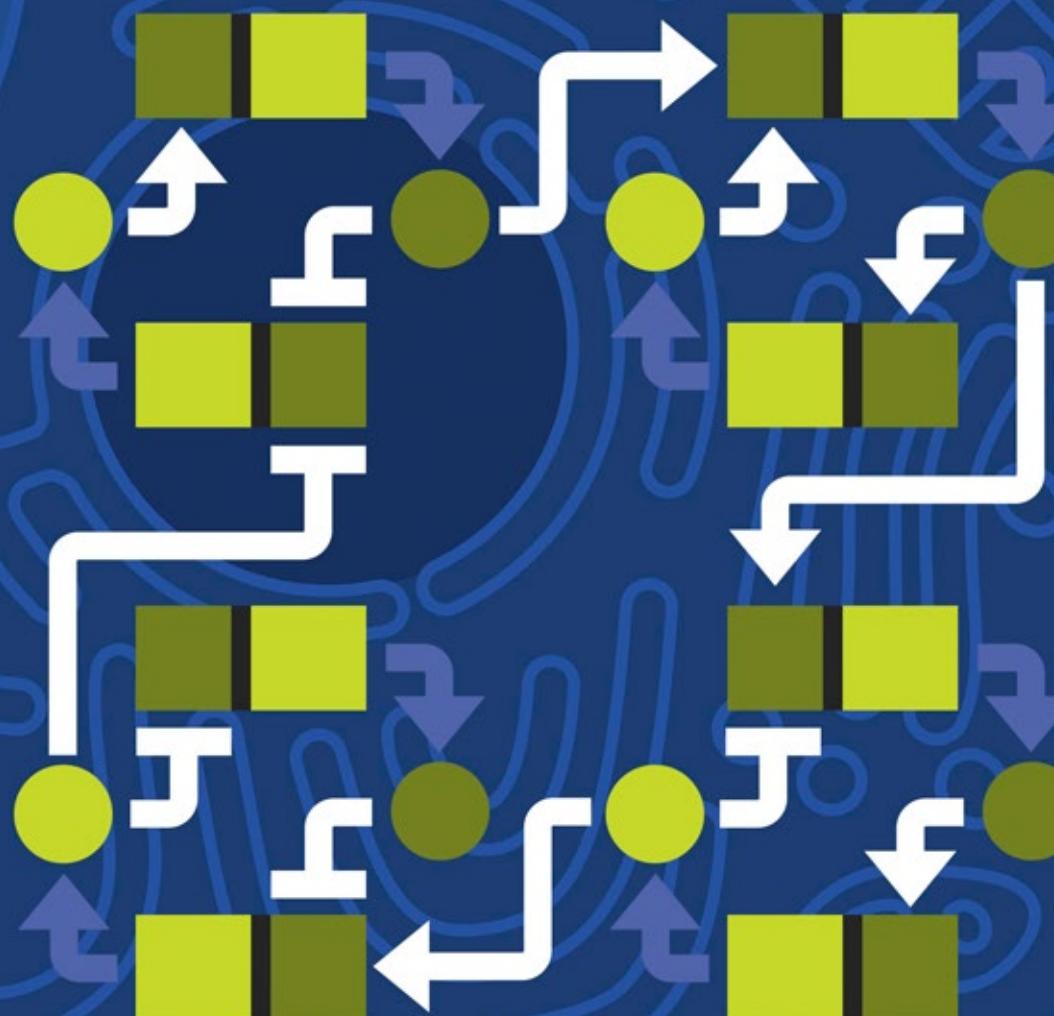


# Molecular Biology of **THE CELL**

Sixth Edition



ALBERTS

JOHNSON

LEWIS

MORGAN

RAFF

ROBERTS

WALTER

# Molecular Biology of **THE CELL**

Sixth Edition

This page intentionally left blank  
to match pagination of print book

# Molecular Biology of THE CELL

Sixth Edition

Bruce Alberts

Alexander Johnson

Julian Lewis

David Morgan

Martin Raff

Keith Roberts

Peter Walter

With problems by

John Wilson

Tim Hunt

## *Garland Science*

Vice President: Denise Schanck  
Associate Editor: Allie Bochicchio  
Production Editor and Layout: EJ Publishing Services  
Senior Production Editor: Georgina Lucas  
Text Editors: Sherry Granum Lewis and Elizabeth Zayatz  
Illustrator: Nigel Orme  
Structures: Tiago Barros  
Designer: Matthew McClements, Blink Studio, Ltd.  
Copyeditor: Jo Clayton  
Proofreader: Sally Huish  
Indexer: Bill Johncocks  
Permissions Coordinator: Sheri Gilbert  
Back Cover Photographer: Photography, Christophe Carlinet;  
Design, Nigel Orme

*Molecular Biology of the Cell* Interactive Media:  
Artistic and Scientific Direction: Peter Walter  
Narration: Julie Theriot  
Director of Digital Publishing: Michael Morales  
Editorial Assistant: Leah Christians  
Production Editor: Natasha Wolfe

© 2015 by Bruce Alberts, Alexander Johnson, Julian Lewis,  
David Morgan, Martin Raff, Keith Roberts, and Peter Walter.

This book contains information obtained from authentic and highly regarded sources. Every effort has been made to trace copyright holders and to obtain their permission for the use of copyright material. Reprinted material is quoted with permission, and sources are indicated. A wide variety of references are listed. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use.

All rights reserved. No part of this book covered by the copyright herein may be reproduced or used in any format in any form or by any means—graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems—with permission of the publisher.

## **Library of Congress Cataloging-in-Publication Data**

Alberts, Bruce, author.

*Molecular biology of the cell* / Bruce Alberts, Alexander Johnson, Julian Lewis, David Morgan, Martin Raff, Keith Roberts, Peter Walter ; with problems by John Wilson, Tim Hunt. -- Sixth edition.

p. ; cm.

Preceded by *Molecular biology of the cell* / Bruce Alberts ... [et al.]. 5th ed. c2008.

Includes bibliographical references and index.

ISBN 978-0-8153-4432-2 (hardcover) -- ISBN 978-0-8153-4464-3 (paperback)

I. Title.

[DNLM: 1. Cells. 2. Molecular Biology. QU 300]

QH581.2

572.8--dc23

2014031818

Published by Garland Science, Taylor & Francis Group, LLC, an informa business,  
711 Third Avenue, New York, NY 10017, US  
3 Park Square, Milton Park, Abingdon, OX14 4RN, UK

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

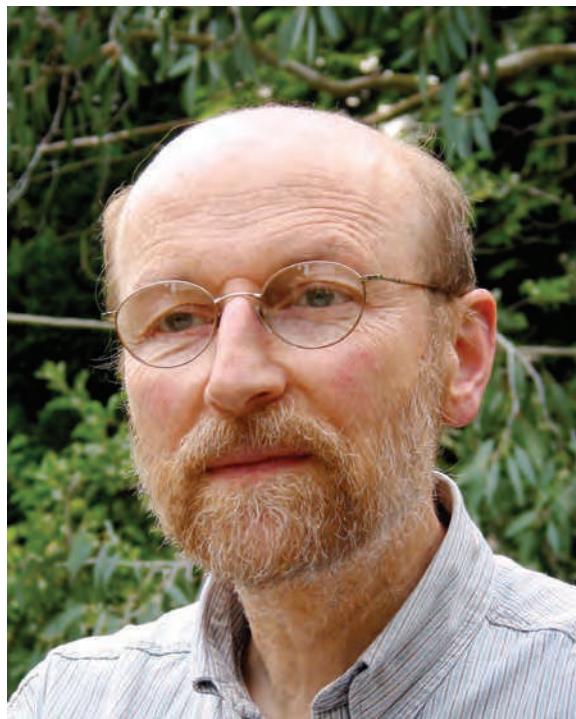


Visit our website at <http://www.garlandscience.com>

## *About the Authors*

**Bruce Alberts** received his PhD from Harvard University and is the Chancellor's Leadership Chair in Biochemistry and Biophysics for Science and Education, University of California, San Francisco. He was the editor-in-chief of *Science* magazine from 2008 until 2013, and for twelve years he served as President of the U.S. National Academy of Sciences (1993–2005). **Alexander Johnson** received his PhD from Harvard University and is Professor of Microbiology and Immunology at the University of California, San Francisco. **Julian Lewis** (1946–2014) received his DPhil from the University of Oxford and was an Emeritus Scientist at the London Research Institute of Cancer Research UK. **David Morgan** received his PhD from the University of California, San Francisco, and is Professor of the Department of Physiology there as well as the Director of the Biochemistry, Cell Biology, Genetics, and Developmental Biology Graduate Program. **Martin Raff** received his MD from McGill University and is Emeritus Professor of Biology at the Medical Research Council Laboratory for Molecular Cell Biology at University College London. **Keith Roberts** received his PhD from the University of Cambridge and was Deputy Director of the John Innes Centre, Norwich. He is Emeritus Professor at the University of East Anglia. **Peter Walter** received his PhD from the Rockefeller University in New York and is Professor of the Department of Biochemistry and Biophysics at the University of California, San Francisco, and an Investigator at the Howard Hughes Medical Institute. **John Wilson** received his PhD from the California Institute of Technology and pursued his postdoctoral work at Stanford University. He is Distinguished Service Professor of Biochemistry and Molecular Biology at Baylor College of Medicine in Houston. **Tim Hunt** received his PhD from the University of Cambridge where he taught biochemistry and cell biology for more than 20 years. He worked at Cancer Research UK until his retirement in 2010. He shared the 2001 Nobel Prize in Physiology or Medicine with Lee Hartwell and Paul Nurse.

**Cover design:** Cell biology is not only about the structure and function of the myriad molecules that comprise a cell, but also about how this complex chemistry is controlled. Understanding the cell's elaborate regulatory feedback networks will require quantitative approaches.



**Julian Hart Lewis**

August 12, 1946—April 30, 2014

This page intentionally left blank  
to match pagination of print book

# Preface

Since the last edition of this book appeared, more than five million scientific papers have been published. There has been a parallel increase in the quantity of digital information: new data on genome sequences, protein interactions, molecular structures, and gene expression—all stored in vast databases. The challenge, for both scientists and textbook writers, is to convert this overwhelming amount of information into an accessible and up-to-date understanding of how cells work.

Help comes from a large increase in the number of review articles that attempt to make raw material easier to digest, although the vast majority of these reviews are still quite narrowly focused. Meanwhile, a rapidly growing collection of online resources tries to convince us that understanding is only a few mouse-clicks away. In some areas this change in the way we access knowledge has been highly successful—in discovering the latest information about our own medical problems, for example. But to understand something of the beauty and complexity of how living cells work, one needs more than just a wiki- this or wiki- that; it is enormously hard to identify the valuable and enduring gems from so much confusing landfill. Much more effective is a carefully wrought narrative that leads logically and progressively through the key ideas, components, and experiments in such a way that readers can build for themselves a memorable, conceptual framework for cell biology—a framework that will allow them to critically evaluate all of the new science and, more importantly, to understand it. That is what we have tried to do in *Molecular Biology of the Cell*.

In preparing this new edition, we have inevitably had to make some difficult decisions. In order to incorporate exciting new discoveries, while at the same time keeping the book portable, much has had to be excised. We have added new sections, such as those on new RNA functions, advances in stem cell biology, new methods for studying proteins and genes and for imaging cells, advances in the genetics and treatment of cancer, and timing, growth control, and morphogenesis in development.

The chemistry of cells is extremely complex, and any list of cell parts and their interactions—no matter how complete—will leave huge gaps in our understanding. We now realize that to produce convincing explanations of cell behavior will require quantitative information about cells that is coupled to sophisticated mathematical/computational approaches—some not yet invented. As a consequence, an emerging goal for cell biologists is to shift their studies more toward quantitative description and mathematical deduction. We highlight this approach and some of its methods in a new section at the end of Chapter 8.

Faced with the immensity of what we have learned about cell biology, it might be tempting for a student to imagine that there is little left to discover. In fact, the more we find out about cells, the more new questions emerge. To emphasize that our understanding of cell biology is incomplete, we have highlighted some of the major gaps in our knowledge by including *What We Don't Know* at the end of each chapter. These brief lists include only a tiny sample of the critical unanswered questions and challenges for the next generation of scientists. We derive great pleasure from the knowledge that some of our readers will provide future answers.

The more than 1500 illustrations have been designed to create a parallel narrative, closely interwoven with the text. We have increased their consistency between chapters, particularly in the use of color and of common icons; membrane pumps and channels are a good example. To avoid interruptions to the text, some material has been moved into new, readily accessible panels. Most of the important protein structures depicted have now been redrawn and consistently colored. In each

case, we now provide the corresponding Protein Data Bank (PDB) code for the protein, which can be used to access online tools that provide more information about it, such as those on the RCSB PDB website ([www.rcsb.org](http://www.rcsb.org)). These connections allow readers of the book to explore more fully the proteins that lie at the core of cell biology.

John Wilson and Tim Hunt have again contributed their distinctive and imaginative problems to help students gain a more active understanding of the text. The problems emphasize quantitative approaches and encourage critical thinking about published experiments; they are now present at the end of all chapters. The answers to these problems, plus more than 1800 additional problems and solutions, all appear in the companion volume that John and Tim have written, *Molecular Biology of the Cell, Sixth Edition: The Problems Book*.

We live in a world that presents us with many complex issues related to cell biology: biodiversity, climate change, food security, environmental degradation, resource depletion, and human disease. We hope that our textbook will help the reader better understand and possibly contribute to meeting these challenges. Knowledge and understanding bring the power to intervene.

We are indebted to a large number of scientists whose generous help we mention separately in the detailed acknowledgments. Here we must mention some particularly significant contributors. For Chapter 8, Hana El-Samad provided the core of the section on Mathematical Analysis of Cell Functions, and Karen Hopkin made valuable contributions to the section on Studying Gene Expression and Function. Werner Kuhlbrandt helped to reorganize and rewrite Chapter 14 (Energy Conversion: Mitochondria and Chloroplasts). Rebecca Heald did the same for Chapter 16 (The Cytoskeleton), as did Alexander Schier for Chapter 21 (Development of Multicellular Organisms), and Matt Welch for Chapter 23 (Pathogens and Infection). Lewis Lanier aided in the writing of Chapter 24 (The Innate and Adaptive Immune Systems). Hossein Amiri generated the enormous online instructor's question bank.

Before starting out on the revision cycle for this edition, we asked a number of scientists who had used the last edition to teach cell biology students to meet with us and suggest improvements. They gave us useful feedback that has helped inform the new edition. We also benefited from the valuable input of groups of students who read most of the chapters in page proofs.

Many people and much effort are needed to convert a long manuscript and a large pile of sketches into a finished textbook. The team at Garland Science that managed this conversion was outstanding. Denise Schanck, directing operations, displayed forbearance, insight, tact, and energy throughout the journey; she guided us all unerringly, ably assisted by Allie Bochicchio and Janette Scobie. Nigel Orme oversaw our revamped illustration program, put all the artwork into its final form, and again enhanced the back cover with his graphics skills. Tiago Barros helped us refresh our presentation of protein structures. Matthew McClements designed the book and its front cover. Emma Jeffcock again laid out the final pages, managing endless rounds of proofs and last-minute changes with remarkable skill and patience; Georgina Lucas provided her with help. Michael Morales, assisted by Leah Christians, produced and assembled the complex web of videos, animations, and other materials that form the core of the online resources that accompany the book. Adam Sendroff provided us with the valuable feedback from book users around the world that informed our revision cycle. Casting expert eyes over the manuscript, Elizabeth Zayatz and Sherry Granum Lewis acted as development editors, Jo Clayton as copyeditor, and Sally Huish as proofreader. Bill Johncocks compiled the index. In London, Emily Preece fed us, while the Garland team's professional help, skills, and energy, together with their friendship, nourished us in every other way throughout the revision, making the whole process a pleasure. The authors are extremely fortunate to be supported so generously.

We thank our spouses, families, friends, and colleagues for their continuing support, which has once again made the writing of this book possible.

Just as we were completing this edition, Julian Lewis, our coauthor, friend, and colleague, finally succumbed to the cancer that he had fought so heroically for ten years. Starting in 1979, Julian made major contributions to all six editions, and, as our most elegant wordsmith, he elevated and enhanced both the style and tone of all the many chapters he touched. Noted for his careful scholarly approach, clarity and simplicity were at the core of his writing. Julian is irreplaceable, and we will all deeply miss his friendship and collaboration. We dedicate this Sixth Edition to his memory.

# Note to the Reader

## Structure of the Book

Although the chapters of this book can be read independently of one another, they are arranged in a logical sequence of five parts. The first three chapters of Part I cover elementary principles and basic biochemistry. They can serve either as an introduction for those who have not studied biochemistry or as a refresher course for those who have. Part II deals with the storage, expression, and transmission of genetic information. Part III presents the principles of the main experimental methods for investigating and analyzing cells; here, a new section entitled “Mathematical Analysis of Cell Functions” in Chapter 8 provides an extra dimension in our understanding of cell regulation and function. Part IV describes the internal organization of the cell. Part V follows the behavior of cells in multicellular systems, starting with development of multicellular organisms and concluding with chapters on pathogens and infection and on the innate and adaptive immune systems.

## End-of-Chapter Problems

A selection of problems, written by John Wilson and Tim Hunt, appears in the text at the end of each chapter. New to this edition are problems for the last four chapters on multicellular organisms. The complete solutions to all of these problems can be found in *Molecular Biology of the Cell, Sixth Edition: The Problems Book*.

## References

A concise list of selected references is included at the end of each chapter. These are arranged in alphabetical order under the main chapter section headings. These references sometimes include the original papers in which important discoveries were first reported.

## Glossary Terms

Throughout the book, boldface type has been used to highlight key terms at the point in a chapter where the main discussion occurs. Italic type is used to set off important terms with a lesser degree of emphasis. At the end of the book is an expanded glossary, covering technical terms that are part of the common currency of cell biology; it should be the first resort for a reader who encounters an unfamiliar term. The complete glossary as well as a set of flashcards is available on the Student Website.

## Nomenclature for Genes and Proteins

Each species has its own conventions for naming genes; the only common feature is that they are always set in italics. In some species (such as humans), gene names are spelled out all in capital letters; in other species (such as zebrafish), all in lowercase; in yet others (most mouse genes), with the first letter in uppercase and rest in lowercase; or (as in *Drosophila*) with different combinations of uppercase and lowercase, according to whether the first mutant allele to be discovered produced a dominant or recessive phenotype. Conventions for naming protein products are equally varied.

This typographical chaos drives everyone crazy. It is not just tiresome and absurd; it is also unsustainable. We cannot independently define a fresh convention for each of the next few million species whose genes we may wish to study.

Moreover, there are many occasions, especially in a book such as this, where we need to refer to a gene generically—without specifying the mouse version, the human version, the chick version, or the hippopotamus version—because they are all equivalent for the purposes of our discussion. What convention then should we use?

We have decided in this book to cast aside the different conventions that are used in individual species and follow a uniform rule: we write all gene names, like the names of people and places, with the first letter in uppercase and the rest in lowercase, but all in italics, thus: *Apc*, *Bazooka*, *Cdc2*, *Dishevelled*, *Egl1*. The corresponding protein, where it is named after the gene, will be written in the same way, but in roman rather than italic letters: Apc, Bazooka, Cdc2, Dishevelled, Egl1. When it is necessary to specify the organism, this can be done with a prefix to the gene name.

For completeness, we list a few further details of naming rules that we shall follow. In some instances, an added letter in the gene name is traditionally used to distinguish between genes that are related by function or evolution; for those genes, we put that letter in uppercase if it is usual to do so (*LacZ*, *RecA*, *HoxA4*). We use no hyphen to separate added letters or numbers from the rest of the name. Proteins are more of a problem. Many of them have names in their own right, assigned to them before the gene was named. Such protein names take many forms, although most of them traditionally begin with a lowercase letter (actin, hemoglobin, catalase), like the names of ordinary substances (cheese, nylon), unless they are acronyms (such as GFP, for Green Fluorescent Protein, or BMP4, for Bone Morphogenetic Protein #4). To force all such protein names into a uniform style would do too much violence to established usages, and we shall simply write them in the traditional way (actin, GFP, and so on). For the corresponding gene names in all these cases, we shall nevertheless follow our standard rule: *Actin*, *Hemoglobin*, *Catalase*, *Bmp4*, *Gfp*. Occasionally in our book we need to highlight a protein name by setting it in italics for emphasis; the intention will generally be clear from the context.

For those who wish to know them, the table below shows some of the official conventions for individual species—conventions that we shall mostly violate in this book, in the manner shown.

Organism	Species-Specific Convention		Unified Convention Used in This Book	
	Gene	Protein	Gene	Protein
Mouse	<i>Hoxa4</i>	Hoxa4	<i>HoxA4</i>	HoxA4
	<i>Bmp4</i>	BMP4	<i>Bmp4</i>	BMP4
	<i>integrin α-1</i> , <i>Itga1</i>	integrin α1	<i>Integrin α1</i> , <i>Itga1</i>	integrin α1
Human	<i>HOXA4</i>	HOXA4	<i>HoxA4</i>	HoxA4
Zebrafish	<i>cyclops</i> , <i>cyc</i>	Cyclops, Cyc	<i>Cyclops</i> , <i>Cyc</i>	Cyclops, Cyc
<i>Caenorhabditis</i>	<i>unc-6</i>	UNC-6	<i>Unc6</i>	Unc6
<i>Drosophila</i>	sevenless, sev (named after recessive phenotype)	Sevenless, SEV	Sevenless, Sev	Sevenless, Sev
	<i>Deformed</i> , <i>Dfd</i> (named after dominant mutant phenotype)	Deformed, DFD	<i>Deformed</i> , <i>Dfd</i>	Deformed, Dfd
Yeast				
<i>Saccharomyces cerevisiae</i> (budding yeast)	<i>CDC28</i>	Cdc28, Cdc28p	<i>Cdc28</i>	Cdc28
<i>Schizosaccharomyces pombe</i> (fission yeast)	<i>Cdc2</i>	Cdc2, Cdc2p	<i>Cdc2</i>	Cdc2
<i>Arabidopsis</i>	<i>GAI</i>	GAI	<i>Gai</i>	GAI
<i>E. coli</i>	<i>uvrA</i>	UvrA	<i>UvrA</i>	UvrA

***Molecular Biology of the Cell, Sixth Edition: The Problems Book***

by John Wilson and Tim Hunt (ISBN: 978-0-8153-4453-7)

*The Problems Book* is designed to help students appreciate the ways in which experiments and simple calculations can lead to an understanding of how cells work. It provides problems to accompany Chapters 1–20 of *Molecular Biology of the Cell*. Each chapter of problems is divided into sections that correspond to those of the main textbook and review key terms, test for understanding basic concepts, pose research-based problems, and now include MCAT-style questions which help students to prepare for standardized medical school admission tests. *Molecular Biology of the Cell, Sixth Edition: The Problems Book* should be useful for homework assignments and as a basis for class discussion. It could even provide ideas for exam questions. Solutions for all of the problems are provided in the book. Solutions for the end-of-chapter problems for Chapters 1–24 in the main textbook are also found in *The Problems Book*.

## RESOURCES FOR INSTRUCTORS AND STUDENTS

The teaching and learning resources for instructors and students are available online. The instructor's resources are password-protected and available only to adopting instructors. The student resources are available to everyone. We hope these resources will enhance student learning and make it easier for instructors to prepare dynamic lectures and activities for the classroom.

### Instructor Resources

Instructor Resources are available on the Garland Science Instructor's Resource Site, located at [www.garlandscience.com/instructors](http://www.garlandscience.com/instructors). The website provides access not only to the teaching resources for this book but also to all other Garland Science textbooks. Adopting instructors can obtain access to the site from their sales representative or by emailing science@garland.com.

#### *Art of Molecular Biology of the Cell, Sixth Edition*

The images from the book are available in two convenient formats: PowerPoint® and JPEG. They have been optimized for display on a computer. Figures are searchable by figure number, by figure name, or by keywords used in the figure legend from the book.

#### *Figure-Integrated Lecture Outlines*

The section headings, concept headings, and figures from the text have been integrated into PowerPoint presentations. These will be useful for instructors who would like a head start creating lectures for their course. Like all of our PowerPoint presentations, the lecture outlines can be customized. For example, the content of these presentations can be combined with videos and questions from the book or Question Bank, in order to create unique lectures that facilitate interactive learning.

#### *Animations and Videos*

The 174 animations and videos that are available to students are also available on the Instructor's Website in two formats. The WMV-formatted movies are created for instructors who wish to use the movies in PowerPoint presentations on Windows® computers; the QuickTime-formatted movies are for use in PowerPoint for Apple computers or Keynote® presentations. The movies can easily be downloaded using the "download" button on the movie preview page. The movies are correlated to each chapter and callouts are highlighted in color.

#### *Media Guide*

This document provides an overview to the multimedia available for students and instructors and contains the text of the voice-over narration for all of the movies.

#### *Question Bank*

Written by Hossein Amiri, University of California, Santa Cruz, this greatly expanded question bank includes a variety of question formats: multiple choice,

short answer, fill-in-the-blank, true-false, and matching. There are 35–60 questions per chapter, and a large number of the multiple-choice questions will be suitable for use with personal response systems (that is, clickers). The Question Bank was created with the philosophy that a good exam should do much more than simply test students' ability to memorize information; it should require them to reflect upon and integrate information as a part of a sound understanding. This resource provides a comprehensive sampling of questions that can be used either directly or as inspiration for instructors to write their own test questions.

#### *Diploma® Test Generator Software*

The questions from the Question Bank have been loaded into the Diploma Test Generator software. The software is easy to use and can scramble questions to create multiple tests. Questions are organized by chapter and type and can be additionally categorized by the instructor according to difficulty or subject. Existing questions can be edited and new ones added. The Test Generator is compatible with several course management systems, including Blackboard®.

#### *Medical Topics Guide*

This document highlights medically relevant topics covered throughout *Molecular Biology of the Cell* and *The Problems Book*. It will be particularly useful for instructors with a large number of premedical, health science, or nursing students.

#### *Blackboard and Learning Management System (LMS) Integration*

The movies, book images, and student assessments that accompany the book can be integrated into Blackboard or other LMSs. These resources are bundled into a “Common Cartridge” or “Upload Package” that facilitates bulk uploading of textbook resources into Blackboard and other LMSs. The LMS Common Cartridge can be obtained on a DVD from your sales representative or by emailing science@garland.com.

### **Resources for Students**

The resources for students are available on the *Molecular Biology of the Cell* Student Website, located at [www.garlandscience.com/MBOC6-students](http://www.garlandscience.com/MBOC6-students).

#### *Animations and Videos*

There are 174 movies, covering a wide range of cell biology topics, which review key concepts in the book and illuminate subcellular processes. The movies are correlated to each chapter and callouts are highlighted in color.

#### *Cell Explorer Slides*

This application teaches cell morphology through interactive micrographs that highlight important cellular structures.

#### *Flashcards*

Each chapter contains a set of flashcards, built into the website, that allow students to review key terms from the text.

#### *Glossary*

The complete glossary from the book is available on the website and can be searched and browsed.

# Acknowledgments

In writing this book we have benefited greatly from the advice of many biologists and biochemists. We would like to thank the following for their suggestions in preparing this edition, as well as those who helped in preparing the first, second, third, fourth, and fifth editions. (Those who helped on this edition are listed first, those who helped with the first, second, third, fourth, and fifth editions follow.)

## **General:**

Steven Cook (Imperial College London), Jose A. Costoya (Universidade de Santiago de Compostela), Arshad Desai (University of California, San Diego), Susan K. Dutcher (Washington University, St. Louis), Michael Elowitz (California Institute of Technology), Benjamin S. Glick (University of Chicago), Gregory Hannon (Cold Spring Harbor Laboratories), Rebecca Heald (University of California, Berkeley), Stefan Kanzok (Loyola University Chicago), Doug Kellogg (University of California, Santa Cruz), David Kimelman (University of Washington, Seattle), Maria Krasilnikova (Pennsylvania State University), Werner Kühlbrandt (Max Planck Institute of Biophysics), Lewis Lanier (University of California, San Francisco), Annette Müller-Taubenberger (Ludwig Maximilians University), Sandra Schmid (University of Texas Southwestern), Ronald D. Vale (University of California, San Francisco), D. Eric Walters (Chicago Medical School), Karsten Weis (Swiss Federal Institute of Technology)

## **Chapter 2:** H. Lill (VU University)

**Chapter 3:** David S. Eisenberg (University of California, Los Angeles), F. Ulrich Hartl (Max Planck Institute of Biochemistry), Louise Johnson (University of Oxford), H. Lill (VU University), Jonathan Weissman (University of California, San Francisco)

**Chapter 4:** Bradley E. Bernstein (Harvard Medical School), Wendy Bickmore (MRC Human Genetics Unit, Edinburgh), Jason Brickner (Northwestern University), Gary Felsenfeld (NIH), Susan M. Gasser (University of Basel), Shiv Grewal (National Cancer Institute), Gary Karpen (University of California, Berkeley), Eugene V. Koonin, (NCBI, NLM, NIH), Hiten Madhani (University of California, San Francisco), Tom Misteli (National Cancer Institute), Geeta Narlikar (University of California, San Francisco), Maynard Olson (University of Washington, Seattle), Stephen Scherer (University of Toronto), Rolf Sternglanz (Stony Brook University), Chris L. Woodcock (University of Massachusetts, Amherst), Johanna Wysocka and lab members (Stanford School of Medicine)

**Chapter 5:** Oscar Aparicio (University of Southern California), Julie P. Cooper (National Cancer Institute), Neil Hunter (Howard Hughes Medical Institute), Karim Labib (University of Manchester), Joachim Li (University of California, San Francisco), Stephen West (Cancer

Research UK), Richard D. Wood (University of Pittsburgh Cancer Institute)

**Chapter 6:** Briana Burton (Harvard University), Richard H. Ebright (Rutgers University), Daniel Finley (Harvard Medical School), Michael R. Green (University of Massachusetts Medical School), Christine Guthrie (University of California, San Francisco), Art Horwich (Yale School of Medicine), Harry Noller (University of California, Santa Cruz), David Tollervey (University of Edinburgh), Alexander J. Varshavsky (California Institute of Technology)

**Chapter 7:** Adrian Bird (The Wellcome Trust Centre, UK), Neil Brockdorff (University of Oxford), Christine Guthrie (University of California, San Francisco), Jeannie Lee (Harvard Medical School), Michael Levine (University of California, Berkeley), Hiten Madhani (University of California, San Francisco), Duncan Odom (Cancer Research UK), Kevin Struhl (Harvard Medical School), Jesper Svejstrup (Cancer Research UK)

**Chapter 8:** Hana El-Samad [major contribution] (University of California, San Francisco), Karen Hopkin [major contribution], Donita Brady (Duke University), David Kashatus (University of Virginia), Melanie McGill (University of Toronto), Alex Mogilner (University of California, Davis), Richard Morris (John Innes Centre, UK), Prasanth Potluri (The Children's Hospital of Philadelphia Research Institute), Danielle Vidaurre (University of Toronto), Carmen Warren (University of California, Los Angeles), Ian Woods (Ithaca College)

**Chapter 9:** Douglas J. Briant (University of Victoria), Werner Kühlbrandt (Max Planck Institute of Biophysics), Jeffrey Lichtman (Harvard University), Jennifer Lippincott-Schwartz (NIH), Albert Pan (Georgia Regents University), Peter Shaw (John Innes Centre, UK), Robert H. Singer (Albert Einstein School of Medicine), Kurt Thorn (University of California, San Francisco)

**Chapter 10:** Ari Helenius (Swiss Federal Institute of Technology), Werner Kühlbrandt (Max Planck Institute of Biophysics), H. Lill (VU University), Satyajit Mayor (National Centre for Biological Sciences, India), Kai Simons (Max Planck Institute of Molecular Cell Biology and Genetics), Gunnar von Heijne (Stockholm University), Tobias Walther (Harvard University)

**Chapter 11:** Graeme Davis (University of California, San Francisco), Robert Edwards (University of California, San

Francisco), Bertil Hille (University of Washington, Seattle), Lindsay Hinck (University of California, Santa Cruz), Werner Kühlbrandt (Max Planck Institute of Biophysics), H. Lill (VU University), Roger Nicoll (University of California, San Francisco), Poul Nissen (Aarhus University), Robert Stroud (University of California, San Francisco), Karel Svoboda (Howard Hughes Medical Institute), Robert Tampé (Goethe-University Frankfurt)

**Chapter 12:** John Aitchison (Institute for System Biology, Seattle), Amber English (University of Colorado at Boulder), Ralf Erdmann (Ruhr University of Bochum), Larry Gerace (The Scripps Research Institute, La Jolla), Ramanujan Hegde (MRC Laboratory of Molecular Biology, Cambridge, UK), Martin W. Hetzer (The Salk Institute), Lindsay Hinck (University of California, Santa Cruz), James A. McNew (Rice University), Nikolaus Pfanner (University of Freiberg), Peter Rehling (University of Göttingen), Michael Rout (The Rockefeller University), Danny J. Schnell (University of Massachusetts, Amherst), Sebastian Schuck (University of Heidelberg), Suresh Subramani (University of California, San Diego), Gia Voeltz (University of Colorado, Boulder), Susan R. Wente (Vanderbilt University School of Medicine)

**Chapter 13:** Douglas J. Briant (University of Victoria, Canada), Scott D. Emr (Cornell University), Susan Ferro-Novick (University of California, San Diego), Benjamin S. Glick (University of Chicago), Ari Helenius (Swiss Federal Institute of Technology), Lindsay Hinck (University of California, Santa Cruz), Reinhard Jahn (Max Planck Institute for Biophysical Chemistry), Ira Mellman (Genentech), Peter Novick (University of California, San Diego), Hugh Pelham (MRC Laboratory of Molecular Biology, Cambridge, UK), Graham Warren (Max F. Perutz Laboratories, Vienna), Marino Zerial (Max Planck Institute of Molecular Cell Biology and Genetics)

**Chapter 14:** Werner Kühlbrandt [major contribution] (Max Planck Institute of Biophysics), Thomas D. Fox (Cornell University), Cynthia Kenyon (University of California, San Francisco), Nils-Göran Larsson (Max Planck Institute for Biology of Aging), Jodi Nunnari (University of California, Davis), Patrick O'Farrell (University of California, San Francisco), Alastair Stewart (The Victor Chang Cardiac Research Institute, Australia), Daniela Stock (The Victor Chang Cardiac Research Institute, Australia), Michael P. Yaffe (California Institute for Regenerative Medicine)

**Chapter 15:** Henry R. Bourne (University of California, San Francisco), Dennis Bray (University of Cambridge), Douglas J. Briant (University of Victoria, Canada), James Briscoe (MRC National Institute for Medical Research, UK), James Ferrell (Stanford University), Matthew Freeman (MRC Laboratory of Molecular Biology, Cambridge, UK), Alan Hall (Memorial Sloan Kettering Cancer Center), Carl-Henrik Heldin (Uppsala University), James A. McNew (Rice University), Roel Nusse (Stanford University), Julie Pitcher (University College London)

**Chapter 16:** Rebecca Heald [major contribution] (University of California, Berkeley), Anna Akhmanova (Utrecht University), Arshad Desai (University of California, San Diego), Velia Fowler (The Scripps Research Institute, La Jolla), Vladimir Gelfand (Northwestern University), Robert Goldman (Northwestern University), Alan Rick Horwitz (University of Virginia), Wallace Marshall (University of California, San Francisco), J. Richard McIntosh

(University of Colorado, Boulder), Maxence Nachury (Stanford School of Medicine), Eva Nogales (University of California, Berkeley), Samara Reck-Peterson (Harvard Medical School), Ronald D. Vale (University of California, San Francisco), Richard B. Vallee (Columbia University), Michael Way (Cancer Research UK), Orion Weiner (University of California, San Francisco), Matthew Welch (University of California, Berkeley)

**Chapter 17:** Douglas J. Briant (University of Victoria, Canada), Lindsay Hinck (University of California, Santa Cruz), James A. McNew (Rice University)

**Chapter 18:** Emily D. Crawford (University of California, San Francisco), James A. McNew (Rice University), Shigekazu Nagata (Kyoto University), Jim Wells (University of California, San Francisco)

**Chapter 19:** Jeffrey Axelrod (Stanford University School of Medicine), John Couchman (University of Copenhagen), Johan de Rooij (The Hubrecht Institute, Utrecht), Benjamin Geiger (Weizmann Institute of Science, Israel), Andrew P. Gilmore (University of Manchester), Tony Harris (University of Toronto), Martin Humphries (University of Manchester), Andreas Prokop (University of Manchester), Charles Streuli (University of Manchester), Masatoshi Takeichi (RIKEN Center for Developmental Biology, Japan), Barry Thompson (Cancer Research UK), Kenneth M. Yamada (NIH), Alpha Yap (The University of Queensland, Australia)

**Chapter 20:** Anton Berns (Netherlands Cancer Institute), J. Michael Bishop (University of California, San Francisco), Trever Bivona (University of California, San Francisco), Fred Bunz (Johns Hopkins University), Paul Edwards (University of Cambridge), Ira Mellman (Genentech), Caetano Reis e Sousa (Cancer Research UK), Marc Shuman (University of California, San Francisco), Mike Stratton (Wellcome Trust Sanger Institute, UK), Ian Tomlinson (Cancer Research UK)

**Chapter 21:** Alex Schier [major contribution] (Harvard University), Markus Affolter (University of Basel), Victor Ambros (University of Massachusetts, Worcester), James Briscoe (MRC National Institute for Medical Research, UK), Donald Brown (Carnegie Institution for Science, Baltimore), Steven Burden (New York University School of Medicine), Moses Chao (New York University School of Medicine), Caroline Dean (John Innes Centre, UK), Chris Doe (University of Oregon, Eugene), Uwe Drescher (King's College London), Gordon Fishell (New York University School of Medicine), Brigid Hogan (Duke University), Phil Ingham (Institute of Molecular and Cell Biology, Singapore), Laura Johnston (Columbia University), David Kingsley (Stanford University), Tom Kornberg (University of California, San Francisco), Richard Mann (Columbia University), Andy McMahon (University of Southern California), Marek Mlodzik (Mount Sinai Hospital, New York), Patrick O'Farrell (University of California, San Francisco), Duoqia Pan (Johns Hopkins Medical School), Olivier Pourquie (Harvard Medical School), Erez Raz (University of Muenster), Chris Rushlow (New York University), Stephen Small (New York University), Marc Tessier-Lavigne (Rockefeller University)

**Chapter 22:** Simon Hughes (King's College London), Rudolf Jaenisch (Massachusetts Institute of Technology), Arnold Kriegstein (University of California, San Francisco), Doug Melton (Harvard University), Stuart Orkin (Harvard

University), Thomas A. Reh (University of Washington, Seattle), Amy Wagers (Harvard University), Fiona M. Watt (Wellcome Trust Centre for Stem Cell Research, UK), Douglas J. Winton (Cancer Research UK), Shinya Yamanaka (Kyoto University)

**Chapter 23:** Matthew Welch [major contribution] (University of California, Berkeley), Ari Helenius (Swiss Federal Institute of Technology), Dan Portnoy (University of California, Berkeley), David Sibley (Washington University, St. Louis), Michael Way (Cancer Research UK)

**Chapter 24:** Lewis Lanier (University of California, San Francisco).

**Readers:** Najla Arshad (Indian Institute of Science), Venice Chiueh (University of California, Berkeley), Quyen Huynh (University of Toronto), Rachel Kooistra (Loyola University, Chicago), Wes Lewis (University of Alabama), Eric Nam (University of Toronto), Vladimir Ryvkin (Stony Brook University), Laasya Samhita (Indian Institute of Science), John Senderak (Jefferson Medical College), Phillipa Simons (Imperial College, UK), Anna Constance Vind (University of Copenhagen), Steve Wellard (Pennsylvania State University), Evan Whitehead (University of California, Berkeley), Carrie Wilczewski (Loyola University, Chicago), Anna Wing (Pennsylvania State University), John Wright (University of Alabama)

**First, second, third, fourth, and fifth editions:**

Jerry Adams (The Walter and Eliza Hall Institute of Medical Research, Australia), Ralf Adams (London Research Institute), David Agard (University of California, San Francisco), Julie Ahringer (The Gurdon Institute, UK), Michael Akam (University of Cambridge), David Allis (The Rockefeller University), Wolfhard Almers (Oregon Health and Science University), Fred Alt (CBR Institute for Biomedical Research, Boston), Linda Amos (MRC Laboratory of Molecular Biology, Cambridge), Raul Andino (University of California, San Francisco), Clay Armstrong (University of Pennsylvania), Martha Arnaud (University of California, San Francisco), Spyros Artavanis-Tsakonas (Harvard Medical School), Michael Ashburner (University of Cambridge), Jonathan Ashmore (University College London), Laura Attardi (Stanford University), Tayna Awabdy (University of California, San Francisco), Jeffrey Axelrod (Stanford University Medical Center), Peter Baker (deceased), David Baldwin (Stanford University), Michael Banda (University of California, San Francisco), Cornelia Bargmann (The Rockefeller University), Ben Barres (Stanford University), David Bartel (Massachusetts Institute of Technology), Konrad Basler (University of Zurich), Wolfgang Baumeister (Max Planck Institute of Biochemistry), Michael Bennett (Albert Einstein College of Medicine), Darwin Berg (University of California, San Diego), Anton Berns (Netherlands Cancer Institute), Merton Bernfield (Harvard Medical School), Michael Berridge (The Babraham Institute, Cambridge, UK), Walter Birchmeier (Max Delbrück Center for Molecular Medicine, Germany), Adrian Bird (Wellcome Trust Centre, UK), David Birk (UMDNJ—Robert Wood Johnson Medical School), Michael Bishop (University of California, San Francisco), Elizabeth Blackburn (University of California, San Francisco), Tim Bliss (National Institute for Medical Research, London), Hans Bode (University of California, Irvine), Piet Borst (Jan Swammerdam Institute, University

of Amsterdam), Henry Bourne (University of California, San Francisco), Alan Boyde (University College London), Martin Brand (University of Cambridge), Carl Branden (deceased), Andre Brandli (Swiss Federal Institute of Technology, Zurich), Dennis Bray (University of Cambridge), Mark Bretscher (MRC Laboratory of Molecular Biology, Cambridge), James Briscoe (National Institute for Medical Research, UK), Marianne Bronner-Fraser (California Institute of Technology), Robert Brooks (King's College London), Barry Brown (King's College London), Michael Brown (University of Oxford), Michael Bulger (University of Rochester Medical Center), Fred Bunz (Johns Hopkins University), Steve Burden (New York University of Medicine), Max Burger (University of Basel), Stephen Burley (SGX Pharmaceuticals), Keith Burridge (University of North Carolina, Chapel Hill), John Cairns (Radcliffe Infirmary, Oxford), Patricia Calarco (University of California, San Francisco), Zacheus Cande (University of California, Berkeley), Lewis Cantley (Harvard Medical School), Charles Cantor (Columbia University), Roderick Capaldi (University of Oregon), Mario Capecchi (University of Utah), Michael Carey (University of California, Los Angeles), Adelaide Carpenter (University of California, San Diego), John Carroll (University College London), Tom Cavalier-Smith (King's College London), Pierre Chambon (University of Strasbourg), Hans Clevers (Hubrecht Institute, The Netherlands), Enrico Coen (John Innes Institute, Norwich, UK), Philip Cohen (University of Dundee, Scotland), Robert Cohen (University of California, San Francisco), Stephen Cohen (EMBL Heidelberg, Germany), Roger Cooke (University of California, San Francisco), John Cooper (Washington University School of Medicine, St. Louis), Michael Cox (University of Wisconsin, Madison), Nancy Craig (Johns Hopkins University), James Crow (University of Wisconsin, Madison), Stuart Cull-Candy (University College London), Leslie Dale (University College London), Caroline Damsky (University of California, San Francisco), Johann De Bono (The Institute of Cancer Research, UK), Anthony DeFranco (University of California, San Francisco), Abby Dernburg (University of California, Berkeley), Arshad Desai (University of California, San Diego), Michael Dexter (The Wellcome Trust, UK), John Dick (University of Toronto, Canada), Christopher Dobson (University of Cambridge), Russell Doolittle (University of California, San Diego), W. Ford Doolittle (Dalhousie University, Canada), Julian Downward (Cancer Research UK), Keith Dudley (King's College London), Graham Dunn (MRC Cell Biophysics Unit, London), Jim Dunwell (John Innes Institute, Norwich, UK), Bruce Edgar (Fred Hutchinson Cancer Research Center, Seattle), Paul Edwards (University of Cambridge), Robert Edwards (University of California, San Francisco), David Eisenberg (University of California, Los Angeles), Sarah Elgin (Washington University, St. Louis), Ruth Ellman (Institute of Cancer Research, Sutton, UK), Beverly Emerson (The Salk Institute), Charles Emerson (University of Virginia), Scott D. Emr (Cornell University), Sharyn Endow (Duke University), Lynn Enquist (Princeton University), Tariq Enver (Institute of Cancer Research, London), David Epel (Stanford University), Gerard Evan (University of California, Comprehensive Cancer Center), Ray Evert (University of Wisconsin, Madison), Matthias Falk (Lehigh University), Stanley Falkow (Stanford

University), Douglas Fearon (University of Cambridge), Gary Felsenfeld (NIH), Stuart Ferguson (University of Oxford), James Ferrell (Stanford University), Christine Field (Harvard Medical School), Daniel Finley (Harvard University), Gary Firestone (University of California, Berkeley), Gerald Fischbach (Columbia University), Robert Fletterick (University of California, San Francisco), Harvey Florman (Tufts University), Judah Folkman (Harvard Medical School), Larry Fowke (University of Saskatchewan, Canada), Jennifer Frazier (Exploratorium®, San Francisco), Matthew Freeman (Laboratory of Molecular Biology, UK), Daniel Friend (University of California, San Francisco), Elaine Fuchs (University of Chicago), Joseph Gall (Carnegie Institution of Washington), Richard Gardner (University of Oxford), Anthony Gardner-Medwin (University College London), Peter Garland (Institute of Cancer Research, London), David Garrod (University of Manchester, UK), Susan M. Gasser (University of Basel), Walter Gehring (Biozentrum, University of Basel), Benny Geiger (Weizmann Institute of Science, Rehovot, Israel), Larry Gerace (The Scripps Research Institute), Holger Gerhardt (London Research Institute), John Gerhart (University of California, Berkeley), Günther Gerisch (Max Planck Institute of Biochemistry), Frank Gertler (Massachusetts Institute of Technology), Sankar Ghosh (Yale University School of Medicine), Alfred Gilman (The University of Texas Southwestern Medical Center), Reid Gilmore (University of Massachusetts, Amherst), Bernie Gilula (deceased), Charles Gilvarg (Princeton University), Benjamin S. Glick (University of Chicago), Michael Glotzer (University of Chicago), Larry Goldstein (University of California, San Diego), Bastien Gomperts (University College Hospital Medical School, London), Daniel Goodenough (Harvard Medical School), Jim Goodrich (University of Colorado, Boulder), Jeffrey Gordon (Washington University, St. Louis), Peter Gould (Middlesex Hospital Medical School, London), Alan Grafen (University of Oxford), Walter Gratzer (King's College London), Michael Gray (Dalhousie University), Douglas Green (St. Jude Children's Hospital), Howard Green (Harvard University), Michael Green (University of Massachusetts, Amherst), Leslie Grivell (University of Amsterdam), Carol Gross (University of California, San Francisco), Frank Grosveld (Erasmus Universiteit, The Netherlands), Michael Grunstein (University of California, Los Angeles), Barry Gumbiner (Memorial Sloan Kettering Cancer Center), Brian Gunning (Australian National University, Canberra), Christine Guthrie (University of California, San Francisco), James Haber (Brandeis University), Ernst Hafen (Universitat Zurich), David Haig (Harvard University), Andrew Halestrap (University of Bristol, UK), Alan Hall (Memorial Sloan Kettering Cancer Center), Jeffrey Hall (Brandeis University), John Hall (University of Southampton, UK), Zach Hall (University of California, San Francisco), Douglas Hanahan (University of California, San Francisco), David Hanke (University of Cambridge), Nicholas Harberd (University of Oxford), Graham Hardie (University of Dundee, Scotland), Richard Harland (University of California, Berkeley), Adrian Harris (Cancer Research UK), John Harris (University of Otago, New Zealand), Stephen Harrison (Harvard University), Leland Hartwell (University of Washington, Seattle), Adrian Harwood (MRC Laboratory for Molecular Cell Biology and Cell Biology Unit, London),

Scott Hawley (Stowers Institute for Medical Research, Kansas City), Rebecca Heald (University of California, Berkeley), John Heath (University of Birmingham, UK), Ramanujan Hegde (NIH), Carl-Henrik Heldin (Uppsala University), Ari Helenius (Swiss Federal Institute of Technology), Richard Henderson (MRC Laboratory of Molecular Biology, Cambridge, UK), Glenn Herrick (University of Utah), Ira Herskowitz (deceased), Bertil Hille (University of Washington, Seattle), Alan Hinnebusch (NIH, Bethesda), Brigid Hogan (Duke University), Nancy Hollingsworth (State University of New York, Stony Brook), Frank Holstege (University Medical Center, The Netherlands), Leroy Hood (Institute for Systems Biology, Seattle), John Hopfield (Princeton University), Robert Horvitz (Massachusetts Institute of Technology), Art Horwich (Yale University School of Medicine), David Housman (Massachusetts Institute of Technology), Joe Howard (Max Planck Institute of Molecular Cell Biology and Genetics), Jonathan Howard (University of Washington, Seattle), James Hudspeth (The Rockefeller University), Simon Hughes (King's College London), Martin Humphries (University of Manchester, UK), Tim Hunt (Cancer Research UK), Neil Hunter (University of California, Davis), Laurence Hurst (University of Bath, UK), Jeremy Hyams (University College London), Tony Hyman (Max Planck Institute of Molecular Cell Biology and Genetics), Richard Hynes (Massachusetts Institute of Technology), Philip Ingham (University of Sheffield, UK), Kenneth Irvine (Rutgers University), Robin Irvine (University of Cambridge), Norman Iscove (Ontario Cancer Institute, Toronto), David Ish-Horowicz (Cancer Research UK), Lily Jan (University of California, San Francisco), Charles Janeway (deceased), Tom Jessell (Columbia University), Arthur Johnson (Texas A&M University), Louise Johnson (deceased), Andy Johnston (John Innes Institute, Norwich, UK), E.G. Jordan (Queen Elizabeth College, London), Ron Kaback (University of California, Los Angeles), Michael Karin (University of California, San Diego), Eric Karsenti (European Molecular Biology Laboratory, Germany), Ken Keegstra (Michigan State University), Ray Keller (University of California, Berkeley), Douglas Kellogg (University of California, Santa Cruz), Regis Kelly (University of California, San Francisco), John Kendrick-Jones (MRC Laboratory of Molecular Biology, Cambridge), Cynthia Kenyon (University of California, San Francisco), Roger Keynes (University of Cambridge), Judith Kimble (University of Wisconsin, Madison), Robert Kingston (Massachusetts General Hospital), Marc Kirschner (Harvard University), Richard Klausner (NIH), Nancy Kleckner (Harvard University), Mike Klymkowsky (University of Colorado, Boulder), Kelly Komachi (University of California, San Francisco), Eugene Koonin (NIH), Juan Korenblot (University of California, San Francisco), Roger Kornberg (Stanford University), Tom Kornberg (University of California, San Francisco), Stuart Kornfeld (Washington University, St. Louis), Daniel Koshland (University of California, Berkeley), Douglas Koshland (Carnegie Institution of Washington, Baltimore), Marilyn Kozak (University of Pittsburgh), Mark Krasnow (Stanford University), Werner Kühlbrandt (Max Planck Institute for Biophysics), John Kuriyan (University of California, Berkeley), Robert Kypta (MRC Laboratory for Molecular Cell Biology, London), Peter Lachmann

(MRC Centre, Cambridge), Ulrich Laemmli (University of Geneva, Switzerland), Trevor Lamb (University of Cambridge), Hartmut Land (Cancer Research UK), David Lane (University of Dundee, Scotland), Jane Langdale (University of Oxford), Lewis Lanier (University of California, San Francisco), Jay Lash (University of Pennsylvania), Peter Lawrence (MRC Laboratory of Molecular Biology, Cambridge), Paul Lazarow (Mount Sinai School of Medicine), Robert J. Lefkowitz (Duke University), Michael Levine (University of California, Berkeley), Warren Levinson (University of California, San Francisco), Alex Levitzki (Hebrew University, Israel), Ottoline Leyser (University of York, UK), Joachim Li (University of California, San Francisco), Tomas Lindahl (Cancer Research UK), Vishu Lingappa (University of California, San Francisco), Jennifer Lippincott-Schwartz (NIH), Joseph Lipsick (Stanford University School of Medicine), Dan Littman (New York University School of Medicine), Clive Lloyd (John Innes Institute, Norwich, UK), Richard Locksley (University of California, San Francisco), Richard Losick (Harvard University), Daniel Louvard (Institut Curie, France), Robin Lovell-Badge (National Institute for Medical Research, London), Scott Lowe (Cold Spring Harbor Laboratory), Shirley Lowe (University of California, San Francisco), Reinhard Lührmann (Max Planck Institute of Biophysical Chemistry), Michael Lynch (Indiana University), Laura Machesky (University of Birmingham, UK), Hiten Madhani (University of California, San Francisco), James Maller (University of Colorado Medical School), Tom Maniatis (Harvard University), Colin Manoil (Harvard Medical School), Elliott Margulies (NIH), Philippa Marrack (National Jewish Medical and Research Center, Denver), Mark Marsh (Institute of Cancer Research, London), Wallace Marshall (University of California, San Francisco), Gail Martin (University of California, San Francisco), Paul Martin (University College London), Joan Massagué (Memorial Sloan Kettering Cancer Center), Christopher Mathews (Oregon State University), Brian McCarthy (University of California, Irvine), Richard McCarty (Cornell University), William McGinnis (University of California, San Diego), Anne McLaren (Wellcome/Cancer Research Campaign Institute, Cambridge), Frank McNally (University of California, Davis), Freiderick Meins (Freiderick Miescher Institut, Basel), Stephanie Mel (University of California, San Diego), Ira Mellman (Genentech), Barbara Meyer (University of California, Berkeley), Elliot Meyerowitz (California Institute of Technology), Chris Miller (Brandeis University), Robert Mishell (University of Birmingham, UK), Avrion Mitchison (University College London), N.A. Mitchison (University College London), Timothy Mitchison (Harvard Medical School), Quinn Mitrovich (University of California, San Francisco), Peter Mombaerts (The Rockefeller University), Mark Mooseker (Yale University), David Morgan (University of California, San Francisco), Michelle Moritz (University of California, San Francisco), Montrose Moses (Duke University), Keith Mostov (University of California, San Francisco), Anne Mudge (University College London), Hans Müller-Eberhard (Scripps Clinic and Research Institute), Alan Munro (University of Cambridge), J. Murdoch Mitchison (Harvard University), Richard Myers (Stanford University), Diana Myles (University of California, Davis), Andrew Murray (Harvard University), Shigekazu

Nagata (Kyoto University, Japan), Geeta Narlikar (University of California, San Francisco), Kim Nasmyth (University of Oxford), Mark E. Nelson (University of Illinois, Urbana-Champaign), Michael Neuberger (deceased), Walter Neupert (University of Munich, Germany), David Nicholls (University of Dundee, Scotland), Roger Nicoll (University of California, San Francisco), Suzanne Noble (University of California, San Francisco), Harry Noller (University of California, Santa Cruz), Jodi Nunnari (University of California, Davis), Paul Nurse (Francis Crick Institute), Roel Nusse (Stanford University), Michael Nussenzweig (Rockefeller University), Duncan O'Dell (deceased), Patrick O'Farrell (University of California, San Francisco), Bjorn Olsen (Harvard Medical School), Maynard Olson (University of Washington, Seattle), Stuart Orkin (Harvard University), Terry Orr-Weaver (Massachusetts Institute of Technology), Erin O'Shea (Harvard University), Dieter Osterhelt (Max Planck Institute of Biochemistry), William Otto (Cancer Research UK), John Owen (University of Birmingham, UK), Dale Oxender (University of Michigan), George Palade (deceased), Barbara Panning (University of California, San Francisco), Roy Parker (University of Arizona, Tucson), William W. Parson (University of Washington, Seattle), Terence Partridge (MRC Clinical Sciences Centre, London), William E. Paul (NIH), Tony Pawson (deceased), Hugh Pelham (MRC, UK), Robert Perry (Institute of Cancer Research, Philadelphia), Gordon Peters (Cancer Research UK), Greg Petsko (Brandeis University), Nikolaus Pfanner (University of Freiburg, Germany), David Phillips (The Rockefeller University), Jeremy Pickett-Heaps (The University of Melbourne, Australia), Jonathan Pines (Gurdon Institute, Cambridge), Julie Pitcher (University College London), Jeffrey Pollard (Albert Einstein College of Medicine), Tom Pollard (Yale University), Bruce Ponder (University of Cambridge), Daniel Portnoy (University of California, Berkeley), James Priess (University of Washington, Seattle), Darwin Prockop (Tulane University), Mark Ptashne (Memorial Sloan Kettering Cancer Center), Dale Purves (Duke University), Efraim Racker (Cornell University), Jordan Raff (University of Oxford), Klaus Rajewsky (Max Delbrück Center for Molecular Medicine, Germany), George Ratcliffe (University of Oxford), Elio Raviola (Harvard Medical School), Martin Rechsteiner (University of Utah, Salt Lake City), David Rees (National Institute for Medical Research, London), Thomas A. Reh (University of Washington, Seattle), Louis Reichardt (University of California, San Francisco), Renee Reijo (University of California, San Francisco), Caetano Reis e Sousa (Cancer Research UK), Fred Richards (Yale University), Conly Rieder (Wadsworth Center, Albany), Phillips Robbins (Massachusetts Institute of Technology), Elizabeth Robertson (The Wellcome Trust Centre for Human Genetics, UK), Elaine Robson (University of Reading, UK), Robert Roeder (The Rockefeller University), Joel Rosenbaum (Yale University), Janet Rossant (Mount Sinai Hospital, Toronto), Jesse Roth (NIH), Jim Rothman (Memorial Sloan Kettering Cancer Center), Rodney Rothstein (Columbia University), Erkki Ruoslahti (La Jolla Cancer Research Foundation), Gary Ruvkun (Massachusetts General Hospital), David Sabatini (New York University), Alan Sachs (University of California, Berkeley), Edward Salmon (University of North Carolina),

Chapel Hill), Aziz Sancar (University of North Carolina, Chapel Hill), Joshua Sanes (Harvard University), Peter Sarnow (Stanford University), Lisa Satterwhite (Duke University Medical School), Robert Sauer (Massachusetts Institute of Technology), Ken Sawin (The Wellcome Trust Centre for Cell Biology, UK), Howard Schachman (University of California, Berkeley), Gerald Schatten (Pittsburgh Development Center), Gottfried Schatz (Biozentrum, University of Basel), Randy Schekman (University of California, Berkeley), Richard Scheller (Stanford University), Giampietro Schiavo (Cancer Research UK), Ueli Schibler (University of Geneva, Switzerland), Joseph Schlessinger (New York University Medical Center), Danny J. Schnell (University of Massachusetts, Amherst), Michael Schramm (Hebrew University, Israel), Robert Schreiber (Washington University School of Medicine), James Schwartz (Columbia University), Ronald Schwartz (NIH), François Schweiguth (Institut Pasteur, France), John Scott (University of Manchester, UK), John Sedat (University of California, San Francisco), Peter Selby (Cancer Research UK), Zvi Sellinger (Hebrew University, Israel), Gregg Semenza (Johns Hopkins University), Philippe Sengel (University of Grenoble, France), Peter Shaw (John Innes Institute, Norwich, UK), Michael Sheetz (Columbia University), Morgan Sheng (Massachusetts Institute of Technology), Charles Sherr (St. Jude Children's Hospital), David Shima (Cancer Research UK), Samuel Silverstein (Columbia University), Melvin I. Simon (California Institute of Technology), Kai Simons (Max Planck Institute of Molecular Cell Biology and Genetics), Jonathan Slack (Cancer Research UK), Alison Smith (John Innes Institute, Norfolk, UK), Austin Smith (University of Edinburgh, UK), Jim Smith (The Gurdon Institute, UK), John Maynard Smith (University of Sussex, UK), Mitchell Sogin (Woods Hole Institute), Frank Solomon (Massachusetts Institute of Technology), Michael Solursh (University of Iowa), Bruce Spiegelman (Harvard Medical School), Timothy Springer (Harvard Medical School), Mathias Sprinzl (University of Bayreuth, Germany), Scott Stachel (University of California, Berkeley), Andrew Staehelin (University of Colorado, Boulder), David Standring (University of California, San Francisco), Margaret Stanley (University of Cambridge), Martha Stark (University of California, San Francisco), Wilfred Stein (Hebrew University, Israel), Malcolm Steinberg (Princeton University), Ralph Steinman (deceased), Len Stephens (The Babraham Institute, UK), Paul Sternberg (California Institute of Technology), Chuck Stevens (The Salk Institute), Murray Stewart (MRC Laboratory of Molecular Biology, Cambridge), Bruce Stillman (Cold Spring Harbor Laboratory), Charles Streuli (University of Manchester, UK), Monroe Strickberger (University of Missouri, St. Louis), Robert Stroud (University of California, San Francisco), Michael Stryker (University of California, San Francisco), William Sullivan (University of California, Santa Cruz), Azim Surani (The Gurdon Institute, University of Cambridge), Daniel Szollosi (Institut National de la Recherche Agronomique, France), Jack Szostak (Harvard Medical School), Clifford Tabin (Harvard Medical School), Masatoshi Takeichi (RIKEN Center for Developmental Biology, Japan), Nicolas Tapon (London Research Institute), Diethard Tautz (University of Cologne, Germany), Julie Theriot (Stanford University),

Roger Thomas (University of Bristol, UK), Craig Thompson (Memorial Sloan Kettering Cancer Center), Janet Thornton (European Bioinformatics Institute, UK), Vernon Thornton (King's College London), Cheryll Tickle (University of Dundee, Scotland), Jim Till (Ontario Cancer Institute, Toronto), Lewis Tilney (University of Pennsylvania), David Tollervy (University of Edinburgh, UK), Ian Tomlinson (Cancer Research UK), Nick Tonks (Cold Spring Harbor Laboratory), Alain Townsend (Institute of Molecular Medicine, John Radcliffe Hospital, Oxford), Paul Travers (Scottish Institute for Regeneration Medicine), Robert Trelstad (UMDNJ—Robert Wood Johnson Medical School), Anthony Trewavas (Edinburgh University, Scotland), Nigel Unwin (MRC Laboratory of Molecular Biology, Cambridge), Victor Vacquier (University of California, San Diego), Ronald D. Vale (University of California, San Francisco), Tom Vanaman (University of Kentucky), Harry van der Westen (Wageningen, The Netherlands), Harold Varmus (National Cancer Institute, United States), Alexander J. Varshavsky (California Institute of Technology), Donald Voet (University of Pennsylvania), Harald von Boehmer (Harvard Medical School), Madhu Wahli (University of California, San Francisco), Virginia Walbot (Stanford University), Frank Walsh (GlaxoSmithKline, UK), Trevor Wang (John Innes Institute, Norwich, UK), Xiaodong Wang (The University of Texas Southwestern Medical School), Yu-Lie Wang (Worcester Foundation for Biomedical Research, MA), Gary Ward (University of Vermont), Anne Warner (University College London), Graham Warren (Yale University School of Medicine), Paul Wasserman (Mount Sinai School of Medicine), Clare Waterman-Storer (The Scripps Research Institute), Fiona Watt (Cancer Research UK), John Watts (John Innes Institute, Norwich, UK), Klaus Weber (Max Planck Institute for Biophysical Chemistry), Martin Weigert (Institute of Cancer Research, Philadelphia), Robert Weinberg (Massachusetts Institute of Technology), Harold Weintraub (deceased), Karsten Weis (Swiss Federal Institute of Technology), Irving Weissman (Stanford University), Jonathan Weissman (University of California, San Francisco), Susan R. Wente (Vanderbilt University School of Medicine), Norman Wessells (University of Oregon, Eugene), Stephen West (Cancer Research UK), Judy White (University of Virginia), William Wickner (Dartmouth College), Michael Wilcox (deceased), Lewis T. Williams (Chiron Corporation), Patrick Williamson (University of Massachusetts, Amherst), Keith Willison (Chester Beatty Laboratories, London), John Wilson (Baylor University), Alan Wolffe (deceased), Richard Wolfenden (University of North Carolina, Chapel Hill), Sandra Wolin (Yale University School of Medicine), Lewis Wolpert (University College London), Richard D. Wood (University of Pittsburgh Cancer Institute), Abraham Worcel (University of Rochester), Nick Wright (Cancer Research UK), John Wyke (Beatson Institute for Cancer Research, Glasgow), Michael P. Yaffe (California Institute for Regenerative Medicine), Kenneth M. Yamada (NIH), Keith Yamamoto (University of California, San Francisco), Charles Yocum (University of Michigan, Ann Arbor), Peter Yurchenco (UMDNJ—Robert Wood Johnson Medical School), Rosalind Zalin (University College London), Patricia Zambryski (University of California, Berkeley), Marino Zerial (Max Planck Institute of Molecular Cell Biology and Genetics).

