysis, the enzyme that catalyzes the reverse aldol condensation of fructose-1,6-bisphosphate (17.2)

**aldose** a sugar that contains an aldehyde group as part of its structure (16.1)

**alleles** corresponding genes on paired chromosomes (13.7)

**allosteric** the property of multisubunit proteins such that a conformational change in one subunit induces a change in another subunit (4.5)

**allosteric effector** a substance—substrate, inhibitor, or activator—that binds to an allosteric enzyme and affects its activity (7.4)

**amino acid** any of the fundamental building blocks of proteins; molecules

that contain an amino group and a carboxyl group (3.1)

**amino acid activation** the formation of an ester bond between an amino acid and its specific tRNA that is catalyzed by a suitable synthetase (12.1)

**amino acid analyzer** an instrument that gives information on the number and kind of amino acids in a protein (5.4)

**aminoacyl-tRNA synthetases** enzymes that catalyze the formation of an ester linkage between an amino acid and tRNA (12.1)

**amino group** the  $\text{NH}_2$  functional group (3.1)

**amphibolic** able to be a part of both anabolism and catabolism (19.1)

**amphipathic** refers to a molecule that has one end with a polar, water-soluble group and another end with a nonpolar hydrocarbon group that is insoluble in water (2.1)

**amylopectin** a form of starch; a branched-chain polymer of glucose (16.4)

**amylose** a form of starch; a linear polymer of glucose (16.4)

**anabolism** the synthesis of biomolecules from simpler compounds (15.3)

**anaerobic glycolysis** the pathway of conversion of glucose to lactate; distinguished from glycolysis, which is the conversion of glucose to pyruvate (17.1)

**analytical ultracentrifugation** the technique for observing the motion of particles as they sediment in a centrifuge (10.2)

**anaplerotic** referring to a reaction that ensures an adequate supply of an important metabolite (19.8)

**anion-exchange chromatography** the type of ion-exchange chromatography in which the column ligand has a net positive charge and binds to negatively charged molecules flowing through it (5.2)

**anomer** one of the possible stereoisomers formed when a sugar assumes the cyclic form (16.1)

**anomeric carbon** the chiral center created when a sugar cyclizes (16.1)

**antibody** a glycoprotein that binds to and immobilizes a substance that the cell recognizes as foreign (14.3)

**anticodon** the sequence of three bases (triplet) in tRNA that hydrogen-bonds with the mRNA triplet that specifies a given amino acid (12.2)

**antigen** a substance that triggers an immune response (14.3)

**antigenic determinant** the portion of a molecule that antibodies recognize as foreign and to which they bind (14.3)

**antioxidant** a strong reducing agent, which is easily oxidized and thus prevents the oxidation of other substances (8.7)

**antisense strand** the DNA strand that is used as a template for RNA synthesis (11.2)

**2·3 antiterminator** a hairpin loop that can form during transcription attenuation, allowing transcription to continue (11.3)

**AP (apurinic) site** a site, lacking a purine base in DNA, that is targeted by repair enzymes (10.5)

**arachidonic acid** a fatty acid that contains 20 carbon atoms and 4 double bonds; the precursor of prostaglandins and leukotrienes (8.8)

**aspartate transcarbamoylase (ATCase)** a classic example of an allosteric enzyme that catalyzes an early reaction in pyrimidine biosynthesis (6.5)

**ATP (adenosine triphosphate)** a universal energy carrier (15.6)

**ATP synthase** the enzyme responsible for production of ATP in mitochondria (20.4)

**attenuation** a type of transcription control in which the transcription is controlled after it has begun via pausing and early release of incomplete RNA sequences (11.3)

**autoimmune diseases** diseases in which the immune system attacks the body's own tissues (14.3)

**autoradiography** the technique of locating radioactively labeled substances by allowing them to expose photographic film (13.1)

**bacterial plasmid** a portion of circular DNA separate from the main genome of the bacterium (13.3)

**bacteriophage** a kind of virus that infects bacteria; bacteriophages are frequently used in molecular biology to transfer DNA between cells (13.2)

**$\beta$ -barrel** a  $\beta$ -pleated sheet extensive enough to fold back on itself (4.3)

**base-excision repair** a process for repairing damaged DNA (10.5)

**base stacking** interactions between bases that are next to each other in a DNA chain (9.3)

**basic-region leucine zipper (bZIP)** a common motif found in transcription factors (11.6)

**B cell** a type of white blood cell that plays an important role in the immune system and plays a role in the production of antibodies (14.3)

**B-DNA** the most common form of the DNA double helix (9.3)

**binding assay** an experimental method for selecting one molecule out of a number of possibilities by specific binding; it is used to determine the nature of many triplets of the genetic code (12.2)

**bioinformatics** the application of computer methods to processing large amounts of information in biochemistry (4.6)

**biotin** a CO<sub>2</sub> carrier molecule (18.2)

**blotting** a technique for transferring a portion of a sample to a membrane for further analysis (13.7)

**blue/white screening** a method for determining whether bacterial cells have incorporated a plasmid that includes a gene of interest (13.3)

**Bohr effect** the decrease in oxygen binding by hemoglobin caused by binding of carbon dioxide and hydrogen ion (4.7)

**buffering capacity** a measure of the amount of acid or base that reacts with a given buffer solution (2.5)

**buffer solution** a solution that resists a change in pH on addition of moderate amounts of strong acid or strong base (2.5)

**Calvin cycle** the pathway of carbon dioxide fixation in photosynthesis (22.5)

**5' cap** a structure found at the 5' end of eukaryotic mRNA (11.6)

**capsid** the protein coat of a virus (14.1)

**carboxyl group** the —COOH functional group that dissociates to give the carboxylate anion, —COO<sup>−</sup>, and a hydrogen ion (3.1)

**$\beta$ -carotene** an unsaturated hydrocarbon; the precursor of vitamin A (8.7)

**carrier protein** a membrane protein to which a substance binds in passive transport into the cell (8.6)

**cascade** a series of steps that take place in hormonal control of metabolism, affecting a series of enzymes and amplifying the effect of a small amount of hormone (24.4)

**catabolism** the breakdown of nutrients to provide energy (15.3)

**catabolite activator protein (CAP)** a protein that can bind to a promoter when complexed with cAMP, allowing RNA polymerase to bind to its entry site on the same promoter (11.3)

**catabolite repression** repression of the synthesis of *lac* proteins by glucose (11.3)

**catalysis** the process of increasing the rate of chemical reactions (1.3)

**cation-exchange chromatography** the type of ion-exchange chromatography in which the column resin has a net negative charge and binds to positively charged molecules flowing through the column (5.2)

**cDNA** complementary DNA; a form of DNA synthesized on an mRNA template, thus directly reflecting the coding sequence (13.5)

**cell membrane** the outer membrane of the cell that separates it from the outside world (1.5)

**cellulose** a polymer of glucose; an important structural material in plants (16.4)

**cell wall** the outer coating of bacterial and plant cells (1.5)

**ceramide** a lipid that contains one fatty acid linked to sphingosine by an amide bond (8.2)

**cerebroside** a glycolipid that contains sphingosine and a fatty acid in addition to the sugar moiety (8.2)

**chaperone** a protein that helps another protein fold into the correct three-dimensional structure and prevents it from associating incorrectly with other proteins (4.6)

**chemiosmotic coupling** the mechanism for coupling electron transport to oxidative phosphorylation; it requires

a proton gradient across the inner mitochondrial membrane (20.5)

**chimeric DNA** DNA from more than one species covalently linked together (13.3)

**chiral** refers to an object that is not superimposable on its mirror image (3.1)

**chlorophyll** the principal photosynthetic pigment responsible for trapping light energy from the Sun (22.1)

**chloroplast** the organelle that is the site of photosynthesis in green plants (1.6, 22.1)

**cholera** a disease caused by the bacterium *Vibrio cholerae* and characterized by dehydration due to excessive  $\text{Na}^+$  transport in epithelial cells (24.3)

**cholesterol** a steroid that occurs in cell membranes; the precursor of other steroids (8.2, 21.8)

**chromatin** a complex of DNA and protein found in eukaryotic nuclei (1.6)

**chromatography** an experimental method for separating substances based on their molecular character (5.2)

**chromosome** a linear structure that contains the genetic material and associated proteins (1.6)

**chymotrypsin** a proteolytic enzyme that preferentially hydrolyzes amide bonds adjacent to aromatic amino acid residues (5.4)

**cis-trans isomerase** an enzyme that catalyzes a *cis-trans* isomerization in the catabolism of unsaturated fatty acids (21.2)

**citrate synthase** the enzyme that catalyzes the first step of the citric acid cycle (19.4)

**citric acid cycle** a central metabolic pathway; part of aerobic metabolism (19.1)

**clonal selection** the process by which the immune system responds selectively to antibodies actually present in an organism (14.3)

**clone** a genetically identical population of organisms, cells, viruses, or DNA molecules (13.3)

**cloning of DNA** the introduction of a section of DNA into a genome in which it can be reproduced many times (13.3)

**closed complex** the complex that initially forms between RNA polymerase and DNA before transcription begins (11.2)

**coding strand** the DNA strand that has the same sequence as the RNA that is synthesized from the template (11.2)

**codon** a sequence of three bases on mRNA that specifies a given amino acid (12.2)

**coenzyme** a nonprotein substance that takes part in an enzymatic reaction and is regenerated at the end of the reaction (7.8)

**coenzyme A** a carrier of carboxylic acids bound to its thiol group by a thioester linkage (15.7)

**coenzyme Q** an oxidation-reduction coenzyme in mitochondrial electron transport (20.3)

**column chromatography** a form of chromatography in which the stationary phase is packed in a column (5.2)

**committed step** in a metabolic pathway, the formation of a substance that can play no other role in metabolism but to undergo the rest of the reactions of the pathway (17.2)

**competitive inhibition** a decrease in enzymatic activity caused by binding of a substrate analogue to the active site (6.7)

**complementary** refers to the specific hydrogen bonding of adenine with thymine (or uracil) and guanine with cytosine in nucleic acids (9.3)

**complete protein** a protein that contains all the essential amino acids (4.2)

**concerted model** a description of allosteric activity in which the conformations of all subunits change simultaneously (7.2)

**configuration** the three-dimensional arrangement of groups around a chiral carbon atom (16.1)

**conformational coupling** a mechanism for coupling electron transport to oxidative phosphorylation that depends on a conformational change in the ATP synthetase (20.5)

**consensus sequences** DNA sequences to which RNA polymerase binds; they are identical in many organisms (11.2)

**constitutive expression** the transcription and expression of genes that are not controlled by anything other than the inherent binding of the RNA polymerase to the promoter (11.3)

**control sites** the operator and promoter elements that modulate the production of proteins whose amino acid sequence

is specified by the structural genes under their control (11.3)

**cooperative binding** binding to several sites such that, when the first ligand is bound, the binding of subsequent ones is easier (4.7)

**cooperative transition** a transition that takes place in an all-or-nothing fashion, such as the melting of a crystal (4.7)

**co-repressor** a substance that binds to a repressor protein, making it active and able to bind to an operator gene (11.3)

**core promoter** in prokaryotic transcription, the portion of the DNA from the transcription start site to the -35 region (11.2)

**Cori cycle** a pathway in carbohydrate metabolism that links glycolysis in the liver with gluconeogenesis in the liver (18.3)

**coupled translation** in prokaryotes, the situation in which a gene is simultaneously transcribed and translated (12.4)

**coupling** the process by which an exergonic reaction provides energy for an endergonic one (15.6)

**C-terminal** the end of a protein or peptide with a carboxyl group not bonded to another amino acid (3.4)

**C-terminal domain** the region of a protein at the C-terminus, especially important in eukaryotic RNA polymerase B (11.4)

**cut and patch** a mechanism for repair of DNA by enzymatically removing incorrect nucleotides and substituting correct ones (10.5)

**cyanogen bromide** a reagent that cleaves proteins at internal methionine residues (5.4)

**cyclic AMP** a nucleotide in which the same phosphate group is esterified to the 3' and 5' hydroxyl groups of a single adenosine; an important second messenger (24.3)

**cyclic-AMP-response element (CRE)** an important eukaryotic response element that is controlled by production of cAMP in the cell (11.5)

**cyclic-AMP-response-element binding protein (CREB)** an important transcription factor in eukaryotes that binds to the CRE and activates transcription (11.5)

**cyclins** proteins that play an important role in control of the cell cycle by regulating the activity of kinases (10.6)



**cytochrome** any one of a group of heme-containing proteins in the electron transport chain (20.3)

**cytokines** soluble protein factors produced by one cell that affect another cell (14.3)

**cytosine** one of the pyrimidine bases found in nucleic acids (9.2)

**cytoskeleton (microtrabecular lattice)** a lattice of fine strands, consisting mostly of protein, that pervades the cytosol (1.6)

**cytosol** the portion of the cell that lies outside the nucleus and the other membrane-enclosed organelles (1.6)

**debranching enzyme** an enzyme that hydrolyzes the linkages in a branched-chain polymer such as amylopectin (18.1)

**degenerate code** a genetic code in which more than one triplet of bases can code for the same amino acid (12.2)

**denaturation** the unraveling of the three-dimensional structure of a macromolecule caused by the breakdown of noncovalent interactions (4.4)

**dendritic cells** components of the innate immune system (14.3)

**denitrification** the process by which nitrates and nitrites are broken down to molecular nitrogen (23.1)

**density-gradient centrifugation** the technique of separating substances in an ultracentrifuge by applying the sample to the top of a tube that contains a solution of varying densities (10.2)

**deoxyribonucleoside** a compound formed when a nucleobase and deoxyribose form a glycosidic bond (9.2)

**deoxyribose** a sugar that is part of the structure of DNA (9.2)

**deoxy sugar** a sugar in which one of the hydroxyl groups has been reduced to a hydrogen (16.2)

**dextran** a complex polysaccharide that is often used in column chromatography resins (5.2)

**diastereomers** nonsuperimposable, non-mirror-image stereoisomers (16.1)

**dimer** a molecule consisting of two subunits (4.5)

**dipoles** molecules with positive and negative ends due to an uneven distribution of electrons in bonds (2.1)

**disaccharide** two monosaccharides (monomeric sugars) linked by a glycosidic bond (16.3)

**DNA** deoxyribonucleic acid; the molecule that contains the genetic code (9.3)

**DNA-binding domain** the part of a transcription factor that binds to the DNA (11.6)

**DNA chip** a microarray of DNA samples on a single computer chip, on which many samples can be examined simultaneously (13.9)

**DNA gyrase** an enzyme that introduces supercoiling into closed circular DNA (9.3)

**DNA library** a collection of clones that include the total genome of an organism (13.5)

**DNA ligase** the enzyme that links separate stretches of DNA (10.3)

**DNA polymerase** the enzyme that forms DNA from deoxyribonucleotides on a DNA template (10.3)

**DNase** deoxyribonuclease; an enzyme that specifically hydrolyzes DNA (13.2)

**domain** a portion of a polypeptide chain that folds independently of other portions of the chain (4.3)

**downstream** in transcription, a portion of the DNA sequence that is nearer the 5' end in the gene to be transcribed, where the DNA is read from the 3' to the 5' end and the RNA is formed from the 5' to the 3' end; in translation, a portion of the RNA sequence that is nearer the 3' end of mRNA (13.1)

**dwarfism** a disease caused by a deficiency of growth hormone (24.4)

**Edman degradation** a method for determining the amino acid sequence of peptides and proteins (5.4)

**electronegativity** a measure of the tendency of an atom to attract electrons to it in a chemical bond (2.1)

**electron transport to oxygen** a series of oxidation–reduction reactions by which the electrons derived from oxidation of nutrients are passed to oxygen (19.1)

**electrophile** an electron-poor substance that tends to react with centers of negative charge or polarization (7.5)

**electrophoresis** a method for separating molecules on the basis of the ratio of charge to size (3.3)

**–35 element (or –35 region)** a portion of DNA that is 35 base pairs upstream from the start of RNA transcription that is important in control of RNA synthesis in bacteria (11.2)

**elongation step** in protein synthesis, the succession of reactions in which the peptide bonds are formed (12.4)

**enantiomers** mirror-image, nonsuperimposable stereoisomers (16.1)

**endergonic** energy-absorbing (1.11)

**endocrine system** the system of ductless glands that release hormones into the bloodstream (24.3)

**endocytosis** the process by which materials are brought into the cell when portions of a cell membrane are pinched off into the cell (8.6)

**endonuclease** an enzyme that hydrolyzes nucleic acids, attacking linkages in the middle of the polynucleotide chain (13.2)

**endoplasmic reticulum (ER)** a continuous single-membrane system throughout the cell (1.6)

**endosymbiosis** a symbiotic relationship in which a smaller organism is completely contained within a larger organism (1.8)

**enhancer** a DNA sequence that binds to a transcription factor and increases the rate of transcription (11.3)

**enthalpy** a thermodynamic quantity measured as the heat of reaction at constant pressure (1.11)

**entropy** a thermodynamic quantity; a measure of the energy dispersal of the Universe (1.11)

**enzyme** a biological catalyst, usually a globular protein, with self-splicing RNA as the only exception (6.1)

**epimerase** an enzyme that catalyzes the inversion of configuration around a single carbon atom

**epimers** stereoisomers that differ only in configuration around one of several chiral carbon atoms (16.1)

**epitope** a binding site for an antibody on an antigen (14.3)

**equilibrium** the state in which a forward process and a reverse process occur at the same rate (1.11)

**essential amino acids** amino acids that cannot be synthesized by the body and must therefore be obtained in the diet (23.4)

**essential fatty acids** the polyunsaturated fatty acids (such as linoleic acid) that the body cannot synthesize; they must be obtained from dietary sources (21.6)

**eukaryote** an organism whose cells have a well-defined nucleus and membrane-enclosed organelles (1.4)



**excision repair** repair of DNA by the enzymatic removal of incorrect nucleotides and their replacement by the correct ones (10.5)

**exergonic** energy-releasing (1.11)

**exon** a DNA sequence that is expressed in the sequence of mRNA (11.7)

**exonuclease** an enzyme that hydrolyzes nucleic acids, starting at the end of the polynucleotide chain (13.2)

**expression cassette** in gene therapy, the assemblage that contains the gene being transferred (14.2)

**expression vector** a plasmid that has the machinery to direct the synthesis of a desired protein (13.4)

**extended promoter** in prokaryotic transcription, the DNA from the transcription start site to the UP element (11.2)

**facilitated diffusion** a process by which substances enter a cell by binding to a carrier protein; this process does not require energy (8.6)

**fatty acid** a compound with a carboxyl group at one end and a long, normally unbranched hydrocarbon tail at the other; the hydrocarbon tail may be saturated or unsaturated (8.1)

**feedback inhibition** the process by which the final product of a series of reactions inhibits the first reaction in the series (7.1, 23.3)

**fibrous protein** a protein whose overall shape is that of a long, narrow rod (4.3)

**filter-binding assay** a method used to determine the base sequence of many mRNA codons (12.2)

**first-order reaction** a reaction whose rate depends on the first power of the concentration of a single reactant (6.3)

**Fischer projection** a two-dimensional representation of the stereochemistry of three-dimensional molecules (16.1)

**Fis site** an enhancer found in prokaryotic transcription of rRNA (11.3)

**fluid-mosaic model** the model for membrane structure in which proteins and a lipid bilayer exist side by side without covalent bonds between the proteins and lipids (8.5)

**fluorescence** a sensitive method for detection and identification of substances that absorb and re-emit light (13.1)

**folate reductase** the enzyme that reduces dihydrofolate to

tetrahydrofolate, a target for cancer chemotherapy (23.11)

**fold purification** a measurement of increased purity taken during a protein-purification experiment (5.1)

**N-formylmethionine-tRNA<sup>fmet</sup>** an essential factor for the start of prokaryotic protein synthesis, the amino acid methionine formylated at the amino group and covalently bonded to its specific tRNA (12.4)

**free energy** a thermodynamic quantity; diagnostic for the spontaneity of a reaction at constant temperature (1.11)

**functional group** one of the groups of atoms that give rise to the characteristic reactions of organic compounds (1.2)

**Fungi** one of the five kingdoms into which living organisms are commonly classified; it includes molds and mushrooms (1.7)

**furanose** a cyclic sugar with a six-membered ring, named for its resemblance to the ring system in furan (16.1)

**furanoside** a glycoside involving a furanose (16.2)

**fusion protein** one that has had extra amino acids added to one of its ends (13.4)

**$\beta$ -galactosidase** the enzyme that hydrolyzes lactose to galactose and glucose, the classic example of an inducible enzyme (11.3)

**gel electrophoresis** a method for separating molecules on the basis of charge-to-size ratio using a gel as a support and sieving material (5.3)

**gel-filtration chromatography** a type of column chromatography in which the molecules are separated according to size as they pass through the column (5.2)

**gene** an individual unit of inheritance (1.4)

**general acid-base catalysis** a form of catalysis that depends on transfer of protons (7.6)

**gene therapy** a method for treating a genetic disease by introducing a good copy of a defective gene (13.4)

**genetic code** the information for the structure and function of all living organisms (1.3)

**genome** the total DNA of the cell (1.4)

**gigantism** a disease caused by overproduction of growth hormone

before the skeleton has stopped growing (24.3)

**globular protein** a protein whose overall shape is more or less spherical (4.3)

**glucocorticoid** a kind of steroid hormone involved in the metabolism of sugars (24.3)

**glucogenic amino acid** an amino acid that has pyruvate or oxaloacetate as a catabolic breakdown product (23.6)

**gluconeogenesis** the pathway of synthesis of glucose from lactate (18.2)

**glucose** a monosaccharide; a ubiquitous metabolite (16.1)

**glyceraldehyde** the simplest carbohydrate that contains a chiral carbon, the starting point of a system of describing optical isomers (16.1)

**glyceraldehyde-3-phosphate** a key intermediate in the reactions of sugars (17.3)

**glyceraldehyde-3-phosphate dehydrogenase** an important enzyme in glycolysis and gluconeogenesis (17.3)

**glycerol phosphate shuttle** a mechanism for transferring electrons from NADH in the cytosol to FADH<sub>2</sub> in the mitochondrion (20.7)

**glycogen** a polymer of glucose; an important energy storage molecule in animals (16.4)

**glycolipid** a lipid to which a sugar moiety is bonded (8.2)

**glycolysis** the anaerobic breakdown of glucose to three-carbon compounds (17.1)

**glycoside** a compound in which one or more sugars is bonded to another molecule (16.2)

**glyoxylate cycle** a pathway in plants that is an alternative to the citric acid cycle and that bypasses several citric acid cycle reactions (19.6)

**glyoxysomes** membrane-enclosed organelles that contain the enzymes of the glyoxylate cycle (19.6)

**Golgi apparatus** a cytoplasmic organelle that consists of flattened membranous sacs, usually involved in secretion of proteins (1.6)

**G protein** a membrane-bound protein that mediates the action of adenylate cyclase (24.3)

**grana** bodies within the chloroplast that contain the thylakoid disks, the site of photosynthesis (22.1)

**guanine** one of the purine bases found in nucleic acids (9.2)

**half reaction** an equation that shows either the oxidative or the reductive part of an oxidation–reduction reaction (15.4)

**Haworth projection formulas** a perspective representation of the cyclic forms of sugars (16.1)

**helicase (rep protein)** a protein that unwinds the double helix of DNA in the process of replication (10.4)

**$\alpha$ -helix** one of the most frequently encountered folding patterns in the protein backbone (4.3)

**helix–turn–helix** a common motif found in the DNA-binding domain of transcription factors (4.3)

**helper T cells** components of the human immune system; the target of the AIDS virus (14.3)

**heme** an iron-containing cyclic compound found in cytochromes, hemoglobin, and myoglobin (4.4)

**hemiacetal** a compound that is formed by reaction of an aldehyde with an alcohol and is found in the cyclic structure of sugars (16.1)

**hemiketal** a compound that is formed by reaction of a ketone with an alcohol and is found in the cyclic structure of sugars (16.1)

**Henderson–Hasselbalch equation** a mathematical relationship between the  $pK_a$  of an acid and the pH of a solution containing the acid and its conjugate base (2.4)

**heteropolysaccharide** a polysaccharide that contains more than one kind of sugar monomer (16.4)

**heterotropic effects** allosteric effects that occur when different substances are bound to a protein (7.1)

**heterozygous** exhibiting differences in a given gene on one chromosome and the corresponding gene on the paired chromosome (13.7)

**hexokinase** the first enzyme of glycolysis (17.2)

**hexose monophosphate shunt** a synonym for the pentose phosphate pathway, in which glucose is converted to five-carbon sugars with concomitant production of NADPH (18.4)

**high-performance liquid**

**chromatography (HPLC)** a sophisticated chromatography technique that gives fast and clean purifications (5.4)

**histones** basic proteins found complexed to eukaryotic DNA (9.3)

**hnRNA** heterogeneous nuclear RNA; the original form of mRNA in eukaryotes that contains intervening sequences (9.5)

**holoenzyme** an enzyme that has all component parts, including coenzymes and all subunits (11.2)

**homeostasis** the balance of biological activities in the body (24.3)

**homology** similarity of monomer sequences in polymers (4.6)

**homopolysaccharide** a polysaccharide that contains only one kind of sugar monomer (16.4)

**homotropic effects** allosteric effects that occur when several identical molecules are bound to a protein (7.1)

**homozygous** exhibiting no differences between a given gene on one chromosome and the corresponding gene on the paired chromosome (13.7)

**hormone** a substance produced by endocrine glands and delivered by the bloodstream to target cells, producing a regulatory effect (24.3)

**hsp70** a protein that acts as a chaperone in *E. coli*. The letters *hsp* stand for “heat-shock protein,” since this chaperone is produced when the bacteria are grown at elevated temperatures (12.6)

**hyaluronic acid** a polysaccharide found in the lubricating fluid of joints (16.4)

**hydrogen bonding** a noncovalent association formed between a hydrogen atom covalently bonded to one electronegative atom and a lone pair of electrons on another electronegative atom (2.2)

**hydrophilic** tending to dissolve in water (2.1)

**hydrophobic** tending not to dissolve in water (2.1)

**hydrophobic bonds** attractions between molecules that are nonpolar; also called *hydrophobic interactions* (2.1)

**$\beta$ -hydroxy- $\beta$ -methylglutaryl-CoA** an intermediate in the biosynthesis of cholesterol (21.8)

**hyperbolic** a characteristic of a curve on a graph such that it rises quickly and then levels off (4.5)

**hyperglycemia** the condition of elevated blood glucose levels (24.3)

**hypoglycemia** the condition of low blood glucose levels (24.3)

**hypothalamus** the portion of the brain that controls, among other things, many of the workings of the endocrine system (24.3)

**immunoglobulins** another name for antibodies, proteins that play a role in the immune system (14.3)

**induced-fit model** a description of substrate binding to an enzyme such that the conformation of the enzyme changes to accommodate the shape of the substrate (6.4)

**inducible enzyme** an enzyme whose synthesis can be triggered by the presence of some substance, which is called the inducer (11.3)

**induction of enzyme synthesis** the triggering of the production of an enzyme by the presence of a specific inducer (11.3)

**inhibitor** a substance that decreases the rate of an enzyme-catalyzed reaction (6.7)

**initial rate** the rate of a reaction immediately after it starts, before any significant accumulation of product (6.6)

**initiation complex** the aggregate of mRNA, *N*-formylmethionine-tRNA, ribosomal subunits, and initiation factors needed at the start of protein synthesis (12.4)

**initiation factor** any of a large group of proteins that bind to DNA and play a role in initiation of eukaryotic RNA synthesis (12.5)

**initiation step** the start of protein synthesis; the formation of the initiation complex (12.4)

**initiator element** a loosely conserved sequence surrounding the transcription start site in eukaryotic DNA (11.4)

**integral protein** a protein that is embedded in a membrane (8.4)

**interleukins** proteins that play a role in the immune system (14.3)

**intermembrane space** the region between the inner and outer mitochondrial membranes (19.2)

**intrinsic termination** the type of transcription termination that is not dependent on the rho protein (11.2)

**intron** an intervening sequence in DNA that does not appear in the final sequence of mRNA (11.7)

**ion-exchange chromatography** a method for separating substances on the basis of charge (5.2)

**ion product constant for water** a measure of the tendency of water to

dissociate to give hydrogen ion and hydroxide ion (2.3)

**irreversible inhibition** covalent binding of an inhibitor to an enzyme, causing permanent inactivation (6.7)

**isoelectric focusing** a method for separating substances on the basis of their isoelectric points (5.3)

**isoelectric point (isoelectric pH)** the pH at which a molecule has no net charge (3.3)

**isoprene** a five-carbon unsaturated group, which is part of the structure of many lipids (21.8)

**isozymes** multiple forms of an enzyme that catalyze the same overall reaction but have subtle physical and kinetic parameters (6.2)

**ketogenic amino acid** amino acid that has acetyl-CoA or acetoacetyl-CoA as a catabolic breakdown product (23.6)

**$\alpha$ -ketoglutarate dehydrogenase complex** one of the enzymes of the citric acid cycle; it catalyzes the conversion of  $\alpha$ -ketoglutarate to succinyl-CoA (19.4)

**ketone body** one of several ketone-based molecules produced in the liver during overutilization of fatty acids when carbohydrates are limited (21.5)

**ketose** a sugar that contains a ketone group as part of its structure (16.1)

**killer T cells** components of the human immune system (14.3)

**kinase** a phosphate-transfer enzyme, with ATP as the usual source of the phosphate (17.3)

**Kozak sequence** the base sequence that identifies the start codon in eukaryotic protein synthesis (12.5)

**Krebs cycle** an alternative name for the citric acid cycle (19.1)

**K system** a combination of an allosteric enzyme and an inhibitor or activator, in which the presence of the inhibitor/activator changes the substrate concentration that yields one-half  $V_{\max}$  (7.1)

**kwashiorkor** a disease caused by serious protein deficiency (23.5)

**labeling** covalent modification of a specific residue on an enzyme (7.5)

**lac operon** the promoter, operator, and structural genes involved in the induction of  $\beta$ -galactosidase and related proteins (11.3)

**lactate dehydrogenase** an NADH-linked dehydrogenase that catalyzes the conversion of pyruvate to lactate (17.4)

**lagging strand** in DNA replication, the strand that is formed in small fragments that are subsequently joined by DNA ligase (10.2)

**L and D amino acids** amino acids whose stereochemistry is the same as the stereochemical standards L- and D-glyceraldehyde, respectively (3.1)

**lanosterol** a precursor of cholesterol (21.8)

**leading strand** in DNA replication, the strand that is continuously formed in one long stretch (10.3)

**leucine zipper (bZIP)** a structural motif found in DNA-binding proteins (11.6)

**leukocytes** white blood cells that play an important role in the functioning of the immune system (14.3)

**leukotriene** a substance derived from leukocytes (white blood cells) that has three double bonds; it is of pharmaceutical importance (8.8)

**lignin** a polymer of coniferyl alcohol; a structural material found in woody plants (16.4)

**Lineweaver–Burk double-reciprocal plot** a graphical method for analyzing the kinetics of enzyme-catalyzed reactions (6.6)

**lipase** an enzyme that hydrolyzes lipids (8.2)

**lipid** a compound insoluble in water but soluble in organic solvents (8.1)

**lipid bilayers** aggregates of lipid molecules in which the polar head groups are in contact with water and the hydrophobic parts are not (8.3)

**liposome** a spherical aggregate of lipids arranged so that the polar head groups are in contact with water and the nonpolar tails are sequestered from water (8.5)

**lock-and-key model** a description of the binding of a substrate to an enzyme such that the active site and the substrate exactly match each other in shape (6.4)

**luminescence** emission of light as a result of a chemical reaction (chemiluminescence) or re-emission of absorbed light (fluorescence) (13.1)

**lymphocytes** a type of white blood cell; a major component of the immune system (14.3)

**lymphokines** soluble protein factors produced by one lymphocyte that affect another cell (14.3)

**lysosomes** membrane-enclosed organelles that contain hydrolytic enzymes (1.6)

**macronutrients** ones needed in large amounts, such as proteins, carbohydrates, or fats (24.2)

**macrophages** components of the innate immune system (14.3)

**major histocompatibility complex (MHC)** protein that displays an antigen on the surface of cells of the immune system (14.3)

**malate–aspartate shuttle** a mechanism for transferring electrons from NADH in the cytosol to NADH in the mitochondrion (20.7)

**Maloney murine leukemia virus (MMLV)** a vector commonly used in gene therapy (14.2)

**malonyl-CoA** a three-carbon intermediate important in the biosynthesis of fatty acids (21.6)

**matrix (mitochondrial)** the part of a mitochondrion enclosed within the inner mitochondrial membrane (1.6, 19.2)

**metabolic water** the water produced as a result of complete oxidation of nutrients; sometimes it is the only water source of desert-dwelling organisms (21.3)

**metabolism** the sum total of all biochemical reactions that take place in an organism (15.3)

**metal–ion catalysis** (Lewis acid–base catalysis) a form of catalysis that depends on the Lewis definition of an acid as an electron-pair acceptor and a base as an electron-pair donor (7.6)

**micelle** an aggregate formed by amphipathic molecules such that their polar ends are in contact with water and their nonpolar portions are on the interior (2.1)

**Michaelis constant** a numerical value for the strength of binding of a substrate to an enzyme; an important parameter in enzyme kinetics (6.6)

**micronutrients** vitamins and minerals that are needed in small amounts (24.2)

**mineralocorticoid** a kind of steroid hormone involved in the regulation of levels of inorganic ions “minerals” (24.3)



**minerals** in nutrition, inorganic substances required as the ion or free element (24.2)

**miRNA (micro RNA)** short stretches of RNA that affect gene expression and that play a role in growth and development (9.5)

**mismatch repair** a process for repairing damaged DNA (10.5)

**mitochondrion** an organelle that contains the apparatus responsible for aerobic oxidation of nutrients (1.6)

**mitogen-activated protein kinase (MAPK)** an enzyme that responds to cell growth and stress signals and phosphorylates key proteins that act as transcription factors (14.4)

**mobile phase (eluent)** in chromatography, the portion of the system in which the mixture to be separated moves (5.2)

**Monera** one of the five kingdoms used to classify living organisms; it includes prokaryotes (1.7)

**monoclonal antibodies** antibodies produced from the progeny of a single cell and specific for a single antigen (14.3)

**monomer** a small molecule that may bond to many others to form a polymer (1.3)

**monosaccharide** a compound that contains a single carbonyl group and two or more hydroxyl groups (16.1)

**motif** a repetitive supersecondary structure (4.3)

**mRNA (messenger RNA)** the kind of RNA that specifies the order of amino acids in a protein (9.5)

**mucopolysaccharide** a polysaccharide that has a gelatinous consistency (16.4)

**multifunctional enzyme** an enzyme in which a single protein catalyzes several reactions (21.6)

**multiple cloning site (MCS)** a region of a bacterial plasmid with many restriction sites (13.3)

**mutagen** an agent that brings about a mutation; such agents include radiation and chemical substances that alter DNA (10.5)

**mutation** a change in DNA, causing subsequent changes in the organism that can be transmitted genetically (10.5)

**myelin** the lipid-rich sheath of nerve cells (8.2)

**native conformation** a three-dimensional shape of a protein with biological activity (4.1)

**natural killer (NK) cells** components of the innate immune system (14.3)

**negative cooperativity** a cooperative effect whereby binding of the first ligand to an enzyme or protein causes the affinity for the next ligand to be lower (7.2)

**nick translation** a process for repairing damaged DNA (10.5)

**nicotinamide adenine dinucleotide** an important coenzyme in metabolism that is found in an oxidized or reduced form (7.8, 15.5)

**nitrification** the conversion of ammonia to nitrates (23.1)

**nitrogenase** the enzyme complex that catalyzes nitrogen fixation (23.2)

**nitrogen fixation** the conversion of molecular nitrogen to ammonia (23.1)

**noncompetitive inhibition** a form of enzyme inactivation in which a substance binds to a place other than the active site but distorts the active site so that the reaction is inhibited (6.7)

**nonheme (iron–sulfur) protein** a protein that contains iron and sulfur but no heme group (19.4)

**nonoverlapping, commaless code** a genetic code in which no bases are shared between the sequences of three bases (triplets) that specify an amino acid, with no intervening, noncoding bases (12.2)

**nonpolar bond** a bond in which two atoms share electrons evenly (2.1)

**nontemplate strand** the DNA strand that has the same sequence as the RNA that is synthesized from the template (11.2)

**N-terminal** the end of a protein or polypeptide with its amino group not linked to another amino acid by a peptide bond (3.4)

**nuclear magnetic resonance (NMR) spectroscopy** a method for determining the three-dimensional shape of proteins in solution (4.4)

**nuclear region** the portion of a prokaryotic cell that contains the DNA (1.5)

**nuclease** an enzyme that hydrolyzes a nucleic acid; it is specific for DNA or RNA (13.2)

**nucleic acid** a macromolecule formed by polymerization of nucleotides (1.3)

**nucleic-acid base (nucleobase)** one of the nitrogen-containing aromatic compounds that makes up the coding portion of a nucleic acid (9.2)

**nucleolus** a portion of the nucleus rich in RNA (1.6)

**nucleophile** an electron-rich substance that tends to react with sites of positive charge or polarization (7.5)

**nucleophilic substitution reaction** a reaction in which one functional group is replaced by another as the result of nucleophilic attack (7.6)

**nucleoside** a purine or pyrimidine base bonded to a sugar (ribose or deoxyribose) (9.2)

**nucleosome** a globular structure in chromatin in which DNA is wrapped around an aggregate of histone molecules (9.3)

**nucleotide** a purine or pyrimidine base bonded to a sugar (ribose or deoxyribose), which in turn is bonded to a phosphate group (9.2)

**nucleotide-excision repair** a process for repairing damaged DNA (10.5)

**nucleus** the organelle that contains the main genetic apparatus in eukaryotes (1.6)

**Okazaki fragments** short stretches of DNA formed in the lagging strand in replication that are subsequently linked by DNA ligase (10.3)

**oligomer** an aggregate of several smaller units (monomers); bonding may be covalent or noncovalent (4.5)

**oligosaccharide** a few sugars linked by glycosidic bonds (16.1)

**oncogene** a gene that causes cancer when a triggering event takes place (14.1, 14.4)

**one-carbon transfers** reactions in which the transfer usually involves carbon dioxide, a methyl group, or a formyl group (23.4)

**open complex** the form of the complex of RNA polymerase and DNA that occurs during transcription (11.2)

**operator** the DNA element to which a repressor of protein synthesis binds (11.3)

**operon** a group of operator, promoter, and structural genes (11.3)

**opsin** a protein in the rod and cone cells of the retina; it plays a crucial role in vision (8.7)

**optical isomers** (see *stereoisomers*) (16.1)  
**order of a reaction** the experimentally determined dependence of the rate of a reaction on substrate concentrations (6.3)

**organelle** a membrane-enclosed portion of a cell with a specific function (1.4)

**organic chemistry** the study of compounds of carbon, especially of carbon and hydrogen and their derivatives (1.2)

**origin of replication** the point at which the DNA double helix begins to unwind at the start of replication (10.2)

**origin recognition complex (ORC)** a protein complex bound to DNA throughout the cell cycle that serves as an attachment site for several proteins that help control replication (10.6)

**oxidation** the loss of electrons (1.9)

**$\beta$ -oxidation** the main pathway of catabolism of fatty acids (21.2)

**oxidative decarboxylation** loss of carbon dioxide accompanied by oxidation (19.2)

**oxidative phosphorylation** a process for generating ATP; it depends on the creation of a pH gradient within the mitochondrion as a result of electron transport (19.1)

**oxidizing agent** a substance that accepts electrons from other substances (15.4)

**oxygen-evolving complex** the part of photosystem II that splits water to produce oxygen (22.2)

**palindrome** a message that reads the same backward or forward (13.2)

**palmitate** a 16-carbon saturated fatty acid; the end product of fatty-acid biosynthesis in mammals (21.6)

**passive transport** the process by which a substance enters a cell without an expenditure of energy by the cell (8.6)  
**pause structure** a hairpin loop that can form during transcription attenuation, causing premature termination of transcription (11.3)

**pectin** a polymer of galacturonic acid; it occurs in the cell walls of plants (16.4)

**pentose phosphate pathway** a pathway in sugar metabolism that gives rise to five-carbon sugars and NADPH (18.4)

**peptide bond** an amide bond between amino acids in a protein (3.4)

**peptides** molecules formed by linking two to several dozen amino acids by amide bonds (3.3)

**peptidoglycan** a polysaccharide that contains peptide crosslinks; it is found in bacterial cell walls (16.4)

**peptidyl transferase** in protein synthesis, the enzyme that catalyzes formation of the peptide bond, part of the 50S ribosomal subunit (12.4)

**percent recovery** a measurement of the amount of an enzyme recovered at each step of a purification experiment (5.1)

**peripheral proteins** proteins loosely bound to the outside of a membrane (8.2)

**peroxisomes** membrane-bounded sacs that contain enzymes involved in the metabolism of hydrogen peroxide ( $\text{H}_2\text{O}_2$ ) (1.6)

**pH** a measure of the acidity of a solution (2.3)

**phenylketonuria** a disease characterized by mental retardation in developing children; it is caused by a lack of the enzyme that converts phenylalanine to tyrosine (3.5)

**pheophytin** a photosynthetic pigment that differs from chlorophyll only in having two hydrogens in place of magnesium (22.2)

**phosphatidic acid** a compound in which two fatty acids and phosphoric acid are esterified to the three hydroxyl groups of glycerol (8.2)

**phosphatidylinositol 4,5-bisphosphate ( $\text{PIP}_2$ )** a membrane-bound substance that mediates the action of  $\text{Ca}^{2+}$  as a second messenger (24.3)

**phosphoaclyglycerol (phosphoglyceride)** a phosphatidic acid (see *above*) with another alcohol esterified to the phosphoric acid moiety (8.2)

**3',5'-phosphodiester bond** a covalent linkage in which phosphoric acid is esterified to the 3' hydroxyl of one nucleoside and the 5' hydroxyl of another nucleoside; it forms the backbone of nucleic acids (9.2)

**phosphofructokinase** the key allosteric control enzyme in glycolysis; it catalyzes the phosphorylation of fructose-6-phosphate (17.2)

**phospholipase** an enzyme that hydrolyzes phospholipids (21.2)

**phosphorolysis** the addition of phosphoric acid across a bond, such as the glycosidic bond in glycogen, giving glucose phosphate and a glycogen

remainder one residue shorter; it is analogous to hydrolysis (addition of water across a bond) (18.1)

**photophosphorylation** the synthesis of ATP coupled to photosynthesis (22.2)

**photorespiration** the process by which plants oxidize carbohydrates aerobically in the light (22.6)

**photosynthesis** the process of using light energy from the Sun to drive the synthesis of carbohydrates (22.1)

**photosynthetic unit** the assemblage of chlorophylls that includes light-harvesting molecules and the special pair that actually carry out the reaction (22.2)

**photosystem I** the portion of the photosynthetic apparatus responsible for the production of NADPH (22.2)

**photosystem II** the portion of the photosynthetic apparatus responsible for the splitting of water to oxygen (22.2)

**pituitary** the gland that releases trophic hormones to specific endocrine glands; it is under the control of the hypothalamus (24.4)

**plasma membrane** another name for the cell membrane; the outer boundary of the cell (1.4)

**plasmid** a small, circular DNA molecule that usually contains genes for antibiotic resistance and is often used for cloning (13.2)

**plastocyanin** a copper-containing protein; it is part of the electron transport chain that links the two photosystems in photosynthesis (22.2)

**plastoquinone** a substance similar to coenzyme Q, part of the electron transport chain that links the two photosystems in photosynthesis (22.2)

**$\beta$ -pleated sheet** one of the most important types of secondary structure, in which the protein backbone is almost fully extended with hydrogen bonding between adjacent strands (4.3)

**polar bond** a bond in which two atoms have an unequal share in the bonding electrons (2.1)

**polyacrylamide gel electrophoresis (PAGE)** a form of electrophoresis in which a polyacrylamide gel serves as both a sieve and a supporting medium (5.2)

**poly A tail** a long sequence of adenosine residues at the 3' end of eukaryotic mRNA (11.7)



**polylinker** a region of a bacterial plasmid with many restriction sites (13.3)

**polymer** a macromolecule formed by the bonding of smaller units (1.3)

**polymerase chain reaction (PCR)**

a method for amplifying a small amount of DNA based on the reaction of isolated enzymes rather than on cloning (13.6)

**polypeptide chain** the backbone of a protein; it is formed by linking amino acids by peptide (amide) bonds (3.4)

**polysaccharide** a polymer of sugars (16.4)

**polysome** the assemblage of several ribosomes bound to one mRNA (12.4)

**P/O ratio** the ratio of ATP produced by oxidative phosphorylation to oxygen atoms consumed in electron transport (20.4)

**porphyrins** large-ring compounds formed by linking four pyrrole rings; they combine with iron ions to form the heme group (4.5)

**positive cooperativity** a cooperative effect whereby binding of the first ligand to an enzyme or protein causes the affinity for the next ligand to be higher (4.5)

**pre-replication complex (pre-RC)** the complex of DNA, recognition protein (ORC), activator protein (RAP), and licensing factors (RLFs) that makes DNA competent for replication in eukaryotes (10.6)

**Pribnow box** a DNA base sequence that is part of a prokaryotic promoter; it is located 10 bases before the transcription start site (11.2)

**primary structure** the order in which the amino acids in a protein are linked by peptide bonds (4.1)

**primer** in DNA replication, a short stretch of RNA hydrogen-bonded to the template DNA to which the growing DNA strand is bonded at the start of synthesis (10.3)

**primosome** the complex at the replication fork in DNA synthesis; it consists of the RNA primer, primase, and helicase (10.4)

**prion** a naturally occurring protein found in nervous tissue and brain that can adopt multiple forms; the abnormal form leads to prion diseases, such as mad-cow disease and

human spongiform encephalopathy (Creutzfeldt-Jakob disease) (4.6)

**probe** a radioactively labeled strand of a nucleic acid used for selecting a complementary strand out of a mixture (13.5)

**processivity** the number of nucleotides incorporated in a growing DNA chain before the DNA polymerase dissociates from the template DNA (10.3)

**prokaryote** a microorganism that lacks a distinct nucleus and membrane-enclosed organelles (1.4)

**promoter** the portion of DNA to which RNA polymerase binds at the start of transcription (11.2)

**proofreading** the process of removing incorrect nucleotides when DNA replication is in progress (10.5)

**propeller-twist** a twisting of the bases in a DNA double helix that allows for stronger base stacking (9.3)

**prostaglandin** one of a group of derivatives of arachidonic acid; it contains a five-membered ring and is of pharmaceutical importance (8.8)

**prosthetic group** a portion of a protein that does not consist of amino acids (4.1)

**protease** an enzyme that hydrolyzes proteins (7.5)

**proteasome** a multisubunit complex of proteins that mediates degradation of other, suitably tagged proteins (12.7)

**protein** a macromolecule formed by polymerization of amino acids (1.3)

**protein kinase** a class of enzymes that modifies a protein by attaching a phosphate group to it (7.3)

**proteome** the total protein content of the cell (13.9)

**proteomics** study of interactions among all the proteins of the cell (13.9)

**Protista** one of the five kingdoms used to classify living organisms; it includes single-celled eukaryotes (1.7)

**proton gradient** the difference between the hydrogen ion concentrations in the mitochondrial matrix and that in the intermembrane space, which is the basis of coupling between oxidation and phosphorylation (20.1)

**purine** a nitrogen-containing aromatic compound that contains a six-membered ring fused to a five-membered ring; the parent compound

of two nucleobases, adenine and guanine (9.2)

**pyranose** a cyclic form of a sugar containing a five-membered ring; it was named for its resemblance to pyran (16.1)

**pyranoside** a glycoside involving a pyranose (16.1)

**pyrimidine** a nitrogen-containing aromatic compound that contains a six-membered ring; the parent compound of several nucleobases (9.2)

**pyrrole ring** a five-membered ring that contains one nitrogen atom; part of the structure of porphyrins and heme (4.5)

**pyruvate dehydrogenase complex** a multienzyme complex that catalyzes the conversion of pyruvate to acetyl-CoA and carbon dioxide (19.3)

**pyruvate kinase** the enzyme that catalyzes the final step common to all forms of glycolysis (17.3)

**Q cycle** a series of reactions in the electron transport chain that provides the link between two-electron transfers and one-electron transfers (20.3)

**quaternary structure** the interaction of several polypeptide chains in a multisubunit protein (4.1)

**rate constant** a proportionality constant in the equation that describes the rate of a reaction (6.3)

**rate-limiting step** the slowest step in a reaction mechanism; it determines the maximum velocity of the reaction (6.3)

**reaction center** the site of the special pair of chlorophylls responsible for trapping light energy from the Sun (22.1)

**reaction order** an experimentally determined number that describes the rate of a reaction in terms of the concentration of a reactant or reactants (6.3)

**reading frame** the starting point for reading of a genetic message (12.2)

**receptor protein** a protein on a cell membrane with specific binding site for extracellular substances (8.4)

**recombinant DNA** DNA that has been produced by linking DNA from two different sources (13.3)

**reducing agent** a substance that gives up electrons to other substances (15.4)

**reducing sugar** a sugar that has a free carbonyl group, one that can react with an oxidizing agent (16.2)

**reduction** the gain of electrons (1.9)

**reduction potential** a standard voltage that indicates the tendency of a reduction half reaction to take place (20.2)

**regulatory gene** a gene that directs the synthesis of a repressor protein (11.3)

**repair** the enzymatic removal of incorrect nucleotides from DNA and their replacement by correct ones (10.5)

**replication** the process of duplication of DNA (10.1)

**replication activator protein (RAP)** the protein whose binding prepares for the start of DNA replication in eukaryotes (10.6)

**replication fork** in DNA replication, the point at which new DNA strands are formed (10.2)

**replication licensing factors (RLFs)** proteins required for DNA replication in eukaryotes (10.6)

**replicator** one of the multiple origins of replication in eukaryotic DNA synthesis (10.6)

**replicon** a part of a chromosome in which DNA synthesis is taking place (10.6)

**replisome** a complex of DNA polymerase, the RNA primer, primase, and helicase at the replication fork (10.4)

**repressor** a protein that binds to an operator gene, blocking the transcription and eventual translation of structural genes under the control of that operator (11.3)

**residue** the portion of a monomer unit included in a polymer after splitting out of water between the linked monomers (3.4)

**resonance structures** structural formulas that differ from each other only in the position of electrons (3.4)

**respiratory complexes** the multienzyme systems in the inner mitochondrial membrane that carry out the reactions of electron transport (20.3)

**response element** a DNA sequence that binds to transcription factors involved in more generalized control of pathways (11.3)

**restriction-fragment length**

**polymorphism (RFLP)** differences in the lengths of DNA fragments from different sources when digested with restriction enzymes; a forensic

technique using DNA to identify biological samples (13.7)

**restriction nuclease** an enzyme that catalyzes a double-strand hydrolysis of DNA at a defined point in a specific sequence (13.2)

**retinal** the aldehyde form of vitamin A (8.7)

**retrovirus** a virus in which the base sequence of RNA directs the synthesis of DNA (14.2)

**reverse transcriptase** the enzyme that directs the synthesis of DNA on an RNA template (14.2)

**reverse turn** a part of a protein where the polypeptide chain folds back on itself (4.3)

**reversible inhibitor** an inhibitor that is not covalently bound to an enzyme; it can be removed with restoration of activity (6.7)

**R group** the side chain of an amino acid that determines its identity (3.1)

**rho-dependent termination** the type of transcription termination that requires the rho protein (11.2)

**rhodopsin** a molecule crucial in vision; it is formed by the reaction of retinal and opsin (8.7)

**ribonucleoside** a compound formed when a nucleobase forms a glycosidic bond with ribose (9.2)

**ribose** a sugar that is part of the structure of RNA (9.2)

**ribosome** the site of protein synthesis in all organisms, consisting of RNA and protein (1.5)

**ribozyme** catalytic RNA (11.8)

**ribulose-1,5-bisphosphate** a key intermediate in the production of sugars in photosynthesis (22.5)

**RNA** ribonucleic acid (9.2)

**RNAi** RNA interference (9.5)

**RNA polymerase** the enzyme that catalyzes the production of RNA on a DNA template (11.2)

**RNA polymerase II** the RNA polymerase in eukaryotes that makes mRNA; also called *RNA polymerase B* (11.4)

**rRNA (ribosomal RNA)** the kind of RNA found in ribosomes (9.5)

**rubisco (ribulose-1,5-bisphosphate carboxylase/ oxygenase)** the enzyme that catalyzes the first step in carbon dioxide fixation in photosynthesis (22.5)

**salting out** a purification technique for proteins based on differential solubility in salt solutions (5.1)

**salvage reactions** reactions that reuse compounds, such as purines, that require a large amount of energy to produce (23.8)

**saponification** the reaction of a triacylglycerol with base to produce glycerol and three molecules of fatty acid (8.2)

**saturated** having all carbon-carbon bonds as single bonds (8.2)

**SDS-polyacrylamide-gel electrophoresis (SDS-PAGE)** an electrophoretic technique that separates proteins on the basis of size (5.3)

**secondary structure** the arrangement in space of the backbone atoms in a polypeptide chain (4.1)

**second messenger** a substance produced or released by a cell in response to hormone binding to a receptor on the cell surface; it elicits the actual response in the cell (24.3)

**semiconservative replication** the mode in which DNA reproduces itself, such that one strand comes from parent DNA and the other strand is newly formed (10.2)

**sense strand** the DNA strand that has the same sequence as the RNA that is synthesized from the template (11.2)

**sequencer** an automated instrument used in determining the amino acid sequence of a peptide or the nucleotide sequence of a nucleic acid (5.4)

**sequential model** a description of the action of allosteric proteins in which a conformational change in one subunit is passed along to the other subunits (7.2)

**serine protease** a proteolytic enzyme in which a serine hydroxyl plays an essential role in catalysis (7.5)

**severe combined immune deficiency (SCID)** a genetic disease that affects DNA synthesis in the cells of the immune system (14.2)

**Shine-Dalgarno sequence** a leader sequence in prokaryotic mRNA that precedes the start signal (12.4)

**side-chain group** the portion of an amino acid that determines its identity (3.1)

**sigmoidal** referring to an S-shaped curve on a graph, characteristic of cooperative interactions (4.5)

**silencer** a DNA sequence that binds to a transcription factor and reduces the level of transcription (11.3)

**simple diffusion** the process of passing through a pore or opening in a membrane without a requirement for a carrier or for the expenditure of energy (8.6)

**single-strand binding (SSB) protein** in DNA replication, a protein that protects exposed single-strand sections of DNA from nucleases (10.4)

**siRNA (small interfering RNA)** short stretches of RNA that control gene expression by selective suppression of genes (9.5)

**S<sub>N</sub>1** a unimolecular nucleophilic substitution reaction; one of the most common types of organic reactions seen in biochemistry; the rate of the reaction follows first-order kinetics (7.6)

**S<sub>N</sub>2** a bimolecular nucleophilic substitution reaction; an important type of organic reaction seen in biochemistry; the rate of the reaction follows second-order kinetics (7.6)

**snRNA (small nuclear RNA)** an RNA type that is found in eukaryotes and is involved in splicing and some regulation of transcription (9.5)

**snRNPs (small nuclear ribonucleoprotein particles)** a protein–RNA complex found in the nucleus that aids in processing RNA molecules for export to the cytosol (9.5)

**sodium–potassium ion pump** the export of sodium ion from the cell with simultaneous inflow of potassium ion, both against concentration gradients (8.6)

**Southern blotting** a technique used for transferring DNA from an agarose gel after electrophoresis onto a membrane, such as one made of nitrocellulose (13.7)

**spacer region** a region of eukaryotic DNA that is between nucleosomes (9.3)

**sphingolipid** a lipid whose structure is based on sphingosine (8.2)

**sphingosine** a long-chain amino alcohol; the basis of the structure of a number of lipids (8.2)

**spliceosome** a large multisubunit particle, similar in size to a ribosome, that is involved in splicing of RNA molecules (11.7)

**split gene** a gene that contains intervening sequences that are not present in the mature RNA (11.7)

**spontaneous** in thermodynamics, characteristic of a reaction or process that takes place without outside intervention (1.10)

**standard state** the standard set of conditions used for comparisons of chemical reactions (15.1)

**starch** a polymer of glucose that plays an energy-storage role in plants (16.4)

**stationary phase** in chromatography, the substance that selectively retards the flow of the sample, effecting the separation (5.2)

**steady state** the condition in which the concentration of an enzyme–substrate complex remains constant in spite of continuous turnover (6.6)

**stereochemistry** the branch of chemistry that deals with the three-dimensional shape of molecules (3.1)

**stereoisomers** molecules that differ from each other only in their configuration (three-dimensional shape); also called *optical isomers* (3.1)

**stereospecific** able to distinguish between stereoisomers (7.6)

**steroid** a lipid with a characteristic fused-ring structure (8.2)

**“sticky ends”** short, single-stranded stretches at the ends of double-stranded DNA; they can provide sites to which other DNA molecules with sticky ends can be linked (13.2)

(–) **strand** the DNA strand that is used as a template for RNA synthesis (11.2)

(+) **strand** the DNA strand that has the same sequence as the RNA that is synthesized from the template (11.2)

**stroma** in a chloroplast, the portion of the organelle that is equivalent to the mitochondrial matrix; the site of production of sugars in photosynthesis (22.1)

**structural gene** a gene that directs the synthesis of a protein under the control of some regulatory gene (11.3)

**substrate** a reactant in an enzyme-catalyzed reaction (6.4)

**substrate cycling** the control process in which opposing reactions are catalyzed by different enzymes (18.3)

**substrate-level phosphorylation** a reaction in which the source of phosphorus is inorganic phosphate ion, not ATP (17.3)

**subunits** the individual parts of a larger molecule (e.g., the individual polypeptide chains that make up a complete protein) (4.1)

**sugar–phosphate backbone** the series of ester bonds between phosphoric acid and deoxyribose (in DNA) or ribose (in RNA) (9.2)

**supercoiling** the presence of extra twists (over and above those of the double helix) in closed circular DNA (9.3)

**supersecondary structure** specific clusters of secondary structural motifs in proteins (4.3)

**suppressor tRNAs** tRNAs that allow translation to continue through a stop codon (12.5)

**TATA box** a promoter element found in eukaryotic transcription that is located 25 bases upstream of the transcription start site (11.4)

**TATA-box binding protein (TBP)** a part of one of the general transcription factors found in eukaryotic transcription; it binds to the TATA-box portion of the promoter (11.4)

**TATA-box binding protein associated factors (TAFs)** proteins that are associated with the TATA-box binding protein itself (11.4)

**T cell** one of two kinds of white blood cells important in the immune system—a killer T cell, which destroys infected cells, or a helper T cell, which is involved in the process of B-cell maturation (14.3)

**template (antisense) strand** the DNA strand that is used as a template for RNA synthesis (11.2)

**termination site** the area on DNA that causes termination of transcription by generating hairpin loops and a zone of weak binding between DNA and RNA (11.2)

**termination step** in protein synthesis, the point at which the stop signal is reached, releasing the newly formed protein from the ribosome (12.4)

**3•4 terminator** a hairpin loop that can form during transcription termination



and that causes premature release of the RNA transcript (11.3)

**tertiary structure** the arrangement in space of all the atoms in a protein (4.1)

**tetrahydrofolate** the metabolically active form of the vitamin folic acid; a carrier of one-carbon groups (23.4)

**tetramer** a molecule consisting of four subunits (4.5)

**thermodynamics** the study of transformations and transfer of energy (1.9)

**thiamine pyrophosphate** a coenzyme involved in the transfer of two-carbon units (17.4)

**thioester** a sulfur-containing analogue of an ester (19.3)

**thylakoid disks** the site of the light-trapping reaction in chloroplasts (22.1)

**thylakoid space** the portion of the chloroplast between the thylakoid disks (22.1)

**thymidylate synthetase** the enzyme that catalyzes the production of thymine nucleotides needed for DNA synthesis; a target for cancer chemotherapy (23.11)

**thymine** one of the pyrimidine bases found in nucleic acids (9.2)

**thymine dimers** a defect in DNA structure caused by the action of ultraviolet light (10.5)

**titration** an experiment in which a measured amount of base is added to an acid (2.4)

**$\alpha$ -tocopherol** the most active form of vitamin E (8.7)

**topoisomerase** an enzyme that relaxes supercoiling in closed circular DNA (9.3)

**torr** a unit of pressure equal to that exerted by a column of mercury 1 mm high at 0°C (4.5)

**transaldolase** an enzyme that transfers a two-carbon unit in reactions of sugars (18.4)

**transamination** the transfer of amino groups from one molecule to another; an important process in the anabolism and catabolism of amino acids (23.4)

**transcription** the process of formation of RNA on a DNA template (9.5)

**transcription-activation domain** the part of a transcription factor that interacts with other proteins and complexes rather than with the DNA directly (11.5)

**transcription bubble** the complex of separated DNA and RNA polymerase in which transcription is actively occurring (11.2)

**transcription-coupled repair** a type of DNA repair that occurs during transcription (11.4)

**transcription factor** a protein or other complex that binds to a DNA sequence and alters the basal level of transcription (11.3)

**transcription start site** the location on the template DNA strand where the first ribonucleotide is used to initiate RNA synthesis (11.2)

**transcriptome** a group of genes that are being transcribed at a given time (13.9)

**transition state** the intermediate stage in a reaction in which old bonds break and new bonds are formed (6.2)

**transition-state analogue** a synthesized compound that mimics the form of the transition state of an enzyme reaction (7.7)

**translation** the process of protein synthesis in which the amino acid sequence of the protein reflects the sequence of bases in the gene that codes for that protein (9.5)

**translocation** in protein synthesis, the motion of the ribosome along the mRNA as the genetic message is being read (12.4)

**transport protein** a component of a membrane that mediates the entry of specific substances into a cell (8.4)

**triacylglycerol (triglyceride)** a lipid formed by esterification of three fatty acids to glycerol (8.2)

**tricarboxylic acid cycle** another name for the citric acid cycle (19.1)

**trimer** a molecule consisting of three subunits (4.5)

**triosephosphate isomerase** the enzyme that catalyzes the conversion of dihydroxyacetone phosphate to glyceraldehyde-3-phosphate (17.2)

**triplet code** a sequence of three bases (a triplet) in mRNA that specifies one amino acid in a protein (12.2)

**tRNA (transfer RNA)** the kind of RNA to which amino acids are bonded as a preliminary step to being incorporated into a growing polypeptide chain (9.5)

**trophic hormones** hormones that are produced by the pituitary gland under the direction of the hypothalamus that,

in turn, cause the release of specific hormones by individual endocrine glands (24.3)

**trypsin** a proteolytic enzyme specific for basic amino acid residues as the site of hydrolysis (5.4)

**tumor suppressor** a gene that encodes a protein that inhibits cell division (14.4)

**turnover number** the number of moles of substrate that react per second per mole of enzyme (6.6)

**uncoupler** a substance that overcomes the proton gradient in mitochondria, allowing electron transport to proceed in the absence of phosphorylation (20.4)

**universal code** the genetic code, which is the same in all organisms (12.2)

**unsaturated** containing some carbon-carbon double or triple bonds (8.2)

**UP element** a prokaryotic promoter element that is 40 to 60 bases upstream of the transcription start site (13.2)

**upstream** in transcription, a portion of the sequence nearer the 3' end than the gene to be transcribed, where the DNA is read from the 3' to the 5' end and the RNA is formed from the 5' to the 3' end; in translation, nearer the 5' end of the mRNA (13.1)

**uracil** one of the pyrimidine bases found in nucleic acids (9.2)

**urea cycle** a pathway that leads to excretion of waste products of nitrogen metabolism, especially those of amino acids (23.6)

**uric acid** a product of catabolism of nitrogen-containing compounds, especially purines; accumulation of uric acid in joints causes gout in humans (23.8)

**vacuole** a cavity within the cytoplasm of a cell, typically enclosed by a single membrane, that may serve secretory, excretory, or storage functions (1.6)

**van der Waals bond** a noncovalent association based on the weak attraction of transient dipoles for one another; also called a *van der Waals interaction* (2.1)

**vector** a carrier molecule for transfer of genes in DNA recombination (13.3)

**virion** a complete virus particle consisting of nucleic acid and coat protein (14.1)

**V system** a combination of an allosteric enzyme and an inhibitor or activator in which the presence of the inhibitor/

activator changes the maximal velocity of the enzyme but not the substrate level that yields one-half  $V_{\max}$  (7.1)  
**“wobble”** the possible variation in the third base of a codon allowed by several acceptable forms of base pairing between mRNA and tRNA (12.2)  
**X-ray crystallography** an experimental method for determining the three-dimensional structure of molecules,

such as the tertiary or quaternary structure of proteins (4.4)  
**Z-DNA** a form of DNA that is a left-handed helix, which has been seen to occur naturally under certain circumstances (9.3)  
**zero order** refers to a reaction that proceeds at a constant rate, independent of the concentration of reactant (6.3)

**zinc-finger motif** a common motif found in the DNA-binding region of transcription factors (11.6)  
**zwitterion** a molecule that has both a positive and a negative charge (2.5)  
**zymogen** an inactive protein that can be activated by specific hydrolysis of peptide bonds (7.4)



# Answers to Questions

## Chapter 1

### 1.1 Basic Themes

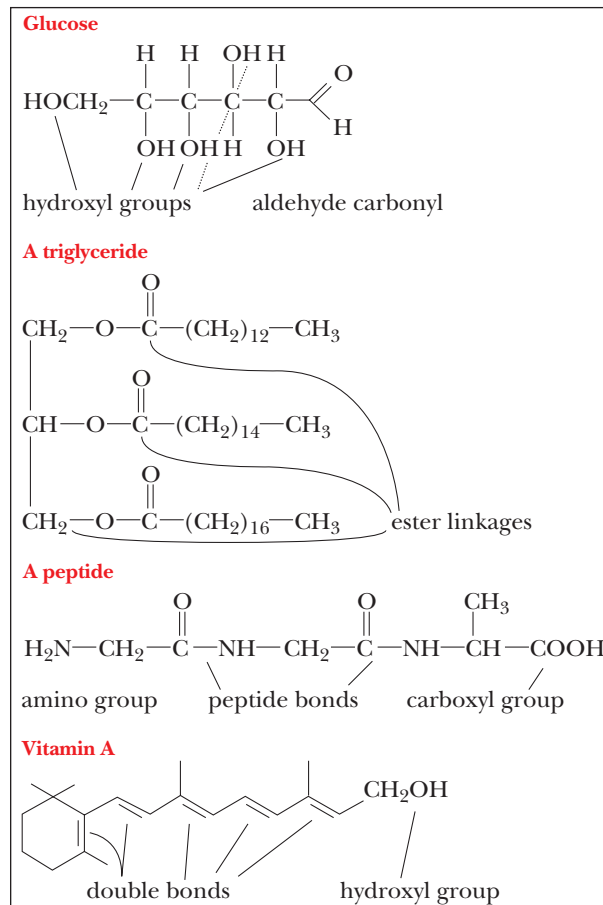
1. A polymer is a very large molecule formed by linking smaller units (monomers) together. A protein is a polymer formed by linking amino acids together. A nucleic acid is a polymer formed by linking nucleotides together. Catalysis is the process that increases the rate of chemical reactions compared with the rate of the uncatalyzed reaction. Biological catalysts are proteins in almost all cases; the only exceptions are a few types of RNA, which can catalyze some of the reactions of their own metabolism. The genetic code is the means by which the information for the structure and function of all living things is passed from one generation to the next. The sequence of purines and pyrimidines in DNA carries the genetic code. (RNA is the coding material in some viruses.)