# Contents

<b>PART I</b>	<b>INTRODUCTION TO THE CELL</b>	<b>1</b>
Chapter 1	Cells and Genomes	1
Chapter 2	Cell Chemistry and Bioenergetics	43
Chapter 3	Proteins	109
<b>PART II</b>	<b>BASIC GENETIC MECHANISMS</b>	<b>173</b>
Chapter 4	DNA, Chromosomes, and Genomes	173
Chapter 5	DNA Replication, Repair, and Recombination	237
Chapter 6	How Cells Read the Genome: From DNA to Protein	299
Chapter 7	Control of Gene Expression	369
<b>PART III</b>	<b>WAYS OF WORKING WITH CELLS</b>	<b>439</b>
Chapter 8	Analyzing Cells, Molecules, and Systems	439
Chapter 9	Visualizing Cells	529
<b>PART IV</b>	<b>INTERNAL ORGANIZATION OF THE CELL</b>	<b>565</b>
Chapter 10	Membrane Structure	565
Chapter 11	Membrane Transport of Small Molecules and the Electrical Properties of Membranes	597
Chapter 12	Intracellular Compartments and Protein Sorting	641
Chapter 13	Intracellular Membrane Traffic	695
Chapter 14	Energy Conversion: Mitochondria and Chloroplasts	753
Chapter 15	Cell Signaling	813
Chapter 16	The Cytoskeleton	889
Chapter 17	The Cell Cycle	963
Chapter 18	Cell Death	1021
<b>PART V</b>	<b>CELLS IN THEIR SOCIAL CONTEXT</b>	<b>1035</b>
Chapter 19	Cell Junctions and the Extracellular Matrix	1035
Chapter 20	Cancer	1091
Chapter 21	Development of Multicellular Organisms	1145
Chapter 22	Stem Cells and Tissue Renewal	1217
Chapter 23	Pathogens and Infection	1263
Chapter 24	The Innate and Adaptive Immune Systems	1297
Glossary		G: 1
Index		I: 1
Tables	The Genetic Code, Amino Acids	T: 1

# Special Features

TABLE 1–2 Some Model Organisms and Their Genomes	29
TABLE 2–1 Covalent and Noncovalent Chemical Bonds	45
TABLE 2–2 Relationship Between the Standard Free-Energy Change, $\Delta G^\circ$ , and the Equilibrium Constant	63
PANEL 2–1 Chemical Bonds and Groups Commonly Encountered in Biological Molecules	90
PANEL 2–2 Water and Its Influence on the Behavior of Biological Molecules	92
PANEL 2–3 The Principal Types of Weak Noncovalent Bonds that Hold Macromolecules Together	94
PANEL 2–4 An Outline of Some of the Types of Sugars Commonly Found in Cells	96
PANEL 2–5 Fatty Acids and Other Lipids	98
PANEL 2–6 A Survey of the Nucleotides	100
PANEL 2–7 Free Energy and Biological Reactions	102
PANEL 2–8 Details of the 10 Steps of Glycolysis	104
PANEL 2–9 The Complete Citric Acid Cycle	106
PANEL 3–1 The 20 Amino Acids Found in Proteins	112
TABLE 3–3 Some Molecules Covalently Attached to Proteins Regulate Protein Function	165
TABLE 4–1 Some Vital Statistics for the Human Genome	184
TABLE 5–4 Three Major Classes of Transposable Elements	288
TABLE 6–1 Principal Types of RNAs Produced in Cells	305
PANEL 7–1 Common Structural Motifs in Transcription Regulators	376
PANEL 8–1 DNA Sequencing Methods	478
PANEL 8–2 Review of Classical Genetics	486
TABLE 11–1 A Comparison of Inorganic Ion Concentrations Inside and Outside a Typical Mammalian Cell	598
PANEL 11–1 The Derivation of the Nernst Equation	616
TABLE 12–1 Relative Volumes Occupied by the Major Intracellular Compartments in a Liver Cell (Hepatocyte)	643
PANEL 14–1 Redox Potentials	765
TABLE 14–1 Product Yields from the Oxidation of Sugars and Fats	775
TABLE 15–3 Four Major Families of Trimeric G Proteins	846
TABLE 15–4 Some Signal Proteins That Act Via RTKs	850
TABLE 15–5 The Ras Superfamily of Monomeric GTPases	854
TABLE 15–6 Some Extracellular Signal Proteins That Act Through Cytokine Receptors and the JAK–STAT Signaling Pathway	864
PANEL 16–2 The Polymerization of Actin and Tubulin	902
TABLE 16–1 Chemical Inhibitors of Actin and Microtubules	904
PANEL 16–3 Actin Filaments	905
PANEL 16–4 Microtubules	933
TABLE 16–2 Major Types of Intermediate Filament Proteins in Vertebrate Cells	944
TABLE 17–1 The Major Cyclins and Cdks of Vertebrates and Budding Yeast	969
TABLE 17–2 Summary of the Major Cell Cycle Regulatory Proteins	973
PANEL 17–1 The Principle Stages of M Phase (Mitosis and Cytokinesis) in an Animal Cell	980
TABLE 19–1 Anchoring Junctions	1037
TABLE 19–2 Some Types of Collagen and Their Properties	1063
TABLE 19–3 Some Types of Integrins	1076
TABLE 22–1 Blood Cells	1241
TABLE 24–2 Properties of the Major Classes of Antibodies in Humans	1318
TABLE 24–3 Properties of Human Class I and Class II MHC Proteins	1330

# Detailed Contents

## Chapter 1 Cells and Genomes

THE UNIVERSAL FEATURES OF CELLS ON EARTH	
All Cells Store Their Hereditary Information in the Same Linear Chemical Code: DNA	
All Cells Replicate Their Hereditary Information by Templated Polymerization	
All Cells Transcribe Portions of Their Hereditary Information into the Same Intermediary Form: RNA	
All Cells Use Proteins as Catalysts	
All Cells Translate RNA into Protein in the Same Way	
Each Protein Is Encoded by a Specific Gene	
Life Requires Free Energy	
All Cells Function as Biochemical Factories Dealing with the Same Basic Molecular Building Blocks	8
All Cells Are Enclosed in a Plasma Membrane Across Which Nutrients and Waste Materials Must Pass	
A Living Cell Can Exist with Fewer Than 500 Genes	
Summary	
THE DIVERSITY OF GENOMES AND THE TREE OF LIFE	
Cells Can Be Powered by a Variety of Free-Energy Sources	
Some Cells Fix Nitrogen and Carbon Dioxide for Others	
The Greatest Biochemical Diversity Exists Among Prokaryotic Cells	
The Tree of Life Has Three Primary Branches: Bacteria, Archaea, and Eukaryotes	
Some Genes Evolve Rapidly; Others Are Highly Conserved	
Most Bacteria and Archaea Have 1000–6000 Genes	
New Genes Are Generated from Preexisting Genes	
Gene Duplications Give Rise to Families of Related Genes Within a Single Cell	
Genes Can Be Transferred Between Organisms, Both in the Laboratory and in Nature	
Sex Results in Horizontal Exchanges of Genetic Information Within a Species	
The Function of a Gene Can Often Be Deduced from Its Sequence	
More Than 200 Gene Families Are Common to All Three Primary Branches of the Tree of Life	
Mutations Reveal the Functions of Genes	
Molecular Biology Began with a Spotlight on <i>E. coli</i>	
Summary	
GENETIC INFORMATION IN EUKARYOTES	
Eukaryotic Cells May Have Originated as Predators	
Modern Eukaryotic Cells Evolved from a Symbiosis	
Eukaryotes Have Hybrid Genomes	
Eukaryotic Genomes Are Big	
Eukaryotic Genomes Are Rich in Regulatory DNA	
The Genome Defines the Program of Multicellular Development	
Many Eukaryotes Live as Solitary Cells	
A Yeast Serves as a Minimal Model Eukaryote	
The Expression Levels of All the Genes of An Organism Can Be Monitored Simultaneously	
<i>Arabidopsis</i> Has Been Chosen Out of 300,000 Species As a Model Plant	
The World of Animal Cells Is Represented By a Worm, a Fly, a Fish, a Mouse, and a Human	
Studies in <i>Drosophila</i> Provide a Key to Vertebrate Development	
The Vertebrate Genome Is a Product of Repeated Duplications	

1	The Frog and the Zebrafish Provide Accessible Models for Vertebrate Development	35
2	The Mouse Is the Predominant Mammalian Model Organism	35
	Humans Report on Their Own Peculiarities	36
2	We Are All Different in Detail	38
	To Understand Cells and Organisms Will Require Mathematics, Computers, and Quantitative Information	38
3	Summary	39
4	Problems	39
5	References	41
6		
7	<b>Chapter 2 Cell Chemistry and Bioenergetics</b>	43
8		
8	<b>THE CHEMICAL COMPONENTS OF A CELL</b>	43
8	Water Is Held Together by Hydrogen Bonds	44
	Four Types of Noncovalent Attractions Help Bring Molecules Together in Cells	44
8	Some Polar Molecules Form Acids and Bases in Water	45
9	A Cell Is Formed from Carbon Compounds	47
10	Cells Contain Four Major Families of Small Organic Molecules	47
10	The Chemistry of Cells Is Dominated by Macromolecules with Remarkable Properties	47
12	Noncovalent Bonds Specify Both the Precise Shape of a Macromolecule and Its Binding to Other Molecules	49
12	Summary	50
14		
14	<b>CATALYSIS AND THE USE OF ENERGY BY CELLS</b>	51
15	Cell Metabolism Is Organized by Enzymes	51
16	Biological Order Is Made Possible by the Release of Heat Energy from Cells	52
16	Cells Obtain Energy by the Oxidation of Organic Molecules	54
17	Oxidation and Reduction Involve Electron Transfers	55
17	Enzymes Lower the Activation-Energy Barriers That Block Chemical Reactions	57
18	Enzymes Can Drive Substrate Molecules Along Specific Reaction Pathways	58
19	How Enzymes Find Their Substrates: The Enormous Rapidity of Molecular Motions	59
20	The Free-Energy Change for a Reaction, $\Delta G$ , Determines Whether It Can Occur Spontaneously	60
21	The Concentration of Reactants Influences the Free-Energy Change and a Reaction's Direction	61
22	The Standard Free-Energy Change, $\Delta G^\circ$ , Makes It Possible to Compare the Energetics of Different Reactions	61
23	The Equilibrium Constant and $\Delta G^\circ$ Are Readily Derived from Each Other	62
24	The Free-Energy Changes of Coupled Reactions Are Additive	63
25	Activated Carrier Molecules Are Essential for Biosynthesis	63
27	The Formation of an Activated Carrier Is Coupled to an Energetically Favorable Reaction	64
28	ATP Is the Most Widely Used Activated Carrier Molecule	65
29	Energy Stored in ATP Is Often Harnessed to Join Two Molecules Together	65
30	NADH and NADPH Are Important Electron Carriers	67
32	There Are Many Other Activated Carrier Molecules in Cells	68
32	The Synthesis of Biological Polymers Is Driven by ATP Hydrolysis	70
33	Summary	73
33	<b>HOW CELLS OBTAIN ENERGY FROM FOOD</b>	73
33	Glycolysis Is a Central ATP-Producing Pathway	74
34	Fermentations Produce ATP in the Absence of Oxygen	75

Glycolysis Illustrates How Enzymes Couple Oxidation to Energy Storage	76	The Regulation of the Src Protein Kinase Reveals How a Protein Can Function as a Microprocessor	155
Organisms Store Food Molecules in Special Reservoirs	78	Proteins That Bind and Hydrolyze GTP Are Ubiquitous Cell Regulators	156
Most Animal Cells Derive Their Energy from Fatty Acids Between Meals	81	Regulatory Proteins GAP and GEF Control the Activity of GTP-Binding Proteins by Determining Whether GTP or GDP Is Bound	157
Sugars and Fats Are Both Degraded to Acetyl CoA in Mitochondria	81	Proteins Can Be Regulated by the Covalent Addition of Other Proteins	157
The Citric Acid Cycle Generates NADH by Oxidizing Acetyl Groups to CO <sub>2</sub>	82	An Elaborate Ubiquitin-Conjugating System Is Used to Mark Proteins	158
Electron Transport Drives the Synthesis of the Majority of the ATP in Most Cells	84	Protein Complexes with Interchangeable Parts Make Efficient Use of Genetic Information	159
Amino Acids and Nucleotides Are Part of the Nitrogen Cycle	85	A GTP-Binding Protein Shows How Large Protein Movements Can Be Generated	160
Metabolism Is Highly Organized and Regulated	87	Motor Proteins Produce Large Movements in Cells	161
Summary	88	Membrane-Bound Transporters Harness Energy to Pump Molecules Through Membranes	163
Problems	88	Proteins Often Form Large Complexes That Function as Protein Machines	164
References	108	Scaffolds Concentrate Sets of Interacting Proteins	164
<b>Chapter 3 Proteins</b>	<b>109</b>	Many Proteins Are Controlled by Covalent Modifications That Direct Them to Specific Sites Inside the Cell	165
THE SHAPE AND STRUCTURE OF PROTEINS	109	A Complex Network of Protein Interactions Underlies Cell Function	166
The Shape of a Protein Is Specified by Its Amino Acid Sequence	109	Summary	169
Proteins Fold into a Conformation of Lowest Energy	114	Problems	170
The $\alpha$ Helix and the $\beta$ Sheet Are Common Folding Patterns	115	References	172
Protein Domains Are Modular Units from Which Larger Proteins Are Built	117	<b>Chapter 4 DNA, Chromosomes, and Genomes</b>	<b>173</b>
Few of the Many Possible Polypeptide Chains Will Be Useful to Cells	118	THE STRUCTURE AND FUNCTION OF DNA	173
Proteins Can Be Classified into Many Families	119	A DNA Molecule Consists of Two Complementary Chains of Nucleotides	175
Some Protein Domains Are Found in Many Different Proteins	121	The Structure of DNA Provides a Mechanism for Heredity	177
Certain Pairs of Domains Are Found Together in Many Proteins	122	In Eukaryotes, DNA Is Enclosed in a Cell Nucleus	178
The Human Genome Encodes a Complex Set of Proteins, Revealing That Much Remains Unknown	122	Summary	179
Larger Protein Molecules Often Contain More Than One Polypeptide Chain	123	<b>CHROMOSOMAL DNA AND ITS PACKAGING IN THE CHROMATIN FIBER</b>	<b>179</b>
Some Globular Proteins Form Long Helical Filaments	123	Eukaryotic DNA Is Packaged into a Set of Chromosomes	180
Many Protein Molecules Have Elongated, Fibrous Shapes	124	Chromosomes Contain Long Strings of Genes	182
Proteins Contain a Surprisingly Large Amount of Intrinsically Disordered Polypeptide Chain	125	The Nucleotide Sequence of the Human Genome Shows How Our Genes Are Arranged	183
Covalent Cross-Linkages Stabilize Extracellular Proteins	127	Each DNA Molecule That Forms a Linear Chromosome Must Contain a Centromere, Two Telomeres, and Replication Origins	185
Protein Molecules Often Serve as Subunits for the Assembly of Large Structures	127	DNA Molecules Are Highly Condensed in Chromosomes	187
Many Structures in Cells Are Capable of Self-Assembly	128	Nucleosomes Are a Basic Unit of Eukaryotic Chromosome Structure	187
Assembly Factors Often Aid the Formation of Complex Biological Structures	130	The Structure of the Nucleosome Core Particle Reveals How DNA Is Packaged	188
Amyloid Fibrils Can Form from Many Proteins	130	Nucleosomes Have a Dynamic Structure, and Are Frequently Subjected to Changes Catalyzed by ATP-Dependent Chromatin Remodeling Complexes	188
Amyloid Structures Can Perform Useful Functions in Cells	132	Nucleosomes Are Usually Packed Together into a Compact Chromatin Fiber	190
Many Proteins Contain Low-complexity Domains that Can Form "Reversible Amyloids"	132	Summary	191
Summary	134	<b>CHROMATIN STRUCTURE AND FUNCTION</b>	<b>194</b>
<b>PROTEIN FUNCTION</b>	<b>134</b>	Heterochromatin Is Highly Organized and Restricts Gene Expression	194
All Proteins Bind to Other Molecules	134	The Heterochromatic State Is Self-Propagating	194
The Surface Conformation of a Protein Determines Its Chemistry	135	The Core Histones Are Covalently Modified at Many Different Sites	196
Sequence Comparisons Between Protein Family Members Highlight Crucial Ligand-Binding Sites	135	Chromatin Acquires Additional Variety Through the Site-Specific Insertion of a Small Set of Histone Variants	198
Proteins Bind to Other Proteins Through Several Types of Interfaces	136	Covalent Modifications and Histone Variants Act in Concert to Control Chromosome Functions	198
Antibody Binding Sites Are Especially Versatile	137	A Complex of Reader and Writer Proteins Can Spread Specific Chromatin Modifications Along a Chromosome	199
The Equilibrium Constant Measures Binding Strength	138	Barrier DNA Sequences Block the Spread of Reader-Writer Complexes and thereby Separate Neighboring Chromatin Domains	202
Enzymes Are Powerful and Highly Specific Catalysts	140	The Chromatin in Centromeres Reveals How Histone Variants Can Create Special Structures	203
Substrate Binding Is the First Step in Enzyme Catalysis	141	Some Chromatin Structures Can Be Directly Inherited	204
Enzymes Speed Reactions by Selectively Stabilizing Transition States	141		
Enzymes Can Use Simultaneous Acid and Base Catalysis	142		
Lysozyme Illustrates How an Enzyme Works	144		
Tightly Bound Small Molecules Add Extra Functions to Proteins	144		
Multienzyme Complexes Help to Increase the Rate of Cell Metabolism	146		
The Cell Regulates the Catalytic Activities of Its Enzymes	148		
Allosteric Enzymes Have Two or More Binding Sites That Interact	149		
Two Ligands Whose Binding Sites Are Coupled Must Reciprocally Affect Each Other's Binding	151		
Symmetric Protein Assemblies Produce Cooperative Allosteric Transitions	151		
Many Changes in Proteins Are Driven by Protein Phosphorylation	152		
A Eukaryotic Cell Contains a Large Collection of Protein Kinases and Protein Phosphatases	153		
	154		

Experiments with Frog Embryos Suggest that both Activating and Repressive Chromatin Structures Can Be Inherited Epigenetically	205	The Proteins at a Replication Fork Cooperate to Form a Replication Machine	249
Chromatin Structures Are Important for Eukaryotic Chromosome Function	206	A Strand-Directed Mismatch Repair System Removes Replication Errors That Escape from the Replication Machine	250
Summary	207	DNA Topoisomerases Prevent DNA Tangling During Replication	251
<b>THE GLOBAL STRUCTURE OF CHROMOSOMES</b>	207	DNA Replication Is Fundamentally Similar in Eukaryotes and Bacteria	253
Chromosomes Are Folded into Large Loops of Chromatin	207	Summary	254
Polytene Chromosomes Are Uniquely Useful for Visualizing Chromatin Structures	208	<b>THE INITIATION AND COMPLETION OF DNA REPLICATION IN CHROMOSOMES</b>	254
There Are Multiple Forms of Chromatin	210	DNA Synthesis Begins at Replication Origins	254
Chromatin Loops Decondense When the Genes Within Them Are Expressed	211	Bacterial Chromosomes Typically Have a Single Origin of DNA Replication	255
Chromatin Can Move to Specific Sites Within the Nucleus to Alter Gene Expression	212	Eukaryotic Chromosomes Contain Multiple Origins of Replication	256
Networks of Macromolecules Form a Set of Distinct Biochemical Environments inside the Nucleus	213	In Eukaryotes, DNA Replication Takes Place During Only One Part of the Cell Cycle	258
Mitotic Chromosomes Are Especially Highly Condensed	214	Different Regions on the Same Chromosome Replicate at Distinct Times in S Phase	258
Summary	216	A Large Multisubunit Complex Binds to Eukaryotic Origins of Replication	259
<b>HOW GENOMES EVOLVE</b>	216	Features of the Human Genome That Specify Origins of Replication Remain to Be Discovered	260
Genome Comparisons Reveal Functional DNA Sequences by their Conservation Throughout Evolution	217	New Nucleosomes Are Assembled Behind the Replication Fork	261
Genome Alterations Are Caused by Failures of the Normal Mechanisms for Copying and Maintaining DNA, as well as by Transposable DNA Elements	217	Telomerase Replicates the Ends of Chromosomes	262
The Genome Sequences of Two Species Differ in Proportion to the Length of Time Since They Have Separately Evolved	218	Telomeres Are Packaged Into Specialized Structures That Protect the Ends of Chromosomes	263
Phylogenetic Trees Constructed from a Comparison of DNA Sequences Trace the Relationships of All Organisms	218	Telomere Length Is Regulated by Cells and Organisms	264
A Comparison of Human and Mouse Chromosomes Shows How the Structures of Genomes Diverge	219	Summary	265
The Size of a Vertebrate Genome Reflects the Relative Rates of DNA Addition and DNA Loss in a Lineage	221	<b>DNA REPAIR</b>	266
We Can Infer the Sequence of Some Ancient Genomes	222	Without DNA Repair, Spontaneous DNA Damage Would Rapidly Change DNA Sequences	267
Multispecies Sequence Comparisons Identify Conserved DNA Sequences of Unknown Function	223	The DNA Double Helix Is Readily Repaired	268
Changes in Previously Conserved Sequences Can Help Decipher Critical Steps in Evolution	224	DNA Damage Can Be Removed by More Than One Pathway	269
Mutations in the DNA Sequences That Control Gene Expression Have Driven Many of the Evolutionary Changes in Vertebrates	226	Coupling Nucleotide Excision Repair to Transcription Ensures That the Cell's Most Important DNA Is Efficiently Repaired	271
Gene Duplication Also Provides an Important Source of Genetic Novelty During Evolution	227	The Chemistry of the DNA Bases Facilitates Damage Detection	271
Duplicated Genes Diverge	227	Special Translesion DNA Polymerases Are Used in Emergencies	273
The Evolution of the Globin Gene Family Shows How DNA Duplications Contribute to the Evolution of Organisms	228	Double-Strand Breaks Are Efficiently Repaired	273
Genes Encoding New Proteins Can Be Created by the Recombination of Exons	229	DNA Damage Delays Progression of the Cell Cycle	276
Neutral Mutations Often Spread to Become Fixed in a Population, with a Probability That Depends on Population Size	230	Summary	276
A Great Deal Can Be Learned from Analyses of the Variation Among Humans	230	<b>HOMOLOGOUS RECOMBINATION</b>	276
Summary	232	Homologous Recombination Has Common Features in All Cells	277
Problems	234	DNA Base-Pairing Guides Homologous Recombination	277
References	234	Homologous Recombination Can Flawlessly Repair Double-Strand Breaks in DNA	278
<b>Chapter 5 DNA Replication, Repair, and Recombination</b>	236	Strand Exchange Is Carried Out by the RecA/Rad51 Protein	279
<b>THE MAINTENANCE OF DNA SEQUENCES</b>	237	Homologous Recombination Can Rescue Broken DNA Replication Forks	280
Mutation Rates Are Extremely Low	237	Cells Carefully Regulate the Use of Homologous Recombination in DNA Repair	280
Low Mutation Rates Are Necessary for Life as We Know It	238	Homologous Recombination Is Crucial for Meiosis	282
Summary	238	Meiotic Recombination Begins with a Programmed Double-Strand Break	282
<b>DNA REPLICATION MECHANISMS</b>	239	Holliday Junctions Are Formed During Meiosis	284
Base-Pairing Underlies DNA Replication and DNA Repair	239	Homologous Recombination Produces Both Crossovers and Non-Crossovers During Meiosis	284
The DNA Replication Fork Is Asymmetrical	240	Homologous Recombination Often Results in Gene Conversion	286
The High Fidelity of DNA Replication Requires Several Proofreading Mechanisms	242	Summary	286
Only DNA Replication in the 5'-to-3' Direction Allows Efficient Error Correction	244	<b>TRANSPOSITION AND CONSERVATIVE SITE-SPECIFIC RECOMBINATION</b>	287
A Special Nucleotide-Polymerizing Enzyme Synthesizes Short RNA Primer Molecules on the Lagging Strand	244	Through Transposition, Mobile Genetic Elements Can Insert Into Any DNA Sequence	288
Special Proteins Help to Open Up the DNA Double Helix in Front of the Replication Fork	245	DNA-Only Transposons Can Move by a Cut-and-Paste Mechanism	288
A Sliding Ring Holds a Moving DNA Polymerase Onto the DNA	246	Some Viruses Use a Transposition Mechanism to Move Themselves Into Host-Cell Chromosomes	290
	246	Retroviral-like Retrotransposons Resemble Retroviruses, but Lack a Protein Coat	291
	246	A Large Fraction of the Human Genome Is Composed of Nonretroviral Retrotransposons	291
	246	Different Transposable Elements Predominate in Different Organisms	292
	246	Genome Sequences Reveal the Approximate Times at Which Transposable Elements Have Moved	292

Conservative Site-Specific Recombination Can Reversibly Rearrange DNA	292	Proteins Are Made on Polyribosomes	349
Conservative Site-Specific Recombination Can Be Used to Turn Genes On or Off	294	There Are Minor Variations in the Standard Genetic Code	349
Bacterial Conservative Site-Specific Recombinases Have Become Powerful Tools for Cell and Developmental Biologists	294	Inhibitors of Prokaryotic Protein Synthesis Are Useful as Antibiotics	351
Summary	295	Quality Control Mechanisms Act to Prevent Translation of Damaged mRNAs	351
Problems	296	Some Proteins Begin to Fold While Still Being Synthesized	353
References	298	Molecular Chaperones Help Guide the Folding of Most Proteins	354
<b>Chapter 6 How Cells Read the Genome: From DNA to Protein</b>	<b>299</b>	Cells Utilize Several Types of Chaperones	355
<b>FROM DNA TO RNA</b>		Exposed Hydrophobic Regions Provide Critical Signals for Protein Quality Control	357
RNA Molecules Are Single-Stranded	301	The Proteasome Is a Compartmentalized Protease with Sequestered Active Sites	357
Transcription Produces RNA Complementary to One Strand of DNA	302	Many Proteins Are Controlled by Regulated Destruction	359
RNA Polymerases Carry Out Transcription	303	There Are Many Steps From DNA to Protein	361
Cells Produce Different Categories of RNA Molecules	305	Summary	362
Signals Encoded in DNA Tell RNA Polymerase Where to Start and Stop	306	<b>THE RNA WORLD AND THE ORIGINS OF LIFE</b>	362
Transcription Start and Stop Signals Are Heterogeneous in Nucleotide Sequence	307	Single-Stranded RNA Molecules Can Fold into Highly Elaborate Structures	363
Transcription Initiation in Eukaryotes Requires Many Proteins	309	RNA Can Both Store Information and Catalyze Chemical Reactions	364
RNA Polymerase II Requires a Set of General Transcription Factors	310	How Did Protein Synthesis Evolve?	365
Polymerase II Also Requires Activator, Mediator, and Chromatin-Modifying Proteins	312	All Present-Day Cells Use DNA as Their Hereditary Material	365
Transcription Elongation in Eukaryotes Requires Accessory Proteins	313	Summary	366
Transcription Creates Superhelical Tension	314	Problems	366
Transcription Elongation in Eukaryotes Is Tightly Coupled to RNA Processing	315	References	368
RNA Capping Is the First Modification of Eukaryotic Pre-mRNAs	316	<b>Chapter 7 Control of Gene Expression</b>	<b>369</b>
RNA Splicing Removes Intron Sequences from Newly Transcribed Pre-mRNAs	317	<b>AN OVERVIEW OF GENE CONTROL</b>	<b>369</b>
Nucleotide Sequences Signal Where Splicing Occurs	319	The Different Cell Types of a Multicellular Organism Contain the Same DNA	369
RNA Splicing Is Performed by the Spliceosome	319	Different Cell Types Synthesize Different Sets of RNAs and Proteins	370
The Spliceosome Uses ATP Hydrolysis to Produce a Complex Series of RNA-RNA Rearrangements	321	External Signals Can Cause a Cell to Change the Expression of Its Genes	372
Other Properties of Pre-mRNA and Its Synthesis Help to Explain the Choice of Proper Splice Sites	321	Gene Expression Can Be Regulated at Many of the Steps in the Pathway from DNA to RNA to Protein	372
Chromatin Structure Affects RNA Splicing	323	Summary	373
RNA Splicing Shows Remarkable Plasticity	323	<b>CONTROL OF TRANSCRIPTION BY SEQUENCE-SPECIFIC DNA-BINDING PROTEINS</b>	<b>373</b>
Spliceosome-Catalyzed RNA Splicing Probably Evolved from Self-splicing Mechanisms	324	The Sequence of Nucleotides in the DNA Double Helix Can Be Read by Proteins	373
RNA-Processing Enzymes Generate the 3' End of Eukaryotic mRNAs	324	Transcription Regulators Contain Structural Motifs That Can Read DNA Sequences	374
Mature Eukaryotic mRNAs Are Selectively Exported from the Nucleus	325	Dimerization of Transcription Regulators Increases Their Affinity and Specificity for DNA	375
Noncoding RNAs Are Also Synthesized and Processed in the Nucleus	327	Transcription Regulators Bind Cooperatively to DNA	378
The Nucleolus Is a Ribosome-Producing Factory	329	Nucleosome Structure Promotes Cooperative Binding of Transcription Regulators	379
The Nucleus Contains a Variety of Subnuclear Aggregates	331	Summary	380
<b>Summary</b>	333	<b>TRANSCRIPTION REGULATORS SWITCH GENES ON AND OFF</b>	<b>380</b>
<b>FROM RNA TO PROTEIN</b>		The Tryptophan Repressor Switches Genes Off	380
An mRNA Sequence Is Decoded in Sets of Three Nucleotides	334	Repressors Turn Genes Off and Activators Turn Them On	381
tRNA Molecules Match Amino Acids to Codons in mRNA	334	An Activator and a Repressor Control the Lac Operon	382
tRNAs Are Covalently Modified Before They Exit from the Nucleus	336	DNA Looping Can Occur During Bacterial Gene Regulation	383
Specific Enzymes Couple Each Amino Acid to Its Appropriate tRNA Molecule	336	Complex Switches Control Gene Transcription in Eukaryotes	384
Editing by tRNA Synthetases Ensures Accuracy	338	A Eukaryotic Gene Control Region Consists of a Promoter Plus Many <i>cis</i> -Regulatory Sequences	384
Amino Acids Are Added to the C-terminal End of a Growing Polypeptide Chain	339	Eukaryotic Transcription Regulators Work in Groups	385
The RNA Message Is Decoded in Ribosomes	340	Activator Proteins Promote the Assembly of RNA Polymerase at the Start Point of Transcription	386
Elongation Factors Drive Translation Forward and Improve Its Accuracy	343	Eukaryotic Transcription Activators Direct the Modification of Local Chromatin Structure	386
Many Biological Processes Overcome the Inherent Limitations of Complementary Base-Pairing	345	Transcription Activators Can Promote Transcription by Releasing RNA Polymerase from Promoters	388
Accuracy in Translation Requires an Expenditure of Free Energy	345	Transcription Activators Work Synergistically	388
The Ribosome Is a Ribozyme	346	Eukaryotic Transcription Repressors Can Inhibit Transcription in Several Ways	389
Nucleotide Sequences in mRNA Signal Where to Start Protein Synthesis	347	Insulator DNA Sequences Prevent Eukaryotic Transcription Regulators from Influencing Distant Genes	391
Stop Codons Mark the End of Translation	348	Summary	392

<b>MOLECULAR GENETIC MECHANISMS THAT CREATE AND MAINTAIN SPECIALIZED CELL TYPES</b>	392	Hybridoma Cell Lines Are Factories That Produce Monoclonal Antibodies	444
Complex Genetic Switches That Regulate <i>Drosophila</i> Development Are Built Up from Smaller Molecules	392	Summary	445
The <i>Drosophila</i> Eve Gene Is Regulated by Combinatorial Controls	394	<b>PURIFYING PROTEINS</b>	445
Transcription Regulators Are Brought Into Play by Extracellular Signals	395	Cells Can Be Separated into Their Component Fractions	445
Combinatorial Gene Control Creates Many Different Cell Types	396	Cell Extracts Provide Accessible Systems to Study Cell Functions	447
Specialized Cell Types Can Be Experimentally Reprogrammed to Become Pluripotent Stem Cells	398	Proteins Can Be Separated by Chromatography	448
Combinations of Master Transcription Regulators Specify Cell Types by Controlling the Expression of Many Genes	398	Immunoprecipitation Is a Rapid Affinity Purification Method	449
Specialized Cells Must Rapidly Turn Sets of Genes On and Off	399	Genetically Engineered Tags Provide an Easy Way to Purify Proteins	450
Differentiated Cells Maintain Their Identity	400	Purified Cell-free Systems Are Required for the Precise Dissection of Molecular Functions	451
Transcription Circuits Allow the Cell to Carry Out Logic Operations	402	Summary	451
Summary	404	<b>ANALYZING PROTEINS</b>	452
<b>MECHANISMS THAT REINFORCE CELL MEMORY IN PLANTS AND ANIMALS</b>	404	Proteins Can Be Separated by SDS Polyacrylamide-Gel Electrophoresis	452
Patterns of DNA Methylation Can Be Inherited When Vertebrate Cells Divide	404	Two-Dimensional Gel Electrophoresis Provides Greater Protein Separation	452
CG-Rich Islands Are Associated with Many Genes in Mammals	405	Specific Proteins Can Be Detected by Blotting with Antibodies	454
Genomic Imprinting Is Based on DNA Methylation	407	Hydrodynamic Measurements Reveal the Size and Shape of a Protein Complex	455
Chromosome-Wide Alterations in Chromatin Structure Can Be Inherited	409	Mass Spectrometry Provides a Highly Sensitive Method for Identifying Unknown Proteins	455
Epigenetic Mechanisms Ensure That Stable Patterns of Gene Expression Can Be Transmitted to Daughter Cells	411	Sets of Interacting Proteins Can Be Identified by Biochemical Methods	457
Summary	413	Optical Methods Can Monitor Protein Interactions	458
<b>POST-TRANSCRIPTIONAL CONTROLS</b>	413	Protein Function Can Be Selectively Disrupted With Small Molecules	459
Transcription Attenuation Causes the Premature Termination of Some RNA Molecules	414	Protein Structure Can Be Determined Using X-Ray Diffraction	460
Riboswitches Probably Represent Ancient Forms of Gene Control	414	NMR Can Be Used to Determine Protein Structure in Solution	461
Alternative RNA Splicing Can Produce Different Forms of a Protein from the Same Gene	415	Protein Sequence and Structure Provide Clues About Protein Function	462
The Definition of a Gene Has Been Modified Since the Discovery of Alternative RNA Splicing	416	Summary	463
A Change in the Site of RNA Transcript Cleavage and Poly-A Addition Can Change the C-terminus of a Protein	417	<b>ANALYZING AND MANIPULATING DNA</b>	463
RNA Editing Can Change the Meaning of the RNA Message	418	Restriction Nucleases Cut Large DNA Molecules into Specific Fragments	464
RNA Transport from the Nucleus Can Be Regulated	419	Gel Electrophoresis Separates DNA Molecules of Different Sizes	465
Some mRNAs Are Localized to Specific Regions of the Cytosol	421	Purified DNA Molecules Can Be Specifically Labeled with Radioisotopes or Chemical Markers <i>in vitro</i>	467
The 5' and 3' Untranslated Regions of mRNAs Control Their Translation	422	Genes Can Be Cloned Using Bacteria	467
The Phosphorylation of an Initiation Factor Regulates Protein Synthesis Globally	423	An Entire Genome Can Be Represented in a DNA Library	469
Initiation at AUG Codons Upstream of the Translation Start Can Regulate Eukaryotic Translation Initiation	424	Genomic and cDNA Libraries Have Different Advantages and Drawbacks	471
Internal Ribosome Entry Sites Provide Opportunities for Translational Control	425	Hybridization Provides a Powerful, But Simple Way to Detect Specific Nucleotide Sequences	472
Changes in mRNA Stability Can Regulate Gene Expression	426	Genes Can Be Cloned <i>in vitro</i> Using PCR	473
Regulation of mRNA Stability Involves P-bodies and Stress Granules	427	PCR Is Also Used for Diagnostic and Forensic Applications	474
Summary	428	Both DNA and RNA Can Be Rapidly Sequenced	477
<b>REGULATION OF GENE EXPRESSION BY NONCODING RNAs</b>	429	To Be Useful, Genome Sequences Must Be Annotated	477
Small Noncoding RNA Transcripts Regulate Many Animal and Plant Genes Through RNA Interference	429	DNA Cloning Allows Any Protein to be Produced in Large Amounts	483
miRNAs Regulate mRNA Translation and Stability	429	Summary	484
RNA Interference Is Also Used as a Cell Defense Mechanism	431	<b>STUDYING GENE EXPRESSION AND FUNCTION</b>	485
RNA Interference Can Direct Heterochromatin Formation	432	Classical Genetics Begins by Disrupting a Cell Process by Random Mutagenesis	485
piRNAs Protect the Germ Line from Transposable Elements	433	Genetic Screens Identify Mutants with Specific Abnormalities	488
RNA Interference Has Become a Powerful Experimental Tool	433	Mutations Can Cause Loss or Gain of Protein Function	489
Bacteria Use Small Noncoding RNAs to Protect Themselves from Viruses	433	Complementation Tests Reveal Whether Two Mutations Are in the Same Gene or Different Genes	490
Long Noncoding RNAs Have Diverse Functions in the Cell	435	Gene Products Can Be Ordered in Pathways by Epistasis Analysis	490
Summary	436	Mutations Responsible for a Phenotype Can Be Identified Through DNA Analysis	491
Problems	436	Rapid and Cheap DNA Sequencing Has Revolutionized Human Genetic Studies	491
References	438	Linked Blocks of Polymorphisms Have Been Passed Down from Our Ancestors	492
<b>Chapter 8 Analyzing Cells, Molecules, and Systems</b>	439	Polymorphisms Can Aid the Search for Mutations Associated with Disease	493
<b>ISOLATING CELLS AND GROWING THEM IN CULTURE</b>	440	Genomics Is Accelerating the Discovery of Rare Mutations That Predispose Us to Serious Disease	493
Cells Can Be Isolated from Tissues	440	Reverse Genetics Begins with a Known Gene and Determines Which Cell Processes Require Its Function	494
Cells Can Be Grown in Culture	440	Animals and Plants Can Be Genetically Altered	495
Eukaryotic Cell Lines Are a Widely Used Source of Homogeneous Cells	442		

The Bacterial CRISPR System Has Been Adapted to Edit Genomes in a Wide Variety of Species	497	Superresolution Fluorescence Techniques Can Overcome Diffraction-Limited Resolution	549
Large Collections of Engineered Mutations Provide a Tool for Examining the Function of Every Gene in an Organism	498	Superresolution Can Also be Achieved Using Single-Molecule Localization Methods	551
RNA Interference Is a Simple and Rapid Way to Test Gene Function	499	Summary	554
Reporter Genes Reveal When and Where a Gene Is Expressed	501	<b>LOOKING AT CELLS AND MOLECULES IN THE ELECTRON MICROSCOPE</b>	554
<i>In situ</i> Hybridization Can Reveal the Location of mRNAs and Noncoding RNAs	502	The Electron Microscope Resolves the Fine Structure of the Cell	554
Expression of Individual Genes Can Be Measured Using Quantitative RT-PCR	502	Biological Specimens Require Special Preparation for Electron Microscopy	555
Analysis of mRNAs by Microarray or RNA-seq Provides a Snapshot of Gene Expression	503	Specific Macromolecules Can Be Localized by Immunogold Electron Microscopy	556
Genome-wide ChIP Identifies Sites on the Genome Occupied by Transcription Regulators	505	Different Views of a Single Object Can Be Combined to Give a Three-Dimensional Reconstruction	557
Ribosome Profiling Reveals Which mRNAs Are Being Translated in the Cell	505	Images of Surfaces Can Be Obtained by Scanning Electron Microscopy	558
Recombinant DNA Methods Have Revolutionized Human Health	506	Negative Staining and Cryo-electron Microscopy Both Allow Macromolecules to Be Viewed at High Resolution	559
Transgenic Plants Are Important for Agriculture	507	Multiple Images Can Be Combined to Increase Resolution	561
Summary	508	Summary	562
<b>MATHEMATICAL ANALYSIS OF CELL FUNCTIONS</b>	509	Problems	563
Regulatory Networks Depend on Molecular Interactions	509	References	564
Differential Equations Help Us Predict Transient Behavior	512		
Both Promoter Activity and Protein Degradation Affect the Rate of Change of Protein Concentration	513	<b>Chapter 10 Membrane Structure</b>	565
The Time Required to Reach Steady State Depends on Protein Lifetime	514	<b>THE LIPID BILAYER</b>	566
Quantitative Methods Are Similar for Transcription Repressors and Activators	514	Phosphoglycerides, Sphingolipids, and Sterols Are the Major Lipids in Cell Membranes	566
Negative Feedback Is a Powerful Strategy in Cell Regulation	515	Phospholipids Spontaneously Form Bilayers	568
Delayed Negative Feedback Can Induce Oscillations	516	The Lipid Bilayer Is a Two-dimensional Fluid	569
DNA Binding By a Repressor or an Activator Can Be Cooperative	516	The Fluidity of a Lipid Bilayer Depends on Its Composition	571
Positive Feedback Is Important for Switchlike Responses and Bistability	518	Despite Their Fluidity, Lipid Bilayers Can Form Domains of Different Compositions	572
Robustness Is an Important Characteristic of Biological Networks	520	Lipid Droplets Are Surrounded by a Phospholipid Monolayer	573
Two Transcription Regulators That Bind to the Same Gene Promoter Can Exert Combinatorial Control	520	The Asymmetry of the Lipid Bilayer Is Functionally Important	573
An Incoherent Feed-forward Interaction Generates Pulses	522	Glycolipids Are Found on the Surface of All Eukaryotic Plasma Membranes	575
A Coherent Feed-forward Interaction Detects Persistent Inputs	522	Summary	576
The Same Network Can Behave Differently in Different Cells Due to Stochastic Effects	523	<b>MEMBRANE PROTEINS</b>	576
Several Computational Approaches Can Be Used to Model the Reactions in Cells	524	Membrane Proteins Can Be Associated with the Lipid Bilayer in Various Ways	576
Statistical Methods Are Critical For the Analysis of Biological Data	524	Lipid Anchors Control the Membrane Localization of Some Signaling Proteins	577
Summary	525	In Most Transmembrane Proteins, the Polypeptide Chain Crosses the Lipid Bilayer in an $\alpha$ -Helical Conformation	579
Problems	525	Transmembrane $\alpha$ Helices Often Interact with One Another	580
References	528	Some $\beta$ Barrels Form Large Channels	580
<b>Chapter 9 Visualizing Cells</b>	529	Many Membrane Proteins Are Glycosylated	582
<b>LOOKING AT CELLS IN THE LIGHT MICROSCOPE</b>	529	Membrane Proteins Can Be Solubilized and Purified in Detergents	583
The Light Microscope Can Resolve Details 0.2 $\mu\text{m}$ Apart	530	Bacteriorhodopsin Is a Light-driven Proton ( $\text{H}^+$ ) Pump That Traverses the Lipid Bilayer as Seven $\alpha$ Helices	586
Photon Noise Creates Additional Limits to Resolution When Light Levels Are Low	532	Membrane Proteins Often Function as Large Complexes	588
Living Cells Are Seen Clearly in a Phase-Contrast or a Differential-Interference-Contrast Microscope	533	Many Membrane Proteins Diffuse in the Plane of the Membrane	588
Images Can Be Enhanced and Analyzed by Digital Techniques	534	Cells Can Confine Proteins and Lipids to Specific Domains Within a Membrane	590
Intact Tissues Are Usually Fixed and Sectioned Before Microscopy	535	The Cortical Cytoskeleton Gives Membranes Mechanical Strength and Restricts Membrane Protein Diffusion	591
Specific Molecules Can Be Located in Cells by Fluorescence Microscopy	536	Membrane-bending Proteins Deform Bilayers	593
Antibodies Can Be Used to Detect Specific Molecules	539	Summary	594
Imaging of Complex Three-Dimensional Objects Is Possible with the Optical Microscope	540	Problems	595
The Confocal Microscope Produces Optical Sections by Excluding Out-of-Focus Light	540	References	596
Individual Proteins Can Be Fluorescently Tagged in Living Cells and Organisms	542	<b>Chapter 11 Membrane Transport of Small Molecules and the Electrical Properties of Membranes</b>	597
Protein Dynamics Can Be Followed in Living Cells	543	<b>PRINCIPLES OF MEMBRANE TRANSPORT</b>	597
Light-Emitting Indicators Can Measure Rapidly Changing Intracellular Ion Concentrations	546	Protein-Free Lipid Bilayers Are Impermeable to Ions	598
Single Molecules Can Be Visualized by Total Internal Reflection Fluorescence Microscopy	547	There Are Two Main Classes of Membrane Transport Proteins: Transporters and Channels	598
Individual Molecules Can Be Touched, Imaged, and Moved Using Atomic Force Microscopy	548	Active Transport Is Mediated by Transporters Coupled to an Energy Source	599
		Summary	600
		<b>TRANSPORTERS AND ACTIVE MEMBRANE TRANSPORT</b>	600
		Active Transport Can Be Driven by Ion-Concentration Gradients	601

Transporters in the Plasma Membrane Regulate Cytosolic pH An Asymmetric Distribution of Transporters in Epithelial Cells Underlies the Transcellular Transport of Solutes	604	Nuclear Import Receptors Bind to Both Nuclear Localization Signals and NPC Proteins	652
There Are Three Classes of ATP-Driven Pumps	605	Nuclear Export Works Like Nuclear Import, But in Reverse	652
A P-type ATPase Pumps $\text{Ca}^{2+}$ into the Sarcoplasmic Reticulum in Muscle Cells	606	The Ran GTPase Imposes Directionality on Transport Through NPCs	653
The Plasma Membrane $\text{Na}^+-\text{K}^+$ Pump Establishes $\text{Na}^+$ and $\text{K}^+$ Gradients Across the Plasma Membrane	606	Transport Through NPCs Can Be Regulated by Controlling Access to the Transport Machinery	654
ABC Transporters Constitute the Largest Family of Membrane Transport Proteins	607	During Mitosis the Nuclear Envelope Disassembles	656
Summary	609	Summary	657
<b>CHANNELS AND THE ELECTRICAL PROPERTIES OF MEMBRANES</b>	611	<b>THE TRANSPORT OF PROTEINS INTO MITOCHONDRIA AND CHLOROPLASTS</b>	658
Aquaporins Are Permeable to Water But Impermeable to Ions	611	Translocation into Mitochondria Depends on Signal Sequences and Protein Translocators	659
Ion Channels Are Ion-Selective and Fluctuate Between Open and Closed States	612	Mitochondrial Precursor Proteins Are Imported as Unfolded Polypeptide Chains	660
The Membrane Potential in Animal Cells Depends Mainly on $\text{K}^+$ Leak Channels and the $\text{K}^+$ Gradient Across the Plasma Membrane	613	ATP Hydrolysis and a Membrane Potential Drive Protein Import Into the Matrix Space	661
The Resting Potential Decays Only Slowly When the $\text{Na}^+-\text{K}^+$ Pump Is Stopped	615	Bacteria and Mitochondria Use Similar Mechanisms to Insert Porins into their Outer Membrane	662
The Three-Dimensional Structure of a Bacterial $\text{K}^+$ Channel Shows How an Ion Channel Can Work	615	Transport Into the Inner Mitochondrial Membrane and Intermembrane Space Occurs Via Several Routes	663
Mechanosensitive Channels Protect Bacterial Cells Against Extreme Osmotic Pressures	617	Two Signal Sequences Direct Proteins to the Thylakoid Membrane in Chloroplasts	664
The Function of a Neuron Depends on Its Elongated Structure	619	Summary	666
Voltage-Gated Cation Channels Generate Action Potentials in Electrically Excitable Cells	620	<b>PEROXISOMES</b>	666
The Use of Channelrhodopsins Has Revolutionized the Study of Neural Circuits	621	Peroxisomes Use Molecular Oxygen and Hydrogen Peroxide to Perform Oxidation Reactions	666
Myelination Increases the Speed and Efficiency of Action Potential Propagation in Nerve Cells	623	A Short Signal Sequence Directs the Import of Proteins into Peroxisomes	667
Patch-Clamp Recording Indicates That Individual Ion Channels Open in an All-or-Nothing Fashion	625	Summary	669
Voltage-Gated Cation Channels Are Evolutionarily and Structurally Related	626	<b>THE ENDOPLASMIC RETICULUM</b>	669
Different Neuron Types Display Characteristic Stable Firing Properties	627	The ER Is Structurally and Functionally Diverse	670
Transmitter-Gated Ion Channels Convert Chemical Signals into Electrical Ones at Chemical Synapses	627	Signal Sequences Were First Discovered in Proteins Imported into the Rough ER	672
Chemical Synapses Can Be Excitatory or Inhibitory	629	A Signal-Recognition Particle (SRP) Directs the ER Signal Sequence to a Specific Receptor in the Rough ER Membrane	673
The Acetylcholine Receptors at the Neuromuscular Junction Are Excitatory Transmitter-Gated Cation Channels	630	The Polypeptide Chain Passes Through an Aqueous Channel in the Translocator	675
Neurons Contain Many Types of Transmitter-Gated Channels	631	Translocation Across the ER Membrane Does Not Always Require Ongoing Polypeptide Chain Elongation	677
Many Psychoactive Drugs Act at Synapses	631	In Single-Pass Transmembrane Proteins, a Single Internal ER Signal Sequence Remains in the Lipid Bilayer as a Membrane-spanning $\alpha$ Helix	677
Neuromuscular Transmission Involves the Sequential Activation of Five Different Sets of Ion Channels	632	Combinations of Start-Transfer and Stop-Transfer Signals Determine the Topology of Multipass Transmembrane Proteins	679
Single Neurons Are Complex Computation Devices	633	ER Tail-anchored Proteins Are Integrated into the ER Membrane by a Special Mechanism	682
Neuronal Computation Requires a Combination of at Least Three Kinds of $\text{K}^+$ Channels	634	Translocated Polypeptide Chains Fold and Assemble in the Lumen of the Rough ER	682
Long-Term Potentiation (LTP) in the Mammalian Hippocampus Depends on $\text{Ca}^{2+}$ Entry Through NMDA-Receptor Channels	636	Most Proteins Synthesized in the Rough ER Are Glycosylated by the Addition of a Common <i>N</i> -Linked Oligosaccharide	683
Summary	637	Oligosaccharides Are Used as Tags to Mark the State of Protein Folding	685
Problems	638	Improperly Folded Proteins Are Exported from the ER and Degraded in the Cytosol	685
References	640	Misfolded Proteins in the ER Activate an Unfolded Protein Response	686
<b>Chapter 12 Intracellular Compartments and Protein Sorting</b>	641	Some Membrane Proteins Acquire a Covalently Attached Glycosylphosphatidylinositol (GPI) Anchor	688
<b>THE COMPARTMENTALIZATION OF CELLS</b>	641	The ER Assembles Most Lipid Bilayers	689
All Eukaryotic Cells Have the Same Basic Set of Membrane-enclosed Organelles	641	Summary	691
Evolutionary Origins May Help Explain the Topological Relationships of Organelles	643	Problems	692
Proteins Can Move Between Compartments in Different Ways	643	References	694
Signal Sequences and Sorting Receptors Direct Proteins to the Correct Cell Address	645	<b>Chapter 13 Intracellular Membrane Traffic</b>	695
Most Organelles Cannot Be Constructed De Novo: They Require Information in the Organelle Itself	647	<b>THE MOLECULAR MECHANISMS OF MEMBRANE TRANSPORT AND THE MAINTENANCE OF COMPARTMENTAL DIVERSITY</b>	697
Summary	648	There Are Various Types of Coated Vesicles	697
<b>THE TRANSPORT OF MOLECULES BETWEEN THE NUCLEUS AND THE CYTOSOL</b>	649	The Assembly of a Clathrin Coat Drives Vesicle Formation	697
Nuclear Pore Complexes Perforate the Nuclear Envelope	649	Adaptor Proteins Select Cargo into Clathrin-Coated Vesicles	698
Nuclear Localization Signals Direct Nuclear Proteins to the Nucleus	650	Phosphoinositides Mark Organelles and Membrane Domains	700