### 1.2 Chemical Foundations of Biochemistry

2. The correct match of functional groups and the compounds containing those functional groups is given in the following list.

Amino group	$\text{CH}_3\text{CH}_2\text{NH}_2$
Carbonyl group (ketone)	$\text{CH}_3\text{COCH}_3$
Hydroxyl group	$\text{CH}_3\text{OH}$
Carbonyl group	$\text{CH}_3\text{COOH}$
Carbonyl group (aldehyde)	$\text{CH}_3\text{CH}_2\text{CHO}$
Thiol group	$\text{CH}_3\text{SH}$
Ester linkage	$\text{CH}_3\text{COOCH}_2\text{CH}_3$
Double bond	$\text{CH}_3\text{CH}=\text{CHCH}_3$
Amide linkage	$\text{CH}_3\text{CON}(\text{CH}_3)_2$
Ether	$\text{CH}_3\text{CH}_2\text{OCH}_2\text{CH}_3$

3. The functional groups in the compounds follow:



4. Before 1828, the concept of vitalism held that organic compounds could be made only by living systems and were beyond the realm of laboratory investigations. Wöhler's synthesis showed that organic compounds, like inorganic ones, did not require a vitalistic explanation, but that, rather, they obeyed the laws of chemistry and physics and thus were subject to laboratory investigation. Subsequently, the concept was extended to the much more complex, but still testable, discipline of biochemistry.
5. Urea, like all organic compounds, has the same molecular structure, whether it is produced by a living organism or not.

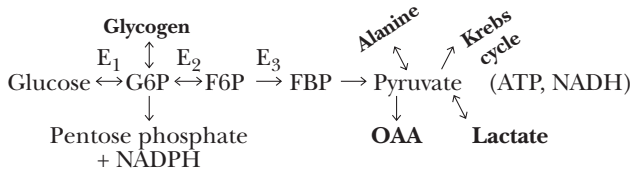
6.

Item	Organic	Biochemical
Solvent	Varies (smelly)	Water (usually)
Concentrations	High	Low (mM, $\mu$ M, nM)
Use catalyst?	Usually not	Almost always (enzymes)
Speed	Min, hr, day	$\mu$ sec, nsec
Temp	Varies (high)	Isothermal, ambient temp
Yield	Poor–good (90%)	High (can be 100%)
Side reactions	Often*	No
Internal control	Little	Very high**—choices
Polymers (product)	Usually not	Commonly (proteins, nucleic acids, saccharides)
Bond strength	High (covalent)	High, weak (in polymers)
Bond distances	Not critical	Critical (close fit)
Compartmented	No	Yes (esp. eukaryotes)
Emphasis	One reaction	Pathways, interconnected (control** choices) <sup>†</sup>
System	Closed or open	Open (overcome + $\Delta G$ )

\* Example of side reactions: Glucose  $\rightarrow$  G6P or G1P or G2P.

\*\* Control levels: enzyme, hormone, gene.

<sup>†</sup> Example of choices:



**1.3 The Beginnings of Biology: Origin of Life**

- 7. It is generally believed that carbon is the likely basis for all life forms, terrestrial or extraterrestrial.
- 8. Eighteen residues would give  $20^{18}$ , or  $2.6 \times 10^{23}$  possibilities. Thus, 19 residues would be necessary to have at least Avogadro's number ( $6.022 \times 10^{23}$ ) of possibilities.
- 9. The number is  $4^{40}$ , or  $1.2 \times 10^{24}$ , which is twice Avogadro's number.
- 10. RNA is capable of both coding and catalysis.
- 11. Catalysis allows living organisms to carry out chemical reactions much more efficiently than without catalysts.
- 12. Two of the most obvious advantages are speed and specificity; they also work at constant temperature or produce little heat.
- 13. Coding allows for reproduction of cells.
- 14. With respect to coding, RNA-type polynucleotides have been produced from monomers in the absence of either a preexisting RNA to be copied or an enzyme to catalyze the process. The observation that some existing RNA molecules can catalyze their own processing suggests a role for RNA in catalysis. With this dual role, RNA may have been the original informational macromolecule in the origin of life.
- 15. It is unlikely that cells could have arisen as bare cytoplasm without a plasma membrane. The presence of the membrane protects cellular components from the environment and prevents them from diffusing away from each other. The molecules within a cell can react more easily if they are closer to each other.

**1.4 The Biggest Biological Distinction—Prokaryotes and Eukaryotes**

- 16. Five differences between prokaryotes and eukaryotes are as follows: (1) Prokaryotes do not have a well-defined nucleus, but eukaryotes have a nucleus marked off from the rest of the cell by a double membrane. (2) Prokaryotes have only a plasma (cell) membrane; eukaryotes have an extensive internal membrane system. (3) Eukaryotic cells contain membrane-bounded organelles, while prokaryotic cells do not. (4) Eukaryotic cells are normally larger than those of prokaryotes. (5) Prokaryotes are single celled organisms, while eukaryotes can be either single-celled or multicellular.
- 17. Protein synthesis takes place on ribosomes both in prokaryotes and in eukaryotes. In eukaryotes, ribosomes may be bound to the endoplasmic reticulum or found free in the cytoplasm; in prokaryotes, ribosomes are only found free in the cytoplasm.

**1.5 Prokaryotic Cells**

- 18. It is unlikely that mitochondria would be found in bacteria. These eukaryotic organelles are enclosed by a double membrane, and bacteria do not have an internal membrane system. The mitochondria found in eukaryotic cells are about the same size as most bacteria.

**1.6 Eukaryotic Cells**

- 19. See Section 1.6 for the functions of the parts of an animal cell, which are shown in Figure 1.10a.
- 20. See Section 1.6 for the functions of the parts of a plant cell, which are shown in Figure 1.10b.
- 21. In green plants photosynthesis takes place in the membrane system of chloroplasts, which are large membrane-enclosed organelles. Photosynthetic bacteria have extensions of the plasma membrane into the interior of the cell called chromatophores, which are the sites of photosynthesis.
- 22. Nuclei, mitochondria, and chloroplasts are all enclosed by a double membrane.
- 23. Nuclei, mitochondria, and chloroplasts all contain DNA. The DNA found in mitochondria and in chloroplasts differs from that found in the nucleus.
- 24. Mitochondria carry out a high percentage of the oxidation (energy-releasing) reactions of the cell. They are the primary sites of ATP synthesis.
- 25. The Golgi apparatus is involved in binding carbohydrates to proteins and in exporting substances from the cell. Lysosomes contain hydrolytic enzymes, peroxisomes contain catalase (needed for the metabolism of peroxides), and glyoxysomes contain enzymes needed by plants for the glyoxylate cycle. All of these organelles have the appearance of flattened sacs, and each is enclosed by a single membrane.

**1.7 Five Kingdoms, Three Domains**

- 26. Monera includes bacteria (e.g., *E. coli*) and cyanobacteria. Protista includes such organisms as *Euglena*, *Volvox*, *Amoeba*, and *Paramecium*. Fungi includes molds and mushrooms. Plantae includes club mosses and oak trees. Animalia includes spiders, earthworms, salmon, rattlesnakes, robins, and dogs.
- 27. The kingdom Monera consists of prokaryotes. Each of the other four kingdoms consists of eukaryotes.
- 28. The kingdom Monera is divided into the domains Eubacteria and Archaea on the basis of biochemical differences. The domain Eukarya consists of the four kingdoms of eukaryotic organisms.

**1.8 Common Ground for All Cells**

- 29. The major advantage of being eukaryotic is that of having compartments (organelles) with specialized functions (and thus

division of labor). Another advantage is that cells can be much larger without surface area-to-volume considerations being critical because of compartmentalization.

30. See the discussion of the endosymbiotic theory in Section 1.8.
31. See Question 30. The division of labor in cells gives rise to greater efficiency and a larger number of individuals. This in turn allows more opportunity for evolution and speciation.

### 1.9 Biochemical Energetics

32. Processes that release energy are favored.

### 1.10 Energy and Change

33. The term *spontaneous* means energetically favored. It does not necessarily mean fast.

### 1.11 Spontaneity in Biochemical Reactions

34. The system consists of the nonpolar solute and water, which become more disordered when a solution is formed;  $\Delta S_{\text{sys}}$  is positive but comparatively small.  $\Delta S_{\text{surr}}$  is negative and comparatively large because it is a reflection of the unfavorable enthalpy change for forming the solution ( $\Delta H_{\text{sys}}$ ).
35. Processes (a) and (b) are spontaneous, whereas processes (c) and (d) are not. The spontaneous processes represent an increase in disorder (increase in the entropy of the Universe) and have a negative  $\Delta G^\circ$  at constant temperature and pressure, while the opposite is true of the nonspontaneous processes.
36. In all cases, there is an increase in entropy, and the final state has more possible random arrangements than the initial state.
37. Since the equation involves multiplication of  $\Delta S$  by  $T$ , the value of  $\Delta G$  is temperature-dependent.
38. If one considers entropy a measure of dispersion of energy, then at higher temperatures, it is logical that molecules would have more possible arrangements due to increased molecular motion.
39. Assuming the value of  $\Delta S$  is positive, an increase in temperature increases the  $-\Delta G$  contribution of the entropy component to the overall energy change.
40. The heat exchange, getting colder, reflects only the enthalpy or  $\Delta H$  component of the total energy change. The entropy change must be high enough to offset the enthalpy component and to add up to an overall  $-\Delta G$ .
41. Entropy would increase. Two molecules, ADP and  $P_i$ , can be randomized in more ways than a single molecule, ATP, can.

### 1.12 Life and Thermodynamics

42. The lowering of entropy needed to give rise to organelles leads to higher entropy in the surroundings, thus increasing the entropy of the Universe as a whole.
43. Compartmentalization in organelles brings components of reactions into proximity with one another. The energy change of the reaction is not affected, but the availability of components allows it to proceed more readily.
44. DNA would have higher entropy with the strands separated. There are two single strands instead of one double strand, and the single strands have more conformational mobility.
45. See the answer to Question 41. It is still unlikely that cells could have arisen as bare cytoplasm, but the question of proximity of reactants is more to the point here than the energy change of a given reaction.
46. It would be unlikely that cells of the kind we know would have evolved on a gas giant. The lack of solids and liquids on which aggregates could form would make a large difference.

47. The available materials differ from those that would have been found on Earth, and conditions of temperature and pressure are very different.
48. Mars, because of conditions more like those on Earth.
49. A number of energetically favorable interactions drive the process of protein folding, ultimately increasing the entropy of the Universe.
50. Photosynthesis is endergonic, requiring light energy from the Sun. The complete aerobic oxidation of glucose is exergonic and is a source of energy for many organisms, including humans. It would be reasonable to expect the two processes to take place differently in order to provide energy for the endergonic one.

## Chapter 2

### 2.1 Water and Polarity

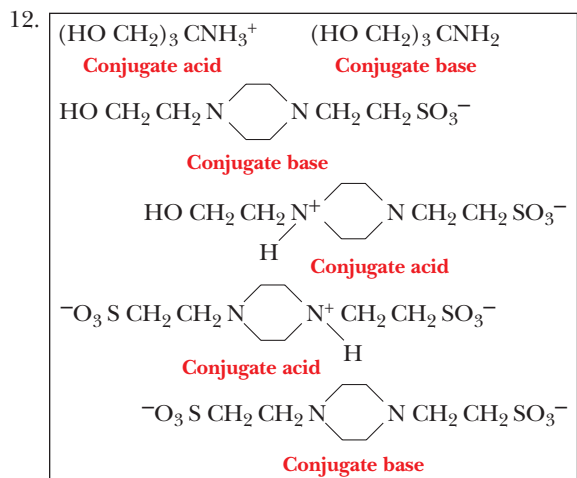
1. The unique fitness of water for forming hydrogen bonds determines the properties of many important biomolecules. Water can also act as an acid and as a base, giving it great versatility in biochemical reactions.
2. If atoms did not differ in electronegativity, there would be no polar bonds. This would drastically affect all reactions that involve functional groups containing oxygen or nitrogen—that is, most biochemical reactions.

### 2.2 Hydrogen Bonds

3. Proteins and nucleic acids have hydrogen bonds as an important part of their structures.
4. Replication of DNA and its transcription to RNA requires hydrogen bonding of complementary bases to the DNA template strand.
5. The C—H bond is not sufficiently polar for greatly unequal distribution of electrons at its two ends. Also, there are no unshared pairs of electrons to serve as hydrogen bond acceptors.
6. Many molecules can form hydrogen bonds. Examples might be  $H_2O$ ,  $CH_3OH$ , or  $NH_3$ .
7. For a bond to be called a hydrogen bond, it must have a hydrogen covalently bonded to O, N, or F. This hydrogen then forms a hydrogen bond with another O, N, or F.
8. In a hydrogen-bonded dimer of acetic acid, the —OH portion of the carboxyl group on molecule 1 is hydrogen-bonded to the —C=O portion of the carboxyl group on molecule 2, and vice versa.
9. Glucose = 17, sorbitol = 18, ribitol = 15; each alcohol group can bond to three water molecules and the ring oxygen binds to two. The sugar alcohols bind more than the corresponding sugars.
10. Positively charged ions bind to nucleic acids as a result of electrostatic attraction to the negatively charged phosphate groups.

### 2.3 Acids, Bases, and pH

11. (a)  $(CH_3)_3NH^+$  (conjugate acid),  $(CH_3)_3N$  (conjugate base)  
(b)  $^+H_3N-CH_2-COOH$  (conjugate acid),  $^+H_3N-CH_2-COO^-$  (conjugate base)  
(c)  $^+H_3N-CH_2-COO^-$  (conjugate acid),  $H_2N-CH_2-COO^-$  (conjugate base)  
(d)  $^-OOC-CH_2-COOH$  (conjugate acid),  $^-OOC-CH_2-COO^-$  (conjugate base)  
(e)  $^-OOC-CH_2-COOH$  (conjugate base),  $HOOC-CH_2-COOH$  (conjugate acid)



13. Aspirin is electrically neutral at the pH of the stomach and can pass through the membrane more easily than in the small intestine.
14. The definition of pH is  $-\log[H^+]$ . By definition of the log function, a change in concentration of 10 leads to a change in pH of 1. The log of 10 is 1, the log of 100 is 2, etc.
15. Blood plasma, pH 7.4  $[H^+] = 4.0 \times 10^{-8} \text{ M}$   
 Orange juice, pH 3.5  $[H^+] = 3.2 \times 10^{-4} \text{ M}$   
 Human urine, pH 6.2  $[H^+] = 6.3 \times 10^{-7} \text{ M}$   
 Household ammonia, pH 11.5  $[H^+] = 3.2 \times 10^{-12} \text{ M}$   
 Gastric juice, pH 1.8  $[H^+] = 1.6 \times 10^{-2} \text{ M}$
16. Saliva, pH 6.5  $[H^+] = 3.2 \times 10^{-7} \text{ M}$   
 Intracellular fluid (liver), pH 6.9  $[H^+] = 1.6 \times 10^{-7} \text{ M}$   
 Tomato juice, pH 4.3  $[H^+] = 5.0 \times 10^{-5} \text{ M}$   
 Grapefruit juice, pH 3.2  $[H^+] = 6.3 \times 10^{-4} \text{ M}$
17. Saliva, pH 6.5  $[OH^-] = 3.2 \times 10^{-8} \text{ M}$   
 Intracellular fluid (liver), pH 6.9  $[OH^-] = 7.9 \times 10^{-8} \text{ M}$   
 Tomato juice, pH 4.3  $[OH^-] = 2.0 \times 10^{-10} \text{ M}$   
 Grapefruit juice, pH 3.2  $[OH^-] = 1.6 \times 10^{-11} \text{ M}$

#### 2.4 Titration Curves

18. (a) The numerical constant equal to the concentration of the products of the dissociation divided by the concentration of the undissociated acid form:  $([H^+][A^-])/[HA]$ .
- (b) The qualitative or quantitative description of how much acid (HA) dissociates to hydrogen ion.
- (c) The property of a molecule that has both a polar region and a nonpolar region.
- (d) The amount of acid or base that can be added to a buffer before experiencing a sharp pH change.
- (e) The point in a titration curve at which the added acid or base equals the amount of buffer originally present.
- (f) The property of a molecule that is readily soluble in water (i.e., water-loving).
- (g) The property of a molecule that is insoluble in water (i.e., water-hating).
- (h) The property of a molecule that is not soluble in water. The property of a covalent bond in which there is even sharing of electrons and no dipole moments (partial charges).
- (i) The property of a molecule that is soluble in water. The property of a covalent bond in which the electrons are not shared evenly and dipole moments (partial charges) exist.
- (j) An experiment in which acid or base is added stepwise to a solution of a compound and the pH is measured as a function of the added substance.

19. To get a titration curve most like the one in Figure 2.15, we have to titrate a compound with a  $pK_a$  as close as possible to that of  $H_2PO_4^-$ . According to Table 2.8, MOPS has a  $pK_a$  of 7.2, which is the closest value.

20. The titration curve for TRIS would be shifted to the right compared to that of phosphate. The crossover point would be at pH 8.3, rather than pH 7.2.

#### 2.5 Buffers

21. The  $pK$  of the buffer should be close to the desired buffer pH, and the substance chosen should not interfere with the reaction being studied.
22. The useful pH range of a buffer is one pH unit above and below its  $pK_a$ .
23. Use the Henderson-Hasselbalch equation:

$$pH = pK_a + \log\left(\frac{[CH_3COO^-]}{[CH_3COOH]}\right)$$

$$5.00 = 4.76 + \log\left(\frac{[CH_3COO^-]}{[CH_3COOH]}\right)$$

$$0.24 = \log\left(\frac{[CH_3COO^-]}{[CH_3COOH]}\right)$$

$$\frac{[CH_3COO^-]}{[CH_3COOH]} = \text{inverse log of } 0.24 = \frac{1.7}{1}$$

24. Use the Henderson-Hasselbalch equation:

$$pH = pK_a + \log\left(\frac{[CH_3COO^-]}{[CH_3COOH]}\right)$$

$$4.00 = 4.76 + \log\left(\frac{[CH_3COO^-]}{[CH_3COOH]}\right)$$

$$-.076 = \log\left(\frac{[CH_3COO^-]}{[CH_3COOH]}\right)$$

$$\frac{[CH_3COO^-]}{[CH_3COOH]} = \text{inverse log of } -0.76 = \frac{0.17}{1}$$

25. Use the Henderson-Hasselbalch equation:

$$pH = pK_a + \log\left(\frac{[TRIS]}{[TRIS-H^+]}\right)$$

$$8.7 = 8.3 + \log\left(\frac{[TRIS]}{[TRIS-H^+]}\right)$$

$$0.4 = \log\left(\frac{[TRIS]}{[TRIS-H^+]}\right)$$

$$\frac{[TRIS]}{[TRIS-H^+]} = \text{inverse log of } 0.4 = \frac{2.5}{1}$$

26. Use the Henderson-Hasselbalch equation:

$$pH = pK_a + \log\left(\frac{[HEPES]}{[HEPES-H^+]}\right)$$

$$7.9 = 7.55 + \log\left(\frac{[HEPES]}{[HEPES-H^+]}\right)$$

$$0.35 = \log\left(\frac{[HEPES]}{[HEPES-H^+]}\right)$$

$$\frac{[HEPES]}{[HEPES-H^+]} = \text{inverse log of } 0.35 = \frac{2.2}{1}$$

27. At pH 7.5, the ratio of  $[HPO_4^{2-}]/[H_2PO_4^-]$  is 2/1 ( $pK_a$  of  $H_2PO_4^- = 7.2$ ), as calculated using the Henderson-Hasselbalch equation.  $K_2HPO_4$  is a source of the base form, and HCl must be added to convert one-third of it to the acid form, according to



the 2/1 base/acid ratio. Weigh out 8.7 g of  $\text{K}_2\text{HPO}_4$  (0.05 mol, based on a formula weight of 174 g/mol), dissolve it in a small quantity of distilled water, add 16.7 mL of 1 M HCl (gives 1/3 of 0.05 mol of hydrogen ion, which converts 1/3 of the 0.05 mol of  $\text{HPO}_4^{2-}$  to  $\text{H}_2\text{PO}_4^-$ ), and dilute the resulting mixture to 1 L.

28. A 2/1 ratio of the base form to acid form is still needed, because the pH of the buffer is the same in both problems.  $\text{NaH}_2\text{PO}_4$  is a source of the acid form, and NaOH must be added to convert two-thirds of it to the base form. Weigh out 6.0 g of  $\text{NaH}_2\text{PO}_4$  (0.05 mol, based on a formula weight of 120 g/mol), dissolve it in a small quantity of distilled water, add 33.3 mL of 1 M NaOH (gives 2/3 of 0.05 mol of hydroxide ion, which converts 2/3 of the 0.05 mol of  $\text{H}_2\text{PO}_4^-$  to  $\text{HPO}_4^{2-}$ ), and dilute the resulting mixture to 1 L.

29. After mixing, the buffer solution (100 mL) contains 0.75 M lactic acid and 0.25 M sodium lactate. The  $\text{p}K_a$  of lactic acid is 3.86. Use the Henderson–Hasselbalch equation:

$$\begin{aligned}\text{pH} &= \text{p}K_a + \log\left(\frac{[\text{CH}_3\text{CHOHCOO}^-]}{[\text{CH}_3\text{CHOHCOOH}]}\right) \\ \text{pH} &= 3.86 + \log\left(\frac{[\text{CH}_3\text{CHOHCOO}^-]}{[\text{CH}_3\text{CHOHCOOH}]}\right) \\ \text{pH} &= 3.86 + \log(0.25 \text{ M}/0.75 \text{ M}) \\ \text{pH} &= 3.86 + (-0.48) \\ \text{pH} &= 3.38\end{aligned}$$

30. After mixing, the buffer solution (100 mL) contains 0.25 M lactic acid and 0.75 M sodium lactate. The  $\text{p}K_a$  of lactic acid is 3.86. Use the Henderson–Hasselbalch equation:

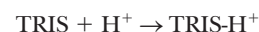
$$\begin{aligned}\text{pH} &= \text{p}K_a + \log\left(\frac{[\text{CH}_3\text{CHOHCOO}^-]}{[\text{CH}_3\text{CHOHCOOH}]}\right) \\ \text{pH} &= 3.86 + \log\left(\frac{[\text{CH}_3\text{CHOHCOO}^-]}{[\text{CH}_3\text{CHOHCOOH}]}\right) \\ \text{pH} &= 3.86 + \log(0.75 \text{ M}/0.25 \text{ M}) \\ \text{pH} &= 3.86 + (0.48) \\ \text{pH} &= 4.34\end{aligned}$$

31. Use the Henderson–Hasselbalch equation:

$$\begin{aligned}\text{pH} &= \text{p}K_a + \log\left(\frac{[\text{CH}_3\text{COO}^-]}{[\text{CH}_3\text{COOH}]}\right) \\ \text{pH} &= 4.76 + \log\left(\frac{[\text{CH}_3\text{COOH}^-]}{[\text{CH}_3\text{COOH}]}\right) \\ \text{pH} &= 4.76 + \log(0.25 \text{ M}/0.10 \text{ M}) \\ \text{pH} &= 4.76 + 0.40 \\ \text{pH} &= 5.16\end{aligned}$$

32. Yes, it is correct; calculate the molar amounts of the two forms and insert into the Henderson–Hasselbalch equation. (2.02 g = 0.0167 mol and 5.25 g = 0.0333 mol.)
33. The solution is a buffer because it contains equal concentrations of TRIS in the acid and free amine forms. When the two solutions are mixed, the concentrations of the resulting solution (in the absence of reaction) are 0.05 M HCl and 0.1 M TRIS because of dilution. The HCl reacts with half the TRIS present, giving 0.05 M TRIS (protonated form) and 0.05 M TRIS (free amine form).
34. Any buffer that has equal concentrations of the acid and basic forms has a pH equal to its  $\text{p}K_a$ . Therefore, the buffer from Question 33 has a pH of 8.3.
35. First calculate the moles of buffer that you have: 100 mL = 0.1 L, and 0.1 L of 0.1 M TRIS buffer is 0.01 mol. Since the buffer is at its  $\text{p}K_a$ , there are equal concentrations of the acid and basic form, so the amount of TRIS is 0.005 mol, and the amount of

TRIS- $\text{H}^+$  is 0.005 mol. If you then add 3 mL of 1 M HCl, you will be adding 0.003 mol of  $\text{H}^+$ . This reacts as shown:



until you run out of something, which will be the  $\text{H}^+$ , since it is the limiting reagent. The new amounts can be calculated as shown:

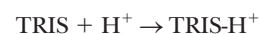
$$\begin{aligned}\text{TRIS-H}^+ &= 0.005 \text{ mol} + 0.003 \text{ mol} = 0.008 \text{ mol} \\ \text{TRIS} &= 0.005 \text{ mol} - 0.003 \text{ mol} = 0.002 \text{ mol}\end{aligned}$$

Now plug these values into the Henderson–Hasselbalch equation:

$$\text{pH} = 8.3 + \log([\text{TRIS}]/[\text{TRIS-H}^+]) = 8.3 + \log(0.002/0.008)$$

$$\text{pH} = 7.70$$

36. First calculate the moles of buffer that you have (we are going to do some rounding off): 100 mL = 0.1 L, and 0.1 L of 0.1 M TRIS buffer is 0.01 mol. Since the buffer is at pH 7.70, we saw in Question 25 that the amount of TRIS is 0.002 mol, and the amount of TRIS- $\text{H}^+$  is 0.008 mol. If you then add 1 mL of 1 M HCl, you will be adding 0.001 mol of  $\text{H}^+$ . This reacts as shown:



until you run out of something, which will be the TRIS, since it is the limiting reagent. All the TRIS is converted to TRIS- $\text{H}^+$ :

$$\begin{aligned}\text{TRIS-H}^+ &= 0.01 \text{ mol} \\ \text{TRIS} &= \sim 0 \text{ mol}\end{aligned}$$

We have used up the buffer capacity of the TRIS. We now have 0.001 mol of  $\text{H}^+$  in approximately 0.1 L of solution. This is approximately 0.01 M  $\text{H}^+$ .

$$\begin{aligned}\text{pH} &= -\log 0.01 \\ \text{pH} &= 2.0\end{aligned}$$

37.  $[\text{H}^+] = [\text{A}^-]$  for pure acid, thus  $K_a = [\text{H}^+]^2/[\text{HA}]$
- $$[\text{H}^+]^2 = K_a \times [\text{HA}] \quad -2 \log [\text{H}^+] = \text{p}K_a - \log [\text{HA}]$$
- $$\text{pH} = \frac{1}{2}(\text{p}K_a - \log [\text{HA}])$$

38. Use the Henderson–Hasselbalch equation:

$$[\text{Acetate ion}]/[\text{acetic acid}] = 2.3/1$$

39. A substance with a  $\text{p}K_a$  of 3.9 has a buffer range of 2.9 to 4.9. It does not buffer effectively at pH 7.5.
40. Use the Henderson–Hasselbalch equation. The ratio of  $[\text{A}^-]/[\text{HA}]$  would be 3981 to 1.
41. In all cases, the suitable buffer range covers a pH range of  $\text{p}K_a \pm 1$  pH units.
- (a) Lactic acid ( $\text{p}K_a = 3.86$ ) and its sodium salt: pH 2.86–4.86.
- (b) Acetic acid ( $\text{p}K_a = 4.76$ ) and its sodium salt: pH 3.76–5.76.
- (c) TRIS (see Table 3.4,  $\text{p}K_a = 8.3$ ) in its protonated form and its free amine form: pH 7.3–9.3.
- (d) HEPES (see Table 3.4,  $\text{p}K_a = 7.55$ ) in its zwitterionic form and its anionic form: pH 6.55–8.55.
42. Several of the buffers would be suitable, namely TES, HEPES, MOPS, and PIPES; but the best buffer would be MOPS, because its  $\text{p}K_a$  of 7.2 is closest to the desired pH of 7.3.
43. Buffer concentrations are typically reported to be the sum of the two ionic forms.



44. At the equivalence point of the titration, a small amount of acetic acid remains because of the equilibrium  $\text{CH}_3\text{COOH} \rightarrow \text{H}^+ + \text{CH}_3\text{COO}^-$ . There is a small, but nonzero, amount of acetic acid left.
45. Buffering capacity is based on the amounts of the acid and base forms present in the buffer solution. A solution with a high buffering capacity can react with a large amount of added acid or base without drastic changes in pH. A solution with a low buffering capacity can react with only comparatively small amounts of acid or base before showing changes in pH. The more concentrated the buffer, the higher is its buffering capacity. Buffer (a) has one-tenth the buffering capacity of buffer (b), which in turn has one-tenth the buffering capacity of buffer (c). All three buffers have the same pH, because they all have the same relative amounts of the acid and base form.
46. It would be more effective to start with the HEPES base. You want a buffer at a pH above the  $\text{pK}_a$ , which means that the base form predominates when you have finished preparing it. It is easier to convert some of the base form to the acid form than most of the acid form to the base form.
47. In a buffer with the pH above the  $\text{pK}_a$ , the base form predominates. This would be useful as a buffer for a reaction that produces  $\text{H}^+$  because plenty of the base form will be available to react with the hydrogen ion produced.
48. Zwitterions tend not to interfere with biochemical reactions.
49. It is useful to have a buffer that maintains a stable pH even if assay conditions change. Dilution is one such possible change.
50. It is useful to have a buffer that maintains a stable pH even if assay conditions change. Temperature variation is one such possible change.
51. The only zwitterion is  $^+\text{H}_3\text{N}-\text{CH}_2-\text{COO}^-$ .
52. Hypoventilation decreases the pH of blood.

## Chapter 3

### 3.1 Amino Acids Exist in a Three-Dimensional World

- D- and L-amino acids have different stereochemistry around the  $\alpha$ -carbon. Peptides that contain D-amino acids are found in bacterial cell walls and in some antibiotics.

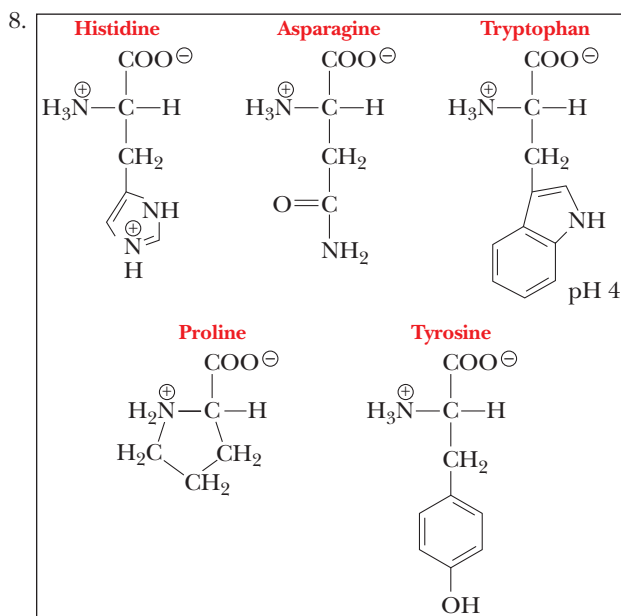
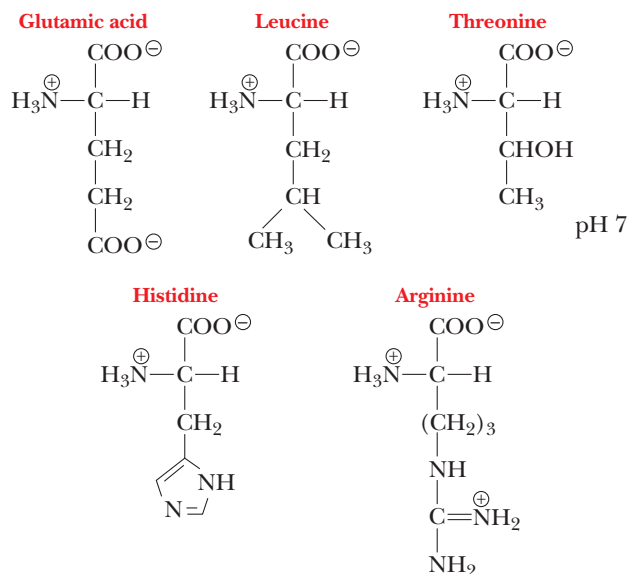
### 3.2 Individual Amino Acids: Their Structures and Properties

- Proline is technically not an amino acid. Glycine contains no chiral carbon atoms.
- Listed here are amino acids in which the R group contains the following: a hydroxyl group (serine, threonine, or tyrosine); a sulfur atom (cysteine or methionine); a second chiral carbon atom (isoleucine or threonine); an amino group (lysine); an amide group (asparagine or glutamine); an acid group (aspartate or glutamate); an aromatic ring (phenylalanine, tyrosine, or tryptophan); a branched side chain (leucine or valine).
- In the peptide Val—Met—Ser—Ile—Phe—Arg—Cys—Tyr—Leu, the polar amino acids are Ser, Arg, Cys, and Tyr; the aromatic amino acids are Phe and Tyr; and the sulfur-containing amino acids are Met and Cys.
- In the peptide Glu—Thr—Val—Asp—Ile—Ser—Ala, the non-polar amino acids are Val, Ile, and Ala; the acidic amino acids are Glu and Asp.
- Amino acids other than the usual 20 are produced by modification of one of the common amino acids. See Figure 3.5 for the

structures of some modified amino acids. Hydroxyproline and hydroxylysine are found in collagen; thyroxine is found in thyroglobulin.

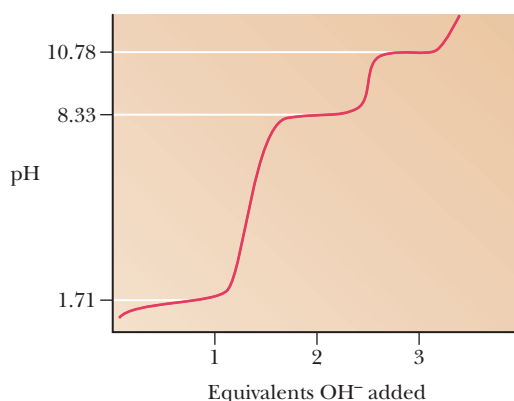
### 3.3 Amino Acids Can Act as Both Acids and Bases

- The ionized forms of each of the following amino acids at pH 7—glutamic acid, leucine, threonine, histidine, and arginine—are as follows:

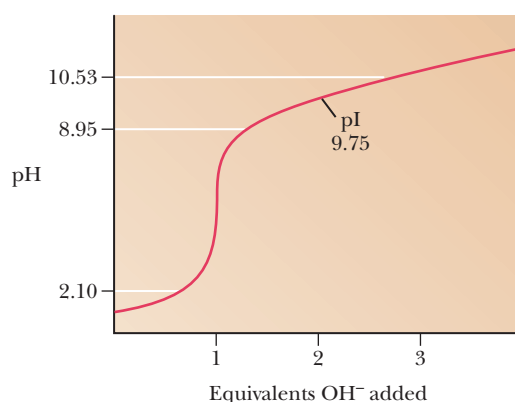


- Histidine: imidazole is deprotonated,  $\alpha$ -amino group is predominantly deprotonated. Asparagine:  $\alpha$ -amino group is deprotonated. Tryptophan:  $\alpha$ -amino group is predominantly deprotonated. Proline:  $\alpha$ -amino group is partially deprotonated. Tyrosine:  $\alpha$ -amino group is predominantly deprotonated, phenolic hydroxyl is approximately a 50–50 mixture of protonated and deprotonated forms.
- Glutamic acid, 3.25; serine, 5.7; histidine, 7.58; lysine, 9.75; tyrosine, 5.65; arginine, 10.75.

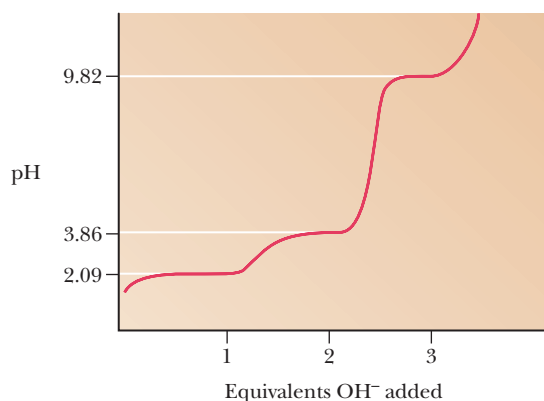
11. Cysteine has no net charge at  $\text{pH } 5.02 = (1.71 + 8.33)/2$  (see titration curve below).



12.

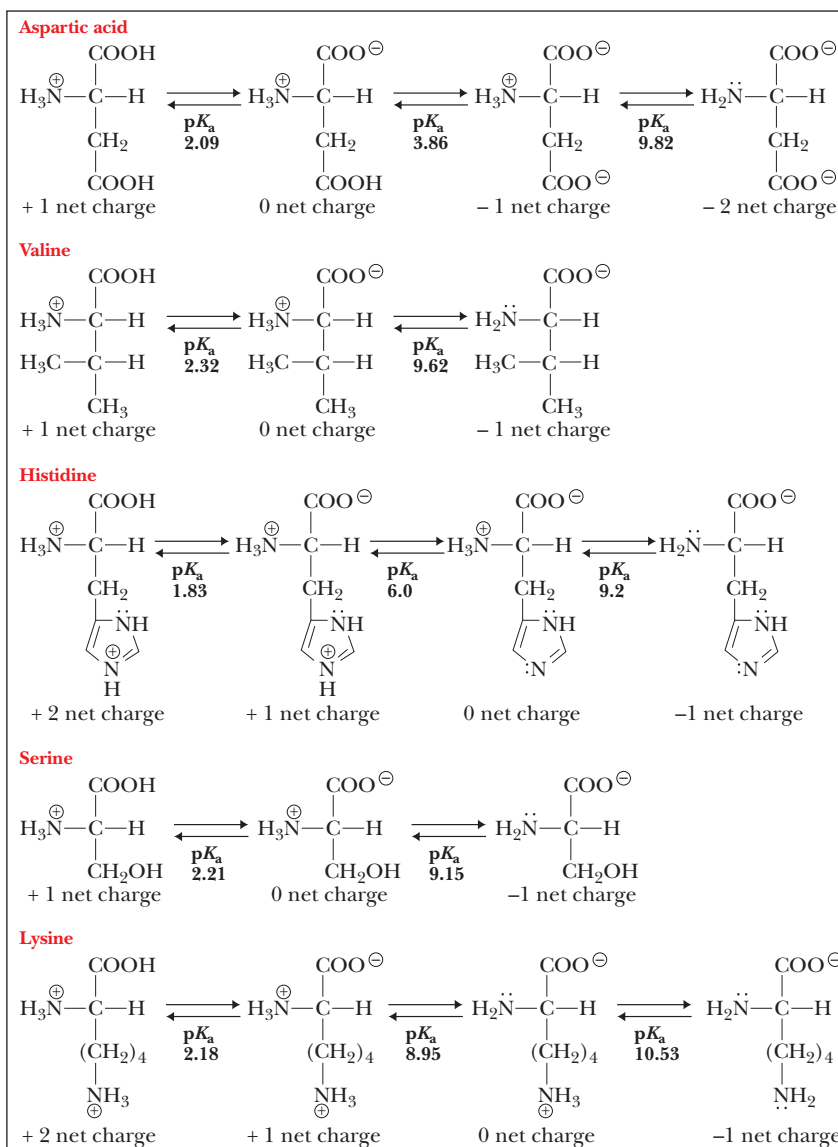


13. In all cases, the yield is  $0.95^n$ . For 10 residues, that means 60% yield; for 50 residues, 8%; and for 100 residues, 0.6%. These are not satisfactory yields. Enzyme specificity gets around the problem.
14. The conjugate acid-base pair acts as a buffer in the pH range 1.09–3.09.



15. They have a net charge at pH extremes, and the molecules tend to repel each other. When the molecular charge is zero, the amino acids can aggregate more easily.

16. The ionic dissociation reactions of the amino acids aspartic acid, valine, histidine, serine, and lysine are as follows:



17. The  $\text{pK}_a$  for the ionization of the thiol group of cysteine is 8.33, so this amino acid could serve as a buffer in the  $-\text{SH}$  and  $\text{S}^{2-}$  forms over the pH range 7.33–9.33. The  $\alpha$ -amino groups of asparagine and lysine have  $\text{pK}_a$  values of 8.80 and 8.95, respectively; these are also possible buffers, but they are both near the end of their buffer ranges.
18. At pH 4, the  $\alpha$ -carboxyl group is deprotonated to a carboxylate, the side-chain carboxyl is still more than 50% protonated, and both amino groups are protonated. At pH 7, both the  $\alpha$ -carboxyl group and the side-chain carboxyl group are deprotonated to a carboxylate, and both amino groups are protonated. At pH 10, both the  $\alpha$ -carboxyl group and the side-chain carboxyl group are deprotonated to a carboxylate, one of the amino groups is primarily deprotonated, and the other amino group is a mixture of the protonated and deprotonated forms.
19. The pI refers to the form in which both carboxyl groups are deprotonated, and both amino groups protonated at pH 6.96.
20. At pH 1, the charged groups are the N-terminal  $\text{NH}_3^+$  on valine and the protonated guanidino group on arginine, giving a net charge of +2. The charged groups at pH 7 are the same as

- those at pH 1, with the addition of the carboxylate group on the C-terminal leucine, giving a net charge of +1.
21. Both peptides, Phe—Glu—Ser—Met and Val—Trp—Cys—Leu, have a charge of +1 at pH 1 because of the protonated N-terminal amino group. At pH 7, the peptide on the right has no net charge because of the protonated N-terminal amino group and the ionized C-terminal carboxylate negative charge. The peptide on the left has a net charge of  $-1$  at pH 7 because of the side-chain carboxylate group on the glutamate in addition to the charges on the N-terminal and C-terminal groups.
  22. (a) Lysine, because of the side-chain amino group.  
(b) Serine, because of the lack of a side-chain carboxyl.
  23. Glycine is frequently used as the basis of a buffer in the acid range near the  $pK$  of its carboxyl group. The useful buffer range is pH 1.3–3.3.
- ### 3.4 The Peptide Bond
24. See Figure 3.10.
  25. The resonance structures contribute to the planar arrangement by giving the CON bond partial double-bond character.
  26. Tyrosine, tryptophan, and their derivatives.
  27. A monoamine oxidase is an enzyme that degrades compounds with an amino group, including neurotransmitters; consequently, it can control a person's mental state.
  28. The two peptides differ in amino acid sequence but not in composition.
  29. The titration curves of the two peptides have the same general shape. The  $pK_a$  values of the  $\alpha$ -amino and  $\alpha$ -carboxyl groups differ. Very careful work will show slight differences in side-chain  $pK_a$  values because of the different distances to the charged groups at the ends of the peptide. Such changes are particularly marked in proteins.
  30. Asp—Leu—Phe; Leu—Asp—Phe; Phe—Asp—Leu; Asp—Phe—Leu; Leu—Phe—Asp; Phe—Leu—Asp
  31. DLF; LDF; FDL; DFL; LFD; FLD
  32. You would get  $20^{100} \approx 1.27 \times 10^{130}$  molecules, which is about  $10^{84}$  Earth volumes. The same calculation for a pentapeptide gives more comprehensible results.
  33. Oxytocin has an isoleucine residue at position 3 of the amino acid sequence and a leucine residue at position 8; vasopressin has a phenylalanine residue at position 3 and an arginine residue at position 8. The primary role of oxytocin is in childbirth and lactation. The primary role of vasopressin is in control of blood pressure.
  34. The high concentration of tryptophan in milk protein may mildly elevate the levels of serotonin, which relaxes the brain.
  35. The tryptophan in milk might make you sleepy, whereas the tyramine in cheese should pep you up.
  36. They are relatively stable because they are zwitterions. They typically have high melting points.
  37. With very little doubt, no. Compare predicting the properties of water from those of hydrogen and oxygen, in either atomic or molecular form. If you knew the properties of the protein, you might be able to do the reverse to some extent.
  38. The amino acids thyroxine and hydroxyproline occur in very few proteins. The genetic code does not include them, so they are produced by modification of tyrosine and proline, respectively.
  39. These two peptides differ chemically. The open chain has a free C-terminal and N-terminal, but the cyclic peptide has only peptide bonds.

40. Both the C-terminal and the N-terminal of the open-chain peptide can be charged at appropriate pH values, which is not the case with the cyclic peptide. This can provide a basis for separation by electrophoresis.
  41. Carbohydrates are not a source of the nitrogen needed for biosynthesis of amino acids.
  42. Suggest that your friend show the carboxyl group as a charged carboxylate ( $-\text{COO}^-$ ) and the amino group in its charged form ( $-\text{NH}_3^+$ ).
  43. Very few side chains have functional groups to form crosslinks.
  44. Many more conformations would be possible because of free rotation around the peptide bond.
  45. There would be no possibility of disulfide crosslinks within or between peptide chains, giving more possible conformations. There would not be the possibility of oxidation–reduction reactions involving sulfhydryl and disulfide groups.
  46. The big difference would be the loss of stereospecificity in the conformation of any peptide or protein. This would have drastic consequences for the kinds of reactions of the protein.
- ### 3.5 Small Peptides with Physiological Activity
47. Oxytocin has an isoleucine at position 3 and a leucine at position 8; it stimulates smooth muscle contraction in the uterus during labor and in the mammary glands during lactation. Vasopressin has a phenylalanine at position 3 and an arginine at position 8; it stimulates resorption of water by the kidneys, thus raising blood pressure.
  48. The reduced form of glutathione consists of three amino acids with a sulfhydryl group; the oxidized form consists of six amino acids and can be considered the result of linking two molecules of reduced glutathione by a disulfide bridge.
  49. Enkephalins are pentapeptides (Y—G—G—F—L, leucine enkephalin, and Y—G—G—F—M, methionine enkephalin), which are naturally occurring analgesics.
  50. In most cases, D-amino acids are toxic. They occur in nature in antibiotics and bacterial cell walls.

## Chapter 4

### 4.1 Protein Structure and Function

1. (a) (iii); (b) (i); (c) (iv); (d) (ii).
2. When a protein is denatured, the interactions that determine secondary, tertiary, and any quaternary structures are overcome by the presence of the denaturing agent. Only the primary structure remains intact.
3. The random portions of a protein do not contain structural motifs that are repeated within the protein, such as  $\alpha$ -helix or  $\beta$ -pleated sheet, but three-dimensional features of these parts of the protein are repeated from one molecule to another. Thus, the term *random* is somewhat of a misnomer.

### 4.2 Primary Structure of Proteins

4. When a protein is covalently modified, its primary structure is changed. The primary structure determines the final three-dimensional structure of the protein. The modification disrupts the folding process.
5. (a) Serine has a small side chain that can fit in any relatively polar environment.  
(b) Tryptophan has the largest side chain of any of the common amino acids, and it tends to require a nonpolar environment.  
(c) Lysine and arginine are both basic amino acids; exchanging one for the other would not affect the side-chain  $pK_a$  in a

significant way. Similar reasoning applies to the substitution of a nonpolar isoleucine for a nonpolar leucine.

6. Glycine is frequently a conserved residue because its side chain is so small that it can fit into spaces that will not accommodate larger ones.
7. When alanine is replaced by isoleucine, there is not enough room in the native conformation for the larger side chain of the isoleucine. Consequently, there is a great enough change in the conformation of the protein that it loses activity. When glycine is substituted in turn for isoleucine, the presence of the smaller side chain leads to a restoration of the active conformation.
8. Meat consists largely of animal proteins and fat. The temperatures involved in cooking meat are usually more than enough to denature the protein portion of the meat.
9. Prion diseases have been linked to the immune system. It is believed that the prion proteins travel in the lymph system bound to lymphocytes and eventually arrive at the nervous tissue, where they begin to transform a normal cellular protein into an abnormal one (a prion).
10. Although there may be a strong genetic predisposition to acquire scrapie, that alone will not cause the disease. The disease must be started by ingesting a prion that already has the altered conformation, PrP<sup>Sc</sup>.

#### 4.3 Secondary Structure of Proteins

11. Shape, solubility, and type of biological function (static, structural versus dynamic, catalytic).
12. The protein efficiency ratio is an arbitrary measurement of the essential amino acid content of a given type of protein.
13. Eggs have the highest PER.
14. The essential amino acids are those that must be consumed in the diet because the body cannot synthesize them in sufficient quantities.
15. Reasons for creating genetically modified foods include increasing their protein content, increasing their shelf life, increasing their resistance to insects or other pests, and decreasing the need for using pesticides in order to grow them economically.
16. The angles of the amide planes as they rotate about the  $\alpha$ -carbon. The angles are both defined as zero when the two planes would be overlapping such that the carbonyl group of one contacts the N—H of the other.
17. A  $\beta$ -bulge is a common nonrepetitive irregularity found in anti-parallel  $\beta$ -sheets. A misalignment occurs between strands of the  $\beta$ -sheet, causing one side to bow outward.
18. A reverse turn is a region of a polypeptide where the direction changes by about  $180^\circ$ . There are two kinds—those that contain proline and those that do not. See Figure 4.6 for examples.
19. The  $\alpha$ -helix is not fully extended, and its hydrogen bonds are parallel to the protein fiber. The  $\beta$ -pleated sheet structure is almost fully extended, and its hydrogen bonds are perpendicular to the protein fiber.
20. The  $\alpha\alpha$  unit, the  $\beta\alpha\beta$  unit, the  $\beta$ -meander, the Greek key, and the  $\beta$ -barrel.
21. The geometry of the proline residue is such that it does not fit into the  $\alpha$ -helix, but it does fit exactly for a reverse turn. See Figure 4.10c.
22. Glycine is the only residue small enough to fit at crucial points in the collagen triple helix.
23. The principal component of wool is the protein keratin, which is a classic example of  $\alpha$ -helical structure. The principal

component of silk is the protein fibroin, which is a classic example of  $\beta$ -pleated sheet structure. The statement is somewhat of an oversimplification, but it is fundamentally valid.

24. Wool, which consists largely of the protein keratin, shrinks because of its  $\alpha$ -helical conformation. It can stretch and then shrink. Silk consists largely of the protein fibroin, which has the fully extended  $\beta$ -sheet conformation, with far less tendency to stretch or shrink.

#### 4.4 Tertiary Structure of Proteins

25. See Figure 4.2 for a hydrogen bond that is part of the  $\alpha$ -helix (secondary structure). See Figure 4.13 for a hydrogen bond that is part of tertiary structure (side-chain hydrogen bonding).
26. See Figure 4.13 for electrostatic interactions, such as might be seen between the side chains of lysine and aspartate.
27. See Figure 4.13 for an example of a disulfide bond.
28. See Figure 4.13 for an example of hydrophobic bonds.
29. *Configuration* refers to the position of groups due to covalent bonding. Examples include *cis* and *trans* isomers and optical isomers. *Conformation* refers to the positioning of groups in space due to rotation around single bonds. An example is the difference between the eclipsed and staggered conformations of ethane.
30. Five possible features limit possible protein configurations and conformations. (1) Although any one of 20 amino acids is possible at each position, only one is used, as dictated by the gene that codes for that protein. (2) Either a D- or an L-amino acid could be used at each position (except for glycine), but only L-amino acids are used. (3) The peptide group is planar, so that only *cis* and *trans* arrangements are observed. The *trans* form is more stable and is the one usually found in proteins. (4) The angles  $\phi$  and  $\psi$  can theoretically take on any value from  $0^\circ$  to  $360^\circ$ , but some angles are not possible because of steric hindrance; angles that are sterically allowed may not have stabilizing interactions, such as those in the  $\alpha$ -helix. (5) The primary structure determines an optimum tertiary structure, according to the “second half of the genetic code.”
31. Technically, collagen has quaternary structure because it has multiple polypeptide chains. However, most discussions of quaternary structure involve subunits of globular proteins, not fibrous ones like collagen. Many scientists consider the collagen triple helix to be an example of a secondary structure.

#### 4.5 Quaternary Structure of Proteins

32. *Similarities*: both contain a heme group; both are oxygen binding; secondary structure is primarily  $\alpha$ -helix. *Differences*: hemoglobin is a tetramer, while myoglobin is a monomer; oxygen binding to hemoglobin is cooperative, but noncooperative to myoglobin.
33. The crucial residues are histidines in both proteins.
34. Myoglobin's highest level of organization is tertiary. Hemoglobin's is quaternary.
35. The function of hemoglobin is oxygen transport; its sigmoidal binding curve reflects the fact that it can bind easily to oxygen at comparatively high pressures and release oxygen at lower pressures. The function of myoglobin is oxygen storage; as a result, it is easily saturated with oxygen at low pressures, as shown by its hyperbolic binding curve.
36. In the presence of  $H^+$  and  $CO_2$ , both of which bind to hemoglobin, the oxygen-binding capacity of hemoglobin decreases.
37. In the absence of 2,3-bisphosphoglycerate, the binding of oxygen by hemoglobin resembles that of myoglobin, characterized by lack of cooperativity. 2,3-Bisphosphoglycerate binds at the center of the hemoglobin molecule, increases cooperativity,



- stabilizes the deoxy conformation of hemoglobin, and modulates the binding of oxygen so that it can easily be released in the capillaries.
38. Fetal hemoglobin binds oxygen more strongly than adult hemoglobin. See Figure 4.25.
  39. Histidine 143 in a  $\beta$ -chain is replaced by a serine in a  $\gamma$ -chain.
  40. Deoxygenated hemoglobin is a weaker acid (has a higher  $pK_a$ ) than oxygenated hemoglobin. In other words, deoxygenated hemoglobin binds more strongly to  $H^+$  than does oxygenated hemoglobin. The binding of  $H^+$  (and of  $CO_2$ ) to hemoglobin favors the change in quaternary structure to the deoxygenated form of hemoglobin.
  41. The primary flaw in your friend's reasoning is a reversal of the definition of pH, which is  $pH = -\log [H^+]$ . If the release or binding of hydrogen ion by hemoglobin were the primary factor in the Bohr effect, the pH changes would be the opposite of those actually observed. The response of hemoglobin to changes in pH is the central point. When the pH increases, the hydrogen ion concentration decreases, and vice versa.
  42. The change of a histidine to a serine in the  $\gamma$ -chain removes a positively charged amino acid that could have interacted with BPG. Thus there are fewer salt bridges to break, so binding is easier than it is in a  $\beta$ -chain.
  43. People with sickle-cell trait have some abnormal hemoglobin. At high altitudes, there is less oxygen, and the concentration of the deoxy form of the abnormal hemoglobin increases. Less oxygen can be bound, causing the observed breathing difficulties.
  44. In fetal hemoglobin, the subunit composition is  $\alpha_2\gamma_2$  with replacement of the  $\beta$ -chains by the  $\gamma$ -chains. The sickle-cell mutation affects the  $\beta$ -chain, so the fetus homozygous for Hb S has normal fetal hemoglobin.
  45. The relative oxygen affinities allow oxygen to be taken by the fetal cells from the maternal Hb.
  46. Because people with sickle-cell disease are chronically anemic, some cells with fetal Hb are produced to help overcome the impaired oxygen delivery system.
  47. The crystalline form changed because oxygen entered under the cover slip, transforming deoxyhemoglobin to oxyhemoglobin.
- #### 4.6 Protein Folding Dynamics
48. This level of sequence homology is marginal for use of comparative modeling. It is best to try that method, but then to compare the results with those obtained from the fold-recognition approach.
  49. Protein folding is driven by many processes. The intuitive ones are the direct interactions of functional groups through covalent bonds, electrostatic attractions, and hydrogen bonds. These explain why parts of the protein are attracted to each other and why a protein would tend to adopt a shape making these interactions possible. However, much of the protein-folding process is also driven by an entropy effect. We refer to hydrophobic interactions as an explanation of why nonpolar regions of the protein tend to cluster together, usually in the interior of the protein. However, in reality, it is not the interaction of nonpolar amino acids that drives this process. It is actually the increase in entropy of the solvent, water. When the hydrophobic regions of the protein are isolated to the interior, the water molecules surrounding the protein are more free to rotate and move in less restricted ways. Thus, what drives much of protein folding is not a  $\Delta H$  change with the bonding of specific amino acids, but rather a  $\Delta S$  increase of the solvent.
  50. See the Protein Data Bank.
  51. A chaperone is a protein that aids another protein in folding correctly and keeps it from associating with other proteins before it has reached its final, mature form.
  52. A prion is a potentially infectious protein found in multiple forms in mammals, often concentrated in nervous tissue. It is an abnormal form of a normal cellular protein. It tends to form plaques that destroy the nervous tissue. Prions have been found to be transmissible across species.
  53. A series of encephalopathies have been found to be caused by prions. In cows, the disease caused by prions is called bovine spongiform encephalopathy, or more commonly mad-cow disease. In sheep, the disease is called scrapie. In humans, it is called Creutzfeldt-Jakob disease.
  54. The normal form of the prion protein has a higher  $\alpha$ -helix content compared to the  $\beta$ -sheet content. The abnormal one has an increased  $\beta$ -sheet content.
  55. Alzheimer's, Parkinson's, and Huntington's diseases are caused by accumulation of protein deposits from aggregates caused by misfolded proteins. This chapter also looked at prion diseases. When prions are misfolded they can cause spongiform encephalopathies, such as mad-cow disease, and the human form, Creutzfeldt-Jakob disease.
  56. Protein aggregates form when there are exposed areas on a protein surface that are nonpolar. Proteins then stick together via these nonpolar regions causing the aggregates. An example is the prion disease in which an area of the normal molecule that should be an  $\alpha$ -helix adopts a  $\beta$ -sheet conformation instead.
  57. The root problem with the globin genes and potential issues with hemoglobin formation is based on the fact that there are two  $\alpha$ -globin genes for every  $\beta$ -globin gene, yet to make hemoglobin they must combine in a 1:1 ratio. Thus, one theoretically possible solution would be if there were not a 2:1 ratio of these genes. Another issue is that the two genes are on different chromosomes. They are then most likely controlled separately. If the two genes were close together on one chromosome, then they could be controlled together by the same signal and produced in the correct amounts.
  58. The sequence of the mutant prion that confers the most extreme sensitivity to conferring a prion disease is the substitution of the amino acid at position 129 to a methionine.
  59. Prion diseases are transmissible, while the other neurodegenerative diseases like Alzheimer's are not.
  60. The spongiform encephalopathies that we know of have characteristics of both inherited diseases and transmissible diseases. On the one hand, animals can be infected by consuming meat or other tissues that are themselves carrying mutant prion proteins. As the example of the New Zealand sheep showed, even those that are very susceptible to a prion disease can remain disease-free if they are never exposed. However, the predisposition to acquire a prion disease has a hereditary component as well. The prion protein has many known mutations, some of which render the individual very susceptible to the disease. These mutations can be tracked and they are passed along family lines.
  61. The two enzymes associated with the disease are called  $\beta$ -secretase and  $\gamma$ -secretase.
  62. Amyloid  $\beta$  and  $\tau$  are the two proteins that form destructive plaques. The former is formed from pieces cut from a precursor protein called Amyloid Precursor Protein.