Membrane-Bending Proteins Help Deform the Membrane During Vesicle Formation	701	Secretory Vesicle Membrane Components Are Quickly Removed from the Plasma Membrane	746
Cytoplasmic Proteins Regulate the Pinching-Off and Uncoating of Coated Vesicles	701	Some Regulated Exocytosis Events Serve to Enlarge the Plasma Membrane	748
Monomeric GTPases Control Coat Assembly	703	Polarized Cells Direct Proteins from the <i>Trans</i> Golgi Network to the Appropriate Domain of the Plasma Membrane	748
Not All Transport Vesicles Are Spherical	704	Summary	750
Rab Proteins Guide Transport Vesicles to Their Target Membrane	705	Problems	750
Rab Cascades Can Change the Identity of an Organelle	707	References	752
SNAREs Mediate Membrane Fusion	708		
Interacting SNAREs Need to Be Pried Apart Before They Can Function Again	709		
Summary	710		
<b>TRANSPORT FROM THE ER THROUGH THE GOLGI APPARATUS</b>	<b>710</b>	<b>Chapter 14 Energy Conversion: Mitochondria and Chloroplasts</b>	<b>753</b>
Proteins Leave the ER in COPII-Coated Transport Vesicles	711	<b>THE MITOCHONDRION</b>	<b>755</b>
Only Proteins That Are Properly Folded and Assembled Can Leave the ER	712	The Mitochondrion Has an Outer Membrane and an Inner Membrane	757
Vesicular Tubular Clusters Mediate Transport from the ER to the Golgi Apparatus	712	The Inner Membrane Cristae Contain the Machinery for Electron Transport and ATP Synthesis	758
The Retrieval Pathway to the ER Uses Sorting Signals	713	The Citric Acid Cycle in the Matrix Produces NADH	758
Many Proteins Are Selectively Retained in the Compartments in Which They Function	714	Mitochondria Have Many Essential Roles in Cellular Metabolism	759
The Golgi Apparatus Consists of an Ordered Series of Compartments	715	A Chemiosmotic Process Couples Oxidation Energy to ATP Production	761
Oligosaccharide Chains Are Processed in the Golgi Apparatus	716	The Energy Derived from Oxidation Is Stored as an Electrochemical Gradient	762
Proteoglycans Are Assembled in the Golgi Apparatus	718	Summary	763
What Is the Purpose of Glycosylation?	719	<b>THE PROTON PUMPS OF THE ELECTRON-TRANSPORT CHAIN</b>	<b>763</b>
Transport Through the Golgi Apparatus May Occur by Cisternal Maturation	720	The Redox Potential Is a Measure of Electron Affinities	763
Golgi Matrix Proteins Help Organize the Stack	721	Electron Transfers Release Large Amounts of Energy	764
Summary	722	Transition Metal Ions and Quinones Accept and Release Electrons Readily	764
<b>TRANSPORT FROM THE TRANS GOLGI NETWORK TO LYSOSOMES</b>	<b>722</b>	NADH Transfers Its Electrons to Oxygen Through Three Large Enzyme Complexes Embedded in the Inner Membrane	766
Lysosomes Are the Principal Sites of Intracellular Digestion	722	The NADH Dehydrogenase Complex Contains Separate Modules for Electron Transport and Proton Pumping	768
Lysosomes Are Heterogeneous	723	Cytochrome c Reductase Takes Up and Releases Protons on the Opposite Side of the Crista Membrane, Thereby Pumping Protons	768
Plant and Fungal Vacuoles Are Remarkably Versatile Lysosomes	724	The Cytochrome c Oxidase Complex Pumps Protons and Reduces O <sub>2</sub> Using a Catalytic Iron–Copper Center	770
Multiple Pathways Deliver Materials to Lysosomes	725	The Respiratory Chain Forms a Supercomplex in the Crista Membrane	772
Autophagy Degrades Unwanted Proteins and Organelles	726	Protons Can Move Rapidly Through Proteins Along Predefined Pathways	773
A Mannose 6-Phosphate Receptor Sorts Lysosomal Hydrolases in the <i>Trans</i> Golgi Network	727	Summary	774
Defects in the GlcNAc Phosphotransferase Cause a Lysosomal Storage Disease in Humans	728	<b>ATP PRODUCTION IN MITOCHONDRIA</b>	<b>774</b>
Some Lysosomes and Multivesicular Bodies Undergo Exocytosis	729	The Large Negative Value of ΔG for ATP Hydrolysis Makes ATP Useful to the Cell	774
Summary	729	The ATP Synthase Is a Nanomachine that Produces ATP by Rotary Catalysis	776
<b>TRANSPORT INTO THE CELL FROM THE PLASMA MEMBRANE: ENDOCYTOSIS</b>	<b>730</b>	Proton-driven Turbines Are of Ancient Origin	777
Pinocytic Vesicles Form from Coated Pits in the Plasma Membrane	731	Mitochondrial Cristae Help to Make ATP Synthesis Efficient	778
Not All Pinocytic Vesicles Are Clathrin-Coated	731	Special Transport Proteins Exchange ATP and ADP Through the Inner Membrane	779
Cells Use Receptor-Mediated Endocytosis to Import Selected Extracellular Macromolecules	732	Chemiosmotic Mechanisms First Arose in Bacteria	780
Specific Proteins Are Retrieved from Early Endosomes and Returned to the Plasma Membrane	734	Summary	782
Plasma Membrane Signaling Receptors are Down-Regulated by Degradation in Lysosomes	735	<b>CHLOROPLASTS AND PHOTOSYNTHESIS</b>	<b>782</b>
Early Endosomes Mature into Late Endosomes	735	Chloroplasts Resemble Mitochondria But Have a Separate Thylakoid Compartment	782
ESCRT Protein Complexes Mediate the Formation of Intraluminal Vesicles in Multivesicular Bodies	736	Chloroplasts Capture Energy from Sunlight and Use It to Fix Carbon	783
Recycling Endosomes Regulate Plasma Membrane Composition	737	Carbon Fixation Uses ATP and NADPH to Convert CO <sub>2</sub> into Sugars	784
Specialized Phagocytic Cells Can Ingest Large Particles	738	Sugars Generated by Carbon Fixation Can Be Stored as Starch or Consumed to Produce ATP	785
Summary	740	The Thylakoid Membranes of Chloroplasts Contain the Protein Complexes Required for Photosynthesis and ATP Generation	786
<b>TRANSPORT FROM THE TRANS GOLGI NETWORK TO THE CELL EXTERIOR: EXOCYTOSIS</b>	<b>741</b>	Chlorophyll-Protein Complexes Can Transfer Either Excitation Energy or Electrons	787
Many Proteins and Lipids Are Carried Automatically from the <i>Trans</i> Golgi Network (TGN) to the Cell Surface	741	A Photosystem Consists of an Antenna Complex and a Reaction Center	788
Secretory Vesicles Bud from the <i>Trans</i> Golgi Network	742	The Thylakoid Membrane Contains Two Different Photosystems Working in Series	789
Precursors of Secretory Proteins Are Proteolytically Processed During the Formation of Secretory Vesicles	743		
Secretory Vesicles Wait Near the Plasma Membrane Until Signaled to Release Their Contents	744		
For Rapid Exocytosis, Synaptic Vesicles Are Primed at the Presynaptic Plasma Membrane	744		
Synaptic Vesicles Can Form Directly from Endocytic Vesicles	746		

Photosystem II Uses a Manganese Cluster to Withdraw Electrons From Water	790	Some G Proteins Signal Via Phospholipids	836
The Cytochrome <i>b</i> <sub>6</sub> - <i>f</i> Complex Connects Photosystem II to Photosystem I	791	Ca <sup>2+</sup> Functions as a Ubiquitous Intracellular Mediator	838
Photosystem I Carries Out the Second Charge-Separation Step in the Z Scheme	792	Feedback Generates Ca <sup>2+</sup> Waves and Oscillations	838
The Chloroplast ATP Synthase Uses the Proton Gradient Generated by the Photosynthetic Light Reactions to Produce ATP	793	Ca <sup>2+</sup> /Calmodulin-Dependent Protein Kinases Mediate Many Responses to Ca <sup>2+</sup> Signals	840
All Photosynthetic Reaction Centers Have Evolved From a Common Ancestor	793	Some G Proteins Directly Regulate Ion Channels	843
The Proton-Motive Force for ATP Production in Mitochondria and Chloroplasts Is Essentially the Same	794	Smell and Vision Depend on GPCRs That Regulate Ion Channels	843
Chemiosmotic Mechanisms Evolved in Stages	794	Nitric Oxide Is a Gaseous Signaling Mediator That Passes Between Cells	846
By Providing an Inexhaustible Source of Reducing Power, Photosynthetic Bacteria Overcame a Major Evolutionary Obstacle	796	Second Messengers and Enzymatic Cascades Amplify Signals	848
The Photosynthetic Electron-Transport Chains of Cyanobacteria Produced Atmospheric Oxygen and Permitted New Life-Forms	796	GPCR Desensitization Depends on Receptor Phosphorylation	848
Summary	798	Summary	849
<b>THE GENETIC SYSTEMS OF MITOCHONDRIA AND CHLOROPLASTS</b>	800	<b>SIGNALING THROUGH ENZYME-COUPLED RECEPTORS</b>	850
The Genetic Systems of Mitochondria and Chloroplasts Resemble Those of Prokaryotes	800	Activated Receptor Tyrosine Kinases (RTKs) Phosphorylate Themselves	850
Over Time, Mitochondria and Chloroplasts Have Exported Most of Their Genes to the Nucleus by Gene Transfer	800	Phosphorylated Tyrosines on RTKs Serve as Docking Sites for Intracellular Signaling Proteins	852
The Fission and Fusion of Mitochondria Are Topologically Complex Processes	801	Proteins with SH2 Domains Bind to Phosphorylated Tyrosines	852
Animal Mitochondria Contain the Simplest Genetic Systems Known	801	The GTPase Ras Mediates Signaling by Most RTKs	854
Mitochondria Have a Relaxed Codon Usage and Can Have a Variant Genetic Code	802	Ras Activates a MAP Kinase Signaling Module	855
Chloroplasts and Bacteria Share Many Striking Similarities	802	Scaffold Proteins Help Prevent Cross-talk Between Parallel MAP Kinase Modules	857
Organelle Genes Are Maternally Inherited in Animals and Plants	803	Rho Family GTPases Functionally Couple Cell-Surface Receptors to the Cytoskeleton	858
Mutations in Mitochondrial DNA Can Cause Severe Inherited Diseases	803	PI 3-Kinase Produces Lipid Docking Sites in the Plasma Membrane	859
The Accumulation of Mitochondrial DNA Mutations Is a Contributor to Aging	804	The PI-3-Kinase–Akt Signaling Pathway Stimulates Animal Cells to Survive and Grow	860
Why Do Mitochondria and Chloroplasts Maintain a Costly Separate System for DNA Transcription and Translation?	804	RTKs and GPCRs Activate Overlapping Signaling Pathways	861
Summary	806	Some Enzyme-Coupled Receptors Associate with Cytoplasmic Tyrosine Kinases	862
Problems	806	Cytokine Receptors Activate the JAK–STAT Signaling Pathway	863
References	807	Protein Tyrosine Phosphatases Reverse Tyrosine Phosphorylations	864
	807	Signal Proteins of the TGF $\beta$ Superfamily Act Through Receptor Serine/Threonine Kinases and Smads	865
	808	Summary	866
<b>Chapter 15 Cell Signaling</b>	808	<b>ALTERNATIVE SIGNALING ROUTES IN GENE REGULATION</b>	867
<b>PRINCIPLES OF CELL SIGNALING</b>	808	The Receptor Notch Is a Latent Transcription Regulatory Protein	867
Extracellular Signals Can Act Over Short or Long Distances	813	Wnt Proteins Bind to Frizzled Receptors and Inhibit the Degradation of $\beta$ -Catenin	868
Extracellular Signal Molecules Bind to Specific Receptors	814	Hedgehog Proteins Bind to Patched, Relieving Its Inhibition of Smoothened	871
Each Cell Is Programmed to Respond to Specific Combinations of Extracellular Signals	815	Many Stressful and Inflammatory Stimuli Act Through an NF $\kappa$ B-Dependent Signaling Pathway	873
There Are Three Major Classes of Cell-Surface Receptor Proteins	816	Nuclear Receptors Are Ligand-Modulated Transcription Regulators	874
Cell-Surface Receptors Relay Signals Via Intracellular Signaling Molecules	818	Circadian Clocks Contain Negative Feedback Loops That Control Gene Expression	876
Intracellular Signals Must Be Specific and Precise in a Noisy Cytoplasm	819	Three Proteins in a Test Tube Can Reconstitute a Cyanobacterial Circadian Clock	878
Intracellular Signaling Complexes Form at Activated Receptors	820	Summary	879
Modular Interaction Domains Mediate Interactions Between Intracellular Signaling Proteins	822	<b>SIGNALING IN PLANTS</b>	880
The Relationship Between Signal and Response Varies in Different Signaling Pathways	822	Multicellularity and Cell Communication Evolved Independently in Plants and Animals	880
The Speed of a Response Depends on the Turnover of Signaling Molecules	824	Receptor Serine/Threonine Kinases Are the Largest Class of Cell-Surface Receptors in Plants	881
Cells Can Respond Abruptly to a Gradually Increasing Signal	825	Ethylene Blocks the Degradation of Specific Transcription Regulatory Proteins in the Nucleus	881
Positive Feedback Can Generate an All-or-None Response	827	Regulated Positioning of Auxin Transporters Patterns Plant Growth	882
Negative Feedback is a Common Motif in Signaling Systems	828	Phytochromes Detect Red Light, and Cryptochromes Detect Blue Light	883
Cells Can Adjust Their Sensitivity to a Signal	829	Summary	885
Summary	830	Problems	886
<b>SIGNALING THROUGH G-PROTEIN-COUPLED RECEPTORS</b>	831	References	887
Trimeric G Proteins Relay Signals From GPCRs	832		
Some G Proteins Regulate the Production of Cyclic AMP	832	<b>Chapter 16 The Cytoskeleton</b>	889
Cyclic-AMP-Dependent Protein Kinase (PKA) Mediates Most of the Effects of Cyclic AMP	833	<b>FUNCTION AND ORIGIN OF THE CYTOSKELETON</b>	889
	834	Cytoskeletal Filaments Adapt to Form Dynamic or Stable Structures	890
		The Cytoskeleton Determines Cellular Organization and Polarity	892
		Filaments Assemble from Protein Subunits That Impart Specific Physical and Dynamic Properties	893

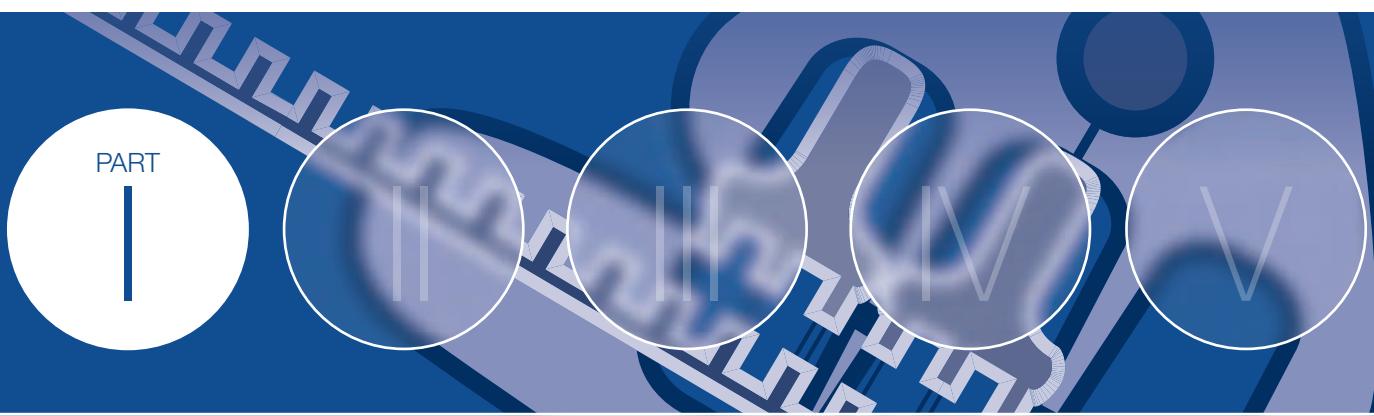
Accessory Proteins and Motors Regulate Cytoskeletal Filaments	894	Cell Polarization Is Controlled by Members of the Rho Protein Family	955
Bacterial Cell Organization and Division Depend on Homologs of Eukaryotic Cytoskeletal Proteins	896	Extracellular Signals Can Activate the Three Rho Protein Family Members	958
Summary	898	External Signals Can Dictate the Direction of Cell Migration	958
<b>ACTIN AND ACTIN-BINDING PROTEINS</b>	<b>898</b>	Communication Among Cytoskeletal Elements Coordinates Whole-Cell Polarization and Locomotion	959
Actin Subunits Assemble Head-to-Tail to Create Flexible, Polar Filaments	898	Summary	960
Nucleation Is the Rate-Limiting Step in the Formation of Actin Filaments	899	Problems	960
Actin Filaments Have Two Distinct Ends That Grow at Different Rates	900	References	962
ATP Hydrolysis Within Actin Filaments Leads to Treadmilling at Steady State	901	<b>Chapter 17 The Cell Cycle</b>	<b>963</b>
The Functions of Actin Filaments Are Inhibited by Both Polymer-stabilizing and Polymer-destabilizing Chemicals	904	<b>OVERVIEW OF THE CELL CYCLE</b>	<b>963</b>
Actin-Binding Proteins Influence Filament Dynamics and Organization	904	The Eukaryotic Cell Cycle Usually Consists of Four Phases	964
Monomer Availability Controls Actin Filament Assembly	906	Cell-Cycle Control Is Similar in All Eukaryotes	965
Actin-Nucleating Factors Accelerate Polymerization and Generate Branched or Straight Filaments	906	Cell-Cycle Progression Can Be Studied in Various Ways	966
Actin-Filament-Binding Proteins Alter Filament Dynamics	907	Summary	967
Severing Proteins Regulate Actin Filament Depolymerization	909	<b>THE CELL-CYCLE CONTROL SYSTEM</b>	<b>967</b>
Higher-Order Actin Filament Arrays Influence Cellular Mechanical Properties and Signaling	911	The Cell-Cycle Control System Triggers the Major Events of the Cell Cycle	967
Bacteria Can Hijack the Host Actin Cytoskeleton	913	The Cell-Cycle Control System Depends on Cyclically Activated Cyclin-Dependent Protein Kinases (Cdks)	968
Summary	914	Cdk Activity Can Be Suppressed By Inhibitory Phosphorylation and Cdk Inhibitor Proteins (Ckls)	970
<b>MYOSIN AND ACTIN</b>	<b>915</b>	Regulated Proteolysis Triggers the Metaphase-to-Anaphase Transition	970
Actin-Based Motor Proteins Are Members of the Myosin Superfamily	916	Cell-Cycle Control Also Depends on Transcriptional Regulation	971
Myosin Generates Force by Coupling ATP Hydrolysis to Conformational Changes	916	The Cell-Cycle Control System Functions as a Network of Biochemical Switches	972
Sliding of Myosin II Along Actin Filaments Causes Muscles to Contract	920	Summary	974
A Sudden Rise in Cytosolic $\text{Ca}^{2+}$ Concentration Initiates Muscle Contraction	923	<b>S PHASE</b>	<b>974</b>
Heart Muscle Is a Precisely Engineered Machine	923	S-Cdk Initiates DNA Replication Once Per Cycle	974
Actin and Myosin Perform a Variety of Functions in Non-Muscle Cells	923	Chromosome Duplication Requires Duplication of Chromatin Structure	975
Summary	925	Cohesins Hold Sister Chromatids Together	977
<b>MICROTUBULES</b>	<b>925</b>	Summary	977
Microtubules Are Hollow Tubes Made of Protofilaments	926	<b>MITOSIS</b>	<b>978</b>
Microtubules Undergo Dynamic Instability	926	M-Cdk Drives Entry Into Mitosis	978
Microtubule Functions Are Inhibited by Both Polymer-stabilizing and Polymer-destabilizing Drugs	927	Dephosphorylation Activates M-Cdk at the Onset of Mitosis	978
A Protein Complex Containing $\gamma$ -Tubulin Nucleates Microtubules	929	Condensin Helps Configure Duplicated Chromosomes for Separation	979
Microtubules Emanate from the Centrosome in Animal Cells	929	The Mitotic Spindle Is a Microtubule-Based Machine	982
Microtubule-Binding Proteins Modulate Filament Dynamics and Organization	930	Microtubule-Dependent Motor Proteins Govern Spindle Assembly and Function	983
Microtubule Plus-End-Binding Proteins Modulate Microtubule Dynamics and Attachments	932	Multiple Mechanisms Collaborate in the Assembly of a Bipolar Mitotic Spindle	984
Tubulin-Sequestering and Microtubule-Severing Proteins Destabilize Microtubules	932	Centrosome Duplication Occurs Early in the Cell Cycle	984
Two Types of Motor Proteins Move Along Microtubules	935	M-Cdk Initiates Spindle Assembly in Prophase	985
Microtubules and Motors Move Organelles and Vesicles	936	The Completion of Spindle Assembly in Animal Cells Requires Nuclear-Envelope Breakdown	985
Construction of Complex Microtubule Assemblies Requires Microtubule Dynamics and Motor Proteins	938	Microtubule Instability Increases Greatly in Mitosis	986
Motile Cilia and Flagella Are Built from Microtubules and Dyneins	940	Mitotic Chromosomes Promote Bipolar Spindle Assembly	986
Primary Cilia Perform Important Signaling Functions in Animal Cells	941	Kinetochores Attach Sister Chromatids to the Spindle	987
Summary	942	Bi-orientation Is Achieved by Trial and Error	988
<b>INTERMEDIATE FILAMENTS AND SEPTINS</b>	<b>943</b>	Multiple Forces Act on Chromosomes in the Spindle	990
Intermediate Filament Structure Depends on the Lateral Bundling and Twisting of Coiled-Coils	944	The APC/C Triggers Sister-Chromatid Separation and the Completion of Mitosis	992
Intermediate Filaments Impart Mechanical Stability to Animal Cells	945	Unattached Chromosomes Block Sister-Chromatid Separation: The Spindle Assembly Checkpoint	993
Linker Proteins Connect Cytoskeletal Filaments and Bridge the Nuclear Envelope	946	Chromosomes Segregate in Anaphase A and B	994
Septins Form Filaments That Regulate Cell Polarity	948	Segregated Chromosomes Are Packaged in Daughter Nuclei at Telophase	995
Summary	949	Summary	995
<b>CELL POLARIZATION AND MIGRATION</b>	<b>950</b>	<b>CYTOKINESIS</b>	<b>996</b>
Many Cells Can Crawl Across a Solid Substratum	951	Actin and Myosin II in the Contractile Ring Generate the Force for Cytokinesis	996
Actin Polymerization Drives Plasma Membrane Protrusion	951	Local Activation of RhoA Triggers Assembly and Contraction of the Contractile Ring	997
Lamellipodia Contain All of the Machinery Required for Cell Motility	953	The Microtubules of the Mitotic Spindle Determine the Plane of Animal Cell Division	997
Myosin Contraction and Cell Adhesion Allow Cells to Pull Themselves Forward	954	The Phragmoplast Guides Cytokinesis in Higher Plants	1000
		Membrane-Enclosed Organelles Must Be Distributed to Daughter Cells During Cytokinesis	1001

Some Cells Reposition Their Spindle to Divide Asymmetrically	1001	Members of the Immunoglobulin Superfamily Mediate Ca <sup>2+</sup> -Independent Cell–Cell Adhesion	1055
Mitosis Can Occur Without Cytokinesis	1002	Summary	1056
The G <sub>1</sub> Phase Is a Stable State of Cdk Inactivity	1002	<b>THE EXTRACELLULAR MATRIX OF ANIMALS</b>	1057
Summary	1004	The Extracellular Matrix Is Made and Oriented by the Cells Within It	1057
<b>MEIOSIS</b>	1004	Glycosaminoglycan (GAG) Chains Occupy Large Amounts of Space and Form Hydrated Gels	1058
Meiosis Includes Two Rounds of Chromosome Segregation	1004	Hyaluronan Acts as a Space Filler During Tissue Morphogenesis and Repair	1059
Duplicated Homologs Pair During Meiotic Prophase	1006	Proteoglycans Are Composed of GAG Chains Covalently Linked to a Core Protein	1059
Homolog Pairing Culminates in the Formation of a Synaptonemal Complex	1006	Collagens Are the Major Proteins of the Extracellular Matrix	1061
Homolog Segregation Depends on Several Unique Features of Meiosis I	1008	Secreted Fibril-Associated Collagens Help Organize the Fibrils	1063
Crossing-Over Is Highly Regulated	1009	Cells Help Organize the Collagen Fibrils They Secrete by Exerting Tension on the Matrix	1064
Meiosis Frequently Goes Wrong	1010	Elastin Gives Tissues Their Elasticity	1065
Summary	1010	Fibronectin and Other Multidomain Glycoproteins Help Organize the Matrix	1066
<b>CONTROL OF CELL DIVISION AND CELL GROWTH</b>	1010	Fibronectin Binds to Integrins	1067
Mitogens Stimulate Cell Division	1011	Tension Exerted by Cells Regulates the Assembly of Fibronectin Fibrils	1068
Cells Can Enter a Specialized Nondividing State	1012	The Basal Lamina Is a Specialized Form of Extracellular Matrix	1068
Mitogens Stimulate G <sub>1</sub> -Cdk and G <sub>1</sub> /S-Cdk Activities	1012	Laminin and Type IV Collagen Are Major Components of the Basal Lamina	1069
DNA Damage Blocks Cell Division: The DNA Damage Response	1014	Basal Laminae Have Diverse Functions	1070
Many Human Cells Have a Built-In Limitation on the Number of Times They Can Divide	1016	Cells Have to Be Able to Degrade Matrix, as Well as Make It	1072
Abnormal Proliferation Signals Cause Cell-Cycle Arrest or Apoptosis, Except in Cancer Cells	1016	Matrix Proteoglycans and Glycoproteins Regulate the Activities of Secreted Proteins	1073
Cell Proliferation is Accompanied by Cell Growth	1016	Summary	1074
Proliferating Cells Usually Coordinate Their Growth and Division	1018	<b>CELL–MATRIX JUNCTIONS</b>	1074
Summary	1018	Integrins Are Transmembrane Heterodimers That Link the Extracellular Matrix to the Cytoskeleton	1075
Problems	1019	Integrin Defects Are Responsible for Many Genetic Diseases	1076
References	1020	Integrins Can Switch Between an Active and an Inactive Conformation	1077
<b>Chapter 18 Cell Death</b>	1021	Integrins Cluster to Form Strong Adhesions	1079
Apoptosis Eliminates Unwanted Cells	1021	Extracellular Matrix Attachments Act Through Integrins to Control Cell Proliferation and Survival	1079
Apoptosis Depends on an Intracellular Proteolytic Cascade That Is Mediated by Caspases	1022	Integrins Recruit Intracellular Signaling Proteins at Sites of Cell–Matrix Adhesion	1079
Cell-Surface Death Receptors Activate the Extrinsic Pathway of Apoptosis	1024	Cell–Matrix Adhesions Respond to Mechanical Forces	1080
The Intrinsic Pathway of Apoptosis Depends on Mitochondria	1025	Summary	1081
Bcl2 Proteins Regulate the Intrinsic Pathway of Apoptosis	1025	<b>THE PLANT CELL WALL</b>	1081
IAPs Help Control Caspases	1029	The Composition of the Cell Wall Depends on the Cell Type	1082
Extracellular Survival Factors Inhibit Apoptosis in Various Ways	1029	The Tensile Strength of the Cell Wall Allows Plant Cells to Develop Turgor Pressure	1083
Phagocytes Remove the Apoptotic Cell	1030	The Primary Cell Wall Is Built from Cellulose Microfibrils Interwoven with a Network of Pectic Polysaccharides	1083
Either Excessive or Insufficient Apoptosis Can Contribute to Disease	1031	Oriented Cell Wall Deposition Controls Plant Cell Growth	1085
Summary	1032	Microtubules Orient Cell Wall Deposition	1086
Problems	1033	Summary	1087
References	1034	Problems	1087
<b>Chapter 19 Cell Junctions and the Extracellular Matrix</b>	1035	References	1089
<b>CELL–CELL JUNCTIONS</b>	1038	<b>Chapter 20 Cancer</b>	1091
Cadherins Form a Diverse Family of Adhesion Molecules	1038	<b>CANCER AS A MICROEVOLUTIONARY PROCESS</b>	1091
Cadherins Mediate Homophilic Adhesion	1038	Cancer Cells Bypass Normal Proliferation Controls and Colonize Other Tissues	1092
Cadherin-Dependent Cell–Cell Adhesion Guides the Organization of Developing Tissues	1040	Most Cancers Derive from a Single Abnormal Cell	1093
Epithelial–Mesenchymal Transitions Depend on Control of Cadherins	1042	Cancer Cells Contain Somatic Mutations	1094
Catenins Link Classical Cadherins to the Actin Cytoskeleton	1042	A Single Mutation Is Not Enough to Change a Normal Cell into a Cancer Cell	1094
Adherens Junctions Respond to Forces Generated by the Actin Cytoskeleton	1043	Cancers Develop Gradually from Increasingly Aberrant Cells	1095
Tissue Remodeling Depends on the Coordination of Actin-Mediated Contraction With Cell–Cell Adhesion	1043	Tumor Progression Involves Successive Rounds of Random Inherited Change Followed by Natural Selection	1096
Desmosomes Give Epithelia Mechanical Strength	1045	Human Cancer Cells Are Genetically Unstable	1097
Tight Junctions Form a Seal Between Cells and a Fence Between Plasma Membrane Domains	1047	Cancer Cells Display an Altered Control of Growth	1098
Tight Junctions Contain Strands of Transmembrane Adhesion Proteins	1047	Cancer Cells Have an Altered Sugar Metabolism	1098
Scaffold Proteins Organize Junctional Protein Complexes	1049	Cancer Cells Have an Abnormal Ability to Survive Stress and DNA Damage	1099
Gap Junctions Couple Cells Both Electrically and Metabolically	1050	Human Cancer Cells Escape a Built-in Limit to Cell Proliferation	1099
A Gap-Junction Connexon Is Made of Six Transmembrane Connexin Subunits	1051	The Tumor Microenvironment Influences Cancer Development	1100
In Plants, Plasmodesmata Perform Many of the Same Functions as Gap Junctions	1053		
Selectins Mediate Transient Cell–Cell Adhesions in the Bloodstream	1054		

Cancer Cells Must Survive and Proliferate in a Foreign Environment	1101	Many Cancers May Be Treatable by Enhancing the Immune Response Against the Specific Tumor	1137
Many Properties Typically Contribute to Cancerous Growth Summary	1103	Cancers Evolve Resistance to Therapies	1139
<b>CANCER-CRITICAL GENES: HOW THEY ARE FOUND AND WHAT THEY DO</b>	1103	Combination Therapies May Succeed Where Treatments with One Drug at a Time Fail	1139
The Identification of Gain-of-Function and Loss-of-Function Cancer Mutations Has Traditionally Required Different Methods	1104	We Now Have the Tools to Devise Combination Therapies Tailored to the Individual Patient	1140
Retroviruses Can Act as Vectors for Oncogenes That Alter Cell Behavior	1104	Summary	1141
Different Searches for Oncogenes Converged on the Same Gene— <i>Ras</i>	1105	Problems	1141
Genes Mutated in Cancer Can Be Made Overactive in Many Ways	1106	References	1143
Studies of Rare Hereditary Cancer Syndromes First Identified Tumor Suppressor Genes	1106	<b>Chapter 21 Development of Multicellular Organisms</b>	<b>1145</b>
Both Genetic and Epigenetic Mechanisms Can Inactivate Tumor Suppressor Genes	1107	<b>OVERVIEW OF DEVELOPMENT</b>	<b>1147</b>
Systematic Sequencing of Cancer Cell Genomes Has Transformed Our Understanding of the Disease	1108	Conserved Mechanisms Establish the Basic Animal Body Plan	1147
Many Cancers Have an Extraordinarily Disrupted Genome	1109	The Developmental Potential of Cells Becomes Progressively Restricted	1148
Many Mutations in Tumor Cells are Merely Passengers	1111	Cell Memory Underlies Cell Decision-Making	1148
About One Percent of the Genes in the Human Genome Are Cancer-Critical	1111	Several Model Organisms Have Been Crucial for Understanding Development	1148
Disruptions in a Handful of Key Pathways Are Common to Many Cancers	1112	Genes Involved in Cell–Cell Communication and Transcriptional Control Are Especially Important for Animal Development	1149
Mutations in the PI3K/Akt/mTOR Pathway Drive Cancer Cells to Grow	1113	Regulatory DNA Seems Largely Responsible for the Differences Between Animal Species	1149
Mutations in the p53 Pathway Enable Cancer Cells to Survive and Proliferate Despite Stress and DNA Damage	1113	Small Numbers of Conserved Cell–Cell Signaling Pathways Coordinate Spatial Patterning	1150
Genome Instability Takes Different Forms in Different Cancers	1114	Through Combinatorial Control and Cell Memory, Simple Signals Can Generate Complex Patterns	1150
Cancers of Specialized Tissues Use Many Different Routes to Target the Common Core Pathways of Cancer	1115	Morphogens Are Long-Range Inductive Signals That Exert Graded Effects	1151
Studies Using Mice Help to Define the Functions of Cancer-Critical Genes	1116	Lateral Inhibition Can Generate Patterns of Different Cell Types	1151
Cancers Become More and More Heterogeneous as They Progress	1117	Short-Range Activation and Long-Range Inhibition Can Generate Complex Cellular Patterns	1152
The Changes in Tumor Cells That Lead to Metastasis Are Still Largely a Mystery	1118	Asymmetric Cell Division Can Also Generate Diversity	1153
A Small Population of Cancer Stem Cells May Maintain Many Tumors	1119	Initial Patterns Are Established in Small Fields of Cells and Refined by Sequential Induction as the Embryo Grows	1153
The Cancer Stem-Cell Phenomenon Adds to the Difficulty of Curing Cancer	1120	Developmental Biology Provides Insights into Disease and Tissue Maintenance	1154
Colorectal Cancers Evolve Slowly Via a Succession of Visible Changes	1121	Summary	1154
A Few Key Genetic Lesions Are Common to a Large Fraction of Colorectal Cancers	1122	<b>MECHANISMS OF PATTERN FORMATION</b>	<b>1155</b>
Some Colorectal Cancers Have Defects in DNA Mismatch Repair	1122	Different Animals Use Different Mechanisms to Establish Their Primary Axes of Polarization	1155
The Steps of Tumor Progression Can Often Be Correlated with Specific Mutations	1123	Studies in <i>Drosophila</i> Have Revealed the Genetic Control Mechanisms Underlying Development	1157
Summary	1123	Egg-Polarity Genes Encode Macromolecules Deposited in the Egg to Organize the Axes of the Early <i>Drosophila</i> Embryo	1157
<b>CANCER PREVENTION AND TREATMENT: PRESENT AND FUTURE</b>	1124	Three Groups of Genes Control <i>Drosophila</i> Segmentation Along the A–P Axis	1159
Epidemiology Reveals That Many Cases of Cancer Are Preventable	1125	A Hierarchy of Gene Regulatory Interactions Subdivides the <i>Drosophila</i> Embryo	1159
Sensitive Assays Can Detect Those Cancer-Causing Agents that Damage DNA	1126	Egg-Polarity, Gap, and Pair-Rule Genes Create a Transient Pattern That Is Remembered by Segment-Polarity and <i>Hox</i> Genes	1160
Fifty Percent of Cancers Could Be Prevented by Changes in Lifestyle	1127	<i>Hox</i> Genes Permanently Pattern the A–P Axis	1162
Viruses and Other Infections Contribute to a Significant Proportion of Human Cancers	1127	<i>Hox</i> Proteins Give Each Segment Its Individuality	1163
Cancers of the Uterine Cervix Can Be Prevented by Vaccination Against Human Papillomavirus	1127	<i>Hox</i> Genes Are Expressed According to Their Order in the <i>Hox</i> Complex	1163
Infectious Agents Can Cause Cancer in a Variety of Ways	1128	Trithorax and Polycomb Group Proteins Enable the <i>Hox</i> Complexes to Maintain a Permanent Record of Positional Information	1164
The Search for Cancer Cures Is Difficult but Not Hopeless	1129	The D–V Signaling Genes Create a Gradient of the Transcription Regulator Dorsal	1164
Traditional Therapies Exploit the Genetic Instability and Loss of Cell-Cycle Checkpoint Responses in Cancer Cells	1131	A Hierarchy of Inductive Interactions Subdivides the Vertebrate Embryo	1166
New Drugs Can Kill Cancer Cells Selectively by Targeting Specific Mutations	1132	A Competition Between Secreted Signaling Proteins Patterns the Vertebrate Embryo	1168
PARP Inhibitors Kill Cancer Cells That Have Defects in <i>Brca1</i> or <i>Brca2</i> Genes	1132	The Insect Dorsoventral Axis Corresponds to the Vertebrate Ventral-Dorsal Axis	1169
Small Molecules Can Be Designed to Inhibit Specific Oncogenic Proteins	1133	<i>Hox</i> Genes Control the Vertebrate A–P Axis	1169
	1133	Some Transcription Regulators Can Activate a Program That Defines a Cell Type or Creates an Entire Organ	1170
	1135	Notch-Mediated Lateral Inhibition Refines Cellular Spacing Patterns	1171

Asymmetric Cell Divisions Make Sister Cells Different	1173	Ephrin-Eph Signaling Drives Segregation of the Different Gut Cell Types	1224
Differences in Regulatory DNA Explain Morphological Differences	1174	Notch Signaling Controls Gut Cell Diversification and Helps Maintain the Stem-Cell State	1224
Summary	1175	The Epidermal Stem-Cell System Maintains a Self-Renewing Waterproof Barrier	1225
<b>DEVELOPMENTAL TIMING</b>	1176	Tissue Renewal That Does Not Depend on Stem Cells: Insulin-Secreting Cells in the Pancreas and Hepatocytes in the Liver	1226
Molecular Lifetimes Play a Critical Part in Developmental Timing	1176	Some Tissues Lack Stem Cells and Are Not Renewable	1227
A Gene-Expression Oscillator Acts as a Clock to Control Vertebrate Segmentation	1177	Summary	1227
Intracellular Developmental Programs Can Help Determine the Time-Course of a Cell's Development	1179	<b>FIBROBLASTS AND THEIR TRANSFORMATIONS: THE CONNECTIVE-TISSUE CELL FAMILY</b>	1228
Cells Rarely Count Cell Divisions to Time Their Development	1180	Fibroblasts Change Their Character in Response to Chemical and Physical Signals	1228
MicroRNAs Often Regulate Developmental Transitions	1180	Osteoblasts Make Bone Matrix	1229
Hormonal Signals Coordinate the Timing of Developmental Transitions	1180	Bone Is Continually Remodeled by the Cells Within It	1230
Environmental Cues Determine the Time of Flowering	1182	Osteoclasts Are Controlled by Signals From Osteoblasts	1232
Summary	1182	Summary	1232
<b>MORPHOGENESIS</b>	1184	<b>GENESIS AND REGENERATION OF SKELETAL MUSCLE</b>	1232
Cell Migration Is Guided by Cues in the Cell's Environment	1185	Myoblasts Fuse to Form New Skeletal Muscle Fibers	1233
The Distribution of Migrant Cells Depends on Survival Factors	1186	Some Myoblasts Persist as Quiescent Stem Cells in the Adult	1234
Changing Patterns of Cell Adhesion Molecules Force Cells Into New Arrangements	1187	Summary	1235
Repulsive Interactions Help Maintain Tissue Boundaries	1188	<b>BLOOD VESSELS, LYMPHATICS, AND ENDOTHELIAL CELLS</b>	1235
Groups of Similar Cells Can Perform Dramatic Collective Rearrangements	1188	Endothelial Cells Line All Blood Vessels and Lymphatics	1235
Planar Cell Polarity Helps Orient Cell Structure and Movement in Developing Epithelia	1189	Endothelial Tip Cells Pioneer Angiogenesis	1236
Interactions Between an Epithelium and Mesenchyme Generate Branching Tubular Structures	1190	Tissues Requiring a Blood Supply Release VEGF	1237
An Epithelium Can Bend During Development to Form a Tube or Vesicle	1192	Signals from Endothelial Cells Control Recruitment of Pericytes and Smooth Muscle Cells to Form the Vessel Wall	1238
Summary	1193	Summary	1238
<b>GROWTH</b>	1193	<b>A HIERARCHICAL STEM-CELL SYSTEM: BLOOD CELL FORMATION</b>	1239
The Proliferation, Death, and Size of Cells Determine Organism Size	1194	Red Blood Cells Are All Alike; White Blood Cells Can Be Grouped in Three Main Classes	1239
Animals and Organs Can Assess and Regulate Total Cell Mass	1194	The Production of Each Type of Blood Cell in the Bone Marrow Is Individually Controlled	1240
Extracellular Signals Stimulate or Inhibit Growth	1196	Bone Marrow Contains Multipotent Hematopoietic Stem Cells, Able to Give Rise to All Classes of Blood Cells	1242
Summary	1197	Commitment Is a Stepwise Process	1243
<b>NEURAL DEVELOPMENT</b>	1198	Divisions of Committed Progenitor Cells Amplify the Number of Specialized Blood Cells	1243
Neurons Are Assigned Different Characters According to the Time and Place of Their Birth	1199	Stem Cells Depend on Contact Signals From Stromal Cells	1244
The Growth Cone Pilots Axons Along Specific Routes Toward Their Targets	1201	Factors That Regulate Hematopoiesis Can Be Analyzed in Culture	1244
A Variety of Extracellular Cues Guide Axons to their Targets	1202	Erythropoiesis Depends on the Hormone Erythropoietin	1244
The Formation of Orderly Neural Maps Depends on Neuronal Specificity	1204	Multiple CSFs Influence Neutrophil and Macrophage Production	1245
Both Dendrites and Axonal Branches From the Same Neuron Avoid One Another	1206	The Behavior of a Hematopoietic Cell Depends Partly on Chance	1245
Target Tissues Release Neurotrophic Factors That Control Nerve Cell Growth and Survival	1208	Regulation of Cell Survival Is as Important as Regulation of Cell Proliferation	1246
Formation of Synapses Depends on Two-Way Communication Between Neurons and Their Target Cells	1209	Summary	1247
Synaptic Pruning Depends on Electrical Activity and Synaptic Signaling	1211	<b>REGENERATION AND REPAIR</b>	1247
Neurons That Fire Together Wire Together	1211	Planarian Worms Contain Stem Cells That Can Regenerate a Whole New Body	1247
Summary	1213	Some Vertebrates Can Regenerate Entire Organs	1248
Problems	1213	Stem Cells Can Be Used Artificially to Replace Cells That Are Diseased or Lost: Therapy for Blood and Epidermis	1249
References	1215	Neural Stem Cells Can Be Manipulated in Culture and Used to Repopulate the Central Nervous System	1250
		Summary	1251
<b>Chapter 22 Stem Cells and Tissue Renewal</b>	1217	<b>CELL REPROGRAMMING AND PLURIPOtent STEM CELLS</b>	1251
<b>STEM CELLS AND RENEWAL IN EPITHELIAL TISSUES</b>	1217	Nuclei Can Be Reprogrammed by Transplantation into Foreign Cytoplasm	1252
The Lining of the Small Intestine Is Continually Renewed Through Cell Proliferation in the Crypts	1218	Reprogramming of a Transplanted Nucleus Involves Drastic Epigenetic Changes	1252
Stem Cells of the Small Intestine Lie at or Near the Base of Each Crypt	1219	Embryonic Stem (ES) Cells Can Generate Any Part of the Body	1253
The Two Daughters of a Stem Cell Face a Choice	1219	A Core Set of Transcription Regulators Defines and Maintains the ES Cell State	1254
Wnt Signaling Maintains the Gut Stem-Cell Compartment	1220	Fibroblasts Can Be Reprogrammed to Create Induced Pluripotent Stem Cells (iPS Cells)	1254
Stem Cells at the Crypt Base Are Multipotent, Giving Rise to the Full Range of Differentiated Intestinal Cell Types	1220	Reprogramming Involves a Massive Upheaval of the Gene Control System	1255
The Two Daughters of a Stem Cell Do Not Always Have to Become Different	1222	An Experimental Manipulation of Factors that Modify Chromatin Can Increase Reprogramming Efficiencies	1255
Paneth Cells Create the Stem-Cell Niche	1222	ES and iPS Cells Can Be Guided to Generate Specific Adult Cell Types and Even Whole Organs	1256
A Single <i>Lgr5</i> -expressing Cell in Culture Can Generate an Entire Organized Crypt-Villus System	1223		1256

Cells of One Specialized Type Can Be Forced to Transdifferentiate Directly Into Another	1258	OVERVIEW OF THE ADAPTIVE IMMUNE SYSTEM	1307
ES and iPS Cells Are Useful for Drug Discovery and Analysis of Disease	1258	B Cells Develop in the Bone Marrow, T Cells in the Thymus	1308
Summary	1260	Immunological Memory Depends On Both Clonal Expansion and Lymphocyte Differentiation	1309
Problems	1260	Lymphocytes Continuously Recirculate Through Peripheral Lymphoid Organs	1311
References	1262	Immunological Self-Tolerance Ensures That B and T Cells Do Not Attack Normal Host Cells and Molecules	1313
<b>Chapter 23 Pathogens and Infection</b>	<b>1263</b>	Summary	1315
<b>INTRODUCTION TO PATHOGENS AND THE HUMAN MICROBIOTA</b>	<b>1263</b>	<b>B CELLS AND IMMUNOGLOBULINS</b>	<b>1315</b>
The Human Microbiota Is a Complex Ecological System That Is Important for Our Development and Health	1264	B Cells Make Immunoglobulins (Igs) as Both Cell-Surface Antigen Receptors and Secreted Antibodies	1315
Pathogens Interact with Their Hosts in Different Ways	1264	Mammals Make Five Classes of Igs	1316
Pathogens Can Contribute to Cancer, Cardiovascular Disease, and Other Chronic Illnesses	1265	Ig Light and Heavy Chains Consist of Constant and Variable Regions	1318
Pathogens Can Be Viruses, Bacteria, or Eukaryotes	1266	Ig Genes Are Assembled From Separate Gene Segments During B Cell Development	1319
Bacteria Are Diverse and Occupy a Remarkable Variety of Ecological Niches	1267	Antigen-Driven Somatic Hypermutation Fine-Tunes Antibody Responses	1321
Bacterial Pathogens Carry Specialized Virulence Genes	1268	B Cells Can Switch the Class of Ig They Make	1322
Bacterial Virulence Genes Encode Effector Proteins and Secretion Systems to Deliver Effector Proteins to Host Cells	1269	Summary	1323
Fungal and Protozoan Parasites Have Complex Life Cycles Involving Multiple Forms	1271	<b>T CELLS AND MHC PROTEINS</b>	<b>1324</b>
All Aspects of Viral Propagation Depend on Host Cell Machinery	1273	T Cell Receptors (TCRs) Are Ig-like Heterodimers	1325
Summary	1275	Activated Dendritic Cells Activate Naïve T Cells	1326
<b>CELL BIOLOGY OF INFECTION</b>	<b>1276</b>	T Cells Recognize Foreign Peptides Bound to MHC Proteins	1326
Pathogens Overcome Epithelial Barriers to Infect the Host	1276	MHC Proteins Are the Most Polymorphic Human Proteins Known	1330
Pathogens That Colonize an Epithelium Must Overcome Its Protective Mechanisms	1276	CD4 and CD8 Co-receptors on T Cells Bind to Invariant Parts of MHC Proteins	1331
Extracellular Pathogens Disturb Host Cells Without Entering Them	1277	Developing Thymocytes Undergo Negative and Positive Selection	1332
Intracellular Pathogens Have Mechanisms for Both Entering and Leaving Host Cells	1278	Cytotoxic T Cells Induce Infected Target Cells to Kill Themselves	1333
Viruses Bind to Virus Receptors at the Host Cell Surface	1279	Effector Helper T Cells Help Activate Other Cells of the Innate and Adaptive Immune Systems	1335
Viruses Enter Host Cells by Membrane Fusion, Pore Formation, or Membrane Disruption	1280	Naïve Helper T Cells Can Differentiate Into Different Types of Effector T Cells	1335
Bacteria Enter Host Cells by Phagocytosis	1281	Both T and B Cells Require Multiple Extracellular Signals For Activation	1336
Intracellular Eukaryotic Parasites Actively Invade Host Cells	1282	Many Cell-Surface Proteins Belong to the Ig Superfamily	1338
Some Intracellular Pathogens Escape from the Phagosome into the Cytosol	1284	Summary	1339
Many Pathogens Alter Membrane Traffic in the Host Cell to Survive and Reproduce	1284	Problems	1340
Viruses and Bacteria Use the Host-Cell Cytoskeleton for Intracellular Movement	1286	References	1342
Viruses Can Take Over the Metabolism of the Host Cell	1288	<b>Glossary</b>	<b>G:1</b>
Pathogens Can Evolve Rapidly by Antigenic Variation	1289	<b>Index</b>	<b>I:1</b>
Error-Prone Replication Dominates Viral Evolution	1291	<b>Tables</b>	<b>T:1</b>
Drug-Resistant Pathogens Are a Growing Problem	1291		
Summary	1294		
Problems	1294		
References	1296		
<b>Chapter 24 The Innate and Adaptive Immune Systems</b>	<b>1297</b>		
<b>THE INNATE IMMUNE SYSTEM</b>	<b>1298</b>		
Epithelial Surfaces Serve as Barriers to Infection	1298		
Pattern Recognition Receptors (PRRs) Recognize Conserved Features of Pathogens	1298		
There Are Multiple Classes of PRRs	1299		
Activated PRRs Trigger an Inflammatory Response at Sites of Infection	1300		
Phagocytic Cells Seek, Engulf, and Destroy Pathogens	1301		
Complement Activation Targets Pathogens for Phagocytosis or Lysis	1302		
Virus-Infected Cells Take Drastic Measures to Prevent Viral Replication	1303		
Natural Killer Cells Induce Virus-Infected Cells to Kill Themselves	1304		
Dendritic Cells Provide the Link Between the Innate and Adaptive Immune Systems	1305		
Summary	1305		



# INTRODUCTION TO THE CELL

CHAPTER

1

## Cells and Genomes

The surface of our planet is populated by living things—curious, intricately organized chemical factories that take in matter from their surroundings and use these raw materials to generate copies of themselves. These living organisms appear extraordinarily diverse. What could be more different than a tiger and a piece of seaweed, or a bacterium and a tree? Yet our ancestors, knowing nothing of cells or DNA, saw that all these things had something in common. They called that something “life,” marveled at it, struggled to define it, and despaired of explaining what it was or how it worked in terms that relate to nonliving matter.

The discoveries of the past century have not diminished the marvel—quite the contrary. But they have removed the central mystery regarding the nature of life. We can now see that all living things are made of cells: small, membrane-enclosed units filled with a concentrated aqueous solution of chemicals and endowed with the extraordinary ability to create copies of themselves by growing and then dividing in two.

Because cells are the fundamental units of life, it is to *cell biology*—the study of the structure, function, and behavior of cells—that we must look for answers to the questions of what life is and how it works. With a deeper understanding of cells and their evolution, we can begin to tackle the grand historical problems of life on Earth: its mysterious origins, its stunning diversity, and its invasion of every conceivable habitat. Indeed, as emphasized long ago by the pioneering cell biologist E. B. Wilson, “the key to every biological problem must finally be sought in the cell; for every living organism is, or at some time has been, a cell.”

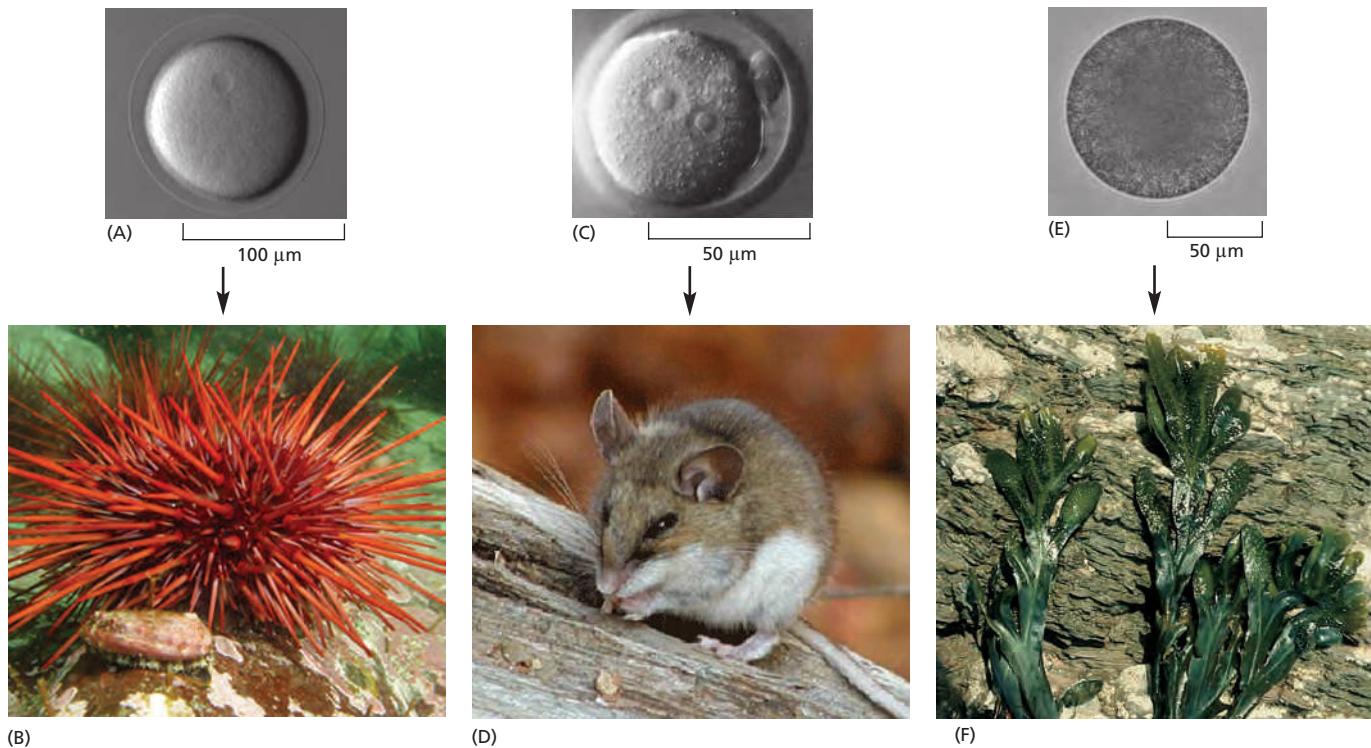
Despite their apparent diversity, living things are fundamentally similar inside. The whole of biology is thus a counterpoint between two themes: astonishing variety in individual particulars; astonishing constancy in fundamental mechanisms. In this first chapter, we begin by outlining the universal features common to all life on our planet. We then survey, briefly, the diversity of cells. And we see how, thanks to the common molecular code in which the specifications for all living organisms are written, it is possible to read, measure, and decipher these specifications to help us achieve a coherent understanding of all the forms of life, from the smallest to the greatest.