63. Alzheimer's begins with the buildup of  $A\beta$ , which is cut from the APP. In the first step, the enzyme  $\beta$ -secretase cuts APP outside the cell membrane. Then the  $\gamma$ -secretase enzyme cuts the remaining portion of the APP inside the membrane, releasing  $A\beta$ .
64.  $\beta$ -secretase has the natural function of being involved in the myelination of nerves.

## Chapter 5

### 5.1 Extracting Pure Proteins from Cells

1. Using a blender, a Potter-Elvehjem homogenizer, or a sonicator.
2. If you needed to maintain the structural integrity of the subcellular organelles, a Potter-Elvehjem homogenizer would be better because it is more gentle. The tissue, such as liver, must be soft enough to use with this device.
3. Salting out is a process whereby a highly ionic salt is used to reduce the solubility of a protein until it comes out of solution and can be centrifuged. The salt forms ion-dipole bonds with the water in the solution, which leaves less water available to hydrate the protein. Nonpolar side chains begin to interact between protein molecules, and they become insoluble.
4. Their amino acid content and arrangements make some proteins more soluble than others. A protein with more highly polar amino acids on the surface is more soluble than one with more hydrophobic ones on the surface.
5. First homogenize the liver cells using a Potter-Elvehjem homogenizer. Then spin the homogenate at  $500 \times g$  to sediment the unbroken cells and nuclei. Centrifuge the supernatant at  $15,000 \times g$  and collect the pellet, which contains the mitochondria.
6. No, peroxisomes and mitochondria have overlapping sedimentation characteristics. Other techniques, such as sucrose-gradient centrifugation, would have to be used to separate the two organelles.
7. If the protein were cytosolic, once the cells were broken open, you could centrifuge at  $100,000 \times g$  and all the organelles would be in the pellet. Your enzyme would be in the supernatant, along with all the other cytosolic ones.
8. Isolate the mitochondria via differential or sucrose-gradient centrifugation. Use another homogenization technique, combined with a strong detergent, to release the enzyme from the membrane.
9. Tables exist to tell you how many grams of ammonium sulfate  $[(NH_4)_2SO_4]$  to add to get a certain percent saturation. A good plan would be to take the homogenate and add enough ammonium sulfate to yield a 20% saturated solution. Let the sample sit for 15 minutes on ice and then centrifuge. Separate the supernatant from the precipitate. Assay both for the protein you are working with. Add more ammonium sulfate to the supernatant to arrive at a 40% saturated solution and repeat the process. In this way, you will find out what the percent saturation in ammonium sulfate needs to be to precipitate the protein.
10. Reasonably harsh homogenization would be able to liberate the soluble protein X from the peroxisomes, which are fragile. Centrifugation at  $15,000 \times g$  would sediment the mitochondria (broken or intact). The supernatant would then have protein X but no protein Y. Freeze/thaw techniques and sonication would accomplish the same thing, or the mitochondria and the peroxisomes could be separated initially by sucrose-gradient centrifugation.

### 5.2 Column Chromatography

11. (a) Size.  
(b) Specific ligand-binding ability.  
(c) Net charge.  
(d) Polarity.
12. The largest proteins elute first; the smallest elute last. Larger proteins are excluded from the interior of the gel bead so they have less available column space to travel. Essentially, they travel a shorter distance and elute first.
13. A compound can be eluted by raising the salt concentration or by adding a mobile ligand that has a higher affinity for the bound protein than the stationary resin ligand does. Salt is cheaper but less specific. A specific ligand may be more specific, but it is likely to be expensive.
14. A compound can be eluted by raising the salt concentration or by changing the pH. Salt is cheap, but it might not be as specific for a particular protein. Changing the pH may be more specific for a tight pI range, but extremes of pH may also denature the protein.
15. Raising the salt concentration is relatively safe. Most proteins will elute this way, and, if the protein is an enzyme, it will still be active. If necessary, the salt can be removed later via dialysis. Changing the pH enough to remove the charge can cause the proteins to become denatured. Many proteins are not soluble at the isoelectric points.
16. The basis of most resins is agarose, cellulose, dextran, or polyacrylamide.
17. See Figure 5.7.
18. Within the fractionation range of a gel-filtration column, molecules elute with a linear relationship of log MW versus their elution volumes. A series of standards can be run to standardize the column, and then an unknown can be determined by measuring its elution volume and comparing it to a standard curve.
19. Both proteins would elute in the void volume together and would not be separated.
20. Yes, the  $\beta$ -amylase would come out in the void volume, but the bovine serum albumin would be included in the column bead and would elute more slowly.
21. In most chromatography systems, the ligands and solvents are polar. In reverse phase HPLC, a solution of nonpolar compounds is put through a column that has a nonpolar liquid immobilized on an inert matrix. A more polar liquid serves as the mobile phase and is passed over the matrix. The solute molecules are eluted in proportion to their solubility in the more polar liquid.
22. Ion Exchange Chromatography is a specific type of separation based on net charge of the molecules being separated. The term HPLC refers to chromatography procedures carried out under high pressure, but the basis of the separation could be ion exchange, gel filtration, reverse phase, or affinity chromatography.
23. Set up an anion-exchange column, such as Q-Sepharose (quaternary amine). Run the column at pH 8.5, a pH at which the protein X has a net negative charge. Put a homogenate containing protein X on the column and wash with the starting buffer. Protein X will bind to the column. Then elute by running a salt gradient.
24. Use a cation-exchange column, such as CM-Sepharose, and run it at pH 6. Protein X will have a positive charge and will stick to the column.

25. With a quaternary amine, the column resin always has a net positive charge, and you don't have to worry about the pH of your buffer altering the form of the column. With a tertiary amine, there is a dissociable hydrogen, and the resin may be positive or neutrally charged, depending on the buffer pH.
26. The easiest way would be to use a sucrose gradient to separate the mitochondria from the peroxisomes first. Then break open the mitochondria via harsh homogenization or sonication, and then centrifuge the mitochondria. The pellet would contain protein B, while the supernatant would contain protein A. Contaminants could still exist, but they could be cleaned away by running gel filtration, on Sephadex G-75 (which would separate enzyme C from enzymes A and B), and then by running ion-exchange chromatography on Q-Sepharose at pH 7.5. Enzyme B would be neutral and would elute, while enzyme A would stick to the column.
27. Glutamic acid will be eluted first because the column pH is close to its pI. Leucine and lysine will be positively charged and will stick to the column. To elute leucine, raise the pH to around 6. To elute lysine, raise the pH to around 11.
28. A nonpolar mobile solvent will move the nonpolar amino acids fastest, so phenylalanine will be the first to elute, followed by glycine and then glutamic acid.
29. The nonpolar amino acids will stick the most to the stationary phase, so glutamic acid will move the fastest, followed by glycine and then phenylalanine.
30. A protein solution from an ammonium sulfate preparation is passed over a gel-filtration column where the proteins of interest will elute in the void volume. The salt, being very small, will move through the column slowly. In this way, the proteins will leave the salt behind and exit the column without it.

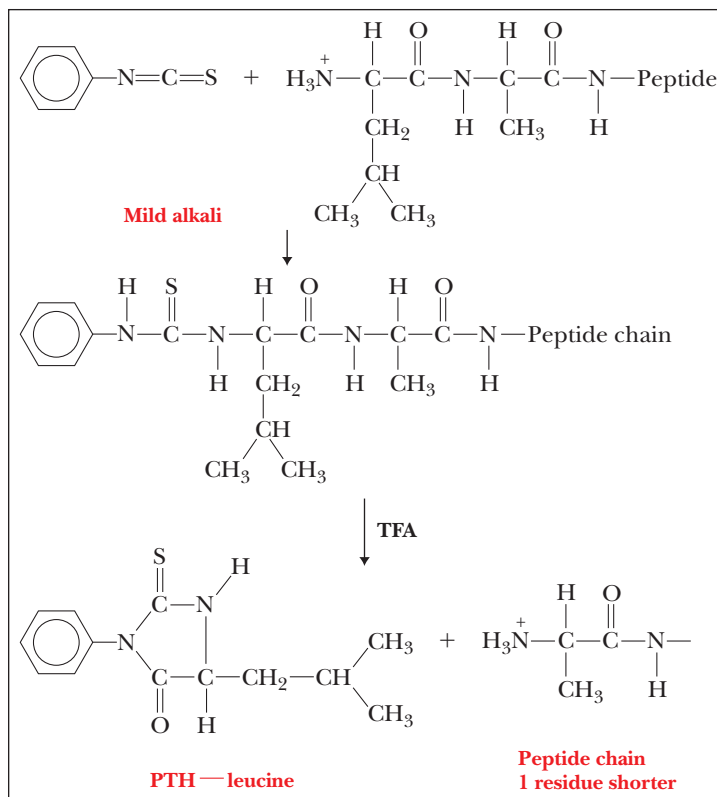
### 5.3 Electrophoresis

31. Size, shape, and charge.
32. Agarose and polyacrylamide.
33. Polyacrylamide.
34. DNA is the molecule most often separated on agarose electrophoresis, although proteins can also be separated.
35. Those with the highest charge/mass ratio would move the fastest. There are three variables to consider, and most electrophoreses are done in a way to eliminate two of the variables so that the separation is by size or by charge, but not by both.
36. Sodium dodecyl-sulfate polyacrylamide gel-electrophoresis. With SDS-PAGE, the charge and shape differences of proteins are eliminated so that the only parameter determining the migration is the size of the protein.
37. SDS binds to the protein in a constant ratio of 1.4 g SDS per gram of protein. It coats the protein with negative charges and puts it into a random coil shape. Thus, charge and shape are eliminated.
38. In a polyacrylamide gel used for gel-filtration chromatography, the larger proteins can travel around the beads, thereby having a shorter path to travel and therefore eluting faster. With electrophoresis, the proteins are forced to go through the matrix, so the larger ones travel more slowly because there is more friction.
39. The MW is 37,000 Da.

### 5.4 Determining the Primary Structure of a Protein

40. The Edman degradation will give the identity of the N-terminal amino acid in its first cycle, so doing a separate experiment is not necessary.
41. It might tell you if the protein were pure or if there were subunits.

42.



43. The amount of Edman reagent must exactly match the amount of N-termini in the first reaction. If there is too little Edman reagent, some of the N-termini will not react. If there is too much, some of the second amino acid will react. In either case, there will be a small amount of contaminating phenylthiohydantoin (PTH) derivatives. This error grows with the number of cycles run until the point that two amino acids are released in equal amounts, and you cannot tell which one was supposed to be the correct one.
44. In the first cycle, the first and second amino acids from the N-terminal end would be reacted and released as PTH derivatives. You would get a double signal and not know which one was the true N-terminus.
45. Val—Leu—Gly—Met—Ser—Arg—Asn—Thr—Trp—Met—Ile—Lys—Gly—Tyr—Met—Gln—Phe
46. Met—Val—Ser—Thr—Lys—Leu—Phe—Asn—Glu—Ser—Arg—Val—Ile—Trp—Thr—Leu—Met—Ile
47. It is possible that your protein is not pure and needs additional purification steps to arrive at a single polypeptide. It is also possible that the protein has subunits, so multiple polypeptide chains could be yielding the contradictory results.
48. There are two fragments that have C-termini that are not lysine or arginine, which is what trypsin is specific for. Normally there would be only one fragment ending with an amino acid that was not Arg or Lys, and we would immediately know that it was the C-terminus. Histidine is a basic amino acid, although it is usually neutral and therefore does not react with trypsin. It is possible that, in the pH environment of the reaction, the histidine was positively charged and was recognized by trypsin.
49. It would tell you a relative concentration of the various amino acids. This is important because it would help you plan your sequencing experiment better. For example, if you had a protein

whose composition showed no aromatic amino acids, it would be a waste of time to use a chymotrypsin digestion.

50. Cyanogen bromide would be useless, because there is no methionine. Trypsin would be little better, because the protein is 35% basic residues. Trypsin would shred the protein into more than 30 pieces, which would be very hard to analyze.
51. Chymotrypsin would be a good choice. There are more than four residues of aromatic amino acids. The protein, containing 100 amino acids, would be cut four times, possibly yielding nice fragments roughly 20–30 amino acids long, which can be sequenced effectively by the Edman degradation.
52. It would work best if the basic residues were spread out in the protein. In that way, fragments in the proper size range would be generated. If all four of the basic residues were in the first 10 amino acids, there would be one long fragment that could not be sequenced.
53. Electrospray Ionization (ESI-MS) and Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF MS).
54. MALDI-TOF MS is very sensitive and very accurate. Attomole ( $10^{-18}$ ) quantities of a molecule can be detected.
55. Proteomics is the systematic analysis of an organism's complete complement of proteins, or its **proteome**. Just as we learned the basic dogma of molecular biology (DNA  $\rightarrow$  RNA  $\rightarrow$  protein), the technology now available has allowed scientists to describe all the DNA of an organism as its genome, all of the RNA as its transcriptome, and all of the proteins produced as its proteome. To understand the flux of proteins in a cell is to understand its metabolism.
56. The bait protein is constructed to have a particular affinity tag. The bait protein interacts with cell proteins of interest and then binds to an affinity column via the tag. In this way, the cell proteins of interest can be found and isolated.
57. There are many assumptions behind the experiment described in the Biochemical Connections on page 135. One must assume that the nature of the tag has not changed the binding of the protein. For example, if adding the tag makes a protein more likely or less likely to bind to it, then the conclusions about cellular protein binding may be incorrect. For example, one might conclude that two proteins bind together to serve their metabolic function, but this binding could be an artifact of the experimental conditions. One must also assume that tagging the proteins has not changed the affinity between the tag and the affinity column.

## Chapter 6

### 6.1 Enzymes Are Effective Biological Catalysts

1. Enzymes are many orders of magnitude more effective as catalysts than are nonenzymatic catalysts.
2. Most enzymes are proteins, but some catalytic RNAs (ribozymes) are known.
3. About 3 seconds ( $1 \text{ year} \times 1 \text{ event}/10^7 \text{ events} \times 365 \text{ days/year} \times 24 \text{ hours/day} \times 3600 \text{ seconds/hour} = 3.15 \text{ seconds}$ ).
4. Enzymes hold the substrates in favorable spatial positions, and they bind effectively to the transition state to stabilize it. Note that *all* catalysts lower the activation energy, so this is not a particular enzyme function.

### 6.2 Kinetics versus Thermodynamics

5. The reaction of glucose with oxygen is thermodynamically favored, as shown by the negative free-energy change. The fact that glucose can be maintained in an oxygen atmosphere is a reflection of the kinetic aspects of the reaction, requiring overcoming an activation-energy barrier.

6. To the first question, most probably: local concentrations (mass-action concepts) could easily dictate the direction. To the second question, probably not: local concentrations would seldom be sufficient to overcome a relatively large  $\Delta G^\circ$  of  $-5.3 \text{ kcal}$  in the reverse reaction. (See, however, the aldolase reaction in glycolysis.)
7. Heating a protein denatures it. Enzymatic activity depends on the correct three-dimensional structure of the protein. The presence of bound substrate can make the protein harder to denature.
8. The results do not prove that the mechanism is correct because results from different experiments could contradict the proposed mechanism. In that case, the mechanism would have to be modified to accommodate the new experimental results.
9. The presence of a catalyst affects the rate of a reaction. The standard free-energy change is a thermodynamic property that does not depend on the reaction rate. Consequently, the presence of the catalyst has no effect.
10. The presence of a catalyst lowers the activation energy of a reaction.
11. Enzymes, like all catalysts, increase the rate of the forward and reverse reaction to the same extent.
12. The amount of product obtained in a reaction depends on the equilibrium constant. A catalyst does not affect that.

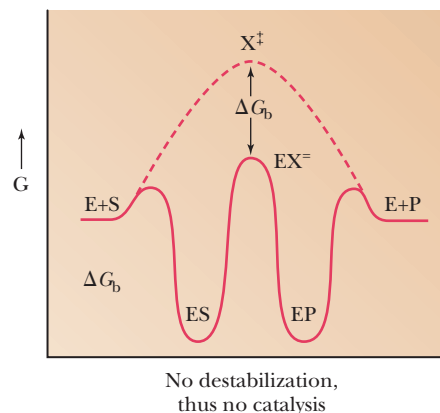
### 6.3 Enzyme Kinetic Equations

13. The reaction is first order with respect to A, first order with respect to B, and second order overall. The detailed mechanism of the reaction is likely to involve one molecule each of A and B.
14. The easiest way to follow the rate of this reaction is to monitor the decrease in absorbance at 340 nm, reflecting the disappearance of NADH.
15. The use of a pH meter would not be a good way to monitor the rate of the reaction. You are probably running this reaction in a buffer solution to keep the pH relatively constant. If you are not running the reaction in a buffer solution, you run the risk of acid denaturation of the enzyme.
16. Enzymes tend to have fairly sharp pH optimum values. It is necessary to ensure that the pH of the reaction mixture stays at the optimum value. This is especially true for reactions that require or produce hydrogen ions.

### 6.4 Enzyme-Substrate Binding

17. In the lock-and-key model, the substrate fits into a comparatively rigid protein that has an active site with a well-defined shape. In the induced-fit model, the enzyme undergoes a conformational change on binding to the substrate. The active site takes shape around the substrate.

18.



19. The ES complex would be in an “energy trough,” with a consequentially large activation energy to the transition state.
20. Amino acids that are far apart in the amino acid sequence can be close to each other in three dimensions because of protein folding. The critical amino acids are in the active site.
21. The overall protein structure is needed to ensure the correct arrangement of amino acids in the active site.
22. The strong inhibition indicates tight binding to the active site. Thus, the compound is very likely to be a transition-state analogue.

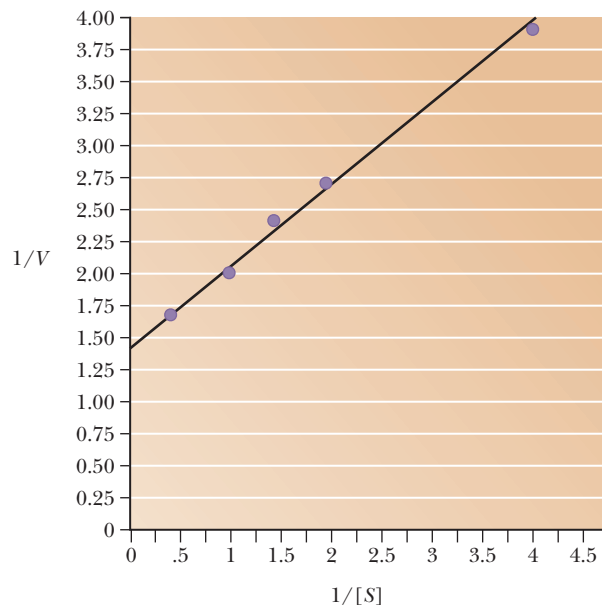
### 6.5 Examples of Enzyme-Catalyzed Reactions

23. See Figures 6.6 and 6.7.
24. Not all enzymes follow Michaelis–Menten kinetics. The kinetic behavior of allosteric enzymes does not obey the Michaelis–Menten equation.
25. The graph of rate against substrate concentration is sigmoidal for an allosteric enzyme but hyperbolic for an enzyme that obeys the Michaelis–Menten equation.

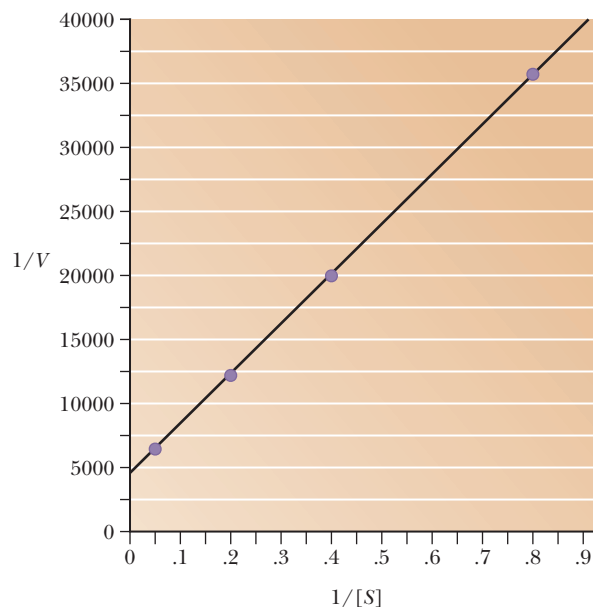
### 6.6 The Michaelis–Menten Approach to Enzyme Kinetics

26. The reaction velocity remains the same with increasing enzyme concentration. It is theoretically possible, but highly unlikely, for a reaction to be saturated with enzyme.
27. The steady-state assumption is that the concentration of the enzyme–substrate complex does not change appreciably over the time in which the experiment takes place. The rate of appearance of the complex is set equal to its rate of disappearance, simplifying the equations for enzyme kinetics.
28. Turnover number =  $V_{\max}/[ET]$ .
29. Use Equation 6.16.
  - (a)  $V = 0.5 V_{\max}$
  - (b)  $V = 0.33 V_{\max}$
  - (c)  $V = 0.09 V_{\max}$
  - (d)  $V = 0.67 V_{\max}$
  - (e)  $V = 0.91 V_{\max}$

30. See graph:  $V_{\max} = 0.681 \text{ mM min}^{-1}$ ,  $K_M = 0.421 \text{ M}$ .

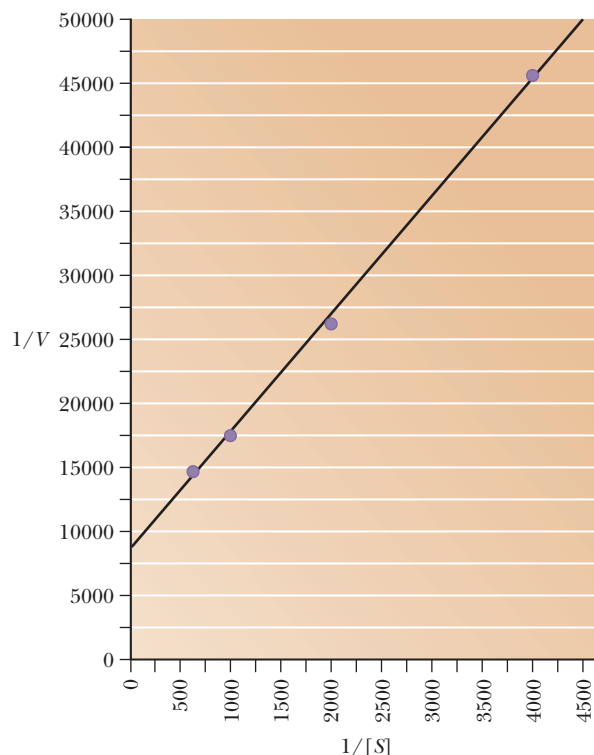


31. See graph:  $V_{\max} = 2.5 \times 10^{-4} \text{ M sec}^{-1}$ ,  $K_M = 1.6 \times 10^8 \text{ M}$ .



32. See graph:  $K_M = 2.86 \times 10^{-2} \text{ M}$ . Concentrations were not determined directly. Absorbance values were used instead as a matter of convenience.

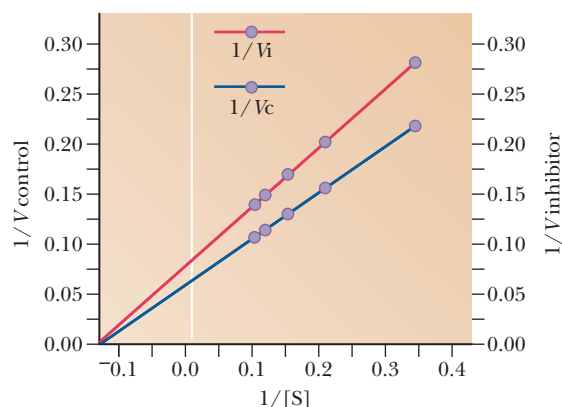
33. See graph:  $V_{\max} = 1.32 \times 10^{-3} \text{ M min}^{-1}$ ,  $K_M = 1.23 \times 10^{-3} \text{ M}$ .



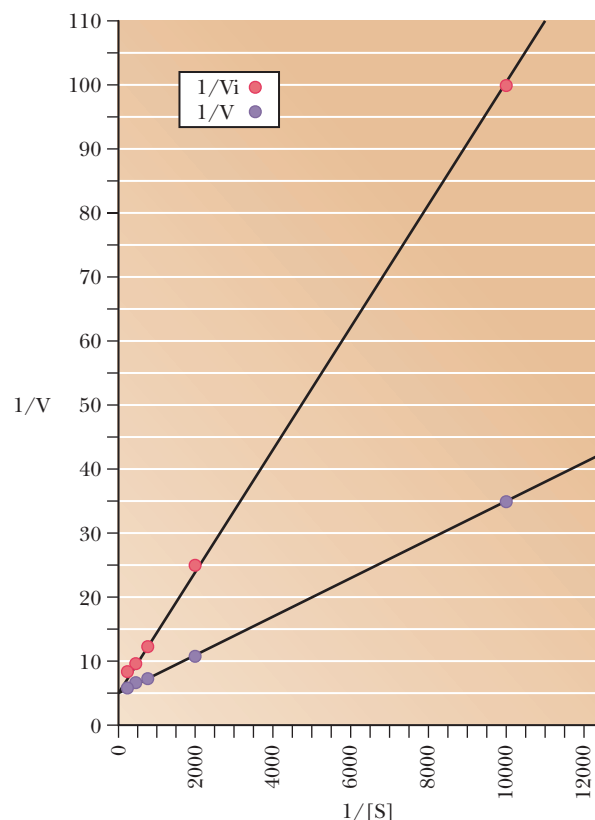
34. The turnover number is  $20.43 \text{ min}^{-1}$ .
35. The number of moles of enzyme is  $1.56 \times 10^{-10}$ . The turnover number is  $10,700 \text{ sec}^{-1}$ .
36. The low  $K_M$  for the aromatic amino acids indicates that they will be oxidized preferentially.



37. It is easier to detect deviations of individual points from a straight line than from a curve.
  38. The assumption that the  $K_M$  is an indication of the binding affinity between the substrate and the enzyme is valid when the rate of dissociation of the enzyme–substrate complex to product and enzyme is much smaller than the rate of dissociation of the complex to enzyme and substrate.
  39. Acetamidazole is an inhibitor of carbonic anhydrase, which is part of a taste receptor that responds to  $\text{CO}_2$ .
  40. Scientists were taking acetamidazole to help fight altitude sickness and noticed their beer tasted terrible. They then studied carbonic anhydrase and found it was a chemical sensor for  $\text{CO}_2$ .
  41. Hexokinase is found predominantly in the muscle and acts during glycolysis of muscle glucose. Glucokinase is found in the liver. The higher  $K_M$  of glucokinase can be explained by the body's need to have the muscle enzyme function at lower glucose levels than the liver enzyme under conditions where quick energy is needed.
  42. Under conditions of low substrate concentration.
- 6.7 Enzyme Inhibition**
43. In the case of competitive inhibition, the value of  $K_M$  increases, while the value of  $K_M$  remains unchanged in noncompetitive inhibition.
  44. A competitive inhibitor blocks binding, not catalysis.
  45. A noncompetitive inhibitor does not change the affinity of the enzyme for its substrate.
  46. A competitive inhibitor binds to the active site of an enzyme, preventing binding of the substrate. A noncompetitive inhibitor binds at a site different from the active site, causing a conformational change, which renders the active site less able to bind substrate and convert it to product.
  47. Competitive inhibition can be overcome by adding enough substrate, but this is not true for all forms of enzyme inhibition.
  48. A Lineweaver–Burk plot is useful because it gives a straight line. It is easier to determine how well points fit to a straight line than to a curve.
  49. In a Lineweaver–Burk plot for competitive inhibition, the lines intersect at the  $y$ -axis intercept, which is equal to  $1/V_{\max}$ . In a Lineweaver–Burk plot for noncompetitive inhibition, the lines intersect at the  $x$ -axis intercept, which is equal to  $-1/K_M$ .
  50.  $K_M = 7.42 \text{ mM}$ ;  $V_{\max} = 15.9 \text{ mmol min}^{-1}$ ; noncompetitive inhibition.



51. Competitive inhibition,  $K_M = 6.5 \times 10^{-4}$ . The key point here is that the  $V_{\max}$  is the same within the limits of error. Some of the concentrations are given to one significant figure.



52. It is *very* good, in the case of noncompetitive inhibitors; much of metabolic control depends on feedback inhibition by downstream noncompetitive inhibitors. The question is perhaps moot in the case of competitive inhibitors, which are much less commonly encountered in vivo. Some antibiotics, however, are competitive inhibitors (good for the sick person, bad for the bacteria).
53. Both the slope and the intercepts will change. The lines will intersect above the  $x$ -axis at negative values of  $1/[S]$ .
54. Not all AIDS drugs are enzyme inhibitors, but an important class of such drugs inhibits the HIV protease. You would need to understand the concepts of substrate binding, inhibition, and inhibitor binding.
55. An irreversible inhibitor is bound by covalent bonds. Noncovalent interactions are relatively weak and easily broken.
56. A noncompetitive inhibitor does not bind to the active site of an enzyme. Its structure need bear no relation to that of the substrate.
57. The production of new virus particles inside the infected cell via inhibition of the HIV protease.
58. Replication of the HIV genome inside an infected cell via inhibition of the HIV integrase.
59. Scientists created a drug called ZIP that blocks the enzyme. They gave rats saccharine-laced water and then artificially induced nausea shortly afterwards. Control rats then had an aversion to saccharine-laced water for weeks afterwards, but the ZIP-treated rats did not.



## Chapter 7

### 7.1 The Behavior of Allosteric Enzymes

1. Allosteric enzymes display sigmoidal kinetics when rates are plotted versus substrate concentration. Michaelis–Menten enzymes exhibit hyperbolic kinetics. Allosteric enzymes usually have multiple subunits, and the binding of substrates or effector molecules to one subunit changes the binding behavior of the other subunits.
2. It is an enzyme used in the early stages of cytidine nucleotide synthesis.
3. ATP acts as a positive effector of ATCase, and CTP acts as an inhibitor.
4. The term  $K_M$  should be used for enzymes that display Michaelis–Menten kinetics. Thus, it is not used with allosteric enzymes. Technically, competitive and noncompetitive inhibition are also terms that are restricted to Michaelis–Menten enzymes, although the concepts are applicable to any enzyme. An inhibitor that binds to an allosteric enzyme at the same site as the substrate is similar to a classical competitive inhibitor. One that binds at a different site is similar to a noncompetitive inhibitor, but the equations and the graphs characteristic of competitive and noncompetitive inhibition don't work the same way with an allosteric enzyme.
5. A K system is an allosteric enzyme in which the binding of inhibitor alters the apparent substrate concentration needed to reach one-half  $V_{\max}$ ,  $S_{0.5}$ .
6. A V system is an allosteric enzyme in which the binding of inhibitor changes the  $V_{\max}$  of the enzyme but not the  $S_{0.5}$ .
7. Homotropic effects are allosteric interactions that occur when several identical molecules are bound to a protein. The binding of substrate molecules to different sites on an enzyme, such as the binding of aspartate to ATCase, is an example of a homotropic effect. Heterotropic effects are allosteric interactions that occur when different substances (such as inhibitor and substrate) are bound to the protein. In the ATCase reaction, inhibition by CTP and activation by ATP are both heterotropic effects.
8. ATCase is made up of two different types of subunits. One of them is the catalytic subunit, and there are six of them organized into two trimers. The other is the regulatory subunit, which consists of six protein subunits organized into three dimers.
9. Enzymes that exhibit cooperativity do not show hyperbolic curves of rate versus substrate concentration. Their curves are sigmoidal. The level of cooperativity can be seen by the shape of the sigmoidal curve.
10. Inhibitors make the shape of the curve more sigmoidal.
11. Activators make the shape of the curve less sigmoidal.
12.  $K_{0.5}$  is the substrate concentration that leads to half of the maximal velocity. This term is used with allosteric enzymes, where the term  $K_M$  is not appropriate.
13. A mercury compound was used to separate the subunits of ATCase. When the subunits were separated, one type of subunit retained catalytic activity but was no longer allosteric and was not inhibited by CTP. The other subunit type had no ATCase activity, but it did bind to CTP and ATP.

### 7.2 The Concerted and Sequential Models for Allosteric Enzymes

14. In the concerted model, all the subunits in an allosteric enzyme are found in the same form, either the T form or the R form.

They are in equilibrium, with each enzyme having a characteristic ratio of the T/R. In the sequential model, the subunits change individually from T to R.

15. The sequential model can explain negative cooperativity, because a substrate binding to the T form could induce other subunits to switch to the T form, thereby reducing binding affinity.
16. Greater cooperativity is favored by having a higher ratio of the T/R form. It is also favored by having a higher dissociation constant for the substrate binding to the T form.
17. The L value is the equilibrium ratio of the T/R form. The  $c$  value is the ratio of the dissociation constants for substrate and the two forms of enzyme, such that  $c = K_R/K_T$ .
18. Many models are possible. We never really know for sure how the enzyme works, rather, we create a model that explains the observed behavior. It is very possible that another model would do so as well.
19. Scientists looked for drugs that would mimic the behavior of signaling molecules, such as hormones and neurotransmitters.
20. Side effects occur because the drug that is meant to affect one type of receptor will likely have unintended effects on several others.
21. First, allosteric effectors modulate the response in a more subtle way than orthosteric ones. Second, an allosteric drug is more specific for one or a few receptor types. Third, allosteric drugs can be safer because they have no effect at all unless the natural ligand is present.
22. Valium is an allosteric drug that binds to a different site on the receptors for  $\gamma$ -aminobutyric acid (GABA). Valium turns up the response of the receptor for the GABA. When Valium is bound, the response to GABA goes up manyfold.
23. Taking too much Valium is not as deadly as taking too much Phenobarbital because the Valium does not have a direct effect. Rather, it modulates the effect of the bound, natural ligand.
24. One is Amgen's Cincalcet, a drug designed to fight chronic kidney failure by improving the action of calcium receptors. The other is an HIV medication by Pfizer called Maraviroc. It interferes with HIV entry into the cells.

### 7.3 Control of Enzyme Activity by Phosphorylation

25. A kinase is an enzyme that phosphorylates a protein using a high-energy phosphate, such as ATP, as the phosphate donor.
26. Serine, threonine, and tyrosine are the three most often phosphorylated amino acids in proteins that are acted upon by kinases. Aspartate is another one often phosphorylated.
27. The allosteric effect can be faster because it is based on simple binding equilibrium. For example, if AMP is an allosteric activator of glycogen phosphorylase, the immediate increase in AMP when muscles contract can cause muscle phosphorylase to become more active and to provide energy for the contracting muscles. The phosphorylation effect requires the hormone cascade beginning with glucagon or epinephrine. There are many steps before the glycogen phosphorylase is phosphorylated, so the response time is slower. However, the cascade effect produces many more activated phosphorylase molecules, so the effects are longer and stronger.
28. As part of the mechanism, the sodium–potassium ATPase has an aspartate residue that becomes phosphorylated. This phosphorylation alters the conformation of the enzyme and causes

it to close on one side of the membrane and open on the other, moving ions in the process.

29. Glycogen phosphorylase is controlled allosterically by several molecules. In the muscle, AMP is an allosteric activator. In the liver, glucose is an allosteric inhibitor. Glycogen phosphorylase also exists in a phosphorylated form and an unphosphorylated form, with the phosphorylated form being more active.

#### 7.4 Zymogens

30. The digestive enzymes trypsin and chymotrypsin are classic examples of regulation by zymogens. The blood-clotting protein thrombin is another.
31. Trypsin, chymotrypsin, and thrombin are all proteases. Trypsin cleaves peptide bonds where there are amino acids with positively charged side chains (Lys and Arg). Chymotrypsin cleaves peptides at amino acids with aromatic side chains. Thrombin cleaves the protein fibrinogen into fibrin.
32. Caspases are a family of homodimer cysteine proteases responsible for many processes in cell biology, including apoptosis, signaling within the immune system, and stem cell differentiation.
33. Chymotrypsinogen is an inactive zymogen. It is acted upon by trypsin, which cleaves peptides at basic residues, like arginine. When trypsin cleaves between the arginine and the isoleucine, chymotrypsinogen becomes semiactive, forming  $\pi$ -chymotrypsin. This molecule digests itself further, forming the active  $\alpha$ -chymotrypsin. As it turns out, the  $\alpha$ -amino group of the isoleucine produced by the first cleavage is near the active site of  $\alpha$ -chymotrypsin and necessary for its activity.
34. Zymogens are often seen with digestive enzymes that are produced in one tissue and used in another. If the enzyme were active immediately upon production, it would digest other cell proteins, where it would cause great damage. By having it produced as a zymogen, it can be safely made and then transported to the digestive tissue, such as the stomach or small intestine, where it can then be activated.
35. This allows for a more rapid response when the hormone is needed. The hormone is already synthesized and usually just requires breaking one or two bonds to make it active. The hormone can be poised and ready to go on demand.
36. Apoptosis is a natural phenomenon of programmed cell death.
37. Disruption of apoptosis can lead to forms of cancer and unwanted cell death, such as cells surrounding neurons that have died from a stroke.

#### 7.5 The Nature of the Active Site

38. Serine and histidine are the two most critical amino acids in the active site of chymotrypsin.
39. The initial phase releases the first product and involves an acyl-enzyme intermediate. This step is faster than the second part, in which water comes into the active site and breaks the acyl-enzyme bond.
40. In the first step of the reaction, the serine hydroxyl is the nucleophile that attacks the substrate peptide bond. In the second step, water is the nucleophile that attacks the acyl-enzyme intermediate.
41. Histidine 57 performs a series of steps involving general base catalysis followed by general acid catalysis. In the first phase, it takes a hydrogen from serine 195, acting as a general base. This is followed immediately by an acid catalysis step, in which it gives

the hydrogen to the amide group of the peptide bond that is breaking. A similar scheme takes place in the second phase of the reaction.

42. The first phase is faster for several reasons. The serine at position 195 is a strong nucleophile for the initial nucleophilic attack. It then forms an acyl-enzyme intermediate. In the second phase, water is the nucleophile, and it takes time for water to diffuse to the right spot to perform its nucleophilic attack. It is also not as strong a nucleophile as the serine. Therefore, it takes longer for water to perform its nucleophilic attack and break the acyl-enzyme intermediate than it takes for serine to create it.
43. Histidine 57 exists in both the protonated and unprotonated form during the chymotrypsin reaction. Its  $pK_a$  of 6.0 makes this possible in the physiological pH range.
44. Instead of a phenylalanine moiety (similar to the usual substrates of chymotrypsin), use a nitrogen-containing basic group similar to the usual substrates of trypsin.

#### 7.6 Chemical Reactions Involved in Enzyme Mechanisms

45. They act as Lewis acids (electron-pair acceptors) and can take part in enzyme catalysis mechanisms of enzymes.
46. The carbon of a carbonyl group is often attacked by a nucleophile.
47. General acid catalysis is the part of an enzyme mechanism in which an amino acid or other molecule donates a hydrogen ion to another molecule.
48.  $S_N1$  stands for unimolecular nucleophilic substitution. The unimolecular part means that it obeys first-order kinetics. If the reaction is  $R:X + Z: \rightarrow R:Z + X:$ , with an  $S_N1$  reaction, the rate depends on the speed with which the X breaks away from the R. The Z group comes in later and quickly, compared with the breakdown of R:X.  $S_N2$  stands for bimolecular nucleophilic substitution. This happens with the same reaction scheme if the Z attacks the R:X molecule before it breaks down. Thus, the concentration of both R:X and Z: are important, and the rate displays second-order kinetics.
49. The  $S_N1$  reaction leads to loss of stereospecificity as the X group leaves before the entering nucleophile. This means that the nucleophile can enter from different angles, leading to different isomers.
50. The results do not prove that the mechanism is correct, because results from different experiments could contradict the proposed mechanism. In that case, the mechanism would have to be modified to accommodate the new experimental results.

#### 7.7 The Active Site and Transition States

51. A good transition-state analogue would have to have a tetrahedral carbon atom where the amide carbonyl group was originally found, since the transition state involves a momentary tetrahedral form. It would also have to have oxygens on the same carbon, so that there would be sufficient specificity for the active site.
52. The induced-fit model assumes that the enzyme and substrate must both move and change to conform to each other perfectly. Thus, the true fit is not between the enzyme and substrate but between the enzyme and the transition state of the substrate on its way to product. A transition-state analogue fits the enzyme nicely in this model.
53. An abzyme is created by injecting a host animal with a transition-state analogue of a reaction of interest. The host animal

makes antibodies to the foreign molecule, and these antibodies have specific binding points that mimic an enzyme surrounding a transition state. The purpose is to create an antibody with catalytic activity.

54. Cocaine blocks the reuptake of the neurotransmitter dopamine at synapses. Thus, dopamine stays in the system longer, overstimulating the neuron and leading to the reward signals in the brain that lead to addiction. Using a drug to block a receptor would be of no use with cocaine addiction and would probably just make removal of dopamine even more unlikely.
55. Cocaine can be degraded by a specific enzyme that hydrolyzes an ester bond that is part of cocaine's structure. In the process of this hydrolysis, the cocaine must pass through a transition state that changes its shape. Catalytic antibodies to the transition state of the hydrolysis of cocaine hydrolyze cocaine to two harmless degradation products—benzoic acid and ecgonine methyl ester. When degraded, the cocaine cannot block dopamine reuptake. No prolongation of the neuronal stimulus occurs, and the addictive effects of the drug vanish over time.

### 7.8 Coenzymes

56. Nicotinamide adenine dinucleotide, oxidation–reduction; flavin adenine dinucleotide, oxidation–reduction; coenzyme A, acyl transfer; pyridoxal phosphate, transamination; biotin, carboxylation; lipoic acid, acyl transfer.
57. Most coenzymes are derivatives of compounds we call vitamins. For example, nicotinamide adenine dinucleotide is produced from the B vitamin niacin. Flavin adenine dinucleotide comes from riboflavin.
58. Vitamin B<sub>6</sub> is the source of pyridoxal phosphate, which is used in transamination reactions.
59. Coenzymes can accomplish the same mechanisms that the amino acids do in a reaction. For example, a metal ion may act as a general acid or base. Parts of a coenzyme, such as the reactive carbanion of thiamine pyrophosphate, may act as a nucleophile to catalyze the reaction.
60. Yes, there would be a preference. Because the coenzyme and the other substrate will be locked into the enzyme, the hydride ion would come from some functional group that had a fixed position. Therefore, the hydride would come from one side.

## Chapter 8

### 8.1 The Definition of a Lipid

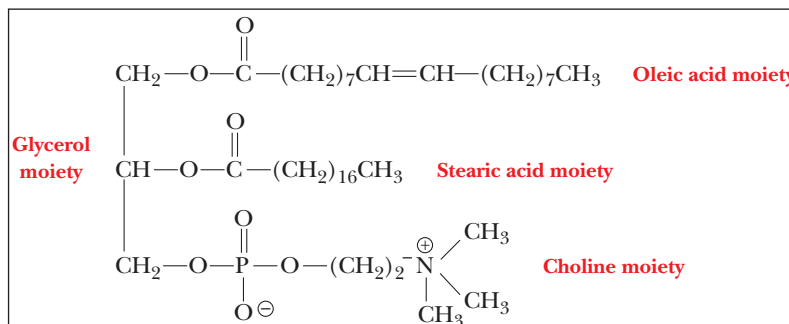
1. Solubility properties (insoluble in aqueous or polar solvents, soluble in nonpolar solvents). Some lipids are not at all structurally related.

### 8.2 The Chemical Natures of the Lipid Types

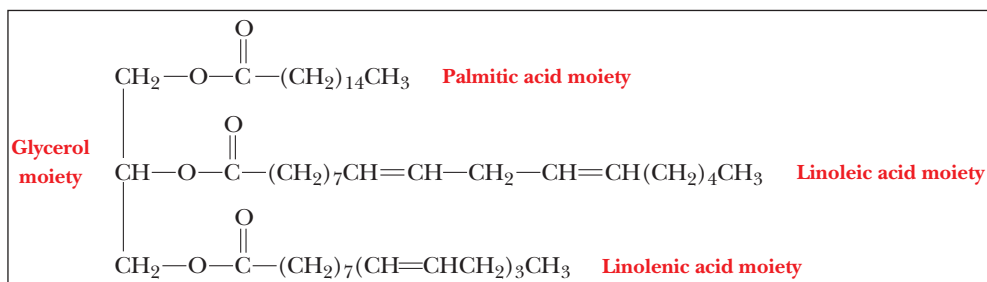
2. In both types of lipids, glycerol is esterified to carboxylic acids, with three such ester linkages formed in triacylglycerols and two in phosphatidyl ethanolamines. The structural difference comes in the nature of the third ester linkage to glycerol. In phosphatidyl ethanolamines, the third hydroxyl group of glycerol is

esterified not to a carboxylic acid but to phosphoric acid. The phosphoric acid moiety is esterified in turn to ethanolamine. (See Figures 8.2 and 8.5.)

3.

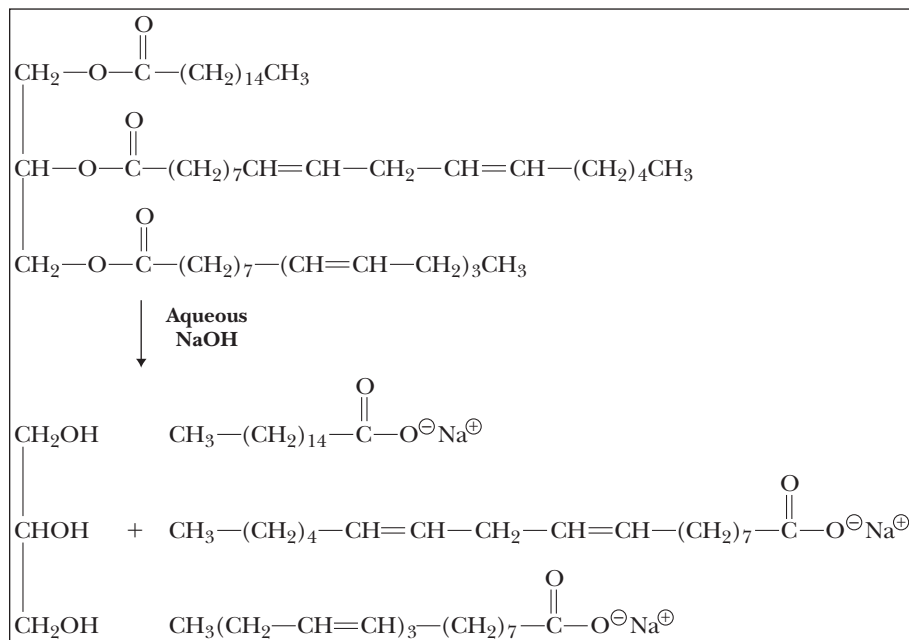


4. Both sphingomyelins and phosphatidylcholines contain phosphoric acid esterified to an amino alcohol, which must be choline in the case of a phosphatidylcholine and may be choline in the case of a sphingomyelin. They differ in the second alcohol to which phosphoric acid is esterified. In phosphatidylcholines, the second alcohol is glycerol, which has also formed ester bonds to two carboxylic acids. In sphingomyelins, the second alcohol is another amino alcohol, sphingosine, which has formed an amide bond to a fatty acid. (See Figure 8.6.)
5. This lipid is a ceramide, which is one kind of sphingolipid.
6. Sphingolipids contain amide bonds, as do proteins. Both can have hydrophobic and hydrophilic parts, and both can occur in cell membranes, but their functions are different.
7. Any combination of fatty acids is possible.



8. Steroids contain a characteristic fused-ring structure, which other lipids do not.
9. Waxes are esters of long-chain carboxylic acids and long-chain alcohols. They tend to be found as protective coatings.
10. Phospholipids are more hydrophilic than cholesterol. The phosphate group is charged, and the attached alcohol is charged or polar. These groups interact readily with water. Cholesterol has only a single polar group, an —OH.

11.



12. The waxy surface coating is a barrier that prevents loss of water.
13. The surface wax keeps produce fresh by preventing loss of water.
14. Cholesterol is not very water-soluble, but lecithin is a good natural detergent, which is actually part of lipoproteins that transport the less soluble fats through the blood.
15. The lecithin in the egg yolks serves as an emulsifying agent by forming closed vesicles. The lipids in the butter (frequently triacylglycerols) are retained in the vesicles and do not form a separate phase.
16. The removal of the oil also removes the natural oils and waxes on the feathers. These oils and waxes must regenerate before the birds can be released.

### 8.3 Biological Membranes

17. Triacylglycerols are not found in animal membranes.
18. Statements (c) and (d) are consistent with what is known about membranes. Covalent bonding between lipids and proteins [statement (e)] occurs in some anchoring motifs, but is not widespread otherwise. Proteins “float” in the lipid bilayers rather than being sandwiched between them [statement (a)]. Bulkier molecules tend to be found in the outer lipid layer [statement (b)].
19. The public is attuned to the idea of polyunsaturated fats as healthful. The *trans* configuration gives a more palatable consistency. Recently, however, concerns have arisen about the extent to which such products mimic saturated fats.
20. Partially hydrogenated vegetable oils have the desired consistency for many foods, such as oleomargarine and components of TV dinners.
21. Many of the double bonds have been saturated. Crisco contains “partially hydrogenated vegetable oils.”
22. Less heart disease is associated with diets low in saturated fatty acids.
23. The transition temperature is lower in a lipid bilayer with mostly unsaturated fatty acids compared with one with a high percentage of saturated fatty acids. The bilayer with the unsaturated

fatty acids is already more disordered than the one with a high percentage of saturated fatty acids.

24. Myelin is a multilayer sheath consisting mainly of lipids (with some proteins) that insulates the axons of nerve cells, facilitating transmission of nerve impulses.
25. At the lower temperature, the membrane would tend to be less fluid. The presence of more unsaturated fatty acids would tend to compensate by increasing the fluidity of the membrane compared to one at the same temperature with a higher proportion of saturated fatty acids.
26. The higher percentage of unsaturated fatty acids in membranes in cold climates is an aid to membrane fluidity.
27. Hydrophobic interactions among the hydrocarbon tails are the main energetic driving force in the formation of lipid bilayers.

### 8.4 The Kinds of Membrane Proteins

28. A glycoprotein is formed by covalent bonding between a carbohydrate and a protein, whereas a glycolipid is formed by covalent bonding between a carbohydrate and a lipid.
29. Proteins that are associated with membranes do not have to span the membrane. Some can be partially embedded in it, and some associate with the membrane by noncovalent interactions with its exterior.
30. A 100-g sample of membrane contains 50 g of protein and 50 g of phosphoglycerides.

$$50 \text{ g lipid} \times \frac{1 \text{ mol lipid}}{800 \text{ g lipid}} = 0.0625 \text{ mol lipid}$$

$$50 \text{ g protein} \times \frac{1 \text{ mol protein}}{50,000 \text{ g protein}} = 0.001 \text{ mol protein}$$

The molar ratio of lipid to protein is 0.0625/0.001 or 62.5/1.

31. Nature chooses what works. This is an efficient use of a large protein and of the energy of ATP.
32. In a protein that spans a membrane, the nonpolar residues are the exterior ones; they interact with the lipids of the cell membrane. The polar residues are in the interior, lining the channel through which the ions enter and leave the cell.



**8.5 The Fluid-Mosaic Model of Membrane Structure**

33. Statements (c) and (d) are correct. Transverse diffusion is only rarely observed [statement (b)], and the term *mosaic* refers to the pattern of distribution of proteins in the lipid bilayer [statement (e)]. Peripheral proteins are also considered part of the membrane [statement (a)].

**8.6 The Functions of Membranes**

34. Biological membranes are highly nonpolar environments. Charged ions tend to be excluded from such environments rather than dissolving in them, as they would have to do to pass through the membrane by simple diffusion.

35. Statements (a) and (c) are correct; statement (b) is not correct because ions and larger molecules, especially polar ones, require channel proteins.

**8.7 Lipid-Soluble Vitamins and Their Functions**

36. Cholesterol is a precursor of vitamin D<sub>3</sub>; the conversion reaction involves ring opening.

37. Vitamin E is an antioxidant.

38. Isoprene units are five-carbon moieties that play a role in the structure of a number of natural products, including fat-soluble vitamins.

39. See Table 8.3.

40. The *cis-trans* isomerization of retinal in rhodopsin triggers the transmission of an impulse to the optic nerve and is the primary photochemical event in vision.

41. Vitamin D can be made in the body.

42. Lipid-soluble vitamins accumulate in fatty tissue, leading to toxic effects. Water-soluble vitamins are excreted, drastically reducing the chances of an overdose.

43. Vitamin K plays a role in the blood-clotting process. Blocking its mode of action can have an anticoagulant effect.

44. Vitamins A and E are known to scavenge free radicals, which can do oxidative damage to cells.

45. Eating carrots is good for both. Vitamin A, which is abundant in carrots, plays a role in vision. Diets that include generous amounts of vegetables are associated with a lower incidence of cancer.

**8.8 Prostaglandins and Leukotrienes**

46. An omega-3 fatty acid has a double bond at the third carbon from the methyl end.

47. Leukotrienes are carboxylic acids with three conjugated double bonds.

48. Prostaglandins are carboxylic acids that include a five-membered ring in their structure.

49. Prostaglandins and leukotrienes are derived from arachidonic acid. They play a role in inflammation and in allergy and asthma attacks.

50. Prostaglandins in blood platelets can inhibit their aggregation. This is one of the important physiological effects of prostaglandins.

**Chapter 9**

**9.1 Levels of Structure in Nucleic Acids**

- 1. (a) Double-stranded DNA is usually thought of as having secondary structure, unless we consider its supercoiling (tertiary) or association with proteins (quaternary).
- (b) tRNA is a tertiary structure with many folds and twists in three dimensions.

(c) mRNA is usually considered a primary structure, as it has little other structure.

**9.2 The Covalent Structure of Polynucleotides**

2. Thymine has a methyl group attached to carbon 5; uracil does not.

3. In adenine, carbon 6 has an amino group attached; in hypoxanthine, carbon 6 is a carbonyl group.

4.

A	Adenine	Adenosine or deoxyadenosine	Adenosine-5'-triphosphate or deoxyadenosine-5'-triphosphate
G	Guanine	Guanosine or deoxyguanosine	Guanosine-5'-triphosphate or deoxyguanosine-5'-triphosphate
C	Cytosine	Cytidine or deoxycytidine	Cytidine-5'-triphosphate or deoxycytidine-5'-triphosphate
T	Thymine	Deoxythymidine	Deoxythymidine-5'-triphosphate
U	Uracil	Uridine	Uridine-5'-triphosphate

5. ATP is made from adenine, ribose, and three phosphates linked the 5'-hydroxyl of the ribose. dATP is the same, except that the sugar is deoxyribose.

6. The sequence on the opposite strand for each of the following (all read 5' → 3') is ACGTAT TGCATA AGATCT TCTAGA ATGGTA TACCAT.

7. They are DNA sequences because of the presence of thymine rather than uracil.

8. (a) Definitely yes! If there is anything that you don't want falling apart, it's your storehouse of genetic instructions. (Compare the effectiveness of a computer if all the \*.exe files were deleted.)

(b) In the case of messenger RNA, yes. The mRNA is the transmitter of information for protein synthesis, but it is needed only as long as a particular protein is needed. If it were long-lived, the protein would continue to be synthesized even when not needed; this would waste energy and could cause more direct detrimental effects. Thus, most mRNAs are short-lived (minutes); if more protein is needed, more mRNA is made.

9. Four different kinds of bases—adenine, cytosine, guanine, and uracil—make up the preponderant majority of the bases found in RNA, but they are not the only ones. Modified bases occur to some extent, principally in tRNA.

10. This speculation arose from the fact that ribose has three hydroxyl groups that can be esterified to phosphoric acid (at the 2', 3', and 5' positions), whereas deoxyribose has free hydroxyls at the 3' and 5' positions alone.

11. The hydrolysis of RNA is greatly enhanced by the formation of a cyclic 2',3'-phosphodiester intermediate. DNA, lacking the 2'-hydroxyl group, cannot form the intermediate and thus is relatively resistant to hydrolysis.

**9.3 The Structure of DNA**

12.

Structure	Kind of Nucleic Acid
A-form helices	Double-stranded RNA
B-form helices	DNA
Z-form helices	DNA with repeating CGCGCG sequences
Nucleosomes	Eukaryotic chromosomes
Circular DNA	Bacterial, mitochondrial, plasmid DNA



13. See Figure 9.8.
14. Statements (c) and (d) are true; statements (a) and (b) are not.
15. Proponents of the patent system say it takes money to drive research. Companies will not want to invest hundreds of thousands, or millions, of dollars in research if they cannot get a tangible gain. Opponents believe a patent on what amounts to information stifles more research and even prevents the advancement of medicine.
16. The idea of patenting information began with a landmark case in 1972 when Ananda M. Chakrabarty, a General Electric engineer, filed for a patent on a strain of *Pseudomonas* bacteria that could break down oil slicks more efficiently.
17. Two genes related to breast cancer, *BRCA 1* and *BRCA 2*. In 2009 a group of patients, doctors, and research professionals brought a suit to invalidate those patents. They argued that the two genes are “products of nature” and should never have been patented in the first place.
18. The major groove and minor groove in B-DNA have very different dimensions (width); those in A-DNA are much closer in width.
19. Statement (c) is true. Statements (a) and (b) are false. Statement (d) is true for the B form of DNA but not for the A and Z forms.
20. Supercoiling refers to twists in DNA over and above those of the double helix. Positive supercoiling refers to an extra twist in DNA caused by overwinding of the helix before sealing the ends to produce circular DNA. A topoisomerase is an enzyme that induces a single-strand break in supercoiled DNA, relaxes the supercoiling, and reseals the break. Negative supercoiling refers to unwinding of the double helix before sealing the ends to produce circular DNA.
21. Propeller-twist is a movement of the two bases in a base pair away from being in the same plane.
22. An AG/CT step is a small section of double-stranded DNA where one strand is 5'-AG-3', and the other is 5'-CT-3'. The exact nature of such steps greatly influences the overall shape of a double helix.
23. Propeller-twist reduces the strength of the hydrogen bond but moves the hydrophobic region of the base out of the aqueous environment, thus being more entropically favorable.
24. B-DNA is a right-handed helix with specified dimensions (10 base pairs per turn, significant differences between major and minor groove, etc.). Z-DNA is a left-handed double helix with different dimensions (12 base pairs per turn, similar major and minor grooves, etc.).
25. Positive supercoils in circular DNA will be left-handed.
26. Chromatin is the complex consisting of DNA and basic proteins found in eukaryotic nuclei (see Figure 9.16).
27. Genome 10K Project proposes to sequence 10,000 genomes in the next 5 years.
28. Negative supercoiling, nucleosome winding, Z-form DNA.
29. It binds to the DNA, forming loops around itself. It then cuts both strands of DNA on one part of the loop, passes the ends across another loop, and reseals.
30. Histones are very basic proteins with many arginine and lysine residues. These residues have positively charged side chains under physiological pH. This is a source of attraction between the DNA and histones because the DNA has negatively charged phosphates: Histone-NH<sub>3</sub><sup>+</sup> attracts <sup>-</sup>O—P—O—DNA chain.
- When the histones become acetylated, they lose their positive charge: Histone—NH—COCH<sub>3</sub>. They therefore have no attraction to the phosphates on the DNA. The situation is even less favorable if they are phosphorylated because now both the histone and the DNA carry negative charges.
31. Adenine–guanine base pairs occupy more space than is available in the interior of the double helix, whereas cytosine–thymine base pairs are too small to span the distance between the sites to which complementary bases are bonded. One would not normally expect to find such base pairs in DNA.
32. The phosphate groups in DNA are negatively charged at physiological pH. If they were grouped together closely, as in the center of a long fiber, the result would be considerable electrostatic repulsion. Such a structure would be unstable.
33. The percentage of cytosine equals that of guanine, 22%. This DNA thus has a 44% G–C content, implying a 56% A–T content. The percentage of adenine equals that of thymine, so adenine and thymine are 28% each.
34. If the DNA were not double stranded, the requirement G=C and A=T would no longer exist.
35. The base distribution would not have A=T and G=C, and total purine would not be equal to total pyrimidine.
36. The purpose of the Human Genome Project was the complete sequencing of the human genome. There are many reasons for doing this. Some are tied to basic research (i.e., the desire to know all that is knowable, especially about our own species). Some are medical in nature (i.e., a better understanding of genetic diseases and how growth and development are controlled). Some are comparative in nature, looking at the similarities and differences between genomes of other species. Our DNA is at least 95% the same as that of a chimpanzee, yet we are clearly different. An understanding of our genome will help us understand what separates humankind from other primates and nonprimates.
37. Human gene therapy has many legal and ethical considerations. Some are moral and philosophical: Do we have the right to manipulate human DNA? Are we playing God? Should “tailor-made” humans be allowed? Some are more scientific: Do we have the knowledge to do it right? What happens if we make a mistake? Will a patient die that would not have died with other treatments?
38. Advantages would be that people could make informed lifestyle choices. A person with a genotype known to lead to atherosclerosis could change his or her diet and exercise habits from an early age to help fight this potential problem and could also seek preventive drug therapies. Disadvantages might involve legal issues over the right to know such information. Employers could discriminate against prospective employees based on a genotype marker that might indicate a susceptibility to drug abuse, alcoholism, or disease. A caste system based on genetics could arise.
39. Because any system involving replication of DNA by DNA polymerases must have a primer to start the reaction, the primer can be RNA or DNA, but it must bind to the template strand being read. Thus, enough of the sequence must be known to create the correct primer.