### IN THIS CHAPTER

THE UNIVERSAL FEATURES OF CELLS ON EARTH

THE DIVERSITY OF GENOMES AND THE TREE OF LIFE

GENETIC INFORMATION IN EUKARYOTES



## THE UNIVERSAL FEATURES OF CELLS ON EARTH

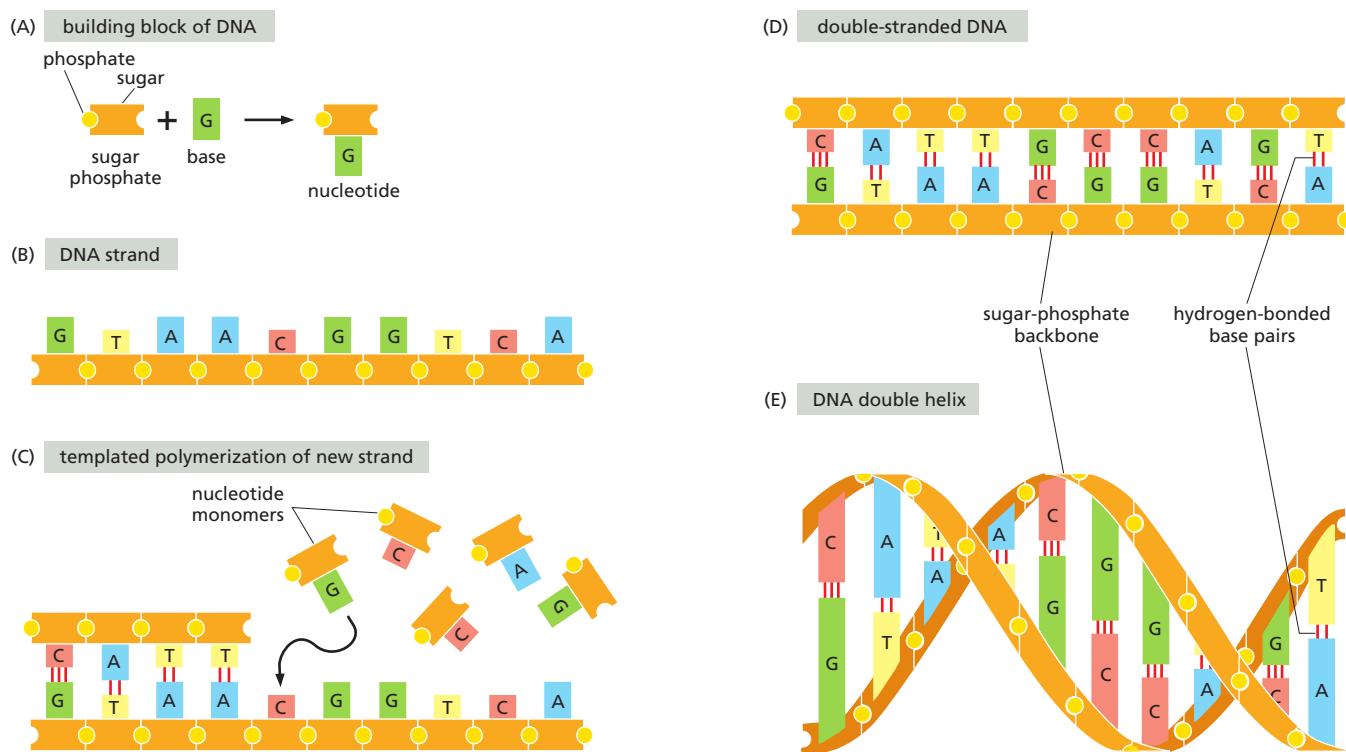
It is estimated that there are more than 10 million—perhaps 100 million—living species on Earth today. Each species is different, and each reproduces itself faithfully, yielding progeny that belong to the same species: the parent organism hands down information specifying, in extraordinary detail, the characteristics that the offspring shall have. This phenomenon of *heredity* is central to the definition of life: it distinguishes life from other processes, such as the growth of a crystal, or the burning of a candle, or the formation of waves on water, in which orderly structures are generated but without the same type of link between the peculiarities of parents and the peculiarities of offspring. Like the candle flame, the living organism must consume free energy to create and maintain its organization. But life employs the free energy to drive a hugely complex system of chemical processes that are specified by hereditary information.

Most living organisms are single cells. Others, such as ourselves, are vast multicellular cities in which groups of cells perform specialized functions linked by intricate systems of communication. But even for the aggregate of more than  $10^{13}$  cells that form a human body, the whole organism has been generated by cell divisions from a single cell. The single cell, therefore, is the vehicle for all of the hereditary information that defines each species (Figure 1-1). This cell includes the machinery to gather raw materials from the environment and to construct from them a new cell in its own image, complete with a new copy of its hereditary information. Each and every cell is truly amazing.

### All Cells Store Their Hereditary Information in the Same Linear Chemical Code: DNA

Computers have made us familiar with the concept of information as a measurable quantity—a million bytes (to record a few hundred pages of text or an image from a digital camera), 600 million bytes for the music on a CD, and so on. Computers have also made us well aware that the same information can be recorded in many different physical forms: the discs and tapes that we used 20 years ago for our electronic archives have become unreadable on present-day machines. Living

**Figure 1-1** The hereditary information in the fertilized egg cell determines the nature of the whole multicellular organism. Although their starting cells look superficially similar, as indicated: a sea urchin egg gives rise to a sea urchin (A and B). A mouse egg gives rise to a mouse (C and D). An egg of the seaweed *Fucus* gives rise to a *Fucus* seaweed (E and F). (A, courtesy of David McClay; B, courtesy of M. Gibbs, Oxford Scientific Films; C, courtesy of Patricia Calarco, from G. Martin, *Science* 209:768–776, 1980. With permission from AAAS; D, courtesy of O. Newman, Oxford Scientific Films; E and F, courtesy of Colin Brownlee.)

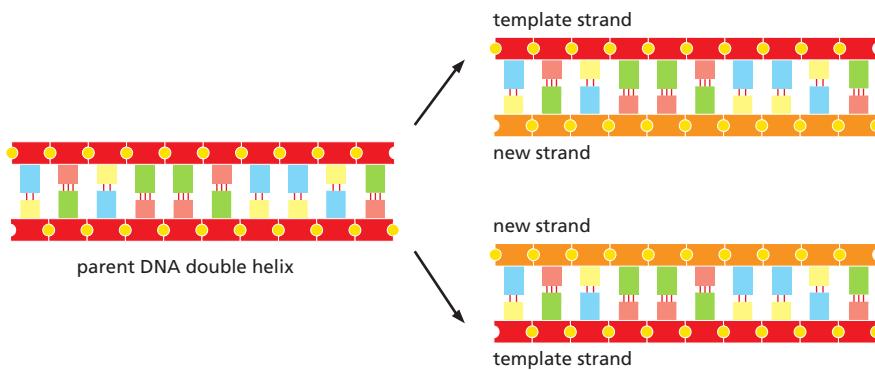


cells, like computers, store information, and it is estimated that they have been evolving and diversifying for over 3.5 billion years. It is scarcely to be expected that they would all store their information in the same form, or that the archives of one type of cell should be readable by the information-handling machinery of another. And yet it is so. All living cells on Earth store their hereditary information in the form of double-stranded molecules of DNA—long, unbranched, paired *polymer* chains, formed always of the same four types of *monomers*. These monomers, chemical compounds known as nucleotides, have nicknames drawn from a four-letter alphabet—A, T, C, G—and they are strung together in a long linear sequence that encodes the genetic information, just as the sequence of 1s and 0s encodes the information in a computer file. We can take a piece of DNA from a human cell and insert it into a bacterium, or a piece of bacterial DNA and insert it into a human cell, and the information will be successfully read, interpreted, and copied. Using chemical methods, scientists have learned how to read out the complete sequence of monomers in any DNA molecule—extending for many millions of nucleotides—and thereby decipher all of the hereditary information that each organism contains.

### All Cells Replicate Their Hereditary Information by Templated Polymerization

The mechanisms that make life possible depend on the structure of the double-stranded DNA molecule. Each monomer in a single DNA strand—that is, each **nucleotide**—consists of two parts: a sugar (deoxyribose) with a phosphate group attached to it, and a *base*, which may be either adenine (A), guanine (G), cytosine (C), or thymine (T) (Figure 1–2). Each sugar is linked to the next via the phosphate group, creating a polymer chain composed of a repetitive sugar-phosphate backbone with a series of bases protruding from it. The DNA polymer is extended by adding monomers at one end. For a single isolated strand, these monomers can, in principle, be added in any order, because each one links to the next in the same way, through the part of the molecule that is the same for all of them. In the living cell, however, DNA is not synthesized as a free strand in isolation, but on a template formed by a preexisting DNA strand. The bases protruding from the

**Figure 1–2** DNA and its building blocks. (A) DNA is made from simple subunits, called nucleotides, each consisting of a sugar-phosphate molecule with a nitrogen-containing side group, or base, attached to it. The bases are of four types (adenine, guanine, cytosine, and thymine), corresponding to four distinct nucleotides, labeled A, G, C, and T. (B) A single strand of DNA consists of nucleotides joined together by sugar-phosphate linkages. Note that the individual sugar-phosphate units are asymmetric, giving the backbone of the strand a definite directionality, or polarity. This directionality guides the molecular processes by which the information in DNA is interpreted and copied in cells: the information is always “read” in a consistent order, just as written English text is read from left to right. (C) Through templated polymerization, the sequence of nucleotides in an existing DNA strand controls the sequence in which nucleotides are joined together in a new DNA strand; T in one strand pairs with A in the other, and G in one strand with C in the other. The new strand has a nucleotide sequence *complementary* to that of the old strand, and a backbone with opposite directionality: corresponding to the GTAA... of the original strand, it has ...TTAC. (D) A normal DNA molecule consists of two such complementary strands. The nucleotides within each strand are linked by strong (covalent) chemical bonds; the complementary nucleotides on opposite strands are held together more weakly, by hydrogen bonds. (E) The two strands twist around each other to form a double helix—a robust structure that can accommodate any sequence of nucleotides without altering its basic structure (see Movie 4.1).



**Figure 1–3** The copying of genetic information by DNA replication. In this process, the two strands of a DNA double helix are pulled apart, and each serves as a template for synthesis of a new complementary strand.

existing strand bind to bases of the strand being synthesized, according to a strict rule defined by the complementary structures of the bases: A binds to T, and C binds to G. This base-pairing holds fresh monomers in place and thereby controls the selection of which one of the four monomers shall be added to the growing strand next. In this way, a double-stranded structure is created, consisting of two exactly complementary sequences of As, Cs, Ts, and Gs. The two strands twist around each other, forming a DNA double helix (Figure 1–2E).

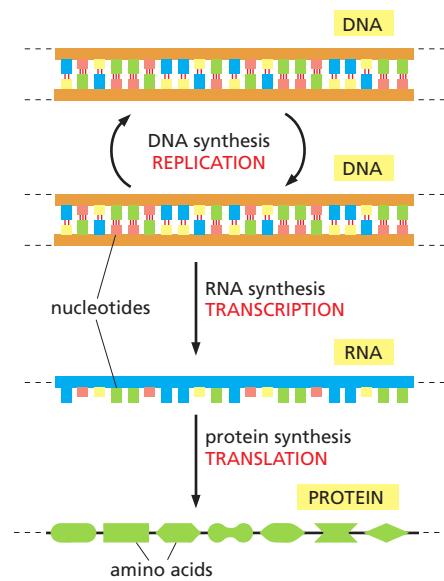
The bonds between the base pairs are weak compared with the sugar-phosphate links, and this allows the two DNA strands to be pulled apart without breakage of their backbones. Each strand then can serve as a template, in the way just described, for the synthesis of a fresh DNA strand complementary to itself—a fresh copy, that is, of the hereditary information (Figure 1–3). In different types of cells, this process of **DNA replication** occurs at different rates, with different controls to start it or stop it, and different auxiliary molecules to help it along. But the basics are universal: DNA is the information store for heredity, and *templated polymerization* is the way in which this information is copied throughout the living world.

### All Cells Transcribe Portions of Their Hereditary Information into the Same Intermediary Form: RNA

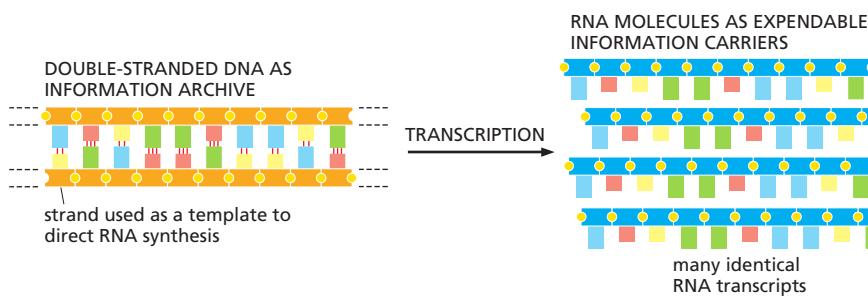
To carry out its information-bearing function, DNA must do more than copy itself. It must also *express* its information, by letting the information guide the synthesis of other molecules in the cell. This expression occurs by a mechanism that is the same in all living organisms, leading first and foremost to the production of two other key classes of polymers: RNAs and proteins. The process (discussed in detail in Chapters 6 and 7) begins with a templated polymerization called **transcription**, in which segments of the DNA sequence are used as templates for the synthesis of shorter molecules of the closely related polymer **ribonucleic acid**, or RNA. Later, in the more complex process of **translation**, many of these RNA molecules direct the synthesis of polymers of a radically different chemical class—the *proteins* (Figure 1–4).

In RNA, the backbone is formed of a slightly different sugar from that of DNA—ribose instead of deoxyribose—and one of the four bases is slightly different—uracil (U) in place of thymine (T). But the other three bases—A, C, and G—are the same, and all four bases pair with their complementary counterparts in DNA—the A, U, C, and G of RNA with the T, A, G, and C of DNA. During transcription, the RNA monomers are lined up and selected for polymerization on a template strand of DNA, just as DNA monomers are selected during replication. The outcome is a polymer molecule whose sequence of nucleotides faithfully represents a portion of the cell's genetic information, even though it is written in a slightly different alphabet—consisting of RNA monomers instead of DNA monomers.

The same segment of DNA can be used repeatedly to guide the synthesis of many identical RNA molecules. Thus, whereas the cell's archive of genetic information in the form of DNA is fixed and sacrosanct, these *RNA transcripts* are



**Figure 1–4** From DNA to protein. Genetic information is read out and put to use through a two-step process. First, in *transcription*, segments of the DNA sequence are used to guide the synthesis of molecules of RNA. Then, in *translation*, the RNA molecules are used to guide the synthesis of molecules of protein.

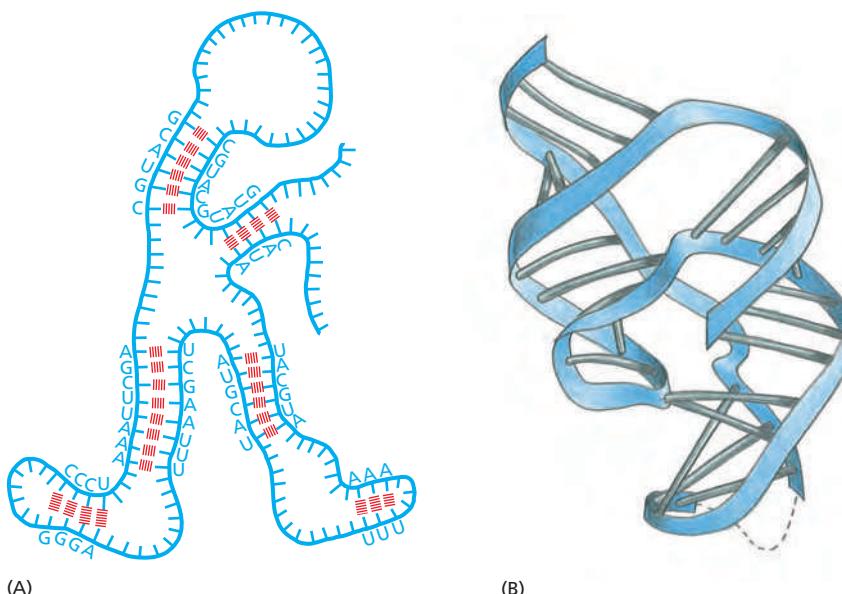


mass-produced and disposable (Figure 1–5). As we shall see, these transcripts function as intermediates in the transfer of genetic information. Most notably, they serve as **messenger RNA (mRNA)** molecules that guide the synthesis of proteins according to the genetic instructions stored in the DNA.

RNA molecules have distinctive structures that can also give them other specialized chemical capabilities. Being single-stranded, their backbone is flexible, so that the polymer chain can bend back on itself to allow one part of the molecule to form weak bonds with another part of the same molecule. This occurs when segments of the sequence are locally complementary: a ...GGGG... segment, for example, will tend to associate with a ...CCCC... segment. These types of internal associations can cause an RNA chain to fold up into a specific shape that is dictated by its sequence (Figure 1–6). The shape of the RNA molecule, in turn, may enable it to recognize other molecules by binding to them selectively—and even, in certain cases, to catalyze chemical changes in the molecules that are bound. In fact, some chemical reactions catalyzed by RNA molecules are crucial for several of the most ancient and fundamental processes in living cells, and it has been suggested that an extensive catalysis by RNA played a central part in the early evolution of life (discussed in Chapter 6).

### All Cells Use Proteins as Catalysts

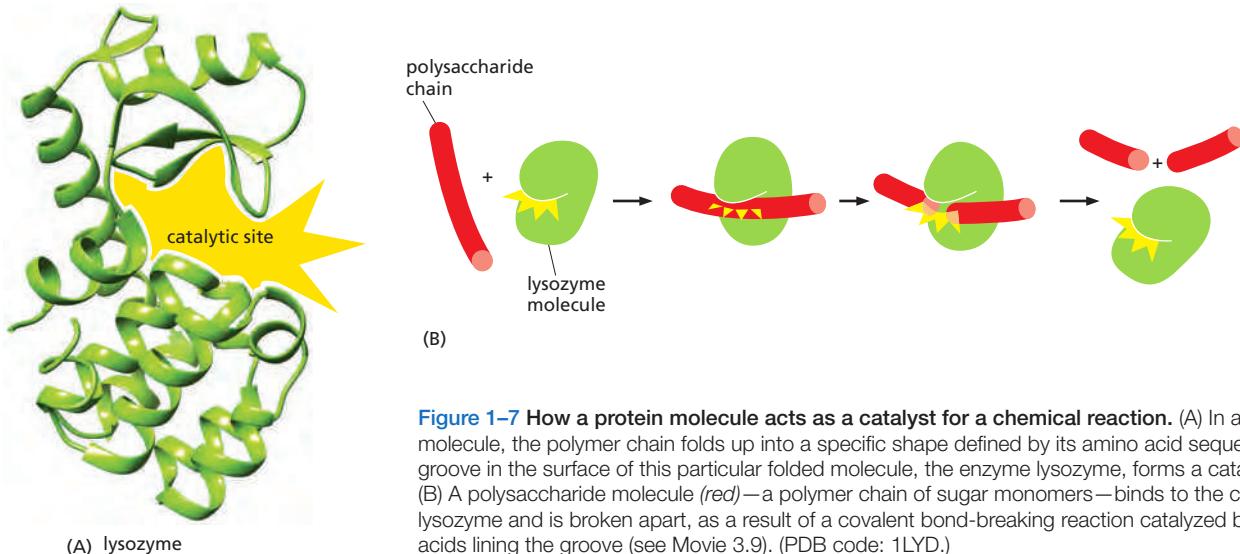
**Protein** molecules, like DNA and RNA molecules, are long unbranched polymer chains, formed by stringing together monomeric building blocks drawn from a standard repertoire that is the same for all living cells. Like DNA and RNA, proteins carry information in the form of a linear sequence of symbols, in the same way as a human message written in an alphabetic script. There are many different protein molecules in each cell, and—leaving out the water—they form most of the cell's mass.



**Figure 1–5** How genetic information is broadcast for use inside the cell.

Each cell contains a fixed set of DNA molecules—its archive of genetic information. A given segment of this DNA guides the synthesis of many identical RNA transcripts, which serve as working copies of the information stored in the archive. Many different sets of RNA molecules can be made by transcribing different parts of a cell's DNA sequences, allowing different types of cells to use the same information store differently.

**Figure 1–6** The conformation of an RNA molecule. (A) Nucleotide pairing between different regions of the same RNA polymer chain causes the molecule to adopt a distinctive shape. (B) The three-dimensional structure of an actual RNA molecule produced by hepatitis delta virus; this RNA can catalyze RNA strand cleavage. The blue ribbon represents the sugar-phosphate backbone and the bars represent base pairs (see Movie 6.1). (B, based on A.R. Ferré-D'Amare, K. Zhou, and J.A. Doudna, *Nature* 395:567–574, 1998. With permission from Macmillan Publishers Ltd.)



**Figure 1–7** How a protein molecule acts as a catalyst for a chemical reaction. (A) In a protein molecule, the polymer chain folds up into a specific shape defined by its amino acid sequence. A groove in the surface of this particular folded molecule, the enzyme lysozyme, forms a catalytic site. (B) A polysaccharide molecule (red)—a polymer chain of sugar monomers—binds to the catalytic site of lysozyme and is broken apart, as a result of a covalent bond-breaking reaction catalyzed by the amino acids lining the groove (see Movie 3.9). (PDB code: 1LYD.)

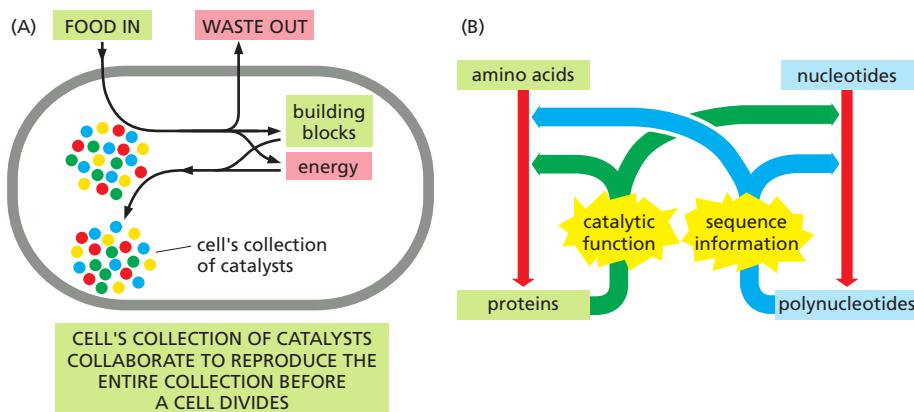
The monomers of protein, the **amino acids**, are quite different from those of DNA and RNA, and there are 20 types instead of 4. Each amino acid is built around the same core structure through which it can be linked in a standard way to any other amino acid in the set; attached to this core is a side group that gives each amino acid a distinctive chemical character. Each of the protein molecules is a **polypeptide**, created by joining its amino acids in a particular sequence. Through billions of years of evolution, this sequence has been selected to give the protein a useful function. Thus, by folding into a precise three-dimensional form with reactive sites on its surface (Figure 1–7A), these amino-acid polymers can bind with high specificity to other molecules and can act as **enzymes** to catalyze reactions that make or break covalent bonds. In this way they direct the vast majority of chemical processes in the cell (Figure 1–7B).

Proteins have many other functions as well—maintaining structures, generating movements, sensing signals, and so on—each protein molecule performing a specific function according to its own genetically specified sequence of amino acids. Proteins, above all, are the main molecules that put the cell's genetic information into action.

Thus, polynucleotides specify the amino acid sequences of proteins. Proteins, in turn, catalyze many chemical reactions, including those by which new DNA molecules are synthesized. From the most fundamental point of view, a living cell is a self-replicating collection of catalysts that takes in food, processes this food to derive both the building blocks and energy needed to make more catalysts, and discards the materials left over as waste (Figure 1–8A). A feedback loop that connects proteins and polynucleotides forms the basis for this autocatalytic, self-reproducing behavior of living organisms (Figure 1–8B).

### All Cells Translate RNA into Protein in the Same Way

How the information in DNA specifies the production of proteins was a complete mystery in the 1950s when the double-stranded structure of DNA was first revealed as the basis of heredity. But in the intervening years, scientists have discovered the elegant mechanisms involved. The translation of genetic information from the 4-letter alphabet of polynucleotides into the 20-letter alphabet of proteins is a complex process. The rules of this translation seem in some respects neat and rational but in other respects strangely arbitrary, given that they are (with minor exceptions) identical in all living things. These arbitrary features, it is thought, reflect frozen accidents in the early history of life. They stem from the chance properties of the earliest organisms that were passed on by heredity and have become so deeply embedded in the constitution of all living cells that they cannot be changed without disastrous effects.



**Figure 1–8** Life as an autocatalytic process. (A) The cell as a self-replicating collection of catalysts. (B) Polynucleotides (the nucleic acids DNA and RNA, which are nucleotide polymers) provide the sequence information, while proteins (amino acid polymers) provide most of the catalytic functions that serve—through a complex set of chemical reactions—to bring about the synthesis of more polynucleotides and proteins of the same types.

It turns out that the information in the sequence of a messenger RNA molecule is read out in groups of three nucleotides at a time: each triplet of nucleotides, or *codon*, specifies (codes for) a single amino acid in a corresponding protein. Since the number of distinct triplets that can be formed from four nucleotides is  $4^3$ , there are 64 possible codons, all of which occur in nature. However, there are only 20 naturally occurring amino acids. That means there are necessarily many cases in which several codons correspond to the same amino acid. This *genetic code* is read out by a special class of small RNA molecules, the *transfer RNAs* (*tRNAs*). Each type of tRNA becomes attached at one end to a specific amino acid, and displays at its other end a specific sequence of three nucleotides—an *anticodon*—that enables it to recognize, through base-pairing, a particular codon or subset of codons in mRNA. The intricate chemistry that enables these tRNAs to translate a specific sequence of A, C, G, and U nucleotides in an mRNA molecule into a specific sequence of amino acids in a protein molecule occurs on the *ribosome*, a large multimolecular machine composed of both protein and *ribosomal RNA*. All of these processes are described in detail in Chapter 6.

### Each Protein Is Encoded by a Specific Gene

DNA molecules as a rule are very large, containing the specifications for thousands of proteins. Special sequences in the DNA serve as punctuation, defining where the information for each protein begins and ends. And individual segments of the long DNA sequence are transcribed into separate mRNA molecules, coding for different proteins. Each such DNA segment represents one **gene**. A complication is that RNA molecules transcribed from the same DNA segment can often be processed in more than one way, so as to give rise to a set of alternative versions of a protein, especially in more complex cells such as those of plants and animals. In addition, some DNA segments—a smaller number—are transcribed into RNA molecules that are not translated but have catalytic, regulatory, or structural functions; such DNA segments also count as genes. A gene therefore is defined as the segment of DNA sequence corresponding to a single protein or set of alternative protein variants or to a single catalytic, regulatory, or structural RNA molecule.

In all cells, the *expression* of individual genes is regulated: instead of manufacturing its full repertoire of possible proteins at full tilt all the time, the cell adjusts the rate of transcription and translation of different genes independently, according to need. Stretches of *regulatory DNA* are interspersed among the segments that code for protein, and these noncoding regions bind to special protein molecules that control the local rate of transcription. The quantity and organization of the regulatory DNA vary widely from one class of organisms to another, but the basic strategy is universal. In this way, the **genome** of the cell—that is, the totality of its genetic information as embodied in its complete DNA sequence—dictates not only the nature of the cell's proteins, but also when and where they are to be made.

## Life Requires Free Energy

A living cell is a dynamic chemical system, operating far from chemical equilibrium. For a cell to grow or to make a new cell in its own image, it must take in free energy from the environment, as well as raw materials, to drive the necessary synthetic reactions. This consumption of free energy is fundamental to life. When it stops, a cell decays toward chemical equilibrium and soon dies.

Genetic information is also fundamental to life, and free energy is required for the propagation of this information. For example, to specify one bit of information—that is, one yes/no choice between two equally probable alternatives—costs a defined amount of free energy that can be calculated. The quantitative relationship involves some deep reasoning and depends on a precise definition of the term “free energy,” as explained in Chapter 2. The basic idea, however, is not difficult to understand intuitively.

Picture the molecules in a cell as a swarm of objects endowed with thermal energy, moving around violently at random, buffeted by collisions with one another. To specify genetic information—in the form of a DNA sequence, for example—molecules from this wild crowd must be captured, arranged in a specific order defined by some preexisting template, and linked together in a fixed relationship. The bonds that hold the molecules in their proper places on the template and join them together must be strong enough to resist the disordering effect of thermal motion. The process is driven forward by consumption of free energy, which is needed to ensure that the correct bonds are made, and made robustly. In the simplest case, the molecules can be compared with spring-loaded traps, ready to snap into a more stable, lower-energy attached state when they meet their proper partners; as they snap together into the bonded arrangement, their available stored energy—their free energy—like the energy of the spring in the trap, is released and dissipated as heat. In a cell, the chemical processes underlying information transfer are more complex, but the same basic principle applies: free energy has to be spent on the creation of order.

To replicate its genetic information faithfully, and indeed to make all its complex molecules according to the correct specifications, the cell therefore requires free energy, which has to be imported somehow from the surroundings. As we shall see in Chapter 2, the free energy required by animal cells is derived from chemical bonds in food molecules that the animals eat, while plants get their free energy from sunlight.

## All Cells Function as Biochemical Factories Dealing with the Same Basic Molecular Building Blocks

Because all cells make DNA, RNA, and protein, all cells have to contain and manipulate a similar collection of small molecules, including simple sugars, nucleotides, and amino acids, as well as other substances that are universally required. All cells, for example, require the phosphorylated nucleotide *ATP* (*adenosine triphosphate*), not only as a building block for the synthesis of DNA and RNA, but also as a carrier of the free energy that is needed to drive a huge number of chemical reactions in the cell.

Although all cells function as biochemical factories of a broadly similar type, many of the details of their small-molecule transactions differ. Some organisms, such as plants, require only the simplest of nutrients and harness the energy of sunlight to make all their own small organic molecules. Other organisms, such as animals, feed on living things and must obtain many of their organic molecules ready-made. We return to this point later.

## All Cells Are Enclosed in a Plasma Membrane Across Which Nutrients and Waste Materials Must Pass

Another universal feature is that each cell is enclosed by a membrane—the **plasma membrane**. This container acts as a selective barrier that enables the cell to concentrate nutrients gathered from its environment and retain the products it

**Figure 1–9 Formation of a membrane by amphiphilic phospholipid molecules.** Phospholipids have a hydrophilic (water-loving, phosphate) head group and a hydrophobic (water-avoiding, hydrocarbon) tail. At an interface between oil and water, they arrange themselves as a single sheet with their head groups facing the water and their tail groups facing the oil. But when immersed in water, they aggregate to form bilayers enclosing aqueous compartments, as indicated.

synthesizes for its own use, while excreting its waste products. Without a plasma membrane, the cell could not maintain its integrity as a coordinated chemical system.

The molecules that form a membrane have the simple physicochemical property of being *amphiphilic*—that is, consisting of one part that is hydrophobic (water-insoluble) and another part that is hydrophilic (water-soluble). Such molecules placed in water aggregate spontaneously, arranging their hydrophobic portions to be as much in contact with one another as possible to hide them from the water, while keeping their hydrophilic portions exposed. Amphiphilic molecules of appropriate shape, such as the phospholipid molecules that comprise most of the plasma membrane, spontaneously aggregate in water to create a *bilayer* that forms small closed vesicles (Figure 1–9). The phenomenon can be demonstrated in a test tube by simply mixing phospholipids and water together; under appropriate conditions, small vesicles form whose aqueous contents are isolated from the external medium.

Although the chemical details vary, the hydrophobic tails of the predominant membrane molecules in all cells are hydrocarbon polymers ( $-\text{CH}_2-\text{CH}_2-\text{CH}_2-$ ), and their spontaneous assembly into a bilayered vesicle is but one of many examples of an important general principle: cells produce molecules whose chemical properties cause them to *self-assemble* into the structures that a cell needs.

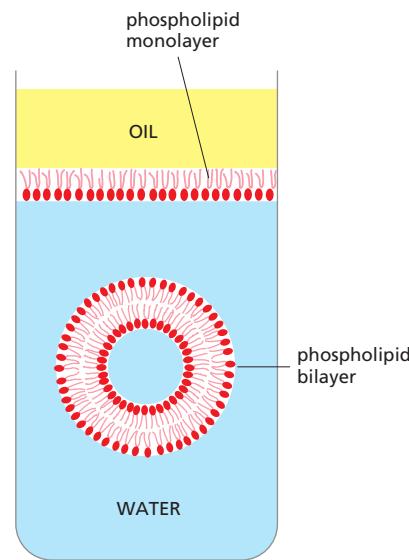
The cell boundary cannot be totally impermeable. If a cell is to grow and reproduce, it must be able to import raw materials and export waste across its plasma membrane. All cells therefore have specialized proteins embedded in their membrane that transport specific molecules from one side to the other. Some of these *membrane transport proteins*, like some of the proteins that catalyze the fundamental small-molecule reactions inside the cell, have been so well preserved over the course of evolution that we can recognize the family resemblances between them in comparisons of even the most distantly related groups of living organisms.

The transport proteins in the membrane largely determine which molecules enter the cell, and the catalytic proteins inside the cell determine the reactions that those molecules undergo. Thus, by specifying the proteins that the cell is to manufacture, the genetic information recorded in the DNA sequence dictates the entire chemistry of the cell; and not only its chemistry, but also its form and its behavior, for these too are chiefly constructed and controlled by the cell's proteins.

### A Living Cell Can Exist with Fewer Than 500 Genes

The basic principles of biological information transfer are simple enough, but how complex are real living cells? In particular, what are the minimum requirements? We can get a rough indication by considering a species that has one of the smallest known genomes—the bacterium *Mycoplasma genitalium* (Figure 1–10). This organism lives as a parasite in mammals, and its environment provides it with many of its small molecules ready-made. Nevertheless, it still has to make all the large molecules—DNA, RNAs, and proteins—required for the basic processes of heredity. It has about 530 genes, about 400 of which are essential. Its genome of 580,070 nucleotide pairs represents 145,018 bytes of information—about as much as it takes to record the text of one chapter of this book. Cell biology may be complicated, but it is not impossibly so.

The minimum number of genes for a viable cell in today's environments is probably not less than 300, although there are only about 60 genes in the core set that is shared by all living species.



## Summary

The individual cell is the minimal self-reproducing unit of living matter, and it consists of a self-replicating collection of catalysts. Central to this reproduction is the transmission of genetic information to progeny cells. Every cell on our planet stores its genetic information in the same chemical form—as double-stranded DNA. The cell replicates its information by separating the paired DNA strands and using each as a template for polymerization to make a new DNA strand with a complementary sequence of nucleotides. The same strategy of templated polymerization is used to transcribe portions of the information from DNA into molecules of the closely related polymer, RNA. These RNA molecules in turn guide the synthesis of protein molecules by the more complex machinery of translation, involving a large multi-molecular machine, the ribosome. Proteins are the principal catalysts for almost all the chemical reactions in the cell; their other functions include the selective import and export of small molecules across the plasma membrane that forms the cell's boundary. The specific function of each protein depends on its amino acid sequence, which is specified by the nucleotide sequence of a corresponding segment of the DNA—the gene that codes for that protein. In this way, the genome of the cell determines its chemistry; and the chemistry of every living cell is fundamentally similar, because it must provide for the synthesis of DNA, RNA, and protein. The simplest known cells can survive with about 400 genes.

## THE DIVERSITY OF GENOMES AND THE TREE OF LIFE

The success of living organisms based on DNA, RNA, and protein has been spectacular. Life has populated the oceans, covered the land, infiltrated the Earth's crust, and molded the surface of our planet. Our oxygen-rich atmosphere, the deposits of coal and oil, the layers of iron ores, the cliffs of chalk and limestone and marble—all these are products, directly or indirectly, of past biological activity on Earth.

Living things are not confined to the familiar temperate realm of land, water, and sunlight inhabited by plants and plant-eating animals. They can be found in the darkest depths of the ocean, in hot volcanic mud, in pools beneath the frozen surface of the Antarctic, and buried kilometers deep in the Earth's crust. The creatures that live in these extreme environments are generally unfamiliar, not only because they are inaccessible, but also because they are mostly microscopic. In more homely habitats, too, most organisms are too small for us to see without special equipment: they tend to go unnoticed, unless they cause a disease or rot the timbers of our houses. Yet microorganisms make up most of the total mass of living matter on our planet. Only recently, through new methods of molecular analysis and specifically through the analysis of DNA sequences, have we begun to get a picture of life on Earth that is not grossly distorted by our biased perspective as large animals living on dry land.

In this section, we consider the diversity of organisms and the relationships among them. Because the genetic information for every organism is written in the universal language of DNA sequences, and the DNA sequence of any given organism can be readily obtained by standard biochemical techniques, it is now possible to characterize, catalog, and compare any set of living organisms with reference to these sequences. From such comparisons we can estimate the place of each organism in the family tree of living species—the “tree of life.” But before describing what this approach reveals, we need first to consider the routes by which cells in different environments obtain the matter and energy they require to survive and proliferate, and the ways in which some classes of organisms depend on others for their basic chemical needs.

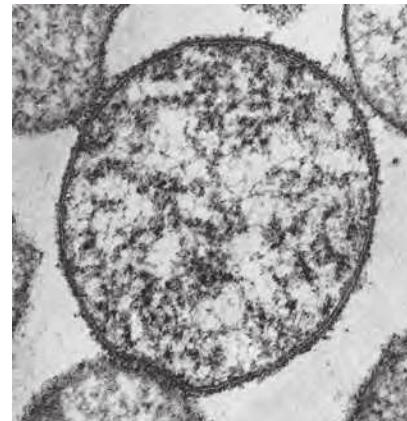
## Cells Can Be Powered by a Variety of Free-Energy Sources

Living organisms obtain their free energy in different ways. Some, such as animals, fungi, and the many different bacteria that live in the human gut, get it by feeding on other living things or the organic chemicals they produce; such organisms



(A)

5 μm



(B)

0.2 μm

**Figure 1–10** *Mycoplasma genitalium*.

(A) Scanning electron micrograph showing the irregular shape of this small bacterium, reflecting the lack of any rigid cell wall. (B) Cross section (transmission electron micrograph) of a *Mycoplasma* cell. Of the 530 genes of *Mycoplasma genitalium*, 43 code for transfer, ribosomal, and other non-messenger RNAs. Functions are known, or can be guessed, for 339 of the genes coding for protein: of these, 154 are involved in replication, transcription, translation, and related processes involving DNA, RNA, and protein; 98 in the membrane and surface structures of the cell; 46 in the transport of nutrients and other molecules across the membrane; 71 in energy conversion and the synthesis and degradation of small molecules; and 12 in the regulation of cell division and other processes. Note that these categories are partly overlapping, so that some genes feature twice. (A, from S. Razin et al., *Infect. Immun.* 30:538–546, 1980. With permission from the American Society for Microbiology; B, courtesy of Roger Cole, in *Medical Microbiology*, 4th ed. [S. Baron ed.]. Galveston: University of Texas Medical Branch, 1996.)

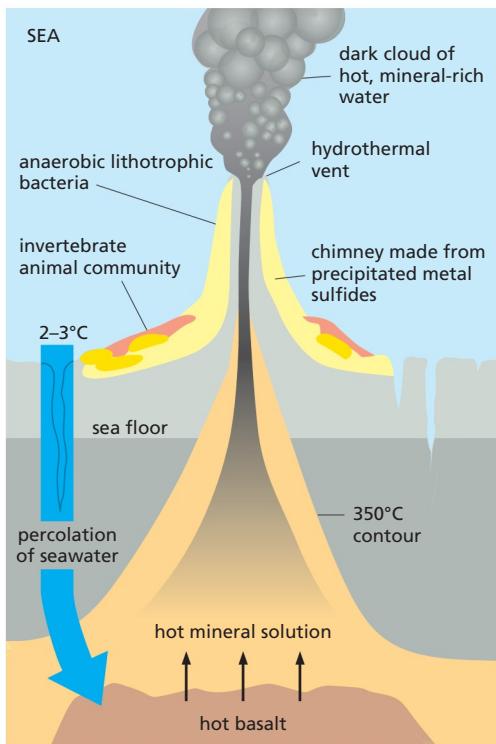
are called *organotrophic* (from the Greek word *trophe*, meaning “food”). Others derive their energy directly from the nonliving world. These primary energy converters fall into two classes: those that harvest the energy of sunlight, and those that capture their energy from energy-rich systems of inorganic chemicals in the environment (chemical systems that are far from chemical equilibrium). Organisms of the former class are called *phototrophic* (feeding on sunlight); those of the latter are called *lithotrophic* (feeding on rock). Organotrophic organisms could not exist without these primary energy converters, which are the most plentiful form of life.

Phototrophic organisms include many types of bacteria, as well as algae and plants, on which we—and virtually all the living things that we ordinarily see around us—depend. Phototrophic organisms have changed the whole chemistry of our environment: the oxygen in the Earth’s atmosphere is a by-product of their biosynthetic activities.

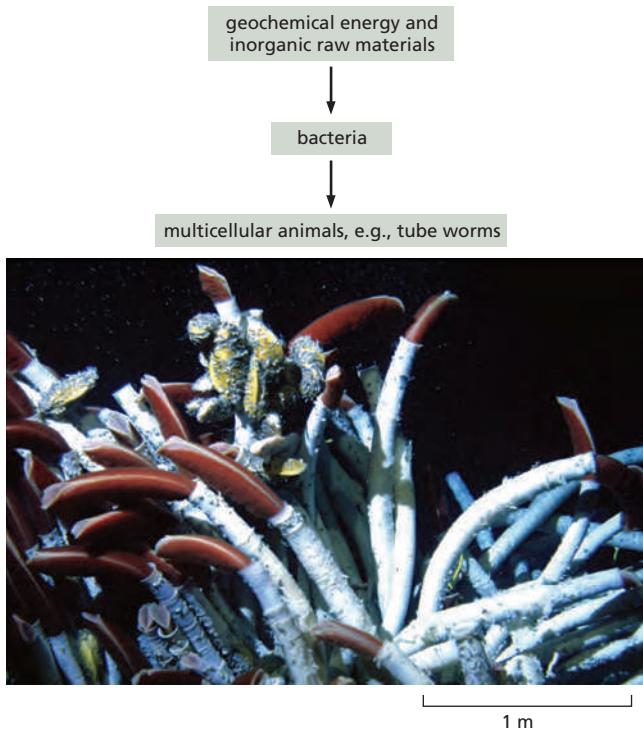
Lithotrophic organisms are not such an obvious feature of our world, because they are microscopic and mostly live in habitats that humans do not frequent—deep in the ocean, buried in the Earth’s crust, or in various other inhospitable environments. But they are a major part of the living world, and they are especially important in any consideration of the history of life on Earth.

Some lithotrophs get energy from *aerobic* reactions, which use molecular oxygen from the environment; since atmospheric O<sub>2</sub> is ultimately the product of living organisms, these aerobic lithotrophs are, in a sense, feeding on the products of past life. There are, however, other lithotrophs that live anaerobically, in places where little or no molecular oxygen is present. These are circumstances similar to those that existed in the early days of life on Earth, before oxygen had accumulated.

The most dramatic of these sites are the hot *hydrothermal vents* on the floor of the Pacific and Atlantic Oceans. They are located where the ocean floor is spreading as new portions of the Earth’s crust form by a gradual upwelling of material from the Earth’s interior (Figure 1–11). Downward-percolating seawater is heated and driven back upward as a submarine geyser, carrying with it a current of chemicals from the hot rocks below. A typical cocktail might include H<sub>2</sub>S, H<sub>2</sub>, CO, Mn<sup>2+</sup>, Fe<sup>2+</sup>, Ni<sup>2+</sup>, CH<sub>4</sub>, NH<sub>4</sub><sup>+</sup>, and phosphorus-containing compounds. A dense



**Figure 1–11** The geology of a hot hydrothermal vent in the ocean floor. As indicated, water percolates down toward the hot molten rock upwelling from the Earth’s interior and is heated and driven back upward, carrying minerals leached from the hot rock. A temperature gradient is set up, from more than 350°C near the core of the vent, down to 2–3°C in the surrounding ocean. Minerals precipitate from the water as it cools, forming a chimney. Different classes of organisms, thriving at different temperatures, live in different neighborhoods of the chimney. A typical chimney might be a few meters tall, spewing out hot, mineral-rich water at a flow rate of 1–2 m/sec.



**Figure 1–12** Organisms living at a depth of 2500 meters near a vent in the ocean floor. Close to the vent, at temperatures up to about 120°C, various lithotrophic species of bacteria and archaea (archaeobacteria) live, directly fueled by geochemical energy. A little further away, where the temperature is lower, various invertebrate animals live by feeding on these microorganisms. Most remarkable are these giant (2 meter) tube worms, *Riftia pachyptila*, which, rather than feed on the lithotrophic cells, live in symbiosis with them: specialized organs in the worms harbor huge numbers of symbiotic sulfur-oxidizing bacteria. These bacteria harness geochemical energy and supply nourishment to their hosts, which have no mouth, gut, or anus. The tube worms are thought to have evolved from more conventional animals, and to have become secondarily adapted to life at hydrothermal vents. (Courtesy of Monika Bright, University of Vienna, Austria.)

population of microbes lives in the neighborhood of the vent, thriving on this austere diet and harvesting free energy from reactions between the available chemicals. Other organisms—clams, mussels, and giant marine worms—in turn live off the microbes at the vent, forming an entire ecosystem analogous to the world of plants and animals that we belong to, but powered by geochemical energy instead of light (**Figure 1–12**).

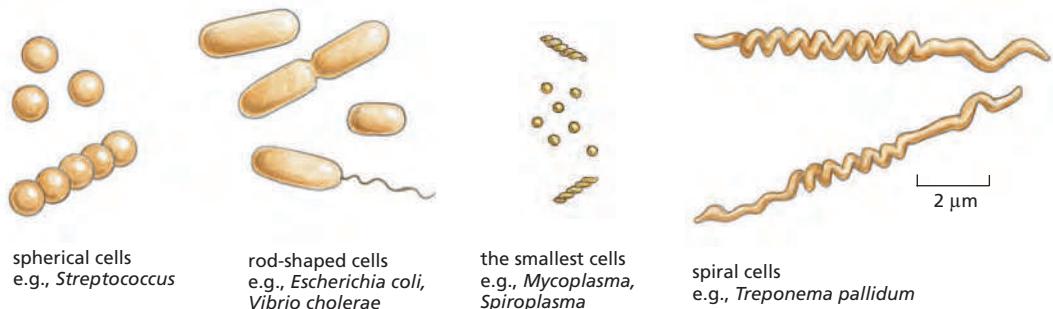
### Some Cells Fix Nitrogen and Carbon Dioxide for Others

To make a living cell requires matter, as well as free energy. DNA, RNA, and protein are composed of just six elements: hydrogen, carbon, nitrogen, oxygen, sulfur, and phosphorus. These are all plentiful in the nonliving environment, in the Earth's rocks, water, and atmosphere. But they are not present in chemical forms that allow easy incorporation into biological molecules. Atmospheric N<sub>2</sub> and CO<sub>2</sub>, in particular, are extremely unreactive. A large amount of free energy is required to drive the reactions that use these inorganic molecules to make the organic compounds needed for further biosynthesis—that is, to *fix* nitrogen and carbon dioxide, so as to make N and C available to living organisms. Many types of living cells lack the biochemical machinery to achieve this fixation; they instead rely on other classes of cells to do the job for them. We animals depend on plants for our supplies of organic carbon and nitrogen compounds. Plants in turn, although they can fix carbon dioxide from the atmosphere, lack the ability to fix atmospheric nitrogen; they depend in part on nitrogen-fixing bacteria to supply their need for nitrogen compounds. Plants of the pea family, for example, harbor symbiotic nitrogen-fixing bacteria in nodules in their roots.

Living cells therefore differ widely in some of the most basic aspects of their biochemistry. Not surprisingly, cells with complementary needs and capabilities have developed close associations. Some of these associations, as we see below, have evolved to the point where the partners have lost their separate identities altogether: they have joined forces to form a single composite cell.

### The Greatest Biochemical Diversity Exists Among Prokaryotic Cells

From simple microscopy, it has long been clear that living organisms can be classified on the basis of cell structure into two groups: the **eukaryotes** and the



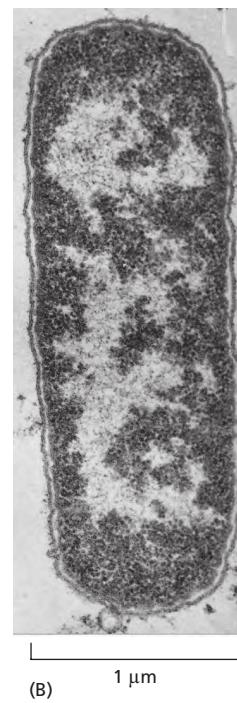
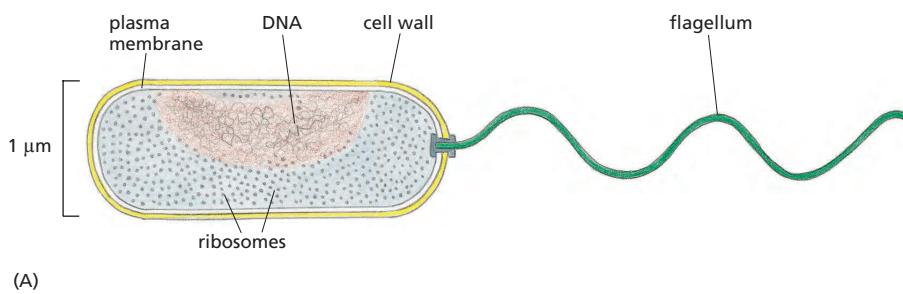
**prokaryotes.** Eukaryotes keep their DNA in a distinct membrane-enclosed intracellular compartment called the nucleus. (The name is from the Greek, meaning “truly nucleated,” from the words *eu*, “well” or “truly,” and *karyon*, “kernel” or “nucleus.”) Prokaryotes have no distinct nuclear compartment to house their DNA. Plants, fungi, and animals are eukaryotes; bacteria are prokaryotes, as are archaea—a separate class of prokaryotic cells, discussed below.

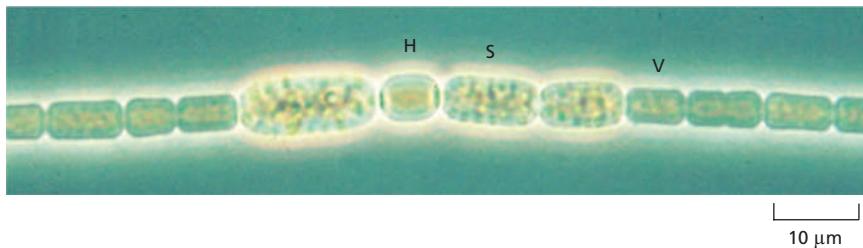
Most prokaryotic cells are small and simple in outward appearance (Figure 1–13), and they live mostly as independent individuals or in loosely organized communities, rather than as multicellular organisms. They are typically spherical or rod-shaped and measure a few micrometers in linear dimension. They often have a tough protective coat, called a *cell wall*, beneath which a plasma membrane encloses a single cytoplasmic compartment containing DNA, RNA, proteins, and the many small molecules needed for life. In the electron microscope, this cell interior appears as a matrix of varying texture without any discernible organized internal structure (Figure 1–14).

Prokaryotic cells live in an enormous variety of ecological niches, and they are astonishingly varied in their biochemical capabilities—far more so than eukaryotic cells. Organotrophic species can utilize virtually any type of organic molecule as food, from sugars and amino acids to hydrocarbons and methane gas. Phototrophic species (Figure 1–15) harvest light energy in a variety of ways, some of them generating oxygen as a by-product, others not. Lithotrophic species can feed on a plain diet of inorganic nutrients, getting their carbon from CO<sub>2</sub>, and relying on H<sub>2</sub>S to fuel their energy needs (Figure 1–16)—or on H<sub>2</sub>, or Fe<sup>2+</sup>, or elemental sulfur, or any of a host of other chemicals that occur in the environment.

**Figure 1–13 Shapes and sizes of some bacteria.** Although most are small, as shown, measuring a few micrometers in linear dimension, there are also some giant species. An extreme example (not shown) is the cigar-shaped bacterium *Epulopiscium fishelsoni*, which lives in the gut of a surgeonfish and can be up to 600 μm long.

**Figure 1–14 The structure of a bacterium.** (A) The bacterium *Vibrio cholerae*, showing its simple internal organization. Like many other species, *Vibrio* has a helical appendage at one end—a flagellum—that rotates as a propeller to drive the cell forward. It can infect the human small intestine to cause cholera; the severe diarrhea that accompanies this disease kills more than 100,000 people a year. (B) An electron micrograph of a longitudinal section through the widely studied bacterium *Escherichia coli* (*E. coli*). The cell’s DNA is concentrated in the lightly stained region. Part of our normal intestinal flora, *E. coli* is related to *Vibrio*, and it has many flagella distributed over its surface that are not visible in this section. (B, courtesy of E. Kellenberger.)





**Figure 1–15** The phototrophic bacterium *Anabaena cylindrica* viewed in the light microscope. The cells of this species form long, multicellular filaments. Most of the cells (labeled V) perform photosynthesis, while others become specialized for nitrogen fixation (labeled H) or develop into resistant spores (labeled S). (Courtesy of Dave G. Adams.)

Much of this world of microscopic organisms is virtually unexplored. Traditional methods of bacteriology have given us an acquaintance with those species that can be isolated and cultured in the laboratory. But DNA sequence analysis of the populations of bacteria and archaea in samples from natural habitats—such as soil or ocean water, or even the human mouth—has opened our eyes to the fact that most species cannot be cultured by standard laboratory techniques. According to one estimate, at least 99% of prokaryotic species remain to be characterized. Detected only by their DNA, it has not yet been possible to grow the vast majority of them in laboratories.

### The Tree of Life Has Three Primary Branches: Bacteria, Archaea, and Eukaryotes

The classification of living things has traditionally depended on comparisons of their outward appearances: we can see that a fish has eyes, jaws, backbone, brain, and so on, just as we do, and that a worm does not; that a rosebush is cousin to an apple tree, but is less similar to a grass. As Darwin showed, we can readily interpret such close family resemblances in terms of evolution from common ancestors, and we can find the remains of many of these ancestors preserved in the fossil record. In this way, it has been possible to begin to draw a family tree of living organisms, showing the various lines of descent, as well as branch points in the history, where the ancestors of one group of species became different from those of another.

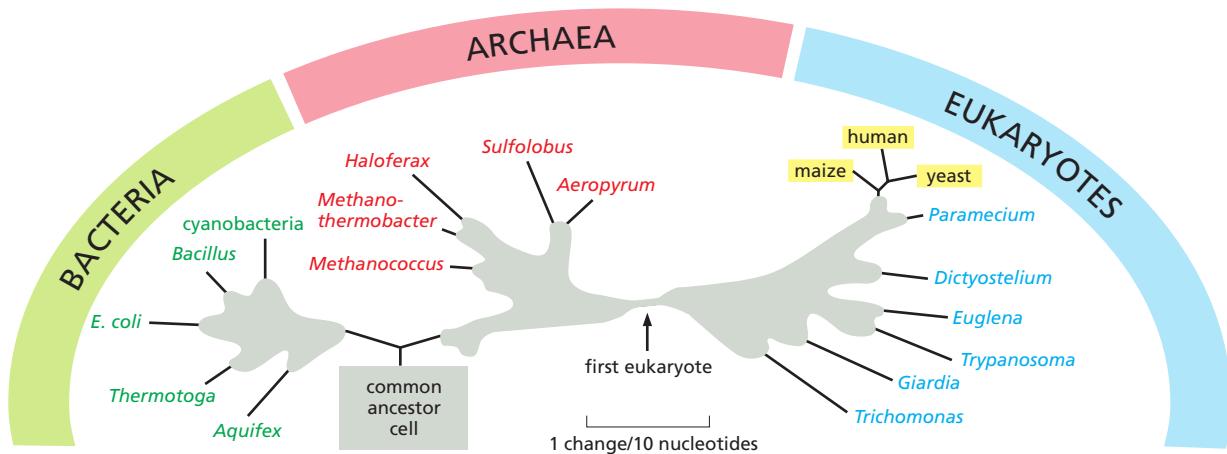
When the disparities between organisms become very great, however, these methods begin to fail. How do we decide whether a fungus is closer kin to a plant or to an animal? When it comes to prokaryotes, the task becomes harder still: one microscopic rod or sphere looks much like another. Microbiologists have therefore sought to classify prokaryotes in terms of their biochemistry and nutritional requirements. But this approach also has its pitfalls. Amid the bewildering variety of biochemical behaviors, it is difficult to know which differences truly reflect differences of evolutionary history.

Genome analysis has now given us a simpler, more direct, and much more powerful way to determine evolutionary relationships. The complete DNA sequence of an organism defines its nature with almost perfect precision and in exhaustive detail. Moreover, this specification is in a digital form—a string of letters—that can be entered straightforwardly into a computer and compared with the corresponding information for any other living thing. Because DNA is subject to random changes that accumulate over long periods of time (as we shall see shortly), the number of differences between the DNA sequences of two organisms can provide a direct, objective, quantitative indication of the evolutionary distance between them.

This approach has shown that the organisms that were traditionally classed together as “bacteria” can be as widely divergent in their evolutionary origins as is any prokaryote from any eukaryote. It is now clear that the prokaryotes comprise two distinct groups that diverged early in the history of life on Earth, before the eukaryotes diverged as a separate group. The two groups of prokaryotes are called the **bacteria** (or eubacteria) and the **archaea** (or archaeabacteria). Detailed genome analyses have recently revealed that the first eukaryotic cell formed after a



**Figure 1–16** A lithotrophic bacterium, *Beggiatoa*, which lives in sulfurous environments, gets its energy by oxidizing  $\text{H}_2\text{S}$  and can fix carbon even in the dark. Note the yellow deposits of sulfur inside the cells. (Courtesy of Ralph W. Wolfe.)



particular type of ancient archaeal cell engulfed an ancient bacterium (see Figure 12–3). Thus, the living world today is considered to consist of three major divisions or *domains*: bacteria, archaea, and eukaryotes (Figure 1–17).

Archaea are often found inhabiting environments that we humans avoid, such as bogs, sewage treatment plants, ocean depths, salt brines, and hot acid springs, although they are also widespread in less extreme and more homely environments, from soils and lakes to the stomachs of cattle. In outward appearance they are not easily distinguished from bacteria. At a molecular level, archaea seem to resemble eukaryotes more closely in their machinery for handling genetic information (replication, transcription, and translation), but bacteria more closely in their apparatus for metabolism and energy conversion. We discuss below how this might be explained.

### Some Genes Evolve Rapidly; Others Are Highly Conserved

Both in the storage and in the copying of genetic information, random accidents and errors occur, altering the nucleotide sequence—that is, creating **mutations**. Therefore, when a cell divides, its two daughters are often not quite identical to one another or to their parent. On rare occasions, the error may represent a change for the better; more probably, it will cause no significant difference in the cell's prospects. But in many cases, the error will cause serious damage—for example, by disrupting the coding sequence for a key protein. Changes due to mistakes of the first type will tend to be perpetuated, because the altered cell has an increased likelihood of reproducing itself. Changes due to mistakes of the second type—*selectively neutral* changes—may be perpetuated or not: in the competition for limited resources, it is a matter of chance whether the altered cell or its cousins will succeed. But changes that cause serious damage lead nowhere: the cell that suffers them dies, leaving no progeny. Through endless repetition of this cycle of error and trial—of *mutation* and *natural selection*—organisms evolve: their genetic specifications change, giving them new ways to exploit the environment more effectively, to survive in competition with others, and to reproduce successfully.

Some parts of the genome will change more easily than others in the course of evolution. A segment of DNA that does not code for protein and has no significant regulatory role is free to change at a rate limited only by the frequency of random errors. In contrast, a gene that codes for a highly optimized essential protein or RNA molecule cannot alter so easily: when mistakes occur, the faulty cells are almost always eliminated. Genes of this latter sort are therefore *highly conserved*. Through 3.5 billion years or more of evolutionary history, many features of the genome have changed beyond all recognition, but the most highly conserved genes remain perfectly recognizable in all living species.

**Figure 1–17** The three major divisions (domains) of the living world. Note that the word *bacteria* was originally used to refer to prokaryotes in general, but more recently has been redefined to refer to eubacteria specifically. The tree shown here is based on comparisons of the nucleotide sequence of a ribosomal RNA (rRNA) subunit in the different species, and the distances in the diagram represent estimates of the numbers of evolutionary changes that have occurred in this molecule in each lineage (see Figure 1–18). The parts of the tree shrouded in gray cloud represent uncertainties about details of the true pattern of species divergence in the course of evolution: comparisons of nucleotide or amino acid sequences of molecules other than rRNA, as well as other arguments, can lead to somewhat different trees. As indicated, the nucleus of the eukaryotic cell is now thought to have emerged from a sub-branch within the archaea, so that in the beginning the tree of life had only two branches—bacteria and archaea.

GTTCCGGGGGAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAAGGAGTGGAGCCTCGGGCTTAATTGACTCAACACGGGAAACCTCACCC	human
GCCGCCTGGGAGTACGGTCGCAAGACTGAAACTTAAAGGAATTGGCGGGGAGCACTACAACGGGTGGAGCCTCGGGTTAATTGATTCAACGCCGGCATCTTACCA	<i>Methanococcus</i>
ACCGCCTGGGAGTACGGCCGCAAGGTTAAACCTCAAATGAATTGACGGGGGCCGC. ACAAGCGGTGGAGCATGTGGTTAACATGCAACCGCGAAGAACCTTACCT	<i>E. coli</i>
GTTCCGGGGGAGTATGGTTGCAAAGCTGAAACTTAAAGGAATTGACGGAAGGGCACCACCAAGGAGTGGAGCCTCGGGCTTAATTGACTCAACACGGGAAACCTCACCC	human

These latter genes are the ones we must examine if we wish to trace family relationships between the most distantly related organisms in the tree of life. The initial studies that led to the classification of the living world into the three domains of bacteria, archaea, and eukaryotes were based chiefly on analysis of one of the rRNA components of the ribosome. Because the translation of RNA into protein is fundamental to all living cells, this component of the ribosome has been very well conserved since early in the history of life on Earth (**Figure 1-18**).

### Most Bacteria and Archaea Have 1000–6000 Genes

Natural selection has generally favored those prokaryotic cells that can reproduce the fastest by taking up raw materials from their environment and replicating themselves most efficiently, at the maximal rate permitted by the available food supplies. Small size implies a large ratio of surface area to volume, thereby helping to maximize the uptake of nutrients across the plasma membrane and boosting a cell's reproductive rate.

Presumably for these reasons, most prokaryotic cells carry very little superfluous baggage; their genomes are small, with genes packed closely together and minimal quantities of regulatory DNA between them. The small genome size has made it easy to use modern DNA sequencing techniques to determine complete genome sequences. We now have this information for thousands of species of bacteria and archaea, as well as for hundreds of species of eukaryotes. Most bacterial and archaeal genomes contain between  $10^6$  and  $10^7$  nucleotide pairs, encoding 1000–6000 genes.

A complete DNA sequence reveals both the genes an organism possesses and the genes it lacks. When we compare the three domains of the living world, we can begin to see which genes are common to all of them and must therefore have been present in the cell that was ancestral to all present-day living things, and which genes are peculiar to a single branch in the tree of life. To explain the findings, however, we need to consider a little more closely how new genes arise and genomes evolve.

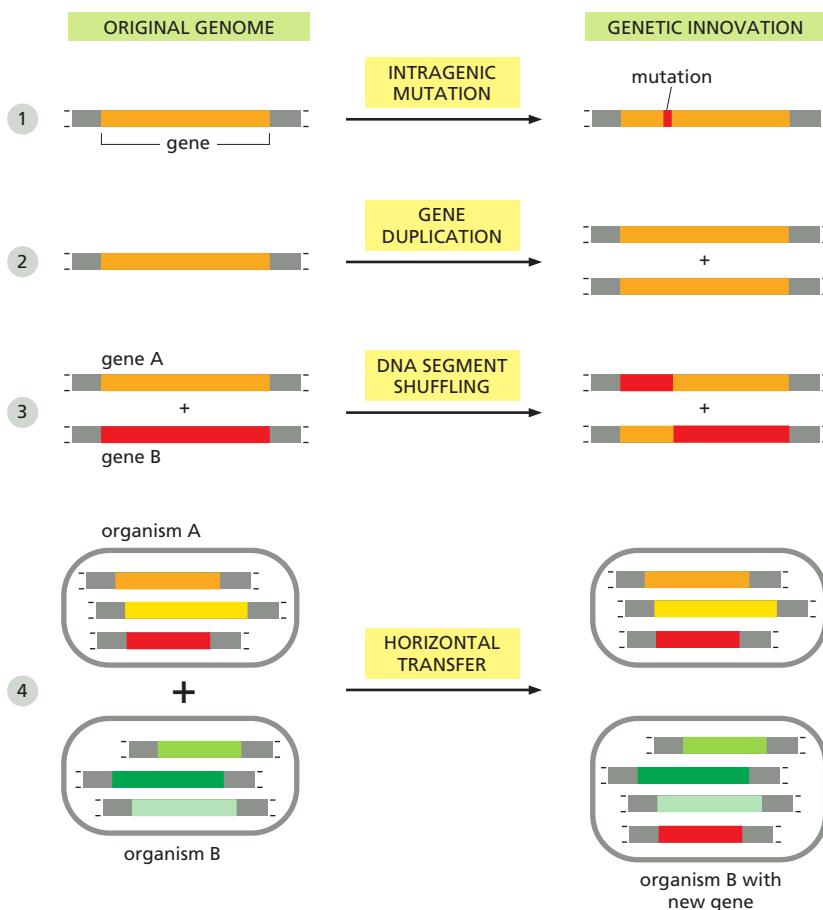
### New Genes Are Generated from Preexisting Genes

The raw material of evolution is the DNA sequence that already exists: there is no natural mechanism for making long stretches of new random sequence. In this sense, no gene is ever entirely new. Innovation can, however, occur in several ways (**Figure 1-19**):

1. *Intragenic mutation*: an existing gene can be randomly modified by changes in its DNA sequence, through various types of error that occur mainly in the process of DNA replication.
2. *Gene duplication*: an existing gene can be accidentally duplicated so as to create a pair of initially identical genes within a single cell; these two genes may then diverge in the course of evolution.
3. *DNA segment shuffling*: two or more existing genes can break and rejoin to make a hybrid gene consisting of DNA segments that originally belonged to separate genes.
4. *Horizontal (intercellular) transfer*: a piece of DNA can be transferred from the genome of one cell to that of another—even to that of another species. This process is in contrast with the usual *vertical transfer* of genetic information from parent to progeny.

Each of these types of change leaves a characteristic trace in the DNA sequence of the organism, and there is clear evidence that all four processes have frequently

**Figure 1-18** Genetic information conserved since the days of the last common ancestor of all living things. A part of the gene for the smaller of the two main rRNA components of the ribosome is shown. (The complete molecule is about 1500–1900 nucleotides long, depending on species.) Corresponding segments of nucleotide sequence from an archaean (*Methanococcus jannaschii*), a bacterium (*Escherichia coli*), and a eukaryote (*Homo sapiens*) are aligned. Sites where the nucleotides are identical between species are indicated by a vertical line; the human sequence is repeated at the bottom of the alignment so that all three two-way comparisons can be seen. A dot halfway along the *E. coli* sequence denotes a site where a nucleotide has been either deleted from the bacterial lineage in the course of evolution or inserted in the other two lineages. Note that the sequences from these three organisms, representative of the three domains of the living world, still retain unmistakable similarities.



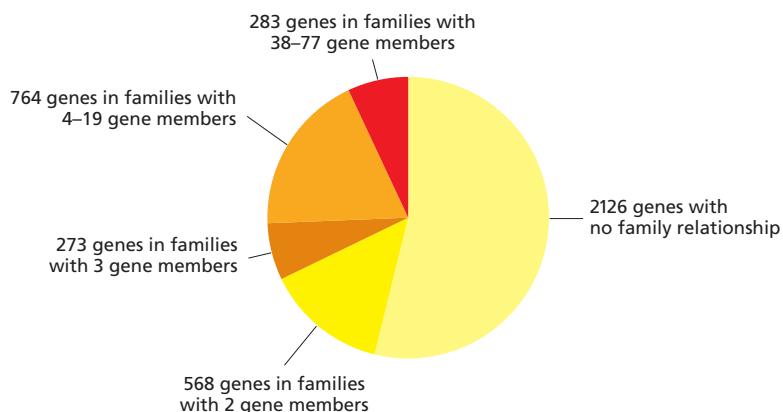
**Figure 1–19** Four modes of genetic innovation and their effects on the DNA sequence of an organism. A special form of horizontal transfer occurs when two different types of cells enter into a permanent symbiotic association. Genes from one of the cells then may be transferred to the genome of the other, as we shall see below when we discuss mitochondria and chloroplasts.

occurred. In later chapters, we discuss the underlying mechanisms, but for the present we focus on the consequences.

### Gene Duplications Give Rise to Families of Related Genes Within a Single Cell

A cell duplicates its entire genome each time it divides into two daughter cells. However, accidents occasionally result in the inappropriate duplication of just part of the genome, with retention of original and duplicate segments in a single cell. Once a gene has been duplicated in this way, one of the two gene copies is free to mutate and become specialized to perform a different function within the same cell. Repeated rounds of this process of duplication and divergence, over many millions of years, have enabled one gene to give rise to a family of genes that may all be found within a single genome. Analysis of the DNA sequence of prokaryotic genomes reveals many examples of such **gene families**: in the bacterium *Bacillus subtilis*, for example, 47% of the genes have one or more obvious relatives (Figure 1–20).

When genes duplicate and diverge in this way, the individuals of one species become endowed with multiple variants of a primordial gene. This evolutionary process has to be distinguished from the genetic divergence that occurs when one species of organism splits into two separate lines of descent at a branch point in the family tree—when the human line of descent became separate from that of chimpanzees, for example. There, the genes gradually become different in the course of evolution, but they are likely to continue to have corresponding functions in the two sister species. Genes that are related by descent in this way—that is, genes in two separate species that derive from the same ancestral gene in the last common ancestor of those two species—are called **orthologs**. Related genes that have resulted from a gene duplication event within a single genome—and



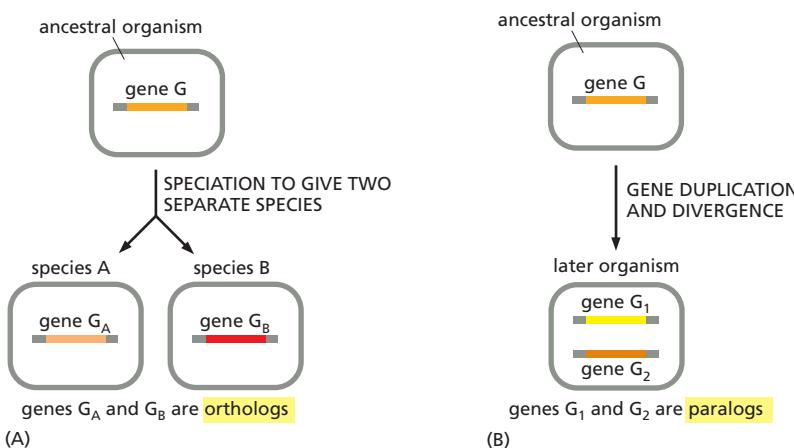
**Figure 1–20** Families of evolutionarily related genes in the genome of *Bacillus subtilis*. The largest gene family in this bacterium consists of 77 genes coding for varieties of ABC transporters—a class of membrane transport proteins found in all three domains of the living world. (Adapted from F. Kunst et al., *Nature* 390:249–256, 1997. With permission from Macmillan Publishers Ltd.)

are likely to have diverged in their function—are called **paralogs**. Genes that are related by descent in either way are called **homologs**, a general term used to cover both types of relationship (Figure 1–21).

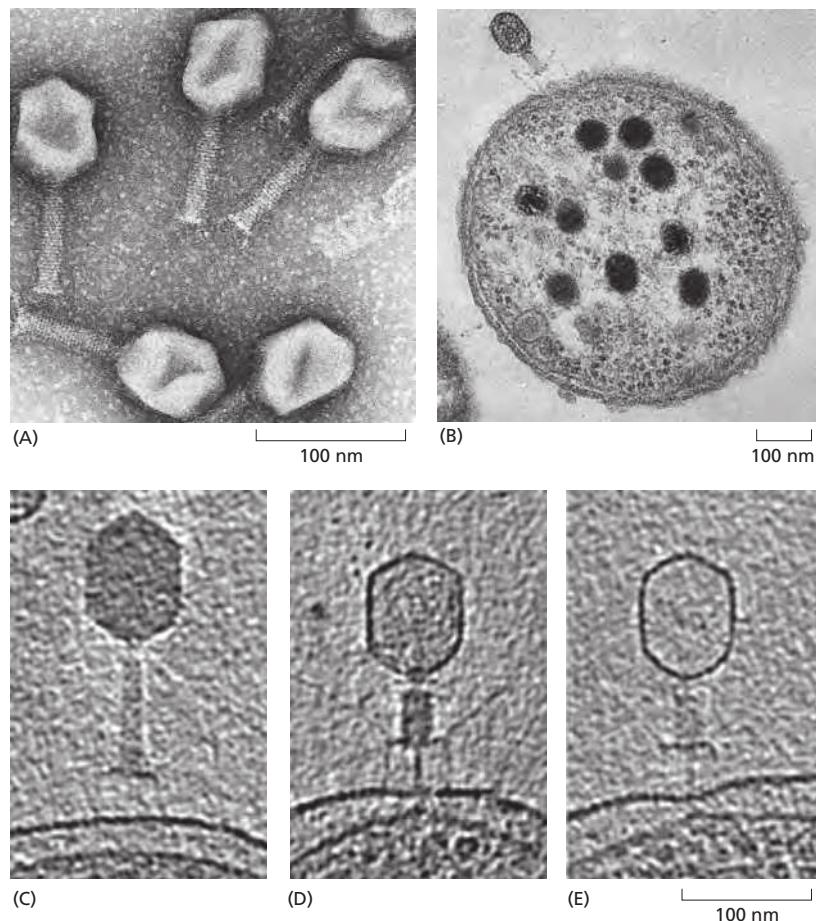
### Genes Can Be Transferred Between Organisms, Both in the Laboratory and in Nature

Prokaryotes provide good examples of the horizontal transfer of genes from one species of cell to another. The most obvious tell-tale signs are sequences recognizable as being derived from viruses, those infecting bacteria being called *bacteriophages* (Figure 1–22). **Viruses** are small packets of genetic material that have evolved as parasites on the reproductive and biosynthetic machinery of host cells. Although not themselves living cells, they often serve as vectors for gene transfer. A virus will replicate in one cell, emerge from it with a protective wrapping, and then enter and infect another cell, which may be of the same or a different species. Often, the infected cell will be killed by the massive proliferation of virus particles inside it; but sometimes, the viral DNA, instead of directly generating these particles, may persist in its host for many cell generations as a relatively innocuous passenger, either as a separate intracellular fragment of DNA, known as a *plasmid*, or as a sequence inserted into the cell's regular genome. In their travels, viruses can accidentally pick up fragments of DNA from the genome of one host cell and ferry them into another cell. Such transfers of genetic material are very common in prokaryotes.

Horizontal transfers of genes between eukaryotic cells of different species are very rare, and they do not seem to have played a significant part in eukaryote evolution (although massive transfers from bacterial to eukaryotic genomes have occurred in the evolution of mitochondria and chloroplasts, as we discuss below).



**Figure 1–21** Paralogous genes and orthologous genes: two types of gene homology based on different evolutionary pathways. (A) Orthologs. (B) Paralogs.



**Figure 1–22** The viral transfer of DNA into a cell. (A) An electron micrograph of particles of a bacterial virus, the T4 bacteriophage. The head of this virus contains the viral DNA; the tail contains the apparatus for injecting the DNA into a host bacterium. (B) A cross section of an *E. coli* bacterium with a T4 bacteriophage latched onto its surface. The large dark objects inside the bacterium are the heads of new T4 particles in the course of assembly. When they are mature, the bacterium will burst open to release them. (C–E) The process of DNA injection into the bacterium, as visualized in unstained, frozen samples by cryoelectron microscopy. (C) Attachment begins. (D) Attached state during DNA injection. (E) Virus head has emptied all of its DNA into the bacterium. (A, courtesy of James Paulson; B, courtesy of Jonathan King and Erika Hartwig from G. Karp, Cell and Molecular Biology, 2nd ed. New York: John Wiley & Sons, 1999. With permission from John Wiley & Sons; C–E, courtesy of Ian Molineux, University of Texas at Austin and Jun Liu, University of Texas Health Science Center, Houston.)

In contrast, horizontal gene transfers occur much more frequently between different species of prokaryotes. Many prokaryotes have a remarkable capacity to take up even nonviral DNA molecules from their surroundings and thereby capture the genetic information these molecules carry. By this route, or by virus-mediated transfer, bacteria and archaea in the wild can acquire genes from neighboring cells relatively easily. Genes that confer resistance to an antibiotic or an ability to produce a toxin, for example, can be transferred from species to species and provide the recipient bacterium with a selective advantage. In this way, new and sometimes dangerous strains of bacteria have been observed to evolve in the bacterial ecosystems that inhabit hospitals or the various niches in the human body. For example, horizontal gene transfer is responsible for the spread, over the past 40 years, of penicillin-resistant strains of *Neisseria gonorrhoeae*, the bacterium that causes gonorrhea. On a longer time scale, the results can be even more profound; it has been estimated that at least 18% of all of the genes in the present-day genome of *E. coli* have been acquired by horizontal transfer from another species within the past 100 million years.

### Sex Results in Horizontal Exchanges of Genetic Information Within a Species

Horizontal gene transfer among prokaryotes has a parallel in a phenomenon familiar to us all: sex. In addition to the usual vertical transfer of genetic material from parent to offspring, sexual reproduction causes a large-scale horizontal transfer of genetic information between two initially separate cell lineages—those of the father and the mother. A key feature of sex, of course, is that the genetic exchange normally occurs only between individuals of the same species. But no matter whether they occur within a species or between species, horizontal gene

transfers leave a characteristic imprint: they result in individuals who are related more closely to one set of relatives with respect to some genes, and more closely to another set of relatives with respect to others. By comparing the DNA sequences of individual human genomes, an intelligent visitor from outer space could deduce that humans reproduce sexually, even if it knew nothing about human behavior.

Sexual reproduction is widespread (although not universal), especially among eukaryotes. Even bacteria indulge from time to time in controlled sexual exchanges of DNA with other members of their own species. Natural selection has clearly favored organisms that can reproduce sexually, although evolutionary theorists dispute precisely what that selective advantage is.

### The Function of a Gene Can Often Be Deduced from Its Sequence

Family relationships among genes are important not just for their historical interest, but because they simplify the task of deciphering gene functions. Once the sequence of a newly discovered gene has been determined, a scientist can tap a few keys on a computer to search the entire database of known gene sequences for genes related to it. In many cases, the function of one or more of these homologs will have been already determined experimentally. Since gene sequence determines gene function, one can frequently make a good guess at the function of the new gene: it is likely to be similar to that of the already known homologs.

In this way, it is possible to decipher a great deal of the biology of an organism simply by analyzing the DNA sequence of its genome and using the information we already have about the functions of genes in other organisms that have been more intensively studied.

### More Than 200 Gene Families Are Common to All Three Primary Branches of the Tree of Life

Given the complete genome sequences of representative organisms from all three domains—archaea, bacteria, and eukaryotes—we can search systematically for homologies that span this enormous evolutionary divide. In this way we can begin to take stock of the common inheritance of all living things. There are considerable difficulties in this enterprise. For example, individual species have often lost some of the ancestral genes; other genes have almost certainly been acquired by horizontal transfer from another species and therefore are not truly ancestral, even though shared. In fact, genome comparisons strongly suggest that both lineage-specific gene loss and horizontal gene transfer, in some cases between evolutionarily distant species, have been major factors of evolution, at least among prokaryotes. Finally, in the course of 2 or 3 billion years, some genes that were initially shared will have changed beyond recognition through mutation.

Because of all these vagaries of the evolutionary process, it seems that only a small proportion of ancestral gene families has been universally retained in a recognizable form. Thus, out of 4873 protein-coding gene families defined by comparing the genomes of 50 species of bacteria, 13 archaea, and 3 unicellular eukaryotes, only 63 are truly ubiquitous (that is, represented in all the genomes analyzed). The great majority of these universal families include components of the translation and transcription systems. This is not likely to be a realistic approximation of an ancestral gene set. A better—though still crude—idea of the latter can be obtained by tallying the gene families that have representatives in multiple, but not necessarily all, species from all three major domains. Such an analysis reveals 264 ancient conserved families. Each family can be assigned a function (at least in terms of general biochemical activity, but usually with more precision). As shown in **Table 1–1**, the largest number of shared gene families are involved in translation and in amino acid metabolism and transport. However, this set of highly conserved gene families represents only a very rough sketch of the common inheritance of all modern life. A more precise reconstruction of the gene complement of the last universal common ancestor will hopefully become feasible with further genome sequencing and more sophisticated forms of comparative analysis.

**TABLE 1–1** The Number of Gene Families, Classified by Function, Common to All Three Domains of the Living World

Information processing		Metabolism	
Translation	63	Energy production and conversion	19
Transcription	7	Carbohydrate transport and metabolism	16
Replication, recombination, and repair	13	Amino acid transport and metabolism	43
Cellular processes and signaling		Nucleotide transport and metabolism	
Cell-cycle control, mitosis, and meiosis	2	Coenzyme transport and metabolism	22
Defense mechanisms	3	Lipid transport and metabolism	9
Signal transduction mechanisms	1	Inorganic ion transport and metabolism	8
Cell wall/membrane biogenesis	2	Secondary metabolite biosynthesis, transport, and catabolism	5
Intracellular trafficking and secretion	4	<b>Poorly characterized</b>	
Post-translational modification, protein turnover, chaperones	8	General biochemical function predicted; specific biological role unknown	24

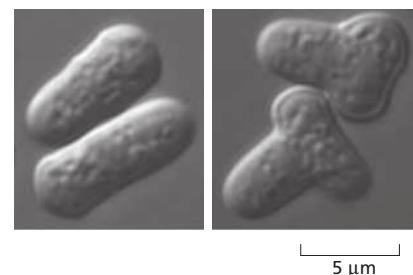
For the purpose of this analysis, gene families are defined as “universal” if they are represented in the genomes of at least two diverse archaea (*Archaeoglobus fulgidus* and *Aeropyrum pernix*), two evolutionarily distant bacteria (*Escherichia coli* and *Bacillus subtilis*), and one eukaryote (yeast, *Saccharomyces cerevisiae*). (Data from R.L. Tatusov, E.V. Koonin and D.J. Lipman, *Science* 278:631–637, 1997; R.L. Tatusov et al., *BMC Bioinformatics* 4:41, 2003; and the COGs database at the US National Library of Medicine.)

## Mutations Reveal the Functions of Genes

Without additional information, no amount of gazing at genome sequences will reveal the functions of genes. We may recognize that gene B is like gene A, but how do we discover the function of gene A in the first place? And even if we know the function of gene A, how do we test whether the function of gene B is truly the same as the sequence similarity suggests? How do we connect the world of abstract genetic information with the world of real living organisms?

The analysis of gene functions depends on two complementary approaches: genetics and biochemistry. Genetics starts with the study of mutants: we either find or make an organism in which a gene is altered, and then examine the effects on the organism’s structure and performance (Figure 1–23). Biochemistry more directly examines the functions of molecules: here we extract molecules from an organism and then study their chemical activities. By combining genetics and biochemistry, it is possible to find those molecules whose production depends on a given gene. At the same time, careful studies of the performance of the mutant organism show us what role those molecules have in the operation of the organism as a whole. Thus, genetics and biochemistry used in combination with cell biology provide the best way to relate genes and molecules to the structure and function of an organism.

In recent years, DNA sequence information and the powerful tools of molecular biology have accelerated progress. From sequence comparisons, we can often identify particular subregions within a gene that have been preserved nearly unchanged over the course of evolution. These conserved subregions are likely to be the most important parts of the gene in terms of function. We can test their individual contributions to the activity of the gene product by creating in the laboratory mutations of specific sites within the gene, or by constructing artificial hybrid genes that combine part of one gene with part of another. Organisms can be engineered to make either the RNA or the protein specified by the gene in large quantities to facilitate biochemical analysis. Specialists in molecular structure can determine the three-dimensional conformation of the gene product, revealing the exact position of every atom in it. Biochemists can determine how each of the



**Figure 1–23** A mutant phenotype reflecting the function of a gene. A normal yeast (of the species *Schizosaccharomyces pombe*) is compared with a mutant in which a change in a single gene has converted the cell from a cigar shape (left) to a T shape (right). The mutant gene therefore has a function in the control of cell shape. But how, in molecular terms, does the gene product perform that function? That is a harder question, and it needs biochemical analysis to answer it. (Courtesy of Kenneth Sawin and Paul Nurse.)

parts of the genetically specified molecule contributes to its chemical behavior. Cell biologists can analyze the behavior of cells that are engineered to express a mutant version of the gene.

There is, however, no one simple recipe for discovering a gene's function, and no simple standard universal format for describing it. We may discover, for example, that the product of a given gene catalyzes a certain chemical reaction, and yet have no idea how or why that reaction is important to the organism. The functional characterization of each new family of gene products, unlike the description of the gene sequences, presents a fresh challenge to the biologist's ingenuity. Moreover, we will never fully understand the function of a gene until we learn its role in the life of the organism as a whole. To make ultimate sense of gene functions, therefore, we have to study whole organisms, not just molecules or cells.

### Molecular Biology Began with a Spotlight on *E. coli*

Because living organisms are so complex, the more we learn about any particular species, the more attractive it becomes as an object for further study. Each discovery raises new questions and provides new tools with which to tackle general questions in the context of the chosen organism. For this reason, large communities of biologists have become dedicated to studying different aspects of the same **model organism**.

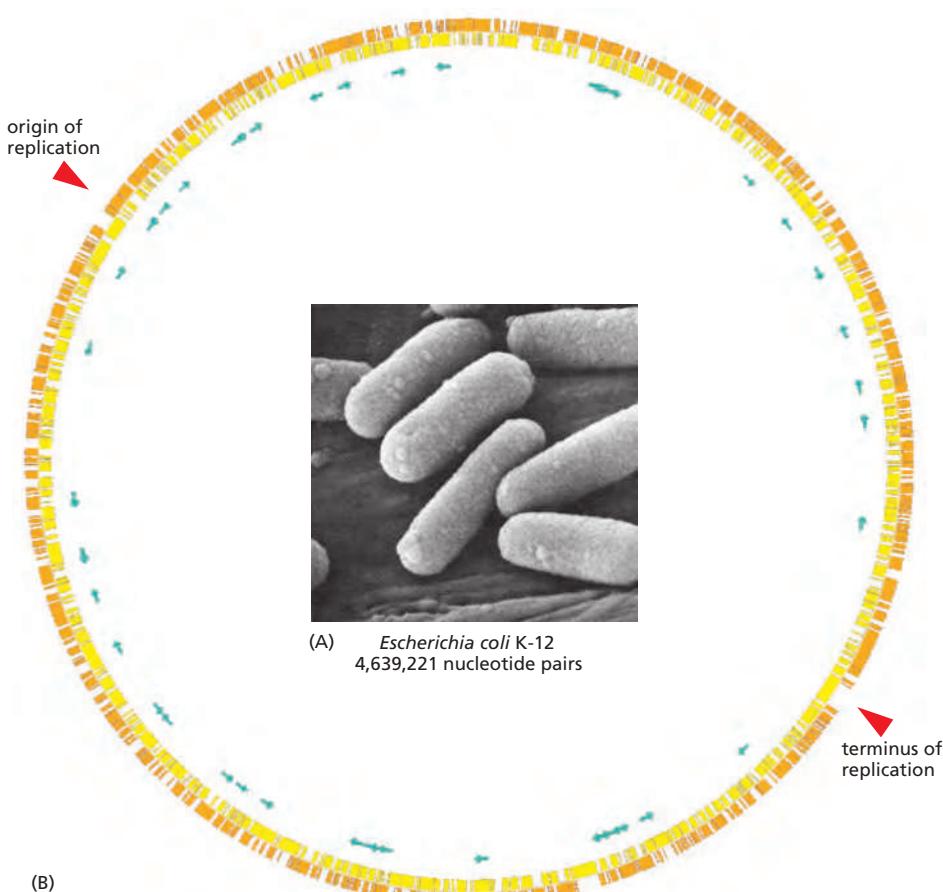
In the early days of molecular biology, the spotlight focused intensely on just one species: the *Escherichia coli*, or *E. coli*, bacterium (see Figures 1–13 and 1–14). This small, rod-shaped bacterial cell normally lives in the gut of humans and other vertebrates, but it can be grown easily in a simple nutrient broth in a culture bottle. It adapts to variable chemical conditions and reproduces rapidly, and it can evolve by mutation and selection at a remarkable speed. As with other bacteria, different strains of *E. coli*, though classified as members of a single species, differ genetically to a much greater degree than do different varieties of a sexually reproducing organism such as a plant or animal. One *E. coli* strain may possess many hundreds of genes that are absent from another, and the two strains could have as little as 50% of their genes in common. The standard laboratory strain *E. coli* K-12 has a genome of approximately 4.6 million nucleotide pairs, contained in a single circular molecule of DNA that codes for about 4300 different kinds of proteins (Figure 1–24).

In molecular terms, we know more about *E. coli* than about any other living organism. Most of our understanding of the fundamental mechanisms of life—for example, how cells replicate their DNA, or how they decode the instructions represented in the DNA to direct the synthesis of specific proteins—initially came from studies of *E. coli*. The basic genetic mechanisms have turned out to be highly conserved throughout evolution: these mechanisms are essentially the same in our own cells as in *E. coli*.

### Summary

Prokaryotes (cells without a distinct nucleus) are biochemically the most diverse organisms and include species that can obtain all their energy and nutrients from inorganic chemical sources, such as the reactive mixtures of minerals released at hydrothermal vents on the ocean floor—the sort of diet that may have nourished the first living cells 3.5 billion years ago. DNA sequence comparisons reveal the family relationships of living organisms and show that the prokaryotes fall into two groups that diverged early in the course of evolution: the bacteria (or eubacteria) and the archaea. Together with the eukaryotes (cells with a membrane-enclosed nucleus), these constitute the three primary branches of the tree of life.

Most bacteria and archaea are small unicellular organisms with compact genomes comprising 1000–6000 genes. Many of the genes within a single organism show strong family resemblances in their DNA sequences, implying that they originated from the same ancestral gene through gene duplication and divergence. Family resemblances (homologies) are also clear when gene sequences are compared between different species, and more than 200 gene families have been so highly



**Figure 1–24** The genome of *E. coli*. (A) A cluster of *E. coli* cells. (B) A diagram of the genome of *E. coli* strain K-12. The diagram is circular because the DNA of *E. coli*, like that of other prokaryotes, forms a single, closed loop. Protein-coding genes are shown as yellow or orange bars, depending on the DNA strand from which they are transcribed; genes encoding only RNA molecules are indicated by green arrows. Some genes are transcribed from one strand of the DNA double helix (in a clockwise direction in this diagram), others from the other strand (counterclockwise). (A, courtesy of Dr. Tony Brain and David Parker/Photo Researchers; B, adapted from F.R. Blattner et al., *Science* 277:1453–1462, 1997.)

conserved that they can be recognized as common to most species from all three domains of the living world. Thus, given the DNA sequence of a newly discovered gene, it is often possible to deduce the gene's function from the known function of a homologous gene in an intensively studied model organism, such as the bacterium *E. coli*.

## GENETIC INFORMATION IN EUKARYOTES

Eukaryotic cells, in general, are bigger and more elaborate than prokaryotic cells, and their genomes are bigger and more elaborate, too. The greater size is accompanied by radical differences in cell structure and function. Moreover, many classes of eukaryotic cells form multicellular organisms that attain levels of complexity unmatched by any prokaryote.

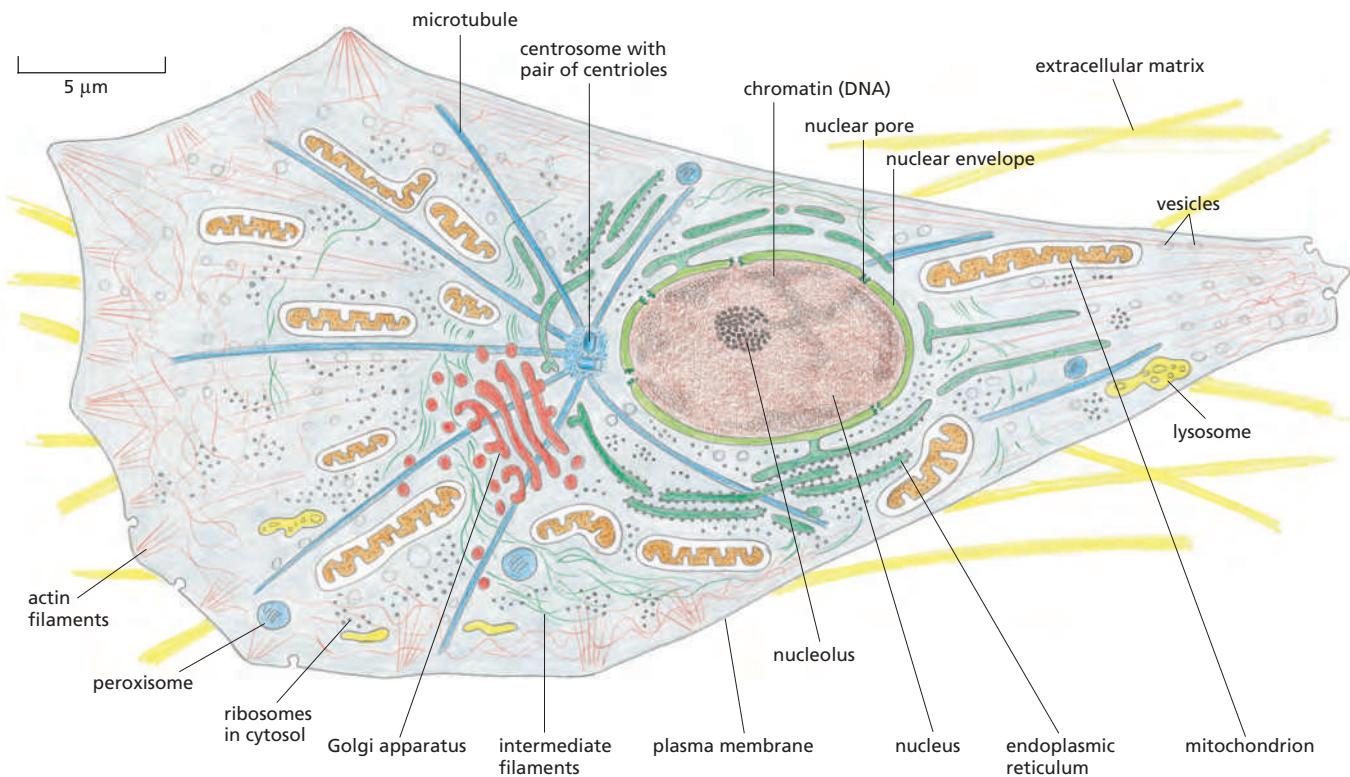
Because they are so complex, eukaryotes confront molecular biologists with a special set of challenges that will concern us in the rest of this book. Increasingly, biologists attempt to meet these challenges through the analysis and manipulation of the genetic information within cells and organisms. It is therefore important at the outset to know something of the special features of the eukaryotic genome. We begin by briefly discussing how eukaryotic cells are organized, how

this reflects their way of life, and how their genomes differ from those of prokaryotes. This leads us to an outline of the strategy by which cell biologists, by exploiting genetic and biochemical information, are attempting to discover how eukaryotic organisms work.

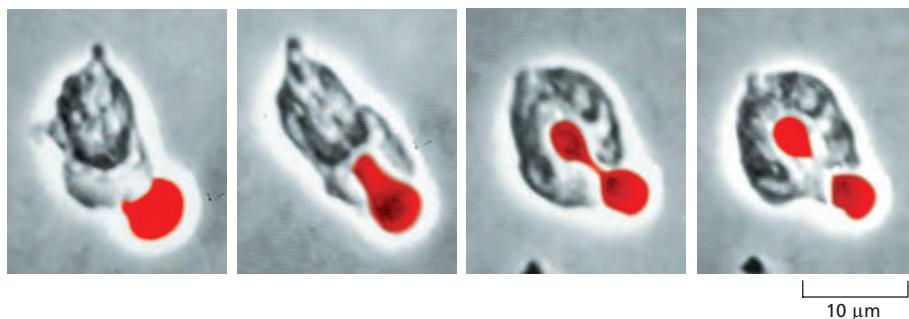
### Eukaryotic Cells May Have Originated as Predators

By definition, eukaryotic cells keep their DNA in an internal compartment called the nucleus. The *nuclear envelope*, a double layer of membrane, surrounds the nucleus and separates the DNA from the cytoplasm. Eukaryotes also have other features that set them apart from prokaryotes (Figure 1-25). Their cells are, typically, 10 times bigger in linear dimension and 1000 times larger in volume. They have an elaborate *cytoskeleton*—a system of protein filaments crisscrossing the cytoplasm and forming, together with the many proteins that attach to them, a system of girders, ropes, and motors that gives the cell mechanical strength, controls its shape, and drives and guides its movements (Movie 1.1). And the nuclear envelope is only one part of a set of *internal membranes*, each structurally similar to the plasma membrane and enclosing different types of spaces inside the cell, many of them involved in digestion and secretion. Lacking the tough cell wall of most bacteria, animal cells and the free-living eukaryotic cells called *protozoa* can change their shape rapidly and engulf other cells and small objects by *phagocytosis* (Figure 1-26).

How all of the unique properties of eukaryotic cells evolved, and in what sequence, is still a mystery. One plausible view, however, is that they are all reflections of the way of life of a primordial cell that was a predator, living by capturing other cells and eating them (Figure 1-27). Such a way of life requires a large cell with a flexible plasma membrane, as well as an elaborate cytoskeleton to support



**Figure 1-25 The major features of eukaryotic cells.** The drawing depicts a typical animal cell, but almost all the same components are found in plants and fungi as well as in single-celled eukaryotes such as yeasts and protozoa. Plant cells contain chloroplasts in addition to the components shown here, and their plasma membrane is surrounded by a tough external wall formed of cellulose.

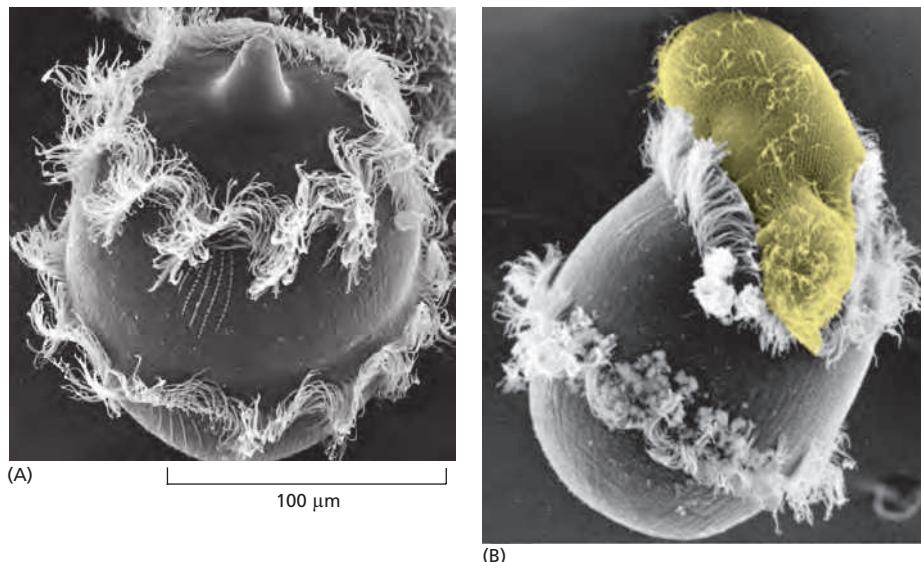


**Figure 1–26 Phagocytosis.** This series of stills from a movie shows a human white blood cell (a neutrophil) engulfing a red blood cell (artificially colored red) that has been treated with an antibody that marks it for destruction (see Movie 13.5). (Courtesy of Stephen E. Malawista and Anne de Boisfleury Chevance.)

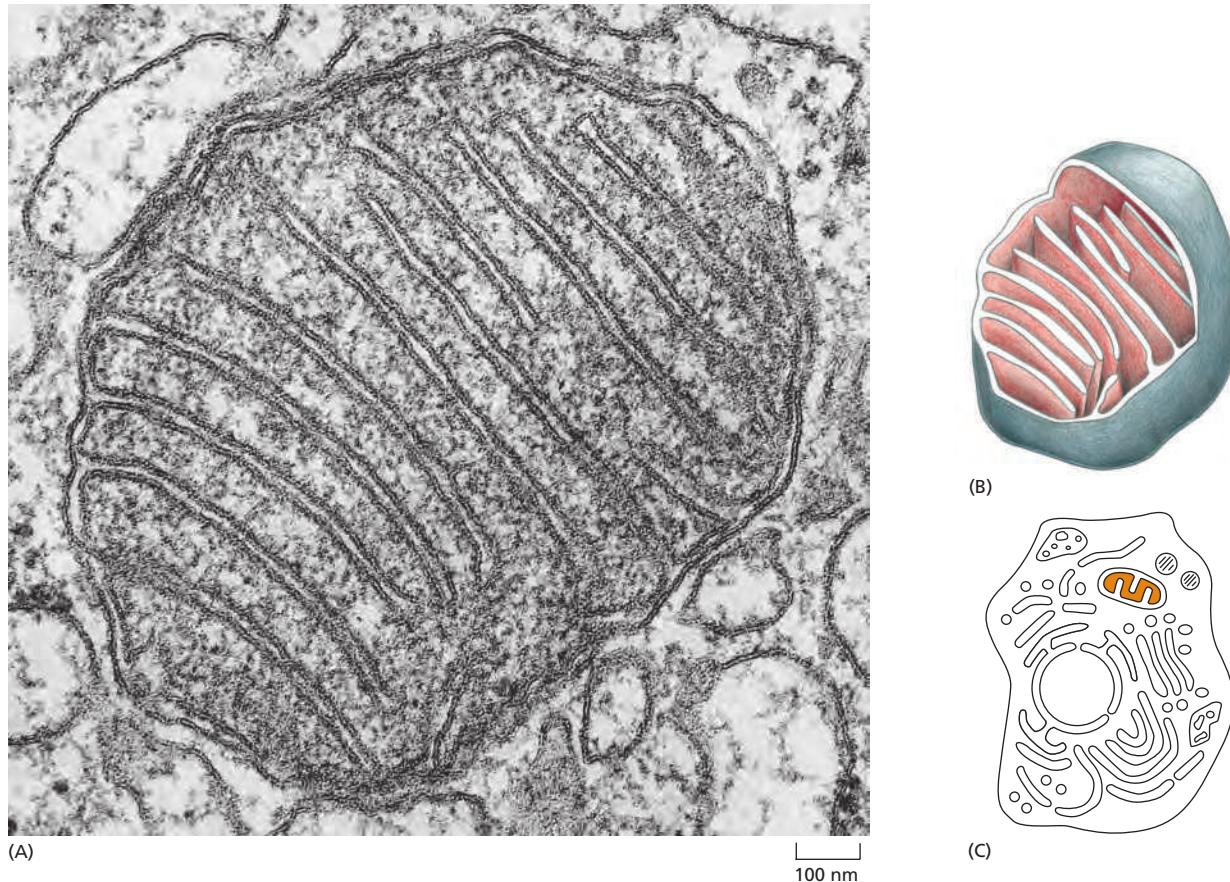
and move this membrane. It may also require that the cell's long, fragile DNA molecules be sequestered in a separate nuclear compartment, to protect the genome from damage by the movements of the cytoskeleton.

### Modern Eukaryotic Cells Evolved from a Symbiosis

A predatory way of life helps to explain another feature of eukaryotic cells. All such cells contain (or at one time did contain) *mitochondria* (Figure 1–28). These small bodies in the cytoplasm, enclosed by a double layer of membrane, take up oxygen and harness energy from the oxidation of food molecules—such as sugars—to produce most of the ATP that powers the cell's activities. Mitochondria are similar in size to small bacteria, and, like bacteria, they have their own genome in the form of a circular DNA molecule, their own ribosomes that differ from those elsewhere in the eukaryotic cell, and their own transfer RNAs. It is now generally accepted that mitochondria originated from free-living oxygen-metabolizing (*aerobic*) bacteria that were engulfed by an ancestral cell that could otherwise make no such use of oxygen (that is, was *anaerobic*). Escaping digestion, these bacteria evolved in symbiosis with the engulfing cell and its progeny, receiving



**Figure 1–27 A single-celled eukaryote that eats other cells.** (A) *Didinium* is a carnivorous protozoan, belonging to the group known as ciliates. It has a globular body, about 150  $\mu\text{m}$  in diameter, encircled by two fringes of cilia—sinuous, whiplike appendages that beat continually; its front end is flattened except for a single protrusion, rather like a snout. (B) A *Didinium* engulfing its prey. *Didinium* normally swims around in the water at high speed by means of the synchronous beating of its cilia. When it encounters a suitable prey (yellow), usually another type of protozoan, it releases numerous small paralyzing darts from its snout region. Then, the *Didinium* attaches to and devours the other cell by phagocytosis, inverting like a hollow ball to engulf its victim, which can be almost as large as itself. (Courtesy of D. Barlow.)



shelter and nourishment in return for the power generation they performed for their hosts. This partnership between a primitive anaerobic predator cell and an aerobic bacterial cell is thought to have been established about 1.5 billion years ago, when the Earth's atmosphere first became rich in oxygen.

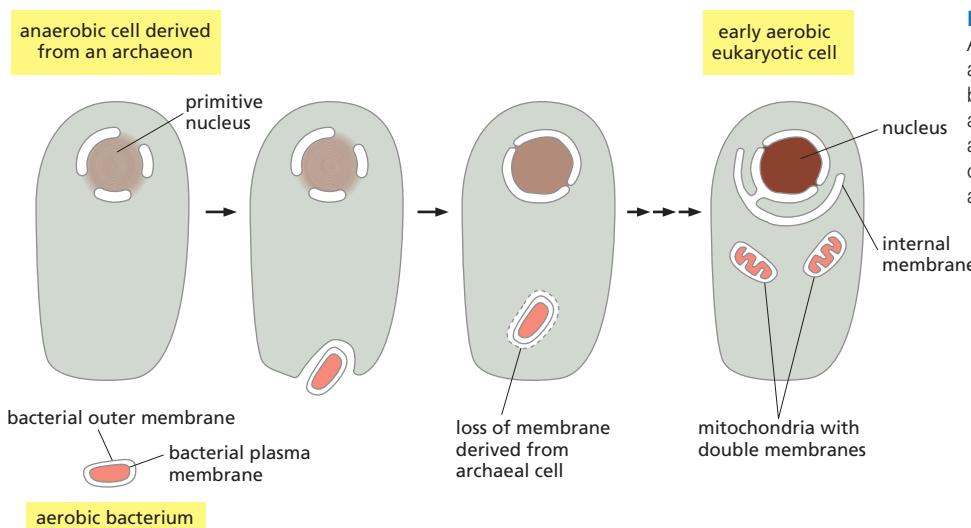
As indicated in [Figure 1-29](#), recent genomic analyses suggest that the first eukaryotic cells formed after an archaeal cell engulfed an aerobic bacterium. This would explain why all eukaryotic cells today, including those that live as strict anaerobes show clear evidence that they once contained mitochondria.

Many eukaryotic cells—specifically, those of plants and algae—also contain another class of small membrane-enclosed organelles somewhat similar to mitochondria—the chloroplasts ([Figure 1-30](#)). Chloroplasts perform photosynthesis, using the energy of sunlight to synthesize carbohydrates from atmospheric carbon dioxide and water, and deliver the products to the host cell as food. Like mitochondria, chloroplasts have their own genome. They almost certainly originated as symbiotic photosynthetic bacteria, acquired by eukaryotic cells that already possessed mitochondria ([Figure 1-31](#)).

A eukaryotic cell equipped with chloroplasts has no need to chase after other cells as prey; it is nourished by the captive chloroplasts it has inherited from its ancestors. Correspondingly, plant cells, although they possess the cytoskeletal equipment for movement, have lost the ability to change shape rapidly and to engulf other cells by phagocytosis. Instead, they create around themselves a tough, protective cell wall. If the first eukaryotic cells were predators on other organisms, we can view plant cells as cells that have made the transition from hunting to farming.

Fungi represent yet another eukaryotic way of life. Fungal cells, like animal cells, possess mitochondria but not chloroplasts; but in contrast with animal cells and protozoa, they have a tough outer wall that limits their ability to move rapidly

**Figure 1–28** A mitochondrion. (A) A cross section, as seen in the electron microscope. (B) A drawing of a mitochondrion with part of it cut away to show the three-dimensional structure ([Movie 1.2](#)). (C) A schematic eukaryotic cell, with the interior space of a mitochondrion, containing the mitochondrial DNA and ribosomes, colored. Note the smooth outer membrane and the convoluted inner membrane, which houses the proteins that generate ATP from the oxidation of food molecules. (A, courtesy of Daniel S. Friend.)

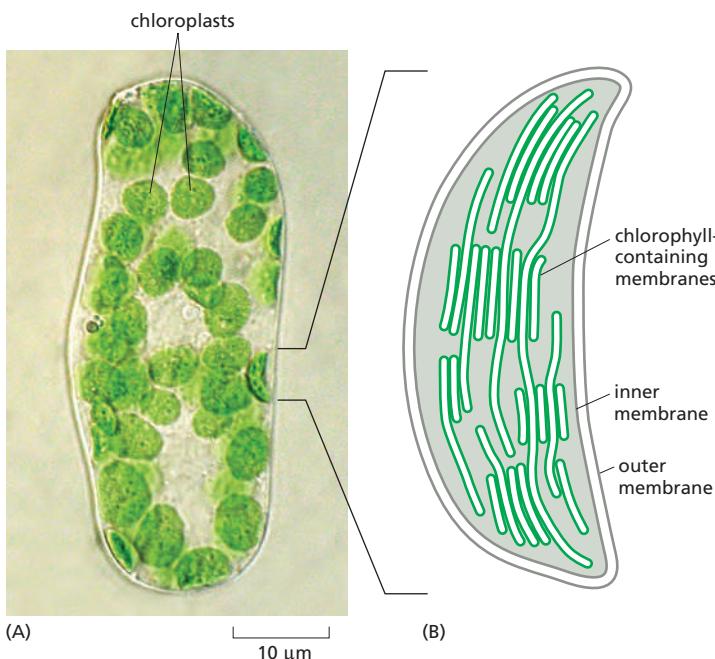


**Figure 1–29 The origin of mitochondria.**  
An ancestral anaerobic predator cell (an archaeon) is thought to have engulfed the bacterial ancestor of mitochondria, initiating a symbiotic relationship. Clear evidence of a dual bacterial and archaeal inheritance can be discerned today in the genomes of all eukaryotes.

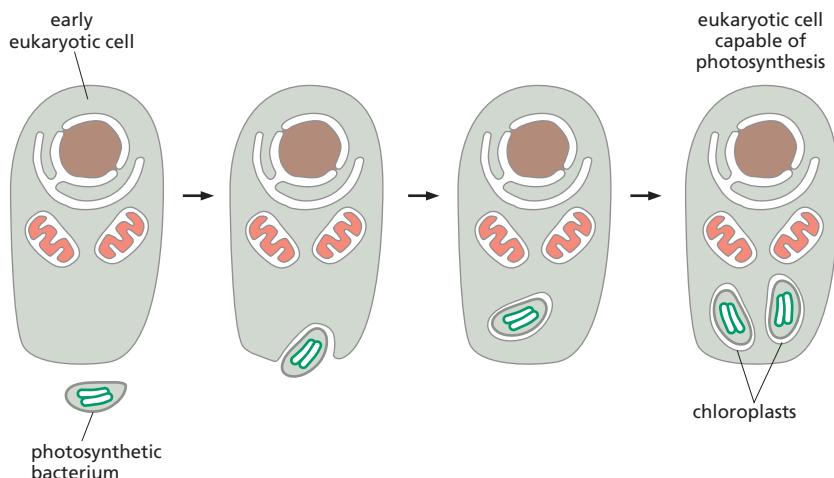
or to swallow up other cells. Fungi, it seems, have turned from hunters into scavengers: other cells secrete nutrient molecules or release them upon death, and fungi feed on these leavings—performing whatever digestion is necessary extracellularly, by secreting digestive enzymes to the exterior.

### Eukaryotes Have Hybrid Genomes

The genetic information of eukaryotic cells has a hybrid origin—from the ancestral anaerobic archaeal cell, and from the bacteria that it adopted as symbionts. Most of this information is stored in the nucleus, but a small amount remains inside the mitochondria and, for plant and algal cells, in the chloroplasts. When mitochondrial DNA and the chloroplast DNA are separated from the nuclear DNA and individually analyzed and sequenced, the mitochondrial and chloroplast genomes are found to be degenerate, cut-down versions of the corresponding bacterial genomes. In a human cell, for example, the mitochondrial genome consists of only 16,569 nucleotide pairs, and codes for only 13 proteins, 2 ribosomal RNA components, and 22 transfer RNAs.



**Figure 1–30 Chloroplasts.** These organelles capture the energy of sunlight in plant cells and some single-celled eukaryotes. (A) A single cell isolated from a leaf of a flowering plant, seen in the light microscope, showing the green chloroplasts (**Movie 1.3** and see **Movie 14.9**). (B) A drawing of one of the chloroplasts, showing the highly folded system of internal membranes containing the chlorophyll molecules by which light is absorbed. (A, courtesy of Preeti Dahiya.)



**Figure 1–31** The origin of chloroplasts.

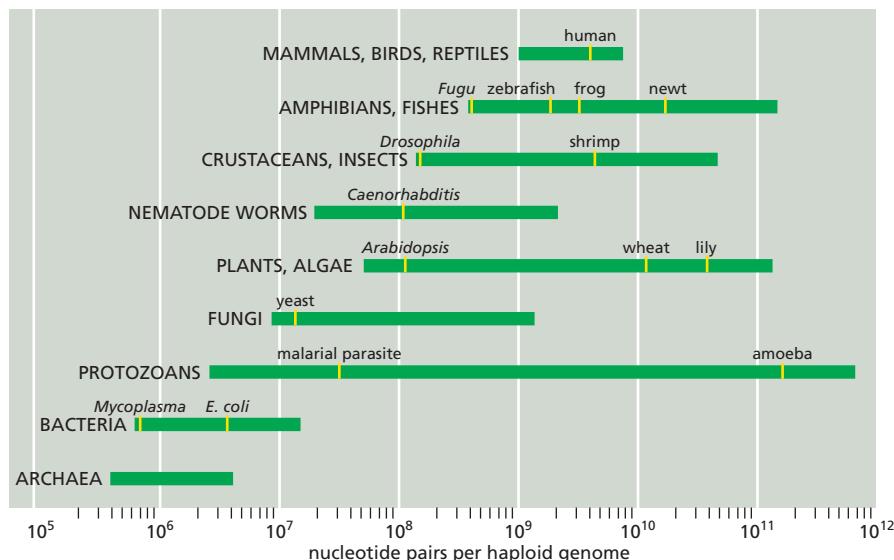
An early eukaryotic cell, already possessing mitochondria, engulfed a photosynthetic bacterium (a cyanobacterium) and retained it in symbiosis. Present-day chloroplasts are thought to trace their ancestry back to a single species of cyanobacterium that was adopted as an internal symbiont (an endosymbiont) over a billion years ago.