#### 9.4 Denaturation of DNA

40. A–T base pairs have two hydrogen bonds, whereas G–C base pairs have three. It takes more energy and higher temperature to disrupt the structure of DNA rich in G–C base pairs.

## 9.5 The Principal Kinds of RNA and Their Structures

41. See Figures 9.20 and 9.25.
42. Small nuclear RNA (snRNA) is found in the eukaryotic nucleus and is involved in splicing reactions of other RNA types. An snRNP is a small nuclear ribonucleoprotein particle. A complex of small nuclear RNA and protein catalyzes splicing of RNA.
43. Ribosomal RNA (rRNA) is the largest. Transfer RNA (tRNA) is the smallest.
44. Messenger RNA (mRNA) has the least amount of secondary structure (hydrogen bonding).
45. The bases in a double-stranded chain are partially hidden from the light beam of a spectrophotometer by the other bases in close proximity, as though they were in the shadow of the other bases. When the strands unwind, these bases become exposed to the light and absorb it; therefore, the absorbance increases.
46. RNA interference is the process by which small RNAs prevent the expression of genes.
47. More extensive hydrogen bonding occurs in tRNA than in mRNA. The folded structure of tRNA, which determines its binding to ribosomes in the course of protein synthesis, depends on its hydrogen-bonded arrangement of atoms. The coding sequences of mRNA must be accessible to direct the order of amino acids in proteins and should not be rendered inaccessible by hydrogen bonding.
48. They prevent intramolecular hydrogen bonding (which occurs in tRNA via the usual A–U and C–G associations), thus permitting loops that are critical for function, the most important being the anticodon loop.
49. Turnover of mRNA should be rapid to ensure that the cell can respond quickly when specific proteins are needed. Ribosomal subunits, including their rRNA component, can be recycled for many rounds of protein synthesis. As a result, mRNA is degraded more rapidly than rRNA.
50. The mistake in the DNA would be more harmful because every cell division would propagate the mistake. A mistake in transcription would lead to one wrong RNA molecule that can be replaced with a correct version with the next transcription.
51. Eukaryotic mRNA is initially formed in the nucleus by transcription of DNA. The mRNA transcript is then spliced to remove introns, a poly-A tail is added at the 3' end, and a 5'-cap is put on. This is the final mRNA, which is then transported, in most cases, out of the nucleus for translation by the ribosomes.
52. The numbers 50S, 30S, etc. refer to a relative rate of sedimentation in an ultracentrifuge and cannot be added directly. Many things besides molecular weight influence the sedimentation characteristics, such as shape and density.

## Chapter 10

### 10.1 The Flow of Genetic Information in the Cell

1. Replication is the production of new DNA from a DNA template. Transcription is the production of RNA from a DNA template. Translation is the synthesis of proteins directed by mRNA, which reflects the base sequence of DNA.
2. False. In retroviruses, the flow of information is RNA → DNA.
3. DNA represents the permanent copy of genetic information, whereas RNA is transient. The cell could survive production of some mutant proteins, but not DNA mutation.

### 10.2 Replication of DNA

4. The semiconservative replication of DNA means that a newly formed DNA molecule has one new strand and one strand from the original DNA. The experimental evidence for semiconservative replication comes from density-gradient centrifugation (Figure 10.3). If replication were a conservative process, the original DNA would have two heavy strands and all newly formed DNA would have light strands.
5. A replication fork is the site of formation of new DNA. The two strands of the original DNA separate, and a new strand is formed on each original strand.
6. An origin of replication consists of a bubble in the DNA. There are two places at opposite ends where new polynucleotide chains are formed (Figure 10.4).
7. Separating the two strands of DNA requires unwinding the helix.
8. If the original Meselson–Stahl experiment had used longer pieces of DNA, the results would not have been as clear-cut. Unless the bacteria were synchronized as to their stage of development, the DNA could have represented several generations at once.
9. Replication requires separating the strands of DNA. This cannot happen unless the DNA is unwound.

### 10.3 DNA Polymerase

10. Most DNA-polymerase enzymes also have exonuclease activity.
11. DNA polymerase I is primarily a repair enzyme. DNA polymerase III is mainly responsible for the synthesis of new DNA. See Table 10.1.
12. The processivity of a DNA polymerase is the number of nucleotides incorporated before the enzyme dissociates from the template. The higher this number, the more efficient the replication process.
13. The reactants are deoxyribonucleotide triphosphates. They provide not only the moiety to be inserted (the deoxyribonucleotide) but also the energy to drive the reaction ( $\text{dNTP} \rightarrow \text{inserted NMP} + \text{PP}_i, \text{PP}_i \rightarrow 2\text{P}_i$ ).
14. Hydrolysis of the pyrophosphate product prevents the reversal of the reaction by removing a product.
15. One strand of newly formed DNA uses the 3'-to-5' strand as a template. The problem arises with the 5'-to-3' strand. Nature deals with this issue by using short stretches of this strand for a number of chunks of newly formed DNA. They are then linked by DNA ligase (Figure 10.5).
16. The free 3' end is needed as the site to which added nucleotides bond. A number of antiviral drugs remove the 3' end in some way.
17. The large negative  $\Delta G^\circ$  ensures that the back reaction of depolymerization does not occur. Energy overkill is a common strategy when it is critically important that the process does not go in the reverse direction.
18. Nucleophilic substitution is a common reaction mechanism, and the hydroxyl group at the 3' end of the growing DNA strand is an example of a frequently encountered nucleophile.
19. Some enzymes have a recognition site that is not the same as the active site. In the specific case of DNA polymerase III, the sliding clamp tethers the rest of the enzyme to the template. This ensures a high degree of processivity.

### 10.4 Proteins Required for DNA Replication

20. All four deoxyribonucleoside triphosphates, template DNA, DNA polymerase, all four ribonucleoside triphosphates, primase, helicase, single-strand binding protein, DNA gyrase, DNA ligase.

21. DNA is synthesized from the 5' end to the 3' end, and the new strand is antiparallel to the template strand. One of the strands is exposed from the 5' end to the 3' end as a result of unwinding. Small stretches of new DNA are synthesized, still in an antiparallel direction from the 5' end to the 3' end and are linked by DNA ligase. See Figure 10.5.
22. DNA gyrase introduces a swivel point in advance of the replication fork. Primase synthesizes the RNA primer. DNA ligase links small, newly formed strands to produce longer ones.
23. In the replication process, the single-stranded portions of DNA are complexed to specific proteins.
24. DNA ligase seals the nicks in newly formed DNA.
25. The primer in DNA replication is a short sequence of RNA to which the growing DNA chain is bonded.
26. Specific enzymes exist to cut the DNA and give a supercoiled configuration at the replication fork that allows replication to proceed.
27. Polymerase III does not insert a deoxyribonucleotide without checking to see that the previous base is correct. It needs a previous base to check even if that base is part of a ribonucleotide.

### 10.5 Proofreading and Repair

28. When an incorrect nucleotide is introduced into a growing DNA chain as a result of mismatched base pairing, DNA polymerase acts as a 3'-exonuclease, removing the incorrect nucleotide. The same enzyme then incorporates the correct nucleotide.
29. In *E. coli*, two different kinds of exonuclease activity are possible for DNA polymerase I, which functions as a repair enzyme.
30. An exonuclease nicks the DNA near the site of the thymine dimers. Polymerase I acts as a nuclease and excises the incorrect nucleotides, then acts as a polymerase to incorporate the correct ones. DNA ligase seals the nick.
31. In DNA, cytosine spontaneously deaminates to uracil. The presence of the extra methyl group is a clear indication that a thymine really belongs in that position, not a cytosine that has been deaminated.
32. About 5000 books:  $10^{10}$  characters/error  $\times$  1 book/( $2 \times 10^6$  characters) =  $5 \times 10^3$  books/error.
33.  $1000 \text{ characters/second} \times 1 \text{ word/5 characters} \times 60 \text{ seconds/minute} = 12,000 \text{ words/minute}$ .
34.  $1 \text{ second/1000 characters} \times 10^{10} \text{ characters/error} \times 107 \text{ seconds/error} = 16.5 \text{ weeks/error nonstop}$ .
35. Prokaryotes methylate their DNA soon after replication. This aids the process of mismatch repair. The enzymes that carry out the process can recognize the correct strand by its methyl groups. The newly formed strand, which contains the incorrect base, does not have methyl groups.
36. DNA is constantly being damaged by environmental factors and by spontaneous mutations. If these mistakes accumulate, deleterious amino acid changes or deletions can arise. As a result, essential proteins, including those that control cell division and programmed cell death, are inactive or overactive, eventually leading to cancer.
37. Prokaryotes have a last-resort mechanism for dealing with drastic DNA damage. This mechanism, called the SOS response, includes the crossing over of DNA. Replication becomes highly error-prone, but it serves the need of the cell to survive.

### 10.6 DNA Recombination

38. Recombination that involves a reaction between homologous sequences.

39. They used two different phages to infect bacteria. One of the phages had light DNA and one had heavy. Without recombination, the light DNA would always package into light virus particles, and the heavy would package into heavy particles. This would lead to only two populations of phages after infection. Their results showed, however, that there were intermediate combinations that had DNA of different weights. This demonstrated that the phage DNA was recombining.
40. Similar to the experiment described in 39 above, using heavy isotopes demonstrated the semiconservative nature of replication. Intermediate-weight products of replication demonstrated that progeny DNA contains one parental strand and one new strand.
41. Recombination occurs by the breakage and reunion of DNA strands so that physical exchange of DNA parts takes place. The mechanism was deduced in 1964 by Robin Holliday and is referred to as the Holliday Model.

### 10.7 Eukaryotic DNA Replication

42. Eukaryotes usually have several origins of replication, whereas prokaryotes have only one.
43. The general features of DNA replication are similar in prokaryotes and eukaryotes. The main differences are that eukaryotic DNA polymerases do not have exonuclease activity. After synthesis, eukaryotic DNA is complexed with proteins; prokaryotic DNA is not.
44. Histones are proteins complexed to eukaryotic DNA. Their synthesis must take place at the same rate as DNA synthesis. The proteins and DNA must then assemble in proper fashion.
45. (a) Eukaryotic DNA replication must deal with histones; the linear DNA molecule in eukaryotes is a much larger molecule and requires special treatment at the ends.  
(b) Special polymerases are used in the organelles.
46. Eukaryotes have more DNA polymerases, which tend to be larger molecules. Eukaryotic DNA polymerases tend not to have exonuclease activity. There are more origins of replication in eukaryotes and shorter Okazaki fragments. See Table 10.5.
47. Mechanisms exist to ensure that DNA synthesis takes place only once in the eukaryotic cell cycle, during the S phase. Preparation for DNA synthesis can and does take place in the G1 phase, but the timing of actual synthesis is strictly controlled.
48. If the telomerase enzyme were inactivated, DNA synthesis would eventually stop. This enzyme maintains the 3' template end strand so that it does not undergo degradation with each round of DNA synthesis. The degradation in turn arises from the removal of the RNA primer with each round of DNA synthesis.
49. If histone synthesis took place faster than DNA synthesis, it would be highly disadvantageous to invest the energy required for protein synthesis. The histones would have no DNA with which to bind.
50. Replication licensing factors (RLFs) are proteins that bind to eukaryotic DNA. They get their name from the fact that replication cannot proceed until they are bound. Some of the RLF proteins have been found to be cytosolic. They have access to the chromosome only when the nuclear membrane dissolves during mitosis. Until they are bound, replication cannot occur. This property links eukaryotic DNA replication and the cell cycle. Once RLFs have bound, the DNA is then competent for replication.
51. It is faster in prokaryotes. The DNA is smaller, and the lack of compartmentalization within the cell facilitates the process. DNA replication in eukaryotes is linked to the cell cycle, and prokaryotic cells proliferate more quickly than those of eukaryotes.



52. In reverse transcriptase action, the single RNA strand serves as a template for the synthesis of a single DNA strand. The DNA strand, in turn, serves as the template for synthesis of the second strand of DNA.
53. Circular DNA does not have ends. This removes the necessity for maintaining the 3' template end on removal of the RNA primer. Telomeres and telomerase are not needed with circular DNA.
54. The presence of a DNA polymerase that operates only in mitochondria is consistent with the view that these organelles are derived from bacteria incorporated by endosymbiosis. The bacteria were originally free-living organisms earlier in evolutionary history.
55. The hypothesis that RNA was the original molecule of heredity, and was the first molecule that took simple compounds and turned them into larger molecules with a function.
56. Because finding that RNA can self-replicate leads credence to the RNA world hypothesis and brings us a step closer to understanding how evolution began.

## Chapter 11

### 11.2 Transcription in Prokaryotes

1. No primer is required for transcription of DNA into RNA.
2. RNA polymerase from *E. coli* has a molecular weight of about 500,000 and four different kinds of subunits. It uses one strand of the DNA template to direct RNA synthesis. It catalyzes polymerization from the 5' end to the 3' end.
3. The subunit composition for the holoenzyme is  $\alpha_2\beta\beta'\sigma$ .
4. The core enzyme lacks the  $\sigma$  subunit; the holoenzyme has it.
5. The strand that the RNA polymerase uses as a template for its RNA is called the template strand, the noncoding strand, the antisense strand, and the (–) strand. The other strand, whose sequence matches the RNA produced except for the T–U change, is called the nontemplate strand, the coding strand, the sense strand, and the (+) strand.
6. The promoter region is the portion of DNA to which RNA polymerase binds at the start of transcription. This region lies upstream (nearer the 3' end of the template DNA) of the actual gene for the RNA. The promoter regions of DNA from many organisms have sequences in common (consensus sequences). The consensus sequences frequently lie 10 base pairs and 35 base pairs upstream of the start of transcription.
7. Moving from 5' to 3' on the coding strand, the order is the following: Fis site, UP element, –35 region, Pribnow box, TSS.
8. Intrinsic termination of transcription involves the formation of a hairpin loop in the RNA being formed, which stalls the RNA polymerase over a region rich in A–U base pairs. This causes termination of transcription and release of the transcript. Rho-dependent termination often involves a similar hairpin loop, but, in addition, a Rho protein binds to the RNA and moves along it toward the transcription bubble. When the Rho protein reaches the transcription bubble, it causes termination.
9. See Figure 11.1. The top DNA strand is the nontemplate strand because it is not used to create the RNA. It is called the coding strand because it has the same sequence as the RNA produced, except for the change of T for U. It is called the sense strand because its sequence would give the correct amino acid sequence of the protein product. It is called the (+) strand again because it has the correct sequence. The bottom strand is called the template strand because it is the one used to make the RNA. It is

also the noncoding strand because its sequence does not match the RNA produced. It is the antisense and the (–) strand for the same reason.

### 11.3 Transcription Regulation in Prokaryotes

10. An inducer is a substance that leads to transcription of the structural genes in an operon. A repressor is a substance that prevents transcription of the structural genes in an operon.
11. The  $\sigma$  factor is a subunit of prokaryotic RNA polymerase. It directs the polymerase to specific promoters and is one of the ways that gene expression is controlled in prokaryotes.
12.  $\sigma^{70}$  is the normal  $\sigma$ -subunit for RNA polymerase in *E. coli*. It directs RNA polymerase to most of the genes that are transcribed under normal circumstances.  $\sigma^{32}$  is an alternate subunit that is produced when the cells are grown at higher temperatures. It directs the RNA polymerase to other genes that need to be expressed during heat shock conditions.
13. The catabolite activator protein is a transcription factor in *E. coli* that stimulates transcription of the *lac* operon structural genes. It responds to cAMP levels such that the *lac* operon is transcribed only when the cells must use lactose as a fuel source.
14. Transcription attenuation is the process found in prokaryotes in which transcription can continue or be prematurely aborted based on the concurrent translation of the mRNA produced. This is often seen in genes whose protein products lead to amino acid synthesis.
15. An operon consists of an operator gene, a promoter gene, and structural genes. When a repressor is bound to the operator, RNA polymerase cannot bind to the promoter to start transcription of the structural genes. When an inducer is present, it binds to the repressor, rendering it inactive. The inactive repressor can no longer bind to the operator. As a result, RNA polymerase can bind to the promoter, leading to the eventual transcription of the structural genes.
16. See Figure 11.5.
17. With phage SPO1, which infects the bacteria *B. subtilis*, the virus has a set of genes called the early genes that are transcribed by the host's RNA polymerase, using its regular  $\sigma$ -subunit. One of the viral early genes codes for a protein called gp28. This protein is another  $\sigma$ -subunit, which directs the RNA polymerase to preferentially transcribe more of the viral genes during the middle phase. Products of the middle phase transcription are gp33 and gp34, which together make up another  $\sigma$  factor that directs the transcription of the late genes.
18. See Figure 11.14. When the level of tryptophan is low, the *trp*-tRNA<sup>trp</sup> becomes limiting. This stalls the ribosome over the tryptophan codons on the mRNA. By stalling the ribosome there, the antitermination loop can form, transcription is not aborted, and the full mRNA is produced. If the ribosome does not stall there, the termination loop forms, and the leader mRNA dissociates.
19. It is the sensing domain of a riboswitch found at the 5' end.
20. It is mRNA that has two functions: sensing and decision making.
21. Translation can be prevented when a hairpin loop forms that blocks the translation initiation site. Another process that halts translation occurs when a terminator hairpin is created, similar to the one that forms during transcription attenuation. Also, in the presence of a certain metabolite, the mRNA can self-destruct.
22. Researchers are hoping to find molecules that can act like a competitive inhibitor and make the riboswitch act as though the natural substrate were present. If the riboswitch controlled

a vital process, then shutting off the riboswitch would kill the pathogen.

#### 11.4 Transcription in Eukaryotes

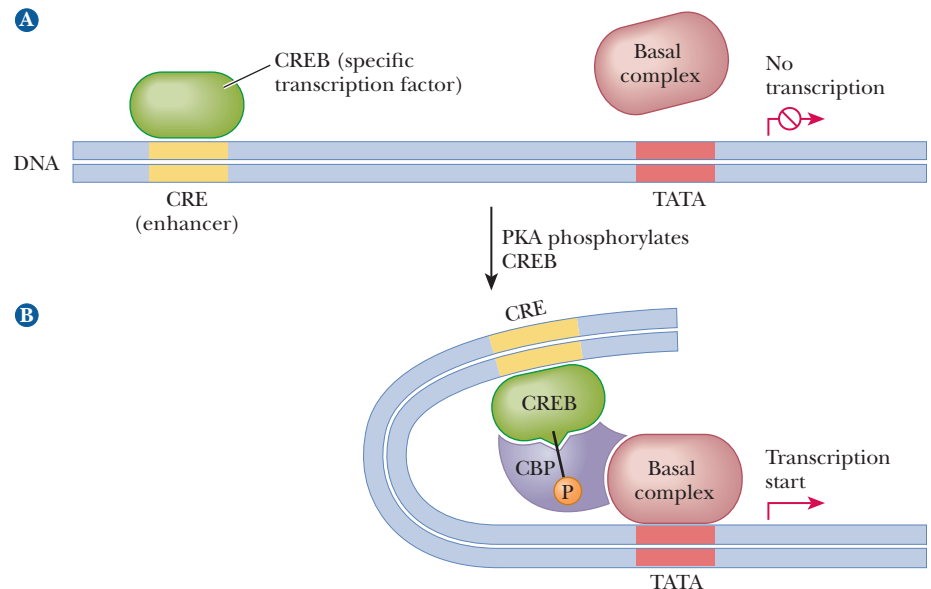
23. Exons are the portions of DNA that are expressed, which means that they are reflected in the base sequence of the final mRNA product. Introns are the intervening sequences that do not appear in the final product, but are removed during the splicing of mRNA.
24. There are three RNA polymerases in eukaryotes, compared with one in prokaryotes. There are many more transcription factors in eukaryotes, including complexes of them necessary for polymerase recruitment. RNA is extensively processed after transcription in eukaryotes, and, in most cases, the mRNA must leave the nucleus to be translated, whereas translation and transcription can occur at the same time in prokaryotes.
25. RNA polymerase I produces most of the rRNA. RNA polymerase II produces mRNA, and RNA polymerase III produces tRNA, the 5S ribosomal subunit, and snRNA.
26. The first component includes a variety of upstream elements, which act as enhancers and silencers. Two common ones are close to the core promoter and are the GC box (−40), which has a consensus sequence of GGGCGG, and the CAAT box (extending to −110), which has a consensus sequence of GGCCAATCT. The second component, found at position −25, is the TATA box, which has a consensus sequence of TATAA(T/A). The third component includes the transcription start site at position +1 and is surrounded by a sequence called the initiator element (*Inr*). The final component is a possible downstream regulator.
27. TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH are the general transcription factors. TFIID is also the TATA-box binding protein and is associated with TAFs (TBP associated factors).
28. Its primary function is as a general transcription factor involved in the formation of the open complex for transcription initiation. It binds to the basal unit and is involved in DNA melting through a helicase activity as well as promoter clearance via phosphorylation of the CTD of RNA polymerase. In addition, it also has a cyclin-dependent kinase activity. Thus, TFIIH is involved in tying transcription and cell division together. It is also involved in DNA repair mechanisms.

#### 11.5 Transcription Regulation in Eukaryotes

29. The heat-shock element responds to increased temperature. The metal-response element responds to the presence of heavy metals, such as cadmium, and the cyclic-AMP-response element controls a wide variety of genes based on cAMP levels in the cell.
30. CREB is a transcription factor that binds to the cAMP-response element. It is involved with the transcription of hundreds of genes based on the cAMP levels of the cell. When there is cAMP, CREB is phosphorylated, which allows it to bind the CREB binding protein, which connects the CRE to the basal transcription machinery, stimulating transcription.
31. Regulation in eukaryotes is much more complicated. Prokaryotic regulation is controlled by the choice of  $\sigma$ -subunit, the nature of the promoters, and the use of repressors/inducers. In eukaryotes, there are many more promoter elements, transcription

factors, and coactivators. In addition, the DNA must be released from histone proteins, so transcription of DNA is linked to histone modifications.

32. As the mRNA is being produced, ribosomes are bound and begin to translate. A leader sequence on the mRNA leads to a leader peptide. Loops can form in the mRNA in different ways. Some loop combinations lead to transcription termination. The speed with which the ribosome is able to move on the mRNA controls which loop combinations form, and this speed is usually governed by the level of a specific tRNA that is available for the translation.
33. Assuming that there is a basal transcription rate for a particular gene, an enhancer would bind to a transcription factor and lead to a greater level of transcription, while a silencer would bind to a transcription factor and reduce the level of transcription below the basal rate.
34. A response element is an enhancer element that binds to a specific transcription factor and increases the level of transcription of target genes. In the case of response elements, however, this is in response to a more general cell signal, such as the presence of cAMP, glucocorticoids, or heavy metals. Response elements may control a large set of genes, and a given gene may be under the control of more than one response element.
35. As seen here, CREB binds to the CRE. When phosphorylated, it also binds to CBP and bridges to the basal transcription complex.



36. TFIID is one of the general transcription factors for RNA polymerase II. Part of it is a protein that binds to the TATA box in eukaryotic promoters. Associated in complex with the TATA box and the TBP are many proteins called TAFs, for TBP associated factors.
37. The statement is untrue. Many eukaryotic promoters do have TATA boxes, but there are also genes that lack one.
38. Transcription elongation in eukaryotes is controlled in several ways. There are pause sites at which RNA polymerase tends to hesitate. There is also antitermination at which RNA polymerase can transcribe past a normal termination point. The general transcription factor TFIIF stimulates elongation as well as initiation by helping RNA polymerase II read through pause sites.



A separate elongation factor, TFIIS, is called an arrest-release factor because it stimulates RNA polymerase to resume transcription once it has hesitated at a pause site. Separate proteins also exist, called P-TEF and N-TEF, that act to positively or negatively affect elongation.

39. CREB is a ubiquitous transcription factor that has been found involved in many genes. It is phosphorylated when cAMP levels are high, which triggers the activation of the genes. CREB-mediated transcription has been implicated in cell proliferation, cell differentiation, spermatogenesis, release of somatostatin, development of mature T cells, protection of nerve cells under hypoxic conditions, circadian rhythms, adaptation to exercise, regulation of gluconeogenesis, transcription regulation of phosphoenolpyruvate carboxykinase and lactate dehydrogenase, and learning and storage in long-term memory.
40. Acidic domains, glutamine-rich domains, and proline-rich domains.

### 11.6 Non-Coding RNAs

41. Micro RNAs are about 22 nucleotides long, and are cut from a longer, hairpin-shaped RNA by the enzyme Dicer. (These miRNAs bind imperfectly to specific mRNAs and block their transcription.)
42. siRNAs are formed in a similar way to miRNA, by the enzyme Dicer. When a cell detects specific double-stranded RNA molecules, Dicer cuts them into small pieces of 21–25 nucleotides. These then bind to mRNA molecules in the process known as RNA interference (RNAi), targeting them for destruction.
43. ncRNAs have been linked to many processes, including regular transcription, gene silencing, replication, RNA processing, RNA modification, translation, protein stability, and protein translocation.
44. siRNAs bind to mRNA molecules, targeting them for destruction.
45. RNA silencing is believed to be an evolutionarily conserved process that is analogous to an immune system for protecting our genomes. Researchers have used a variety of techniques to establish the importance of RNA silencing to the health of the organism, including creating strains of mice that lacked the proteins that made miRNA. This resulted in a variety of health problems for the mice, including heart disease and cancer.
46. Loss of miRNA-101 leads to overexpression of a particular histone methyltransferase that helps the progression of prostate cancer.
47. In normal mice, sciatic nerve injury results in loss of nerve function in the muscle and leads to an increase in miRNA-206. Using a line of mice that had ALS and inactivated miRNA-206 led to shortening of the time from onset of the disease, indicating this miRNA had a protective effect on nerves in the muscle.
48. Work with ALS mice supports a growing body of evidence of the importance of miRNA to neurological function and susceptibility to disease. Micro RNA networks have been implicated in Parkinson's disease, Huntington's disease, and Alzheimer's disease.

### 11.7 Structural Motifs in DNA-Binding Proteins

49. Helix–turn–helix motifs, zinc fingers, and basic-region leucine zippers.
50. The major DNA binding protein motifs are helix–turn–helix, zinc fingers, and basic-region leucine zippers. The helix–turn–helix motifs are organized so that the two helices of the protein fit into the major groove of the DNA. Zinc fingers are formed by combinations of cysteine and/or histidine complexed with zinc ions. A loop of protein forms around this complex, and these loops fit into

the major groove of the DNA. Several such loops can be found spiraling around the DNA with the major groove. The basic-region leucine zipper has two domains. One is an area of leucines spaced out every seven amino acids. This puts them on the same side of an  $\alpha$ -helix, which allows them to dimerize with another such protein. The basic region is high in lysine and arginine, which bind tightly to the DNA backbone via electrostatic attraction.

### 11.8 Posttranscriptional RNA Modification

51. Introns are spliced out. Bases are modified. A poly-A tail is put on the 3' end of mRNA. A 5'-cap is put on mRNA.
52. One of the snURPs, U-1, is the target for destruction by the body's own immune system.
53. They both have multiple isoforms created by differential splicing of mRNA.
54. Trimming is necessary to obtain RNA transcripts of the proper size. Frequently, several tRNAs are transcribed in one long RNA molecule and must be trimmed to obtain active tRNAs.
55. Capping, polyadenylation, and splicing of coding sequences take place in the processing of eukaryotic mRNA.
56. The snRNPs are small nuclear ribonucleoprotein particles. They are the site of mRNA splicing.
57. Besides its traditional role in mRNA, tRNA, and rRNA, RNA serves other functions, such as splicing reactions, trimming reactions, and the peptide synthesis reaction of peptidyl transferase. It also has been shown that some small RNAs are produced; they act as gene silencers by binding to specific DNA sequences and blocking their transcription.
58. See Figure 11.34.
59. The Human Genome Project concluded that humans had far fewer genes than previously thought, yet we seem to be more biologically and biochemically complex. One possibility suggested to explain how so few genes could lead to so many proteins is that more proteins may be produced via differential splicing of mRNA. Thus, the same amount of DNA could lead to more gene products.

### 11.9 Ribozymes

60. A ribozyme is RNA that has catalytic activity without the intervention of protein at the active site. The catalytic portion of RNase P is a ribozyme. The self-splicing rRNA of *Tetrahymena* is the classic example, and it has recently been shown that the peptidyl transferase activity of the ribosome is actually a ribozyme.
61. Two mechanisms for RNA self-splicing are known. In Group I ribozymes, an external guanosine is covalently bonded at the splice site, releasing one end of the intron. The free end of the exon thus produced attacks the end of the other exon to splice the two. The intron cyclizes in the process. (See Figure 11.34.) Group II ribozymes display a lariat mechanism. The 2'-OH of an internal adenosine attacks the splice site. (See Figure 11.36.)
62. Proteins are more efficient catalysts than RNA because they have wider variations in structure and thus can tailor the active site for maximum efficiency for a given reaction.

## Chapter 12

### 12.1 Translating the Genetic Message

1. See Figure 12.1.

### 12.2 The Genetic Code

2. A code in which two bases code for a single amino acid allows for only 16 ( $4 \times 4$ ) possible codons, which is not adequate to code for 20 amino acids.

3. A degenerate code is one in which more than one triplet can specify a given amino acid.
  4. In the binding assay technique, various tRNA molecules, one of which is radioactively labeled with  $^{14}\text{C}$ , are mixed with ribosomes and synthetic trinucleotides bound to a filter. If the radioactive label is detected on the filter, then it is known that the particular tRNA bound to that triplet. The binding experiments can be repeated until all the triplets are assigned.
  5. The wobble base can be uracil, guanine, or hypoxanthine.
  6. The codons UAA, UAG, and UGA are the stop signals. These codons are not recognized by any tRNAs, but they are recognized by proteins called release factors. A release factor not only blocks the binding of a new aminoacyl-tRNA but also affects the activity of the peptidyl transferase, so that the bond between the carboxyl end of the peptide and the tRNA is hydrolyzed.
  7. Note that the sequence in the codon of mRNA is reversed because mRNA synthesis is antiparallel.
    - (a) Position 1 has an intermediate effect. For purine changes, a different amino acid results in all cases. The changes tend to be conservative, with only four of the 16 possible changes leading to hydrophobic-hydrophilic differences. For our purposes, glycine is considered neither hydrophobic nor hydrophilic. The resulting protein would have a better chance of functioning than in a second-base change, but a lesser probability than in a third-base change.
    - (b) Position 2 is the most informational: a different amino acid results from any change. In this case, however, the chances are high (75%) that the mutation would be a conservative one, with one hydrophobic amino acid replacing another one, so the protein would still have a good chance of functioning. A change involving serine or threonine (25% chance) would alter the polarity but would not introduce a charge on the side chain; the protein might still function.
    - (c) There is a high probability of a change in the type of amino acid, including differences in charge; the probability of the resulting protein having proper function is considerably lower.
    - (d) Position 3 is the least informational. There is a high probability of getting the same amino acid. The protein thus has a very good chance of functioning.
  8. The concept of wobble specifies that the first two bases of a codon remain the same, while there is room for variation in the third base. This is precisely what is observed experimentally.
  9. Hypoxanthine is the most versatile of the wobble bases; it can base pair with adenine, cytosine, or uracil.
  10. It is quite reasonable. When codons for a given amino acid have one or two nucleotides in common, a mutation is less likely to give rise to a nonfunctional protein. The survival value of such a feature guarantees its selection in evolution.
  11. An ambiguous code would allow for variation in the amino acid sequence of proteins. Consequently, there would be variation in function, including a number of nonfunctional proteins.
  12. Variations in the genetic code in mitochondria support the idea of their existence as free-living bacteria early in evolutionary history.
- ### 12.3 Amino Acid Activation
13. The hydrolysis of ATP to AMP and  $\text{PP}_i$  provides the energy to drive the activation step.
  14. Proofreading in amino acid activation takes place in two stages. The first requires a hydrolytic site on the aminoacyl-tRNA synthetase; incorrect amino acids that have become esterified to the tRNA are removed here. The second stage of proofreading requires the recognition site on the aminoacyl-tRNA synthetase for the tRNA itself. The incorrect tRNA does not bind tightly to the enzyme.
  15. The following factors ensure fidelity in protein synthesis. Aminoacyl-tRNA formation includes a high degree of enzyme specificity to connect the right amino acid to the right tRNA, proofreading in the formation of some aminoacyl-adenylates, and energy “overkill.” Other factors include proper hydrogen bonding of mRNA to the ribosome and between codon and anticodon. (The latter is a relatively slow association, allowing time for mismatches to dissociate before the peptide bond is formed.) The fidelity of protein synthesis is low compared with DNA synthesis, which has proofreading procedures in addition to energy overkill and proper base pairing. The fidelity of protein synthesis is relatively high compared with RNA synthesis, which has only energy overkill and proper base pairing.
  16. A separate synthetase exists for each amino acid, and this synthetase functions for all of the different tRNA molecules for that amino acid.
  17. The linkage of amino acids to tRNA is as an aminoacyl ester.
  18. Proofreading at the activation step allows for selection of both the amino acid and the tRNA. If proofreading took place at the level of codon-anticodon recognition, there would not be a mechanism to ensure that the correct amino acid has been esterified to the tRNA.
  19. The overall process of amino acid activation is energetically favored because of the energy contributed by the hydrolysis of two phosphate bonds. Without that input of energy, it would not be favorable.
- ### 12.4 Prokaryotic Translation
20. Peptidyl transferase catalyzes the formation of a new peptide bond in protein synthesis. The elongation factors, EF-Tu and EF-Ts, are required for binding of aminoacyl tRNA to the A site. The third elongation factor, EF-G, is needed for the translocation step in which the mRNA moves with respect to the ribosome, exposing the codon for the next amino acid. EF-P is thought to help catalyze the formation of the first peptide bond.
  21. The initiation complex in *E. coli* requires mRNA, the 30S ribosomal subunit, fmet-tRNA<sup>fmet</sup>, GTP, and three protein-initiation factors, called IF-1, IF-2, and IF-3. The IF-3 protein is needed for the binding of mRNA to the ribosomal subunit. The other two protein factors are required for the binding of fmet-tRNA<sup>fmet</sup> to the mRNA-30S complex.
  22. Attachment of the 50S ribosomal subunit to the 30S subunit in the initiation complex is needed for protein synthesis to proceed to the elongation phase.
  23. The A site and the P site on the ribosome are both binding sites for charged tRNAs taking part in protein synthesis. The P (peptidyl) site binds a tRNA to which the growing polypeptide chain is bonded. The A (aminoacyl) site binds to an aminoacyl tRNA. The amino acid moiety is the next one added to the nascent protein. The E (exit) site binds the uncharged tRNA until it is released from the ribosome.
  24. Puromycin terminates the growing polypeptide chain by forming a peptide bond with its C-terminus, which prevents the formation of new peptide bonds (see Figure 12.14).
  25. The stop codons bind to release factors, proteins that block binding of aminoacyl tRNAs to the ribosome, and to release the newly formed protein.

26. In the course of protein synthesis, mRNA binds to the smaller ribosomal subunit.
27. The Shine–Dalgarno sequence is a purine-rich leader segment of prokaryotic mRNA. It binds to a pyrimidine-rich sequence on the 16S rRNA part of the 30S ribosomal subunit and aligns it for proper translation, beginning with the AUG start codon.
28. Your friend is mistaken. The hydrogen-bonded regions contribute to the overall shape of the tRNA. Hydrogen-bonded regions are also important in the recognition of tRNAs by aminoacyl-tRNA synthetases.
29. Methionine bound to tRNA<sup>fmet</sup> can be formylated, but methionine bonded to tRNA<sup>met</sup> cannot be.
30. Different tRNAs and different factors are involved. Initiation requires IF-2, which recognizes fmet-tRNA<sup>fmet</sup> but not met-tRNA<sup>fmet</sup>. Conversely, in elongation, EF-Tu recognizes met-tRNA<sup>met</sup> but not fmet-tRNA<sup>fmet</sup>.
31. The methionine anticodon (UAC) on the tRNA base pairs with the methionine codon AUG in the mRNA sequence that signals the start of protein synthesis.
32. The fidelity of protein synthesis is assured twice during protein synthesis—the first time during activation of the amino acids and the second time during the matching of the codon to the anticodon on the mRNA.
33. (a) Activation cycles needed for a protein with 150 AA: 150.  
(b) Initiation cycles needed for a protein with 150 AA: 1.  
(c) Elongation cycles needed for a protein with 150 AA: 149.  
(d) Termination cycles needed for a protein with 150 AA: 1.
34. Four high-energy phosphate bonds per amino acid: two in aminoacyl-tRNA formation, one in elongation with EF-Tu, and one in translocation from the A to the P site, involving EF-G. Forming a peptide bond requires about 5 kcal/mol. This is an expenditure of about 30 kcal/mol peptide bonds. This is the price of low entropy and high fidelity.
35. Not very precisely. Ignoring any editing or proofreading costs, a maximum value can be calculated in terms of high-energy phosphate bonds. We will designate each phosphate bond as ~P. Four are needed per amino acid, and two are needed per ribonucleotide or deoxyribonucleotide. Therefore, four ~P per amino acid  $\times$  six ~P per codon  $\times$  six ~P per DNA triplet = 144 ~P per amino acid (approximately 1050 kcal per mole of amino acid). However, the actual value would be much less because of several factors. A single mRNA molecule can be involved in the synthesis of several to many protein molecules before it is degraded. One gene can be involved in the synthesis of many mRNA molecules, with replication taking place only once per cell generation. In addition, rRNA and tRNA are relatively long-lived and available for repeated protein syntheses.
36. The fact that peptidyl transferase is one of the most conserved sequences in all of biology may indicate that it evolved very early in evolution and that it is so critical for all living organisms that it cannot be modified.
37. The less highly purified ribosome preparations contained polyosomes, which are more active in protein synthesis than single ribosomes.
38. At first, peptide-bond formation was catalyzed by RNA. In time, as protein catalysts developed and became more efficient, proteins became an integral part of the ribosome.
39. Electron microscopy can give information about ribosomal structure and function, but X-ray crystallography has given far more detailed information.
40. Because the tRNAs are bound in proximity to each other on the ribosome, the growing polypeptide chain and the amino acid to be added are also close to each other. This facilitates formation of the next peptide bond.
41. A virus takes over the protein-synthesizing machinery of the cell. It uses its own nucleic acids and the cell's ribosomes.
42. Pro-X-Thr is conserved in RF-1 and Ser-Pro-Phe is conserved in RF-2.
43. The sequence Gly-Gly-Gln.
44. The other amino acids found in proteins are created by modifying one of the twenty standard amino acids after the protein is made. Selenocysteine is formed while the amino acid is bound to tRNA. Thus, this amino acid is inserted into the protein during translation just like the standard 20 amino acids are.
45. It is unique chemically because it has a selenium ion in it, which takes the place of a sulfur in cysteine. It is also unique in that appears to be coded for in the DNA sequence, even though the codon would normally be a stop codon.

### 12.5 Eukaryotic Translation

46. Similarities between protein synthesis in bacteria and protein synthesis in eukaryotes: same start and stop codons; same genetic code; same chemical mechanisms of synthesis; interchangeable tRNAs. Major differences: in prokaryotes, the Shine–Dalgarno sequence and no introns; in eukaryotes, the 5'-cap and 3'-tail on mRNA and introns have been spliced out.
47. The original N-terminal methionine can be removed by post-translational modification.
48. Puromycin would be useful for treatment of a viral infection, but chloramphenicol would not. Viral mRNAs are translated by eukaryotic translation systems, so one must use an antibiotic active on eukaryotic systems.
49. Protein synthesis in prokaryotes takes place as a coupled process with simultaneous transcription of mRNA and translation of the message in protein synthesis. This is possible because of the lack of compartmentalization in prokaryotic cells. In eukaryotes, mRNA is transcribed and processed in the nucleus and only then exported to the cytoplasm to direct protein synthesis.
50. Some mutations can introduce stop codons. It is useful to a cell to have some mechanism to suppress the formation of incomplete proteins.
51. New protein synthesis is involved in long-term memories.
52. A critical role is played by the transcription factor CREB (Chapter 11) in turning short-term memories into long-term ones.
53. Animals given drugs that block protein synthesis cannot form new long-term memories, yet their ability to make short-term memories is preserved.
54. The strength, whether due to a single strong impulse or repeated impulses, depolarizes the cell membrane of the nerve. Only ones that are strong enough will be passed along to a receiving neuron. This controls whether a memory can become long-term.

### 12.6 Posttranslational Modification of Proteins

55. Hydroxyproline is formed from proline, an amino acid for which there are four codons, by posttranslational modification of the collagen precursor.

### 12.7 Protein Degradation

56. Ubiquitin is a small polypeptide (76 amino acids) that is highly conserved in eukaryotes. When ubiquitin is linked to a protein, it marks that protein for degradation in a proteasome.



57. If proteins to be degraded did not have some signal marking them, the process would take place more randomly and thus be less efficient.
58. If protein degradation took place at any location in a cell, indiscriminate breakdown of functional proteins could take place, so this is an unlikely occurrence. It is much more useful to the cell to have a mechanism for tagging proteins to be degraded and to do so at a specific location in the cell.
59. They are specific sequences that tell the splicing machinery where to splice out introns. A mutation in an ESE could lead to incorrect removal of introns and an entire exon being left out of the final mRNA.
60. A silent mutation is a change in the DNA sequence of a codon that should lead to the same amino acid being inserted. It is a misnomer because we now know that sometimes differences in codons for the same amino acid do affect the overall protein product.
61. If the silent mutation is in an exonic splicing enhancer, then the splicing out of introns could be incorrect and the correct exon could be skipped.
62. Silent mutations in the mRNA for the enzyme control secondary structures of the mRNA, which controls fast and how often the mRNA is translated, leading to different levels of the enzyme related to pain tolerance.
63. Marfan syndrome, androgen-insensitivity syndrome, cholesteryl ester storage disease, McArdle disease, and phenylketonurea.
9. They are all palindromes (ignoring punctuation and spacing in the latter two cases), analogous to palindromic sequences of bases in DNA. Just as the five examples are distinguished by being pronounced differently, different palindromes in DNA are distinguished and acted on by different, very specific restriction endonucleases.
10. GGATCC, GAATTC, AAGCTT (remember that these are listed 5' to 3', so you must read the complementary strand 5' to 3' to see that the sequence is the same).
11. *Hae*III cuts at a sequence of four bases, cuts in the middle of the sequence, and leaves blunt ends. *Bam*HI cuts at a sequence of six bases, cuts on the second base from the 5' end, and leaves sticky ends.
12. Sticky ends are short regions of single-stranded DNA extending from the ends of double-stranded DNA molecules. These are produced by some restriction enzymes or can be added chemically to blunt-ended double-stranded DNA. They are important because they provide a means for DNA from different sources (e.g., "foreign" gene and plasmid, both containing sticky ends) to find each other by hydrogen bonding between complementary bases. A ligase is then used to covalently link the two molecules.
13. An advantage of using *Hae*III is that it yields blunt ends. Thus, one could combine DNA cut with this enzyme with any other DNA that also had blunt ends. Enzymes exist that quickly remove the sticky overhangs from other restriction enzymes. The disadvantage is that *Hae*III is specific for a four-base sequence that is likely to occur many times in a genome, so the target DNA may also be cleaved somewhere in the middle. Also, the blunt ends make it more difficult to get specific ligation of two DNA types.

## Chapter 13

### 13.1 Purification and Detection of Nucleic Acids

1. Safety, no need for special licensing, and convenience of disposal.
2. DNA is labeled with  $^{32}\text{P}$  and run on a gel. The gel is placed next to X-ray paper, which is then developed. The radioactivity shows up as black bands on the X-ray paper. This is called an autoradiograph.
3. The DNA run on electrophoresis gels is usually cleaved with restriction enzymes to give linear pieces; thus the shape is uniform for DNA. The charge is a constant for DNA in that every nucleotide has the same charge due to the phosphate groups; thus, DNA has a uniform shape and a uniform charge-to-mass ratio, so it separates solely on size, with the shorter fragments traveling fastest through the gel.

### 13.2 Restriction Endonucleases

4. The use of restriction endonucleases with different specificities gives overlapping sequences that can be combined to give an overall sequence.
5. Restriction endonucleases do not hydrolyze a methylated restriction site.
6. The restriction site of the DNA of the organism that produces a restriction endonuclease is modified, usually by methylation.
7. The restriction fragments of different sizes (restriction-fragment length polymorphisms, or RFLPs) that come about as a result of different base sequences on paired chromosomes were used as genetic markers to determine the exact position of the cystic fibrosis gene on chromosome 7.
8. An endonuclease is an enzyme that cuts nucleic acid chains in the middle, as opposed to cleaving from the ends inward. The term *restriction* came from the restricted growth seen in host cells that are infected by bacteriophages when the bacteria have restriction enzymes that can cleave the viral DNA.

### 13.3 Cloning

14. A portion of exogenous DNA is introduced into a suitable vector, frequently a bacterial plasmid, and many copies of the DNA are produced when the bacteria grow. Viruses are also commonly used as vectors.
15. The most common vectors are bacterial plasmids. Viruses and cosmids can also be used, depending on the size of the foreign DNA that must be inserted.
16. The plasmid to be used as a vector needs markers both for uptake of the target DNA sequence into the plasmid and for insertion of the plasmid into host cells. Typically, a plasmid has a gene for ampicillin resistance. Only cells that have taken up a plasmid can grow on ampicillin plates. The foreign DNA is usually inserted into a second marker to select for those plasmids that took up the target DNA. This second marker may be another antibiotic-resistant gene or some other gene, such as the  $\beta$ -galactosidase gene.
17. The key feature of a plasmid capable of blue/white screening is the gene for the  $\alpha$ -subunit of the enzyme  $\beta$ -galactosidase. These plasmids are used with a strain of *E. coli* that are deficient in the  $\alpha$ -subunit of this enzyme.  $\beta$ -Galactosidase can convert a colorless sugar derivative, called X-gal, to a blue one. The site for cleavage of the plasmid by a restriction endonuclease lies within the  $\beta$ -galactosidase gene. Cells that have acquired a plasmid can grow on ampicillin. If the plasmid reclosed on itself without the target DNA, the colonies that took up that plasmid grow blue. Cells that have acquired the DNA insert cannot produce a blue color.
18. Restriction enzymes to cut DNA, DNA ligase to rejoin DNA, a suitable vector to carry the foreign DNA, a cell line to accept the vector, and a way of selecting for the correct transformants.

19. Since most recombinant DNA occurs with bacterial and viral vectors, a big concern is that a mutated virus or bacteria will be released that can infect other species and that may be resistant to drugs, thereby creating a new, potentially lethal disease. Precautions include frequent sterilization of cultures to make sure that they are all dead before disposal, working in laminar hoods that isolate the recombinant DNA from the outside, and care in the choice of vectors. Some vectors that are replication-deficient outside certain cell types are used so that they cannot replicate outside the lab environment.

### 13.4 Genetic Engineering

20. To increase disease resistance, resistance to pests, shelf life, level of nitrogen fixation (protein content), and resistance to temperature extremes.
21. Insulin, human growth hormone, tissue plasminogen activator, enterokinase, erythropoietin, and interferon.
22. The corn being grown in the field has been genetically engineered. The gene that was introduced came from the bacterium *Bacillus thuringiensis*.
23. LDH 3 has the subunit composition  $H_2M_2$ . Each of the subunits is coded for by a separate gene, so in order to clone LDH 3, one would have to clone the gene for the M subunit and the gene for the H subunit. These would be separate cloning experiments. Each gene would be cloned into an expression cell line, and the proteins would be expressed. The individual subunits could then be combined, and they would form tetramers, some of which would be LDH 3. This could be verified by native gel electrophoresis.
24. An expression vector, such as pET 5 plasmid, has the components of any normal cloning vector (e.g., origin of replication, selectable marker, multiple cloning site), but it also has the ability to have the inserted DNA be transcribed. It has a promoter for RNA polymerase, such as T7 polymerase, and a termination sequence. These border the multiple cloning site. These vectors are used with a cell line that makes T7 RNA polymerase when induced.
25. A fusion protein is a combination of a protein coded for by an expression vector and the target gene. A common one is a histidine tag and enterokinase, which are linked to the target protein when transcribed and translated. They are used to help with the eventual purification of the target protein. The over-expressed target protein can be quickly separated from the rest of the host's proteins by purifying the fusion protein, which has characteristics that make it easy to purify.
26. The bovine growth hormone is a protein that is denatured and digested in the intestinal tract. Also, all cow's milk contains some of the hormone.
27. The DNA sequence to be inserted in the bacterial plasmid to direct the production of  $\alpha$ -globin should be cDNA, which is a sequence complementary to the mRNA for  $\alpha$ -globin. The cDNA can be produced on the mRNA template in a reaction catalyzed by reverse transcriptase.
28. Isolate the DNA that codes for the growth factor by means of suitable probes. Introduce the DNA into a bacterial genome. Allow the bacteria to grow and to produce human growth hormone.
29. The public is concerned about contamination with prions, which come from mammalian sources. If a mammalian protein can be expressed in large quantities in bacteria, there will be no risk of prion contamination.

### 13.5 DNA Libraries

30. A DNA library is a collection of cells that carry cloned pieces of the entire DNA genome of an organism. A cDNA library is made by taking the mRNA from an organism, converting it to cDNA, and cloning that for the library. In this way, the active DNA sequence is stored.
31. If a DNA library is to represent the total genome of an organism, it must contain at least one clone for each DNA sequence. This requires several hundred thousand separate clones to ensure that every sequence is represented.
32. The amount of work involved in constructing a DNA library makes it desirable to have such libraries available to the entire scientific community, thus avoiding duplication of effort.

### 13.6 The Polymerase Chain Reaction

33. The polymerase chain reaction depends on repeated cycles of separation of DNA strands followed by annealing of primers. The first step requires a significantly higher temperature than the second, giving rise to the requirement for strict temperature control.
34. Part of the procedure of the polymerase chain reaction requires the use of high temperatures. When a temperature-stable RNA polymerase is used, there is no need to add fresh batches of enzyme for each round of amplification. This would need to be the case, however, if the RNA polymerase could not withstand the high temperatures.
35. Good primers have similar G–C contents for the forward and reverse reactions, have minimal secondary structure possibilities with each other or with themselves, and are long enough to give sufficient specificity for the gene to be duplicated without costing too much.
36. The contaminating DNA as well as the desired DNA is amplified at each stage of the polymerase chain reaction, giving rise to an impure product.
37. (a) The primers have very different G–C contents.  
(b) The forward primer will have significant secondary structure with itself (hairpin loop due to inverted Gs and Cs on end).  
(c) The forward and reverse primers will bind to each other.
38. It is a technique that allows the PCR reaction to generate time-point data that can be used to determine how much of the DNA was in the cell originally.
39. Regular PCR is designed to create large quantities of DNA, so the reaction is allowed to go to completion. With qPCR, the reaction does not go to completion as it is the time-point data that are needed in order to determine the amount of the starting material.

### 13.7 DNA Fingerprinting

40. The polymerase chain reaction can increase the amount of a desired DNA sample by a considerable factor, making possible definite identification of DNA samples that were too small to be characterized by other means. It can be used on hair and blood samples found at the scene of a crime to establish the presence of a suspect. This method can also be used to identify remains of possible murder victims.
41. It is easier to show that two DNA samples do not match than to prove that they are identical.

### 13.8 Sequencing DNA

42. 5'GATGCCTACG3'
43. Two factors are involved here. First, large polymers must be cleaved into smaller, manageable fragments for sequencing.



Enzymes (endoproteases) that cleave proteins, while showing some specificity, are far from absolutely specific, and messy mixtures result. On the other hand, restriction endonucleases are absolutely specific for palindromic base sequences in DNA, and “clean” cuts result, allowing easier purification. (Note that if the gene for a protein isn’t available but the mRNA is, the resulting cDNA can be made using reverse transcriptase.) A second factor is that only relatively short fragments of protein can be sequenced without additional internal cleavage. For example, the Edman degradation is limited to peptides of about 50 amino acids or fewer. With DNA, the dideoxy method coupled with polyacrylamide-gel separation can handle DNA fragments 10 to 20 times longer.

44. DNA often has introns in the gene, so knowing the DNA sequence may give the wrong answer for the final protein sequence. Also, proteins are modified posttranslationally, so there may be modifications to the protein sequence not reflected in the DNA.
45. Open-ended answer.
46. *Benefits:* A person at risk for future heart disease could be more careful with diet and exercise. Such a person might also take a drug beforehand that would help prevent the condition from developing. Doctors with access to such information would be able to make better diagnoses and to suggest quicker treatments.  
*Detriments:* Employment could be based on a preconceived idea of what a good genotype is. Health and life insurance could be denied to people considered to have a risky genotype. A new type of prejudice against the “genotypically challenged” could arise.

### 13.9 Genomics and Proteomics

47. The genome is the total DNA of a cell, containing all the genes of that organism. The proteome is the total complement of proteins.
48. A proteomic analysis has been done on the fruit fly *Drosophila melanogaster*.
49. Using robotic technology, a slide or “chip” is loaded with thousands of specific single-stranded DNA sequences. RNA is collected from samples to be tested and converted to cDNA carrying a fluorescent tag. The sample is placed over the chip and the cDNA allowed to bind. A fluorometer measures the fluorescence from the chip and indicates which DNA sequences were bound with their corresponding cDNA. This tells researchers which genes were active as only the active genes would produce RNA.
50. Yeast could be grown under the two conditions and the mRNA collected. The mRNA could then be converted to cDNA and each population could be labeled with a different color fluorescent marker. These samples could then be overlaid on a gene chip containing the yeast genome. The color of the fluorescence on the gene chip would then tell which genes were active under the two conditions.
51. Cancerous cells have altered metabolism at the genetic level. The gene expression patterns seen in patients with known types of cancer act like a fingerprint of that type of cancer. Tissue samples from patients to be diagnosed can be used to collect the RNA and convert it to cDNA. These cDNA samples are then overlaid on a gene chip of the human genome and the binding pattern analyzed through fluorescence. The pattern seen can then be compared to the patterns seen in the known cancers to aid in the diagnosis.
52. DNA microarrays have thousands of bound single-stranded DNA spots. They are used to test for the presence of the corresponding mRNA in a biological sample via cDNA produced from the mRNA. Protein arrays, on the other hand, have applied samples of very specific and pure antibodies. Biological tissue samples are placed on the protein chip. If the antigens for the specific antibodies are present, they bind to the antibodies. Another set of antibodies with fluorescent labels is then added, and the chip analyzed with a fluorometer. The patterns seen show which antigens the tissue sample had, which can be used to diagnose the patient.

## Chapter 14

### 14.1 Viruses

1. Some viruses have DNA and some have RNA. In some cases, a viral genome is single-stranded and in others it is double-stranded.
2. (a) The virion is the entire virus particle.  
(b) The capsid is the protein coat that surrounds the viral nucleic acid.  
(c) The nucleocapsid is the combination of the nucleic acid and the capsid.  
(d) A protein spike is a membrane-bound protein that is used to help the virus attach to its host.
3. The main factors determining the family of a virus is whether its genome is DNA or RNA and whether it has a membrane envelope. Whether the nucleic acid is single- or double-stranded and the method of incorporation of the virus are also considered.
4. The virus attaches to a specific protein on the host cell’s membrane and injects its nucleic acid inside the cell.
5. In the lytic pathway, the viral nucleic acid is replicated in the host cell and packaged into new virus particles that lyse the host cell. In the lysogenic pathway, the viral DNA is incorporated into the host DNA.
6. There is no correlation. Some viruses, such as Ebola virus, are fast acting and very lethal; others, such as HIV, are slow and just as lethal. The influenza virus is fast-acting, but it is rarely lethal these days.
7. One good choice would be a drug that attacks one of the specific protein spikes on the virus. This may be an antibody that attacks it, or a drug that blocks its ability to attach to the host cell. Another choice would be a drug that inhibits a key viral enzyme, such as the reverse transcriptase of a retrovirus, or the enzymes involved in repackaging the viruses.
8. Viruses can often switch from one pathway to another, based on the condition of the host cells. If the host is healthy, there is sufficient material to allow the virus to replicate and to produce new virions. If the host cell is starved or unhealthy, there may be insufficient energy and material to do so. In this case, lysogeny allows the DNA to incorporate in the host cell, where it can wait until the cell’s health improves.
9. One example would be someone who had helper T-cells lacking a CD4 receptor. The HIV virus must bind to the CD4 receptor as part of its attachment process.
10. AIDS is a relatively new disease, having been around since the early 1980s. Flu has been around for centuries and has killed many more people than AIDS.
11. One is called hemagglutinin (HA). The second is neuraminidase (NA).

12. H1N1 is very transmissible, but it causes fewer fatalities of the people that get it than H5N1. H5N1 is not very transmissible, but is deadly to those that catch it.
13. Mutations occur frequently with viruses, and the biggest worry is that a strain with a high mortality could mutate into one that is also very transmissible.

#### 14.2 Retroviruses

14. A retrovirus has an RNA genome that must pass through a stage in which it is reverse-transcribed to DNA, and this DNA must recombine with the host's DNA.
15. Reverse transcriptase.
16. The first is that retroviruses have been linked to cancer. The second is that human immunodeficiency virus (HIV) is a retrovirus. The third is that retroviruses can be used in gene therapy.
17. Gene therapy is the process of introducing a gene into the cells of an organism that was missing functional copies of the gene.
18. Ex vivo gene therapy, in which the cells are removed from the patient before being infected with the virus carrying the therapeutic gene, and in vivo gene therapy, in which the patient is directly infected with the virus carrying the gene.
19. The two most common are the Maloney murine leukemia virus (MLLV) and adenovirus. Both must be manipulated so that the critical genes for replication are removed and replaced with an expression cassette containing the therapeutic gene.
20. When retroviruses, such as MMLV, are used, there is the danger that the therapeutic gene will incorporate in a place that will disrupt another gene. In more cases than would be predicted by random chance, this seems to occur in a place that disrupts a tumor-suppressor gene, causing cancer. There is also the danger that the patient will have a strong reaction to the virus used to introduce the therapeutic gene. In at least one case, this has had fatal consequences.
21. The biggest consideration is where the therapeutic gene has to go. Some viruses are very specific to their target cells, so if the problem is in the lungs, then a virus that is good at attacking lung cells, such as adenovirus, is a good choice. In this case, in vivo delivery would be superior, because lung cells cannot be removed from the body and then replaced. However, if the problem is in an immune cell, then bone marrow cells can be removed and transformed and later given back to the patient, making ex vivo delivery an option.
22. There are dangers inherent to all forms of gene therapy. People who have SCID have such compromised immune systems that they cannot lead normal lives, and few other remedies allow them to lead normal lives. That made SCID a prime candidate for experimental techniques. Diabetes can be controlled effectively by other techniques that are well established and not as risky.

#### 14.3 The Immune System

23. AIDS is the most well-known problem of a malfunctioning immune system. SCID is also high on the list. All allergies are immune system problems, as are autoimmune diseases. Many forms of diabetes are caused by an autoimmune disease in which a person's pancreatic cells are attacked by the immune system.
24. Innate immunity refers to a variety of protective processes, including skin, mucus, and tears as a first line of defense, and dendritic cells, phagocytes, macrophages, and natural killer cells as a second line of defense. These are always present, and the innate-immunity cells are always circulating in the body.

Acquired immunity refers to the processes involving B cells and T cells, in which specific sets of them are activated in response to an antigen challenge, and these subsets then multiply.

25. One part includes physical barriers, such as skin, mucus, and tears. The cells of the innate immune system are dendritic cells, macrophages, and natural killer (NK) cells.
26. B cells, which make antibodies, killer T cells, which attack infected cells, and helper T cells, which help activate B cells.
27. MHCs are receptors on antigen-presenting cells. They bind to fragments of antigens that have been degraded by the infected cell and display it on their surface. T cells then bind to the infected cells.
28. Clonal selection refers to the process in which a particular T cell or B cell is stimulated to divide. Only the one bearing the correct receptor for the antigens being presented is selected.
29. The cells of the innate system initially attack a pathogen, such as a virus, bacteria, or even a cancerous cell. They then present antigens from the pathogen on their surfaces via their MHC proteins. The acquired immunity cells then recognize the MHC/antigen complex, bind to it, and begin the involvement of the acquired immunity system.
30. Interferon is a cytokine produced in very small quantities that stimulates natural killer cells, which attack cancerous cells. One of the first treatments for cancer was to give the patient interferon to stimulate NK cells. Having a large supply of cloned interferon is helpful, therefore, in fighting cancer.
31. When T cells and B cells are developing, they are, in a sense, "trained." If they contain receptors that recognize self-antigens, they are eliminated when they are still young. If they don't ever see any antigens they recognize, then they die by neglect. This leaves a set of precursors to T cells and B cells with receptors that recognize foreign antigens but not self-antigens.
32. Macrophages, part of the innate immune system, are the "double-edged sword." Their presence is important to attack cancer cells, and if they do a thorough job, then the cancer cells are all destroyed. However, they also cause inflammation, which has recently been shown to indirectly lead to the progression of the cancer cells that survive.
33. The small noncoding RNA (ncRNA) of the herpes virus has been linked with its ability to evade the immune system.
34. The herpes virus produces a ncRNA that stabilizes the respiratory chain of the mitochondria of the host cell. This prevents the early destruction of the infected cell by the host's immune system. At the same time, a micro RNA (miRNA) produced by the virus inhibits production of a protein on the surface of the cell that would otherwise attract NK cells.
35. The part of the hemagglutinin protein bound by the antibodies is relatively constant and does not change between strains.
36. They tested their antibodies in mice both before and after they were dosed with lethal quantities of avian flu. Most of the rodents survived, indicating that these antibodies would work as prevention or as cure.

#### 14.4 Cancer

37. Cancer cells continue to grow and divide in situations in which normal cells do not, such as when they are not receiving growth signals from surrounding cells. They also continue to grow even if surrounding tissues are sending out "stop growth" signals. Cancer cells can co-opt the body's vascular system, causing the growth of new blood vessels to supply the cancerous cells with

- nutrients. Cancer cells are essentially immortal. They can continue to grow and to divide indefinitely. Cancer cells can break loose, travel to other parts of the body, and create new cancerous areas, a process known as metastasis.
38. A tumor suppressor is a molecule that restricts the ability of a cell to grow and to divide. An oncogene is a gene whose product stimulates a cell to grow and to divide.
  39. The protein called p53 is a tumor suppressor. Mutations of p53 have been found in more than half of all human cancers. Ras is involved in cell division, and mutations in this protein are involved in 30% of human tumors.
  40. Viruses have been implicated in many cancers. Retroviruses are particularly dangerous because they insert their DNA into the host's DNA. When this happens in a tumor-suppressor gene, the tumor suppressor is inactivated, causing cancer. Also, the homology between proto-oncogenes and oncogenes makes it likely that the infection cycle of viruses may be responsible for some proto-oncogenes becoming oncogenic.
  41. Virotherapy is the process of using a virus to attempt to treat cancer. There are two strategies for virotherapy. One is to use the virus to attack and kill cancer cells directly. In this case, the virus has a protein on its surface that is specific for a cancer cell. Once inside, it kills the cancer cell. The second is to have the virus ferry a gene into the cancer cell that makes the cell more susceptible to a chemotherapeutic agent.
  42. If smoking caused cancer, then everyone who smokes would have cancer, but this is not true. Smoking has been linked to cancer, and it is a strong predictor of future cancer, but cancer is the result of many things going wrong in a cell, and there is no single, definitive cause.
  43. A tumor suppressor is a protein that helps control cell growth and division. It is like the brakes on a car, trying to slow down a process. Many cancers are related to mutation of tumor suppressors. An oncogene produces something that stimulates growth and division. This is like the accelerator of the car. Many other cancers are caused ultimately by overactivation of an oncogene.
  44. Ras, Jun, and Fos are all considered oncogenes. In the process of cell division, Ras is a necessary component, but it is usually active only when the cell should be dividing. Oncogenic forms of Ras are overactive and lead to too much cell division. Ras is an early step in the process. Jun and Fos are transcription factors that together make up AP-1, which is involved in the transcription activation pathway involving CBP.
  45. Many of the early trials involved specific delivery of an active p53 gene via gene therapy. However, such delivery is impractical for human patients in many cases. Now researchers are looking for drugs that can be taken that will increase the levels of p53.
  46. Two drugs, Prima-1 and CP-31398 reactivate mutant p53; nutlins inhibit a protein called MDM2, which is itself a natural inhibitor of p53.
  47. p53 can be restored in several ways. One way would be through gene therapy to give the patient functioning copies of the p53 gene if he or she lacks it. Another is to neutralize molecules that naturally inhibit p53. Another is to give the patient drugs that stimulate the production of p53 by stimulating the transcription of the p53 gene. Finally, one could use drugs that would inhibit the transcription of molecules that act as inhibitors of p53.
  48. The innate immune system is instrumental in fighting cancer cells. Cells that turn cancerous display specific molecules on their surfaces that act as a help signal. Cells of the innate immune system such as macrophages and natural killer cells attack cells that display these cancer-linked antigens on their surfaces. Often they destroy the cancerous cell, ending the threat. However, if they do not, the presence of the innate immune cell can lead to inflammation. More and more research is showing that inflammation is the switch that takes a precancerous cell and turns it into a full-fledged cancer cell. Thus, innate immune cells that attack a cancer cell but fail to kill it may just make it stronger.
  49. The realization that cancer's progression is fueled by inflammation has led to a theory by some scientists that we should spend more time focusing on the symptoms rather than the cure. They believe it is possible that even though potential cancer cells exist, they may not ever grow and spread if we could stop the inflammation.
  50. They found more than 30,000 mutations in the melanoma genome and more than 23,000 in lung cancer. This information will make it possible to diagnose cancer much earlier and will lead to more effective treatment. For individual patients, it will be possible to see which drugs are likely to be effective in treating the cancer and which ones are not.
  51. Most of the mutations associated with melanoma arise from too much exposure to the sun. Likewise, smoking causes most of the DNA errors in lung cancer.

## Chapter 15

### 15.1 Standard States for Free-Energy Changes

1. There is a connection, and it is one of the most important points in this chapter. It can be expressed in the equation  $\Delta G^\circ = -RT \ln K_{\text{eq}}$ .
2. Reaction (a) would take place only if it is coupled to an exergonic reaction. Reaction (b) would proceed only if coupled to an exergonic reaction. Reaction (c) would proceed as written.
3. The information given here deals with the thermodynamics of the reaction, not the kinetics. It is not possible to predict the rate of the reaction.

### 15.2 A Modified Standard State for Biochemical Applications

4. The usual thermodynamic standard state refers to pH = 0. This is not very useful in biochemistry.
5. Statement (a) is true, but statement (b) is not. The standard state of solutes is normally defined as unit activity (1 M for all but the most careful work). In biological systems, the pH is frequently in the neutral range (i.e.,  $\text{H}^+$  is close to  $10^{-7}$  M); the modification is a matter of convenience. Water is the solvent, not a solute, and its standard state is the pure liquid.
6. The designation  $\Delta G^\circ'$  indicates a biological standard state. If the prime is omitted, then it is for chemical standard states.
7. No, there is no relationship between the thermodynamic quantity  $\Delta G^\circ$  and the speed. The  $\Delta G^\circ$  reflects the thermodynamic possibility under standard states. Speed is a kinetic quantity that is based on the ability of an enzyme to catalyze the reaction and the real substrate concentrations in the cell.
8. Assuming one significant figure,  $20 \text{ kJ mol}^{-1}$ ,  $0 \text{ kJ mol}^{-1}$ ,  $+30 \text{ kJ mol}^{-1}$ .
9.  $\Delta G^\circ' = \Delta H^\circ' - T\Delta S^\circ'$  and  $\Delta S^\circ' = 34.9 \text{ J mol}^{-1} \text{ K}^{-1} = 8.39 \text{ cal mol}^{-1} \text{ K}^{-1}$ . There are two particles on the reactant side of the equation and three on the product side, representing an increase in disorder.



10. Assuming 298 K and one significant figure:  
 (a)  $-50 \text{ kJ}$   
 (b)  $-20 \text{ kJ}$   
 (c)  $-20 \text{ kJ}$
11. The levels of substrates and products can affect the true  $\Delta G$  of a reaction, changing it from zero to a high number as in part (a).  $\Delta G$  is negative when there is a larger amount of substrate than product.
12. The overall  $\Delta G^{\circ'} = -260.4 \text{ kJ mol}^{-1}$  or  $-62.3 \text{ kcal mol}^{-1}$ . The reaction is exergonic, because it has a large, negative  $\Delta G^{\circ'}$ .
13. Greater than 3333 to 1.
14. Reaction (a) will not proceed as written;  $\Delta G^{\circ'} = +12.6 \text{ kJ}$ . Reaction (b) will proceed as written;  $\Delta G^{\circ'} = -20.8 \text{ kJ}$ . Reaction (c) will not proceed as written;  $\Delta G^{\circ'} = +31.4 \text{ kJ}$ . Reaction (d) will proceed as written;  $\Delta G^{\circ'} = -18.0 \text{ kJ}$ .
15. Yes, if you correct for the difference in temperature and concentrations from the standard values.
16. Two aspects are involved here. (a) Very rarely, if ever, are in vivo concentrations standard concentrations; actual  $\Delta G$  (not  $\Delta G^{\circ}$ ) values are very dependent on local concentrations, especially if the number of reactant molecules and product molecules is not the same. (b) Values of  $\Delta G^{\circ}$  rigorously apply only to *closed* systems that can reach equilibrium. Biochemical systems, however, are *open* systems and do not reach equilibrium. If you were at equilibrium, you would be dead. Metabolic pathways involve series of reactions, and the metabolic pathways themselves are interconnected, including processes that take in materials from the surroundings and release waste products to the surroundings.

### 15.3 The Nature of Metabolism

17. Group 1: catabolism, oxidative, energy-yielding. Group 2: anabolism, reductive, energy-requiring.
18. The local decrease in entropy associated with living organisms is balanced by the increase in the entropy of the surroundings caused by their presence. Coupling of reactions leads to overall dispersal of energy in the Universe.
19. The synthesis of sugars by plants in photosynthesis is endergonic and requires light energy from the Sun.
20. The biosynthesis of proteins is endergonic and is accompanied by a large decrease in entropy.
21. The ATP constantly generated by living organisms is used as a source of chemical energy for endergonic processes. There is a good deal of turnover of molecules, but no net change.

### 15.4 The Role of Oxidation and Reduction in Metabolism

22. (a) NADH is oxidized,  $\text{H}^+ + \text{NADH} \rightarrow \text{NAD}^+ + 2\text{e}^- + 2\text{H}^+$ . The aldehyde is reduced,  $\text{CH}_3\text{CH}_2\text{CHO} + 2\text{e}^- + 2\text{H}^+ \rightarrow \text{CH}_3\text{CH}_2\text{CH}_2\text{OH}$ .  
 (b)  $\text{Fe}^{2+}$  is oxidized,  $\text{Fe}^{2+} \rightarrow \text{Fe}^{3+} + \text{e}^-$ .  $\text{Cu}^{2+}$  is reduced,  $\text{Cu}^{2+} + \text{e}^- \rightarrow \text{Cu}^+$ .
23. (a) The aldehyde is the oxidizing agent; NADH is the reducing agent.  
 (b)  $\text{Cu}^{2+}$  is the oxidizing agent;  $\text{Fe}^{2+}$  is the reducing agent.

### 15.5 Coenzymes in Biologically Important Oxidation-Reduction Reactions

24.  $\text{NAD}^+$ ,  $\text{NADP}^+$ , and  $\text{FAD}$  all contain an ADP moiety.
25. In NADPH, the 2' hydroxyl of the ribose attached to the adenine has a phosphate attached.
26. There is little effect in the reactions. Both are coenzymes involved in oxidation-reduction reactions. The presence of the

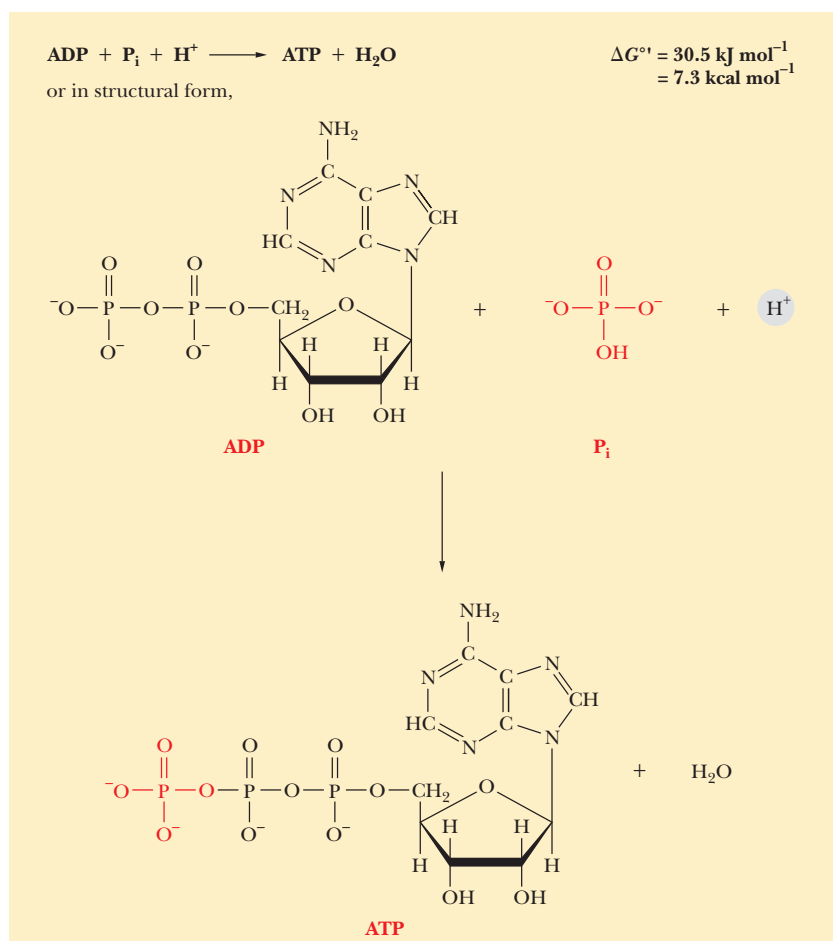
phosphate distinguishes two separate pools of coenzymes so that different ratios of  $\text{NADPH}/\text{NADP}^+$  versus  $\text{NADH}/\text{NAD}^+$  can be maintained.

27. None of these statements is true. Some coenzymes are involved in group-transfer reactions (recall this from Chapter 7). Many coenzymes contain phosphate groups, and CoA contains sulfur. ATP does not represent stored energy, but is generated on demand.
28. Redox reactions.  $\text{NAD}^+$ , or NADPH in an anabolic process, would likely be used. FAD probably would not be used because its free-energy change is too low.
29. The second half reaction (the one involving NADH) is that of oxidation; the first half reaction (the one involving  $\text{O}_2$ ) is that of reduction. The overall reaction is  $\frac{1}{2} \text{O}_2 + \text{NADH} + \text{H}^+ \rightarrow \text{H}_2\text{O} + \text{NAD}^+$ .  $\text{O}_2$  is the oxidizing agent and NADH is the reducing reagent.
30. See Figures 15.3 and 15.4.
31. Glucose-6-phosphate is oxidized, and  $\text{NADP}^+$  is reduced.  $\text{NADP}^+$  is the oxidizing agent, and glucose-6-phosphate is the reducing agent.
32. FAD is reduced, and succinate is oxidized. FAD is the oxidizing agent, and succinate is the reducing agent.
33. It is important to have two different pools of redox coenzymes. In the cytosol, the  $\text{NAD}^+/\text{NADH}$  ratio is high, but the  $\text{NADPH}/\text{NADP}^+$  ratio is also high. This means that anabolic reactions can take place in the cytosol, while catabolic reactions, such as glycolysis, can also take place. If there were not two different pools of these coenzymes, no single cell location could have both catabolism and anabolism. Having two different, but structurally related, reducing agents helps keep anabolic and catabolic reactions distinct from each other.

### 15.6 Coupling of Production and Use of Energy

34. The ratio of substrates to products would have to be 321,258 to 1.
35. Creatine phosphate + ADP  $\rightarrow$  Creatine + ATP;  
 $\Delta G^{\circ'} = -12.6 \text{ kJ}$   
 ATP + Glycerol  $\rightarrow$  ADP + Glycerol-3-phosphate;  
 $\Delta G^{\circ'} = -20.8 \text{ kJ}$   
 Creatine phosphate + Glycerol  $\rightarrow$  Creatine + Glycerol-3-phosphate;  
 $\Delta G^{\circ'} \text{ overall} = -33.4 \text{ kJ}$
36. Glucose-1-phosphate  $\rightarrow$  Glucose +  $\text{P}_i$ ;  
 $\Delta G^{\circ'} = -20.9 \text{ kJ mol}^{-1}$   
 Glucose +  $\text{P}_i \rightarrow$  Glucose-6-phosphate;  
 $\Delta G^{\circ'} = +12.5 \text{ kJ mol}^{-1}$   
 Glucose-1-phosphate  $\rightarrow$  Glucose 6-phosphate;  
 $\Delta G^{\circ'} = -8.4 \text{ kJ mol}^{-1}$
37. In both pathways, the overall reaction is  $\text{ATP} + 2 \text{H}_2\text{O} \rightarrow \text{AMP} + 2 \text{P}_i$ . Thermodynamic parameters, such as energy, are additive. The overall energy is the same because the overall pathway is the same.
38. Phosphoarginine + ADP  $\rightarrow$  Arginine + ATP;  
 $\Delta G^{\circ'} = -1.7 \text{ kJ}$   
 ATP +  $\text{H}_2\text{O} \rightarrow$  ADP +  $\text{P}_i$ ;  
 $\Delta G^{\circ'} = -30.5 \text{ kJ}$   
 Phosphoarginine +  $\text{H}_2\text{O} \rightarrow$  Arginine +  $\text{P}_i$ ;  
 $\Delta G^{\circ'} = -32.2 \text{ kJ}$
39. ATP is less stable than ADP and  $\text{P}_i$  because of the charge distribution and loss of the resonance stabilization in the phosphate

ion. There is stabilization (dispersal of energy) when ATP is hydrolyzed, leading to a negative free-energy change.



40. It is intermediate; thus, ATP is ideally positioned to serve as a phosphate donor or (as ADP) a phosphate acceptor, depending on local concentrations.
41. Creatine phosphate can phosphorylate ADP to ATP. There is a biochemical “germ of truth” here, but the effectiveness of such a supplement is another matter.
42. There is a large increase in entropy accompanying the hydrolysis of one molecule to five separate molecules.
43. PEP is a high-energy compound because energy is released upon its hydrolysis, owing to the resonance stabilization of the inorganic phosphate released and the possible keto–enol tautomerization of its product, pyruvate. See Figure 15.8.
44. The fact that a reaction is thermodynamically favorable does not mean that it will occur biologically. Even though there appears to be ample energy to catalyze the production of 2 ATPs from PEP, there is no enzyme that catalyzes this reaction.
45. Sprints and similar short periods of exercise rely on anaerobic metabolism as a source of energy, producing lactic acid. Longer periods of exercise also draw on aerobic metabolism.

#### 15.7 Coenzyme A in Activation of Metabolic Pathways

46. An activation step leads to an exergonic next step in a pathway. It is similar to the way in which organic chemists want to attach a good leaving group for the next step in a series of reactions.
47. Small energy changes generally involve mild conditions. Also, such reactions are more sensitive to relatively small changes in concentration and thus are easier to control.

48. Thioesters are high-energy compounds. The possible dissociation of the products after hydrolysis and resonance structures of the products facilitate reaction.
49. Coenzyme A serves several purposes. It is a high-energy compound, activating the initial steps of the metabolic pathway. It is used as a tag to “earmark” a molecule for a particular pathway. It is large and cannot cross membranes, so compartmentalization of pathways can be affected by binding metabolites to coenzyme A.
50. The size and complexity of the molecule make it more specific for particular enzyme-catalyzed reactions. In addition, it cannot cross membranes, so acyl-CoA molecules and other CoA derivatives can be segregated.

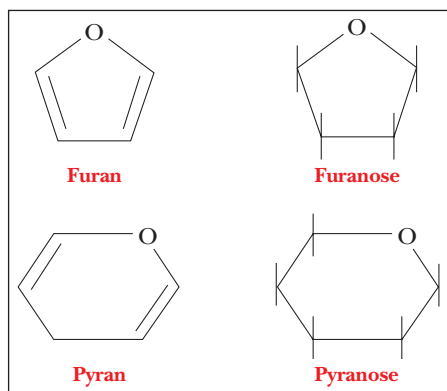
## Chapter 16

### 16.1 Sugars: Their Structures and Stereochemistry

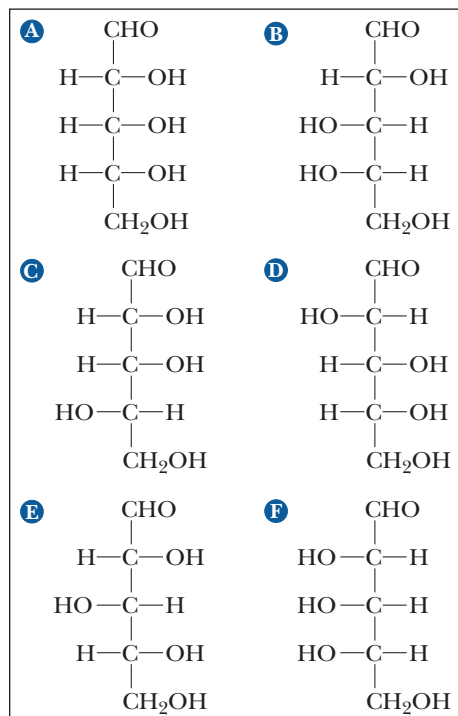
1. A polysaccharide is a polymer of simple sugars, which are compounds that contain a single carbonyl group and several hydroxyl groups. A furanose is a cyclic sugar that contains a five-membered ring similar to that in furan. A pyranose is a cyclic sugar that contains a six-membered ring similar to that in pyran. An aldose is a sugar that contains an aldehyde group; a ketose is a sugar that contains a ketone group. A glycosidic bond is the acetal linkage that joins two sugars. An oligosaccharide is a compound formed by the linking of several simple sugars (monosaccharides) by glycosidic bonds.



A glycoprotein is formed by the covalent bonding of sugars to a protein.

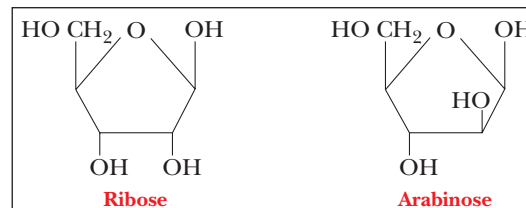


2. D-Mannose and D-galactose are both epimers of D-glucose, with inversion of configuration around carbon atoms 2 and 4, respectively; D-ribose has only five carbons, but the rest of the sugars named in this question have six.
3. All groups are aldose–ketose pairs.
4. Enantiomers are nonsuperimposable, mirror-image stereoisomers. Diastereomers are nonsuperimposable, nonmirror-image stereoisomers.
5. Four epimers of D-glucose exist, with inversion of configuration at a single carbon. The possible carbons at which this is possible are those numbered two through five.
6. Furanoses and pyranoses have five-membered and six-membered rings, respectively. It is well known from organic chemistry that rings of this size are the most stable and the most readily formed.
7. There are four chiral centers in the open-chain form of glucose (carbons two through five). Cyclization introduces another chiral center at the carbon involved in hemiacetal formation, giving a total of five chiral centers in the cyclic form.
8. Enantiomers: (a) and (f), (b) and (d). Epimers: (a) and (c), (a) and (d), (a) and (e), (b) and (f).



9. L-Sorbitol was named early in biochemical history as a derivative of L-sorbose. Reduction of D-glucose gives a hydroxy sugar that could easily be named D-glucitol, but it was originally named L-sorbitol and the name stuck.

10. Arabinose is an epimer of ribose. Nucleosides in which arabinose is substituted for ribose act as inhibitors in reactions of ribonucleosides.



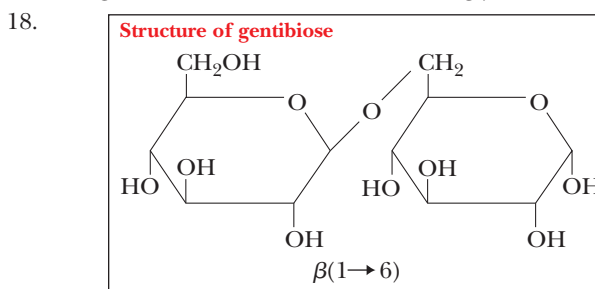
11. Converting a sugar to an epimer requires inversion of configuration at a chiral center. This can be done only by breaking and re-forming covalent bonds.
12. Two different orientations with respect to the sugar ring are possible for the hydroxyl group at the anomeric carbon. The two possibilities give rise to the new chiral center.

### 16.2 Reactions of Monosaccharides

13. This compound contains a lactic acid side chain.
14. In a sugar phosphate, an ester bond is formed between one of the sugar hydroxyls and phosphoric acid. A glycosidic bond is an acetal, which can be hydrolyzed to regenerate the two original sugar hydroxyls.
15. A reducing sugar is one that has a free aldehyde group. The aldehyde is easily oxidized, thus reducing the oxidizing agent.
16. Vitamin C is a lactone (a cyclic ester) with a double bond between two of the ring carbons. The presence of the double bond makes it susceptible to air oxidation.

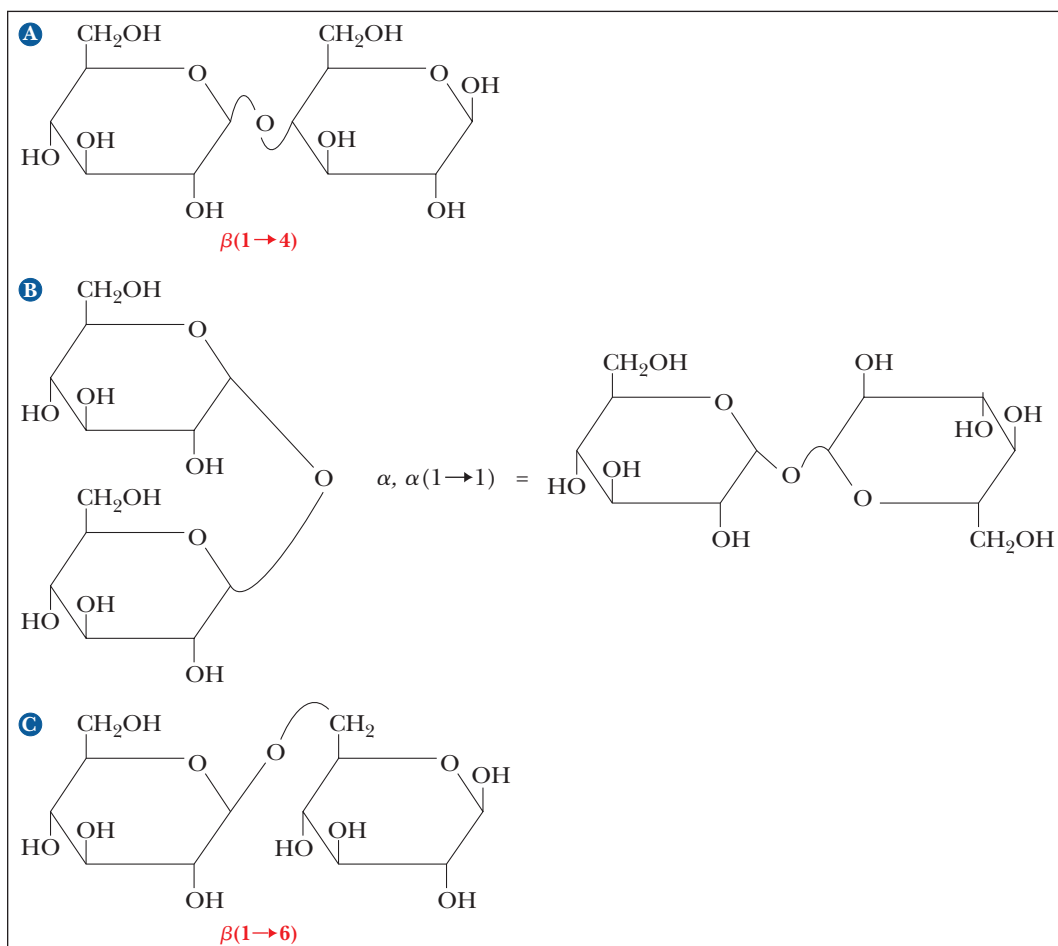
### 16.3 Some Important Oligosaccharides

17. Similarities: sucrose and lactose are both disaccharides, and both contain glucose. Differences: sucrose contains fructose, whereas lactose contains galactose. Sucrose has an  $\alpha, \beta(1 \rightarrow 2)$  glycosidic linkage, whereas lactose has a  $\beta(1 \rightarrow 4)$  glycosidic linkage.



19. In some cases, the enzyme that degrades lactose (milk sugar) to its components—glucose and galactose—is missing. In other cases, the enzyme isomerizes galactose to glucose for further metabolic breakdown.

20.

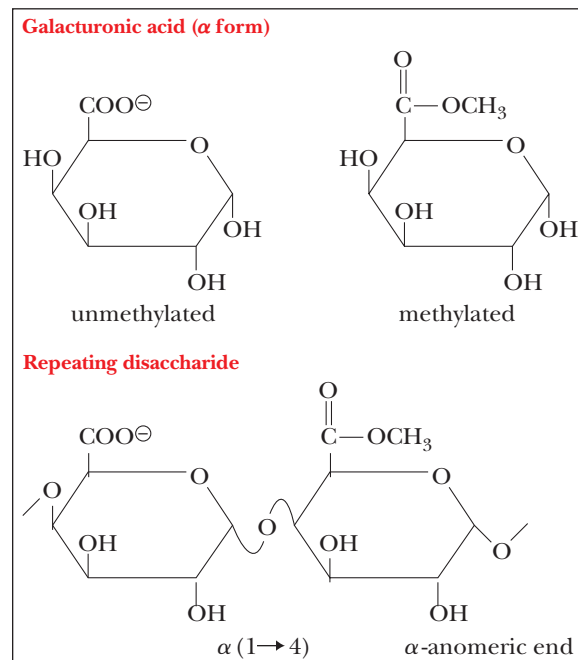


21. Milk contains lactose. Many people are sensitive to lactose and require an alternative beverage.

#### 16.4 Structures and Functions of Polysaccharides

22. The cell walls of plants consist mainly of cellulose, whereas those of bacteria consist mainly of polysaccharides with peptide cross-links.
23. Chitin is a polymer of *N*-acetyl- $\beta$ -D-glucosamine, whereas cellulose is a polymer of D-glucose. Both polymers play a structural role, but chitin occurs in the exoskeletons of invertebrates and cellulose primarily in plants.
24. Glycogen and starch differ mainly in the degree of chain branching. Both polymers serve as vehicles for energy storage, glycogen in animals and starch in plants.
25. Both cellulose and starch are polymers of glucose. In cellulose, the monomers are joined by a  $\beta$ -glycosidic linkage, whereas in starch they are joined by an  $\alpha$ -glycosidic linkage.
26. Glycogen exists as a highly branched polymer. Starch can have both a linear and a branched form, which is not as highly branched as that of glycogen.
27. Plant cell walls consist almost exclusively of carbohydrates, whereas bacterial cell walls contain peptides.

28. Repeating disaccharide of pectin:

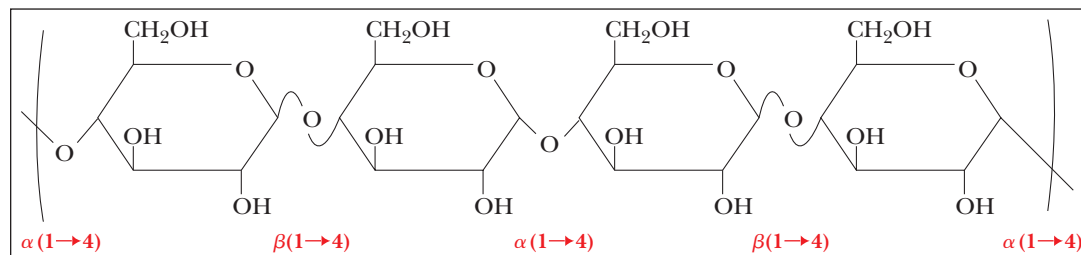


29. Glucose and fructose.

30. Differences in structure: cellulose consists of linear fibers, but starch has a coil form. Differences in function: cellulose has a structural role, but starch is used for energy storage.

31. The concentration of reducing groups is too small to detect.

32. To 2500, one place (0.02%). To 1000, four places (0.08%). To 200, 24 places (0.48%).
33. This polymer would be expected to have a structural role. The presence of the  $\beta$ -glycosidic linkage makes it useful as food only to termites or to ruminants, such as cows and horses; these animals harbor bacteria capable of attacking the  $\beta$ -linkage in their digestive tracts.



34. Because of the branching, the glycogen molecule gives rise to a number of available glucose molecules at a time when it is being hydrolyzed to provide energy. A linear molecule could produce only one available glucose at a time.
35. The digestive tract of these animals contains bacteria that have the enzyme to hydrolyze cellulose.
36. Humans lack the enzyme to hydrolyze cellulose. In addition, the fibrous structure of cellulose makes it too insoluble to digest, even if humans had the necessary enzyme.
37. The enzyme  $\beta$ -amylase is an exoglycosidase, degrading polysaccharides from the ends. The enzyme  $\alpha$ -amylase is an endoglycosidase, cleaving internal glycosidic bonds.
38. Fiber binds many toxic substances in the gut and decreases the transit time of ingested food in the digestive tract, so that harmful compounds such as carcinogens are removed from the body more quickly than would be the case with a low-fiber diet.
39. A cellulase (an enzyme that degrades cellulose) needs an active site that can recognize glucose residues joined in a  $\beta$ -glycosidic linkage and hydrolyze that linkage. An enzyme that degrades starch has the same requirements with regard to glucose residues joined in an  $\alpha$ -glycosidic linkage.
40. Cross-linking can be expected to play a role in the structures of polysaccharides where mechanical strength is an issue. Examples include cellulose and chitin. These crosslinks can be readily formed by extensive hydrogen bonding. (See Figure 16.19.)
41. The sequence of monomers in a polysaccharide is not genetically coded, and, in this sense, it does not contain information.
42. It can be useful for polysaccharides to have a number of ends, characteristic of a branched polymer, rather than the two ends of a linear polymer. This would be the case when it is necessary to release residues from the ends as quickly as possible. Polysaccharides achieve this by having  $1 \rightarrow 4$  and  $1 \rightarrow 6$  glycosidic linkages to a residue at a branch point.
43. Chitin is a suitable material for the exoskeleton of invertebrates because of its mechanical strength. Individual polymer strands are cross-linked by hydrogen bonding, accounting for the strength. Cellulose is another polysaccharide cross-linked in the same way, and it can play a similar role.
44. Bacterial cell walls are not likely to consist largely of protein. Polysaccharides are easily formed and confer considerable mechanical strength. They are likely to play a large role.
45. Athletes try to increase their stores of glycogen before an event. The most direct way to increase the amount of this polymer of glucose is to eat carbohydrates.
46. Iodine is the reagent that will be added to the reaction mixture in the titration. When the end point is reached, the next drop of iodine will produce a characteristic blue color in the presence of the indicator.
47. Heparin is an anticoagulant. Its presence prevents blood clotting.
48. Glycosidic bonds can be formed between the side-chain hydroxyls of serine or threonine residues and the sugar hydroxyls. In addition, there is the possibility of ester bonds forming between the side-chain carboxyl groups of aspartate or glutamate and the sugar hydroxyls.

### 16.5 Glycoproteins

49. Glycoproteins are ones in which carbohydrates are covalently bonded to proteins. They play a role in eukaryotic cell membranes, frequently as recognition sites for external molecules. Antibodies (immunoglobulins) are glycoproteins.
50. The sugar portions of the blood-group glycoproteins are the source of the antigenic difference.

## Chapter 17

### 17.1 The Overall Pathway in Glycolysis

- Reactions that require ATP: phosphorylation of glucose to give glucose-6-phosphate and phosphorylation of fructose-6-phosphate to give fructose-1,6-bisphosphate. Reactions that produce ATP: transfer of phosphate group from 1,3-bisphosphoglycerate to ADP and transfer of phosphate group from phosphoenolpyruvate to ADP. Enzymes that catalyze reactions requiring ATP: hexokinase, glucokinase, and phosphofructokinase. Enzymes that catalyze reactions producing ATP: phosphoglycerate kinase and pyruvate kinase.
- Reactions that require NADH: reduction of pyruvate to lactate and reduction of acetaldehyde to ethanol. Reactions that require  $\text{NAD}^+$ : oxidation of glyceraldehyde-3-phosphate to give 1,3-diphosphoglycerate. Enzymes that catalyze reactions requiring NADH: lactate dehydrogenase and alcohol dehydrogenase. Enzymes that catalyze reactions requiring  $\text{NAD}^+$ : glyceraldehyde-3-phosphate dehydrogenase.
- Pyruvate can be converted to lactate, ethanol, or acetyl-CoA.

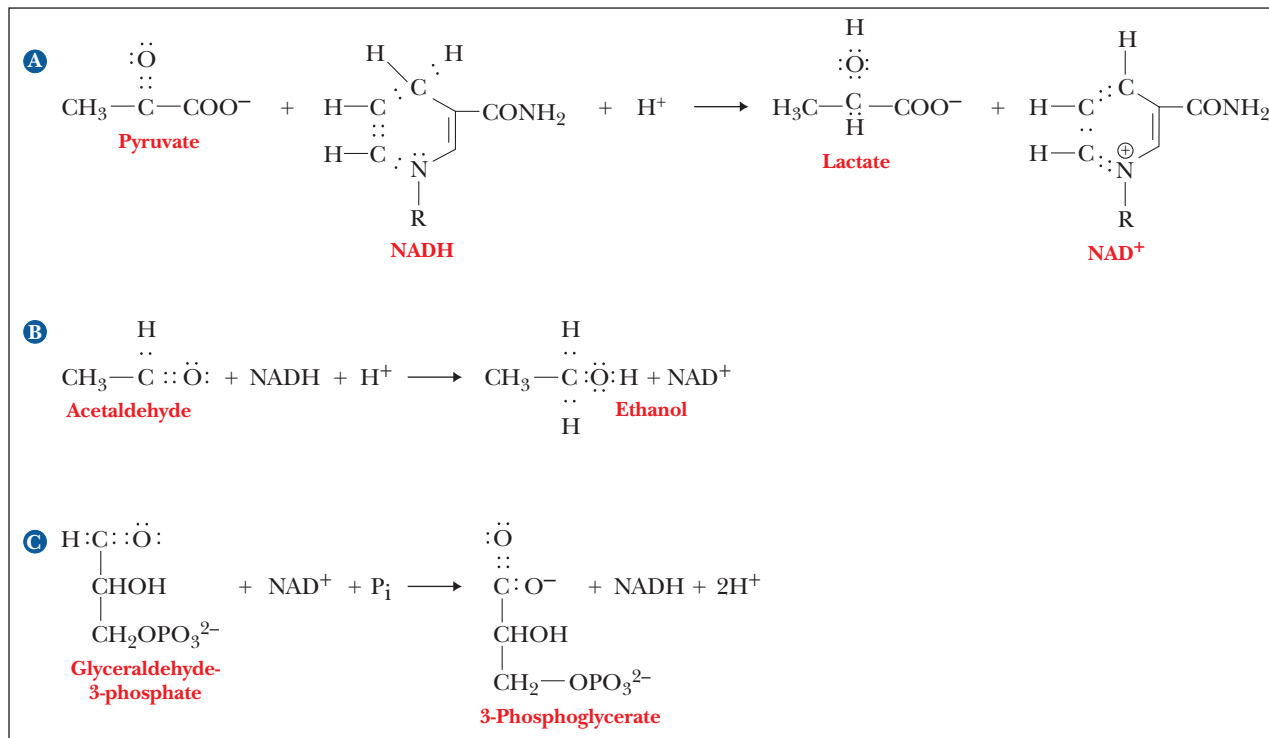
### 17.2 Conversion of Six-Carbon Glucose to Three-Carbon Glyceraldehyde-3-Phosphate

- Aldolase catalyzes the reverse aldol condensation of fructose-1,6-bisphosphate to glyceraldehyde-3-phosphate and dihydroxyacetone phosphate.

5. Isozymes are oligomeric enzymes that have slightly different amino acid compositions in different organs. Lactate dehydrogenase is an example, as is phosphofructokinase.
  6. Isozymes allow for subtle control of the enzyme to respond to different cellular needs. For example, in the liver, lactate dehydrogenase is most often used to convert lactate to pyruvate, but the reaction is often reversed in the muscle. Having a different isozyme in the muscle and liver allows for those reactions to be optimized.
  7. Fructose-1,6-bisphosphate can only undergo the reactions of glycolysis. The components of the pathway up to this point can have other metabolic fates.
  8. Add the  $\Delta G'$  mol<sup>-1</sup> values for the reactions from glucose to glyceraldehyde-3-phosphate. The result is 2.5 kJ mol<sup>-1</sup> = 0.6 kcal mol<sup>-1</sup>.
  9. The two enzymes can have different tissue locations and kinetic parameters. The glucokinase has a higher  $K_M$  for glucose than hexokinase. Thus, under conditions of low glucose, the liver does not convert glucose to glucose-6-phosphate, using the substrate that is needed elsewhere. When the glucose concentration is much higher, however, glucokinase helps phosphorylate glucose so that it can be stored as glycogen.
  10. Individuals who lack the gene that directs the synthesis of the M form of the enzyme can carry on glycolysis in their livers but experience muscle weakness because they lack the enzyme in muscle.
  11. The hexokinase molecule changes shape drastically on binding to substrate, consistent with the induced-fit theory of an enzyme adapting itself to its substrate.
  12. ATP inhibits phosphofructokinase, consistent with the fact that ATP is produced by later reactions of glycolysis.
- 17.3 Glyceraldehyde-3-Phosphate Is Converted to Pyruvate**
13. From the point at which aldolase splits fructose-1,6-bisphosphate into dihydroxyacetone phosphate and glyceraldehyde-3-phosphate; all reactions of the pathway are doubled (only the path from one glyceraldehyde-3-phosphate is usually shown).
  14. NADH-linked dehydrogenases: Glyceraldehyde-3-phosphate dehydrogenase, lactate dehydrogenase, and alcohol dehydrogenase.
  15. The free energy of hydrolysis of a substrate is the energetic driving force in substrate-level phosphorylation. An example is the conversion of glyceraldehyde-3-phosphate to 1,3-bisphosphoglycerate.
  16. The control points in glycolysis are the reactions catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase.
  17. Hexokinase is inhibited by glucose-6-phosphate. Phosphofructokinase is inhibited by ATP and citrate. Pyruvate kinase is inhibited by ATP, acetyl-CoA, and alanine. Phosphofructokinase is stimulated by AMP and fructose-2,6-bisphosphate. Pyruvate kinase is stimulated by AMP and fructose-1,6-bisphosphate.
  18. The part of the active site that binds to NADH would be the part that is most conserved, since many dehydrogenases use that coenzyme.
  19. (a) Using a high-energy phosphate to phosphorylate a substrate.  
(b) Changing the form of a molecule without changing its empirical formula (i.e., replacing one isomer with another).  
(c) Performing an aldol cleavage of a sugar to yield two smaller sugars or sugar derivatives.  
(d) Changing the oxidation state of a substrate by removing hydrogens while simultaneously changing the oxidation state of a coenzyme (NADH, FADH<sub>2</sub>, etc.).
  20. An isomerase is a general term for an enzyme that changes the form of a substrate without changing its empirical formula. A mutase is an enzyme that moves a functional group, such as a phosphate, to a new location in a substrate molecule.
  21. The reaction of 2-phosphoglycerate to phosphoenolpyruvate is a dehydration (loss of water) rather than a redox reaction.
  22. Carbon-1 of glyceraldehyde is the aldehyde group. It changes oxidation state to a carboxylic acid, which is phosphorylated simultaneously.
  23. ATP is an inhibitor of several steps of glycolysis as well as other catabolic pathways. The purpose of catabolic pathways is to produce energy, and high levels of ATP mean the cell already has sufficient energy. Glucose-6-phosphate inhibits hexokinase and is an example of product inhibition. If the glucose-6-phosphate level is high, it may indicate that sufficient glucose is available from glycogen breakdown or that the subsequent enzymatic steps of glycolysis are going slowly. Either way, there is no reason to produce more glucose-6-phosphate. Phosphofructokinase is inhibited by a special effector molecule, fructose-2,6-bisphosphate, whose levels are controlled by hormones. It is also inhibited by citrate, which indicates that there is sufficient energy from the citric acid cycle, probably from fat and amino acid degradation. Pyruvate kinase is also inhibited by acetyl-CoA, the presence of which indicates that fatty acids are being used to generate energy for the citric acid cycle. The main function of glycolysis is to feed carbon units to the citric acid cycle. When these carbon skeletons can come from other sources, glycolysis is inhibited to spare glucose for other purposes.
  24. There would be 15 possible isozymes of LDH, combining three different subunits into combinations of four. Besides the five isozymes containing only M and H, there would also be C<sub>4</sub>, CH<sub>3</sub>, C<sub>2</sub>H<sub>2</sub>, C<sub>3</sub>H, CH<sub>2</sub>M, C<sub>2</sub>HM, C<sub>3</sub>M, CHM<sub>2</sub>, C<sub>2</sub>M<sub>2</sub>, and CM<sub>3</sub>.
  25. Glutamic acid has an acidic side chain with a pK<sub>a</sub> of 4.25. Therefore, it would be negatively charged at pH 8.6, and the H subunit would move more toward the anode (+) than the M subunit. Thus, LDH 1, which is H<sub>4</sub>, would move the farthest. LDH 5, which is M<sub>4</sub>, would move the least, with the other isozymes migrating between those two extremes proportional to their H content.
  26. The formation of fructose-1,6-bisphosphate is the committed step in the glycolytic pathway. It is also one of the energy-requiring steps of the pathway.
  27. Glucose-6-phosphate inhibits hexokinase, the enzyme responsible for its own formation. Because G-6-P is used up by additional reactions of glycolysis, the inhibition is relieved.
  28. With few exceptions, a biochemical reaction typically results in only one chemical modification of the substrate. Accordingly, several to many steps are needed to reach the ultimate goal.
  29. The enzyme contains a phosphate group on a suitable amino acid, such as serine, threonine, and histidine. The substrate donates its phosphate group from the C-3 position to another amino acid on the enzyme, subsequently receiving the one that started out on the enzyme. Thus, the <sup>32</sup>P that was on the substrate is transferred to the enzyme, while an unlabeled phosphorus is put on the C-2 position.
- 17.4 Anaerobic Metabolism of Pyruvate**
30. The bubbles in beer are CO<sub>2</sub>, produced by alcoholic fermentation. Tired and aching muscles are caused in part by a buildup of lactic acid, a product of anaerobic glycolysis.

31. The problem with lactic acid is that it is an acid. The  $\text{H}^+$  produced from lactic acid formation causes the burning muscle sensation. Sodium lactate is the conjugate weak base of lactic acid. It is reconverted to glucose by gluconeogenesis in the liver. Giving sodium lactate intravenously is a good way to supply an indirect source of blood glucose.

33.



34. Thiamine pyrophosphate is a coenzyme in the transfer of two-carbon units. It is required for catalysis by pyruvate decarboxylase in alcoholic fermentation.

35. The important part of TPP is the five-membered ring, in which a carbon is found between a nitrogen and a sulfur. This carbon forms a carbanion and is extremely reactive, making it able to perform a nucleophilic attack on carbonyl groups, leading to decarboxylation of several compounds in different pathways.

36. Thiamine pyrophosphate is a coenzyme required in the reaction catalyzed by pyruvate carboxylase. Because this reaction is a part of the metabolism of ethanol, less will be available to serve as a coenzyme in the reactions of other enzymes that require it.

37. Animals that have been run to death have accumulated large amounts of lactic acid in their muscle tissue, accounting for the sour taste of the meat.

38. Conversion of glucose to lactate rather than pyruvate recycles NADH.

39. This is possible, and it is done. These poisons also affect other tissues, including skin, hair, cells of the intestinal lining, and especially the immune system and red blood cells. People on chemotherapy are usually more susceptible to infectious diseases than healthy people and are often somewhat anemic.

### 17.5 Energy Production in Glycolysis

40. The energy released by all the reactions of glycolysis is  $184.5 \text{ kJ mol glucose}^{-1}$ . The energy released by glycolysis drives the phosphorylation of two ADP to ATP for each molecule of glucose,

32. The purpose of the step that produces lactic acid is to reduce pyruvate so that NADH can be oxidized to  $\text{NAD}^+$ , which is needed for the step catalyzed by glyceraldehyde-3-phosphate dehydrogenase.

trapping  $61.0 \text{ kJ mol glucose}^{-1}$ . The estimate of 33% efficiency comes from the calculation  $(61.0/184.5) \times 100 = 33\%$ .

41. There is a net gain of two ATP molecules per glucose molecule consumed in glycolysis.

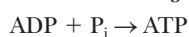
42. The gross yield is four ATP molecules per glucose molecule, but the reactions of glycolysis require two ATP per glucose.

43. The reactions catalyzed by hexokinase, phosphofructokinase, glyceraldehyde-3-phosphate dehydrogenase, phosphoglycerokinase, and pyruvate kinase.

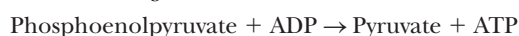
44. The steps catalyzed by hexokinase, phosphofructokinase, and pyruvate kinase.

45. Phosphoenolpyruvate  $\rightarrow$  pyruvate +  $\text{P}_i$

$$\Delta G^\circ = -61.9 \text{ kJ mol}^{-1} = -14.8 \text{ kcal mol}^{-1}$$



$$\Delta G^\circ = 30.5 \text{ kJ mol}^{-1} = 7.3 \text{ kcal mol}^{-1}$$



$$\Delta G^\circ = -31.4 \text{ kJ mol}^{-1} = -7.5 \text{ kcal mol}^{-1}$$

46. The net yield of ATP from glycolysis is the same, two ATP, when either of the three substrates is used. The energetics of the conversion of hexoses to pyruvate are the same, regardless of hexose type.

47. Starting with glucose-1-phosphate, the net yield is three ATP, because one of the priming reactions is no longer used. Thus, glycogen is a more efficient fuel for glycolysis than free glucose.



48. Phosphoenolpyruvate + ADP  $\rightarrow$  Pyruvate + ATP
- |  |  |
|--|--|
|  | $\Delta G^{\circ'} = -31.4 \text{ kJ/mol}$ |
| ADP + P <sub>i</sub> $\rightarrow$ ATP | $\Delta G^{\circ'} = 30.5 \text{ kJ/mol}$  |
| Sum                                    | $\Delta G^{\circ'} = -0.9 \text{ kJ/mol}$  |
- Thus, the reaction is thermodynamically possible under standard conditions.
49. No, the reaction shown in Question 48 does not occur in nature. We can assume that no enzyme evolved that could catalyze it. Nature is not 100% efficient.
50. A positive  $\Delta G^{\circ'}$  does not necessarily mean that the reaction has a positive  $\Delta G$ . Substrate concentrations can make a negative  $\Delta G$  out of a positive  $\Delta G^{\circ'}$ .
51. The entire pathway can be looked at as a large coupled reaction. Thus, if the overall pathway has a negative  $\Delta G$ , an individual step may be able to have a positive  $\Delta G$ , and the pathway can still continue.

## Chapter 18

### 18.1 How Glycogen Is Produced and Degraded

- These two pathways occur in the same cellular compartment, and, if both are on at the same time, a futile ATP hydrolysis cycle results. Using the same mechanism to turn them on/off or off/on is highly efficient.
- In phosphorylation, a bond is cleaved by adding the elements of phosphoric acid across that bond, whereas in hydrolysis, the cleavage takes place by adding the elements of water across the bond.
- Glucose-6-phosphate is already phosphorylated. This saves one ATP equivalent in the early stages of glycolysis.
- Each glucose residue is added to the growing glycogen molecule by transfer from UDPG.
- Glycogen synthase is subject to covalent modification and to allosteric control. The enzyme is active in its phosphorylated form and inactive when dephosphorylated. AMP is an allosteric inhibitor of glycogen synthase, whereas ATP and glucose-6-phosphate are allosteric activators.
- There is a net gain of three, rather than two, ATP when glycogen, not glucose, is the starting material of glycolysis.
- It “costs” one ATP equivalent (UTP to UDP) to add a glucose residue to glycogen. In degradation, about 90% of the glucose residues do not require ATP to produce glucose-1-phosphate. The other 10% require ATP to phosphorylate glucose. On average, this is another 0.1 ATP. Thus, the overall “cost” is 1.1 ATP, compared with the three ATP that can be derived from glucose-6-phosphate by glycolysis.
- The ATP cost is the same, but more than 30 ATP can be derived from aerobic metabolism.
- Eating high-carbohydrate foods for several days before strenuous activity is intended to build up glycogen stores in the body. Glycogen will be available to supply required energy.
- The disaccharide sucrose can be hydrolyzed to glucose and fructose, which can both be readily converted to glucose-1-phosphate, the immediate precursor of glycogen. This is not the usual form of “glycogen loading.”
- Probably not, because the sugar spike initially results in a rapid increase in insulin levels, which results in lowering blood glucose levels and increased glycogen storage in the liver.
- The sprint is essentially anaerobic and produces lactate from glucose by glycolysis. Lactate is then recycled to glucose by gluconeogenesis.
- It is unlikely that this finding will be confirmed by other researchers. The highly branched structure of glycogen is optimized for release of glucose on demand.
- Each glucose residue added to a growing phosphate chain comes from uridine diphosphate glucose. The cleavage of the phosphate ester bond to the nucleoside diphosphate moiety supplies the needed energy.
- The enzyme that catalyzes addition of glucose residues to a growing glycogen chain cannot form a bond between isolated glucose residues; thus we have the need for a primer.
- The glycogen synthase reaction is exergonic overall because it is coupled to phosphate ester hydrolysis.
- (a) Increasing the level of ATP favors both gluconeogenesis and glycogen synthesis.  
(b) Decreasing the level of fructose-1,6-bisphosphate would tend to stimulate glycolysis, rather than gluconeogenesis or glycogen synthesis.  
(c) Levels of fructose-6-phosphate do not have a marked regulatory effect on these pathways of carbohydrate metabolism.
- “Going for the burn” in a workout refers to the sensation that accompanies lactic acid buildup. This in turn arises from anaerobic metabolism of glucose in muscle.
- Sugar nucleotides are diphosphates. The net result is hydrolysis to two phosphate ions, releasing more energy and driving the addition of glucose residues to glycogen in the direction of polymerization.

### 18.2 Gluconeogenesis Produces Glucose from Pyruvate

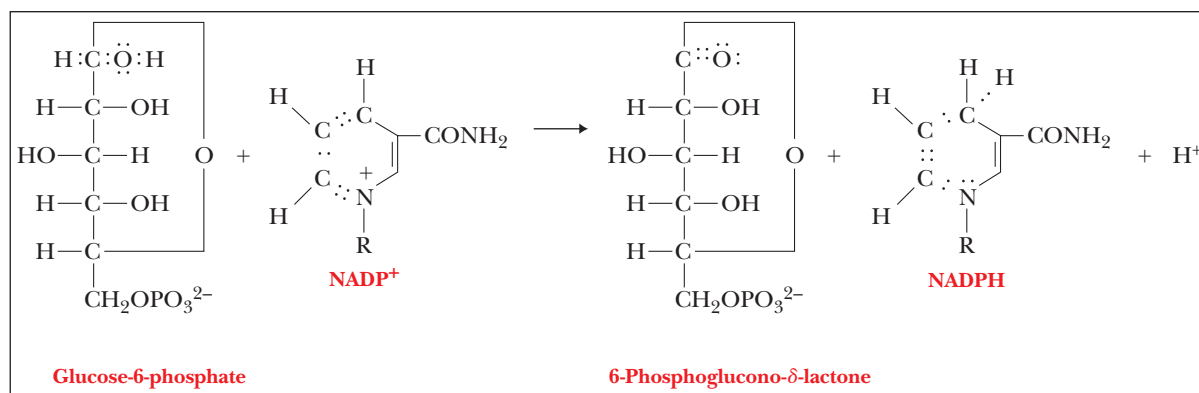
- Reactions that require acetyl-CoA: none. Reactions that require biotin: carboxylation of pyruvate to oxaloacetate.
- Three reactions of glycolysis are irreversible under physiological conditions. They are the production of pyruvate and ATP from phosphoenolpyruvate, the production of fructose-1,6-bisphosphate from fructose-6-phosphate, and the production of glucose-6-phosphate from glucose. These reactions are bypassed in gluconeogenesis; the reactions of gluconeogenesis differ from those of glycolysis at these points and are catalyzed by different enzymes.
- Biotin is the molecule to which carbon dioxide is attached to the process of being transferred to pyruvate. The reaction produces oxaloacetate, which then undergoes further reactions of gluconeogenesis.
- In gluconeogenesis, glucose-6-phosphate is dephosphorylated to glucose (the last step of the pathway); in glycolysis, it isomerizes to fructose-6-phosphate (an early step in the pathway).
- Of the three processes—glycogen formation, gluconeogenesis, and the pentose phosphate pathway—only one, gluconeogenesis, involves an enzyme that requires biotin. The enzyme in question is pyruvate carboxylase, which catalyzes the conversion of pyruvate to oxaloacetate, an early step in gluconeogenesis.
- The hydrolysis of fructose-1,6-bisphosphate is a strongly exergonic reaction. The reverse reaction in glycolysis, phosphorylation of fructose-6-phosphate, is irreversible because of the energy supplied by ATP hydrolysis.

### 18.3 Control of Carbohydrate Metabolism

- Reactions that require ATP: formation of UDP-glucose from glucose-1-phosphate and UTP (indirect requirement, because ATP is needed to regenerate UTP), regeneration of UTP, and carboxylation of pyruvate to oxaloacetate. Reactions that produce ATP: none. Enzymes that catalyze ATP-requiring reactions:

- UDP-glucose phosphorylase (indirect requirement), nucleoside phosphate kinase, and pyruvate carboxylase. Enzymes that catalyze ATP-producing reactions: none.
27. Fructose-2,6-bisphosphate is an allosteric activator of phosphofructokinase (a glycolytic enzyme) and an allosteric inhibitor of fructose bisphosphate phosphatase (an enzyme in the pathway of gluconeogenesis).
  28. Hexokinase can add a phosphate group to any of several six-carbon sugars, whereas glucokinase is specific for glucose. Glucokinase has a lower affinity for glucose than does hexokinase. Consequently, glucokinase tends to deal with an excess of glucose, particularly in the liver. Hexokinase is the usual enzyme for phosphorylating six-carbon sugars.
  29. The Cori cycle is a pathway in which there is cycling of glucose due to glycolysis in muscle and gluconeogenesis in liver. The blood transports lactate from muscle to liver and glucose from liver to muscle.
  30. Substrate cycles are futile in the sense that there is no net change except for the hydrolysis of ATP. However, substrate cycles allow for increased control over opposing reactions when they are catalyzed by different enzymes.
  31. Having two control mechanisms allows for fine-tuning of control and for the possibility of amplification. Both mechanisms are capable of rapid response to conditions, milliseconds in the case of allosteric control and seconds to minutes in the case of covalent modification.
  32. Different control mechanisms have inherently different time scales. Allosteric control can take place in milliseconds, whereas covalent control takes seconds to minutes. Genetic control has a longer time scale than either.
  33. The most important aspect of the amplification scheme is that the control mechanisms affect agents that are catalysts themselves. An enhancement by several powers of ten is itself increased by several powers of ten.
  34. Enzymes, like all catalysts, speed up the forward and reverse reaction to the same extent. Having different catalysts is the only way to ensure independent control over the rates of the forward and reverse process.
  35. Muscle tissue uses large quantities of glucose, producing lactate in the process. The liver is an important site of gluconeogenesis to recycle the lactate to glucose.
  36. Fructose-2,6-bisphosphate is an allosteric activator of phosphofructokinase (a glycolytic enzyme) and an allosteric inhibitor of fructose bisphosphate phosphatase (an enzyme in the pathway

46.



of gluconeogenesis). It thus plays a role in two pathways that are not exactly the reverse of each other.

37. The concentration of fructose-2,6-bisphosphate in a cell depends on the balance between its synthesis (catalyzed by phosphofructokinase-2) and its breakdown (catalyzed by fructose bisphosphatase-2). The separate enzymes that control the formation and breakdown of fructose-2,6-bisphosphate are themselves controlled by a phosphorylation/dephosphorylation mechanism.
  38. Glycogen is more extensively branched than starch. It is a more useful storage form of glucose for animals because the glucose can be mobilized more easily when there is a need for energy.
- 18.4 Glucose Is Sometimes Diverted through the Pentose Phosphate Pathway**
39. NADPH has one more phosphate group than NADH (at the 2' position of the ribose ring of the adenine nucleotide portion of the molecule). NADH is produced in oxidative reactions that give rise to ATP. NADPH is a reducing agent in biosynthesis. The enzymes that use NADH as a coenzyme are different from those that require NADPH.
  40. Glucose-6-phosphate can be converted to glucose (gluconeogenesis), glycogen, pentose phosphates (pentose phosphate pathway), or pyruvate (glycolysis).
  41. Hemolytic anemia is caused by defective working of the pentose phosphate pathway. There is a deficiency of NADPH, which indirectly contributes to the integrity of the red blood cells. The pentose phosphate pathway is the only source of NADPH in red blood cells.
  42. (a) By using only the oxidative reactions.  
(b) By using the oxidative reactions, the transaldolase and transketolase reactions, and gluconeogenesis.  
(c) By using glycolytic reactions and the transaldolase and transketolase reactions in reverse.
  43. Transketolase catalyzes the transfer of a two-carbon unit, whereas transaldolase catalyzes the transfer of a three-carbon unit.
  44. In red blood cells, the presence of the reduced form of glutathione is necessary for the maintenance of the sulfhydryl groups of hemoglobin and other proteins in their reduced forms, as well as for keeping the Fe(II) of hemoglobin in its reduced form. Glutathione also maintains the integrity of red cells by reacting with peroxides that would otherwise degrade fatty-acid side chains in the cell membrane.
  45. Thiamine pyrophosphate is a cofactor necessary for the function of transketolase, an enzyme that catalyzes one of the reactions in the nonoxidative part of the pentose phosphate pathway.