Many of the genes that are missing from the mitochondria and chloroplasts have not been lost; instead, they have moved from the symbiont genome into the DNA of the host cell nucleus. The nuclear DNA of humans contains many genes coding for proteins that serve essential functions inside the mitochondria; in plants, the nuclear DNA also contains many genes specifying proteins required in chloroplasts. In both cases, the DNA sequences of these nuclear genes show clear evidence of their origin from the bacterial ancestor of the respective organelle.

### Eukaryotic Genomes Are Big

Natural selection has evidently favored mitochondria with small genomes. By contrast, the nuclear genomes of most eukaryotes seem to have been free to enlarge. Perhaps the eukaryotic way of life has made large size an advantage: predators typically need to be bigger than their prey, and cell size generally increases in proportion to genome size. Whatever the reason, aided by a massive accumulation of DNA segments derived from parasitic transposable elements (discussed in Chapter 5), the genomes of most eukaryotes have become orders of magnitude larger than those of bacteria and archaea (Figure 1–32).

The freedom to be extravagant with DNA has had profound implications. Eukaryotes not only have more genes than prokaryotes; they also have vastly more DNA that does not code for protein. The human genome contains 1000 times as many nucleotide pairs as the genome of a typical bacterium, perhaps 10 times as



**Figure 1–32** Genome sizes compared.

Genome size is measured in nucleotide pairs of DNA per haploid genome, that is, per single copy of the genome. (The cells of sexually reproducing organisms such as ourselves are generally diploid: they contain two copies of the genome, one inherited from the mother, the other from the father.) Closely related organisms can vary widely in the quantity of DNA in their genomes, even though they contain similar numbers of functionally distinct genes. (Data from W.H. Li, Molecular Evolution, pp. 380–383. Sunderland, MA: Sinauer, 1997.)

**TABLE 1–2** Some Model Organisms and Their Genomes

Organism	Genome size* (nucleotide pairs)	Approximate number of genes
<i>Escherichia coli</i> (bacterium)	$4.6 \times 10^6$	4300
<i>Saccharomyces cerevisiae</i> (yeast)	$13 \times 10^6$	6600
<i>Caenorhabditis elegans</i> (roundworm)	$130 \times 10^6$	21,000
<i>Arabidopsis thaliana</i> (plant)	$220 \times 10^6$	29,000
<i>Drosophila melanogaster</i> (fruit fly)	$200 \times 10^6$	15,000
<i>Danio rerio</i> (zebrafish)	$1400 \times 10^6$	32,000
<i>Mus musculus</i> (mouse)	$2800 \times 10^6$	30,000
<i>Homo sapiens</i> (human)	$3200 \times 10^6$	30,000

\*Genome size includes an estimate for the amount of highly repeated DNA sequence not in genome databases.

many genes, and a great deal more noncoding DNA (~98.5% of the genome for a human does not code for proteins, as opposed to 11% of the genome for the bacterium *E. coli*). The estimated genome sizes and gene numbers for some eukaryotes are compiled for easy comparison with *E. coli* in **Table 1–2**; we shall discuss how each of these eukaryotes serves as a model organism shortly.

### Eukaryotic Genomes Are Rich in Regulatory DNA

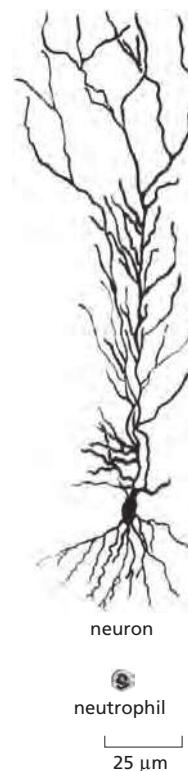
Much of our noncoding DNA is almost certainly dispensable junk, retained like a mass of old papers because, when there is little pressure to keep an archive small, it is easier to retain everything than to sort out the valuable information and discard the rest. Certain exceptional eukaryotic species, such as the puffer fish, bear witness to the profligacy of their relatives; they have somehow managed to rid themselves of large quantities of noncoding DNA. Yet they appear similar in structure, behavior, and fitness to related species that have vastly more such DNA (see Figure 4–71).

Even in compact eukaryotic genomes such as that of puffer fish, there is more noncoding DNA than coding DNA, and at least some of the noncoding DNA certainly has important functions. In particular, it regulates the expression of adjacent genes. With this *regulatory DNA*, eukaryotes have evolved distinctive ways of controlling when and where a gene is brought into play. This sophisticated gene regulation is crucial for the formation of complex multicellular organisms.

### The Genome Defines the Program of Multicellular Development

The cells in an individual animal or plant are extraordinarily varied. Fat cells, skin cells, bone cells, nerve cells—they seem as dissimilar as any cells could be (**Figure 1–33**). Yet all these cell types are the descendants of a single fertilized egg cell, and all (with minor exceptions) contain identical copies of the genome of the species.

The differences result from the way in which the cells make selective use of their genetic instructions according to the cues they get from their surroundings in the developing embryo. The DNA is not just a shopping list specifying the molecules that every cell must have, and the cell is not an assembly of all the items on the list. Rather, the cell behaves as a multipurpose machine, with sensors to receive environmental signals and with highly developed abilities to call different sets of genes into action according to the sequences of signals to which the cell has been exposed. The genome in each cell is big enough to accommodate the



**Figure 1–33** Cell types can vary enormously in size and shape. An animal nerve cell is compared here with a neutrophil, a type of white blood cell. Both are drawn to scale.



**Figure 1–34** Genetic control of the program of multicellular development.

The role of a regulatory gene is demonstrated in the snapdragon *Antirrhinum*. In this example, a mutation in a single gene coding for a regulatory protein causes leafy shoots to develop in place of flowers: because a regulatory protein has been changed, the cells adopt characters that would be appropriate to a different location in the normal plant. The mutant is on the left, the normal plant on the right. (Courtesy of Enrico Coen and Rosemary Carpenter.)

information that specifies an entire multicellular organism, but in any individual cell only part of that information is used.

A large number of genes in the eukaryotic genome code for proteins that regulate the activities of other genes. Most of these *transcription regulators* act by binding, directly or indirectly, to the regulatory DNA adjacent to the genes that are to be controlled, or by interfering with the abilities of other proteins to do so. The expanded genome of eukaryotes therefore not only specifies the hardware of the cell, but also stores the software that controls how that hardware is used (Figure 1–34).

Cells do not just passively receive signals; rather, they actively exchange signals with their neighbors. Thus, in a developing multicellular organism, the same control system governs each cell, but with different consequences depending on the messages exchanged. The outcome, astonishingly, is a precisely patterned array of cells in different states, each displaying a character appropriate to its position in the multicellular structure.

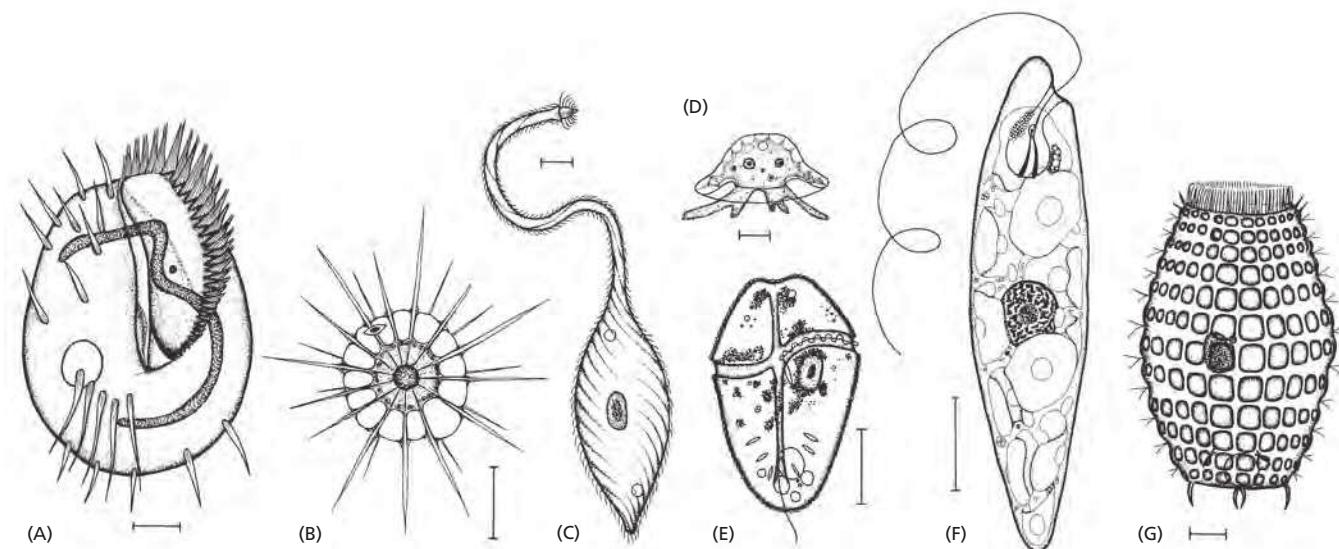
### Many Eukaryotes Live as Solitary Cells

Many species of eukaryotic cells lead a solitary life—some as hunters (the *protozoa*), some as photosynthesizers (the unicellular *algae*), some as scavengers (the unicellular fungi, or *yeasts*). Figure 1–35 conveys something of the astonishing variety of the single-celled eukaryotes. The anatomy of protozoa, especially, is often elaborate and includes such structures as sensory bristles, photoreceptors, sinuously beating cilia, leglike appendages, mouth parts, stinging darts, and musclelike contractile bundles. Although they are single cells, protozoa can be as intricate, as versatile, and as complex in their behavior as many multicellular organisms (see Figure 1–27, Movie 1.4, and Movie 1.5).

In terms of their ancestry and DNA sequences, the unicellular eukaryotes are far more diverse than the multicellular animals, plants, and fungi, which arose as three comparatively late branches of the eukaryotic pedigree (see Figure 1–17). As with prokaryotes, humans have tended to neglect them because they are microscopic. Only now, with the help of genome analysis, are we beginning to understand their positions in the tree of life, and to put into context the glimpses these strange creatures can offer us of our distant evolutionary past.

### A Yeast Serves as a Minimal Model Eukaryote

The molecular and genetic complexity of eukaryotes is daunting. Even more than for prokaryotes, biologists need to concentrate their limited resources on a few selected model organisms to unravel this complexity.

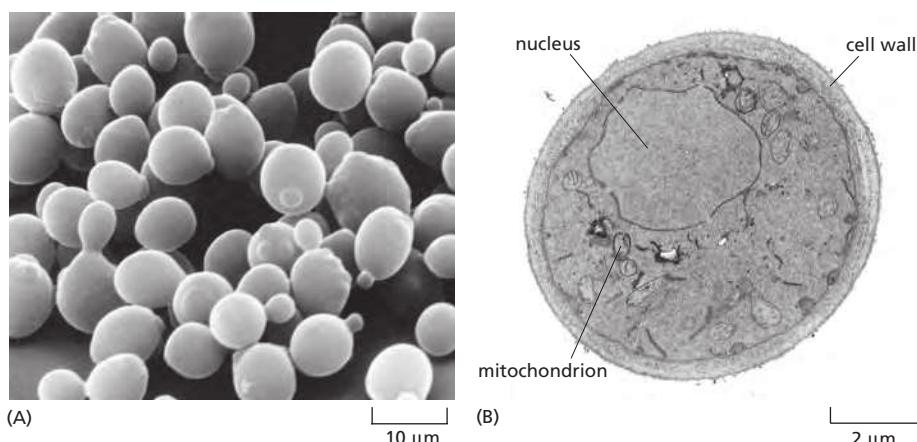


To analyze the internal workings of the eukaryotic cell without the additional problems of multicellular development, it makes sense to use a species that is unicellular and as simple as possible. The popular choice for this role of minimal model eukaryote has been the yeast *Saccharomyces cerevisiae* (Figure 1-36)—the same species that is used by brewers of beer and bakers of bread.

*S. cerevisiae* is a small, single-celled member of the kingdom of fungi and thus, according to modern views, is at least as closely related to animals as it is to plants. It is robust and easy to grow in a simple nutrient medium. Like other fungi, it has a tough cell wall, is relatively immobile, and possesses mitochondria but not chloroplasts. When nutrients are plentiful, it grows and divides almost as rapidly as a bacterium. It can reproduce either vegetatively (that is, by simple cell division), or sexually: two yeast cells that are *haploid* (possessing a single copy of the genome) can fuse to create a cell that is *diploid* (containing a double genome); and the diploid cell can undergo *meiosis* (a reduction division) to produce cells that are once again haploid (Figure 1-37). In contrast with higher plants and animals, the yeast can divide indefinitely in either the haploid or the diploid state, and the process leading from one state to the other can be induced at will by changing the growth conditions.

In addition to these features, the yeast has a further property that makes it a convenient organism for genetic studies: its genome, by eukaryotic standards, is exceptionally small. Nevertheless, it suffices for all the basic tasks that every eukaryotic cell must perform. Mutants are available for essentially every gene,

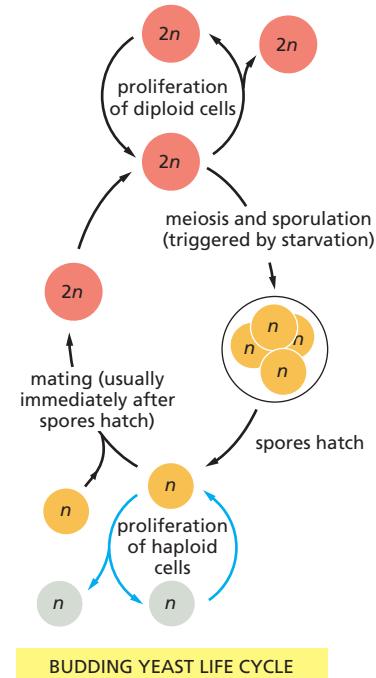
**Figure 1–35** An assortment of protozoa: a small sample of an extremely diverse class of organisms. The drawings are done to different scales, but in each case the scale bar represents 10 µm. The organisms in (A), (C), and (G) are ciliates; (B) is a heliozoan; (D) is an amoeba; (E) is a dinoflagellate; and (F) is a euglenoid. (From M.A. Sleigh, Biology of Protozoa. Cambridge, UK: Cambridge University Press, 1973.)



**Figure 1–36** The yeast *Saccharomyces cerevisiae*. (A) A scanning electron micrograph of a cluster of the cells. This species is also known as budding yeast; it proliferates by forming a protrusion or bud that enlarges and then separates from the rest of the original cell. Many cells with buds are visible in this micrograph. (B) A transmission electron micrograph of a cross section of a yeast cell, showing its nucleus, mitochondrion, and thick cell wall. (A, courtesy of Ira Herskowitz and Eric Schabatach.)

**Figure 1–37** The reproductive cycles of the yeast *S. cerevisiae*.

Depending on environmental conditions and on details of the genotype, cells of this species can exist in either a diploid ( $2n$ ) state, with a double chromosome set, or a haploid ( $n$ ) state, with a single chromosome set. The diploid form can either proliferate by ordinary cell-division cycles or undergo meiosis to produce haploid cells. The haploid form can either proliferate by ordinary cell-division cycles or undergo sexual fusion with another haploid cell to become diploid. Meiosis is triggered by starvation and gives rise to spores—haploid cells in a dormant state, resistant to harsh environmental conditions.



BUDDING YEAST LIFE CYCLE

and studies on yeasts (using both *S. cerevisiae* and other species) have provided a key to many crucial processes, including the eukaryotic cell-division cycle—the critical chain of events by which the nucleus and all the other components of a cell are duplicated and parceled out to create two daughter cells from one. The control system that governs this process has been so well conserved over the course of evolution that many of its components can function interchangeably in yeast and human cells: if a mutant yeast lacking an essential yeast cell-division-cycle gene is supplied with a copy of the homologous cell-division-cycle gene from a human, the yeast is cured of its defect and becomes able to divide normally.

### The Expression Levels of All the Genes of An Organism Can Be Monitored Simultaneously

The complete genome sequence of *S. cerevisiae*, determined in 1997, consists of approximately 13,117,000 nucleotide pairs, including the small contribution (78,520 nucleotide pairs) of the mitochondrial DNA. This total is only about 2.5 times as much DNA as there is in *E. coli*, and it codes for only 1.5 times as many distinct proteins (about 6600 in all). The way of life of *S. cerevisiae* is similar in many ways to that of a bacterium, and it seems that this yeast has likewise been subject to selection pressures that have kept its genome compact.

Knowledge of the complete genome sequence of any organism—be it a yeast or a human—opens up new perspectives on the workings of the cell: things that once seemed impossibly complex now seem within our grasp. Using techniques described in Chapter 8, it is now possible, for example, to monitor, simultaneously, the amount of mRNA transcript that is produced from every gene in the yeast genome under any chosen conditions, and to see how this whole pattern of gene activity changes when conditions change. The analysis can be repeated with mRNA prepared from mutant cells lacking a chosen gene—any gene that we care to test. In principle, this approach provides a way to reveal the entire system of control relationships that govern gene expression—not only in yeast cells, but in any organism whose genome sequence is known.

### *Arabidopsis* Has Been Chosen Out of 300,000 Species As a Model Plant

The large multicellular organisms that we see around us—the flowers and trees and animals—seem fantastically varied, but they are much closer to one another in their evolutionary origins, and more similar in their basic cell biology, than the great host of microscopic single-celled organisms. Thus, while bacteria and archaea are separated by perhaps 3.5 billion years of evolution, vertebrates and insects are separated by about 700 million years, fish and mammals by about 450 million years, and the different species of flowering plants by only about 150 million years.

Because of the close evolutionary relationship between all flowering plants, we can, once again, get insight into the cell and molecular biology of this whole class of organisms by focusing on just one or a few species for detailed analysis. Out of the several hundred thousand species of flowering plants on Earth today, molecular biologists have chosen to concentrate their efforts on a small weed,

the common Thale cress *Arabidopsis thaliana* (Figure 1–38), which can be grown indoors in large numbers and produces thousands of offspring per plant after 8–10 weeks. *Arabidopsis* has a total genome size of approximately 220 million nucleotide pairs, about 17 times the size of yeast's (see Table 1–2).

### The World of Animal Cells Is Represented By a Worm, a Fly, a Fish, a Mouse, and a Human

Multicellular animals account for the majority of all named species of living organisms, and for the largest part of the biological research effort. Five species have emerged as the foremost model organisms for molecular genetic studies. In order of increasing size, they are the nematode worm *Caenorhabditis elegans*, the fly *Drosophila melanogaster*, the zebrafish *Danio rerio*, the mouse *Mus musculus*, and the human, *Homo sapiens*. Each has had its genome sequenced.

*Caenorhabditis elegans* (Figure 1–39) is a small, harmless relative of the eel-worm that attacks crops. With a life cycle of only a few days, an ability to survive in a freezer indefinitely in a state of suspended animation, a simple body plan, and an unusual life cycle that is well suited for genetic studies (described in Chapter 21), it is an ideal model organism. *C. elegans* develops with clockwork precision from a fertilized egg cell into an adult worm with exactly 959 body cells (plus a variable number of egg and sperm cells)—an unusual degree of regularity for an animal. We now have a minutely detailed description of the sequence of events by which this occurs, as the cells divide, move, and change their character according to strict and predictable rules. The genome of 130 million nucleotide pairs codes for about 21,000 proteins, and many mutants and other tools are available for the testing of gene functions. Although the worm has a body plan very different from our own, the conservation of biological mechanisms has been sufficient for the worm to be a model for many of the developmental and cell-biological processes that occur in the human body. Thus, for example, studies of the worm have been critical for helping us to understand the programs of cell division and cell death that determine the number of cells in the body—a topic of great importance for both developmental biology and cancer research.

### Studies in *Drosophila* Provide a Key to Vertebrate Development

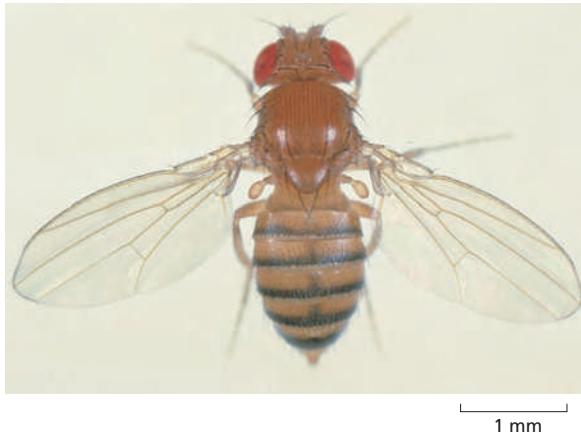
The fruit fly *Drosophila melanogaster* (Figure 1–40) has been used as a model genetic organism for longer than any other; in fact, the foundations of classical genetics were built to a large extent on studies of this insect. Over 80 years ago, it provided, for example, definitive proof that genes—the abstract units of hereditary information—are carried on chromosomes, concrete physical objects whose behavior had been closely followed in the eukaryotic cell with the light microscope, but whose function was at first unknown. The proof depended on one of the many features that make *Drosophila* peculiarly convenient for genetics—the giant chromosomes, with characteristic banded appearance, that are visible in



**Figure 1–38** *Arabidopsis thaliana*, the plant chosen as the primary model for studying plant molecular genetics. (Courtesy of Toni Hayden and the John Innes Foundation.)



**Figure 1–39** *Caenorhabditis elegans*, the first multicellular organism to have its complete genome sequence determined. This small nematode, about 1 mm long, lives in the soil. Most individuals are hermaphrodites, producing both eggs and sperm. (Courtesy of Maria Gallegos, University of Wisconsin, Madison.)



**Figure 1–40** *Drosophila melanogaster*. Molecular genetic studies on this fly have provided the main key to understanding how all animals develop from a fertilized egg into an adult. (From E.B. Lewis, *Science* 221:cover, 1983. With permission from AAAS.)

some of its cells (Figure 1–41). Specific changes in the hereditary information, manifest in families of mutant flies, were found to correlate exactly with the loss or alteration of specific giant-chromosome bands.

In more recent times, *Drosophila*, more than any other organism, has shown us how to trace the chain of cause and effect from the genetic instructions encoded in the chromosomal DNA to the structure of the adult multicellular body. *Drosophila* mutants with body parts strangely misplaced or mispatterned provided the key to the identification and characterization of the genes required to make a properly structured body, with gut, limbs, eyes, and all the other parts in their correct places. Once these *Drosophila* genes were sequenced, the genomes of vertebrates could be scanned for homologs. These were found, and their functions in vertebrates were then tested by analyzing mice in which the genes had been mutated. The results, as we see later in the book, reveal an astonishing degree of similarity in the molecular mechanisms that govern insect and vertebrate development (discussed in Chapter 21).

The majority of all named species of living organisms are insects. Even if *Drosophila* had nothing in common with vertebrates, but only with insects, it would still be an important model organism. But if understanding the molecular genetics of vertebrates is the goal, why not simply tackle the problem head-on? Why sidle up to it obliquely, through studies in *Drosophila*?

*Drosophila* requires only 9 days to progress from a fertilized egg to an adult; it is vastly easier and cheaper to breed than any vertebrate, and its genome is much smaller—about 200 million nucleotide pairs, compared with 3200 million for a human. This genome codes for about 15,000 proteins, and mutants can now be obtained for essentially any gene. But there is also another, deeper reason why genetic mechanisms that are hard to discover in a vertebrate are often readily revealed in the fly. This relates, as we now explain, to the frequency of gene duplication, which is substantially greater in vertebrate genomes than in the fly genome and has probably been crucial in making vertebrates the complex and subtle creatures that they are.

### The Vertebrate Genome Is a Product of Repeated Duplications

Almost every gene in the vertebrate genome has paralogs—other genes in the same genome that are unmistakably related and must have arisen by gene duplication. In many cases, a whole cluster of genes is closely related to similar clusters present elsewhere in the genome, suggesting that genes have been duplicated in linked groups rather than as isolated individuals. According to one hypothesis, at an early stage in the evolution of the vertebrates, the entire genome underwent duplication twice in succession, giving rise to four copies of every gene.

The precise course of vertebrate genome evolution remains uncertain, because many further evolutionary changes have occurred since these ancient events.



**Figure 1–41** Giant chromosomes from salivary gland cells of *Drosophila*. Because many rounds of DNA replication have occurred without an intervening cell division, each of the chromosomes in these unusual cells contains over 1000 identical DNA molecules, all aligned in register. This makes them easy to see in the light microscope, where they display a characteristic and reproducible banding pattern. Specific bands can be identified as the locations of specific genes: a mutant fly with a region of the banding pattern missing shows a phenotype reflecting loss of the genes in that region. Genes that are being transcribed at a high rate correspond to bands with a “puffed” appearance. The bands stained dark brown in the micrograph are sites where a particular regulatory protein is bound to the DNA. (Courtesy of B. Zink and R. Paro, from R. Paro, *Trends Genet.* 6:416–421, 1990. With permission from Elsevier.)

Genes that were once identical have diverged; many of the gene copies have been lost through disruptive mutations; some have undergone further rounds of local duplication; and the genome, in each branch of the vertebrate family tree, has suffered repeated rearrangements, breaking up most of the original gene orderings. Comparison of the gene order in two related organisms, such as the human and the mouse, reveals that—on the time scale of vertebrate evolution—chromosomes frequently fuse and fragment to move large blocks of DNA sequence around. Indeed, it is possible, as discussed in Chapter 4, that the present state of affairs is the result of many separate duplications of fragments of the genome, rather than duplications of the genome as a whole.

There is, however, no doubt that such whole-genome duplications do occur from time to time in evolution, for we can see recent instances in which duplicated chromosome sets are still clearly identifiable as such. The frog genus *Xenopus*, for example, comprises a set of closely similar species related to one another by repeated duplications or triplications of the whole genome. Among these frogs are *X. tropicalis*, with an ordinary diploid genome; the common laboratory species *X. laevis*, with a duplicated genome and twice as much DNA per cell; and *X. ruwenzoriensis*, with a sixfold reduplication of the original genome and six times as much DNA per cell (108 chromosomes, compared with 36 in *X. laevis*, for example). These species are estimated to have diverged from one another within the past 120 million years ([Figure 1–42](#)).

### The Frog and the Zebrafish Provide Accessible Models for Vertebrate Development

Frogs have long been used to study the early steps of embryonic development in vertebrates, because their eggs are big, easy to manipulate, and fertilized outside of the animal, so that the subsequent development of the early embryo is easily followed ([Figure 1–43](#)). *Xenopus laevis*, in particular, continues to be an important model organism, even though it is poorly suited for genetic analysis ([Movie 1.6](#) and see Movie 21.1).

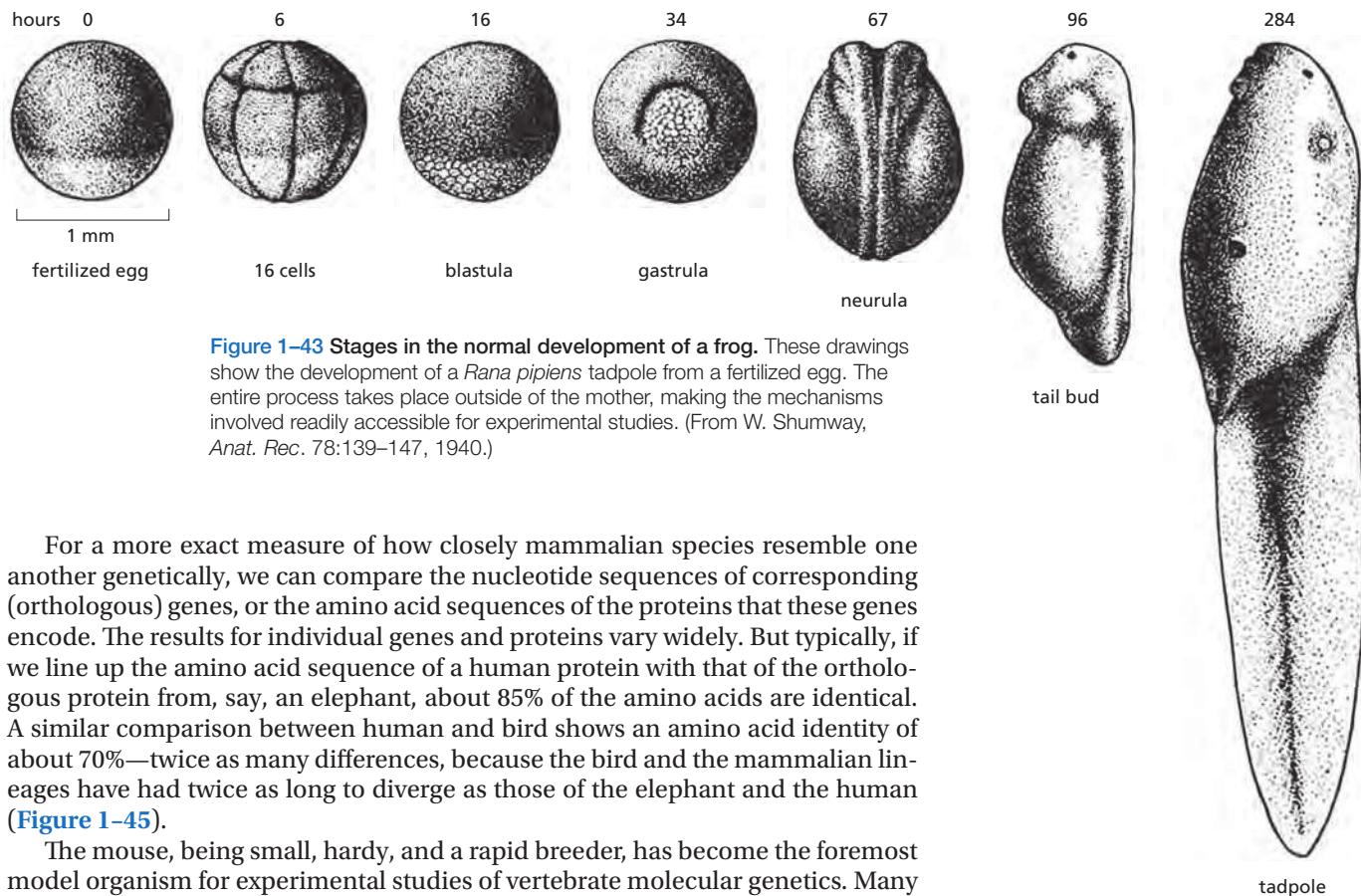
The zebrafish *Danio rerio* has similar advantages, but without this drawback. Its genome is compact—only half as big as that of a mouse or a human—and it has a generation time of only about three months. Many mutants are known, and genetic engineering is relatively easy. The zebrafish has the added virtue that it is transparent for the first two weeks of its life, so that one can watch the behavior of individual cells in the living organism (see Movie 21.2). All this has made it an increasingly important model vertebrate ([Figure 1–44](#)).

### The Mouse Is the Predominant Mammalian Model Organism

Mammals have typically two times as many genes as *Drosophila*, a genome that is 16 times larger, and millions or billions of times as many cells in their adult bodies. In terms of genome size and function, cell biology, and molecular mechanisms, mammals are nevertheless a highly uniform group of organisms. Even anatomically, the differences among mammals are chiefly a matter of size and proportions; it is hard to think of a human body part that does not have a counterpart in elephants and mice, and vice versa. Evolution plays freely with quantitative features, but it does not readily change the logic of the structure.

**Figure 1–42** Two species of the frog genus *Xenopus*. *X. tropicalis*, above, has an ordinary diploid genome; *X. laevis*, below, has twice as much DNA per cell. From the banding patterns of their chromosomes and the arrangement of genes along them, as well as from comparisons of gene sequences, it is clear that the large-genome species have evolved through duplications of the whole genome. These duplications are thought to have occurred in the aftermath of matings between frogs of slightly divergent *Xenopus* species. (Courtesy of E. Amaya, M. Offield, and R. Grainger, *Trends Genet.* 14:253–255, 1998. With permission from Elsevier.)





**Figure 1–43** Stages in the normal development of a frog. These drawings show the development of a *Rana pipiens* tadpole from a fertilized egg. The entire process takes place outside of the mother, making the mechanisms involved readily accessible for experimental studies. (From W. Shumway, *Anat. Rec.* 78:139–147, 1940.)

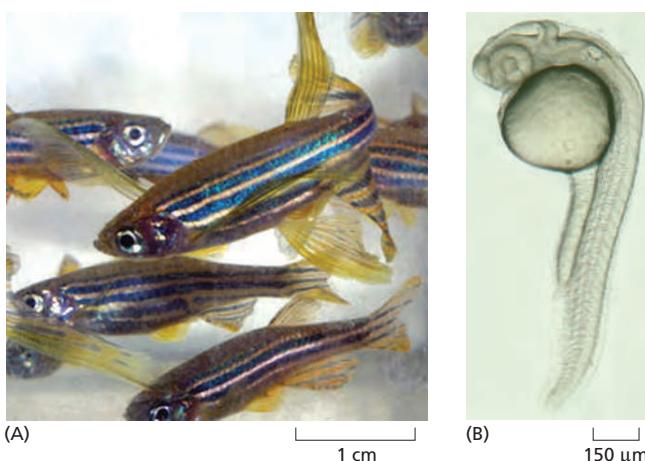
For a more exact measure of how closely mammalian species resemble one another genetically, we can compare the nucleotide sequences of corresponding (orthologous) genes, or the amino acid sequences of the proteins that these genes encode. The results for individual genes and proteins vary widely. But typically, if we line up the amino acid sequence of a human protein with that of the orthologous protein from, say, an elephant, about 85% of the amino acids are identical. A similar comparison between human and bird shows an amino acid identity of about 70%—twice as many differences, because the bird and the mammalian lineages have had twice as long to diverge as those of the elephant and the human (Figure 1–45).

The mouse, being small, hardy, and a rapid breeder, has become the foremost model organism for experimental studies of vertebrate molecular genetics. Many naturally occurring mutations are known, often mimicking the effects of corresponding mutations in humans (Figure 1–46). Methods have been developed, moreover, to test the function of any chosen mouse gene, or of any noncoding portion of the mouse genome, by artificially creating mutations in it, as we explain later in the book.

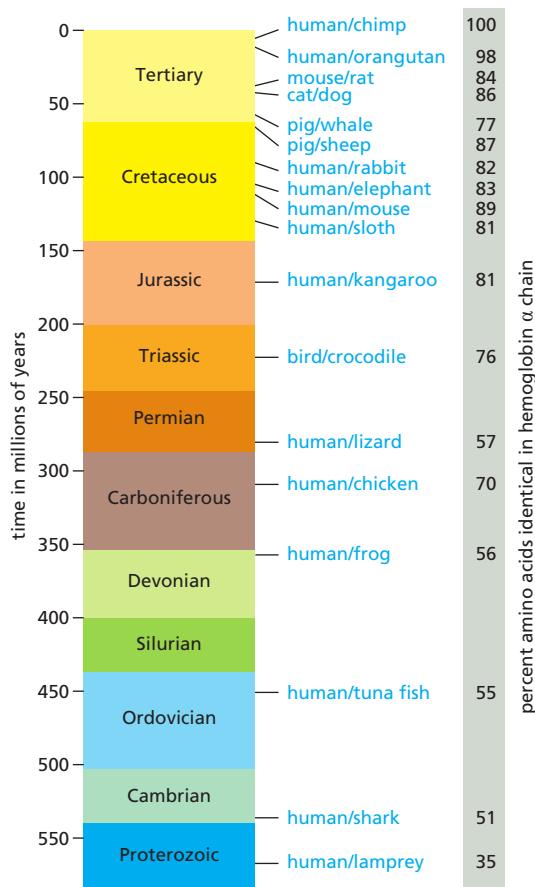
Just one made-to-order mutant mouse can provide a wealth of information for the cell biologist. It reveals the effects of the chosen mutation in a host of different contexts, simultaneously testing the action of the gene in all the different kinds of cells in the body that could in principle be affected.

### Humans Report on Their Own Peculiarities

As humans, we have a special interest in the human genome. We want to know the full set of parts from which we are made, and to discover how they work. But even



**Figure 1–44** Zebrafish as a model for studies of vertebrate development. These small, hardy tropical fish are convenient for genetic studies. Additionally, they have transparent embryos that develop outside of the mother, so that one can clearly observe cells moving and changing their character in the living organism throughout its development. (A) Adult fish. (B) An embryo 24 hours after fertilization. (A, with permission from Steve Baskau; B, from M. Rhinn et al., *Neural Dev.* 4:12, 2009.)



**Figure 1–45 Times of divergence of different vertebrates.** The scale on the left shows the estimated date and geological era of the last common ancestor of each specified pair of animals. Each time estimate is based on comparisons of the amino acid sequences of orthologous proteins; the longer the animals of a pair have had to evolve independently, the smaller the percentage of amino acids that remain identical. The time scale has been calibrated to match the fossil evidence showing that the last common ancestor of mammals and birds lived 310 million years ago.

The figures on the right give data on sequence divergence for one particular protein—the  $\alpha$  chain of hemoglobin. Note that although there is a clear general trend of increasing divergence with increasing time for this protein, there are irregularities that are thought to reflect the action of natural selection driving especially rapid changes of hemoglobin sequence when the organisms experienced special physiological demands. Some proteins, subject to stricter functional constraints, evolve much more slowly than hemoglobin, others as much as five times faster. All this gives rise to substantial uncertainties in estimates of divergence times, and some experts believe that the major groups of mammals diverged from one another as much as 60 million years more recently than shown here. (Adapted from S. Kumar and S.B. Hedges, *Nature* 392:917–920, 1998. With permission from Macmillan Publishers Ltd.)

if you were a mouse, preoccupied with the molecular biology of mice, humans would be attractive as model genetic organisms, because of one special property: through medical examinations and self-reporting, we catalog our own genetic (and other) disorders. The human population is enormous, consisting today of some 7 billion individuals, and this self-documenting property means that a huge database of information exists on human mutations. The human genome sequence of more than 3 billion nucleotide pairs has been determined for thousands of different people, making it easier than ever before to identify at a molecular level the precise genetic change responsible for any given human mutant phenotype.

By drawing together the insights from humans, mice, fish, flies, worms, yeasts, plants, and bacteria—using gene sequence similarities to map out the correspondences between one model organism and another—we are enriching our understanding of them all.



**Figure 1–46 Human and mouse: similar genes and similar development.** The human baby and the mouse shown here have similar white patches on their foreheads because both have mutations in the same gene (called *Kit*), required for the development and maintenance of pigment cells. (Courtesy of R.A. Fleischman.)

## We Are All Different in Detail

What precisely do we mean when we speak of *the* human genome? Whose genome? On average, any two people taken at random differ in about one or two in every 1000 nucleotide pairs in their DNA sequence. The genome of the human species is, properly speaking, a very complex thing, embracing the entire pool of variant genes found in the human population. Knowledge of this variation is helping us to understand, for example, why some people are prone to one disease, others to another; why some respond well to a drug, others badly. It is also providing clues to our history—the population movements and minglings of our ancestors, the infections they suffered, the diets they ate. All these things have left traces in the variant forms of genes that survive today in the human communities that populate the globe.

## To Understand Cells and Organisms Will Require Mathematics, Computers, and Quantitative Information

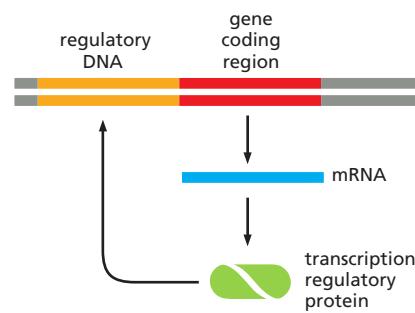
Empowered by knowledge of complete genome sequences, we can list the genes, proteins, and RNA molecules in a cell, and we have methods that allow us to begin to depict the complex web of interactions between them. But how are we to turn all this information into an understanding of how cells work? Even for a single cell type belonging to a single species of organism, the current deluge of data seems overwhelming. The sort of informal reasoning on which biologists usually rely seems totally inadequate in the face of such complexity.

In fact, the difficulty is more than just a matter of information overload. Biological systems are, for example, full of feedback loops, and the behavior of even the simplest of systems with feedback is remarkably difficult to predict by intuition alone (Figure 1–47); small changes in parameters can cause radical changes in outcome. To go from a circuit diagram to a prediction of the behavior of the system, we need detailed quantitative information, and to draw deductions from that information we need mathematics and computers.

Such tools for quantitative reasoning are essential, but they are not all-powerful. You might think that, knowing how each protein influences each other protein, and how the expression of each gene is regulated by the products of others, we should soon be able to calculate how the cell as a whole will behave, just as an astronomer can calculate the orbits of the planets, or a chemical engineer can calculate the flows through a chemical plant. But any attempt to perform this feat for anything close to an entire living cell rapidly reveals the limits of our present knowledge. The information we have, plentiful as it is, is full of gaps and uncertainties. Moreover, it is largely qualitative rather than quantitative. Most often, cell biologists studying the cell's control systems sum up their knowledge in simple schematic diagrams—this book is full of them—rather than in numbers, graphs, and differential equations.

To progress from qualitative descriptions and intuitive reasoning to quantitative descriptions and mathematical deduction is one of the biggest challenges for contemporary cell biology. So far, the challenge has been met only for a few very simple fragments of the machinery of living cells—subsystems involving a handful of different proteins, or two or three cross-regulatory genes, where theory and experiment go closely hand in hand. We discuss some of these examples later in the book and devote the entire final section of Chapter 8 to the role of quantitation in cell biology.

Knowledge and understanding bring the power to intervene—with humans, to avoid or prevent disease; with plants, to create better crops; with bacteria, to turn them to our own uses. All these biological enterprises are linked, because the genetic information of all living organisms is written in the same language. The new-found ability of molecular biologists to read and decipher this language has already begun to transform our relationship to the living world. The account of cell biology in the subsequent chapters will, we hope, equip the reader to understand, and possibly to contribute to, the great scientific adventure of the twenty-first century.



**Figure 1–47** A very simple regulatory circuit—a single gene regulating its own expression by the binding of its protein product to its own regulatory DNA. Simple schematic diagrams such as this are found throughout this book. They are often used to summarize what we know, but they leave many questions unanswered. When the protein binds, does it inhibit or stimulate transcription from the gene? How steeply does the transcription rate depend on the protein concentration? How long, on average, does a molecule of the protein remain bound to the DNA? How long does it take to make each molecule of mRNA or protein, and how quickly does each type of molecule get degraded? As explained in Chapter 8, mathematical modeling shows that we need quantitative answers to all these and other questions before we can predict the behavior of even this single-gene system. For different parameter values, the system may settle to a unique steady state; or it may behave as a switch, capable of existing in one or another of a set of alternative states; or it may oscillate; or it may show large random fluctuations.

## Summary

*Eukaryotic cells, by definition, keep their DNA in a separate membrane-enclosed compartment, the nucleus. They have, in addition, a cytoskeleton for support and movement, elaborate intracellular compartments for digestion and secretion, the capacity (in many species) to engulf other cells, and a metabolism that depends on the oxidation of organic molecules by mitochondria. These properties suggest that eukaryotes may have originated as predators on other cells. Mitochondria—and, in plants, chloroplasts—contain their own genetic material, and they evidently evolved from bacteria that were taken up into the cytoplasm of ancient cells and survived as symbionts.*

*Eukaryotic cells typically have 3–30 times as many genes as prokaryotes, and often thousands of times more noncoding DNA. The noncoding DNA allows for great complexity in the regulation of gene expression, as required for the construction of complex multicellular organisms. Many eukaryotes are, however, unicellular—among them the yeast *Saccharomyces cerevisiae*, which serves as a simple model organism for eukaryotic cell biology, revealing the molecular basis of many fundamental processes that have been strikingly conserved during a billion years of evolution. A small number of other organisms have also been chosen for intensive study: a worm, a fly, a fish, and the mouse serve as “model organisms” for multicellular animals; and a small milkweed serves as a model for plants.*

*Powerful new technologies such as genome sequencing are producing striking advances in our knowledge of human beings, and they are helping to advance our understanding of human health and disease. But living systems are incredibly complex, and mammalian genomes contain multiple closely related homologs of most genes. This genetic redundancy has allowed diversification and specialization of genes for new purposes, but it also makes biological mechanisms harder to decipher. For this reason, simpler model organisms have played a key part in revealing universal genetic mechanisms of animal development, and research using these systems remains critical for driving scientific and medical advances.*

## WHAT WE DON'T KNOW

- What new approaches might provide a clearer view of the anaerobic archaeon that is thought to have formed the nucleus of the first eukaryotic cell? How did its symbiosis with an aerobic bacterium lead to the mitochondrion? Somewhere on Earth, are there cells not yet identified that can fill in the details of how eukaryotic cells originated?
- DNA sequencing has revealed a rich and previously undiscovered world of microbial cells, the vast majority of which fail to grow in a laboratory. How might these cells be made more accessible for detailed study?
- What new model cells or organisms should be developed for scientists to study? Why might a concerted focus on these models speed progress toward understanding a critical aspect of cell function that is poorly understood?
- How did the first cell membranes arise?

## PROBLEMS

Which statements are true? Explain why or why not.

**1–1** Each member of the human hemoglobin gene family, which consists of seven genes arranged in two clusters on different chromosomes, is an ortholog to all of the other members.

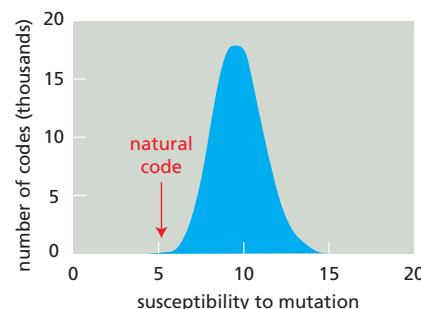
**1–2** Horizontal gene transfer is more prevalent in single-celled organisms than in multicellular organisms.

**1–3** Most of the DNA sequences in a bacterial genome code for proteins, whereas most of the DNA sequences in the human genome do not.

Discuss the following problems.

**1–4** Since it was deciphered four decades ago, some have claimed that the genetic code must be a frozen accident, while others have argued that it was shaped by natural selection. A striking feature of the genetic code is its inherent resistance to the effects of mutation. For example, a change in the third position of a codon often specifies the same amino acid or one with similar chemical properties. The natural code resists mutation more effectively (is less susceptible to error) than most other possible versions, as

illustrated in **Figure Q1–1**. Only one in a million computer-generated “random” codes is more error-resistant than the natural genetic code. Does the extraordinary mutation resistance of the genetic code argue in favor of its origin as a frozen accident or as a result of natural selection? Explain your reasoning.



**Figure Q1–1** Susceptibility to mutation of the natural code shown relative to that of millions of computer-generated alternative genetic codes (Problem 1–4). Susceptibility measures the average change in amino acid properties caused by random mutations in a genetic code. A small value indicates that mutations tend to cause minor changes. (Data courtesy of Steve Freeland.)

**1–5** You have begun to characterize a sample obtained from the depths of the oceans on Europa, one of Jupiter's moons. Much to your surprise, the sample contains a life-form that grows well in a rich broth. Your preliminary analysis shows that it is cellular and contains DNA, RNA, and protein. When you show your results to a colleague, she suggests that your sample was contaminated with an organism from Earth. What approaches might you try to distinguish between contamination and a novel cellular life-form based on DNA, RNA, and protein?

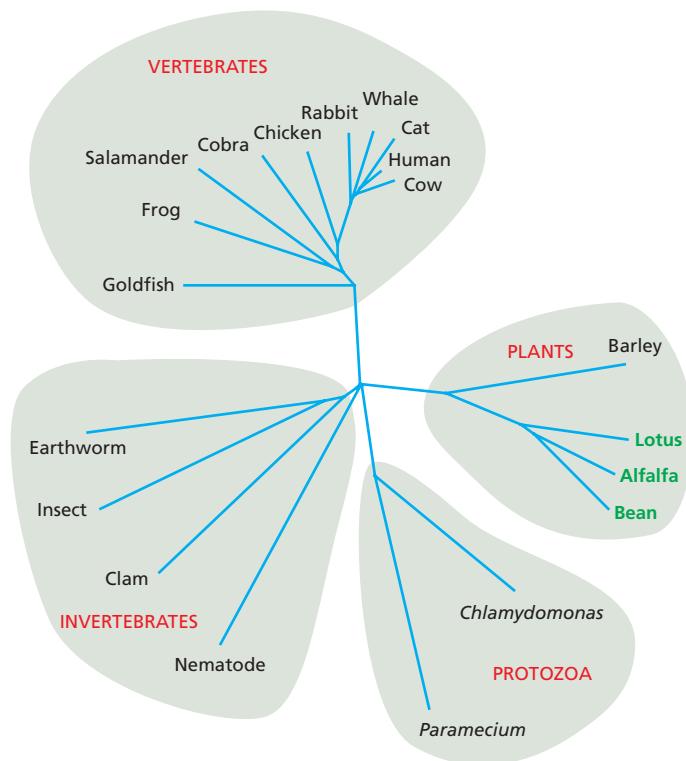
**1–6** It is not so difficult to imagine what it means to feed on the organic molecules that living things produce. That is, after all, what we do. But what does it mean to "feed" on sunlight, as phototrophs do? Or, even stranger, to "feed" on rocks, as lithotrophs do? Where is the "food," for example, in the mixture of chemicals ( $H_2S$ ,  $H_2$ , CO,  $Mn^+$ ,  $Fe^{2+}$ ,  $Ni^{2+}$ ,  $CH_4$ , and  $NH_4^+$ ) that spews from a hydrothermal vent?

**1–7** How many possible different trees (branching patterns) can in theory be drawn to display the evolution of bacteria, archaea, and eukaryotes, assuming that they all arose from a common ancestor?

**1–8** The genes for ribosomal RNA are highly conserved (relatively few sequence changes) in all organisms on Earth; thus, they have evolved very slowly over time. Were ribosomal RNA genes "born" perfect?

**1–9** Genes participating in informational processes such as replication, transcription, and translation are transferred between species much less often than are genes involved in metabolism. The basis for this inequality is unclear at present, but one suggestion is that it relates to the underlying complexity of the two types of processes. Informational processes tend to involve large aggregates of different gene products, whereas metabolic reactions are usually catalyzed by enzymes composed of a single protein. Why would the complexity of the underlying process—informational or metabolic—have any effect on the rate of horizontal gene transfer?

**1–10** Animal cells have neither cell walls nor chloroplasts, whereas plant cells have both. Fungal cells are somewhere in between; they have cell walls but lack chloroplasts. Are fungal cells more likely to be animal cells that gained the ability to make cell walls, or plant cells that lost their chloroplasts? This question represented a difficult issue for early investigators who sought to assign evolutionary relationships based solely on cell characteristics and morphology. How do you suppose that this question was eventually decided?



**Figure Q1–2** Phylogenetic tree for hemoglobin genes from a variety of species (Problem 1–11). The legumes are highlighted in green. The lengths of lines that connect the present-day species represent the evolutionary distances that separate them.

**1–11** When plant hemoglobin genes were first discovered in legumes, it was so surprising to find a gene typical of animal blood that it was hypothesized that the plant gene arose by horizontal transfer from an animal. Many more hemoglobin genes have now been sequenced, and a phylogenetic tree based on some of these sequences is shown in **Figure Q1–2**.

- A.** Does this tree support or refute the hypothesis that the plant hemoglobins arose by horizontal gene transfer?
- B.** Supposing that the plant hemoglobin genes were originally derived from a parasitic nematode, for example, what would you expect the phylogenetic tree to look like?

**1–12** Rates of evolution appear to vary in different lineages. For example, the rate of evolution in the rat lineage is significantly higher than in the human lineage. These rate differences are apparent whether one looks at changes in nucleotide sequences that encode proteins and are subject to selective pressure or at changes in noncoding nucleotide sequences, which are not under obvious selection pressure. Can you offer one or more possible explanations for the slower rate of evolutionary change in the human lineage versus the rat lineage?

## REFERENCES

### General

- Alberts B, Bray D, Hopkin K et al. (2014) Essential Cell Biology, 4th ed. New York: Garland Science.
- Barton NH, Briggs DEG, Eisen JA et al. (2007) Evolution. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Darwin C (1859) On the Origin of Species. London: Murray.
- Graur D & Li W-H (1999) Fundamentals of Molecular Evolution, 2nd ed. Sunderland, MA: Sinauer Associates.
- Madigan MT, Martinko JM, Stahl D et al. (2010) Brock Biology of Microorganisms, 13th ed. Menlo Park, CA: Benjamin-Cummings.
- Margulis L & Chapman MJ (2009) Kingdoms and Domains: An Illustrated Guide to the Phyla of Life on Earth, 1st ed. San Diego: Academic Press.
- Moore JA (1993) Science As a Way of Knowing. Cambridge, MA: Harvard University Press.
- Moore JA (1972) Heredity and Development, 2nd ed. New York: Oxford University Press. (Free download at [www.nap.edu](http://www.nap.edu))
- Yang Z (2014) Molecular Evolution: A Statistical Approach. Oxford: Oxford University Press.

### The Universal Features of Cells On Earth

- Andersson SGE (2006) The bacterial world gets smaller. *Science* 314, 259–260.
- Brenner S, Jacob F & Meselson M (1961) An unstable intermediate carrying information from genes to ribosomes for protein synthesis. *Nature* 190, 576–581.
- Deamer D & Szostak JW eds. (2010) The Origins of Life (Cold Spring Harbor Perspectives in Biology). NY: Cold Spring Harbor Laboratory Press.
- Gibson DG, Benders GA, Andrews-Pfannkoch C et al. (2008) Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome. *Science* 319, 1215–1220.
- Glass JI, Assad-Garcia N, Alperovich N et al. (2006) Essential genes of a minimal bacterium. *Proc. Natl Acad. Sci. USA* 103, 425–430.
- Harris JK, Kelley ST, Spiegelman GB et al. (2003) The genetic core of the universal ancestor. *Genome Res.* 13, 407–413.
- Koonin EV (2005) Orthologs, paralogs, and evolutionary genomics. *Annu. Rev. Genet.* 39, 309–338.
- Noller H (2005) RNA structure: reading the ribosome. *Science* 309, 1508–1514.
- Rinke C, Schwientek P, Sczyrba A et al. (2013) Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499, 431–437.
- Watson JD & Crick FHC (1953) Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. *Nature* 171, 737–738.

### The Diversity of Genomes and the Tree of Life

- Blattner FR, Plunkett G, Bloch CA et al. (1997) The complete genome sequence of *Escherichia coli* K-12. *Science* 277, 1453–1474.
- Boucher Y, Douady CJ, Papke RT et al. (2003) Lateral gene transfer and the origins of prokaryotic groups. *Annu. Rev. Genet.* 37, 283–328.
- Cavicchioli R (2010) Archaea—timeline of the third domain. *Nat. Rev. Microbiol.* 9, 51–61.
- Choudhuri S (2014) Bioinformatics for Beginners: Genes, Genomes, Molecular Evolution, Databases and Analytical Tools, 1st ed. San Diego: Academic Press.
- Dixon B (1997) Power Unseen: How Microbes Rule the World. Oxford: Oxford University Press.
- Handelsman J (2004) Metagenomics: applications of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* 68, 669–685.
- Kerr RA (1997) Life goes to extremes in the deep earth—and elsewhere? *Science* 276, 703–704.
- Lee TI, Rinaldi NJ, Robert F et al. (2002) Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804.

Olsen GJ & Woese CR (1997) Archaeal genomics: an overview. *Cell* 89:991–994.

- Williams TA, Foster PG, Cox CJ & Embley TM (2013) An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504, 231–235.
- Woese C (1998) The universal ancestor. *Proc. Natl Acad. Sci. USA* 95, 6854–6859.

### Genetic Information in Eukaryotes

- Adams MD, Celtniker SE, Holt RA et al. (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195.
- Amborella Genome Project (2013) The *Amborella* genome and the evolution of flowering plants. *Science* 342, 1241093.
- Andersson SG, Zomorodipour A, Andersson JO et al. (1998) The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* 396, 133–140.
- The *Arabidopsis* Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815.
- Carroll SB, Grenier JK & Weatherbee SD (2005) From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design, 2nd ed. Maldon, MA: Blackwell Science.
- de Duve C (2007) The origin of eukaryotes: a reappraisal. *Nat. Rev. Genet.* 8, 395–403.
- Delsuc F, Brinkmann H & Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375.
- DeRisi JL, Iyer VR & Brown PO (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* 278, 680–686.
- Gabriel SB, Schaffner SF, Nguyen H et al. (2002) The structure of haplotype blocks in the human genome. *Science* 296, 2225–2229.
- Goffeau A, Barrell BG, Bussey H et al. (1996) Life with 6000 genes. *Science* 274, 546–567.
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Keeling PJ & Koonin EV eds. (2014) The Origin and Evolution of Eukaryotes (Cold Spring Harbor Perspectives in Biology). NY: Cold Spring Harbor Laboratory Press.
- Lander ES (2011) Initial impact of the sequencing of the human genome. *Nature* 470, 187–197.
- Lynch M & Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155.
- National Center for Biotechnology Information. <http://www.ncbi.nlm.nih.gov/>
- Owens K & King MC (1999) Genomic views of human history. *Science* 286, 451–453.
- Palmer JD & Delwiche CF (1996) Second-hand chloroplasts and the case of the disappearing nucleus. *Proc. Natl Acad. Sci. USA* 93, 7432–7435.
- Reed FA & Tishkoff SA (2006) African human diversity, origins and migrations. *Curr. Opin. Genet. Dev.* 16, 597–605.
- Rine J (2014) A future of the model organism model. *Mol. Biol. Cell* 25, 549–553.
- Rubin GM, Yandell MD, Wortman JR et al. (2000) Comparative genomics of the eukaryotes. *Science* 287, 2204–2215.
- Shen Y, Yue F, McCleary D et al. (2012) A map of the *cis*-regulatory sequences in the mouse genome. *Nature* 488, 116–120.
- The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018.
- Tinsley RC & Kobel HR eds. (1996) The Biology of *Xenopus*. Oxford: Clarendon Press.
- Tyson JJ, Chen KC & Novak B (2003) Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* 15, 221–231.
- Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. *Science* 291, 1304–1351.