47. Having different reducing agents for anabolic and catabolic pathways keeps the pathways separate metabolically. Thus, they are subject to independent control and do not waste energy.
48. If a cell needs NADPH, all the reactions of the pentose phosphate pathway take place. If a cell needs ribose-5-phosphate, the oxidative portion of the pathway can be bypassed; only the nonoxidative reshuffling reactions take place. The pentose phosphate pathway does not have a significant effect on the cell's supply of ATP.
49. The ester bond is more easily broken than any of the other bonds that form the sugar ring. Hydrolysis of that bond is the next step in the pathway.
50. The reshuffling reactions of the pentose phosphate pathway have both an epimerase and an isomerase. Without an isomerase, all the sugars involved are keto sugars, which are not substrates for transaldolase, one of the key enzymes in the reshuffling process.

## Chapter 19

### 19.1 The Central Role of the Citric Acid Cycle in Metabolism

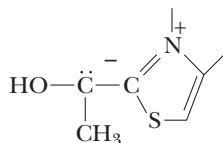
1. Anaerobic glycolysis is the principal pathway for the anaerobic metabolism of glucose. The pentose phosphate pathway can also be considered. Aerobic glycolysis and the citric acid cycle are responsible for the aerobic metabolism of glucose.
2. Anaerobically, two ATPs can be produced from one glucose molecule. Aerobically, this figure is 30 to 32, depending on in which tissue it is occurring.
3. The citric acid cycle is also called the Krebs cycle, the tricarboxylic acid cycle, and the TCA cycle.
4. Amphibolic means that the pathway is involved in both catabolism and anabolism.

### 19.2 The Overall Pathway of the Citric Acid Cycle

5. The citric acid cycle takes place in the mitochondrial matrix. Glycolysis takes place in the cytosol.
6. There is a transporter on the inner mitochondrial matrix that allows pyruvate from the cytosol to pass into the mitochondria.
7.  $\text{NAD}^+$  and FAD are the primary electron acceptors of the citric acid cycle.
8. NADH and  $\text{FADH}_2$  are indirect sources of energy produced in the TCA cycle. GTP is a direct source of energy.

### 19.3 How Pyruvate Is Converted to Acetyl-CoA

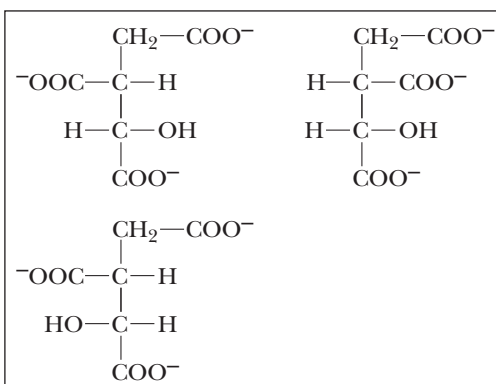
9. Five enzymes are involved in the pyruvate dehydrogenase complex of mammals. Pyruvate dehydrogenase transfers a two-carbon unit to TPP and releases  $\text{CO}_2$ . Dihydrolipoyl transacetylase transfers the two-carbon acetyl unit to lipoic acid and then to coenzyme A. Dihydrolipoyl dehydrogenase reoxidizes lipoic acid and reduces  $\text{NAD}^+$  to NADH. Pyruvate dehydrogenase kinase phosphorylates PDH. PDH phosphatase removes the phosphate.
10. Lipoic acid plays a role both in redox and in acetyl-transfer reactions.
11. Five enzymes are all in close proximity for efficient shuttling of the acetyl unit between molecules and efficient control of the complex by phosphorylation.
12. Thiamine pyrophosphate comes from the B vitamin thiamine. Lipoic acid is a vitamin.  $\text{NAD}^+$  comes from the B vitamin niacin. FAD comes from the B vitamin riboflavin.
- 13.



14. See Figure 19.4.

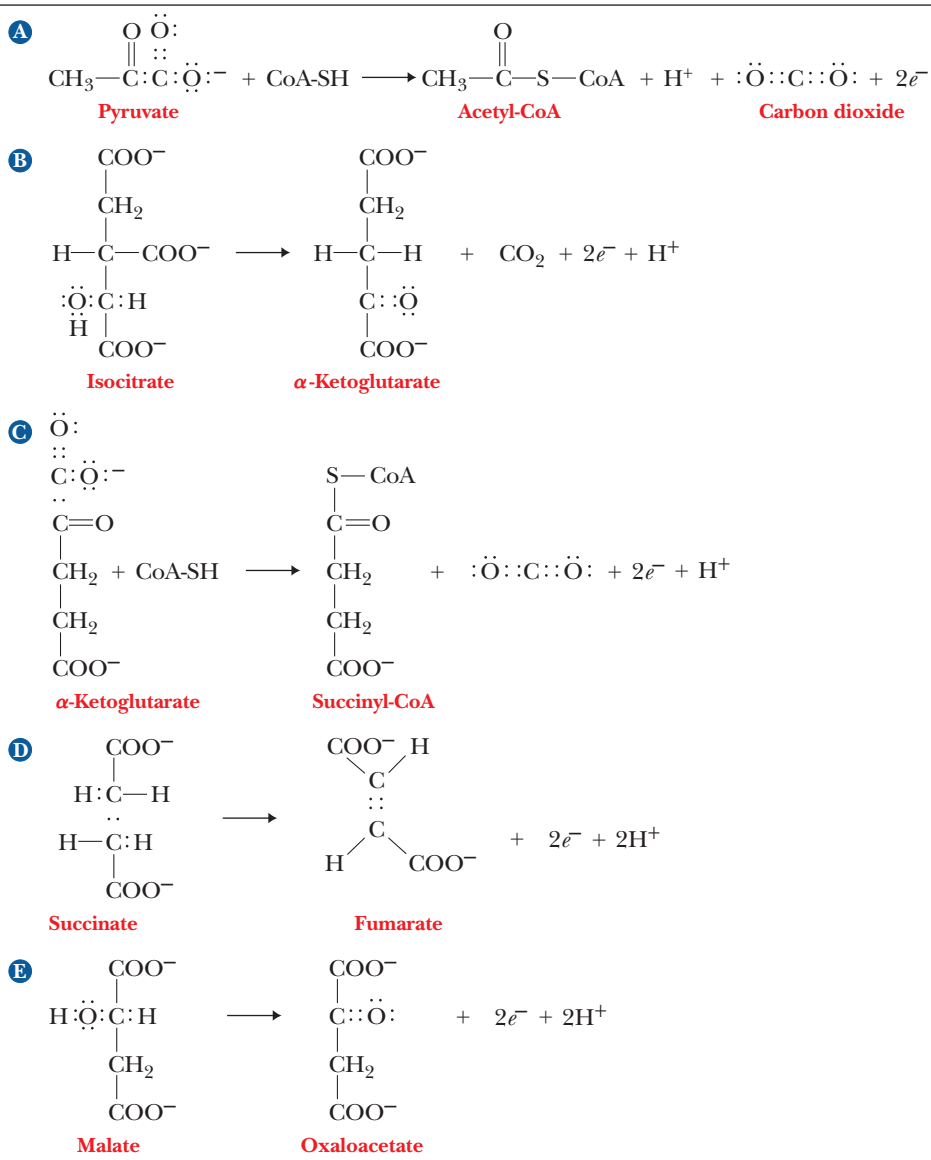
### 19.4 The Individual Reactions of the Citric Acid Cycle

15. A condensation reaction is one in which a new carbon–carbon bond is formed. The reaction of acetyl-CoA and oxaloacetate to produce citrate involves formation of such a carbon–carbon bond.
16. It means that the reaction catalyzed by the enzyme produces the product that is part of the name and does not require a direct input of energy from a high-energy phosphate. Thus, citrate synthase catalyzes the synthesis of citrate without using ATP to do it.
17. Fluoroacetate is a poison that is produced naturally in some plants and is also used as a poison against undesirable pests. It is poisonous because it is used by citrate synthase to make fluorocitrate, which is an inhibitor of the citric acid cycle.
18. The reaction involves an achiral molecule (citrate) being converted to a chiral one (isocitrate).
19. Conversion of pyruvate to acetyl-CoA, conversion of isocitrate to  $\alpha$ -ketoglutarate, and conversion of  $\alpha$ -ketoglutarate to succinyl-CoA.
20. Conversion of pyruvate to acetyl-CoA, conversion of isocitrate to  $\alpha$ -ketoglutarate, conversion of  $\alpha$ -ketoglutarate to succinyl-CoA, conversion of succinate to fumarate, and conversion of malate to oxaloacetate.
21. These enzymes catalyze oxidative decarboxylations.
22. The reactions proceed by the same mechanism and use the same cofactors. The difference is the initial substrate, which is pyruvate or  $\alpha$ -ketoglutarate. During the course of the reaction, pyruvate dehydrogenase shuttles an acetyl unit through the reaction while  $\alpha$ -ketoglutarate dehydrogenase shuttles a succinyl unit.
23. A synthetase is an enzyme that synthesizes a molecule and uses a high-energy phosphate in the process.
24. GTP is equivalent to ATP because an enzyme, nucleoside diphosphate kinase, is able to interconvert GTP and ATP.
25. The enzymes that reduce  $\text{NAD}^+$  are all soluble, matrix enzymes, while succinate dehydrogenase is membrane-bound. The  $\text{NAD}^+$ -linked dehydrogenases all catalyze oxidations that involve carbons and oxygens, such as an alcohol group being oxidized to an aldehyde or aldehyde to carboxylic acid. The FAD-linked dehydrogenase oxidizes a carbon–carbon single bond to a double bond.
26. There is an adenine nucleotide portion in the structure of NADH, with a specific binding site on NADH-linked dehydrogenases for this portion of NADH.
27. The conversion of fumarate to malate is a hydration reaction, not a redox reaction.
- 28.



## Answers to Questions

29.



## 19.5 Energetics and Control of the Citric Acid Cycle

30. The reactions are catalyzed by pyruvate dehydrogenase, citrate synthase, isocitrate dehydrogenase, and  $\alpha$ -ketoglutarate dehydrogenase.
31. PDH is controlled allosterically. It is inhibited by ATP, acetyl-CoA, and NADH. In addition, it is subject to control by phosphorylation. When PDH kinase phosphorylates PDH, it becomes inactive. Removing the phosphate with the PDH phosphatase reactivates it.
32. ATP and NADH are the two most common inhibitors.
33. If the amount of ADP in a cell increases relative to the amount of ATP, the cell needs energy (ATP). This situation not only favors the reactions of the citric acid cycle, which release energy, activating isocitrate dehydrogenase, but also stimulates the formation of NADH and FADH<sub>2</sub> for ATP production by electron transport and oxidative phosphorylation.
34. If the amount of NADH in a cell increases relative to the amount of NAD<sup>+</sup>, the cell has completed a number of energy-releasing reactions. There is less need for the citric acid cycle to be active; as a result, the activity of pyruvate dehydrogenase is decreased.
35. ADP ratio and a high NADH/NAD<sup>+</sup> ratio. Both ratios indicate a high “energy charge” in the cell, indicating less of a need for the energy-releasing reactions of the citric acid cycle.
36. Thioesters are “high-energy” compounds that play a role in group-transfer reactions; consequently, their  $\Delta G^\circ$  of hydrolysis is large and negative to provide energy for the reaction.
37. The energy released by hydrolysis of acetyl-CoA is needed for the condensation reaction that links the acetyl moiety to oxaloacetate, yielding citrate. The energy released by hydrolysis of succinyl-CoA drives the phosphorylation of GDP, yielding GTP.
38. Table 19.2 shows that the sum of the energies of the individual reactions is  $-44.3$  kJ ( $-10.6$  kcal) for each mole of acetyl-CoA that enters the cycle.
39. The expression would relate to the intensive extraction of energy from intermediate compounds by redox reactions. Including the pyruvate dehydrogenase reaction, 5 of 9 reactions are redox reactions (in contrast with only 1 of 10 in glycolysis). Accordingly, energy is rapidly extracted from carbon compounds (yielding the energyless CO<sub>2</sub>) and is transferred to NAD<sup>+</sup> and FAD for subsequent utilization.



40. Lactose is a disaccharide of glucose and galactose. There is no energy cost in the hydrolysis of the bond between the two monosaccharides, so essentially there are two hexoses to consider. Because the processing of any of the hexoses yields the same amount of energy, the aerobic processing of lactose would lead to 60 to 64 ATPs, depending on the tissue and on the shuttle system used.

### 19.6 The Glyoxylate Cycle: A Related Pathway

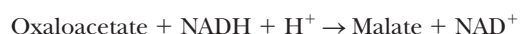
41. Isocitrate dehydrogenase,  $\alpha$ -ketoglutarate dehydrogenase, and succinyl-CoA synthetase.
42. The conversion of isocitrate to succinate and glyoxylate catalyzed by isocitrate lyase and the conversion of glyoxylate and acetyl-CoA to malate catalyzed by malate synthase.
43. Bacteria that have a glyoxylate cycle can convert the acetic acid to amino acids, carbohydrates, and lipids, but humans can use the acetic acid only as an energy source or to make lipids.

### 19.7 The Citric Acid Cycle in Catabolism

44. The citric acid cycle is the central metabolic pathway and indirect producer of energy. It receives fuels from the other pathways at many points and generates reduced electron carriers that go into the electron transport chain. It is also involved in anabolism, as many of its intermediates can be drawn off to synthesize other compounds.
45. The citric acid cycle occurs in the mitochondrial matrix, which is more selective in its permeability than the plasma membrane.
46. In oxidative decarboxylation, the molecule that is oxidized loses a carboxyl group as carbon dioxide. Examples of oxidative decarboxylation include the conversion of pyruvate to acetyl-CoA, isocitrate to  $\alpha$ -ketoglutarate, and  $\alpha$ -ketoglutarate to succinyl-CoA.
47. Yes, not only is citric acid completely degraded to carbon dioxide and water, but it is also readily absorbed into the mitochondrion.

### 19.8 The Citric Acid Cycle in Anabolism

48. The following series of reactions exchanges NADH for NADPH.



49. A variety of reactions in which amino acids are converted to citric acid cycle intermediates are considered anaplerotic. In addition, pyruvate +  $\text{CO}_2$  can form oxaloacetate via pyruvate carboxylase.
50. Many compounds can form acetyl-CoA, such as fats, carbohydrates, and many amino acids. Acetyl-CoA can also form fats and ketone bodies, as well as feed directly into the citric acid cycle.

### 19.9 The Link to Oxygen

51. The NADH and  $\text{FADH}_2$  produced by the citric acid cycle are the electron donors in the electron transport chain linked to oxygen. Because of this connection, the citric acid cycle is considered part of aerobic metabolism.

## Chapter 20

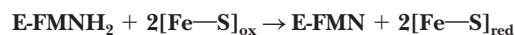
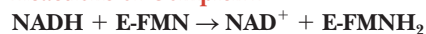
### 20.1 The Role of Electron Transport in Metabolism

- Electrons are passed from NADH to a flavin-containing protein to coenzyme Q. From coenzyme Q, the electrons pass to cytochrome *b*, then to cytochrome *c*, via the Q cycle, followed by cytochromes *a* and *a*<sub>3</sub>. From the cytochrome *aa*<sub>3</sub> complex, the electrons are finally passed to oxygen.
- Electron transport and oxidative phosphorylation are different processes. Electron transport requires the respiratory complexes

of the inner mitochondrial membrane, whereas oxidative phosphorylation requires ATP synthase, also located on the inner mitochondrial membrane. Electron transport can take place in the absence of oxidative phosphorylation.

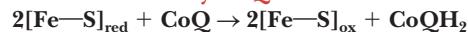
- In all reactions, electrons are passed from the reduced form of one reactant to the oxidized form of the next reactant in the chain. The notation [Fe—S] refers to any one of a number of iron-sulfur proteins.

#### Reactions of Complex I



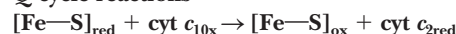
**Liberation of enough energy to produce ATP**

#### Transfer to Coenzyme Q



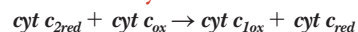
#### Reactions of Complex III

##### Q cycle reactions

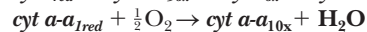
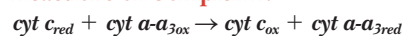


**Liberation of enough energy to produce ATP**

#### Transfer to cytochrome c

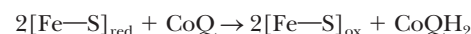
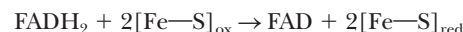


#### Reactions of Complex IV



**Liberation of enough energy to produce ATP**

- When  $\text{FADH}_2$  is the starting point for electron transport, electrons are passed from  $\text{FADH}_2$  to coenzyme Q in a reaction carried out by Complex II that bypasses Complex I.



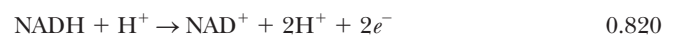
- Mitochondrial structure confines the reduced electron carriers produced by the citric acid cycle to the matrix. There they are close to the respiratory complexes of the electron transport chain that will pass the electrons from the carriers produced by the citric acid cycle to oxygen, the ultimate recipient of electrons and hydrogens.

### 20.2 Reduction Potentials in the Electron Transport Chain

- The electron transport chain translocates charged particles by chemical means. Interconversion of chemical and electrical energy is exactly what a battery does.
- The reactions are all written in the same direction for purposes of comparison. By convention, they are written as reduction, rather than oxidation, reactions.
- $\Delta G^\circ = -60 \text{ kJ/mol}$
- We fundamentally add the half reactions in Table 20.1.



This is the wrong direction, so we reverse the equation and the sign of the potential difference.

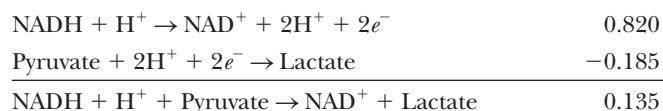


- We fundamentally add the half reactions in Table 20.1.





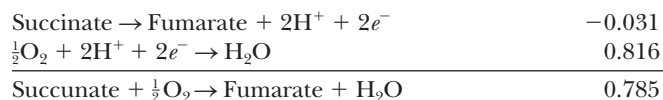
This is the wrong direction, so we reverse the equation and the sign of the potential difference.



11. We fundamentally add the half reactions in Table 20.1.



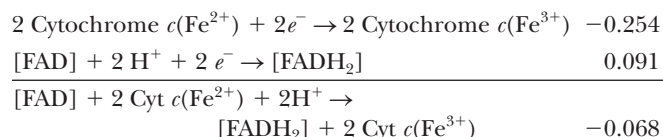
This is the wrong direction, so we reverse the equation and the sign of the potential difference.



12. The cytochrome is the electron donor, and the flavin moiety is the electron acceptor. Once again, we add the half reactions in Table 20.1.



This is the wrong direction, so we reverse the equation and the sign of the potential difference.

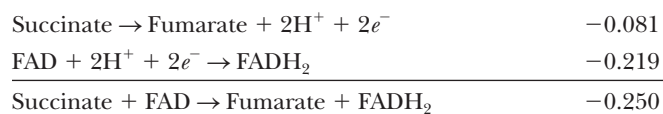


This was the maximum value for a bound flavin. The negative sign indicates that this reaction will not take place as written because it is not energetically favorable.

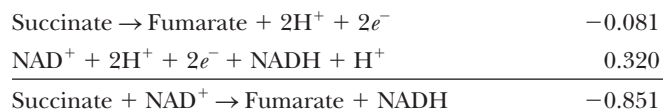
13. Here is an illustration based on standard reduction potentials.



This is the wrong direction, so we reverse the equation and the sign of the potential difference.



The other possibility can be calculated in the same way.



Both reduction potentials indicate a reaction that is not energetically favorable, but less so with FAD than with  $\text{NAD}^+$ . Other factors enter into consideration, however, in a living cell. The first is that the reactions do not take place under standard conditions, altering the values of reduction potentials. The second is that the reduced electron carriers ( $\text{NADH}$  and  $\text{FADH}_2$ ) are reoxidized. Coupling the reactions we have looked at here to others also makes them less unfavorable.

14. The half reaction of oxidation  $\text{NADH} + \text{H}^+ \rightarrow \text{NAD}^+ + 2\text{H}^+ + 2e^-$  is strongly exergonic ( $\Delta G^{\circ'} = -61.3 \text{ kJ mol}^{-1} = -14.8 \text{ kcal mol}^{-1}$ ), as is the overall reaction  $\text{Pyruvate} + \text{NADH} + \text{H}^+ \rightarrow \text{Lactate} + \text{NAD}^+$  ( $\Delta G^{\circ'} = -25.1 \text{ kJ mol}^{-1} = -6.0 \text{ kcal mol}^{-1}$ ).

## 20.3 Organization of Electron Transport Complexes

- They all contain the heme group, with minor differences in the heme side chains in most cytochromes.
- Cytochromes are proteins of electron transport; the heme ion alternates between the  $\text{Fe(II)}$  and  $\text{Fe(III)}$  states. The function of hemoglobin and myoglobin is oxygen transport and storage, respectively. The iron remains in the  $\text{Fe(II)}$  state.
- Coenzyme Q is not bound to any of the respiratory complexes. It moves freely in the inner mitochondrial membrane.
- A part of Complex II catalyzes the conversion of succinate to fumarate in the citric acid cycle.
- Three of the four respiratory complexes generate enough energy to phosphorylate ADP to ATP. Complex II is the sole exception.
- Cytochrome  $c$  is not tightly bound to the mitochondrial membrane and can easily be lost in the course of cell fractionation. This protein is so similar in most aerobic organisms that cytochrome  $c$  from one source can easily be substituted for that from another source.
- $\text{Succinate} + \frac{1}{2}\text{O}_2 \rightarrow \text{Fumarate} + \text{H}_2\text{O}$
- The components are in the proper orientation for the electrons to be transferred rapidly from one component to the next; if the components were in solution, speed would be limited to the rate of diffusion. A second advantage, which is actually a necessity, is that the components are properly positioned to facilitate the transport of protons from the matrix to the intermembrane space.
- From an evolutionary standpoint, two different functions can be performed by identical structures or by structures that are close to identical, with only minor differences in the protein moieties. The organism saves a considerable amount of energy by not having to evolve—and to operate—two pathways.
- The key point here is not the active site, which has a low tolerance for mutations, but the molecules with which the proteins in question are associated. Cytochromes are membrane-bound and must associate with other members of the electron transport chain; most mutations are likely to interfere with the close fit, and thus they are not preserved (because they are lethal). Globins, although soluble, still form some associations, so more mutations can be tolerated, with some limits. Hydrolytic enzymes are soluble and not likely to associate with other polypeptides except substrates. They can tolerate a higher proportion of mutations.
- Having mobile electron carriers in addition to membrane-bound respiratory complexes allows electron transport to use the most readily available complex rather than to use the same one all the time.
- The Q cycle allows for a smooth transition from two-electron carriers ( $\text{NADH}$  and  $\text{FADH}_2$ ) to one-electron carriers (cytochromes).
- The protein environment of the iron differs in each of the cytochromes, causing differences in the reduction potential.
- All the reactions in the electron transport chain are electron-transfer reactions, but some of the reactants and products inherently transfer either one or two electrons, as the case may be.
- The heme groups differ slightly in the various kinds of cytochromes. This is the main difference, with some modification due to the different protein environments.
- Respiratory complexes contain a number of proteins, some of them quite large. This is the first difficulty. Like most proteins

bound to membranes, the components of respiratory complexes are easily denatured on removal from their environment.

#### 20.4 The Connection between Electron Transport and Phosphorylation

31. The  $F_1$  portion of the mitochondrial ATP synthase, which projects into the matrix, is the site of ATP synthesis.
32. The  $F_0$  portion of mitochondrial ATP synthase lies within the inner mitochondrial membrane, but the  $F_1$  portion projects into the matrix.
33. The P/O ratio gives the number of moles of  $P_i$  consumed in the reaction  $ADP + P_i \rightarrow ATP$  for each mole of oxygen atoms consumed in the reaction  $\frac{1}{2}O_2 + 2H^+ \rightarrow 2H_2O$ . It is a measure of the coupling of ATP production to electron transport.
34. The  $F_1$  part of mitochondrial ATP synthase has a stationary domain (the  $\alpha_3\beta_3\delta$  domain) and a domain that rotates (the  $\gamma\epsilon$  domain). This is exactly the arrangement needed for a motor.
35. A P/O ratio of 1.5 can be expected because oxidation of succinate passes electrons to coenzyme Q via a flavoprotein intermediate, bypassing the first respiratory complex.
36. Exact values for P/O ratios are difficult to determine because of the complexity of the systems that pump protons and phosphorylate ADP. The number of ADP molecules phosphorylated is directly related to the number of protons pumped across the membrane. This figure has been a matter of some controversy. It has been difficult for chemists and biochemists to accept uncertain stoichiometry.
37. The difficulties in determining the number of protons pumped across the inner mitochondrial membrane by respiratory complexes are those inherent in working with large assemblies of proteins that must be bound in a membrane environment to be active. As experimental methods improve, the task becomes less difficult.

#### 20.5 The Mechanism of Coupling in Oxidative Phosphorylation

38. The chemiosmotic coupling mechanism is based on the difference in hydrogen ion concentration between the intermembrane space and the matrix of actively respiring mitochondria. The hydrogen ion gradient is created by the proton pumping that accompanies the transfer of electrons. The flow of hydrogen ions back into the matrix through a channel in the ATP synthase is directly coupled to the phosphorylation of ADP.
39. An intact mitochondrial membrane is necessary for compartmentalization, which in turn is necessary for proton pumping.
40. Uncouplers overcome the proton gradient on which oxidative phosphorylation depends.
41. In chemiosmotic coupling, the proton gradient is related to ATP production. The proton gradient leads to conformational changes in a number of proteins, releasing tightly bound ATP from the synthase as a result of the conformational change.
42. Dinitrophenol is an uncoupler of oxidative phosphorylation. The rationale was to dissipate energy as heat.
43. The energy released as protons pass through the  $F$  particles is actually used to cause conformational changes in the  $F_1$  proteins, thereby releasing ATP. The “tight” conformation (one of three) provides a hydrophobic environment in which ADP is phosphorylated by adding  $P_i$  without requiring *immediate* energy.

#### 20.6 Respiratory Inhibitors Can Be Used to Study Electron Transport

44. (a) Azide inhibits the transfer of electrons from cytochrome  $aa_3$  to oxygen.  
(b) Antimycin A inhibits the transfer of electrons from cytochrome  $b$  to coenzyme Q in the Q cycle.

- (c) Amytal inhibits the transfer of electrons from NADH reductase to coenzyme Q.
- (d) Rotenone inhibits the transfer of electrons from NADH reductase to coenzyme Q.
- (e) Dinitrophenol is an uncoupler of oxidative phosphorylation.
- (f) Gramicidin A is an uncoupler of oxidative phosphorylation.
- (g) Carbon monoxide inhibits the transfer of electrons from cytochrome  $aa_3$  to oxygen.

45. Methods exist to determine the amounts of the oxidized and reduced components of the electron transport chain present in a sample. If a respiratory inhibitor is added, the reduced form of the component before the blockage point in the chain accumulates as does the oxidized form of the component immediately after the blockage point.

46. Uncouplers overcome the proton gradient created by electron transport, whereas respiratory inhibitors block the flow of electrons.

#### 20.7 Shuttle Mechanisms

47. The complete oxidation of glucose produces 30 molecules of ATP in muscle and brain and 32 ATP in liver, heart, and kidney. The underlying reason is the different shuttle mechanisms for transfer to mitochondria of electrons from the NADH produced in the cytosol by glycolysis.
48. The transport “product” (in the matrix) of the malate–aspartate shuttle is NADH, whereas that of the glycerol–phosphate shuttle is  $FADH_2$ . The latter shuttle can thus go *against* a transmembrane NADH concentration gradient, whereas the former cannot.

#### 20.8 The ATP Yield from Complete Oxidation of Glucose

49. (a) 34  
(b) 32  
(c) 13.5  
(d) 17  
(e) 2.5  
(f) 12.5
50. The maximum yield of ATP, to the nearest whole number, is 3.

$$102.3 \text{ kJ released} \times \frac{1 \text{ ATP}}{30.5 \text{ kJ}} = 3.35 \text{ ATP}$$

One ATP is actually produced, so the efficiency of the process is

$$\frac{1 \text{ ATP}}{3 \text{ ATP}} \times 100 = 33.3\%$$

## Chapter 21

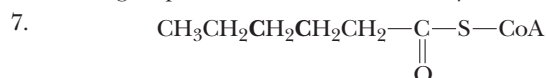
### 21.1 Lipids Are Involved in the Generation and Storage of Energy

1. (a) For mobile organisms—such as a migrating hummingbird—weight can be a critical factor, and packing the most energy into the least weight is decidedly advantageous. A 2.5-g hummingbird needs to add about 2 g of fat for migration energy, which would increase body weight by 80%. The equivalent amount of energy stored as glycogen would be about 5 g, which would increase its body weight by 200%; the bird would never get off the ground!  
(b) For immobile plants, weight is not a critical factor, and it takes more energy to make fat or oil than it does to make starch. (The second law of thermodynamics would dictate that the energy obtained from oil would be less than that expended making oil. You can verify this numerically if you wish.) In the case of plant *seeds*, “compact” energy is beneficial, because

the seed must be self-sufficient until enough growth has occurred to permit photosynthesis.

## 21.2 Catabolism of Lipids

- Phospholipase A<sub>1</sub> hydrolyzes the ester bond to carbon-1 of the glycerol backbone; phospholipase A<sub>2</sub> hydrolyzes the ester bond to carbon-2 of the backbone.
- A hormone signal activates adenylate cyclase, which makes cAMP. This activates protein kinases, which phosphorylate the lipases, thereby activating them.
- Acyl-CoAs are high-energy compounds. An acyl-CoA has sufficient energy to initiate the  $\beta$ -oxidation process. The CoA is also a tag indicating that the molecule is destined for oxidation.
- Acyl groups are esterified to carnitine to cross the inner mitochondrial membrane. There are transesterification reactions from the acyl-CoA to carnitine and from acylcarnitine to CoA (see Figure 21.5).
- Acyl-CoA dehydrogenase removes hydrogens from adjacent carbons, creating a double bond and using FAD as coenzyme.  $\beta$ -Hydroxy-CoA dehydrogenase oxidizes an alcohol group to a ketone group and uses NAD<sup>+</sup> as a coenzyme.



The two carbons shown in boldface type are the ones that will have the double bond between them. The orientation will be *trans*.

- Seven carbon-carbon bonds are broken in the course of  $\beta$ -oxidation (see Figure 21.6).
- In the liver, glycogen breakdown and gluconeogenesis would occur. In the muscle, glycogen breakdown and glycolysis would occur.

## 21.3 The Energy Yield from the Oxidation of Fatty Acids

- One obtains 6.7 ATP per carbon and 0.42 ATP per gram for stearic acid versus 5 ATP per carbon and 0.17 ATP per gram for glucose. More energy is available from stearic acid than from glucose.
- The processing of the acetyl-CoA through the citric acid cycle and the electron transport chain produces more energy than the processing of the NADH and FADH<sub>2</sub> produced during  $\beta$ -oxidation.
- From seven cycles of  $\beta$ -oxidation: 8 acetyl-CoA, 7 FADH<sub>2</sub>, and 7 NADH. From the processing of 8 acetyl-CoA in the citric acid cycle: 8 FADH<sub>2</sub>, 24 NADH, and 8 GTP. From reoxidation of all FADH<sub>2</sub> and NADH: 22.5 ATP from 15 FADH<sub>2</sub>, 77.5 ATP from 31 NADH. From 8 GTP: 8 ATP. Subtotal: 108 ATP. A 2-ATP equivalent was used in the activation step. Grand total: 106 ATP. The grand total for stearic acid was 120 ATP.
- The humps of camels contain lipids that can be degraded as a source of metabolic water, rather than water as such.

## 21.4 Catabolism of Unsaturated Fatty Acids and Odd-Carbon Fatty Acids

- For an odd-chain fatty acid,  $\beta$ -oxidation proceeds normally until the last round. When five carbons are left, that round of  $\beta$ -oxidation releases one acetyl-CoA and one propionyl-CoA. Propionyl-CoA cannot be further metabolized by  $\beta$ -oxidation; however, a separate set of enzymes converts propionyl-CoA into succinyl-CoA, which can then enter the citric acid cycle.
- False. The oxidation of unsaturated fatty acids to acetyl-CoA requires a *cis-trans* isomerization and an epimerization, reactions that are not found in the oxidation of saturated fatty acids.
- For a monounsaturated fatty acid, an additional enzyme is needed, the enoyl-CoA isomerase.

- For a polyunsaturated fatty acid, two additional enzymes are needed, the enoyl-CoA isomerase and 2,4-dienoyl-CoA reductase.
- From seven cycles of  $\beta$ -oxidation: 7 acetyl-CoA, 1 propionyl-CoA, 7 FADH<sub>2</sub>, 7 NADH. From the processing of 7 acetyl-CoA in the citric acid cycle: 7 FADH<sub>2</sub>, 21 NADH, and 7 GTP. From the processing of the propionyl-CoA: -1 ATP for conversion to succinyl-CoA, -1 GTP from the citric acid cycle, and 1 NADH and 1 FADH<sub>2</sub> from the citric acid cycle. From reoxidation of all FADH<sub>2</sub> and NADH: 22.5 ATP from 15 FADH<sub>2</sub>, and 72.5 ATP from 29 NADH. From 8 GTP: 8 ATP. Subtotal: 103 ATP. Subtract a 2-ATP equivalent used in activation step and a 1-ATP equivalent used in the conversion to succinyl-CoA for a grand total of 100 ATP.
- An 18-carbon saturated fatty acid yields 120 ATP. For a mono-unsaturated fatty acid, the double bond eliminates the step that produces FADH<sub>2</sub>, so there would be 1.5 ATP less for oleic acid, or 118.5 ATP total.
- An 18-carbon saturated fatty acid yields 120 ATP. For a diunsaturated fatty acid with the bonds in the  $\Delta^9$  and  $\Delta^{12}$  positions, the first double bond eliminates an FADH<sub>2</sub>. The second double bond uses an NADPH, which we are guessing is the same cost as using an NADH. Thus a total of 4 ATP are lost, compared with a saturated fatty acid, so the total is 116 ATP.
- It would take seven cycles of  $\beta$ -oxidation to release 14 carbons as acetyl-CoA, with the last three being released as propionyl-CoA.
- Fats cannot produce a net yield of glucose because they must enter the citric acid cycle as the two-carbon unit acetyl-CoA. In the first few steps, two carbons are released as CO<sub>2</sub>. However, an odd-chain fatty acid can be considered partially glucogenic because the final three carbons become succinyl-CoA and enter the citric acid cycle after the decarboxylation steps. Thus, if an extra succinyl-CoA is added, it can then be drawn off later as malate and used for gluconeogenesis without removing the steady-state level of citric acid cycle intermediates.

## 21.5 Ketone Bodies

- Ketones are produced when there is an imbalance in lipid catabolism, compared with carbohydrate catabolism. If fatty acids are being  $\beta$ -oxidized to produce acetyl-CoA, but there is insufficient oxaloacetate because it is being drawn off for gluconeogenesis, the acetyl-CoA molecules combine to form ketone bodies.
- Two acetyl-CoA molecules combine to form acetoacetyl-CoA. This can then release coenzyme A to yield acetoacetate, which can be converted either to  $\beta$ -hydroxybutyrate or to acetone.
- If the reason for passing out is uncontrolled diabetes, the doctor expects to smell acetone on the breath, since the otherwise unused sugars are being converted to fats and ketone bodies.
- Ethanol is converted to acetaldehyde and then to acetic acid. Humans can use that acetic acid only for energy, or they can convert it to fatty acids and other lipids.
- The metallic taste may be due to acetone, which means that your friend may have a mild state of ketosis. Ask if your friend has consulted a doctor about the diet regimen, and perhaps recommend either backing off from such a low-calorie diet or drinking more water to flush the system more thoroughly.

## 21.6 Fatty-Acid Biosynthesis

- The two pathways have in common the involvement of acetyl-CoA and thioesters, and each round of breakdown or synthesis involves two-carbon units. The differences are many: malonyl-CoA is involved in biosynthesis, not in breakdown; thioesters involve CoA in breakdown and involve acyl carrier proteins in biosynthesis; biosynthesis occurs in the cytosol, but breakdown



- occurs in the mitochondrial matrix; breakdown is an oxidative process that requires  $\text{NAD}^+$  and FAD and produces ATP by electron transport and oxidative phosphorylation, whereas biosynthesis is a reductive process that requires NADPH and ATP.
29. Step 1: biotin is carboxylated using bicarbonate ion ( $\text{HCO}_3^-$ ) as the source of the carboxyl group. Step 2: the carboxylated biotin is brought into proximity with enzyme-bound acetyl-CoA by a biotin carrier protein. Step 3: the carboxyl group is transferred to acetyl-CoA, forming malonyl-CoA.
  30. It is a molecule that commits itself to fatty-acid synthesis. It is also a potent inhibitor of carnitine acyltransferase I, thereby shutting down  $\beta$ -oxidation.
  31. ACP, citrate, cytosol, *trans* double bonds, D-alcohols,  $\beta$ -reduction, NADPH, malonyl-CoA (except for one acetyl-CoA), and a multifunctional enzyme complex.
  32. In  $\beta$ -oxidation, FAD is the coenzyme for the first oxidation reaction, while  $\text{NAD}^+$  is the coenzyme for the second. In fatty-acid synthesis, NADPH is the coenzyme for both. The  $\beta$ -hydroxy-acyl group in  $\beta$ -oxidation has the L-configuration, while it has the D-configuration in fatty acid synthesis.
  33. Both have a phosphopantetheine group at the active end. In coenzyme A, this group is attached to 2'-phospho-AMP; in ACP, it is attached to a serine residue of a protein.
  34. ACP is a molecule that earmarks acyl groups for fatty-acid synthesis. It can be managed separately from acyl-CoA groups. Also, the ACP attaches to the acyl groups like a "swinging arm" that tethers it to the fatty-acid synthase complex.
  35. Linoleate and linolenate cannot be synthesized by the body and must therefore be obtained from dietary sources. Mammals cannot produce a double bond beyond carbon atom 9 of fatty acids.
  36. Acyl-CoA intermediates are essential in the conversion of fatty acids to other lipids.
  37. Acetyl groups condense with oxaloacetate to form citrate, which can cross the mitochondrial membrane. Acetyl groups are regenerated in the cytosol by the reverse reaction.
  38. If acetyl-carnitine forms in the matrix of the mitochondrion, it can be translocated to the cytosol via the carnitine translocase. Thus, this could represent another way of shuttling acetyl units out of the mitochondria for synthesis.
  39. Energy is needed to condense an acetyl group to the growing fatty acid. In theory, such could be done with acetyl-CoA, using ATP. In practice, the ATP is used to convert acetyl-CoA to malonyl-CoA; the condensation of the acetyl moiety of malonyl-CoA is driven in part by the accompanying decarboxylation and requires no additional energy. A possible reason for this is to avoid a metabolic confusion of pathways, perhaps particularly important in (uncompartmented) prokaryotes; one could envision an acetyl-CoA from degradation being used immediately for synthesis. Malonyl-CoA says "synthesis"; acetyl-CoA says "degradation."
  40. (a) The lipoate "swinging arm" of the pyruvate dehydrogenase complex.  
(b) The "arm" or ACP carries the group to be acted on from one enzyme to another (avoiding a diffusion-limited process and also positioning key groups correctly). In the case of the ACP, the group to be acted on ( $\beta$ -carbon) is always the same distance from the ACP, regardless of the length of the growing fatty acid, and thus the critical group is always in proximity to the active sites of the several pertinent enzymes.
- ### 21.7 Synthesis of Acylglycerols and Compound Lipids
41. The glycerol comes from degradation of other acylglycerols or from glycerol-3-phosphate derived from glycolysis.
  42. The activating group found on the acylglycerol is cytidine diphosphate.
  43. In prokaryotes, CTP reacts with phosphatidic acid to give a CDP-diacylglycerol. This reacts with serine to give phosphatidylserine, which decarboxylates to phosphatidylethanolamine. In eukaryotes, CDP-ethanolamine reacts with a diacylglycerol to give phosphatidylethanolamine.
- ### 21.8 Cholesterol Biosynthesis
44. In steroid biosynthesis, three acetyl-CoA molecules condense to form the six-carbon mevalonate, which then gives rise to a five-carbon isoprenoid unit. A second and then a third isoprenoid unit condense, giving rise to a 10-carbon and then a 15-carbon unit. Two of the 15-carbon units condense, forming a 30-carbon precursor of cholesterol.
  45. See Figure 21.24.
  46. Bile acids and steroid hormones.
  47. All steroids have a characteristic fused-ring structure, implying a common biosynthetic origin.
  48. One oxygen atom from  $\text{O}_2$  is needed to form the epoxide. The NADPH is needed to reduce the other oxygen atom to water.
  49. Cholesterol is nonpolar and cannot dissolve in blood, which is an aqueous medium.
  50. Bile salts are made from cholesterol, and cholesterol is taken from the body into the intestine in the bile fluid.
- ## Chapter 22
- ### 22.1 Chloroplasts Are the Site of Photosynthesis
1. In the fall, the chlorophyll in leaves is lost, and the red and yellow colors of the accessory pigments become visible, accounting for fall foliage colors.
  2. The bean sprouts are grown in the dark to prevent them from turning green; most customers will not purchase green sprouts.
  3. Iron and manganese in chloroplasts; iron and copper in mitochondria. Note that all these are transition metals, which can easily undergo redox reactions.
  4. Both chloroplasts and mitochondria have an inner and outer membrane. Both have their own DNA and ribosomes. Chloroplasts, however, have a third membrane, the thylakoid membrane.
  5. Chlorophyll has a cyclopentanone ring fused to the tetrapyrrole ring, a feature that does not exist in heme. Chlorophyll contains magnesium, whereas heme contains iron. Chlorophyll has a long side chain based on isoprenoid units, which is not found in heme.
  6. Only a relatively small portion of the visible spectrum is absorbed by chlorophylls. The accessory pigments absorb light at additional wavelengths. As a result, most of the visible spectrum can be harnessed in light-dependent reactions.
  7. It is one more piece of evidence that is consistent with the evolution of chloroplasts from independent bacterial organisms.
- ### 22.2 Photosystems I and II and the Light Reactions of Photosynthesis
8. By and large, the synthesis of NADPH in chloroplasts is the reverse of NADH oxidation in mitochondria. The net electron flow in chloroplasts is the reverse of that in the mitochondria, although different carriers are involved.

9. When light impinges on the reaction center of *Rhodospseudomonas*, the special pair of chlorophylls there is raised to an excited energy level. An electron is passed from the special pair to accessory pigments, first pheophytin, then menaquinone, and finally to ubiquinone. The electron lost by the special pair of chlorophylls is replaced by a soluble cytochrome, which diffuses away. The separation of charge represents stored energy (see Figure 21.9).
10. In photosystem I and in photosystem II, light energy is needed to raise the reaction-center chlorophylls to a higher energy level. Energy is needed to generate strong enough reducing agents to pass electrons to the next of the series of components in the pathway.
11. No. Most chlorophylls are light-harvesting molecules that transfer energy to the special pair that takes part in the light reactions.
12. The electron transport chain in chloroplasts, like that in mitochondria, consists of proteins, such as plastocyanin, and protein complexes, such as the cytochrome  $b_6-f$  complex. It also contains mobile electron carriers, such as pheophytin and plastoquinone (equivalent to coenzyme Q), which is also true of the mitochondrial electron transport chain.
13. Probably the electron transport chain in chloroplasts. Chloroplasts generate molecular oxygen; mitochondria use it. The early atmosphere almost certainly lacked molecular oxygen. Only when photosynthesis introduced oxygen into the atmosphere would oxygen be needed.
14. Electron transport and ATP production are coupled to each other by the same mechanism in mitochondria and chloroplasts. In both cases, the coupling depends on the generation of a proton gradient across the inner mitochondrial membrane or across the thylakoid membrane, as the case may be.
15. In mitochondria, both a proton gradient (chemical) and an electrochemical gradient (based on charge) are formed, both contributing to the total potential energy. In chloroplasts, only a proton gradient is formed, because ions move across the thylakoid membrane and neutralize charge. The proton gradient alone is considerably less efficient.
16. With very few exceptions, life directly or indirectly depends on photosynthesis. The electric current is the flow of electrons from water to  $\text{NADP}^+$ , a light-requiring process. The "current" continues in the light-independent reactions, with electrons flowing from NADPH to bisphosphoglycerate, which ultimately yields glucose.
17. Photosystem II requires more energy than photosystem I. The shorter wavelength of light means a higher frequency. Frequency, in turn, is directly proportional to energy.
18. It is quite reasonable to list reduction potentials for the electron-transfer reactions of photosynthesis. They are entirely analogous to the electron-transfer reactions in mitochondria, for which we listed standard reduction potentials in Chapter 20.
19. A photosynthetic reaction center is analogous to a battery because its reactions produce a charge separation. The charge separation is comparable to the stored energy of the battery.
20. The electron transport chains of mitochondria and chloroplasts are similar. In mitochondria, antimycin A inhibits electron transfer from cytochrome  $b$  to coenzyme Q in the Q cycle. By analogy, it can be argued that antimycin A inhibits electron flow from plastoquinone to cytochrome  $b_6-f$ . A Q cycle may also operate in chloroplasts.
21. Oxygen produced in photosynthesis comes from water. The oxygen-evolving complex is part of the series of electron-transfer

reactions from water to NADPH. Carbon dioxide is involved in the dark reactions, which are different reactions that take place in another part of the chloroplast.

22. It is well established that the path of electrons in photosynthesis goes from photosystem II to photosystem I. The reason for the nomenclature is that photosystem I is easier to isolate than photosystem II and was studied more extensively at an earlier date.
23. It would take much work to establish the number of protons pumped across the thylakoid membrane. This is partly the result of experience with mitochondria and partly a prediction based on the greater complexity of structure in the chloroplast.
24. The oxygen-evolving complex of photosystem II passes through a series of five oxidation states (designated as  $S_0$  through  $S_4$ ) in the transfer of four electrons in the process of evolving oxygen (Figure 22.6). One electron is passed from water to photosystem II for each quantum of light. In the process, the components of the reaction center go successively through oxidation states  $S_1$  through  $S_4$ . The  $S_4$  decays spontaneously to the  $S_0$  state and, in the process, oxidizes two water molecules to one oxygen molecule. Four protons are released simultaneously.
25. When the loosely bound cytochrome diffuses away, a charge separation is induced. This separation of charge represents stored energy.
26. The similarity of ATP synthase in chloroplasts and mitochondria supports the idea that both may have arisen from free-living bacteria.

### 22.3 Photosynthesis and ATP Production

27. In cyclic photophosphorylation, the excited chlorophyll of photosystem I passes electrons directly to the electron transport chain that normally links photosystem II to photosystem I. This electron transport chain is coupled to ATP production (see Figure 22.8).
28. Both depend on a proton gradient, resulting from the flow of electrons. In chloroplasts, protons come from the splitting of water to produce oxygen. In mitochondria, protons come from the oxidation of NADH and ultimately consume oxygen and produce water.
29. The proton gradient is created by the operation of the electron transport chain that links the two photosystems in noncyclic photophosphorylation.
30. ATP can be produced by chloroplasts in the absence of light if some way exists to form a proton gradient.
31. Cyclic photophosphorylation can take place when the plant needs ATP but does not have a great need for NADPH. Noncyclic photophosphorylation can take place when the plant needs both.

### 22.4 Evolutionary Implications of Photosynthesis with and without Oxygen

32. Many electron donors other than water are possible in photosynthesis. This is especially the case in bacteria, whose photosystems do not have strong enough oxidizing agents to oxidize water. Some of the alternative electron donors are  $\text{H}_2\text{S}$  and organic compounds.
33. A prokaryotic organism that contains both chlorophyll  $a$  and chlorophyll  $b$  could be a relic of an evolutionary way station in the development of chloroplasts.

### 22.5 Dark Reactions of Photosynthesis Fix $\text{CO}_2$

34. Rubisco is the principal protein in chloroplasts in all green plants. This wide distribution makes it likely to be the most abundant protein in nature.



35. The amino acid sequence of the catalytic subunits of rubisco is encoded by chloroplast genes, whereas that of the regulatory subunits is encoded by nuclear genes.
36. Gluconeogenesis and the pentose phosphate pathway have a number of reactions similar to those of the dark reactions of photosynthesis.
37. From the standpoint of thermodynamics, the production of sugars in photosynthesis is the reverse of the complete oxidation of a sugar such as glucose to  $\text{CO}_2$  and water. The complete oxidation reaction produces six moles of  $\text{CO}_2$  for each mole of glucose oxidized. To get the energy change for the fixation of one mole of  $\text{CO}_2$ , change the sign of the energy for the complete oxidation of glucose and divide by 6.
38. Glucose synthesized by photosynthesis is not uniformly labeled because only one molecule of  $\text{CO}_2$  is incorporated into each molecule of ribulose-1,5-bisphosphate, which then goes on to give rise to sugars.
39. If rubisco was one of the first protein enzymes to arise early in the evolution of life, it may not have the efficiency of protein enzymes that evolved later, when evolution was more dependent on modifying and adapting existing proteins.
40. Their DNA is circular. Their ribosomes are more like those of bacteria than those of eukaryotes. Their aminoacyl-tRNA synthetases use bacterial tRNAs but not eukaryotic tRNAs. In general, they do not have introns in their genomes. Their mRNA uses a Shine-Dalgarno sequence.
41. The pathway borrows heavily from the nonoxidative branch of the pentose phosphate pathway and from gluconeogenesis. Without doubt, the pathways yield sugars as well as NADPH for reductive biosynthesis. Thus, only a few new enzymes would have to evolve through mutations to enable the complete Calvin cycle to function.
42. Atmospheric oxygen is a consequence of photosynthesis. Rubisco evolved before there was a significant amount of oxygen in the atmosphere.
43. The condensation of ribulose-1,5-bisphosphate with carbon dioxide to form two molecules of 3-phosphoglycerate is the actual carbon dioxide fixation. The rest of the Calvin cycle regenerates ribulose-1,5-bisphosphate.
44. Organisms would need only a few mutations giving rise to the enzymes unique to the Calvin cycle. The rest of the pathway is already in place.
45. Six molecules of carbon dioxide fixed in the Calvin cycle do not end up in the same glucose molecule. However, labeling experiments show that six carbon atoms are incorporated into sugars for every six carbon dioxide molecules that enter the Calvin cycle.

## 22.6 $\text{CO}_2$ Fixation in Tropical Plants

46. In tropical plants, the  $\text{C}_4$  pathway is operative in addition to the Calvin cycle.
47. In  $\text{C}_4$  plants, when  $\text{CO}_2$  enters the leaf through pores in the outer cells, it reacts first with phosphoenolpyruvate to produce oxaloacetate and  $\text{P}_i$  in the mesophyll cells of the leaf. Oxaloacetate is reduced to malate, with the concomitant oxidation of NADPH. Malate is then transported to the bundle-sheath cells (the next layer) through channels that connect two kinds of cells. These reactions do not take place in  $\text{C}_3$  plants.
48. Photorespiration is a pathway in which glycolate is a substrate oxidized by rubisco acting as an oxygenase, rather than as a carboxylase. Photorespiration is not completely understood.

49. Three reasons come to mind. (1) Light energy is usually not limiting. (2) The plants have small pores to prevent water loss, but this also limits  $\text{CO}_2$  uptake. (3) The  $\text{C}_4$  pathway allows for increasing the  $\text{CO}_2$  concentration in the inner chloroplast, which would not be otherwise possible with the small pores.
50. Most plants would be more productive in the absence of photorespiration. There is another side to this picture, however. The oxygenase activity appears to be an unavoidable, wasteful activity of rubisco. Photorespiration is a salvage pathway that saves some of the carbon that would be lost due to the oxygenase activity of rubisco. Photorespiration is essential to plants even though the plant pays the price in loss of ATP and reducing power; mutations that affect this pathway can be lethal.

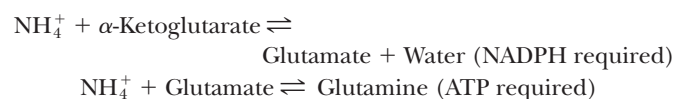
## Chapter 23

### 23.1 Nitrogen Metabolism: An Overview

1. Nitrogen-fixing bacteria (symbiotic organisms that form nodules on the roots of leguminous plants, such as beans and alfalfa) and some free-living microbes and cyanobacteria can fix nitrogen. Plants and animals cannot.

### 23.2 Nitrogen Fixation

2. Nitrogen is fixed by the nitrogenase reaction, in which  $\text{N}_2$  is converted to  $\text{NH}_4^+$ . Very few organisms have this enzyme, which can catalyze the breaking of the triple bond in molecular nitrogen. The glutamate dehydrogenase reaction and the glutamine synthase reactions assimilate nitrogen:



3. The chemical synthesis of ammonia from  $\text{H}_2$  and  $\text{N}_2$ .
4.  $\text{N}_2 + 8e^- + 16\text{ATP} + 10\text{H}^+ \rightarrow 2\text{NH}_4^+ + 16\text{ADP} + 16\text{P}_i + \text{H}_2$  is the half reaction for reduction via nitrogenase. The oxidation reaction varies with species.
5. The nitrogenase complex is made up of ferredoxin, dinitrogenase reductase, and nitrogenase. Dinitrogenase reductase is an iron-sulfur protein, whereas nitrogenase is an iron-molybdenum protein. The Fe-S protein is a dimer ("the iron butterfly"), with the iron-sulfur cluster located at the butterfly's head. The nitrogenase is even more complicated, with several types of subunits arranged into tetramers.

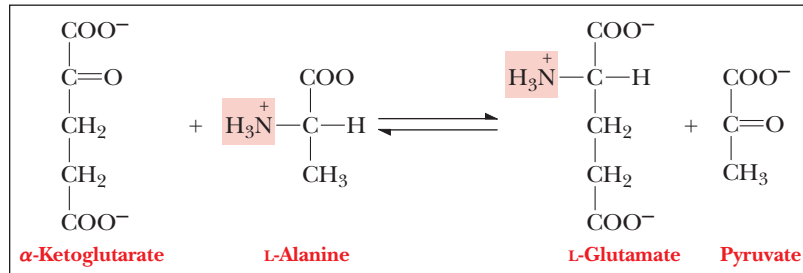
### 23.3 Feedback Inhibition in Nitrogen Metabolism

6. Pathways that use nitrogen to make amino acids, purines, and pyrimidines are controlled by feedback inhibition. The final product, such as CTP, inhibits the first or an early step in its synthesis.
7. Feedback control mechanisms slow down long biosynthetic pathways at or near their beginnings, saving energy for the organism.
8. Because all the components of a cycle are regenerated, only small amounts ("catalytic quantities") are needed. This is important from an energy standpoint and, perhaps with some compounds, because of insolubility problems.

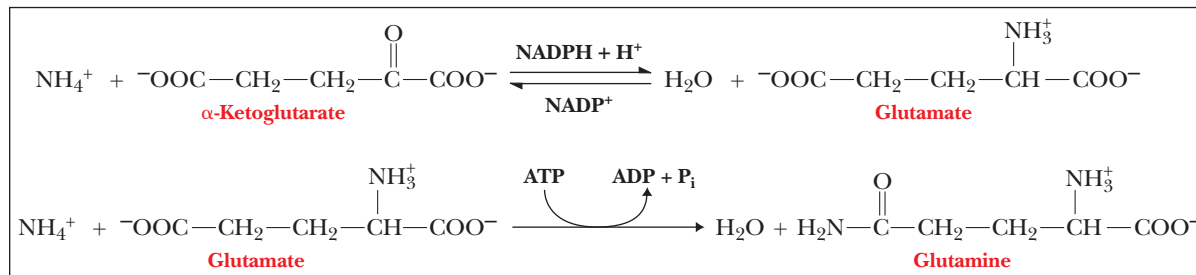
### 23.4 Amino Acid Biosynthesis

9. They are all interrelated.  $\alpha$ -Ketoglutarate can be changed to glutamate via transamination or glutamate dehydrogenase. Glutamine synthetase makes glutamine out of glutamate.

10.



11.



12. Glutamine synthetase catalyzes the following reaction and uses energy:  $\text{NH}_4^+ + \text{Glutamate} + \text{ATP} \rightarrow \text{Glutamine} + \text{ADP} + \text{P}_i + \text{H}_2\text{O}$ . Glutaminase catalyzes the following reaction and does not use energy directly:  $\text{Glutamine} + \text{H}_2\text{O} \rightarrow \text{Glutamate} + \text{NH}_4^+$ .
13. See Figure 23.8.
14. The principal ones are tetrahydrofolate and Sadenosylmethionine.
15. See Figure 23.11.
16. Conversion of homocysteine to methionine using Sadenosylmethionine as the methyl donor gives no net gain; one methionine is needed to produce another methionine.
17.  $\text{Glutamate} + \alpha\text{-Keto acid} \rightarrow \alpha\text{-Ketoglutarate} + \text{Amino acid}$
18. See the Sadenosylmethionine structure in Figure 23.15. The reactive methyl group is indicated.
19. Sulfanilamide inhibits folic acid biosynthesis.
20. Methionine can play a dual role. In addition to providing a hydrophobic group, methionine (in the form of Sadenosylmethionine) can act as a methyl group donor.
- 23.5 Essential Amino Acids**
21. The essential amino acids are those with branched chains, aromatic rings, or basic side chains.
22. In both cases, the requirements are those given in Table 23.1.
- 23.6 Amino Acid Catabolism**
23. Five  $\alpha$ -amino acids are involved directly in the urea cycle (ornithine, citrulline, aspartate, arginosuccinate, and arginine). Of those, only aspartate and arginine are also found in proteins.
24.  $\text{H}^+ + \text{HCO}_3^- + 2\text{NH}_3 + 3\text{ATP} \rightarrow \text{NH}_2\text{CONH}_2 + 2\text{ADP} + 2\text{P}_i + \text{AMP} + \text{PP}_i + 2\text{H}_2\text{O}$ . The urea cycle is linked to the citric acid cycle by fumarate and by aspartate, which can be converted to malate by transamination (see Figure 23.19).
25. Ornithine is similar to lysine, but it has one fewer methylene group in the side chain. Citrulline is a keto version of arginine with a side chain  $\text{C}=\text{NH}_2^+$  replaced by  $\text{C}=\text{O}$ .
26. Aspartate and arginosuccinate are the amino acids that link the two pathways. Aspartate is made by transamination of OAA. The aspartate then combines with citrulline to form arginosuccinate, which then releases a fumarate to go back to the TCA cycle.
27. Each round of the urea cycle costs 4 ATP, two to make carbamoyl-phosphate and effectively two ( $\text{ATP} \rightarrow \text{AMP}$ ) to make arginosuccinate.
28. It is controlled by a special effector molecule, N-acetylglutamate, which is itself controlled by levels of arginine.
29. When arginine levels build up, it means that the urea cycle is going too slow and not enough carbamoyl-phosphate is available to react with ornithine.
30. Glutamate brings ammonia groups to the matrix of the mitochondria for the urea cycle. High levels of glutamate stimulate the urea cycle.
31. Glucogenic amino acids are degraded to pyruvate or one of the citric acid cycle intermediates found after the decarboxylation steps, such as succinate or malate. Ketogenic amino acids are degraded to acetyl-CoA or acetoacetyl-CoA.
32. (a) Glucogenic  
(b) Glucogenic  
(c) Glucogenic  
(d) Ketogenic  
(e) Glucogenic  
(f) Ketogenic
33. Fish excrete excess nitrogen as ammonia, and birds excrete it as uric acid. Mammals excrete it as urea.
34. Because ostriches don't fly, one could argue that they would excrete their excess nitrogen as urea. On the other hand, they are birds, and as such probably have the same metabolism of their lighter counterparts, and might likely excrete it as uric acid.
35. The amounts of arginine necessary in the urea cycle are only catalytic. If arginine from the cycle is used for protein synthesis, the cycle becomes depleted.
36. A high-protein diet leads to increased production of urea. Drinking more water increases the volume of urine, ensuring elimination of the urea from the body with less strain on the kidneys than if urea were at a higher concentration.
37. The metabolism of amino acids encourages urine formation and actually a greater thirst and need for water.

38. Several enzymes, resulting from mutations, are needed for the urea cycle. Most mutations tend to be lost unless they provide some survival value. It seems improbable that all the mutations needed for all the enzymes of the cycle would arise nearly simultaneously. However, the origin of the cycle can be rather easily explained on the premise that only one new enzyme (arginase) was needed. The other enzymes of the cycle are needed for the biosynthesis of arginine. As a component of proteins, arginine was presumably needed before there was a need for a urea cycle. This is an example of nature using features already available to bring about a new function.

### 23.7 Purine Biosynthesis

39. Since folic acid is critical to the formation of purines, antagonists of folic acid metabolism are used as chemotherapy drugs to inhibit nucleic acid synthesis and cell growth. Rapidly dividing cells, such as those found in cancer and tumors, are more susceptible to these antagonists.
40. All four nitrogen atoms of the purine ring are derived from amino acids: two from glutamine, one from aspartate, and one from glycine. Two of the five carbon atoms (adjacent to the glycine nitrogen) also come from glycine, two more come from tetrahydrofolate derivatives, and the fifth comes from  $\text{CO}_2$ .
41. In inosine, carbon-6 of the ring is a ketone group; in adenosine, carbon-6 is bound to an amino group.
42. Tetrahydrofolate is a carrier of carbon groups. Two of the carbons in the purine ring are donated by tetrahydrofolate.
43. The conversion of IMP to GMP produces one NADH and uses the equivalent of 2 ATP because an ATP is converted to AMP. Because NADH gives rise to 2.5 ATP if it goes into the electron transport chain, we can say that the conversion results in a net production of ATP.
44. There is a complicated system of feedback inhibition for the production of purine-containing nucleotides. The final products, ATP and GTP, feed back to inhibit the first steps starting from ribose-5-phosphate. In addition, each intermediate, such as AMP or ADP, can also inhibit the first step. Also, each of the three forms for each nucleotide inhibit the committed reaction from IMP that eventually decides which purine nucleotide is made.

### 23.8 Purine Catabolism

45. The purine salvage reaction that produces GMP requires the equivalent of 2 ATP. The pathway to IMP and then to GMP requires the equivalent of 8 ATP.
46. In most mammals, uric acid is converted to allantoinic acid, which is much more water soluble than uric acid.

### 23.9 Pyrimidine Biosynthesis and Catabolism

47. In purine nucleotide biosynthesis, the growing purine ring is covalently bonded to ribose; the ribose is added after the ring is synthesized in pyrimidine nucleotide biosynthesis.
48. Purines break down to various products, depending on the species. These products are then excreted, representing a major means of nitrogen excretion for many organisms. Pyrimidine catabolism yields, in addition to  $\text{NH}_4^+$  and  $\text{CO}_2$ , the salvageable product  $\beta$ -alanine, which is a breakdown product of both cytosine and uracil.

### 23.10 Conversion of Ribonucleotides to Deoxyribonucleotides

49. Both thioredoxin and thioredoxin reductase are proteins involved in the conversion of ribonucleotides to deoxyribonucleotides. Thioredoxin is an intermediate carrier of

electrons and hydrogens, and thioredoxin reductase is the enzyme that catalyzes the process.

### 23.11 Conversion of dUTP to dTTP

50. Fluorouracil substitutes for thymine in DNA synthesis. In rapidly dividing cells, such as cancer cells, the result is the production of defective DNA, causing cell death.
51. The DNA of fast-growing cells, such as those of the hair follicles, is damaged by chemotherapeutic agents.

## Chapter 24

### 24.1 Connections between Metabolic Pathways

- ATP and NADPH are the two molecules that link the most pathways.
- Acetyl-CoA, pyruvate, PEP,  $\alpha$ -ketoglutarate, succinyl-CoA, oxaloacetate, and several sugar phosphates, such as glucose-6-phosphate and fructose-6-phosphate.
- (a) Fructose-6-phosphate—from the pentose phosphate pathway (PPP).  
(b) Oxaloacetate—to phosphoenolpyruvate in gluconeogenesis, to and from aspartate, to the glyoxylate cycle via citrate.  
(c) Glucose-6-phosphate—to PPP, to and from glycogen in animals, to starch in plants.  
(d) Acetyl-CoA—to and from fatty acids, to steroids (and isoprenoids), some amino acid degradations, to the glyoxylate cycle via citrate.  
(e) Glyceraldehyde-3-phosphate—to reverse PPP.  
(f)  $\alpha$ -Ketoglutarate—to and from glutamate.  
(g) Dihydroxyacetone phosphate—to and from the glycerol moiety of triacylglycerols and phosphoacylglycerols.  
(h) Succinyl-CoA—degradation of fatty acids with odd numbers of carbon atoms, some amino acid degradation.  
(i) 3-Phosphoglycerate—appears in the Calvin cycle.  
(j) Fumarate—some amino acid degradations.  
(k) Phosphoenolpyruvate—from oxaloacetate in gluconeogenesis.  
(l) Citrate—to the glyoxylate cycle, transport across the mitochondrial membrane for fatty acid and steroid synthesis.  
(m) Pyruvate—fermentation, to gluconeogenesis, also to and from alanine.
- When the body breaks down proteins to supply material for gluconeogenesis, the increased urea output results in greater urine production, which uses water stored in the body. Fat metabolism also produces much metabolic water.
- (a) High ATP or NADH concentration and the citric acid cycle: isocitrate dehydrogenase (and the citric acid cycle) would be inhibited. The resulting pileup of acetyl-CoA (or citrate) would stimulate fatty acid and steroid synthesis, gluconeogenesis, and (in plants and some microorganisms) the glyoxylate cycle.  
(b) High ATP concentration and glycolysis: phosphofructokinase-1 (and glycolysis) would be inhibited. Glucose-6-phosphate would pile up, stimulating glycogen (or starch) synthesis, the oxidative pentose phosphate pathway, or glucose formation. (c) High NADPH concentration and the pentose phosphate pathway: the oxidative branch of the pentose phosphate pathway would be inhibited, thus making glucose-6-phosphate available for other purposes. These include glycolysis, glycogen synthesis, glucose synthesis, and



the “reverse” pentose phosphate pathway (yielding only pentose phosphate). (d) High fructose-2,6-bisphosphate concentration and gluconeogenesis: fructose-2,6-bisphosphate inhibits fructose-1,6-bisphosphatase and activates phosphofructokinase-1. Gluconeogenesis would thus be inhibited and glycolysis would be stimulated, as would the reverse pentose phosphate pathway and the production of glycerol phosphate for lipids.

6. Many compounds, such as oxaloacetate, pyruvate, and acetyl-CoA, play a role in a number of reactions. More to the point, the end products of some pathways are the starting points of others. Each pathway is one aspect of an overall metabolic scheme.
7. The *effect* of biochemical pathways can be reversed. Examples include glycolysis and gluconeogenesis, glycogen formation and synthesis, and the pentose phosphate pathway. The details are not completely reversible. An irreversible step in one pathway tends to be replaced with another reaction, catalyzed by another enzyme.
8. Transport processes are especially important for substances, such as oxaloacetate, that cannot cross the mitochondrial membrane. The same is true for electrons. Shuttle mechanisms must exist to transport electrons as the reduced form of important compounds. Compounds that cannot cross the membrane must be converted to ones that can, and then must be converted back to their original form on the other side of the membrane.
9. When a pathway has a number of steps, it is possible for energy changes to take place in steps of manageable size. It also allows for control of a pathway to be exercised at more points than would be the case if there were only a few steps.
10. The possibilities are limitless. Even more to the point, some discovery that no one expects can open even more possibilities.

## 24.2 Biochemistry and Nutrition

11. The old pyramid assumed that all carbohydrates and fats were the same and that carbohydrates were good and all fats were bad. The new pyramid recognizes that not all carbohydrates are good and not all fats are bad. Complex carbohydrates are placed lower down on the new pyramid, whereas processed ones are placed higher. Essential fats and oils are included as necessary food types. Also, dairy consumption recommendations have been reduced.
12. Fats and carbohydrates can be stored when they are consumed in excess. Fats are stored as triacylglycerols and carbohydrates are stored as glycogen. However, proteins consumed in excess are not stored. The extra protein is broken down. The amino groups are released as urea and the carbon skeletons are stored as carbohydrate or fat.
13. Saturated fatty acids have been correlated with increased levels of LDL, which have been shown to be an indicator of high risk for heart disease.
14. Leptin is a hormone that affects metabolism. It affects the brain to suppress appetite and it affects metabolism directly by stimulating fatty-acid oxidation and inhibiting fatty-acid synthesis.
15. Yes, cholecalciferol is made in the body, and many of its functions are hormone-like in nature.
16. Carbohydrates are the main energy source. Excess fat consumption can lead to the formation of “ketone bodies” and to atherosclerosis. Diets extremely high in protein can put a strain on the kidneys.
17. The liver is the primary organ for alcohol metabolism and for disposing of drugs (legal, illegal, and accidental) and halocarbon compounds. When the liver spends its time dealing with these

other tasks, it may not be able to carry out its other normal functions; in essence, prolonged exposure to any such “toxin” overworks the liver.

18. Vitamin A is a lipid-soluble vitamin, which can accumulate in the body. Overdoses of this vitamin can be toxic.
19. Low levels of iodine in the diet often lead to hypothyroidism and an enlarged thyroid gland (goiter). This condition has largely been eliminated by the addition of sodium iodide to commercial table salt.
20. Lucullus breaks down the protein in the tuna to amino acids, which in turn undergo the urea cycle and the breakdown of the carbon skeleton described in Chapter 23, eventually leading to the citric acid cycle and electron transport. In addition to protein catabolism, Griselda breaks down the carbohydrates to sugars, which then undergo glycolysis and enter the citric acid cycle. (Gratuitous information: Lucullus was a notorious Roman gourmand. In medieval literature, Griselda was the name usually given to a forbearing, long-suffering woman.)
21. All amino acids must be present at the same time for protein synthesis to occur. Newly synthesized proteins are necessary for growth in the immature rats.
22. The weight loss is due to correction of the bloating caused by retention of liquids.
23. After a person is fully grown, many amino acids are scavenged and recycled by the body. Because all proteins contain at least some of these two amino acids, there are enough to maintain the body. It should be noted that both again become essential if there is disease or tissue damage and that arginine is required for sperm production in males.
24. The early colonists always cooked in iron pots; enough iron is leached out to supply required amounts, as long as the body is able to absorb it. (Glass cookware did not become available until after World War I, and aluminum cookware was not available until after World War II.)
25. Diets high in fiber are usually lower in fats, especially saturated fats; fiber adsorbs many potentially toxic substances, such as cholesterol and halocarbons, preventing their absorption into the body; fiber decreases transit time through the intestine, so any toxic materials in food remain in the body for less time and have a smaller chance of being absorbed or otherwise causing problems.
26. This claim has a chemical basis. Calcium carbonate dissolves in stomach acid, releasing calcium ion in its usual hydrated form. Calcium citrate is likely to have the calcium ion bound to the citrate in a manner similar to iron in heme. Consequently, the charge of the calcium ion is effectively decreased. Calcium bound to citrate can pass a cell membrane more easily than a hydrated calcium ion.
27. Alcohol provides calories but does not provide vitamins. This is one of the leading causes of malnutrition. Metabolizing alcohol involves an enzyme (alcohol dehydrogenase) with thiamine pyrophosphate (TPP) as a cofactor. The cofactor, in turn, is a metabolite of vitamin B<sub>1</sub>, leading to severe deficiencies.
28. Metal ions play a role in the structure and function of proteins and some coenzymes. They tend to do so because they operate as Lewis acids.
29. Severe depletion of glycogen often results in a rebound effect, in which so much is made that some is stored in inappropriate tissues, including the heart, and mineral imbalances often occur. It is best to exercise moderately before the glycogen loading because then the glycogen is stored more effectively and safely in the liver and muscle tissue where it is most needed.

30. Nutrients and water turnover in the body, sometimes very frequently. This implies that an organism is an open system. Equilibrium requires a closed system. Consequently, an organism can reach a steady state, but never equilibrium.

### 24.3 Hormones and Second Messengers

31. Hormones can have several different kinds of chemical structures, including steroids, polypeptides, and amino acid derivatives.
32. The anterior pituitary stimulates release of trophic hormones, which in turn stimulate specific endocrine glands; the workings of the adrenal cortex, the thyroid, and the gonads can all be affected as a result. The adrenal cortex produces adrenocortical hormones, including glucocorticoids (involved in carbohydrate metabolism, inflammatory reactions, and reaction to stress) and mineralocorticoids, which control the level of excretion of water and salt by the kidney. If the adrenal cortex does not function adequately, one result is Addison's disease, characterized by hypoglycemia, weakness, and increased susceptibility to stress. The opposite condition, hyperadrenocorticism, is Cushing's syndrome.
33. The hypothalamus secretes hormone-releasing factors. Under the influence of these factors, the pituitary secretes trophic hormones, which act on specific endocrine glands. Individual hormones are then released by the specific endocrine glands.
34. Thyroxine is an amino acid derivative and is absorbed directly from the gut into the bloodstream. If insulin were taken orally, it would be hydrolyzed to amino acids in the stomach and intestine.
35. G proteins get their name because they bind GTP as part of their effect. An example is the G protein that is linked to the epinephrine receptor and leads to the production of cAMP as a second messenger. Receptor tyrosine kinases have a different mode of action. When they bind their hormone, they phosphorylate tyrosine residues on themselves and other target proteins, which then act as a second messenger. Insulin is an example of a hormone that binds to a receptor tyrosine kinase.
36. cAMP,  $\text{Ca}^{2+}$ , insulin receptor substrate.
37. Human growth hormone is a peptide hormone. If it were taken orally, the peptide would be degraded to its component amino acids in the small intestine and would be rendered useless.

### 24.4 Hormones and the Control of Metabolism

38. Epinephrine and glucagon are the two that were discussed the most in this book.
39. Glucagon causes the activation of glycogen phosphorylase, inhibition of glycogen synthase, and inhibition of phosphofructokinase-1.
40. Epinephrine has the same effect on glycogen phosphorylase and glycogen synthase, but it has the opposite effect on phosphofructokinase-1.
41. The G protein is bound to GTP. Eventually, the GTP is hydrolyzed to GDP, which causes it to dissociate from adenylate cyclase. This stops the hormone response until the hormone dissociates from the receptor; the G protein trimers are rejoined, and the process starts over again.
42.  $\text{IP}_3$  is a polar compound and can dissolve in the aqueous environment of the cytosol; DAG is nonpolar and interacts with the side chains of the membrane phospholipids.
43. When a stimulatory hormone binds to its receptor on a cell surface, it stimulates the action of adenylate cyclase, mediated by the G protein. The cAMP that is produced elicits the desired

effect on the cell by stimulating a kinase that phosphorylates a target enzyme.

44. See Table 24.2.

45. It is most unlikely that a metabolic pathway could exist without control mechanisms. Many pathways require energy, so it is advantageous for an organism to shut down a pathway when its products are not needed. Even if a pathway does not require large amounts of energy, the many connections among pathways make it likely that control is established over the levels of important metabolites.

46. In cholera, adenylate cyclase is permanently "turned on." This in turn stimulates active transport of  $\text{Na}^+$  and water from epithelial cells, leading to diarrhea.

47. (a) Stoichiometric amounts of cAMP are required to activate cAMP-dependent protein kinase.

(b) Six catalytic steps, including the reaction catalyzed by glycogen phosphorylase, with 10 molecules acted on in each step, would result in  $10^6$  (one million) G-1-P molecules for each epinephrine.

(c) A major factor is speed. It is important to be able to use stored energy rapidly in "fight or flight" situations. A second factor is control. Note that glycogen phosphorylase is activated by kinases. The competing process of glycogen storage, catalyzed by glycogen synthetase, is inactivated by kinases. A third factor is economy. A single molecule of epinephrine activates many molecules of glycogen phosphorylase and yet more molecules of G-1-P.

48. A phosphatase dephosphorylates glycogen phosphorylase and glycogen synthetase, inactivating and activating them, respectively. The phosphatase becomes active in response to high concentrations of glucose.

49. Low-carbohydrate diets are designed to prevent the high blood sugar levels that arise when large quantities of carbohydrates are consumed. High blood sugar leads to a rapid rise in insulin. Insulin is known to stimulate fat synthesis and to inhibit fatty-acid oxidation. Thus, low-carbohydrate diets are thought to help fight weight gain.

### 24.5 Insulin and Its Effects

50. Insulin's primary function is to stimulate the transport of glucose out of the blood and into the cell.
51. The second messenger is a protein called the insulin receptor substrate, which is phosphorylated on a tyrosine by the insulin receptor kinase.

52. When insulin binds to its receptor, the  $\beta$ -subunit of the receptor kinase autophosphorylates. When this happens, the receptor kinase is able to phosphorylate tyrosines on the insulin receptor kinase.

53. Insulin causes the following effects:

- (a) Glycogen breakdown is decreased.
- (b) Glycogen synthesis is increased.
- (c) Glycolysis is increased.
- (d) Fatty-acid synthesis is increased.
- (e) Fatty-acid storage is increased.

54. Insulin and epinephrine normally have opposite effects, but they both stimulate muscle glycolysis. Epinephrine is the hormone that signals the need for quick energy, which means the muscle cells must be able to use glucose via glycolysis. Insulin stimulates pathways that use up glucose so that the blood glucose lowers, so it makes sense for it to stimulate glycolysis



- as well. Epinephrine stimulates muscle glycolysis by activating adenylate cyclase, which makes cAMP; cAMP then activates protein kinase A, which phosphorylates phosphofructokinase-2 and fructose-*bis*phosphatase-2. In the muscle, phosphorylation of phosphofructokinase-2 activates it, producing more fructose-2,6-*bis*phosphate, which activates phosphofructokinase-1 and glycolysis. In muscle, insulin stimulates glycolysis by activating phosphofructokinase and pyruvate dehydrogenase.
55. Prerace diet can be critical to a runner. If the race is at 9 AM, and the runner gets up at 7 AM and then eats a typical American breakfast of cereal, toast, or pancakes, she will have a high blood-sugar level within half an hour, which will lead to a high insulin level shortly thereafter. In that scenario, by the time the runner gets to the starting line, she will have a metabolism dedicated to fat and glycogen synthesis and will not be burning fat or carbohydrates. The runner will be like a car with a full tank of gas and a clogged fuel line.
  56. It has been shown that the GLUT4 transporter responds to physical activity. When a person is active, the transporter is active and responds well to insulin. After a few days of detraining, this transporter shows only half of the activity it did before.
  57. GLUT4 is one of the glucose transporters on muscle cells. It responds to insulin by moving glucose out of the blood and into the cell. In type II diabetes, insulin is present, but it does not have the same effect. It takes more insulin to accomplish the same movement out of the blood and into the cell. People with type II diabetes often show classical signs of obesity, and there is a correlation between diminishing GLUT4 activity, obesity, and diabetes.
  58. It had been known for some time that calorie restriction could promote longevity in many species. Scientists discovered a family of proteins called sirtuins that seemed to be at the center of this phenomenon. Sirt1 is a mammalian version of a sirtuin studied very extensively in yeast. This protein is a histone deacetylase, but also seems involved in many other processes. It reacts to scarcity in a way that primes an organism for survival. Transgenic organisms with multiple copies of the gene for Sirt1 show greater longevity and any other stimulus that increases the expression of the gene does too.
  59. The direct target of rapamycin is a protein called mammalian target of rapamycin (mTOR). Both calorie restriction and rapamycin lower the activity of the mTOR enzyme.
  60. Sirt1 regulates several key cellular proteins, such as p53, NF- $\kappa$ B, HSF-1, FOXO1, 3, and 4, and PGC-1 $\alpha$ .

# Index

Page numbers in **boldface** indicate a major discussion of the entry, page numbers followed by an *f* indicate a figure, page numbers followed by a *t* indicate a table, page numbers followed by a *ph* indicate a photograph, HT preceding a page number refers to the Hot Topics section of the book. Positional and configurational designations in chemical names (e.g. 3-, *N*,  $\alpha$ -) are ignored in alphabetization.

## A

- ABC excinuclease, 268
- abiotic synthesis of nucleotides, 8*f*
- abiotical reaction, 6
- abortive transcription, 285
- absolute specificity, 184
- abzymes, 186*f*
- accessory pigments, 629*f*, 630
- acetaldehyde, 435–436, 501
- acetate, 178
- acetic acid, 46*t*, 47–48
- acetoacetate, 601
- acetone, 601
- acetyl groups, 603*f*
- acetylation, 248
- N*-acetyl-D-glucosamine, 464*f*, 472*f*
- acetylcholinesterase, 154*t*
- acetyl-CoA
  - carboxylase, 612
  - carboxylase complex, 603
  - carboxylase reaction, 604*f*
  - citric acid cycle, 534*f*, 535–540, 548*f*, 549
  - glyoxylate cycle, 550
  - hydrolysis of, 445
  - ketone bodies, 601
  - palmitate synthesis, 606*f*
  - production, 709*t*
- N*-acetylmuramic acid, 464*f*, 472
- achiral, 61
- acidic domains, 311
- acidic N-termini, 350
- acid(s)
  - biochemical buffers, form of, 56*t*
  - buffering, 54
  - chymotrypsin, reactions catalyzed by, 146
  - defined, 44–49
  - dissociation constant, 44, 46*t*
  - lactic, 57
  - strength, 44
  - titration curve, 47–49, 52–53, 54*f*, 67, 70
- aconitase, 541, 543
- acquired immunity, 406
- acquired immunodeficiency syndrome (AIDS), 160, 395, 402
- acromegaly, 699
- activating transcription factor 1 (ATF-1), 309
- activation, metabolic, 444
- activation, molecular, 592
- activation energy, 139–141
- activation energy profile, 140
- activation step, 444
- activators, 302
- active cysteine residue, 496*f*
- active site, 144, 177, 179–182, 184–187
- active transport, 209–210
- active tumor targeting, 422
- acyl carrier protein (ACP), 605
- acyl-CoA synthetase, 592, 593*f*, 594
- acyl-enzyme intermediate, 178
- acylglycerols synthesis, 609–613
- N*-acylsphingosine, 613
- adaptive system, 424*f*
- addiction, 689
- Addison's disease, 698
- adenine, 187, 228, 676*f*
- adenosine deaminase (ADA), 403
- adenosine diphosphate (ADP), 438*f*, 439–443, 549
- adenosine monophosphate (AMP), 673–674
- adenosine triphosphate (ATP). *See also*
  - electron transport
  - chloroplasts, production in, 637–639
  - citric acid cycle, 544–547
  - glucose oxidation, yield from, 586
  - high-energy bonds, 438–439
  - hydrolysis of, 440–443
  - metabolism, production in, 446
  - mitochondrial matrix and, 579
  - phosphate ester of glucose, formation of, 461*f*
  - production, mitochondrial structure in, 563–564
  - ratio, metabolic state of a cell and, 549
  - reactions in formation of, 5*t*
- adenoviruses, 397*t*, 403
- adenylate cyclase, 699, 700*f*–701*f*
- adipose tissue, 593*f*
- adrenalin, 68, 704
- adrenocortical hormones, 698
- adrenocorticotrophic hormone (ACTH), 697*t*
- adult-onset diabetes, 709
- Aequorea victoria*, HT3
- aerobic metabolism, 514, 573
- aerobic oxidation, 482*f*
- affinity chromatography, 121, 122*f*, 375*f*
- affinity resins, 122*t*
- African origins, HT7–HT8
- African swine fever virus, 397*t*
- agarose, 120
- agarose gel, 126
- aging, 710–711
- agriculture, 369, 640–641
- AIDS, 160, 395, 402
- alanine, 62*f*, 65*t*, 70
- alcohol, 4*t*, 146
- alcohol consumption, 689
- alcohol dehydrogenase, 502
- alcoholic fermentation, 501
- aldehyde, 4*t*, 459
- alditols, 460
- aldolase, 487*f*, 490, 645
- aldose, 451, 453, 455*f*
- aldosterone, 620
- aldotetrose, 454*f*
- alfalfa roots, 655
- aliphatic, 62
- alkenes, 4*t*
- alleles, 382
- allergies, 404
- allied health
  - amino acids and peptides, 77–78
  - carbohydrate metabolism, 528
  - carbohydrates, 474, 476
  - electron transport and oxidative phosphorylation, 585
  - glycolysis, 488, 502–503
  - lipid metabolism, 622
  - lipids and proteins, 200
  - metabolism, 689, 709–711
  - nucleic acids, 274–275, 373
  - proteins, 185, 189
  - transcription of the genetic code, 303

- allopurinol, 676
- allosteric
  - control, 175, 512*f*
  - effector, 169
  - effects, 520
  - enzymes, 165, 166*f*, 167–170, 171*f*, 172–175, 490
  - proteins, 100
  - regulation, 657*f*
  - sites, 173
  - transitions, 170*f*
- allosterism, 173
- all-*trans*-retinal, 217
- alternative RNA splicing, 316–317
- alternative  $\sigma$  factors, 287–288
- altitude, 56, 351
- Alzheimer's disease, 112–113
- amide, 4*t*, 72
- amine, 4*t*, 146
- L-amino acid, 61–62
- D-amino acid, 61–62
- amino acids, **61–81**. *See also* peptides; proteins
  - acidic, 63*f*
  - acids and bases, as both, 67–72
  - activation, 323, 329
  - $\beta$ -alanine, 74*f*
  - basic, 63*f*–64*f*
  - biosynthesis, 557*f*, 656–666
  - branched chain, 75
  - breakdown, 666–667
  - catabolism, 666–672
  - common, values of, 71*t*
  - derivatives, 697*t*
  - directionality in, 10*f*
  - DNA interactions, 310*f*
  - DNA-binding proteins with zippers, sequences of, 312*f*
  - essential, 75, 666
  - human requirements, 666*t*
  - hydrophilic substances, 38*t*
  - hydroxylysine, 66*f*
  - hydroxyproline, 66*f*
  - ionization of, 69*f*
  - lysine, 66*f*
  - metabolism, 556, 659*f*
  - names and abbreviations, 65*t*
  - nitrogen metabolism, 653
  - non-polar (hydrophobic), 63*f*–64*f*
  - peptide bond, 72–74
  - polar, 63*f*–64*f*
  - proline, 66*f*
  - protein, building blocks of, 63*f*–64*f*
  - residues, 72, 178, 180–181, 473*f*
  - selenocysteine, 339
  - separation, 128*f*
  - sequences, 310*t*
  - side chains, 61–62, 65–66, 94
  - small peptides, 74–79
  - structures and properties, 62–67
  - three-dimensional structure, 61–62
  - thyroxine, 66*f*
  - titration curve, 67, 70–72
  - transamination reaction, 661
  - tryptophan, 68–69
  - tyrosine, 66*f*, 68–69
  - uncommon, 66–67
- amino group, 61
- amino sugars, 463
- A (aminoacyl) site, 334
- aminoacyl-tRNA synthetase reaction, 330*f*
- aminoacyl-tRNA synthetases, 323
- $\delta$ -aminolevulinic acid, 557*f*
- ammonia, 41*t*, 653, 667
- ammonium ion, 46*t*
- amphibolic, 533
- amphipathic, 37
- amphipathic compounds, 193
- amphipathic molecule, 39*f*
- amphiphilic molecule, 38*f*
- amprenavir (VX-478), 160
- $\beta$ -amylase, 470
- amylopectin, 471*f*
- amylose, 470*f*–471*f*
- anabolic processes, 688*f*
- anabolic steroids, HT28
- anabolism, 432, 552–557
- anabolism of inosine monophosphate, 672
- anabolism of pyrimidine nucleotides, 678
- anaerobic alcoholic fermentation, 482*f*
- anaerobic glycolysis, 481, 482*f*
- anaerobic metabolism, 499–503
- anaerobically, 514
- analogue unit (ADX), HT13
- analytical chemistry, 375
- analytical ultracentrifugation, 245, 246*f*
- anaplerotic, 556
- anaplerotic reaction, 552
- ancient DNA, 232
- androgens, 697*t*, 699
- aneuploid, HT23
- aneuploidy, 270
- angiogenesis, 351
- Animalia, 22
- animals, 556, 640–641, 665*f*
- anion exchanger, 123
- $\alpha$ -anomer, 469*f*
- anomeric carbon, 454
- anomers, 454
- Anopheles gambiae*, 367
- antenna molecules, 630*f*
- anterograde signaling, 646
- antibiotics, 671
- antibodies, 410–416, 475
- anticodon, 326
- antigen, 406, 408
- antigen–antibody reaction, 411*f*
- antigenic determinants, 475
- antigen-presenting cells (APCs), 405, 409
- anti-malarial plants, 638
- antioxidant, 218
- antisense strand, 282
- antitermination, 301
- 2·3 antiterminator, 294–295
- anylopectin, 470*f*
- AP endonuclease, 267
- AP site, 267
- AP-1 (activating protein 1), 305, 418
- apoptosis, 177
- applied genetics, 638
- aptamer, 295
- arachidic acid, 195*t*
- arachidonic acid, 195*t*, 220
- archaeobacteria, 22–23
- arenaviruses, 397*t*
- arginine, 65*t*, 670
- argininosuccinate, 670
- Armstrong, Lance, 585, HT30–HT31
- arrays, 389*f*
- arrest release factors, 301
- arteriviruses, 397*t*
- ascorbic acid, 461
- asparagine, 65*t*
- aspartate concentration, [S], 147*f*
- aspartate transcarbamoylase (ATCase), 146–147, 168*f*
- aspartic acid, 65*t*
- aspirin poisoning, 56
- astroviruses, 397*t*
- asymmetric binding site, 184
- atherosclerosis, 198, 622
- athletes, 514, HT28–HT31
- atmosphere, 654
- atomic force microscopy (AFM), HT17–HT18
- ATP (adenosine triphosphate). *See* adenosine triphosphate (ATP)
- ATP synthase, 574–575, 579
- attenuation mechanism, 296*f*
- 5'-AUG-3', 332
- autocatalytic, HT3
- autoimmune disease, 404, 414
- autoradiograph, 356, 357*f*
- autoradiography, 356
- autoregulation, 294
- autotrophs, 640
- Azotobacter vinelandii* Fe–protein dimer, 656

## B

- B cells, 405–406, 407*f*, 410*f*
- backbone hydrogen bonding, 94
- bacteria, 14*f*, 23, 368, 556, 611*f*, 665*f*
- bacterial cell walls, 472, 473*f*
- bacterial DNA gyrase (topoisomerase II), 237
- bacterial plasmid, 363*f*
- bacterial promoter, 290*f*
- bacteriochlorophyll, 627–630
- bacteriology, 295
- bacteriophages, 357
- bacteriorhodopsin, 578*f*
- bacterium, 16*f*
- $\beta$ -barrel arrangements, 92*f*
- basal level, 302
- base modification, 312

base pairing, 234, 235*f*, 236, 328  
 base stacking, 236  
 base-excision repair, 267, 268*f*  
 bases, 44–49, 54, 56*t*, 227–228  
 basic side chain, 66  
 basic-region leucine zipper (bZIP), 309, 311, 312*f*  
 benzoic acid, 46*t*  
 bidentate, 219  
 bidirectional replication, 257*f*  
 big-bang cosmology, 5–6  
 bile acids, 618  
 binding activators, 171*f*  
 binding sites, 291  
 biochemical buffer, 56*t*  
 Biochemical Connections. *See also* allied health; genetic(s); nutrition  
   acids and sports, 57  
   agriculture, 640–641  
   analytical chemistry, 375  
   applied genetics, 638  
   bacteriology, 295  
   biophysical chemistry, 346  
   biotechnology, 23, 208, 422  
   cancer research, 523  
   cell biology, 677  
   chemistry, 43, 55–56  
   chemistry of the blood, 56  
   dentistry, 502  
   endocrinology, 309  
   environmental, 190  
   environmental science, 484  
   epidemiology, 400–401, 413  
   evolution, 556  
   evolutionary biology, 277, 319  
   exercise physiology, 514  
   forensics, 379  
   gene expression, 602  
   genomics, 249  
   health, 457  
   health sciences, 142  
   immunology, 424  
   instrumentation, 134  
   law, 239  
   medical genetics, 572  
   medicine, 111–113, 160, 173, 308, 403, 671  
   microbiology, 271  
   molecular biology, 385  
   molecular genetics, 339  
   neurology, 344  
   neurophysiology, 68–69  
   neuroscience, 79, 147, 154, 217  
   oncology, 424  
   organic chemistry, 75  
   physical organic chemistry, 157  
   physics, 630  
   physiology, 213, 351, 441, 668  
   plant science, 369, 466, 655  
   proteomics, 135  
   thermodynamics, 29, 431, 433  
   toxicology, 543  
   virology, 415

biochemical energetics, 26–27  
 biochemical reactions, 28, **35–60**  
 biochemistry  
   cells, and the organization of, **1–33**  
   chemical foundations of, 3–4  
   DNA testing, forensic uses of, 379  
   energy, 26–27  
   and nutrition, 687–695  
   robotic technology meets, 388  
 bioenergetics, 429  
 biofuels, 484  
 bioinformatics, 106  
 biological membranes, **193–225**  
 biological systems, 27*f*  
 biology, origin of life, 4–13  
 biomolecules, 1–3, 6, 7*f*, 434*f*  
 biophysical chemistry, 346  
 biosphere, 654*f*  
 biosynthesis  
   amino acids, 658  
   cholesterol, 615  
   cysteine, 665*f*  
   glutamate and glutamine, 657*f*  
   of nucleic acids, **253–279**  
   of phosphatidylethanolamine in bacteria, 611*f*  
   pyrimidine nucleotide, 680*f*  
   of RNA, **281–322**  
   serine, 662  
 biotechnology, 23, 208, 422  
 biotechnology techniques, **355–393**  
 biotin, 187*t*, 516  
 biotin carboxylase, 603  
 biotin carrier protein, 603  
 birnaviruses, 397*t*  
 blood, 56, 476, 696  
 blood doping, HT29–HT30  
 blood transfusions, 476, HT29–HT30  
 bloodgroup antigenic determinants, 476  
 blue/white screening, 366  
 Bohr effect, 102, 103*f*, 104*t*  
 Boltzmann, Ludwig, 29*ph*  
 bond energies, 41*t*  
 Bonds, Barry, HT28, HT29*ph*  
 brain, HT12  
 branch points, HT13  
 branch site, 315  
 branched-chain amino acids, 75  
 branched-chain polymer, 463*f*  
 branched-chain structure of glycogen, 507*f*  
 branching enzyme, 510, 512*f*  
 brown adipose tissue (BAT), 580  
 bubble boy syndrome, 403  
 buffer, 55–56  
 buffer solution, 49–53  
 buffering capacity, 54  
 buffers, 49–57  
 $\beta$ -bulge, 87, 89*f*  
 bunyaviruses, 397*t*  
 butter, 204  
 bZIP transcription factor, 313

## C

C domain, 411  
 C<sub>4</sub> pathway, 647*f*, 648  
*Caenorhabditis elegans*, 24*f*, 200, 240, 307, 387, HT3  
 calcium, 6*t*  
 calcium ion, 700  
 calciviruses, 397*t*  
 California encephalitis virus, 397*t*  
 Calvin cycle, 642, 643*f*, 644–645  
 cAMP-dependent protein kinase (protein kinase A), 304, 701*f*, 704–705, 706*f*  
 cancer, **395–425**  
   cause of, 417  
   cells, 200, 416  
   cervical, HT20–HT21  
   chemotherapy, 671  
   curing, 421–422  
   defined, 416  
   fighting, 421  
   human genome, 417  
   leukemia, HT4  
   MDR1, 348  
   research, 523, HT15–HT16  
   symptoms, 424  
   telomerase and, 274–275  
   treatment, 424  
   viruses and, 421  
 canine parvovirus, 397*t*  
 5'-cap, 247, 342  
 capping, 314  
 capsid, 395, 396*f*  
 carbamoyl phosphate, 667  
 carbamoyl-phosphate synthetase I (CPS-I), 671  
 carbanion, 503*f*  
 carbohydrate metabolism, **507–531**, 543, 704  
 carbohydrates, **451–479**, 557*f*  
 carbon, 6*t*, 35*t*, 190  
 carbon atoms, 434*f*, 453, 455*f*, 598  
 carbon dioxide, 541–542  
 carbon monoxide, 99*f*  
 carbon skeleton, 666–667  
 carbonic acid, 46*t*  
 carbonic anhydrase, 154*t*  
 carboxyl group, 61, 66  
 carboxyl transferase, 603  
 carboxylic acids, 4*t*  
 carnitine, 592  
 carnitine acyltransferase, 592  
 carnitine palmitoyltransferase (CPT-I), 592  
 carnitine palmitoyltransferase (CPT-II), 592  
 carnitine translocase, 592, 594*f*  
 carnosine, 74*f*  
 $\beta$ -carotene, 214, 215*f*  
 cascade, 704  
 caspases, 177  
 catabolic processes, 688*f*  
 catabolism  
   citric acid cycle in, 534*f*, 551–552

- defined, 432
- of lipids, 591–596
- of odd-carbon fatty acids, 598–600
- purine, 675
- pyrimidine, 678–680
- of unsaturated fatty acids, 598–600
- uric acid to CO<sub>2</sub>, 668
- catabolite activator protein (CAP), 291, 292*f*
- catabolite repression, 291, 292*f*
- catalase, 19, 154*t*
- catalysis, 9, 139, 141, 179, 209
- catalytic activity, 9
- catalytic antibodies, 189
- catalytic trimers, 168*f*
- catechol-*O*-methyltransferase (COMT), 349
- cation exchanger, 123, 124*f*
- cattle, 470*ph*
- CD4 cells, 409
- CD8 cells, 408
- cDNA library, 376–377
- cell membrane
  - animal, 17*f*
  - double, 19*f*
  - organelles and their functions, 21*t*
  - in the origin of life, 12*f*
  - plant, 17*f*
  - prokaryotes and eukaryotes, comparison of, 15*t*
  - prokaryotic cells, 16
- cellobiose, 468
- cell(s)
  - animal, 17*f*, 608*f*
  - biology, 677
  - cancer, 416
  - cholesterol in the, 621
  - cloning, 361–362
  - common ground, 24–25
  - components, 19, 119*f*
  - differentiation, 309
  - division, 270–271
  - DP cell, 414
  - endothelial, HT21
  - epithelial, HT20
  - feeder, HT23
  - genetic information, 253
  - homogenate, 119*f*
  - induced pluripotent stem cells (iPS cells), HT25
  - information transfer, 243*f*
  - mammalian, 20*f*
  - membrane, 16
  - metabolic state of a, 549*t*
  - muscle, HT5
  - nerve, 309
  - origin of, 2–3
  - plant, 17*f*
  - progenitor, HT23
  - prokaryotic, 17*f*
  - proliferation, 309
  - proteins, extracting pure, 117–118
  - signaling, 441
  - stem, HT23–HT27
  - tobacco leaf, 18*f*
  - viral attachment, 399
  - walls, 16, 17*f*, 21*t*, 472, 473*f*
- cellular signaling, **687–714**
- cellulases, 468
- cellulose, 468, 469*f*, 474
- central vacuole, 21
- ceramides, 198, 200, 613
- cerebroside, 198, 202*f*, 613
- cervical cancer, HT20–HT21
- chain elongation, 284–285, 286*f*, 323, 334–335, 343–345
- chain initiation, 284–285, 323, 331–332, 342
- chain termination, 287, 323, 336, 345
- chair confirmation, 458*f*
- champagne, 154
- chaperones, 346
- chaperonins, 346
- chemical reactions, 182
- chemiosmotic coupling, 563, 577
- chemistry
  - analytical, 375
  - Biochemical Connections, 43, 55
  - biophysical, 346
  - of the blood, 56
  - green, 190
  - hydrogen bonding, 43
  - organic, 3–4, 75, 157
  - physical organic, 157
- chemotherapy, 671
- chimeric DNA, 360
- Chinese restaurant syndrome, 75
- chiral, 61
- chitin, 471–472
- chlorophyll, 627–632
- chloroplast genes, 646–647
- chloroplasts, 15–16, 17*f*–19*f*, 21*t*, 627–631
- cholera, 700
- cholesterol, 38*t*, 198, 202*f*, 203, 614, 617*f*, 618, 619*f*, 620
- cholesterol biosynthesis, 613–623
- chromatin, 17, 238, 248
- chromatin granules, 18*f*
- chromatography, 118–125, 375
- chromatophores, 15
- chromophore, HT3
- chromosome pairing, 269
- chromosomes, 17
- chymotrypsin, 129, 130*f*, 146–147, 154*t*, 176–182, 185
- chymotrypsinogen, 176
- cinnamaldehyde, 466
- circadian rhythms, 309
- circoviruses, 397*t*
- citrate, 541
- citrate formation, 540
- citrate synthase, 536*f*, 538, 540
- citric acid, 46*t*
- citric acid cycle, 482*f*, **533–561**, 659*f*, 670, 687
- citrulline, 667
- classification, 21–24
- cleavage, 482, 594
- cleavage sites, 359*t*
- clonal selection, 408
- clone, 361–362
- cloning, 360–362, 363*f*, 365*f*, 366–367
- cloning vectors, 370–371
- closed complex, 284
- closed system, 433
- closed vesicles, 578
- cloverleaf structure, 245
- CO<sub>2</sub>, 668
- CO<sub>2</sub> fixation, 646–649
- coactivator, 305
- coat proteins (CP), 402
- cocaine, 189
- coding strand, 282
- codon, 324, 326
- coenzyme A, 187*t*, 443–446, 607*f*
- coenzyme Q (CoQ), 567–571
- coenzymes, 187, 188*f*, 434–437, 502
- co-inducer, 290–292
- collagen triple helix, 90, 92
- colony hybridization, 376*f*
- colors, 466
- column chromatography, 118–125
- commaless code, 324
- comparative modeling, 106
- competitive inhibition, 155–156, 168–169
- complementary base pairing, 233
- complementary proteins, 85
- complementary strands, 233
- complement-control protein module, 91*f*
- complete covalent structure, 94
- complete protein, 85
- complexed side chains, 94
- compound lipids, 609–613
- COMT gene variants, 349
- concentration gradient, 209
- concentrations, 51
- concerted model, 169–170, 171*f*, 173
- configuration, 452
- confocal fluorescence microscopy, HT18
- conformational coupling, 579
- coniferyl alcohol, 475*f*
- connective tissue, 475
- consensus sequence –<sub>3</sub>ACCAUGG+<sub>4</sub>, 343
- consensus sequences, 284, 298
- constant region, 411
- constitutive, 290–292
- control
  - carbohydrate metabolism, **507–531**
  - of the citric acid cycle, 547–549
  - mechanisms, 293*f*
  - points, 498, 548*f*, 549
  - sites, 291
- cooperative binding, 101
- copy number variants, HT7
- CoQH2-cytochrome *c* oxidoreductase, 569
- core enzyme, 282
- core promoter, 284
- co-repressor, 290–292
- Cori cycle, 521, 522*f*



coronaviruses, 397*t*  
 correct folding, 109–110  
 corticotropin-releasing factor (CRF), 697*t*  
 cortisone, 620  
 coupled transcription, 345  
 coupled translation, 341  
 coupling, 442, 574, 577–580  
 covalent bonds, 41*t*  
 covalent modification, 175, 512*f*, 520*t*  
 CP-31398, 421  
 CREB-binding protein (CBP), 304–306  
 Creutzfeld-Jakob disease, 112  
 cristae, 17, 18*f*  
 cross-linked polyacrylamide, 121*f*  
 crown gall, 372  
 crude extract, 117*t*  
 C-terminal amino acid residue, 74  
 C-terminal domain (CTD), 297  
 CTF-1 activator, 312  
 Cushing's syndrome, 698  
 cut-and-patch process, 264  
 cyanidin chloride, 466  
 cyanobacteria, 14  
 cyanogen bromide (CNBr), 129  
*Cyanophora paradoxa*, 24  
 Cyclic AMP (adenosine-3',5'-monophosphate, cAMP), 699  
 cyclic electron, 635  
 cyclic hemiacetal, 456*f*  
 cyclic molecule, 454  
 cyclic photophosphorylation, 635  
 cyclic structure, 86  
 cyclic-AMP-response element (CRE), 304  
 cyclic-AMP-response-element binding protein (CREB), 304–306, 309  
 cyclin-dependent protein kinases (CDKs), 272, 419–420  
 cyclins, 272  
 Cys<sub>2</sub>His<sub>2</sub> zinc-finger, 311  
 cysteine, 65*t*, 665*f*  
 cysteine protease, 185  
 cystic fibrosis (CF), 384*f*  
 cytidine triphosphate (CTP), 610–611, 680*f*  
 cytochrome *c* oxidase, 570–571  
 cytochromes, 568*f*, 569–571, 574*f*  
 cytokines, 405  
 cytomegalovirus, 397*t*  
 cytoplasm, 15  
 cytoplasmic alarm system, 677  
 cytoplasmic defenses, 677  
 cytosine, 227, 267  
 cytoskeleton, 20, 21*t*  
 cytosol, 15, 20, 21*t*, 554*f*–555*f*, 603*f*, 669*f*  
 cytotoxic T cells (killer T cells), 405–406, 407*f*, 408

## D

dairy, 467  
 dalton, 111, 245  
*Danio rerio*, HT4  
 data points, 96*f*

de novo prediction, 106  
 debranching enzymes, 470, 508, 509*f*  
 decarboxylation, 538*f*  
 degenerate code, 324  
 dehydration, 484  
*Deinococcus radiodurans*, HT4  
 denaturation, 99–100, 101*f*, 241–242  
 dendritic cells, 405, 406*f*  
 dengue virus, 397*t*  
 denitrification, 653  
 density-gradient centrifugation, 254  
 dentistry, 502  
 deoxy sugars, 459–460  
 deoxyhemoglobin, 103*f*, 105*f*  
 deoxyribonucleoside, 228, 229*f*–230*f*  
 deoxyribonucleotide, 681–682  
 detection, 355  
 detection methods, 356–358  
 detergents, 100  
 dextran, 120  
 diabetes, 488, 708–711, HT10–HT12  
 diacylglycerol (DAG), 610, 700), HT11  
 diarylsulfonamide activator, 523  
 diastereomers, 453  
 dicer, 385  
 dideoxy nucleotides, 384  
 dietary fiber, 474  
 diets, low-carbohydrate, 457, 708  
 differential centrifugation, 118, 119*f*  
 diffraction pattern, 95  
 digitalis, 466  
 dihydrofolate reductase, 682  
 dihydrolipoyl dehydrogenase, 537–539  
 dihydrolipoyl transacetylase, 537–539  
 dihydroxyacetone, 451–452  
 dihydroxyacetone phosphate, 490–491, 499*f*  
 diisopropylphosphofluoridate (DIPF), 178  
 dimer, 100  
 dimethylallyl pyrophosphate, 615  
 dinitrogenase reductase, 656  
 diphenylphosphatidyl glycerol (cardiolipin), 196  
 dipole-dipole interaction, 36, 37*f*  
 dipole-induced dipole interaction, 37  
 dipoles, 35–36  
 direct conversion, HT26  
 direct programming, HT25  
 disaccharides, 462*f*, 464, 465*f*, 467–468, 469*f*  
 disease. *See also* cancer; viruses  
   alcoholism, 689  
   Alzheimer's disease, 111–113, HT11–HT12, HT23  
   asthma, HT11  
   autoimmune, 404, 414, 416, 708  
   CREB-binding protein (CBP), 268  
   Creutzfeld-Jakob disease, 111–112  
   diabetes, 488, 708, HT10–HT12  
   enzymes as markers for, 142  
   genetic, 268, 303  
   genetic engineering, 367–369  
   high blood sugar, HT11  
   Huntington's, 111, HT12  
   hypertension, HT11

immunology, 424  
 leukemia, HT4  
 mad cow, 111–112  
 micro RNA networks, 308–309, 349  
 muscular dystrophy, 218  
 neurodegenerative, 112–113  
 neurotransmitters and, 68–69  
 Parkinson's, HT12  
 prion, 111–112  
 protein folding, 109, 111–112  
 resistance, 369  
 schizophrenia, 69  
 sexually transmitted disease (STD), HT20–HT22  
 sickle-cell anemia, 84  
 strokes, HT11  
 vertebrate viruses and, 397*t*  
 viral, 397*t*  
 disulfide bonds, 94  
 A-DNA, 234, 235*f*  
 B-DNA, 234, 235*f*  
 Z-DNA, 235–236  
 DNA (deoxyribonucleic acid)  
   amino acid interactions, 310*f*  
   ancient, 232  
   antisense strand, 282  
   automated sequencing, 387*f*  
   cells, organization of, 13  
   chain, 231*f*, 257*f*  
   circular, 261  
   cloning, 362, 363*f*  
   coding strand, 282  
   denaturation of, 241–242  
   detection methods, 356–358  
   family trees, HT6–HT9  
   fingerprinting, 379, 381  
   glycosylase, 267  
   gyrase, 237, 261  
   human mt DNA map, 572  
   hydrolysis, 360*f*  
   *lac* repressor and, 292*f*  
   library, 374–377  
   ligase, 258, 360  
   loops, 302–303  
   methylation, 358*f*  
   microarray, 388  
   mitochondrial, HT6–HT7  
   nanostructures, HT15  
   nanotechnology, HT13–HT16  
   nucleic acids, 227–230, 253  
   origin of life, 9–11  
   paternity test, 379, 383  
   plasmids, 362  
   Poll II promoters, 298–299  
   polymerase, 256, 258–260  
   polymerase proofreading, 264  
   polymerase repair, 265*f*  
   polymerase III holoenzyme, 260  
   prokaryotic cell, 17*f*  
   recombinant DNA, 360, 363*f*  
   recombination, 268–270, 368  
   replication, 254–256, 258–263

- retroviruses, 401–403
- scanning tunneling microscopy (STM), HT17–HT19
- semidiscontinuous DNA replication, 257*f*–258*f*
- sense strand, 282
- sequencing of, 384
- single-stranded, 262, HT18–HT19
- (–) strand, 282
- (+) strand, 282
- strands, 254, 258, 263, 282
- structure and function, 258
- structure of, 232–238
- synthesizing, 254, 263
- template strand, 282
- thymine and, 267
- viruses and, 395, 401
- DnaB protein, 262
- DNA-binding domain, 309
- DNA-binding proteins, 309–310, 312*f*
- DNA-dependent RNA polymerase, 281*t*
- dolphins, 488
- domains, 83, 88
- dopamine, 68, 189
- doping, HT28–HT31
- cis* double bond, 193
- double bonds, 202*f*
- double cell membrane, 19*f*
- double crossover (DX) units, HT13
- double membrane, 18*f*
- double-helical structure, 232–234
- double-helix, 242*f*
- double-origin theory, 11
- downstream, 287
- DP cell, 414
- Drosophila melanogaster*, HT4
- drug delivery, 208
- drug targets, 421–422
- dTTP, 682
- duDP, 682
- dwarfism, 699
- E**
- E (exit) site, 334
- E. coli*. *See* *Escherichia coli*
- earth, 6–7
- Eastern equine encephalomyelitis (EEE) virus, 397*t*
- ebola virus, 395, 397*t*
- EcoRI, 358
- Edman degradation, 129–133
- effector, 169
- EF-Tu, 334
- Eicosapentenoic acid (EPA), 222
- elastin, 185
- electron
  - micrograph, 208, 472*f*, 575*f*
  - microscope, HT17
  - oxidation and reduction, 26
  - transfer, 429–449, 639
  - transfer reactions, 665*f*
  - transport, 533, 534*f*, **563–589**
  - transport chain, 563–564, 568*f*
  - transport complexes, 566–574
  - transport reactions, 573*t*
- electronegativity, 35–36
- electrophile, 180–181
- electrophoresis, 71, 126–130
- electroporation, 362
- electrospray Ionization (ESI-MS), 134
- electrostatic attraction, 94
- electrostatic repulsion, 439*f*
- elongation, 301
- eluent, 118
- embryonal carcinoma (EC) cells, HT23
- embryonic stem (ES) cells, HT23
- enantiomers, 452
- endergonic, 28
- endergonic reactions, 431
- endocrine glands, 697
- endocrine system, 695–696
- endocrinology, 309
- endocytosis, 212
- endoglycosidase, 470
- endonucleases, 357
- endoplasmic reticulum (ER), 15, 17*f*, 18, 21*t*, 347
- $\beta$ -endorphin, 697*t*
- endosymbiosis, 24–25
- endothelial cells, HT21
- energetically favorable, 27
- energetics of conversion of pyruvate to CO<sub>2</sub>, 547–549
- energy. *See also* free-energy
  - AMP and GMP production, 674
  - balance, 431
  - changes, 27–28, **429–449**
  - generation, 591
  - life processes, 26
  - of light, 630
  - plants and, 640–641
  - production, 503–504
  - storage, 591
  - use, 438–443
  - yield, 596–598
- enhancers, 289, 298, 302
- enolase, 493*f*, 497
- enterokinase (EK), 373, 375
- enthalpy, 29
- entropy, 29–30, 440*f*
- env* gene, 403
- envelope membrane, 395, 396*f*
- envelope proteins (EP), 402
- environmental science, 484
- environmental toxicology, 190
- enzymatic activity, 47*f*, 142, 175
- enzyme-catalyzed reactions, 142, 146–147, 595*f*
- enzyme(s)
  - allosteric, 165, 166*f*, 167–170, 171*f*, 172–175, 490
  - allosteric, behavior of, 165, 173
  - asymmetric binding site, 184
  - catalytic activity, 9, **139–154**
  - inhibition, 155–160
  - insulin effects on, 708
  - key, control of, 518
  - kinetic equation, 142–144
  - kinetics, 148–153, 156
  - markers for disease, 142
  - mechanisms, 182–184
  - memory, 147
  - model compound, 178
  - proteases, 185
  - proteins, behavior of, **165–192**
  - pyruvate to acetyl-CoA, to convert, 537
  - turnover number, 153–154, 157
  - turnover numbers and K<sub>M</sub>, 154*t*
- enzyme-substrate binding, 144–145
- enzyme-substrate complex, 145
- epidemiology, 400–401, 413
- epigenetic mechanism, HT24
- epigenetics, 248
- epimers, 454
- epinephrine, 68, 697*t*, 704–705
- epithelial cells, HT20
- epitopes, 411
- epogen, 373
- Epstein–Barr virus (EBV), 397*t*
- equilibrium, 28, 430–431
- equine arteritis virus, 397*t*
- equivalence point, 47–48
- erythropoietin (EPO), 351, 373
- Escherichia coli*
  - amino acids, 666
  - bacterial promoter, 290*f*
  - chain elongation, 334
  - chain initiation, 331–332
  - chaperonins, 346
  - DNA polymerase, 258–259, 260*t*, 263*t*
  - DNA recombination, 270–271
  - DNA replication, 254–255
  - fatty acid synthase, 607
  - green fluorescent protein (GFP), HT3
  - human proteins and, 368
  - mismatch repair, 266*f*
  - mutualistic symbiosis, 24
  - operons, 291–292
  - plasmids, 362
  - primase reaction, 262
  - prokaryotic systems, research on, 245
  - protein folding chaperones, 110
  - protein synthesis, 339*t*
  - ribosomes, 334
  - RNA polymerase, 282–283
  - transcription regulation, 288–289
  - trp* operon, 293*f*
- essential amino acid, 662
- essential fatty acids, 608
- ester, 4*t*, 146
- esterases, 189
- esterification reactions, 460
- estradiol, 619
- estrogens, 697*t*, 699
- ethanol, 435–436, 484

ethanol/acetaldehyde couple, 565  
 ethers, 4*t*  
 eubacteria, 22–23  
 eukaryote  
   cell cycle, 270*f*  
   cells, 16–21  
   DNA (deoxyribonucleic acid), 238  
   DNA polymerases, 273–274, 276  
   DNA replication, 257*f*, 270, 272–273, 276  
   genetic engineering in, 372–374  
   prokaryote and, 13–16, 23–25, 244*f*  
   replication fork, 274–275, 276*f*  
   transcription, 296–297, 299–302  
   transcription regulation, 302  
   translation, 341–343, 345  
 eukaryotic initiation factor eIF, 342  
 eukaryotic mRNA, 341*f*  
 evolutionary biology, 277, 319  
*ex vivo*, 403  
 excision exonuclease, 267  
 exercise, 709  
 exercise physiology, 514  
 exergonic, 28  
 exergonic reaction, 444  
 exoglycosidase, 470  
 exon skipping, 349  
 exonic splicing enhancers (ESE), 348  
 exons, 314, 368  
 exonuclease I, 267  
 exonucleases, 357  
 expression cassette, 403  
 expression profiles, 390*f*  
 expression vector, 371  
 extended promoter, 284  
 extremophiles, 23

## F

facilitated diffusion, 209, 210*f*  
 FAD-linked oxidation, 545  
 familial hypercholesterolemia, 623  
 family trees, HT6–HT7  
 farnesyl pyrophosphate, 615  
 fat cell, 213  
 fatty acid(s)  
   anabolism, 555  
   biosynthesis, 602–608, 609*t*  
   defined, 193–195  
   degradation, 609*t*  
   hydrocarbon tails, 202*f*  
   hydrophobic example, 38*t*  
   insulin resistance, HT11  
   metabolism, 608*f*  
   oxidation, 592, 596–598  
   synthase, 605  
   synthesis, 709*t*  
   triacylglycerols in adipocytes, release  
     from, 591, 592*f*–593*f*  
 feedback control system, 696*f*  
 feedback inhibition, 165, 166*f*, 656, 674*f*,  
   680*f*  
 feeder cells, HT23

fermentation, 484  
 fertilizers, 655  
 fetal alcohol syndrome, 503  
 fetal hemoglobin, 105  
 fiber, 474  
 fibroblasts, HT23  
 fibronectin type I module, 91*f*  
 fibrous proteins, 93  
 filoviruses, 397*t*  
 filter-binding assay, 326, 327*f*  
 first order, 143  
 first-order kinetics, 148  
*Fis* sites, 289  
 Fischer, Emil, 454*ph*  
 Fischer projection, 453  
 Fischer projection formulas, 456*f*, 458*f*  
 five-kingdom classification system, 21–24  
 flavin adenine dinucleotide (FAD), 437*f*  
 flavin coenzymes, 187*t*  
 flavin mononucleotide (FMN), 437*f*, 567  
 flaviviruses, 397*t*  
 flowers, 466  
 flu virus, 413  
 fluid-mosaic model, 207–208  
 fluorescence, 356, HT3–HT5, HT18  
 fluorescent labeling, 387*f*  
 fluorine compounds, 543  
 fluoroacetyl-CoA, 543  
 fold recognition, 106  
 folic acid, 663*f*–664*f*, 671  
 follicle-stimulating hormone (FSH), 697*t*  
 food pyramid, 692–695  
 foreign DNA, 362  
 forensic analysis, 382  
 forensics, 379  
 formic acid, 46*t*  
 formyl-L-tryptophan, 180  
 N-formylmethionine-tRNA<sup>fm<sup>t</sup></sup>  
   (fm<sup>t</sup>-tRNA<sup>fm<sup>t</sup></sup>), 332  
 Fos, 419  
 Fourier series, 95–96  
 Fox, Michael J., HT26*f*, HT27  
 foxglove, 466  
 free energy, 28, 145, 440*t*, 443  
 free energy changes, 429–430  
 free energy profiles, 140*f*  
 free radicals, 218  
 freeze-fractured membrane, 208*f*  
 frequency, 630  
 frost-free plants, 369  
 fructose- biphosphatase-2 (FBPase-2), 519  
 fructose-1,6-biphosphatase, 518, 520–521  
 fructose-1,6-bisphosphate, 490, 499*f*  
 fructose-2,6-bisphosphate (F2,6P), 519–520  
 fruits, 466  
 full acetal, 460–461  
 fumarase, 546  
 fumarate, 545, 670  
 fumarate/succinate couple, 565  
 functional groups, 3, 4*t*, HT15  
 fungal fatty acid synthase, 608*f*  
 fungi, 22

furanose, 456  
 furanose structures, 458*f*  
 furanosides, 461  
 fusion proteins, 375

## G

G proteins, 699  
*gag* gene, 403  
*Gal4*, 311–312  
 galactoside permease, 212*f*  
 gangliosides, 199*f*, 202*f*, 613  
*GC box*, 311–312  
 gel electrophoresis, 126, 355, 356*f*  
 gel-filtration chromatography, 120, 121*f*  
 gel-to-liquid crystalline phase transition,  
   203*f*  
 gene expression, 306*f*, 602  
 gene therapy, 367, 403  
 general acid–base catalysis, 183  
 general transcription factors (GTFs), 299  
 genes, 13, 239, 288, 293*f*, 362, 403, 609, 695  
 genetically engineered, 367–368  
 genetic(s)  
   anticipation, 306  
   cancer and, 417  
   chloroplast genes, 646–647  
   code, 9, 281–322, 324–326, 327*f*  
   control, 520*t*  
   diversity, HT7–HT8  
   DNA (deoxyribonucleic acid), 232, 267  
   endocrinology and, 309  
   engineering, 367–369, 372, 374  
   Human Genome Project (HGP), 240  
   lipid metabolism, 609  
   message, 323–354  
   nucleic acids, 232, 240, 248, 267  
   photosynthesis, 646–647  
   protein synthesis, 348–349  
   recombination, 268–269, 373  
   silent mutations, 348–349  
   Transcription factor TFIID, 303  
   transcription of the genetic code, 303,  
     309  
   twins, 248  
   viruses, cancer, and immunology, 417  
 geneX, 362  
 Genghis Khan, HT6  
 genome, 13, 24*f*, 303  
 genomic library, 376*f*  
 genomics, 249, 387–390  
 germ-line, 372  
 Gibbs, J. Willard, 28*ph*  
 Gibbs free energy, 431  
 gigantism, 699  
 globular proteins, 93  
 glucagon, 697*t*  
 glucagon action, 706*f*  
 glucocerebroside, 198*f*  
 glucocorticoid-response element (GRE),  
   304  
 glucocorticoids, 620, 697*t*, 698