This page intentionally left blank  
to match pagination of print book

# Cell Chemistry and Bioenergetics

## CHAPTER 2

It is at first sight difficult to accept the idea that living creatures are merely chemical systems. Their incredible diversity of form, their seemingly purposeful behavior, and their ability to grow and reproduce all seem to set them apart from the world of solids, liquids, and gases that chemistry normally describes. Indeed, until the nineteenth century animals were believed to contain a Vital Force—an “animus”—that was responsible for their distinctive properties.

We now know that there is nothing in living organisms that disobeys chemical or physical laws. However, the chemistry of life is indeed special. First, it is based overwhelmingly on carbon compounds, the study of which is known as *organic chemistry*. Second, cells are 70% water, and life depends largely on chemical reactions that take place in aqueous solution. Third, and most important, cell chemistry is enormously complex: even the simplest cell is vastly more complicated in its chemistry than any other chemical system known. In particular, although cells contain a variety of small carbon-containing molecules, most of the carbon atoms present are incorporated into enormous polymeric molecules—chains of chemical subunits linked end-to-end. It is the unique properties of these *macromolecules* that enable cells and organisms to grow and reproduce—as well as to do all the other things that are characteristic of life.

## THE CHEMICAL COMPONENTS OF A CELL

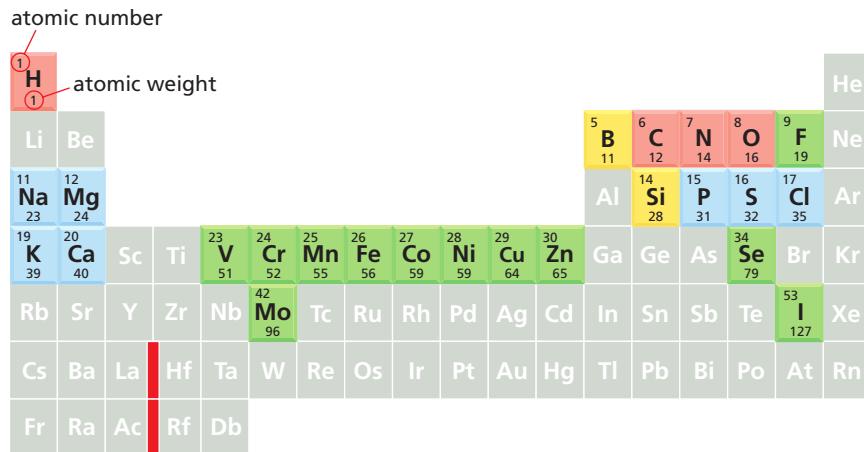
Living organisms are made of only a small selection of the 92 naturally occurring elements, four of which—carbon (C), hydrogen (H), nitrogen (N), and oxygen (O)—make up 96.5% of an organism’s weight (Figure 2–1). The atoms of these elements are linked together by **covalent bonds** to form *molecules* (see Panel 2–1, pp. 90–91). Because covalent bonds are typically 100 times stronger than the thermal energies within a cell, they resist being pulled apart by thermal motions, and they are normally broken only during specific chemical reactions with other atoms and molecules. Two different molecules can be held together by *noncovalent bonds*,

### IN THIS CHAPTER

#### THE CHEMICAL COMPONENTS OF A CELL

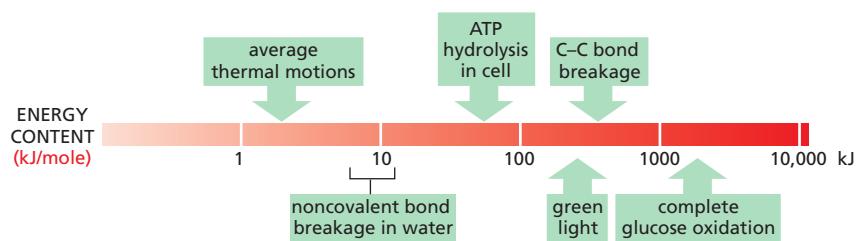
#### CATALYSIS AND THE USE OF ENERGY BY CELLS

#### HOW CELLS OBTAIN ENERGY FROM FOOD



**Figure 2–1** The main elements in cells, highlighted in the periodic table. When ordered by their atomic number and arranged in this manner, elements fall into vertical columns that show similar properties. Atoms in the same vertical column must gain (or lose) the same number of electrons to attain a filled outer shell, and they thus behave similarly in bond or ion formation. Thus, for example, Mg and Ca tend to give away the two electrons in their outer shells. C, N, and O occur in the same horizontal row, and tend to complete their second shells by sharing electrons.

The four elements highlighted in red constitute 99% of the total number of atoms present in the human body. An additional seven elements, highlighted in blue, together represent about 0.9% of the total. The elements shown in green are required in trace amounts by humans. It remains unclear whether those elements shown in yellow are essential in humans. The chemistry of life, it seems, is therefore predominantly the chemistry of lighter elements. The atomic weights shown here are those of the most common isotope of each element.



which are much weaker (**Figure 2–2**). We shall see later that noncovalent bonds are important in the many situations where molecules have to associate and dissociate readily to carry out their biological functions.

### Water Is Held Together by Hydrogen Bonds

The reactions inside a cell occur in an aqueous environment. Life on Earth began in the ocean, and the conditions in that primeval environment put a permanent stamp on the chemistry of living things. Life therefore hinges on the chemical properties of water, which are reviewed in **Panel 2–2**, pp. 92–93.

In each water molecule ( $\text{H}_2\text{O}$ ) the two H atoms are linked to the O atom by covalent bonds. The two bonds are highly polar because the O is strongly attractive for electrons, whereas the H is only weakly attractive. Consequently, there is an unequal distribution of electrons in a water molecule, with a preponderance of positive charge on the two H atoms and of negative charge on the O. When a positively charged region of one water molecule (that is, one of its H atoms) approaches a negatively charged region (that is, the O) of a second water molecule, the electrical attraction between them can result in a *hydrogen bond*. These bonds are much weaker than covalent bonds and are easily broken by the random thermal motions that reflect the heat energy of the molecules. Thus, each bond lasts only a short time. But the combined effect of many weak bonds can be profound. For example, each water molecule can form hydrogen bonds through its two H atoms to two other water molecules, producing a network in which hydrogen bonds are being continually broken and formed. It is only because of the hydrogen bonds that link water molecules together that water is a liquid at room temperature—with a high boiling point and high surface tension—rather than a gas.

Molecules, such as alcohols, that contain polar bonds and that can form hydrogen bonds with water dissolve readily in water. Molecules carrying charges (ions) likewise interact favorably with water. Such molecules are termed **hydrophilic**, meaning that they are water-loving. Many of the molecules in the aqueous environment of a cell necessarily fall into this category, including sugars, DNA, RNA, and most proteins. **Hydrophobic** (water-hating) molecules, by contrast, are uncharged and form few or no hydrogen bonds, and so do not dissolve in water. Hydrocarbons are an important example. In these molecules all of the H atoms are covalently linked to C atoms by a largely nonpolar bond; thus they cannot form effective hydrogen bonds to other molecules (see **Panel 2–1**, p. 90). This makes the hydrocarbon as a whole hydrophobic—a property that is exploited in cells, whose membranes are constructed from molecules that have long hydrocarbon tails, as we see in Chapter 10.

### Four Types of Noncovalent Attractions Help Bring Molecules Together in Cells

Much of biology depends on the specific binding of different molecules caused by three types of noncovalent bonds: **electrostatic attractions** (ionic bonds), **hydrogen bonds**, and **van der Waals attractions**; and on a fourth factor that can push molecules together: the **hydrophobic force**. The properties of the four types of noncovalent attractions are presented in **Panel 2–3** (pp. 94–95). Although each

**Figure 2–2 Some energies important for cells.** A crucial property of any bond—covalent or noncovalent—is its strength. *Bond strength* is measured by the amount of energy that must be supplied to break it, expressed in units of either kilojoules per mole (kJ/mole) or kilocalories per mole (kcal/mole). Thus if 100 kJ of energy must be supplied to break  $6 \times 10^{23}$  bonds of a specific type (that is, 1 mole of these bonds), then the strength of that bond is 100 kJ/mole. Note that, in this diagram, energies are compared on a logarithmic scale. Typical strengths and lengths of the main classes of chemical bonds are given in Table 2–1.

One joule (J) is the amount of energy required to move an object a distance of one meter against a force of one Newton. This measure of energy is derived from the SI units (Système Internationale d'Unités) universally employed by physical scientists. A second unit of energy, often used by cell biologists, is the kilocalorie (kcal); one calorie is the amount of energy needed to raise the temperature of 1 gram of water by  $1^{\circ}\text{C}$ . One kJ is equal to 0.239 kcal (1 kcal = 4.18 kJ).

**Figure 2–3 Schematic indicating how two macromolecules with complementary surfaces can bind tightly to one another through noncovalent interactions.** Noncovalent chemical bonds have less than 1/20 the strength of a covalent bond. They are able to produce tight binding only when many of them are formed simultaneously. Although only electrostatic attractions are illustrated here, in reality all four noncovalent forces often contribute to holding two macromolecules together (Movie 2.1).

individual noncovalent attraction would be much too weak to be effective in the face of thermal motions, their energies can sum to create a strong force between two separate molecules. Thus sets of noncovalent attractions often allow the complementary surfaces of two macromolecules to hold those two macromolecules together (**Figure 2–3**).

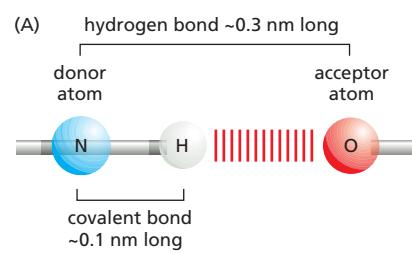
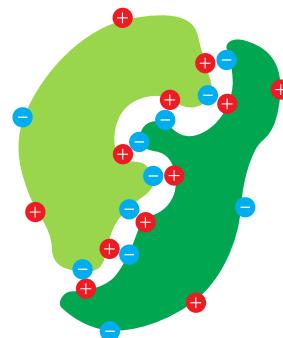
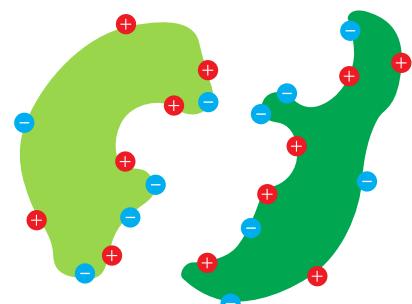
**Table 2–1** compares noncovalent bond strengths to that of a typical covalent bond, both in the presence and in the absence of water. Note that, by forming competing interactions with the involved molecules, water greatly reduces the strength of both electrostatic attractions and hydrogen bonds.

The structure of a typical hydrogen bond is illustrated in **Figure 2–4**. This bond represents a special form of polar interaction in which an electropositive hydrogen atom is shared by two electronegative atoms. Its hydrogen can be viewed as a proton that has partially dissociated from a donor atom, allowing it to be shared by a second acceptor atom. Unlike a typical electrostatic interaction, this bond is highly directional—being strongest when a straight line can be drawn between all three of the involved atoms.

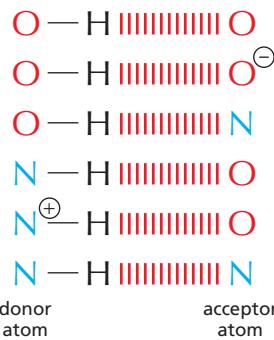
The fourth effect that often brings molecules together in water is not, strictly speaking, a bond at all. However, a very important hydrophobic force is caused by a pushing of nonpolar surfaces out of the hydrogen-bonded water network, where they would otherwise physically interfere with the highly favorable interactions between water molecules. Bringing any two nonpolar surfaces together reduces their contact with water; in this sense, the force is nonspecific. Nevertheless, we shall see in Chapter 3 that hydrophobic forces are central to the proper folding of protein molecules.

### Some Polar Molecules Form Acids and Bases in Water

One of the simplest kinds of chemical reaction, and one that has profound significance in cells, takes place when a molecule containing a highly polar covalent bond between a hydrogen and another atom dissolves in water. The hydrogen atom in such a molecule has given up its electron almost entirely to the companion atom, and so exists as an almost naked positively charged hydrogen nucleus—in



(B)

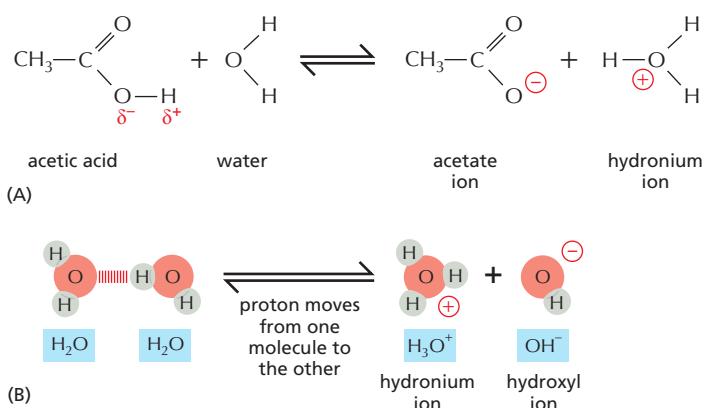


**Figure 2–4 Hydrogen bonds.** (A) Ball-and-stick model of a typical hydrogen bond. The distance between the hydrogen and the oxygen atom here is less than the sum of their van der Waals radii, indicating a partial sharing of electrons. (B) The most common hydrogen bonds in cells.

**TABLE 2–1 Covalent and Noncovalent Chemical Bonds**

Bond type	Length (nm)	Strength kJ/mole**	
		in vacuum	in water
Covalent	0.15	377 (90)	377 (90)
Noncovalent	ionic*	335 (80)	12.6 (3)
	hydrogen	16.7 (4)	4.2 (1)
	van der Waals attraction (per atom)	0.4 (0.1)	0.4 (0.1)

\*An ionic bond is an electrostatic attraction between two fully charged atoms. \*\*Values in parentheses are kcal/mole. 1 kJ = 0.239 kcal and 1 kcal = 4.18 kJ.



**Figure 2–5 Protons readily move in aqueous solutions.** (A) The reaction that takes place when a molecule of acetic acid dissolves in water. At pH 7, nearly all of the acetic acid is present as acetate ion. (B) Water molecules are continuously exchanging protons with each other to form hydronium and hydroxyl ions. These ions in turn rapidly recombine to form water molecules.

other words, a **proton** ( $\text{H}^+$ ). When the polar molecule becomes surrounded by water molecules, the proton will be attracted to the partial negative charge on the O atom of an adjacent water molecule. This proton can easily dissociate from its original partner and associate instead with the oxygen atom of the water molecule, generating a **hydronium ion** ( $\text{H}_3\text{O}^+$ ) (Figure 2–5A). The reverse reaction also takes place very readily, so in the aqueous solution protons are constantly flitting to and fro between one molecule and another.

Substances that release protons when they dissolve in water, thus forming  $\text{H}_3\text{O}^+$ , are termed **acids**. The higher the concentration of  $\text{H}_3\text{O}^+$ , the more acidic the solution.  $\text{H}_3\text{O}^+$  is present even in pure water, at a concentration of  $10^{-7}$  M, as a result of the movement of protons from one water molecule to another (Figure 2–5B). By convention, the  $\text{H}_3\text{O}^+$  concentration is usually referred to as the  $\text{H}^+$  concentration, even though most protons in an aqueous solution are present as  $\text{H}_3\text{O}^+$ . To avoid the use of unwieldy numbers, the concentration of  $\text{H}_3\text{O}^+$  is expressed using a logarithmic scale called the **pH scale**. Pure water has a pH of 7.0 and is said to be neutral—that is, neither acidic (pH < 7) nor basic (pH > 7).

Acids are characterized as being strong or weak, depending on how readily they give up their protons to water. Strong acids, such as hydrochloric acid (HCl), lose their protons quickly. Acetic acid, on the other hand, is a weak acid because it holds on to its proton more tightly when dissolved in water. Many of the acids important in the cell—such as molecules containing a carboxyl ( $\text{COOH}$ ) group—are weak acids (see Panel 2–2, pp. 92–93).

Because the proton of a hydronium ion can be passed readily to many types of molecules in cells, altering their character, the concentration of  $\text{H}_3\text{O}^+$  inside a cell (the acidity) must be closely regulated. Acids—especially weak acids—will give up their protons more readily if the concentration of  $\text{H}_3\text{O}^+$  in solution is low and will tend to receive them back if the concentration in solution is high.

The opposite of an acid is a **base**. Any molecule capable of accepting a proton from a water molecule is called a base. Sodium hydroxide ( $\text{NaOH}$ ) is basic (the term *alkaline* is also used) because it dissociates readily in aqueous solution to form  $\text{Na}^+$  ions and  $\text{OH}^-$  ions. Because of this property,  $\text{NaOH}$  is called a strong base. More important in living cells, however, are the weak bases—those that have a weak tendency to reversibly accept a proton from water. Many biologically important molecules contain an amino ( $\text{NH}_2$ ) group. This group is a weak base that can generate  $\text{OH}^-$  by taking a proton from water:  $-\text{NH}_2 + \text{H}_2\text{O} \rightarrow -\text{NH}_3^+ + \text{OH}^-$  (see Panel 2–2, pp. 92–93).

Because an  $\text{OH}^-$  ion combines with a  $\text{H}_3\text{O}^+$  ion to form two water molecules, an increase in the  $\text{OH}^-$  concentration forces a decrease in the concentration of  $\text{H}_3\text{O}^+$ , and vice versa. A pure solution of water contains an equal concentration ( $10^{-7}$  M) of both ions, rendering it neutral. The interior of a cell is also kept close to neutrality by the presence of **buffers**: weak acids and bases that can release or take up protons near pH 7, keeping the environment of the cell relatively constant under a variety of conditions.

## A Cell Is Formed from Carbon Compounds

Having reviewed the ways atoms combine into molecules and how these molecules behave in an aqueous environment, we now examine the main classes of small molecules found in cells. We shall see that a few categories of molecules, formed from a handful of different elements, give rise to all the extraordinary richness of form and behavior shown by living things.

If we disregard water and inorganic ions such as potassium, nearly all the molecules in a cell are based on carbon. Carbon is outstanding among all the elements in its ability to form large molecules; silicon is a poor second. Because carbon is small and has four electrons and four vacancies in its outermost shell, a carbon atom can form four covalent bonds with other atoms. Most important, one carbon atom can join to other carbon atoms through highly stable covalent C–C bonds to form chains and rings and hence generate large and complex molecules with no obvious upper limit to their size. The carbon compounds made by cells are called *organic molecules*. In contrast, all other molecules, including water, are said to be *inorganic*.

Certain combinations of atoms, such as the methyl ( $-\text{CH}_3$ ), hydroxyl ( $-\text{OH}$ ), carboxyl ( $-\text{COOH}$ ), carbonyl ( $-\text{C}=\text{O}$ ), phosphate ( $-\text{PO}_3^{2-}$ ), sulphydryl ( $-\text{SH}$ ), and amino ( $-\text{NH}_2$ ) groups, occur repeatedly in the molecules made by cells. Each such **chemical group** has distinct chemical and physical properties that influence the behavior of the molecule in which the group occurs. The most common chemical groups and some of their properties are summarized in Panel 2–1, pp. 90–91.

## Cells Contain Four Major Families of Small Organic Molecules

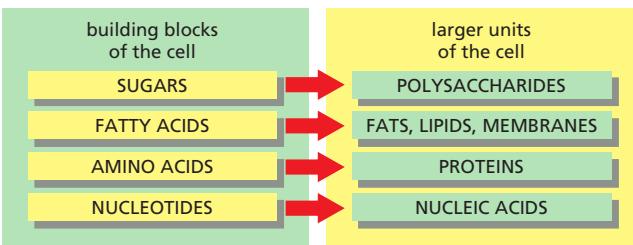
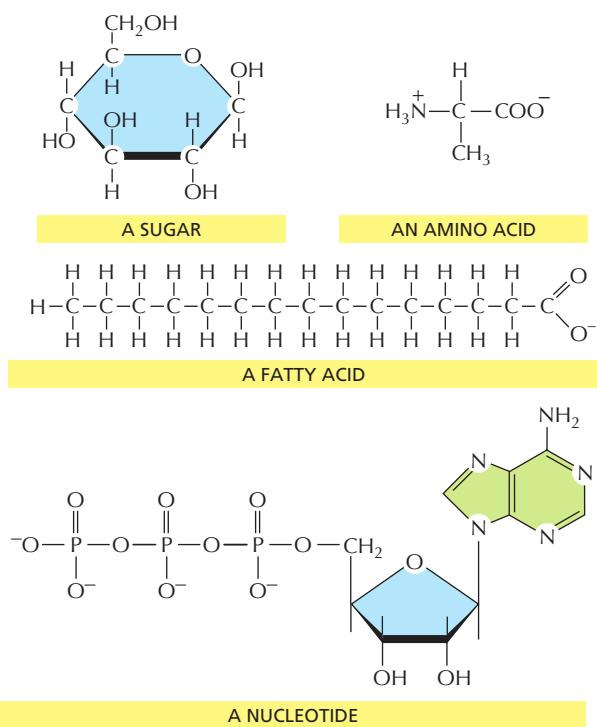
The small organic molecules of the cell are carbon-based compounds that have molecular weights in the range of 100–1000 and contain up to 30 or so carbon atoms. They are usually found free in solution and have many different fates. Some are used as *monomer* subunits to construct giant polymeric *macromolecules*—proteins, nucleic acids, and large polysaccharides. Others act as energy sources and are broken down and transformed into other small molecules in a maze of intracellular metabolic pathways. Many small molecules have more than one role in the cell—for example, acting both as a potential subunit for a macromolecule and as an energy source. Small organic molecules are much less abundant than the organic macromolecules, accounting for only about one-tenth of the total mass of organic matter in a cell. As a rough guess, there may be a thousand different kinds of these small molecules in a typical cell.

All organic molecules are synthesized from and are broken down into the same set of simple compounds. As a consequence, the compounds in a cell are chemically related and most can be classified into a few distinct families. Broadly speaking, cells contain four major families of small organic molecules: the *sugars*, the *fatty acids*, the *nucleotides*, and the *amino acids* (Figure 2–6). Although many compounds present in cells do not fit into these categories, these four families of small organic molecules, together with the macromolecules made by linking them into long chains, account for a large fraction of the cell mass.

Amino acids and the proteins that they form will be the subject of Chapter 3. A summary of the structures and properties of the remaining three families—sugars, fatty acids, and nucleotides—is presented in Panels 2–4, 2–5, and 2–6, respectively (see pages 96–101).

## The Chemistry of Cells Is Dominated by Macromolecules with Remarkable Properties

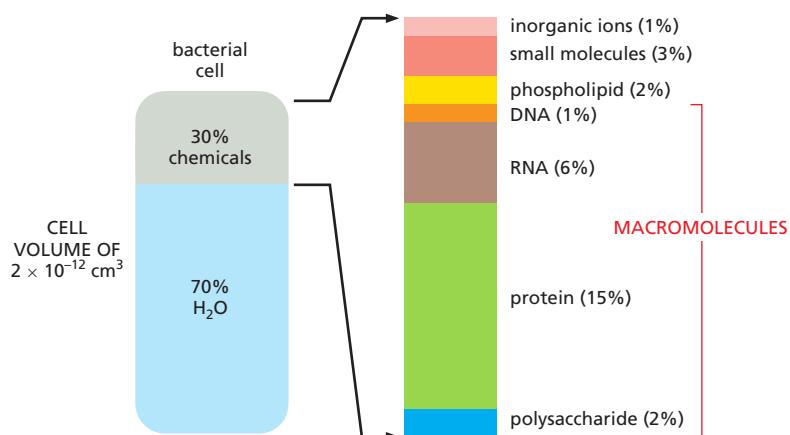
By weight, **macromolecules** are the most abundant carbon-containing molecules in a living cell (Figure 2–7). They are the principal building blocks from which a cell is constructed and also the components that confer the most distinctive properties of living things. The macromolecules in cells are polymers that are constructed by covalently linking small organic molecules (called *monomers*) into



**Figure 2–6** The four main families of small organic molecules in cells. These small molecules form the monomeric building blocks, or subunits, for most of the macromolecules and other assemblies of the cell. Some, such as the sugars and the fatty acids, are also energy sources. Their structures are outlined here and shown in more detail in the Panels at the end of this chapter and in Chapter 3.

long chains (Figure 2–8). They have remarkable properties that could not have been predicted from their simple constituents.

Proteins are abundant and spectacularly versatile, performing thousands of distinct functions in cells. Many proteins serve as *enzymes*, the catalysts that facilitate the many covalent bond-making and bond-breaking reactions that the cell needs. Enzymes catalyze all of the reactions whereby cells extract energy from food molecules, for example, and an enzyme called ribulose bisphosphate carboxylase helps to convert CO<sub>2</sub> to sugars in photosynthetic organisms, producing most of the organic matter needed for life on Earth. Other proteins are used to build structural components, such as tubulin, a protein that self-assembles to make the cell's long microtubules, or histones, proteins that compact the DNA in chromosomes. Yet other proteins act as molecular motors to produce force and



**Figure 2–7** The distribution of molecules in cells. The approximate composition of a bacterial cell is shown by weight. The composition of an animal cell is similar, even though its volume is roughly 1000 times greater. Note that macromolecules dominate. The major inorganic ions include Na<sup>+</sup>, K<sup>+</sup>, Mg<sup>2+</sup>, Ca<sup>2+</sup>, and Cl<sup>-</sup>.

movement, as for myosin in muscle. Proteins perform many other functions, and we shall examine the molecular basis for many of them later in this book.

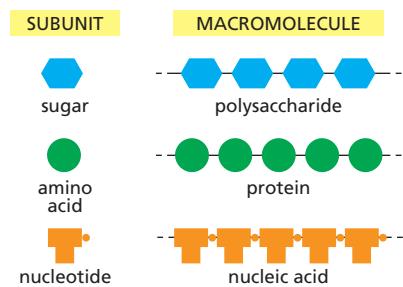
Although the chemical reactions for adding subunits to each polymer are different in detail for proteins, nucleic acids, and polysaccharides, they share important features. Each polymer grows by the addition of a monomer onto the end of a growing chain in a *condensation reaction*, in which one molecule of water is lost with each subunit added (Figure 2–9). The stepwise polymerization of monomers into a long chain is a simple way to manufacture a large, complex molecule, since the subunits are added by the same reaction performed over and over again by the same set of enzymes. Apart from some of the polysaccharides, most macromolecules are made from a limited set of monomers that are slightly different from one another—for example, the 20 different amino acids from which proteins are made. It is critical to life that the polymer chain is not assembled at random from these subunits; instead the subunits are added in a precise order, or *sequence*. The elaborate mechanisms that allow enzymes to accomplish this task are described in detail in Chapters 5 and 6.

### Noncovalent Bonds Specify Both the Precise Shape of a Macromolecule and Its Binding to Other Molecules

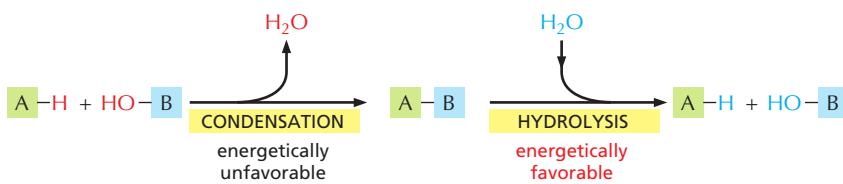
Most of the covalent bonds in a macromolecule allow rotation of the atoms they join, giving the polymer chain great flexibility. In principle, this allows a macromolecule to adopt an almost unlimited number of shapes, or *conformations*, as random thermal energy causes the polymer chain to writhe and rotate. However, the shapes of most biological macromolecules are highly constrained because of the many weak *noncovalent bonds* that form between different parts of the same molecule. If these noncovalent bonds are formed in sufficient numbers, the polymer chain can strongly prefer one particular conformation, determined by the linear sequence of monomers in its chain. Most protein molecules and many of the small RNA molecules found in cells fold tightly into a highly preferred conformation in this way (Figure 2–10).

The four types of noncovalent interactions important in biological molecules were presented earlier, and they are discussed further in Panel 2–3 (pp. 94–95). In addition to folding biological macromolecules into unique shapes, they can also add up to create a strong attraction between two different molecules (see Figure 2–3). This form of molecular interaction provides for great specificity, inasmuch as the close multipoint contacts required for strong binding make it possible for a macromolecule to select out—through binding—just one of the many thousands of other types of molecules present inside a cell. Moreover, because the strength of the binding depends on the number of noncovalent bonds that are formed, interactions of almost any affinity are possible—allowing rapid dissociation where appropriate.

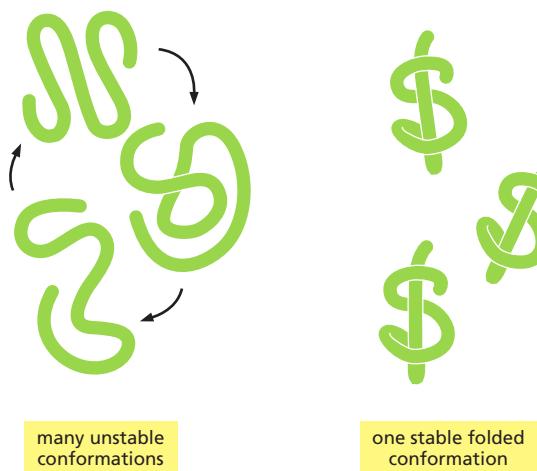
As we discuss next, binding of this type underlies all biological catalysis, making it possible for proteins to function as enzymes. In addition, noncovalent interactions allow macromolecules to be used as building blocks for the formation of



**Figure 2–8** Three families of macromolecules. Each is a polymer formed from small molecules (called monomers) linked together by covalent bonds.



**Figure 2–9** Condensation and hydrolysis as opposite reactions. The macromolecules of the cell are polymers that are formed from subunits (or monomers) by a condensation reaction, and they are broken down by hydrolysis. The condensation reactions are all energetically unfavorable; thus polymer formation requires an energy input, as will be described in the text.



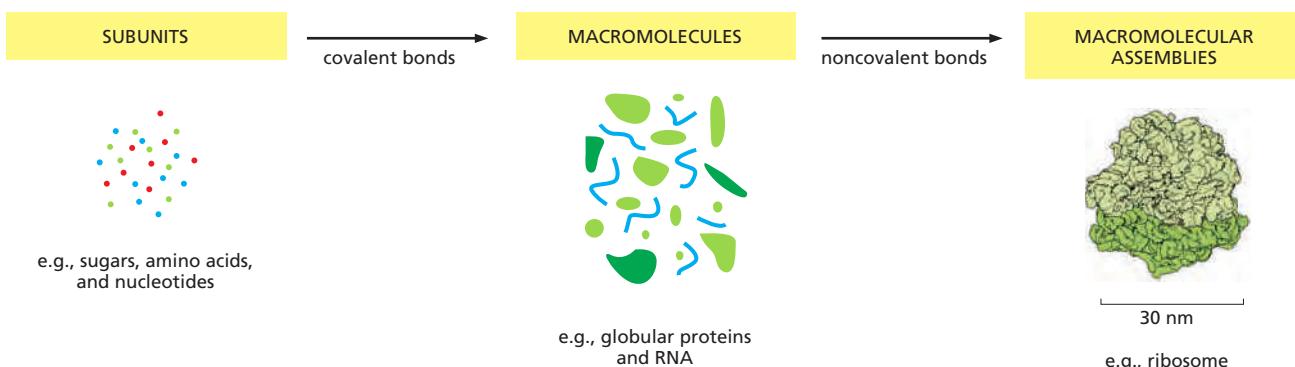
**Figure 2–10** The folding of proteins and RNA molecules into a particularly stable three-dimensional shape, or **conformation**. If the noncovalent bonds maintaining the stable conformation are disrupted, the molecule becomes a flexible chain that loses its biological activity.

larger structures, thereby forming intricate machines with multiple moving parts that perform such complex tasks as DNA replication and protein synthesis (**Figure 2–11**).

## Summary

Living organisms are autonomous, self-propagating chemical systems. They are formed from a distinctive and restricted set of small carbon-based molecules that are essentially the same for every living species. Each of these small molecules is composed of a small set of atoms linked to each other in a precise configuration through covalent bonds. The main categories are sugars, fatty acids, amino acids, and nucleotides. Sugars are a primary source of chemical energy for cells and can be incorporated into polysaccharides for energy storage. Fatty acids are also important for energy storage, but their most critical function is in the formation of cell membranes. Long chains of amino acids form the remarkably diverse and versatile macromolecules known as proteins. Nucleotides play a central part in energy transfer, while also serving as the subunits for the informational macromolecules, RNA and DNA.

Most of the dry mass of a cell consists of macromolecules that have been produced as linear polymers of amino acids (proteins) or nucleotides (DNA and RNA), covalently linked to each other in an exact order. Most of the protein molecules and many of the RNAs fold into a unique conformation that is determined by their sequence of subunits. This folding process creates unique surfaces, and it depends on a large set of weak attractions produced by noncovalent forces between atoms.



**Figure 2–11** Small molecules become covalently linked to form macromolecules, which in turn assemble through noncovalent interactions to form large complexes. Small molecules, proteins, and a ribosome are drawn approximately to scale. Ribosomes are a central part of the machinery that the cell uses to make proteins: each ribosome is formed as a complex of about 90 macromolecules (protein and RNA molecules).

These forces are of four types: electrostatic attractions, hydrogen bonds, van der Waals attractions, and an interaction between nonpolar groups caused by their hydrophobic expulsion from water. The same set of weak forces governs the specific binding of other molecules to macromolecules, making possible the myriad associations between biological molecules that produce the structure and the chemistry of a cell.

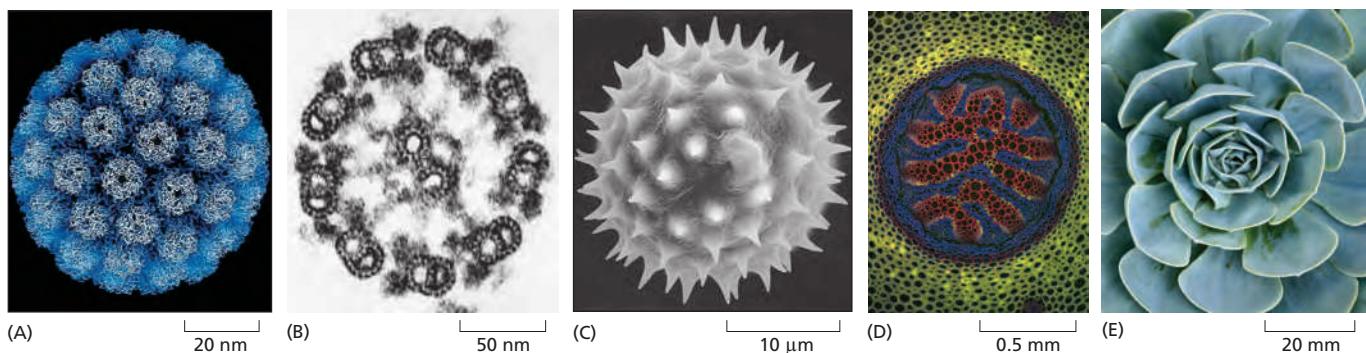
## CATALYSIS AND THE USE OF ENERGY BY CELLS

One property of living things above all makes them seem almost miraculously different from nonliving matter: they create and maintain order, in a universe that is tending always to greater disorder (Figure 2–12). To create this order, the cells in a living organism must perform a never-ending stream of chemical reactions. In some of these reactions, small organic molecules—amino acids, sugars, nucleotides, and lipids—are being taken apart or modified to supply the many other small molecules that the cell requires. In other reactions, small molecules are being used to construct an enormously diverse range of proteins, nucleic acids, and other macromolecules that endow living systems with all of their most distinctive properties. Each cell can be viewed as a tiny chemical factory, performing many millions of reactions every second.

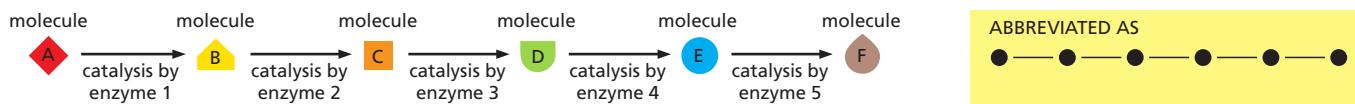
### Cell Metabolism Is Organized by Enzymes

The chemical reactions that a cell carries out would normally occur only at much higher temperatures than those existing inside cells. For this reason, each reaction requires a specific boost in chemical reactivity. This requirement is crucial, because it allows the cell to control its chemistry. The control is exerted through specialized biological *catalysts*. These are almost always proteins called *enzymes*, although RNA catalysts also exist, called *ribozymes*. Each enzyme accelerates, or *catalyzes*, just one of the many possible kinds of reactions that a particular molecule might undergo. Enzyme-catalyzed reactions are connected in series, so that the product of one reaction becomes the starting material, or *substrate*, for the next (Figure 2–13). Long linear reaction pathways are in turn linked to one another, forming a maze of interconnected reactions that enable the cell to survive, grow, and reproduce.

Two opposing streams of chemical reactions occur in cells: (1) the *catabolic* pathways break down foodstuffs into smaller molecules, thereby generating both a useful form of energy for the cell and some of the small molecules that the cell needs as building blocks, and (2) the *anabolic*, or *biosynthetic*, pathways use the



**Figure 2–12 Biological structures are highly ordered.** Well-defined, ornate, and beautiful spatial patterns can be found at every level of organization in living organisms. In order of increasing size: (A) protein molecules in the coat of a virus (a parasite that, although not technically alive, contains the same types of molecules as those found in living cells); (B) the regular array of microtubules seen in a cross section of a sperm tail; (C) surface contours of a pollen grain (a single cell); (D) cross section of a fern stem, showing the patterned arrangement of cells; and (E) a spiral arrangement of leaves in a succulent plant. (A, courtesy of Robert Grant, Stéphane Crainic, and James M. Hogle; B, courtesy of Lewis Tilney; C, courtesy of Colin MacFarlane and Chris Jeffree; D, courtesy of Jim Haseloff.)



**Figure 2-13** How a set of enzyme-catalyzed reactions generates a metabolic pathway. Each enzyme catalyzes a particular chemical reaction, leaving the enzyme unchanged. In this example, a set of enzymes acting in series converts molecule A to molecule F, forming a metabolic pathway. (For a diagram of many of the reactions in a human cell, abbreviated as shown, see Figure 2-63.)

small molecules and the energy harnessed by catabolism to drive the synthesis of the many other molecules that form the cell. Together these two sets of reactions constitute the **metabolism** of the cell (Figure 2-14).

The details of cell metabolism form the traditional subject of *biochemistry* and most of them need not concern us here. But the general principles by which cells obtain energy from their environment and use it to create order are central to cell biology. We begin with a discussion of why a constant input of energy is needed to sustain all living things.

### Biological Order Is Made Possible by the Release of Heat Energy from Cells

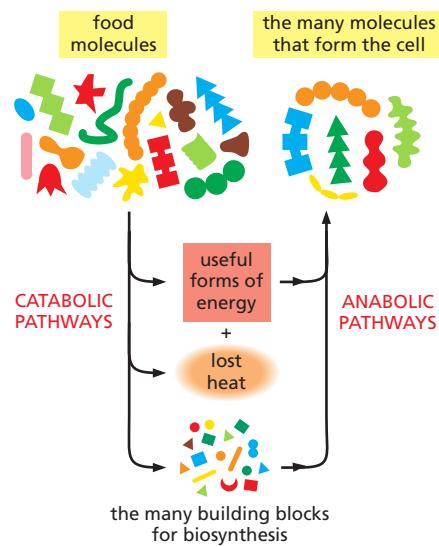
The universal tendency of things to become disordered is a fundamental law of physics—the *second law of thermodynamics*—which states that in the universe, or in any isolated system (a collection of matter that is completely isolated from the rest of the universe), the degree of disorder always increases. This law has such profound implications for life that we will restate it in several ways.

For example, we can present the second law in terms of probability by stating that systems will change spontaneously toward those arrangements that have the greatest probability. If we consider a box of 100 coins all lying heads up, a series of accidents that disturbs the box will tend to move the arrangement toward a mixture of 50 heads and 50 tails. The reason is simple: there is a huge number of possible arrangements of the individual coins in the mixture that can achieve the 50–50 result, but only one possible arrangement that keeps all of the coins oriented heads up. Because the 50–50 mixture is therefore the most probable, we say that it is more “disordered.” For the same reason, it is a common experience that one’s living space will become increasingly disordered without intentional effort: the movement toward disorder is a *spontaneous process*, requiring a periodic effort to reverse it (Figure 2-15).

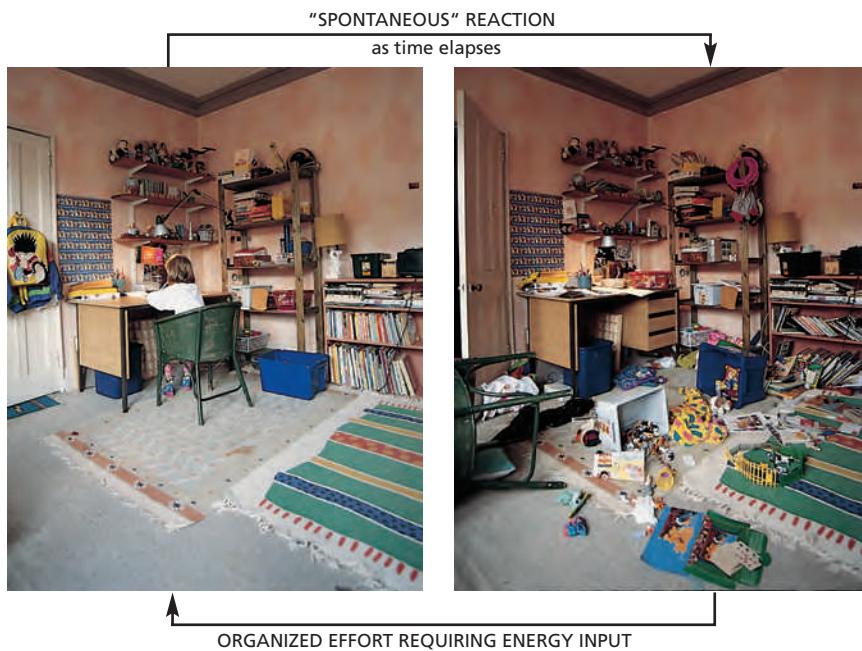
The amount of disorder in a system can be quantified and expressed as the **entropy** of the system: the greater the disorder, the greater the entropy. Thus, another way to express the second law of thermodynamics is to say that systems will change spontaneously toward arrangements with greater entropy.

Living cells—by surviving, growing, and forming complex organisms—are generating order and thus might appear to defy the second law of thermodynamics. How is this possible? The answer is that a cell is not an isolated system: it takes in energy from its environment in the form of food, or as photons from the sun (or even, as in some chemosynthetic bacteria, from inorganic molecules alone). It then uses this energy to generate order within itself. In the course of the chemical reactions that generate order, the cell converts part of the energy it uses into heat. The heat is discharged into the cell’s environment and disorders the surroundings. As a result, the total entropy—that of the cell plus its surroundings—increases, as demanded by the second law of thermodynamics.

To understand the principles governing these energy conversions, think of a cell surrounded by a sea of matter representing the rest of the universe. As the cell lives and grows, it creates internal order. But it constantly releases heat energy as it synthesizes molecules and assembles them into cell structures. Heat is energy in its most disordered form—the random jostling of molecules. When



**Figure 2-14** Schematic representation of the relationship between catabolic and anabolic pathways in metabolism. As suggested in this diagram, a major portion of the energy stored in the chemical bonds of food molecules is dissipated as heat. In addition, the mass of food required by any organism that derives all of its energy from catabolism is much greater than the mass of the molecules that it can produce by anabolism.

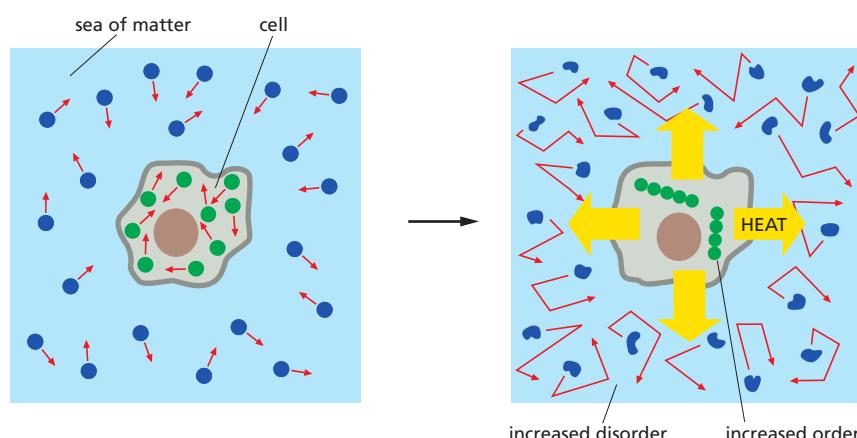


**Figure 2–15** An everyday illustration of the spontaneous drive toward disorder. Reversing this tendency toward disorder requires an intentional effort and an input of energy: it is not spontaneous. In fact, from the second law of thermodynamics, we can be certain that the human intervention required will release enough heat to the environment to more than compensate for the reordering of the items in this room.

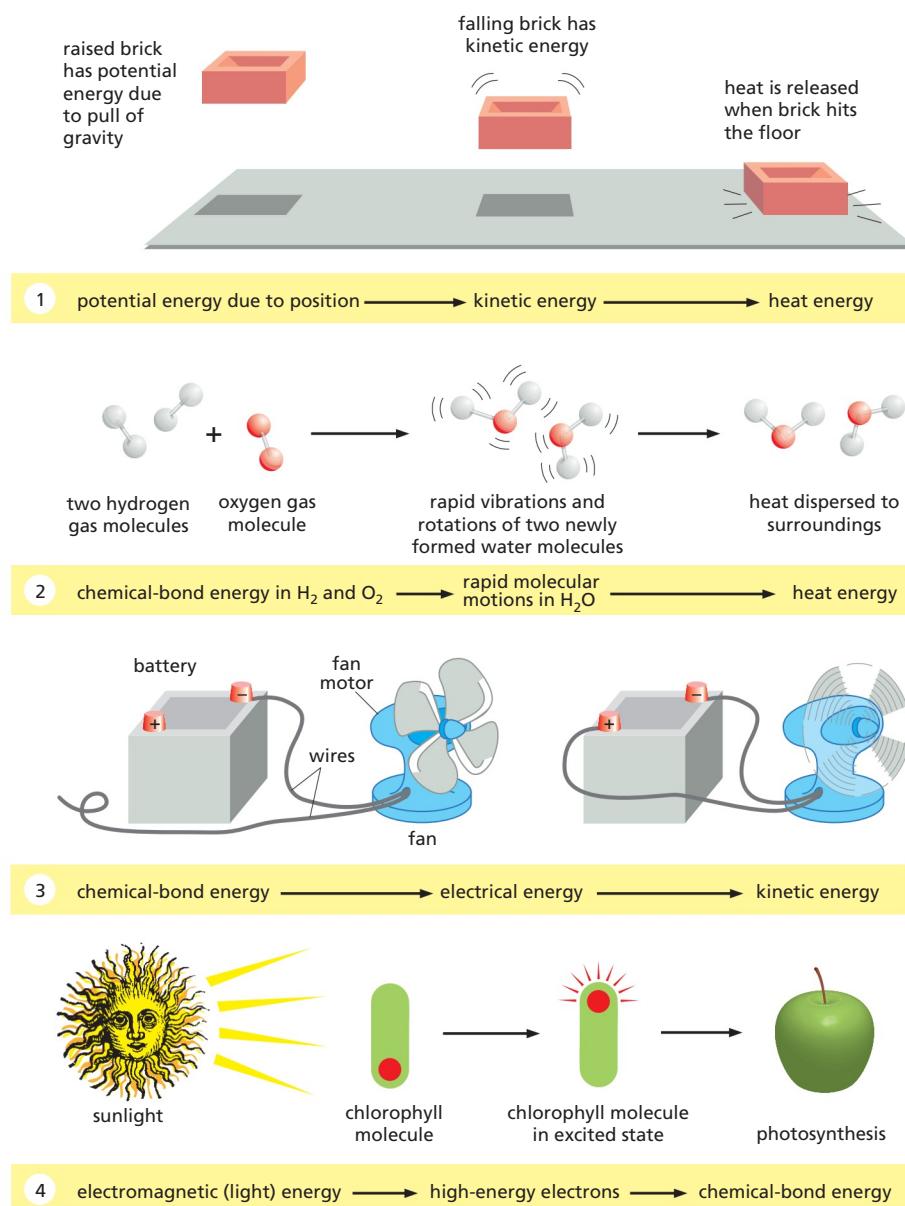
the cell releases heat to the sea, it increases the intensity of molecular motions there (thermal motion)—thereby increasing the randomness, or disorder, of the sea. The second law of thermodynamics is satisfied because the increase in the amount of order inside the cell is always more than compensated for by an even greater decrease in order (increase in entropy) in the surrounding sea of matter (**Figure 2–16**).

Where does the heat that the cell releases come from? Here we encounter another important law of thermodynamics. The *first law of thermodynamics* states that energy can be converted from one form to another, but that it cannot be created or destroyed. **Figure 2–17** illustrates some interconversions between different forms of energy. The amount of energy in different forms will change as a result of the chemical reactions inside the cell, but the first law tells us that the total amount of energy must always be the same. For example, an animal cell takes in foodstuffs and converts some of the energy present in the chemical bonds between the atoms of these food molecules (chemical-bond energy) into the random thermal motion of molecules (heat energy).

The cell cannot derive any benefit from the heat energy it releases unless the heat-generating reactions inside the cell are directly linked to the processes that generate molecular order. It is the tight *coupling* of heat production to an increase



**Figure 2–16** A simple thermodynamic analysis of a living cell. In the diagram on the left, the molecules of both the cell and the rest of the universe (the sea of matter) are depicted in a relatively disordered state. In the diagram on the right, the cell has taken in energy from food molecules and released heat through reactions that order the molecules the cell contains. The heat released increases the disorder in the environment around the cell (depicted by jagged arrows and distorted molecules, indicating increased molecular motions caused by heat). As a result, the second law of thermodynamics—which states that the amount of disorder in the universe must always increase—is satisfied as the cell grows and divides. For a detailed discussion, see Panel 2–7 (pp. 102–103).

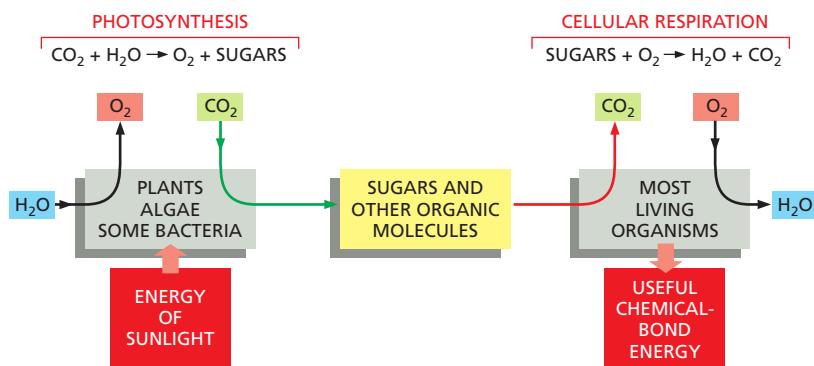


in order that distinguishes the metabolism of a cell from the wasteful burning of fuel in a fire. Later, we illustrate how this coupling occurs. For now, it is sufficient to recognize that a direct linkage of the “controlled burning” of food molecules to the generation of biological order is required for cells to create and maintain an island of order in a universe tending toward chaos.

### Cells Obtain Energy by the Oxidation of Organic Molecules

All animal and plant cells are powered by energy stored in the chemical bonds of organic molecules, whether they are sugars that a plant has photosynthesized as food for itself or the mixture of large and small molecules that an animal has eaten. Organisms must extract this energy in usable form to live, grow, and reproduce. In both plants and animals, energy is extracted from food molecules by a process of gradual oxidation, or controlled burning.

The Earth's atmosphere contains a great deal of oxygen, and in the presence of oxygen the most energetically stable form of carbon is CO<sub>2</sub> and that of hydrogen



**Figure 2–18** Photosynthesis and respiration as complementary processes in the living world. Photosynthesis converts the electromagnetic energy in sunlight into chemical-bond energy in sugars and other organic molecules. Plants, algae, and cyanobacteria obtain the carbon atoms that they need for this purpose from atmospheric CO<sub>2</sub> and the hydrogen from water, releasing O<sub>2</sub> gas as a by-product. The organic molecules produced by photosynthesis in turn serve as food for other organisms. Many of these organisms carry out aerobic respiration, a process that uses O<sub>2</sub> to form CO<sub>2</sub> from the same carbon atoms that had been taken up as CO<sub>2</sub> and converted into sugars by photosynthesis. In the process, the organisms that respire obtain the chemical-bond energy that they need to survive.

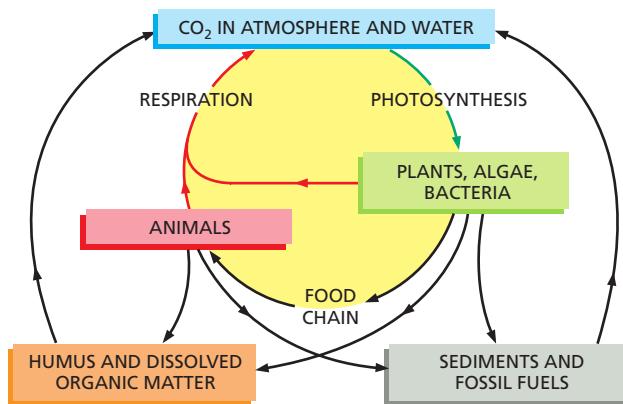
The first cells on the Earth are thought to have been capable of neither photosynthesis nor respiration (discussed in Chapter 14). However, photosynthesis must have preceded respiration on the Earth, since there is strong evidence that billions of years of photosynthesis were required before O<sub>2</sub> had been released in sufficient quantity to create an atmosphere rich in this gas. (The Earth's atmosphere currently contains 20% O<sub>2</sub>.)

is H<sub>2</sub>O. A cell is therefore able to obtain energy from sugars or other organic molecules by allowing their carbon and hydrogen atoms to combine with oxygen to produce CO<sub>2</sub> and H<sub>2</sub>O, respectively—a process called **aerobic respiration**.

Photosynthesis (discussed in detail in Chapter 14) and respiration are complementary processes (Figure 2–18). This means that the transactions between plants and animals are not all one way. Plants, animals, and microorganisms have existed together on this planet for so long that many of them have become an essential part of the others' environments. The oxygen released by photosynthesis is consumed in the combustion of organic molecules during aerobic respiration. And some of the CO<sub>2</sub> molecules that are fixed today into organic molecules by photosynthesis in a green leaf were yesterday released into the atmosphere by the respiration of an animal—or by the respiration of a fungus or bacterium decomposing dead organic matter. We therefore see that carbon utilization forms a huge cycle that involves the *biosphere* (all of the living organisms on Earth) as a whole (Figure 2–19). Similarly, atoms of nitrogen, phosphorus, and sulfur move between the living and nonliving worlds in cycles that involve plants, animals, fungi, and bacteria.

### Oxidation and Reduction Involve Electron Transfers

The cell does not oxidize organic molecules in one step, as occurs when organic material is burned in a fire. Through the use of enzyme catalysts, metabolism takes these molecules through a large number of reactions that only rarely involve the direct addition of oxygen. Before we consider some of these reactions and their purpose, we discuss what is meant by the process of oxidation.



**Figure 2–19** The carbon cycle. Individual carbon atoms are incorporated into organic molecules of the living world by the photosynthetic activity of bacteria, algae, and plants. They pass to animals, microorganisms, and organic material in soil and oceans in cyclic paths. CO<sub>2</sub> is restored to the atmosphere when organic molecules are oxidized by cells or burned by humans as fuels.

**Oxidation** refers to more than the addition of oxygen atoms; the term applies more generally to any reaction in which electrons are transferred from one atom to another. Oxidation in this sense refers to the removal of electrons, and **reduction**—the converse of oxidation—means the addition of electrons. Thus,  $\text{Fe}^{2+}$  is oxidized if it loses an electron to become  $\text{Fe}^{3+}$ , and a chlorine atom is reduced if it gains an electron to become  $\text{Cl}^-$ . Since the number of electrons is conserved (no loss or gain) in a chemical reaction, oxidation and reduction always occur simultaneously: that is, if one molecule gains an electron in a reaction (reduction), a second molecule loses the electron (oxidation). When a sugar molecule is oxidized to  $\text{CO}_2$  and  $\text{H}_2\text{O}$ , for example, the  $\text{O}_2$  molecules involved in forming  $\text{H}_2\text{O}$  gain electrons and thus are said to have been reduced.