- glucogenic amino acid, 666–667  
 gluconeogenesis, 514–519, 522, 554*f*  
 gluconeogenesis regulation, 309  
 *$\beta$ -D-Glucopyranose*, 456*f*  
 glucopyranose, 458*f*, 461*f*  
 D-glucose, 456*f*  
 $\alpha$ -D-glucose, 462*f*–464*f*  
 $\beta$ -D-glucose, 462*f*  
 glucose  
   breakdown, 709*t*  
   complex, 489*f*  
   Fischer projection formulas, 456*f*  
   gluconeogenesis, 514–519  
   glycogen formation from, 509  
   glycolysis, 481–482, 483*f*  
   oxidation, 586  
   pentose phosphate pathway, 524–529  
   phosphate ester of, 461*f*  
   starch monomer, 469*f*  
   uptake, 707, 709*t*  
 glucose-1-phosphate, 508  
 glucose-6-phosphatase, 518  
 glucose-6-phosphate, 485  
 glucose-6-phosphate dehydrogenase, 524–525  
 glucosephosphate isomerase, 487*f*, 489  
 GLUT4, 707  
 glutamate, 656, 657*f*  
 glutamate dehydrogenase (GDH), 659  
 glutamic acid, 65*t*, 75  
 glutamine, 65*t*, 656, 657*f*  
 glutamine synthetase activity, 657*f*  
 glutamine synthetase (GS), 659–660  
 glutamine-rich domains, 311–312  
 glutathione, 75, 76*f*, 528  
 glyceraldehyde, 451–452  
 glyceraldehyde-3-phosphate, 485, 486*f*, 490–499  
 glyceraldehyde-3-phosphate dehydrogenase, 495, 496*f*  
 glycerol, 195*f*  
 glycerolipids, 197*f*  
 glycerol-phosphate shuttle, 583–584  
 glycine, 62*f*, 65*t*, 664*f*  
 glycine residues, 473*f*  
 glycogen  
   breakdown, 509*f*, 709*t*  
   granules, 472*f*  
   loading, 514  
   metabolism, 511  
   phosphorylase, 508  
   phosphorylase activity, 175, 512*f*  
   produced and degraded, 507–514  
   starch and, 471  
   synthase, 510, 511*f*  
   synthesis, 512*f*, 709*t*  
 glycogen synthase D, 513  
 glycogen synthase I, 513  
 glycogenin, 510  
 glycolipid, 193, 198  
 glycolysis, **481–506**, 709*t*  
 glycolytic pathway, 483*f*, 498  
 glycoprotein, 459, 475–476  
 glycosaminoglycans, 475, 475*f*  
 glycosides, 460–461  
 glycosidic bond, 228, 461–462  
 glycosyl residues, 471*f*  
 glyoxylate cycle, 19, 550–551  
 glyoxysomes, 19, 550  
 GMP synthesis, 673–674  
 goiter, 698  
 Golgi apparatus, 17*f*, 19, 20*f*, 21*t*, 347  
 gonadotropin-releasing factor (GnRF), 697*t*  
 gout, 675–677  
*The Graduate*, HT13  
 gramicidin A, 576  
 gramicidin S, 78*f*  
 grana, 19, 627  
 GRB2, 418  
 Greek key, 88–89, 90*f*, 92*f*  
 green chemistry, 190  
 green fluorescent protein (GFP), HT3–HT5  
 green monkeys, HT3  
 GroEL, 346  
 GroES, 346  
 Group I ribozymes, 318  
 Group II ribozymes, 318  
 group-specific affinity chromatography, 122*t*  
 group-transfer reactions, 526  
 growth hormone-releasing factor (GRF), 697*t*  
 growth-factor module, 91*f*  
 guanidine hydrochloride, 100  
 guanine, 228, 676*f*  
 GUN1, 647
- ## H
- HadII*, 358  
 half reaction, 434–436  
*Halophiles*, 23  
 hantaviruses, 397*t*  
 haploid genotype, HT7–HT8  
 Hatch–Slack pathway, 646  
 Haworth projection formulas, 456–457  
 Hayworth representations, 458*f*  
 health, 457  
 health sciences, 142  
 heart disease, 620  
 heat, 100  
 heat of a reaction at constant pressure, 29  
 heat-shock element (HSE), 304  
 70,000 MW heat-shock protein, 110  
 heavy chains, 412  
 helical structure, 95*f*  
 helical twist, 236*f*  
 helical wheel structure, 312*f*  
 helicase, 261  
 helices, 236*f*  
 $\alpha$ -helix, 84–87  
 helix, 471*f*  
 helix-turn-helix (HTH), 309–311  
 helper T cells (T<sub>H</sub> cells), 400*f*, 405, 407*f*, 409, 410*f*  
 hematocrit, HT29  
 heme group, 97–99, 574*f*  
 hemiacetal, 454  
 hemiketal, 454  
 hemoglobin, 100–102, 103*f*, 104–106, 110  
 $\beta$ -hemoglobin subunit, 87*f*  
 hemolytic anemia, 528  
 Henderson–Hasselbalch equation, 47, 51, 54  
 hepadnaviruses, 397*t*  
 hepatitis a virus, 397*t*  
 hepatitis c virus, 397*t*  
 herpes simplex virus (HSV), 397*t*, 415  
 herpesviruses, 397*t*  
 heterogeneous nuclear RNA (hnRNA), 246  
 heteropolysaccharide, 468  
 heterotrimeric G proteins, 700*f*  
 heterotrophs, 639  
 heterotropic effects, 169  
 heterozygous, 382  
 hexagonal virus, 396*f*  
 hexokinase, 486, 489*f*  
 hexokinase glucokinase, 487*f*  
 hexokinase-glucose complex, 489*f*  
 hexose monophosphate shunt, 527  
 HGPRT reaction, 677*f*  
 high altitudes, 56  
 high-energy bonds, 438*f*  
 high-performance liquid chromatography (HPLC), 124, 127  
*his-tag* sequence, 375  
 histamine, 75  
 histidine, 65*t*, 70–71, 75  
 histidine (57), 179  
 histones, 238  
 HIV, 160, 395, 397*t*, 400*f*, 402  
 HIV-1 protease, 185  
 HMG-CoA, 615  
 hnRNA (heterogeneous nuclear RNA), 246  
 Holliday Model, 269  
 holoenzyme, 282  
 homeostasis, 696  
 homogenization, 117  
 homologous, 101  
 homologous recombination, 268–269  
 homology, 106  
 homopolysaccharide, 468  
 homotropic effects, 169  
 homozygous, 382  
 hormonal control, 698*f*  
 hormone action, 698*f*  
 hormones, 695–706  
 hot spots, 269  
 HPLC chromatogram, 128*f*  
 HPV vaccines, HT20–HT22  
*Hsp60*, 346  
*Hsp70*, 346  
 human body, 2*f*  
 human chromosome 7, 384*f*  
 human cytomegalovirus (HCMV), 415  
 human genetic diversity, HT7–HT8  
 human genome, 417



- Human Genome Project (HGP), 240, HT8, HT18–HT19
- human growth hormone (HGH), 373
- human immunodeficiency virus (HIV), 160, 395, 397*t*, 400*f*, 402
- human migration, HT7
- human mitochondrial genome (mtDNA), 572
- human papillomavirus (HPV), 397*t*, HT20–HT22
- human parvovirus B19, 397*t*
- human plasma, 620
- human proteins, 373
- Human T-cell lymphotropic viruses (HTLVs), 397*t*
- Huntington's disease, 111
- hybridoma, 412
- hydration, 594
- hydration shells, 36*f*
- hydrocarbon, non polarity of, 37
- hydrogen, 6*t*, 35*t*, 190
- hydrogen bonding, 39–43, 88*f*
- hydrogen ion in water, 44
- hydrogen peroxide decomposition, 141
- hydrogen-bond acceptor, 39
- hydrogen-bond donor, 39
- hydrolysis, 439, 440*f*, 440*t*, 441–442, 445
- hydrophilic, 36
- hydrophilic substances, 38*t*
- hydrophobic, 37
- hydrophobic bonds, 37
- hydrophobic interactions, 37, 94, 107–109
- hydrophobic substances, 38*t*
- $\beta$ -hydroxy- $\beta$ -methylglutaryl-CoA, 615
- $\beta$ -hydroxybutyrate, 601
- hydroxyethyl-TPP, 538*f*
- hydroxylysine, 66*f*
- hydroxyproline, 66*f*
- 5-hydroxytryptophan, 68
- hyperbolic, 101
- hyperchromicity, 241
- hyperfunction of the adrenal cortex, 698
- hyperthyroidism, 698
- hyperventilation, 56
- hypoglycemia, 696
- hypothalamus, 696, 698*f*
- hypothyroidism, 697–698
- hypoxanthine, 228*f*, 676*f*
- hypoxia, 351
- hypoxia inducible factors (HIFs), 351
- I**
- identical twins, 248
- immune paradox, 424*f*
- immune response, 476
- immune system, 404–414
- immunoaffinity chromatography, 117*t*
- immunoglobulin module, 91*f*
- immunoglobulins, 410
- immunology, **395–425**
- imprinting, HT24
- induced-fit model, 144–145
- induced pluripotent stem cells (iPS cells), HT25–HT26
- inducer, 289
- inducible enzyme, 289
- inducible operon, 291–292
- induction, 289
- inflammation, HT11
- influenza virion, 400–401
- informational macromolecules, 9*f*
- inhibitor, 155, 167, 171*f*, 172
- inhibitory G protein, 700
- inimo acid, 62
- initiation complex, 332, 333*f*
- initiator element (*Inr*), 298
- innate immunity, 405–406
- innate system, 424*f*
- inosine, 228*f*
- inosine monophosphate (IMP), 672–674
- inositol 1,4,5-triphosphate (IP3), 700
- insects, 232
- insert, 362
- instrumentation, 134
- insulin, 373, 697*t*, 706–711, HT10–HT12
- insulin, synthesis of, 370*f*
- insulin growth factor 2 (IGF2), 351
- insulin resistance, HT10–HT12
- insulin sequence, 707*f*
- insulin-degrading enzyme (IDE), 709
- insulin–glucose index, 709
- insulin-receptor substrates (IRSs), 706–707
- insulin-sensitive protein kinase, 703
- integral proteins, 205, 206*f*
- integrase, 160
- interchain bonds, 87
- intergranal lamellae, 627
- interleukins, 408
- intermembrane space, 534–535
- internet resources, 107, 662
- intrachain bonds, 87
- intramolecular reaction, 456*f*
- intrinsic termination, 287
- introns, 246, 314–315, 368
- ion channels, HT15
- ion product constant for water, 45
- ion-dipole interactions, 36, 37*f*
- ion-exchange chromatography, 117*t*, 123, 124*f*
- ionic compounds, 36, 38*t*
- ionization of water, 44
- ionophores, 578
- ions, 36*f*
- iridoviruses, 397*t*
- iron, 6*t*, 190, 692
- iron-sulfur protein, 546
- isocitrate, 540–541
- isocitrate dehydrogenase, 541, 542*f*
- isocitrate lyase, 550
- isoelectric focusing, 126
- isoelectric pH, 70*f*, 71
- isoforms, 316–317
- isolated system, 433
- isoleucine, 65*t*
- cis-trans* isomerase, 599
- cis-trans* isomerization, 214, 217
- isomerization, 482, 484, 540, 644
- isopentenyl pyrophosphate, 615
- isoprene unit, 218–219, 614
- isozyme, 142, 490
- J**
- jellyfish, HT3
- Jenner, Edward, 405*ph*
- Johnson, Ben, HT28
- Jun transcription factor, 419
- K**
- K systems, 168–169
- kangaroo rat, 668*ph*
- $k_{cat}/K_M$ , 157
- $\alpha$ -keto acid, 661
- ketogenic amino acid, 666–667
- $\alpha$ -ketoglutarate, 541–542
- $\alpha$ -ketoglutarate dehydrogenase complex, 543
- ketone bodies, 601
- ketones, 4*t*
- ketose, 451, 453
- ketosis, 601
- kinase, 147, 174, 486, 496
- kinetic data, 157
- kinetics, 139, 148, 178
- Klf4 gene, HT25
- $K_M$ , 151–153, 154*t*, 157
- Koagulation, 218
- Kozak sequence, 343
- Krebs, Edwin G., 174*ph*
- Krebs cycle, 533, 536*f*
- kringle module, 91*f*
- kwashiorkor, 666
- L**
- labeling, 178
- lac* operon, 290*f*, 291, 371
- lac* repressor, 290*f*, 291, 292*f*
- lactate, 482*f*, 499
- lactate dehydrogenase (LDH), 142, 500
- lactic acid, 46*t*, 57
- lactic acid fermentation, 501*f*
- Lactobacillus*, 481
- lactone, 459
- lactose, 289, 467
- lactose intolerance, 467
- laetrile, 466
- Landis, Floyd, HT30
- lanosterol, 615, 617*f*
- lariats, 315–316
- lassa virus, 397*t*
- lauric acid, 195*t*
- law, 239
- law of mass action, 517
- LDL particle, 620–623



LDL receptor, 213*f*  
 leaving group, 182  
 leguminous plant, 24*f*  
 LeMond, Greg, 585  
 leptin, 213, 695  
 lesions, 271  
*let-7*, 307  
 leucine, 65*t*  
 leu-enkephalin, 697*t*  
 leukemia, HT4  
 leukemia inhibitory factor (LIF), HT23  
 leukocytes, 404  
 Leukotriene C, 221  
 leukotrienes, 219–222  
 life and thermodynamics, 28–30  
 life processes, 1, 26  
 ligand, 121, 219  
 light, 630  
 light chains, 412  
 light reactions, 631–637  
 lignin, 474–475  
*lin-4*, 307  
 lineage switching, HT25–HT26  
 linear hydrogen bond, 39*f*  
 linear polyglucose chain, 463*f*  
 linear template, 275  
 Lineweaver–Burk double-reciprocal plot, 151–152, 156, 158–159  
 linkage, 9  
 linoleic acid, 195*t*, 600*f*  
 linolenic acid, 195*t*  
 lipases, 195, 591  
 lipid(s)  
   anabolism, 554, 555*f*  
   anchors, 206  
   bilayers, 200–201, 202*f*, 203–205, 207–208  
   biosynthesis, 602  
   citric acid cycle, 557*f*  
   defined, 193  
   droplets, 213  
   metabolism, **591–625**  
   prostaglandins and leukotrienes and, 219–222  
   proteins and, **193–225**  
   types, chemical nature of, 194–198  
 lipid-soluble vitamins, 214–219  
 lipoic acid, 187*t*, 537–538  
 lipoproteins, 620  
 liposome, 107, 108*f*, 208  
 lipotoxicity, HT11  
 liver pyruvate kinase, 524*f*  
 living organisms, 54–55  
 lock-and-key model, 144  
 longevity, 710–711  
 loops, 245  
 lovastatin, 615  
 low-carbohydrate diets, 457, 708  
 luminescence, 356  
 luminescent jellyfish, HT3  
 luteinizing hormone (LH), 697*t*  
 T-lymphocytes, 309  
 lymphocytes, 405, 407*f*

lymphoid stem cell, 405  
 lysine, 65*t*, 66*f*, HT15  
 lysogeny, 397  
 lysophosphatidate, 609  
 lysosome, 17*f*, 19, 21*t*  
 lysozyme, 47*f*, 154*t*  
 lytic pathway, 397

## M

macromolecules, 9*f*–10*f*  
 macronutrients, 688  
 macrophages, 405  
 mad cow disease, 111–112  
 magnesium, 6*t*  
 major groove, 234  
 major histocompatibility complex (MHC), 405  
 L-Malate, 546  
 malate dehydrogenase, 536*f*, 546  
 malate synthase, 550  
 malate–aspartate shuttle, 583–584  
 malic acid, 46*t*  
 Maloney murine leukemia virus (MMLV), 403  
 malonic acid, 46*t*  
 malonyl-CoA, 603, 604*f*, 606  
 maltose, 468  
 mammalian cell, 20*f*  
 mammalian target of rapamycin (mTOR), 711  
 mammals, 553*f*  
 manganese, 6*t*  
 MAP kinase signal transduction, 419*f*  
 marburg virus, 397*t*  
 margarine, 204  
 marmosets, HT5  
 mass spectrometry (MS), 134  
 maternal hemoglobin, 105  
 matrix, 17, 18*f*  
 Matrix-Assisted Laser Desorption Ionization-Time of Flight (MALDI-TOF MS), 134  
*β*-meander, 88, 90*f*, 92*f*  
 measles virus, 397*t*  
 mediator, 305  
 medical genetics, 572  
 medicine, 111–113, 160, 173, 308, 403, 671, HT15–HT16  
 meiotic DX recombination intermediate, HT13  
 melting, 241  
 membrane, 18*f*, **193–225**  
 membrane proteins, 205–207  
 membrane receptors, 212  
 membrane structure, 207–208  
 membranous vesicles, 20*f*  
 memory, 344  
 memory, immune system, 404  
 memory molecule, 147  
 menaquinone, 637*f*  
*β*-Mercaptoethanol, 100  
 metabolic control, 520*t*

metabolic enzymes, 309  
 metabolic intermediates, 553*f*  
 metabolic pathways, 443–446, 687  
 metabolic syndrome, HT11  
 metabolic water, 598  
 metabolism  
   anaerobic, 499–503  
   citric acid cycle, 533  
   control, 703–706  
   electron transfer role in, 446*f*  
   electron transports role in, 563–564  
   energy changes/electron transfers in, **429–449**  
   insulin effects on, 709*t*  
   integration of, **687–714**  
   lipids, **591–625**  
   nature of, 432  
   nitrogen, **653–685**  
   oxidation and reduction, 432–434  
   sports and, 585  
 metal-ion catalysis, 183  
 metal-response element (MRE), 304  
 metastasis, 416  
 met-enkephalin, 697*t*  
 methane, 41*t*  
*Methanococcus jannaschii*, 24  
*methanogens*, 23  
 methionine, 65*t*  
 methyl alcohol, 461*f*  
 methylation of DNA, 248, 265  
 5-methylcytosine, 248  
 mevalonate, 614–615, 616*f*  
 mevinolinic acid, 615*f*  
 micelle formation, 39*f*  
 micelles, 38  
 Michaelis constant, 149–150, 153  
 Michaelis–Menten Approach, 148–154  
 micro RNAs (miRNA), 244*f*, 307–308  
 microarrays, 388  
 microbiology, 271  
 micronutrients, 690–691  
 microsatellites, HT7  
 microscopy, HT17–HT19  
 microtrabecular lattice, 20  
 middle phase, 288  
 milk, 467  
 milk production, 369  
 Miller–Urey experiment, 7  
 mineralocorticoids, 620, 697*t*, 698  
 minerals, 691–692  
 minor groove, 234  
 mirror images, 455*f*  
 mismatch repair, 264, 266*f*  
 mitochondria/mitochondrion  
   acetyl groups, transfer from, 603*f*  
   animal cell, 17*f*  
   disease, 572  
   electron transport, 637, 639  
   endoplasmic reticulum (ER), 19*f*  
   eukaryotic cells, 15  
   eukaryotic organelle, 17, 18*f*  
   gluconeogenesis, 554*f*

- lipid anabolism transfer from, 555*f*
  - matrix, 534–535, 579*f*, 594*f*
  - matrix space of, 575*f*
  - membrane, 564*f*, 570*f*
  - organelles and their functions, 21*t*
  - plant cell, 17*f*
  - prokaryotes and eukaryotes, comparison of, 15*t*
  - prokaryotic cell, 17*f*
  - structure, 563
  - structure of, 534*f*
  - mitogen activated protein kinase (MAPK), 305, 418
  - mitogen-activated protein kinase kinase (MAPKK), 419
  - mixed inhibition, 159
  - mobile phase, 118
  - modified standard state, 430
  - modules, 91*f*
  - molecular
    - biology, 385
    - genetics, 339
    - immune system, aspects of the, 411–414
    - nature of life processes, 1
    - oxygen, 573
    - taxonomy, 232
    - weight, 127*f*
  - molecular-sieve chromatography, 117*t*
  - molecule(s)
    - amphipathic, 39*f*
    - amphiphilic, 38*f*
    - antenna, 630*f*
    - to cells, 10–11
    - hydrogen bonds found in, 43*t*
    - memory, 147
    - origination of, 2–3
    - single, HT17
    - water, 109
  - monera, 22
  - monkeypox virus, 397*t*
  - monoclonal antibodies, 412–413
  - Monod–Wyman–Changeux (MWC) model, 170, 171*f*
  - monomers, 7–8
  - monosaccharides, 451, 455*f*, 459–464, 470*f*
  - motif, 90, 91*f*
  - mouse liver cells rough endoplasmic reticulum, 19*f*
  - mouse liver mitochondria, 18*f*
  - mRNA (messenger RNA), 244*f*, 246, 294, 314–315
  - mRNA looping, 343*f*
  - mt DNA (human mitochondrial genome), 572
  - mucosal membrane, HT20
  - Mullis, Kary B., 377*ph*
  - multifunctional enzyme, 607
  - multiple cloning site (MCS), 364
  - multipotent adult progenitor (MAP) cells, HT25
  - multipotent stem cells, HT23
  - mumps virus, 397*t*
  - muscle, 499
  - muscle cells, HT5
  - muscular dystrophy, 218
  - mutagens, 264
  - mutations, 263
  - mutualism, 24
  - c-Myc gene, HT25
  - myoglobin, 96–98, 99*f*, 100–102, 104*f*
  - myohemerythrin, 87*f*
  - myristic acid, 195*t*
- N**
- Na<sup>+</sup>/K<sup>+</sup> ATPase, 211*f*
  - NAD<sup>+</sup>, 187, 501
  - NADH, 187, 501, 507, 524, 528
  - NADH-CoQ oxidoreductase, 549, 555, 567–568, 570*f*
  - NADH-linked dehydrogenase, 495
  - NADP<sup>+</sup>, 634–635
  - NADPH (nicotinamide adenine dinucleotide phosphate), 524
  - Naill (Ireland), HT6
  - nanoparticles, 422, HT16
  - nanotech, 422
  - nanotechnology, HT13–HT16
  - native conformations, 83
  - native gel, 126
  - natural killer (NK) cell, 405–406
  - Neanderthals, HT8–HT9
  - negative cooperativity, 172–173
  - negative regulation, 291
  - negative selection, 415
  - negative supercoils, 237
  - nene geese, 232
  - nerve cells, 309
  - nerve synapses, 308
  - net chain growth, 255
  - net reaction, 481
  - neurodegenerative diseases, 112–113
  - neurology, 344
  - neuroscience, 79, 147, 154, 217
  - neutral pH values, 45
  - neutral polar side chains, 65
  - Nicholas II (Russia), HT6
  - nick translation, 264
  - nicotinamide adenine dinucleotide (NADH), 187, 501, 507, 524, 528
  - nicotinamide adenine dinucleotide phosphate (NADPH), 524
  - nicotinamide coenzymes, 435–436
  - nicotinamide ring, 188*f*
  - Nitella*, 19*f*
  - nitrification, 653
  - nitrogen, 6*t*, 35*t*, 190, **653–685**
  - nitrogen fixation, 369, 653–656
  - nitrogen wastes, 668
  - nitrogenase enzyme complex, 654
  - nitrogenase reaction, 655
  - p*-nitrophenolate, 146, 178
  - p*-nitrophenyl acetate, 178
  - p*-nitrophenylacetate concentration, [S], 146*f*
  - NLRP3 protein, 677
  - no net oxidation, 555
  - non-coding RNA, 307
  - noncompetitive inhibition, 155, 158–159, 168–169
  - noncovalent bonds, 41*t*
  - nonheme iron protein, 546, 573, 574*f*
  - nonhomologous recombination, 268–269
  - nonlinear hydrogen bond, 39*f*
  - nonoverlapping code, 324
  - nonoxidative reactions, 526
  - nonphagocytic cells, 406*f*
  - non-polar (hydrophobic) amino acids, 63*f*–64*f*
  - nonpolar bonds, 35
  - nonpolar covalent compounds, 38*t*
  - nonpolar side chains, 62
  - nonpolar solute, 109
  - nonsteroid hormones, 698*f*
  - nontemplate strand, 282
  - norepinephrine, 697*t*
  - normal RNA splicing, 349
  - nuclear double membrane, 17
  - nuclear magnetic resonance (NMR) spectroscopy, 95–96
  - nuclear region, 16
  - nuclear transfer, HT26
  - nucleases, 254, 357
  - nuclei, 6
  - nucleic acid bases, 227–228
  - nucleic acids, 8–9, 10*f*, **227–251**, **253–279**, **355–393**, 396*f*
  - nucleobases, 227–228
  - nucleocapsid, 395
  - nucleolus, 17, 18*f*
  - nucleophile, 180–181
  - nucleophilic substitution reactions, 182
  - nucleoside, 228
  - nucleosidediphosphate kinase, 544
  - nucleosome, 238
  - nucleotide-excision repair, 267–268
  - nucleotides, 8*f*, 11, 227–229, 230*f*, 257*f*, 314*f*
  - nucleotides (dNTPs), HT18–HT19
  - nucleus, 15–17, 18*f*, 21*t*
  - nutlins, 421
  - nutrients, 688–690
  - nutrition
    - biochemistry and, 687–695
    - butter *v.* margarine, 204
    - complete protein and, 85
    - insulin and low-carbohydrate diets, 708
    - iron, 692
    - lactose intolerance, 467
    - low-carbohydrate diets, 457, 708
    - obesity, 580, 612
    - protein, 85
    - salmon, 222
    - vitamin C, 461
    - weight loss, 558

## O

obesity, 580, 609, 612, 695, 709,  
HT10–HT12

obesity (ob) gene, 695

Oct3/4 gene, HT25

odd-carbon fatty acids, 598–600

oleic acid, 195*t*

oligomer, 100

oligonucleotides, 355

oligosaccharides, 451, 464–468

OmpF porin crystal, HT18

oncogene, 399, 417–420

oncology, 424

one-carbon transfer, 661–662

open complex, 284–285, 286*f*, 301

open system, 433

operator (*O*), 291

operons, 289

opsin, 214, 217

optical isomers (stereoisomers), 451

organ transplants, 404

organelle, 15–17, 21*t*

organic chemistry, 3, 75, 157

organophosphates, 440*t*

origin of replication, 255

origin recognition complex (ORC), 270

ornithine, 78*f*, 667

orthomyxoviruses, 397*t*

oryzias latipes, 238

overall order, 143

oxalic acid, 46*t*

oxaloacetate, 516, 546, 554*f*, 557*f*

$\beta$ -oxidation, 594, 595*f*, 596–599, 600*f*

oxidation

- citric acid cycle, 541
- damage, 266*f*
- defined, 26
- fatty acids, 594, 596–598
- glucose, yield of ATP from, 586
- glycolysis reactions, 484
- metabolism, role in, 432–437
- polyunsaturated fatty acids, 600*f*
- unsaturated fatty acid, 599

oxidation–reduction reactions, 188*f*, 432, 434–437, 459

oxidative decarboxylation, 535

oxidative phosphorylation, 482*f*, 533, 563–589

oxidative reactions, 524–525

oxidizing agent, 432

oxygen

- binding, 101–102, 105–106
- Bohr effect, 103*f*, 104
- citric acid cycle, 557–558
- electronegativity, 35*t*
- elements relative to carbon, 6*t*
- green chemistry, 190
- heme group, binding to the, 97–99
- hypoxia, 200, 351
- photosynthesis and, 639–642
- oxygen-binding site, 99*f*

oxygen-evolving complex, 632

oxyhemoglobin, 103*f*

oxytocin, 77, 79, 697*t*

## P

P (peptidyl) site, 334

p21<sup>ras</sup>, 418

p300 protein, 306*f*

p53 tumor suppressor gene, 419–420, HT25

Paabo, Svante, HT8–HT9

palindrome, 358

palmitate synthesis, 605, 606*f*

palmitic acid, 195*t*

palmitoleic acid, 195*t*

Pantani, Marco, HT29–HT30

Pap smear, HT20

papain, 185

papillomaviruses, 397*t*

papovaviruses, 397*t*

paramecium, 15

paramyxoviruses, 397*t*

parasitic symbiosis, 24

parvoviruses, 397*t*

passive diffusion, 209*f*–210*f*

passive transport, 209

passive tumor targeting, 422

paternity, 379

pathogens, 295

pause sites, 301

1-2 pause structure, 294–295

payoff phase, 491

pBR322 plasmid, 364

PCNA homotrimer, 273

pectin, 474

pentaglycine cross-links, 473*f*

pentapeptide, 473*f*, 474

pentose phosphate pathway, 524–529

pepsin, 47*f*

peptide bond, 334

peptide chain termination, 338*f*

peptide nucleic acids (PNA), 13

peptide(s)

- amino acids and, 61–81
- backbone, 84–85
- bond, 72–74, 336
- chain, 73*f*
- chymotrypsin, reactions catalyzed by, 146
- digestion, 130*f*
- hormones, 77, 79
- sequence of, 75, 129–130, 132–134
- small, 74–79

peptidoglycan, 473*f*, 474, 476

peptidyl transferase, 334

percent recovery, 117

perilipin, 213

peripheral proteins, 205, 206*f*

peroxisomes, 19, 21*t*

pET expression vectors, 371*f*

pH, 44–49, 54, 57–58

pH extremes, 100

phagemids, HT14

phages, 357

phenylacetate, 78

phenylalanine, 65*t*, 68

phenylalanine hydroxylase, 78

phenylisothiocyanate (PITC), 132

phenylketonuria (PKU), 78

phenyllactate, 78

phenylpyruvate (a phenyl ketone), 78

pheophytin (Pheo), 634

phi angle, 84

phosphate ester, 38*t*, 461*f*

phosphate ion, 508

phosphatidate, 610

phosphatidic acid, 196

phosphatidyl choline (lecithin), 196

phosphatidyl ester, 196

phosphatidyl ethanolamine (cephalin), 196

phosphatidyl glycerol, 196

phosphatidyl inositol, 196

phosphatidyl serine, 196

phosphatidylcholine, 197*f*, 592*f*

phosphatidylethanolamine, 610–611, 612*f*, 613

phosphatidylglycerols, 197*f*

phosphatidylinositol, 197*f*

phosphatidylinositol 4,5-bisphosphate, 700

phosphatidylserine, 613

phosphoacylglycerols, 193, 196, 197*f*, 202*f*, 592*f*, 610–611

3',5'-phosphodiester bond, 229

phosphoenolpyruvate (PEP), 440*f*, 497–498, 499*f*, 516, 554*f*

phosphoenolpyruvate carboxykinase (PEPCK), 516

phosphofructokinase, 487*f*, 490

phosphofructokinase-2 (PFK-2), 519

phosphoglucosyltransferase, 508

6-phosphogluconate dehydrogenase, 526

2-phosphoglycerate, 497, 499*f*

3-phosphoglycerate, 494–495, 497, 499*f*

phosphoglycerate kinase, 493*f*, 496

phosphoglycerides, 200

phosphoglyceromutase, 493*f*, 497

phospholipase C, 700

phospholipases, 591, 592*f*

phospholipid bilayer, 201*f*–202*f*

phosphopantetheine group, 607*f*

phosphopentose isomerase, 526

phosphopentose-3-epimerase, 526

phosphoprotein phosphatase, 511, 548

phosphoribosylpyrophosphate (PRPP), 676*f*

phosphoribulose kinase, 645

phosphoric acid, 46*t*

phosphoric acid anhydrides, 4*t*

phosphoric acid esters, 4*t*

phosphoric anhydride bonds, 438*f*

phosphorolysis, 508

phosphorus, 6*t*, 35*t*

phosphorylase *a*, 511

phosphorylase *b*, 511

phosphorylase kinase, 511

- phosphorylation, 173–175, 481–482, 484, 524*f*, 574–577, 644
- photomicrograph, 406*f*
- photophosphorylation, 631, 638*f*
- photorespiration, 648–649
- photosynthesis, **627–651**
- photosynthesis fix CO<sub>2</sub>, 642
- photosynthetic anaerobe, 639
- photosynthetic electron transport, 638*f*
- photosynthetic reaction centers, 636
- photosynthetic unit, 630*f*
- photosystem I (PSI), 631, 633*f*, 635
- photosystem II (PSII), 631, 633*f*, 634
- photosystems, 630–637
- physical organic chemistry, 157
- physics, 630
- physiology, 351, 668
- Pic1 gene, 419–420
- picornaviruses, 397*t*
- pineal gland, metabolism of the, 309
- PIP<sub>2</sub> second-messenger scheme, 702*f*
- pitch, 85
- pituitary, 697–698
- PKM $\zeta$ , 147
- planar group, 74*f*
- plant cell walls, 474
- plant science, 369, 466, 655
- Plantae, 22
- plants, 556, 640–641, 665*f*
- plaque, 361
- plaque hybridization, 376*f*
- plasma, 620
- plasma membrane, 17*f*
- plasma membrane receptors, 698*f*
- plasmid, 362, 363*f*, 364
- plastocyanin (PC), 634
- plastoquinone (PQ), 634
- $\beta$ -pleated sheet, 84–85, 87, 88*f*
- pluripotent stem cell, HT23
- P/O ratio, 576
- pol* gene, 403
- Pol II promoters, 298–299, 300*f*, 301–302
- polar bonds, 35
- polar compounds, 36
- polar covalent compounds, 38*t*
- polar groups, 42*f*
- polarity, 35–38
- polioviruses, 397*t*
- poly A binding protein (Pab1p), 342–343
- poly A tail, 343
- 3' poly(A) tails, 247
- polyacrylamide, 120
- polyacrylamide gel, 126
- polyadenylate tail, 314
- polyadenylating, 314
- polylinker, 364
- polymer assembly, HT14
- polymerase chain reaction (PCR), 377, 378*f*, 379
- polymerase proofreading, 264
- polymerase repair, 265*f*
- polymeric structure, 469*f*, 472*f*
- polymers, 8–9
- polymorphisms, 383, HT7
- polynucleotides, 11, 227–228
- polypeptide backbone arrangement, 92*f*
- polypeptide chain, 72
- polypeptides, 697*t*
- polyribosome, 341
- polysaccharides, 10*f*, 468–475
- polysomes, 340
- poly-T tail, 314
- porins, HT18
- porphyrins, 653
- positive cooperativity, 101, 165
- positive regulation, 291
- positive supercoils, 237
- postranscriptional RNA modification, 312
- postranslational modification of proteins, 345–346
- potassium, 6*t*, 173–174
- poxviruses, 397*t*
- pp60<sup>src</sup>, 418
- PRDM9 gene, 270
- precipitate, 411*f*
- predator attraction, 369
- pregnenolone, 619
- preinitiation complex, 299
- preparation, Calvin cycle, 644
- preparation phase of glycolysis, 485
- pre-replication complex (pre-RC), 272
- Pribnow box, 284
- Prigogine, Ilya, 433
- Prima-1, 421
- primary active transport, 210
- primary structure, 84–85, 127, 128*f*, 129–135, 227
- primase, 262
- primase reaction, 262
- primer, 260
- primer (P1), HT19
- primosome, 263
- prion diseases, 111–112
- probability, 29
- processivity, 258
- production, 438–443
- progenitor cells, HT23
- progesterone, 619
- proinsulin, 706–707
- prokaryote
- cells, 16
  - DNA replication, 257*f*, 263*t*, 276
  - DNA supercoil, 237
  - eukaryote and, 13–16, 24–25
  - mRNA molecules in, 244*f*
  - polymerases, 272
  - ribosomes, 247*f*
  - transcription in, 282–287
  - transcription regulation, 287–289, 290*f*, 291–292, 293*f*, 294–296
  - translation, 331
- proline, 65*t*, 66*f*
- proline hydroxylase (PH), 351
- proline racemase reaction, 186*f*
- proline-rich domain, 312
- promoter, 289, 290*f*
- promoter locus, 283
- promoter structure, 283
- proofreading, 260, 263–268, 319
- propeller-twist, 236
- prostaglandins, 219–222
- prosthetic groups, 83
- protease, 160
- proteasomes, 347
- protein antigen X, 413*f*
- G-protein coupled receptors, HT15
- protein efficiency ratio (PER), 85
- protein kinase C, 701
- protein-folding chaperones, 110–111
- protein(s). *See also* enzyme(s)
- aggregation, 110*f*
  - amino acids as building blocks of, 63*f*–64*f*
  - arrays, 390–391
  - backbone, 84–85
  - behavior of, 139–141, **165–192**
  - biosynthesis, 323
  - from cells, extracting pure, 117–118
  - characterization techniques, **117–138**
  - cleavage at internal methionine residues
    - by cyanogen bromide, 131*f*
  - cleavage by chymotrypsin, 130*f*
  - cleavage by trypsin, 130*f*
  - cleavage into peptides, 129
  - column chromatography, 118–125
  - complementary, 85
  - complete, 85
  - conformation, 106*f*
  - control, **165–192**
  - cytochrome c, 108*f*
  - degradation, 347, 350
  - denaturation, 100
  - DNA replication, 261–263
  - electrophoresis, 126–129
  - expression vector, 370–371
  - factories, 368
  - fibrous, 93
  - folding, 348
  - folding diseases, 111–112
  - folding dynamics, 106–113
  - folding pathways, 346*f*
  - globular, 93
  - green fluorescent protein (GFP), HT3–HT5
  - integral, 205, 206*f*
  - interactions, 135
  - interactions, analyzing, 135
  - iron-containing, 573
  - kinases, 174
  - lipids and, **193–225**
  - mechanisms, **165–192**
  - membrane, 205–207
  - mobility, 127*f*
  - nanotechnology, HT14
  - nucleic acids and, 8–9
  - peripheral, 205, 206*f*
  - postranslational modification of, 345–346



- primary structure, 84–85, 127, 128*f*, 129–135
  - purification, **117–138**, 375*f*
  - purification scheme, 117*t*
  - quaternary structure, 100–102, 103*f*, 104–106
  - receptor, 207
  - Sadenosylmethionine (SAM), 662
  - secondary structure of proteins, 84–93
  - sequence, 131*f*
  - spikes, 395, 396*f*
  - structure and function, 83, 107*f*
  - synthesis, 243, 245–246, **323–354**
  - tertiary structure, 93–100, 106
  - three-dimensional structure, **83–115**
  - transport, 207
  - proteoglycans, 475*f*, 476
  - proteolytic activation, 176
  - proteome, 135, 388
  - proteomics, 135, 387–390
  - prothrombin, 220*f*
  - Protista, 22
  - proton gradient, 563–564, 577
  - proton pumps, 211, 564*f*, 578
  - proto-oncogenes, 418
  - Pseudomonas* bacteria, 239
  - psi angle, 84
  - P-TEF (positive-transcription elongation factor), 301
  - pUC plasmids, 365
  - purification, 355, 375
  - purine(s)
    - bases, 228
    - biosynthesis, 672–674
    - catabolism, 674–677
    - nitrogen metabolism, 653
    - nucleotide biosynthesis, 672*f*, 674*f*
    - nucleotide structure, 227–228
    - ring, 672*f*
    - salvage, 676*f*–677*f*
  - puromycin, 337*f*
  - pyranose, 456
  - pyranose structures, 458*f*
  - pyranosides, 461
  - pyridoxal phosphate, 187*t*, 188*f*, 660
  - pyrimidine bases, 227–228
  - pyrimidine biosynthesis, 678–680
  - pyrimidine biosynthetic pathway, 679*f*
  - pyrimidine catabolism, 678
  - pyrimidine nucleotide biosynthesis, 680*f*
  - pyrimidine nucleotides, 678
  - pyrimidines, 227–228, 653, 680*f*
  - pyruvate
    - acetyl-CoA, conversion to, 548*f*, 549
    - anaerobic metabolism, 499–503
    - carboxylase, 516, 517*f*, 552
    - citric acid cycle, 535–540
    - decarboxylase, 501, 503*f*
    - dehydrogenase complex, 535, 537, 539*f*, 540, 546–547
    - dehydrogenase kinase, 537, 548
    - dehydrogenase phosphatase, 537
    - dehydrogenase reaction, 548
    - gluconeogenesis glucose production, 514–519
    - glyceraldehyde-3-phosphate, conversion to, 486*f*, 491–499
    - glycolysis, 481–482, 483*f*, 492*f*
    - kinase, 493*f*, 498
    - pyruvate dehydrogenase (PDH), 537, 538*f*
    - pyruvic acid, 46*t*
- Q**
- Q cycle, 570–571
  - Quantitative PCR (qPCR), 379–380
  - quaternary structure, 83, 100–106, 165
- R**
- Rabies virus, 397*t*
  - random coil, 83
  - Ras protein, 418
  - rate constant, 143
  - reaction center, 630
  - reaction velocity,  $V$ , 146*f*–147*f*, 150*f*, 167*f*
  - reactions, dark, 642–646
  - reactions, predicting, 29
  - reactive hypoglycemia, 708
  - reading frames, 326
  - RecA protein, 270
  - RecBCD enzyme complex, 270
  - receptor property, 209
  - receptor proteins, 207
  - receptor tyrosine kinase, 701, 703*f*
  - recombinant DNA, 360, 363*f*
  - recombinant erythropoietin (EPO), HT29
  - recombinant human insulin, 371*f*
  - recombination, 271
  - redox, 432, 435
  - redox couple, 565*f*
  - reduced adenine dinucleotide phosphate (NADPH), 555
  - reducing agent, 432
  - reducing sugars, 459, 464*f*
  - reduction, 26, 432–437
  - reduction half reaction, 565*t*
  - reduction potential, 564–566
  - refolding, 99–100, 101*f*
  - regular structures, 87
  - regulatory dimers, 168*f*
  - regulatory gene, 289
  - relative specificity, 184
  - relatives, HT6–HT7
  - reoviruses, 397*t*
  - rep protein, 262
  - repair, 261, 263–268, 271
  - repeating disaccharide, 472*f*–473*f*, 475*f*
  - replication, 13*f*, 243*f*, **253–279**
  - replication activator protein (RAP), 270
  - replication fidelity, 264
  - replication forks, 255, 262, 274–275
  - replication licensing factors (RLFs), 272
  - replicators, 270
  - replicons, 270
  - replisome, 263
  - repressible operon, 291–292
  - repression, 291
  - repressor, 289
  - reproduction, 13*f*
  - reshuffling, Calvin cycle, 644
  - resins, 123
  - resonance structures, 72, 74*f*
  - resonance-stabilized phosphate ion, 438*f*
  - respiratory complexes, 566–567, 568*f*, 570, 580–581
  - respiratory inhibitors, 580, 582*f*, 583–584
  - respiratory syncytial virus (RSV), 397*t*
  - response elements, 289, 304
  - restricted growth, 357
  - restriction endonucleases, 357–358, 359*t*, 360*f*
  - restriction-fragment length polymorphisms, 382–384
  - retinal, 214
  - 11-*cis*-retinal, 215*f*–216*f*, 217
  - retinol, 214
  - retrograde signaling, 646
  - retrovirus genes, 402
  - retrovirus(es), 253, 397*t*, 401–403
  - Reverse Phase HPLC, 124
  - reverse transcriptase (RT), 160, 253, 275, 402
  - reverse turns, 88, 89*f*
  - reversibility, 431
  - rhabdoviruses, 397*t*
  - rhinovirus, 395, 397*t*
  - Rhodopseudomonas viridis*, 636
  - rhodopsin, 214, 216*f*, 217
  - rho-factor mechanism, 287, 288*f*
  - riboflavin, 437*f*
  - ribofuranose, 458*f*
  - ribonuclease, 101*f*
  - ribonucleoprotein particles (RNPs), 16, 315
  - ribonucleoside, 228, 229*f*
  - ribonucleotide reductase, 681
  - ribonucleotides, 681–682
  - ribose, 187
  - ribose-5-phosphate isomerase, 645
  - ribosomal architecture, 331
  - ribosomal RNA (rRNA), 244*f*, 245, 247*f*, 313–314
  - ribosomal S6 protein kinase (RSK), 711
  - ribosome structure, 331*f*, 340*f*
  - ribosomes, 15–16, 17*f*–19*f*, 21*t*, 227, 333–337
  - riboswitches, 295
  - ribozymes, 317–318, 336–337
  - ribulose-1,5-bisphosphate carboxylase/oxygenase (rubisco), 642, 643*f*, 644
  - ricketts, 218
  - RNA (ribonucleic acid). *See also* nucleic acids; transcription
    - biosynthesis, 281
    - chain, 229, 231*f*
    - detection methods, 356
    - discovery of, 10–11
    - DNA (deoxyribonucleic acid), difference between, 227



- evolutionary stages in, 12*f*–14*f*
  - gene product formation, 253
  - kinds and roles of, 244*t*
  - kinds and structures of, 242–243
  - messenger, 247–248
  - non-coding, 307
  - normal RNA splicing, 349
  - posttranscriptional RNA modification, 312
  - primase reaction, 262
  - purine bases, 228
  - retroviruses, 401–403
  - ribosomal, 245–247
  - secondary structure, 294
  - self-replicating, 277
  - silencing, 307–308
  - small nuclear, 248
  - synthesis, 281*t*
  - transcriptional proofreading, 319
  - viral, 415
  - viruses and, 395, 396*f*, 397, 401
  - RNA interference (RNAi), 243, 247, 307, 385
  - RNA polymerase, 282–285, 286*f*, 287–292
  - RNA polymerase B (RPB), 297
  - RNA polymerase II, 297, 303
  - RNA-induced silencing complex (RISC), 385
  - RNase P*, 313
  - robotic technology, 388
  - Romanov family, HT6
  - rotating molecular motor, 575*f*
  - rotaviruses, 397*t*
  - rough endoplasmic reticulum, 18–19
  - Rous sarcoma virus, 402, 418
  - rubella virus, 397*t*
  - RuvA protein, 270
  - RuvB protein, 270
  - RuvC protein, 270
- S**
- S6K1 [ribosomal S6 protein kinase (RSK)], 711
  - sal ammoniac, 653
  - salmon, 222
  - salt precipitate, 117*t*
  - salting out, 118
  - salvage reactions, 676
  - Sanger–Coulson method, 386*f*, HT18
  - Sap-1a, 305
  - saponification, 196
  - saturated fatty acid, 193–194, 195*t*, 595*f*
  - scanning mechanism, 342
  - scanning tunneling microscopy (STM), HT17–HT19
  - SDS-polyacrylamide-gel electrophoresis (SDS-PAGE), 126–127
  - second genetic code, 329, 331
  - second messengers, 699–703
  - second order, 143
  - secondary active transport, 210, 212*f*
  - secondary structure, 227, 247*f*, 294
  - secondary structure of proteins, 83–84
  - sedimentation coefficient, 245
  - sedoheptulose biphosphatase, 645
  - selectable marker, 362
  - selection, 362
  - selenocysteine (Sec), 339
  - self-assembly of ribosomes, 246
  - self-cleavage, 295
  - self-replicating RNA, 277
  - semiconservative replication, 254–255, 256*f*
  - semidiscontinuous DNA replication, 257*f*–258*f*
  - semiquinone anion, 571
  - sense strand, 282
  - separation techniques, 355–356
  - sequence homology, 106
  - sequencer, 132
  - sequencing by separation, HT18
  - sequencing by synthesis, HT18–HT19
  - sequential model, 169, 172–173
  - serine, 65*t*, 664*f*
  - serine acyltransferase, 662
  - serine hydroxymethylase, 662
  - serine proteases, 178
  - serotonin, 68, 75
  - severe combined immune deficiency (SCID), 403
  - sexually transmitted disease (STD), HT20–HT22
  - sheet structure, 95*f*
  - Shine–Dalgarno sequence (5'-GGAGGU-3'), 333, 334*f*
  - shuttle mechanisms, 583–586
  - sickle-cell anemia, 84
  - side chain group, 61–62, 66, 94
  - sigmoidal, 101
  - silencer, 289, 298, 302
  - silent mutation, 324, 348–349
  - simian virus 40 (SV40), 397–398, 399*f*
  - simple diffusion, 209
  - Simpson, Tom, HT28
  - single-molecule fluorescence spectroscopy, 205*f*
  - single-nucleotide polymorphism (SNP), HT7
  - single-strand binding protein (SSB), 262
  - siRNA (Small Interfering RNA), 307
  - sirtuins, 710–711
  - six carbon atoms, 481
  - six-carbon glucose, 485–491
  - size-exclusion chromatography, 120
  - Small Interfering RNA (siRNA), 307
  - small nuclear ribonucleoprotein particles (snRNPs), 247
  - small nuclear RNA, 244*f*, 247
  - small peptides, 74–79
  - smooth endoplasmic reticulum, 19
  - S<sub>N</sub>1 (substitution nucleophilic unimolecular), 182
  - snurps, 315–316
  - sodium, 6*t*, 173–174
  - sodium palmitate, 38*f*
  - sodium–potassium ion pump, 210, 211*f*
  - sodium–potassium pump, 174*f*
  - solvent, 35–60
  - somatic, 372
  - somatic cell nuclear transfer, 356*f*
  - somatic nucleus, HT24
  - somatostatin, release of, 309
  - sorbitol, 460
  - sorbose, 460
  - Sos protein, 418
  - SOS response, *E. coli*, 271
  - Southern blot, 381*f*, 382
  - Sox2 gene, HT25
  - Sp1* transcription factor, 311–312
  - spacer regions, 238
  - spermatogenesis, 309
  - sphingolipids, 193, 197–198, 613
  - sphingomyelin, 198, 202*f*, 613
  - sphingosine, 613
  - splice site, 315–316
  - spliceosome, 316
  - splicing, 314–316
  - split genes, 314–315
  - spontaneity, 431
  - spontaneous, 27, 140
  - sports, 57, 585, 710–711, HT28–HT31
  - squalene, 615, 616*f*–617*f*
  - squalene epoxide, 615
  - St. Louis encephalitis virus, 397*t*
  - standard conditions, 429
  - standard free energy change, 139, 488*t*
  - standard state for biochemical applications, 430–431
  - standard states, 429–430
  - standard states for free energy change, 429–430
  - Staphylococcus aureus*, 472, 473*f*
  - starch granules, 472*f*
  - starches, 468, 469*f*, 470–471
  - starch–iodine complex, 471*f*
  - start signal, 332
  - state function, 442
  - stationary phase, 118
  - steady state, 149, 433
  - steady-state theory, 149
  - stearic acid, 195*t*, 596*f*, 597
  - stem cells, 407*f*, HT23–HT27
  - stems, 245
  - step, 236
  - stereochemistry, 61, 62*f*, 455*f*
  - steroids, HT28
  - stereoisomers, 61, 454*f*
  - stereospecific enzymes, 184
  - stereochemistry, 451–459
  - steroid hormones, 618–619, 698*f*
  - steroids, 193, 198, 199*f*, 618, 697*t*
  - sticky ends, 358–360, HT13
  - storage, 102, 309
  - storage mechanisms, 507–531
  - (–) strand, 282
  - (+) strand, 282
  - strand invasion, 269
  - strokes, 200
  - stroma, 627, 628*f*

- Stromatolite fossils, 25*f*  
 structural gene, 289  
 submarine gel, 356  
 substrate, 144–145, 155*t*, 172  
 substrate concentration, [S], 148–154  
 substrate cycling, 520  
 substrate-level phosphorylation, 496  
 subunits, 83, 100, 172, 282–283  
 $\sigma$  subunits, 289  
 succinate, 544, 550  
 succinate- CoQ oxidoreductase, 568–569  
 succinate dehydrogenase, 545  
 succinic acid, 46*t*  
 succinyl-CoA synthetase, 542, 544  
 sucralose, 467  
 sucrose, 464–465  
 sugar– phosphate backbone, 229  
 sugar phosphates, 518  
 sugar structures, 458*f*  
 sugars, 38*t*, 451–459  
*Sulfolobus islandicus* virus 2 (SIRV2), HT14  
 sulfonamides, 671  
 sulfur, 6*t*, 35*t*, 665*f*  
 Supercoiled DNA topology, 237*f*  
 supercoils, 237–238, 261  
 supernovas, 6  
 supersecondary structure, 83, 88, 90*f*  
 suppressor tRNA, 345  
 survival of the fittest, 277  
 Svedberg units (S), 245  
 swine flu, 400–401  
 symbiosis, 24–25  
 synthetic genome, 249  
 synthetic mRNA, 325  
 synvinolin, 615*f*
- T**  
 T7 polymerase, 371  
 TAMLs, 190  
 Tandem Mass Spectrometry, 134  
*Taq* polymerase, 377  
 target issues, 698*f*  
 TATA box, 298–299, 300*f*, 301  
 TATA-binding protein (TBP), 299  
 TBP-associated factors (TAFIIIs), 299  
 TCA cycle, 536*f*  
 T-cell functions, 406–410  
 T-cell memory, 411  
 T-cell receptors (TCRs), 406  
 T-cells, 406, 407*f*, 414  
 N-TEF negative-transcription elongation factor, 301  
 telomerase, 274–275  
 telomere replication, 274  
 temperature, 142  
 template mechanism, role in synthesis of polynucleotides, 11*f*  
 template strand, 282  
 N-terminal amino acid residue, 74  
 terminal sequences, 312  
 termination, 301  
 termination sites, 287  
 3·4 terminator, 294–295  
 termites, 470*ph*  
 terpenes, 614  
 tertiary structure, 96*f*, 180*f*, 227, 237, 331*f*  
 tertiary structure of proteins, 83, 93–100  
 testosterone, 619  
 tetrahedral hydrogen bonding, 41*f*  
 tetrahedral intermediate in the HMG-CoA reductase mechanism, 615*f*  
 tetrahydrofolate (THF), 662, 663*f*–664*f*, 682  
 tetrahydrofolic acid, 187*t*  
*Tetrahymena*, 282  
 tetramer, 100  
 tetrapeptide side chain, 473*f*, 474  
 TFIIA, TFIIB, TFIID, TFIIIE, TFIIF, TFIIH, 299, 300*f*, 301  
 TFIIH, 303  
 TFIS, 301  
 thalassemia, 111  
*Thermacidophiles*, 23  
 thermodynamics, 26, 28–30, 107, 139–142, 431, 433  
*Thermus aquaticus* (Taq polymerase), 23, 377  
 thiamine (vitamin B1), 502  
 thiamine pyrophosphate (TPP), 187*t*, 501–502  
 thioester, 537  
 thiols, 4*t*  
 thioredoxin, 681  
 thioredoxin reductase, 681  
 -35 element, 284  
 -35 region, 284  
 three carbon atoms, 481  
 three-carbon glyceraldehyde-3-phosphate, 485  
 three-dimensional structure, 87, 95, 246, 347  
 three-domain classification scheme, 23–24  
 threonine, 65*t*  
 thromboxane A<sub>2</sub>, 221  
 thylakoid disks, 627–628  
 thylakoid space, 627  
 thymidylate synthase, 682  
 thymine, 227, 267  
 thymine bases, 265*f*  
 thyrotropin (TSH), 697*t*  
 thyrotropin-releasing factor (TRF), 697*t*  
 thyroxine, 66*f*, 697*t*  
 tissue plasminogen activator (TPA), 373  
 titration, 47–48  
 titration curve, 47–49, 53*f*, 67, 70–71  
 tobacco leaf cell, 18*f*  
 tobacco mosaic virus, 396*f*  
 $\alpha$ -tocopherol, 218  
 togaviruses, 397*t*  
 Tollens reagent, 459  
 tomatoes, 369  
 topoisomerases, 237  
 torr, 102  
 toxic cycle, 306*f*  
 toxicology, 190, 543  
 trans fatty acids, 193–194, 204  
 transaldolase, 526  
 transaminase, 78  
 transamination reaction, 188*f*, 656, 659–661  
 transcription  
   abbreviations, 305*t*  
   activation, 305*f*  
   activators, 602  
   attenuation, 294  
   elongation, 286*f*  
   in eukaryotes, 296–299, 301  
   factors, 289, 299  
   of the genetic code, **281–322**  
   information transfer in cells, 243*f*–244*f*, 253  
   of metabolic enzymes, 309  
   mRNA, 314  
   order of events of, 300*f*  
   in prokaryotes, 282–285  
   prokaryotes, regulation in, 287  
   proofreading in, 319  
   regulation, 302  
   termination, 288*f*, 295  
 transcription start site (TSS), 284  
 transcription-activation domain, 309, 311–312  
 transcriptional targeting, 423, 425*f*  
 transcription-coupled repair (TCR), 303  
 transcriptome, 388  
 transductional targeting, 423  
 transfer of a phosphate group, 484  
 transformed bacteria, 362  
 transforming growth factor  $\alpha$  (TGF $\alpha$ ), 351  
 transgenic tomato plant, 372*f*  
 transition state, 140, 184–187  
 transition-state analogs, 185  
 transketolase, 526, 645  
 translation, 242, 243*f*–244*f*, 253, **323–354**  
 translation kinetics, 348  
 translation prevention, 295  
 translocation, 334  
 transport, 102, 209–211  
 transport proteins, 207  
 triacylglycerols, 193–196, 592*f*–593*f*, 609, 610*f*, 709*t*, HT11  
 tricarboxylic acid cycle (or TCA cycle), 533, 536*f*  
 triglyceride, 194  
 trimer, 100  
 trimming, 312  
 triosephosphate isomerase, 487*f*, 491  
 trioses, 452*f*  
 triple helix, 92  
 triple reassortant, 401  
 triplet code, 324  
 tRNA (transfer RNA), 244*f*–245*f*, 313  
 tRNA molecules, 313*f*, 326, 327*f*, 328–329, 330*f*, 331  
 trophic hormones, 696–697  
 tropical plants, 646  
 tropocollagen, 92  
 troponin T gene, 317*f*

*trp* operator, 294  
*trp* operon, 290–292, 293*f*, 294, 296*f*  
*trp* promoter, 294  
 trypsin, 47*f*, 129, 130*f*, 185  
 tryptophan, 65*t*, 68–69  
 tumor cells, 416  
 tumor suppressor, 417, 419–420  
 turnover number, 153, 154*t*, 157  
 twins, 248  
 2,3-bisphosphoglycerate (BPG), 104–105  
 2,4-dienoyl-CoA reductase, 599  
 2,4-dinitrophenol, 576  
 two-dimensional electrophoresis, 127*f*  
 2'-*O*-methylribosyl group, 314*f*  
 two-pass sequencing, HT18–HT19  
 Type 2 diabetes, HT11  
 tyrocidine A, 78*f*  
 tyrosine, 65*t*, 66*f*, 68–69, 704, HT15  
 tyrosyl tRNA synthetase, 173

## U

ubiquinone, 637*f*  
 ubiquitin ligase (UL), 351  
 ubiquitin-activating enzyme (E1), 347  
 ubiquitin-carrier protein (E2), 347  
 ubiquitin-protein ligase (E3), 347  
 ubiquitinylation, 347  
 UDP-glucose pyrophosphorylase, 510  
 UMP, 679*f*  
 uncharged tRNA, 334  
 uncompetitive inhibition, 159  
 uncouplers, 576  
 uninducible, 290–292  
 $\alpha\alpha$  unit, 88, 90*f*  
 $\beta\alpha\beta$  unit, 88, 90*f*, 92*f*  
 unit activity, 429  
 units per second, 153  
 universal code, 324, 328–329  
 unsaturated fatty acid, 193–194, 195*t*,  
     598–600  
 UP element, 284  
 upstream binding site, 283  
 upstream elements, 298  
 uracil, 227, 267  
 urea, 100, 667  
 urea cycle, 667, 668*f*, 670  
 uric acid, 667–668, 674

uridine diphosphate glucose (UDPG), 510  
 UTP, 679*f*–680*f*  
 UV irradiation, 265*f*

## V

V domain, 411  
 V systems, 168–169  
 vaccine, 404, HT20–HT22  
 vacuole, 17*f*–18*f*, 21  
 valine, 65*t*  
 valinomycin, 576  
 van der Waals bond, 38  
 van der Waals interaction, 38  
 vanillin, 466  
 variable region, 411  
 varicella-zoster virus (herpes zoster virus),  
     397*t*  
 variola major virus, 397*t*  
 vascular endothelial growth factor (VEGF),  
     351  
 vasopressin, 77, 79, 697*t*  
 vector, 362, 363*f*, 364, 370–371  
 velocity curves, 147  
 vesicles, 578  
 Vinokourov, Alexandre, HT30  
 viral RNA, 415  
 viral sarcoma, 418  
 viral vector, 363*f*  
 virion, 395  
 virotherapy, 422–423, 425*f*  
 virus, 361  
 virus particle, 396*f*  
 viruses, **395–425**. *See also* cancer; disease  
     acquired immunodeficiency syndrome  
       (AIDS), 160, 395, 402  
     cancer and, 421  
     cell attachment, 399  
     diseases, 397*t*  
     families of, 395–396  
     flu, 413  
     human papillomavirus (HPV),  
       HT20–HT22  
     life cycles of, 396–397, 398*f*, 399–401  
     particles, 399*f*  
     vertebrate, 397*t*  
 vision, 217  
 vitamin A, 214, 215*f*

vitamin B6, 188*f*  
 vitamin C, 461  
 vitamin D, 214, 216*f*, 218  
 vitamin E, 214*t*, 218  
 vitamin K, 214*t*, 218–219, 220*f*  
 vitamin K epoxide reductase (VKOR), 385  
 vitamins, 690–691  
*in vitro*, 55–56, 277  
*in vivo*, 55, 403  
 Vmax, 151–153

## W

water, **35–60**, 668  
 water ligand, 190  
 wavelength, 630  
 waxes, 197, 198*f*  
 weight loss, 558  
 wobble, 326, 328  
 Wöhler, Friedrich, 3  
 world wide web, 107, 662

## X

xeroderma pigmentosum, 268  
 X-ray crystallography, 95, 331*f*  
 xylitol, 460  
 xylulose, 460  
 xylulose-5-phosphate epimerase, 645

## Y

Y chromosome, HT6–HT7  
 yeast phenylalanine tRNA, 246*f*  
 yeast RNA polymerase II subunits, 297*t*, 298*f*  
 yeast TATA-binding protein (TBP), 301  
 yellow fever virus, 397*t*  
 YFG (your favorite gene), 362

## Z

Z scheme, 632, 633*f*, 634  
 zebrafish, HT4–HT5  
 zero order, 144  
 zero-order kinetics, 148  
 zinc, 6*t*  
 zinc fingers, 309, 311  
 zipper, 311  
 zwitterions, 55  
 zymogens, 175–177

The Standard Genetic Code					
First Position (5' End)	Second Position				Third Position (3' End)
	U	C	A	G	
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA Stop UAG Stop	UGU Cys UGC Cys UGA Stop UGG Trp	U C A G
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg	U C A G
A	AUU Ile AUC Ile AUA Ile AUG Met*	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg	U C A G
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly	U C A G

\* AUG forms part of the initiation signal as well as coding for internal methionine residues.

Names and Abbreviations of the Common Amino Acids		
Amino Acid	Three-Letter Abbreviation	One-Letter Abbreviation
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V