The terms “oxidation” and “reduction” apply even when there is only a partial shift of electrons between atoms linked by a covalent bond (Figure 2–20). When a carbon atom becomes covalently bonded to an atom with a strong affinity for electrons, such as oxygen, chlorine, or sulfur, for example, it gives up more than its equal share of electrons and forms a *polar* covalent bond. Because the positive charge of the carbon nucleus is now somewhat greater than the negative charge of its electrons, the atom acquires a partial positive charge and is said to be oxidized. Conversely, a carbon atom in a C-H linkage has slightly more than its share of electrons, and so it is said to be reduced.

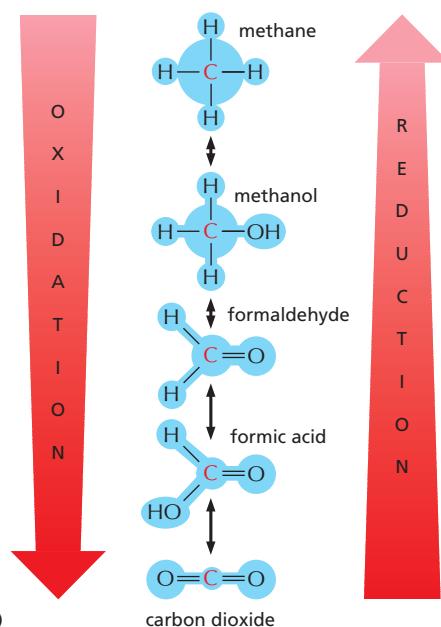
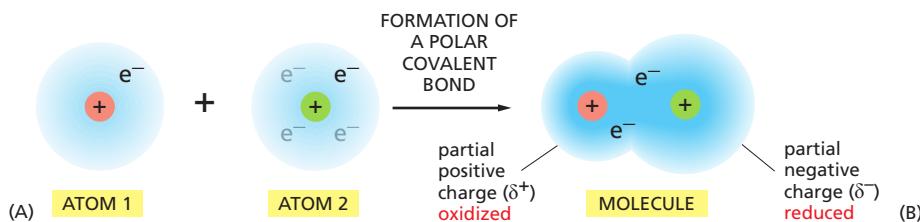
When a molecule in a cell picks up an electron ( $e^-$ ), it often picks up a proton ( $\text{H}^+$ ) at the same time (protons being freely available in water). The net effect in this case is to add a hydrogen atom to the molecule.



Even though a proton plus an electron is involved (instead of just an electron), such *hydrogenation* reactions are reductions, and the reverse, *dehydrogenation* reactions are oxidations. It is especially easy to tell whether an organic molecule is being oxidized or reduced: reduction is occurring if its number of C-H bonds increases, whereas oxidation is occurring if its number of C-H bonds decreases (see Figure 2–20B).

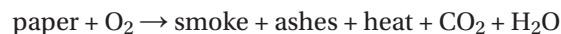
Cells use enzymes to catalyze the oxidation of organic molecules in small steps, through a sequence of reactions that allows useful energy to be harvested. We now need to explain how enzymes work and some of the constraints under which they operate.

**Figure 2–20 Oxidation and reduction.** (A) When two atoms form a *polar* covalent bond, the atom ending up with a greater share of electrons is said to be reduced, while the other atom acquires a lesser share of electrons and is said to be oxidized. The reduced atom has acquired a partial negative charge ( $\delta^-$ ) as the positive charge on the atomic nucleus is now more than equaled by the total charge of the electrons surrounding it, and conversely, the oxidized atom has acquired a partial positive charge ( $\delta^+$ ). (B) The single carbon atom of methane can be converted to that of carbon dioxide by the successive replacement of its covalently bonded hydrogen atoms with oxygen atoms. With each step, electrons are shifted away from the carbon (as indicated by the blue shading), and the carbon atom becomes progressively more oxidized. Each of these steps is energetically favorable under the conditions present inside a cell.



## Enzymes Lower the Activation-Energy Barriers That Block Chemical Reactions

Consider the reaction

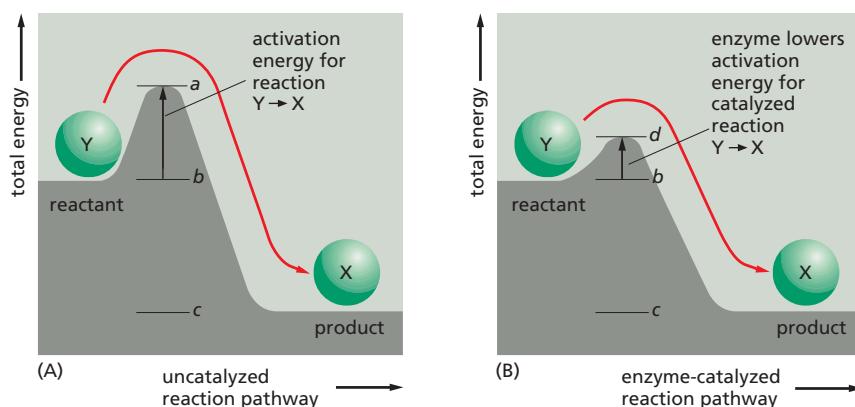


Once ignited, the paper burns readily, releasing to the atmosphere both energy as heat and water and carbon dioxide as gases. The reaction is irreversible, since the smoke and ashes never spontaneously retrieve these entities from the heated atmosphere and reconstitute themselves into paper. When the paper burns, its chemical energy is dissipated as heat—not lost from the universe, since energy can never be created or destroyed, but irretrievably dispersed in the chaotic random thermal motions of molecules. At the same time, the atoms and molecules of the paper become dispersed and disordered. In the language of thermodynamics, there has been a loss of *free energy*; that is, of energy that can be harnessed to do work or drive chemical reactions. This loss reflects a reduction of orderliness in the way the energy and molecules were stored in the paper.

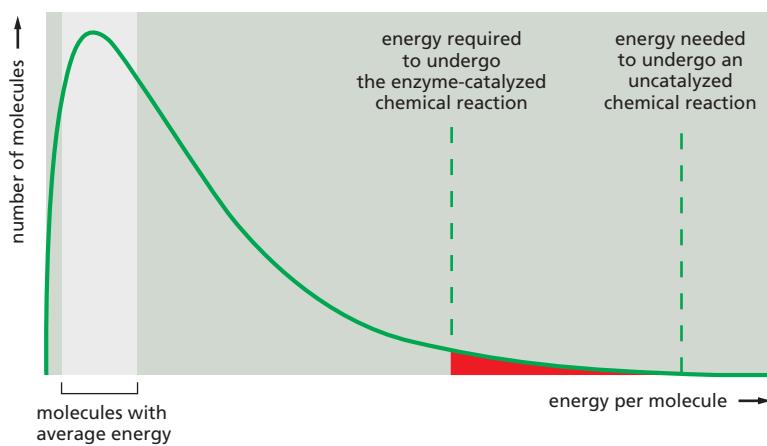
We shall discuss free energy in more detail shortly, but the general principle is clear enough intuitively: chemical reactions proceed spontaneously only in the direction that leads to a loss of free energy. In other words, the spontaneous direction for any reaction is the direction that goes “downhill,” where a “downhill” reaction is one that is *energetically favorable*.

Although the most energetically favorable form of carbon under ordinary conditions is CO<sub>2</sub>, and that of hydrogen is H<sub>2</sub>O, a living organism does not disappear in a puff of smoke, and the paper book in your hands does not burst into flames. This is because the molecules both in the living organism and in the book are in a relatively stable state, and they cannot be changed to a state of lower energy without an input of energy: in other words, a molecule requires **activation energy**—a kick over an energy barrier—before it can undergo a chemical reaction that leaves it in a more stable state (Figure 2–21). In the case of a burning book, the activation energy can be provided by the heat of a lighted match. For the molecules in the watery solution inside a cell, the kick is delivered by an unusually energetic random collision with surrounding molecules—collisions that become more violent as the temperature is raised.

The chemistry in a living cell is tightly controlled, because the kick over energy barriers is greatly aided by a specialized class of proteins—the **enzymes**. Each enzyme binds tightly to one or more molecules, called **substrates**, and holds them in a way that greatly reduces the activation energy of a particular chemical reaction that the bound substrates can undergo. A substance that can lower the activation energy of a reaction is termed a **catalyst**; catalysts increase the rate of chemical reactions because they allow a much larger proportion of the random collisions with surrounding molecules to kick the substrates over the energy barrier, as illustrated in Figure 2–22. Enzymes are among the most effective catalysts



**Figure 2-21** The important principle of activation energy. (A) Compound Y (a reactant) is in a relatively stable state, and energy is required to convert it to compound X (a product), even though X is at a lower overall energy level than Y. This conversion will not take place, therefore, unless compound Y can acquire enough activation energy (*energy a minus energy b*) from its surroundings to undergo the reaction that converts it into compound X. This energy may be provided by means of an unusually energetic collision with other molecules. For the reverse reaction, X → Y, the activation energy will be much larger (*energy a minus energy c*); this reaction will therefore occur much more rarely. Activation energies are always positive; note, however, that the total energy change for the energetically favorable reaction Y → X is *energy c minus energy b*, a negative number. (B) Energy barriers for specific reactions can be lowered by catalysts, as indicated by the line marked d. Enzymes are particularly effective catalysts because they greatly reduce the activation energy for the reactions they perform.



**Figure 2-22 Lowering the activation energy greatly increases the probability of a reaction.** At any given instant, a population of identical substrate molecules will have a range of energies, distributed as shown on the graph. The varying energies come from collisions with surrounding molecules, which make the substrate molecules jingle, vibrate, and spin. For a molecule to undergo a chemical reaction, the energy of the molecule must exceed the activation-energy barrier for that reaction (dashed lines). For most biological reactions, this almost never happens without enzyme catalysis. Even with enzyme catalysis, the substrate molecules must experience a particularly energetic collision to react (red shaded area). Raising the temperature will also increase the number of molecules with sufficient energy to overcome the activation energy needed for a reaction; but in marked contrast to enzyme catalysis, this effect is nonselective, speeding up all reactions ([Movie 2.2](#)).

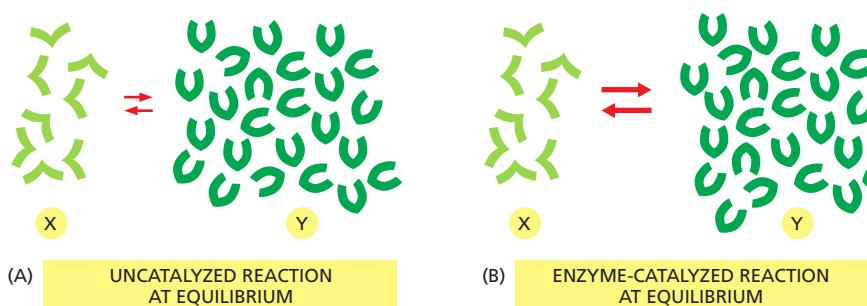
known: some are capable of speeding up reactions by factors of  $10^{14}$  or more. Enzymes thereby allow reactions that would not otherwise occur to proceed rapidly at normal temperatures.

### Enzymes Can Drive Substrate Molecules Along Specific Reaction Pathways

An enzyme cannot change the equilibrium point for a reaction. The reason is simple: when an enzyme (or any catalyst) lowers the activation energy for the reaction  $Y \rightarrow X$ , of necessity it also lowers the activation energy for the reaction  $X \rightarrow Y$  by exactly the same amount (see Figure 2-21). The forward and backward reactions will therefore be accelerated by the same factor by an enzyme, and the equilibrium point for the reaction will be unchanged (Figure 2-23). Thus no matter how much an enzyme speeds up a reaction, it cannot change its direction.

Despite the above limitation, enzymes steer all of the reactions in cells through specific reaction paths. This is because enzymes are both highly selective and very precise, usually catalyzing only one particular reaction. In other words, each enzyme selectively lowers the activation energy of only one of the several possible chemical reactions that its bound substrate molecules could undergo. In this way, sets of enzymes can direct each of the many different molecules in a cell along a particular reaction pathway (Figure 2-24).

The success of living organisms is attributable to a cell's ability to make enzymes of many types, each with precisely specified properties. Each enzyme



**Figure 2-23 Enzymes cannot change the equilibrium point for reactions.** Enzymes, like all catalysts, speed up the forward and backward rates of a reaction by the same factor. Therefore, for both the catalyzed and the uncatalyzed reactions shown here, the number of molecules undergoing the transition  $X \rightarrow Y$  is equal to the number of molecules undergoing the transition  $Y \rightarrow X$  when the ratio of  $Y$  molecules to  $X$  molecules is 3 to 1. In other words, the two reactions reach equilibrium at exactly the same point.

**Figure 2-24** Directing substrate molecules through a specific reaction pathway by enzyme catalysis. A substrate molecule in a cell (green ball) is converted into a different molecule (blue ball) by means of a series of enzyme-catalyzed reactions. As indicated (yellow box), several reactions are energetically favorable at each step, but only one is catalyzed by each enzyme. Sets of enzymes thereby determine the exact reaction pathway that is followed by each molecule inside the cell.

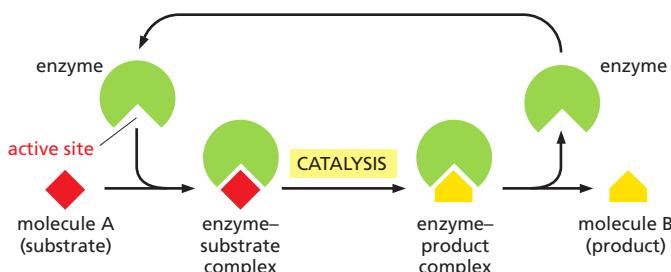
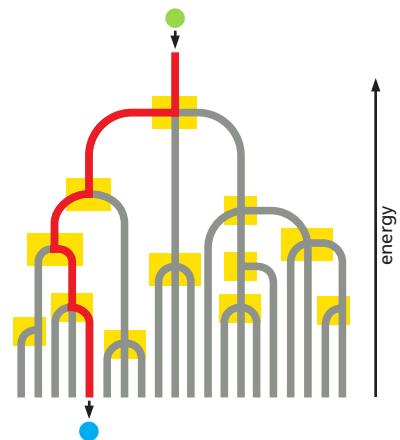
has a unique shape containing an *active site*, a pocket or groove in the enzyme into which only particular substrates will fit (Figure 2-25). Like all other catalysts, enzyme molecules themselves remain unchanged after participating in a reaction and therefore can function over and over again. In Chapter 3, we discuss further how enzymes work.

### How Enzymes Find Their Substrates: The Enormous Rapidity of Molecular Motions

An enzyme will often catalyze the reaction of thousands of substrate molecules every second. This means that it must be able to bind a new substrate molecule in a fraction of a millisecond. But both enzymes and their substrates are present in relatively small numbers in a cell. How do they find each other so fast? Rapid binding is possible because the motions caused by heat energy are enormously fast at the molecular level. These molecular motions can be classified broadly into three kinds: (1) the movement of a molecule from one place to another (*translational motion*), (2) the rapid back-and-forth movement of covalently linked atoms with respect to one another (vibrations), and (3) rotations. All of these motions help to bring the surfaces of interacting molecules together.

The rates of molecular motions can be measured by a variety of spectroscopic techniques. A large globular protein is constantly tumbling, rotating about its axis about a million times per second. Molecules are also in constant translational motion, which causes them to explore the space inside the cell very efficiently by wandering through it—a process called **diffusion**. In this way, every molecule in a cell collides with a huge number of other molecules each second. As the molecules in a liquid collide and bounce off one another, an individual molecule moves first one way and then another, its path constituting a *random walk* (Figure 2-26). In such a walk, the average net distance that each molecule travels (as the “crow flies”) from its starting point is proportional to the square root of the time involved: that is, if it takes a molecule 1 second on average to travel 1  $\mu\text{m}$ , it takes 4 seconds to travel 2  $\mu\text{m}$ , 100 seconds to travel 10  $\mu\text{m}$ , and so on.

The inside of a cell is very crowded (Figure 2-27). Nevertheless, experiments in which fluorescent dyes and other labeled molecules are injected into cells show that small organic molecules diffuse through the watery gel of the cytosol nearly



**Figure 2-25** How enzymes work. Each enzyme has an active site to which one or more *substrate* molecules bind, forming an *enzyme–substrate complex*. A reaction occurs at the active site, producing an *enzyme–product complex*. The *product* is then released, allowing the enzyme to bind further substrate molecules.

as rapidly as they do through water. A small organic molecule, for example, takes only about one-fifth of a second on average to diffuse a distance of 10  $\mu\text{m}$ . Diffusion is therefore an efficient way for small molecules to move the limited distances in the cell (a typical animal cell is 15  $\mu\text{m}$  in diameter).

Since enzymes move more slowly than substrates in cells, we can think of them as sitting still. The rate of encounter of each enzyme molecule with its substrate will depend on the concentration of the substrate molecule. For example, some abundant substrates are present at a concentration of 0.5 mM. Since pure water is 55.5 M, there is only about one such substrate molecule in the cell for every  $10^5$  water molecules. Nevertheless, the active site on an enzyme molecule that binds this substrate will be bombarded by about 500,000 random collisions with the substrate molecule per second. (For a substrate concentration tenfold lower, the number of collisions drops to 50,000 per second, and so on.) A random collision between the active site of an enzyme and the matching surface of its substrate molecule often leads immediately to the formation of an enzyme–substrate complex. A reaction in which a covalent bond is broken or formed can now occur extremely rapidly. When one appreciates how quickly molecules move and react, the observed rates of enzymatic catalysis do not seem so amazing.

Two molecules that are held together by noncovalent bonds can also dissociate. The multiple weak noncovalent bonds that they form with each other will persist until random thermal motion causes the two molecules to separate. In general, the stronger the binding of the enzyme and substrate, the slower their rate of dissociation. In contrast, whenever two colliding molecules have poorly matching surfaces, they form few noncovalent bonds and the total energy of association will be negligible compared with that of thermal motion. In this case, the two molecules dissociate as rapidly as they come together, preventing incorrect and unwanted associations between mismatched molecules, such as between an enzyme and the wrong substrate.

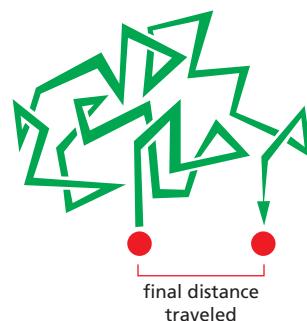
### The Free-Energy Change for a Reaction, $\Delta G$ , Determines Whether It Can Occur Spontaneously

Although enzymes speed up reactions, they cannot by themselves force energetically unfavorable reactions to occur. In terms of a water analogy, enzymes by themselves cannot make water run uphill. Cells, however, must do just that in order to grow and divide: they must build highly ordered and energy-rich molecules from small and simple ones. We shall see that this is done through enzymes that directly *couple* energetically favorable reactions, which release energy and produce heat, to energetically unfavorable reactions, which produce biological order.

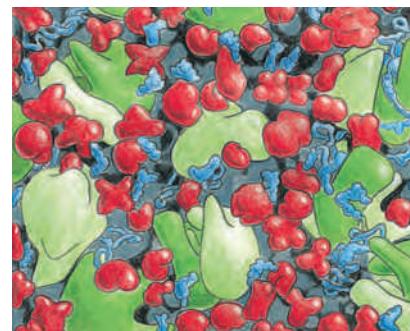
What do cell biologists mean by the term “energetically favorable,” and how can this be quantified? According to the second law of thermodynamics the universe tends toward maximum disorder (largest *entropy* or greatest probability). Thus, a chemical reaction can proceed spontaneously only if it results in a net increase in the disorder of the universe (see Figure 2–16). This disorder of the universe can be expressed most conveniently in terms of the *free energy* of a system, a concept we touched on earlier.

**Free energy,  $G$ ,** is an expression of the *energy available to do work*—for example, the work of driving chemical reactions. The value of  $G$  is of interest only when a system undergoes a *change*, denoted  $\Delta G$  (delta  $G$ ). The change in  $G$  is critical because, as explained in **Panel 2–7** (pp. 102–103),  $\Delta G$  is a direct measure of the

**Figure 2–27** The structure of the cytoplasm. The drawing is approximately to scale and emphasizes the crowding in the cytoplasm. Only the macromolecules are shown: RNAs are shown in blue, ribosomes in green, and proteins in red. Enzymes and other macromolecules diffuse relatively slowly in the cytoplasm, in part because they interact with many other macromolecules; small molecules, by contrast, diffuse nearly as rapidly as they do in water (**Movie 2.4**). (Adapted from D.S. Goodsell, *Trends Biochem. Sci.* 16:203–206, 1991. With permission from Elsevier.)



**Figure 2–26** A random walk. Molecules in solution move in a random fashion as a result of the continual buffeting they receive in collisions with other molecules. This movement allows small molecules to diffuse rapidly from one part of the cell to another, as described in the text (**Movie 2.3**).



100 nm

amount of disorder created in the universe when a reaction takes place. *Energetically favorable reactions*, by definition, are those that decrease free energy; in other words, they have a *negative  $\Delta G$*  and disorder the universe (Figure 2–28).

An example of an energetically favorable reaction on a macroscopic scale is the “reaction” by which a compressed spring relaxes to an expanded state, releasing its stored elastic energy as heat to its surroundings; an example on a microscopic scale is salt dissolving in water. Conversely, *energetically unfavorable reactions* with a *positive  $\Delta G$* —such as the joining of two amino acids to form a peptide bond—by themselves create order in the universe. Therefore, these reactions can take place only if they are coupled to a second reaction with a *negative  $\Delta G$*  so large that the  $\Delta G$  of the overall process is negative (Figure 2–29).

### The Concentration of Reactants Influences the Free-Energy Change and a Reaction’s Direction

As we have just described, a reaction  $Y \leftrightarrow X$  will go in the direction  $Y \rightarrow X$  when the associated free-energy change,  $\Delta G$ , is negative, just as a tensed spring left to itself will relax and lose its stored energy to its surroundings as heat. For a chemical reaction, however,  $\Delta G$  depends not only on the energy stored in each individual molecule, but also on the concentrations of the molecules in the reaction mixture. Remember that  $\Delta G$  reflects the degree to which a reaction creates a more disordered—in other words, a more probable—state of the universe. Recalling our coin analogy, it is very likely that a coin will flip from a head to a tail orientation if a jiggling box contains 90 heads and 10 tails, but this is a less probable event if the box has 10 heads and 90 tails.

The same is true for a chemical reaction. For a reversible reaction  $Y \leftrightarrow X$ , a large excess of  $Y$  over  $X$  will tend to drive the reaction in the direction  $Y \rightarrow X$ . Therefore, as the ratio of  $Y$  to  $X$  increases, the  $\Delta G$  becomes more negative for the transition  $Y \rightarrow X$  (and more positive for the transition  $X \rightarrow Y$ ).

The amount of concentration difference that is needed to compensate for a given decrease in chemical-bond energy (and accompanying heat release) is not intuitively obvious. In the late nineteenth century, the relationship was determined through a thermodynamic analysis that makes it possible to separate the concentration-dependent and the concentration-independent parts of the free-energy change, as we describe next.

### The Standard Free-Energy Change, $\Delta G^\circ$ , Makes It Possible to Compare the Energetics of Different Reactions

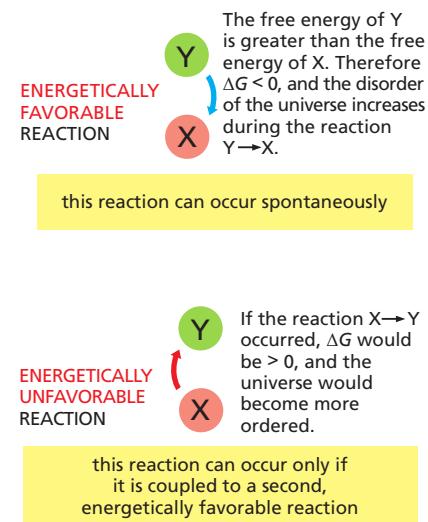
Because  $\Delta G$  depends on the concentrations of the molecules in the reaction mixture at any given time, it is not a particularly useful value for comparing the relative energies of different types of reactions. To place reactions on a comparable basis, we need to turn to the **standard free-energy change** of a reaction,  $\Delta G^\circ$ . The  $\Delta G^\circ$  is the change in free energy under a standard condition, defined as that where the concentrations of all the reactants are set to the same fixed value of 1 mole/liter. Defined in this way,  $\Delta G^\circ$  depends only on the intrinsic characters of the reacting molecules.

For the simple reaction  $Y \rightarrow X$  at 37°C,  $\Delta G^\circ$  is related to  $\Delta G$  as follows:

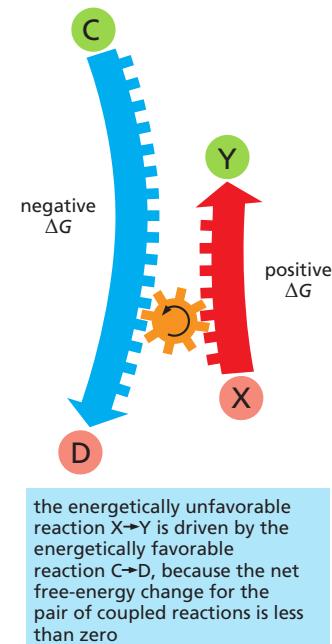
$$\Delta G = \Delta G^\circ + RT \ln \frac{[X]}{[Y]}$$

where  $\Delta G$  is in kilojoules per mole,  $[Y]$  and  $[X]$  denote the concentrations of  $Y$  and  $X$  in moles/liter,  $\ln$  is the *natural logarithm*, and  $RT$  is the product of the gas constant,  $R$ , and the absolute temperature,  $T$ . At 37°C,  $RT = 2.58 \text{ J mole}^{-1}$ . (A mole is  $6 \times 10^{23}$  molecules of a substance.)

A large body of thermodynamic data has been collected that has made it possible to determine the standard free-energy change,  $\Delta G^\circ$ , for the important metabolic reactions of a cell. Given these  $\Delta G^\circ$  values, combined with additional information about metabolite concentrations and reaction pathways, it is possible to quantitatively predict the course of most biological reactions.

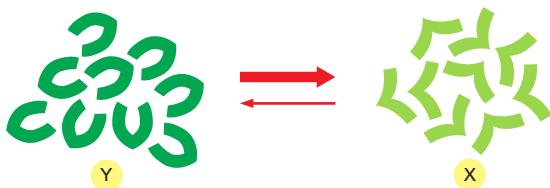


**Figure 2–28** The distinction between energetically favorable and energetically unfavorable reactions.



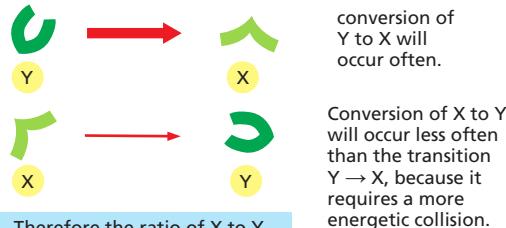
**Figure 2–29** How reaction coupling is used to drive energetically unfavorable reactions.

FOR THE ENERGETICALLY FAVORABLE REACTION  $Y \rightarrow X$ ,



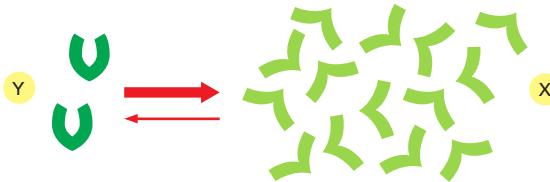
when  $X$  and  $Y$  are at equal concentrations,  $[Y] = [X]$ , the formation of  $X$  is energetically favored. In other words, the  $\Delta G$  of  $Y \rightarrow X$  is negative and the  $\Delta G$  of  $X \rightarrow Y$  is positive. But because of thermal bombardments, there will always be some  $X$  converting to  $Y$ .

THUS, FOR EACH INDIVIDUAL MOLECULE,



Therefore the ratio of  $X$  to  $Y$  molecules will increase with time

EVENTUALLY, there will be a large enough excess of  $X$  over  $Y$  to just compensate for the slow rate of  $X \rightarrow Y$ , such that the number of  $Y$  molecules being converted to  $X$  molecules each second is exactly equal to the number of  $X$  molecules being converted to  $Y$  molecules each second. At this point, the reaction will be at equilibrium.



AT EQUILIBRIUM, there is no net change in the ratio of  $Y$  to  $X$ , and the  $\Delta G$  for both forward and backward reactions is zero.

### The Equilibrium Constant and $\Delta G^\circ$ Are Readily Derived from Each Other

Inspection of the above equation reveals that the  $\Delta G$  equals the value of  $\Delta G^\circ$  when the concentrations of  $Y$  and  $X$  are equal. But as any favorable reaction proceeds, the concentrations of the products will increase as the concentration of the substrates decreases. This change in relative concentrations will cause  $[X]/[Y]$  to become increasingly large, making the initially favorable  $\Delta G$  less and less negative (the logarithm of a number  $x$  is positive for  $x > 1$ , negative for  $x < 1$ , and zero for  $x = 1$ ). Eventually, when  $\Delta G = 0$ , a chemical **equilibrium** will be attained; here there is no net change in free energy to drive the reaction in either direction, inasmuch as the concentration effect just balances the push given to the reaction by  $\Delta G^\circ$ . As a result, the ratio of product to substrate reaches a constant value at chemical equilibrium (**Figure 2–30**).

We can define the **equilibrium constant**,  $K$ , for the reaction  $Y \rightarrow X$  as

$$K = \frac{[X]}{[Y]}$$

where  $[X]$  is the concentration of the product and  $[Y]$  is the concentration of the reactant at equilibrium. Remembering that  $\Delta G = \Delta G^\circ + RT \ln [X]/[Y]$ , and that  $\Delta G = 0$  at equilibrium, we see that

$$\Delta G^\circ = -RT \ln \frac{[X]}{[Y]} = -RT \ln K$$

At 37°C, where  $RT = 2.58$ , the equilibrium equation is therefore:

$$\Delta G^\circ = -2.58 \ln K$$

**Figure 2–30** Chemical equilibrium. When a reaction reaches equilibrium, the forward and backward fluxes of reacting molecules are equal and opposite.

Converting this equation from the natural logarithm ( $\ln$ ) to the more commonly used base 10 logarithm ( $\log$ ), we get

$$\Delta G^\circ = -5.94 \log K$$

The above equation reveals how the equilibrium ratio of X to Y (expressed as the equilibrium constant,  $K$ ) depends on the intrinsic character of the molecules, (as expressed in the value of  $\Delta G^\circ$  in kilojoules per mole). Note that for every 5.94 kJ/mole difference in free energy at 37°C, the equilibrium constant changes by a factor of 10 (Table 2–2). Thus, the more energetically favorable a reaction, the more product will accumulate if the reaction proceeds to equilibrium.

More generally, for a reaction that has multiple reactants and products, such as  $A + B \rightarrow C + D$ ,

$$K = \frac{[C][D]}{[A][B]}$$

The concentrations of the two reactants and the two products are multiplied because the rate of the forward reaction depends on the collision of A and B and the rate of the backward reaction depends on the collision of C and D. Thus, at 37°C,

$$\Delta G^\circ = -5.94 \log \frac{[C][D]}{[A][B]}$$

where  $\Delta G^\circ$  is in kilojoules per mole, and [A], [B], [C], and [D] denote the concentrations of the reactants and products in moles/liter.

### The Free-Energy Changes of Coupled Reactions Are Additive

We have pointed out that unfavorable reactions can be coupled to favorable ones to drive the unfavorable ones forward (see Figure 2–29). In thermodynamic terms, this is possible because the overall free-energy change for a set of coupled reactions is the sum of the free-energy changes in each of its component steps. Consider, as a simple example, two sequential reactions



whose  $\Delta G^\circ$  values are +5 and -13 kJ/mole, respectively. If these two reactions occur sequentially, the  $\Delta G^\circ$  for the coupled reaction will be -8 kJ/mole. This means that, with appropriate conditions, the unfavorable reaction  $X \rightarrow Y$  can be driven by the favorable reaction  $Y \rightarrow Z$ , provided that this second reaction follows the first. For example, several of the reactions in the long pathway that converts sugars into  $\text{CO}_2$  and  $\text{H}_2\text{O}$  have positive  $\Delta G^\circ$  values. But the pathway nevertheless proceeds because the total  $\Delta G^\circ$  for the series of sequential reactions has a large negative value.

Forming a sequential pathway is not adequate for many purposes. Often the desired pathway is simply  $X \rightarrow Y$ , without further conversion of Y to some other product. Fortunately, there are other more general ways of using enzymes to couple reactions together. These often involve the activated carrier molecules that we discuss next.

### Activated Carrier Molecules Are Essential for Biosynthesis

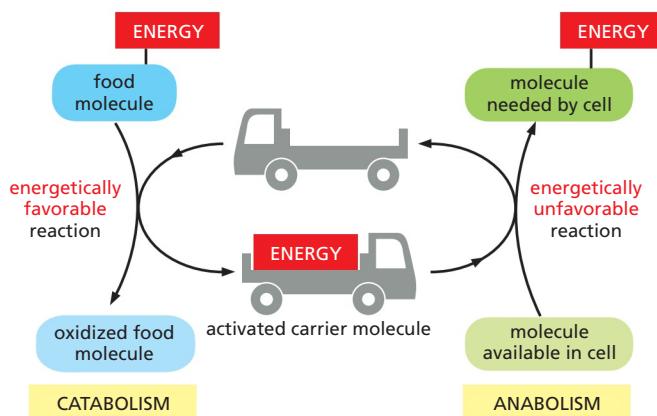
The energy released by the oxidation of food molecules must be stored temporarily before it can be channeled into the construction of the many other molecules needed by the cell. In most cases, the energy is stored as chemical-bond energy in a small set of activated “carrier molecules,” which contain one or more energy-rich covalent bonds. These molecules diffuse rapidly throughout the cell and thereby carry their bond energy from sites of energy generation to the sites where the energy will be used for biosynthesis and other cell activities (Figure 2–31).

The **activated carriers** store energy in an easily exchangeable form, either as a readily transferable chemical group or as electrons held at a high energy level, and they can serve a dual role as a source of both energy and chemical groups in biosynthetic reactions. For historical reasons, these molecules are also sometimes referred to as *coenzymes*. The most important of the activated carrier molecules

**TABLE 2–2 Relationship Between the Standard Free-Energy Change,  $\Delta G^\circ$ , and the Equilibrium Constant**

Equilibrium constant $\frac{[X]}{[Y]} = K$	Free energy of X minus free energy of Y [kJ/mole (kcal/mole)]
$10^5$	-29.7 (-7.1)
$10^4$	-23.8 (-5.7)
$10^3$	-17.8 (-4.3)
$10^2$	-11.9 (-2.8)
$10^1$	-5.9 (-1.4)
1	0 (0)
$10^{-1}$	5.9 (1.4)
$10^{-2}$	11.9 (2.8)
$10^{-3}$	17.8 (4.3)
$10^{-4}$	23.8 (5.7)
$10^{-5}$	29.7 (7.1)

Values of the equilibrium constant were calculated for the simple chemical reaction  $Y \leftrightarrow X$  using the equation given in the text. The  $\Delta G^\circ$  given here is in kilojoules per mole at 37°C, with kilocalories per mole in parentheses. One kilojoule (kJ) is equal to 0.239 kilocalories (kcal) (1 kcal = 4.18 kJ). As explained in the text,  $\Delta G^\circ$  represents the free-energy difference under standard conditions (where all components are present at a concentration of 1.0 mole/liter). From this table, we see that if there is a favorable standard free-energy change ( $\Delta G^\circ$ ) of -17.8 kJ/mole (-4.3 kcal/mole) for the transition  $Y \rightarrow X$ , there will be 1000 times more molecules in state X than in state Y at equilibrium ( $K = 1000$ ).

**Figure 2–31** Energy transfer and the role of activated carriers in metabolism.

By serving as energy shuttles, activated carrier molecules perform their function as go-betweens that link the breakdown of food molecules and the release of energy (catabolism) to the energy-requiring biosynthesis of small and large organic molecules (anabolism).

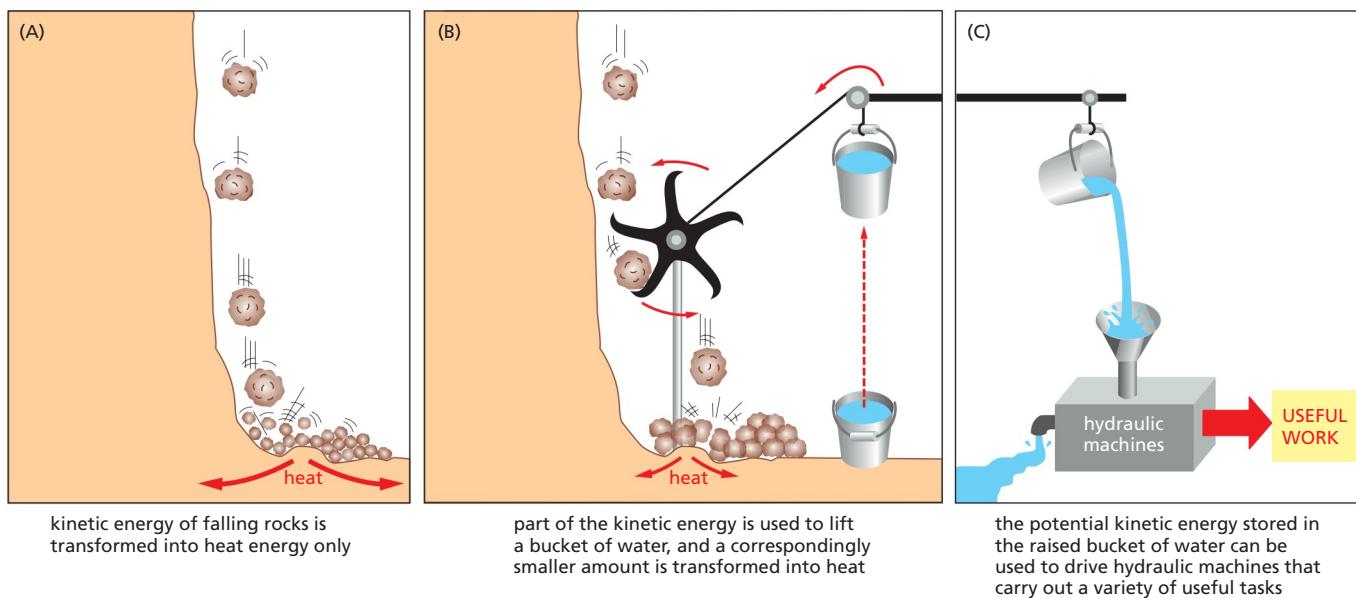
are ATP and two molecules that are closely related to each other, NADH and NADPH. Cells use such activated carrier molecules like money to pay for reactions that otherwise could not take place.

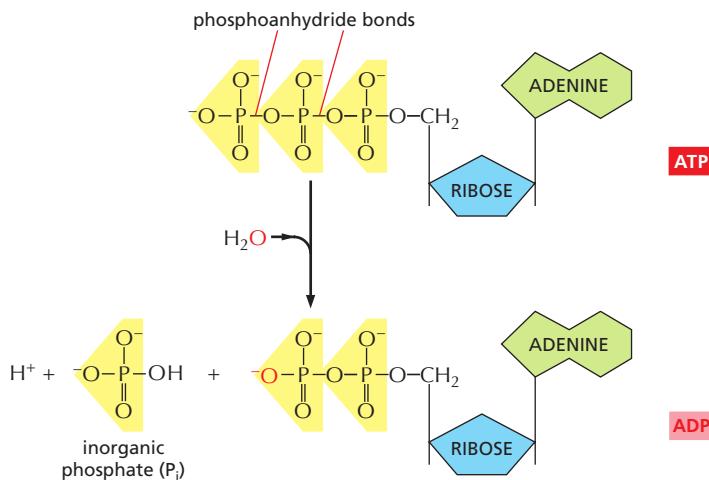
### The Formation of an Activated Carrier Is Coupled to an Energetically Favorable Reaction

Coupling mechanisms require enzymes and are fundamental to all the energy transactions of the cell. The nature of a **coupled reaction** is illustrated by a mechanical analogy in **Figure 2–32**, in which an energetically favorable chemical reaction is represented by rocks falling from a cliff. The energy of falling rocks would normally be entirely wasted in the form of heat generated by friction when the rocks hit the ground (see the falling-brick diagram in Figure 2–17). By careful design, however, part of this energy could be used instead to drive a paddle wheel that lifts a bucket of water (Figure 2–32B). Because the rocks can now reach the ground only after moving the paddle wheel, we say that the energetically favorable reaction of rock falling has been directly *coupled* to the energetically unfavorable reaction of lifting the bucket of water. Note that because part of the energy is used to do work in Figure 2–32B, the rocks hit the ground with less velocity than in Figure 2–32A, and correspondingly less energy is dissipated as heat.

Similar processes occur in cells, where enzymes play the role of the paddle wheel. By mechanisms that we discuss later in this chapter, enzymes couple an

**Figure 2–32** A mechanical model illustrating the principle of coupled chemical reactions. The spontaneous reaction shown in (A) could serve as an analogy for the direct oxidation of glucose to  $\text{CO}_2$  and  $\text{H}_2\text{O}$ , which produces heat only. In (B), the same reaction is coupled to a second reaction; this second reaction is analogous to the synthesis of activated carrier molecules. The energy produced in (B) is in a more useful form than in (A) and can be used to drive a variety of otherwise energetically unfavorable reactions (C).





**Figure 2–33** The hydrolysis of ATP to ADP and inorganic phosphate. The two outermost phosphates in ATP are held to the rest of the molecule by high-energy phosphoanhydride bonds and are readily transferred. As indicated, water can be added to ATP to form ADP and inorganic phosphate ( $\text{P}_i$ ). Hydrolysis of the terminal phosphate of ATP yields between 46 and 54 kJ/mole of usable energy, depending on the intracellular conditions. The large negative  $\Delta G$  of this reaction arises from several factors: release of the terminal phosphate group removes an unfavorable repulsion between adjacent negative charges, and the inorganic phosphate ion ( $\text{P}_i$ ) released is stabilized by resonance and by favorable hydrogen-bond formation with water.

energetically favorable reaction, such as the oxidation of foodstuffs, to an energetically unfavorable reaction, such as the generation of an activated carrier molecule. In this example, the amount of heat released by the oxidation reaction is reduced by exactly the amount of energy stored in the energy-rich covalent bonds of the activated carrier molecule. And the activated carrier molecule picks up a packet of energy of a size sufficient to power a chemical reaction elsewhere in the cell.

### ATP Is the Most Widely Used Activated Carrier Molecule

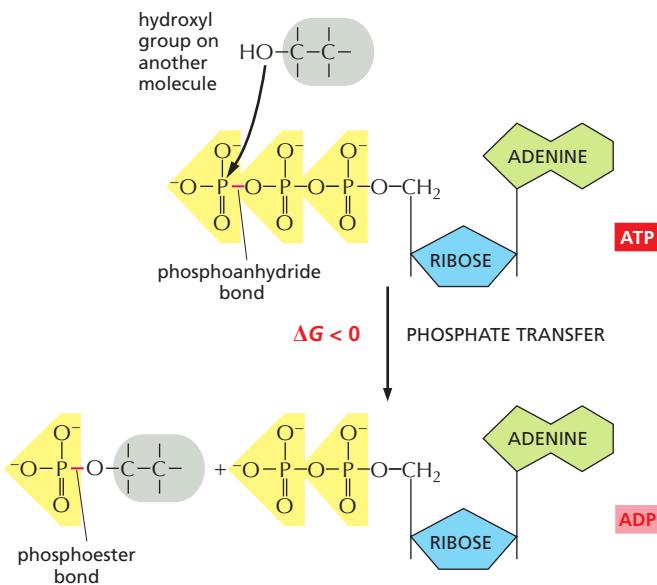
The most important and versatile of the activated carriers in cells is ATP (adenosine triphosphate). Just as the energy stored in the raised bucket of water in Figure 2–32B can drive a wide variety of hydraulic machines, ATP is a convenient and versatile store, or currency, of energy used to drive a variety of chemical reactions in cells. ATP is synthesized in an energetically unfavorable phosphorylation reaction in which a phosphate group is added to ADP (adenosine diphosphate). When required, ATP gives up its energy packet through its energetically favorable hydrolysis to ADP and inorganic phosphate (Figure 2–33). The regenerated ADP is then available to be used for another round of the phosphorylation reaction that forms ATP.

The energetically favorable reaction of ATP hydrolysis is coupled to many otherwise unfavorable reactions through which other molecules are synthesized. Many of these coupled reactions involve the transfer of the terminal phosphate in ATP to another molecule, as illustrated by the phosphorylation reaction in Figure 2–34.

As the most abundant activated carrier in cells, ATP is the principle energy currency. To give just two examples, it supplies energy for many of the pumps that transport substances into and out of the cell (discussed in Chapter 11), and it powers the molecular motors that enable muscle cells to contract and nerve cells to transport materials from one end of their long axons to another (discussed in Chapter 16).

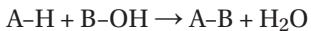
### Energy Stored in ATP Is Often Harnessed to Join Two Molecules Together

We have previously discussed one way in which an energetically favorable reaction can be coupled to an energetically unfavorable reaction,  $X \rightarrow Y$ , so as to enable it to occur. In that scheme, a second enzyme catalyzes the energetically favorable reaction  $Y \rightarrow Z$ , pulling all of the  $X$  to  $Y$  in the process. But when the required product is  $Y$  and not  $Z$ , this mechanism is not useful.



**Figure 2–34** An example of a phosphate transfer reaction. Because an energy-rich phosphoanhydride bond in ATP is converted to a phosphoester bond, this reaction is energetically favorable, having a large negative  $\Delta G$ . Reactions of this type are involved in the synthesis of phospholipids and in the initial steps of reactions that catabolize sugars.

A typical biosynthetic reaction is one in which two molecules, A and B, are joined together to produce A-B in the energetically unfavorable *condensation* reaction

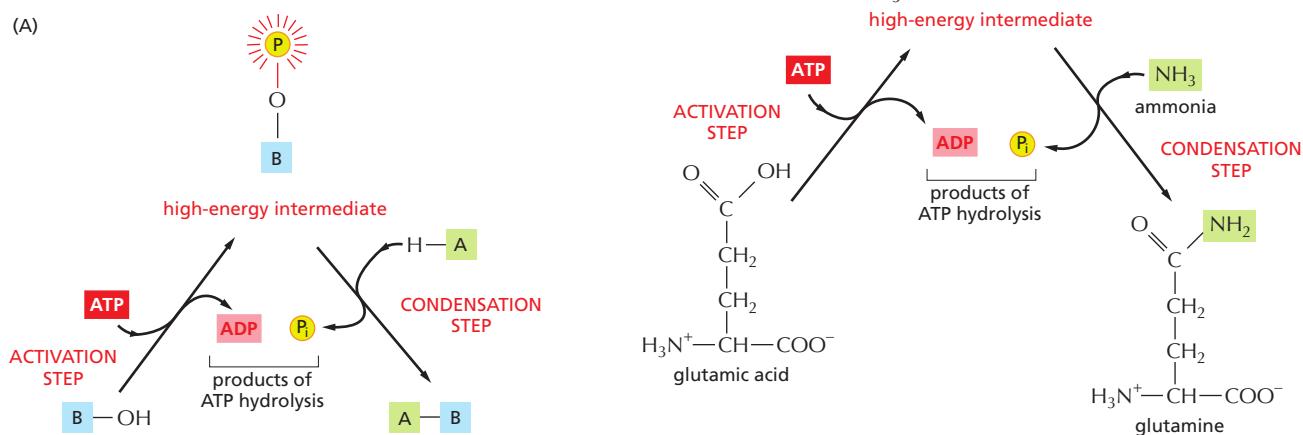


There is an indirect pathway that allows A-H and B-OH to form A-B, in which a coupling to ATP hydrolysis makes the reaction go. Here, energy from ATP hydrolysis is first used to convert B-OH to a higher-energy intermediate compound, which then reacts directly with A-H to give A-B. The simplest possible mechanism involves the transfer of a phosphate from ATP to B-OH to make B-O-PO<sub>3</sub>, in which case the reaction pathway contains only two steps:

1. B-OH + ATP → B-O-PO<sub>3</sub> + ADP
  2. A-H + B-O-PO<sub>3</sub> → A-B + P<sub>i</sub>
- Net result: B-OH + ATP + A-H → A-B + ADP + P<sub>i</sub>

The condensation reaction, which by itself is energetically unfavorable, is forced to occur by being directly coupled to ATP hydrolysis in an enzyme-catalyzed reaction pathway (**Figure 2–35A**).

A biosynthetic reaction of exactly this type synthesizes the amino acid glutamine (Figure 2–35B). We will see shortly that similar (but more complex) mechanisms are also used to produce nearly all of the large molecules of the cell.



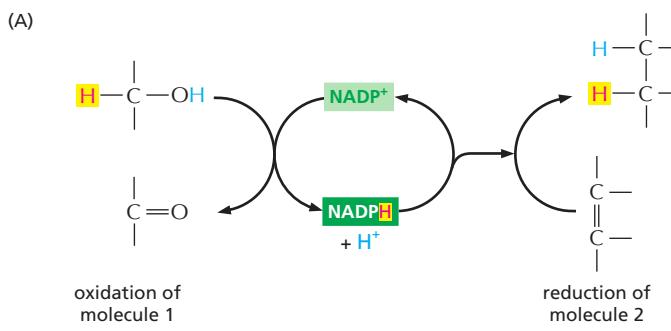
**Figure 2–35** An example of an energetically unfavorable biosynthetic reaction driven by ATP hydrolysis. (A) Schematic illustration of the formation of A-B in the condensation reaction described in the text. (B) The biosynthesis of the common amino acid glutamine from glutamic acid and ammonia. Glutamic acid is first converted to a high-energy phosphorylated intermediate (corresponding to the compound B-O-PO<sub>3</sub> described in the text), which then reacts with ammonia (corresponding to A-H) to form glutamine. In this example, both steps occur on the surface of the same enzyme, *glutamine synthetase*. The high-energy bonds are shaded red; here, as elsewhere throughout the book, the symbol P<sub>i</sub> = HPO<sub>4</sub><sup>2-</sup>, and a yellow “circled P” = PO<sub>3</sub><sup>2-</sup>.

## NADH and NADPH Are Important Electron Carriers

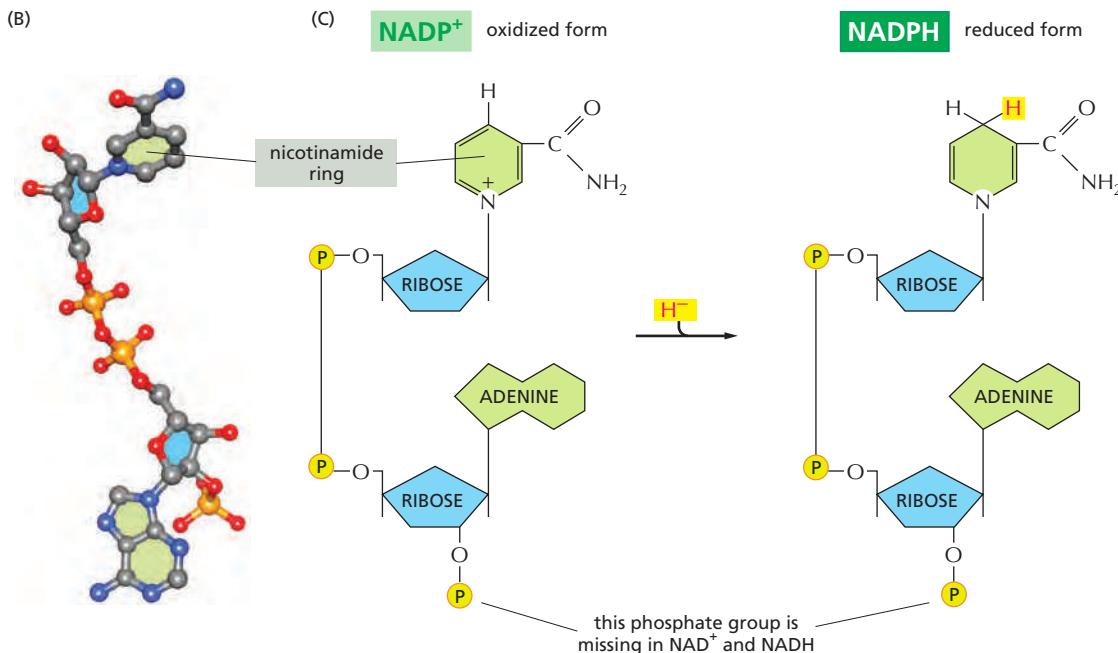
Other important activated carrier molecules participate in oxidation-reduction reactions and are commonly part of coupled reactions in cells. These activated carriers are specialized to carry electrons held at a high energy level (sometimes called “high-energy” electrons) and hydrogen atoms. The most important of these electron carriers are **NAD<sup>+</sup>** (nicotinamide adenine dinucleotide) and the closely related molecule **NADP<sup>+</sup>** (nicotinamide adenine dinucleotide phosphate). Each picks up a “packet of energy” corresponding to two electrons plus a proton ( $H^+$ ), and they are thereby converted to **NADH** (*reduced* nicotinamide adenine dinucleotide) and **NADPH** (*reduced* nicotinamide adenine dinucleotide phosphate), respectively (Figure 2–36). These molecules can therefore be regarded as carriers of hydride ions (the  $H^+$  plus two electrons, or  $H^-$ ).

Like ATP, NADPH is an activated carrier that participates in many important biosynthetic reactions that would otherwise be energetically unfavorable. The NADPH is produced according to the general scheme shown in Figure 2–36A. During a special set of energy-yielding catabolic reactions, two hydrogen atoms are removed from a substrate molecule. Both electrons but just one proton (that is, a hydride ion,  $H^-$ ) are added to the nicotinamide ring of NADP<sup>+</sup> to form NADPH; the second proton ( $H^+$ ) is released into solution. This is a typical oxidation-reduction reaction, in which the substrate is oxidized and NADP<sup>+</sup> is reduced.

NADPH readily gives up the hydride ion it carries in a subsequent oxidation-reduction reaction, because the nicotinamide ring can achieve a more stable arrangement of electrons without it. In this subsequent reaction, which



**Figure 2–36 NADPH, an important carrier of electrons.**  
 (A) NADPH is produced in reactions of the general type shown on the left, in which two hydrogen atoms are removed from a substrate. The oxidized form of the carrier molecule, NADP<sup>+</sup>, receives one hydrogen atom plus an electron (a hydride ion); the proton ( $H^+$ ) from the other H atom is released into solution. Because NADPH holds its hydride ion in a high-energy linkage, the hydride ion can easily be transferred to other molecules, as shown on the right. (B) and (C) The structures of NADP<sup>+</sup> and NADPH. The part of the NADP<sup>+</sup> molecule known as the nicotinamide ring accepts the hydride ion,  $H^-$ , forming NADPH. The molecules NAD<sup>+</sup> and NADH are identical in structure to NADP<sup>+</sup> and NADPH, respectively, except that they lack the indicated phosphate group.



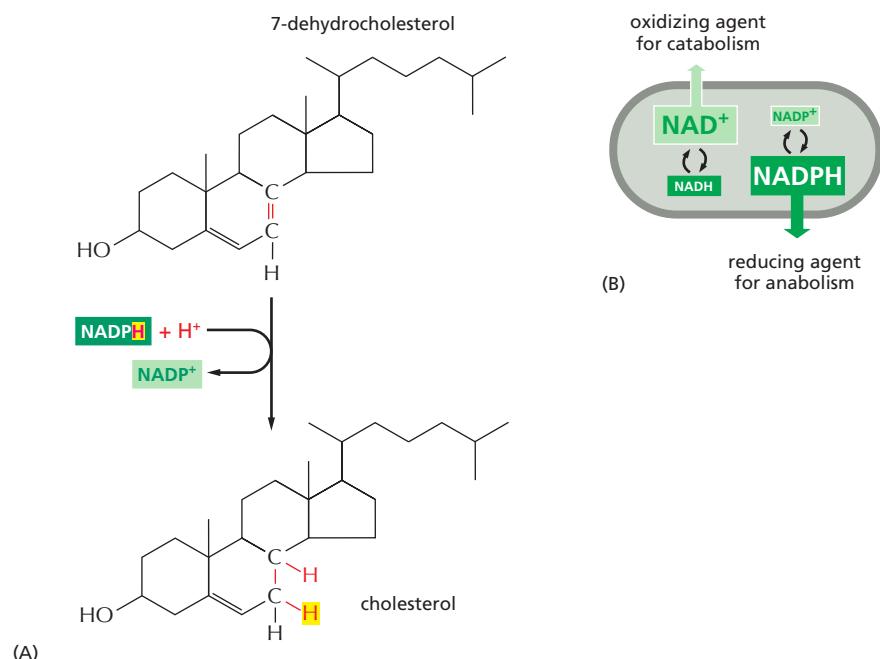
regenerates NADP<sup>+</sup>, it is the NADPH that is oxidized and the substrate that is reduced. The NADPH is an effective donor of its hydride ion to other molecules for the same reason that ATP readily transfers a phosphate: in both cases the transfer is accompanied by a large negative free-energy change. One example of the use of NADPH in biosynthesis is shown in **Figure 2–37**.

The extra phosphate group on NADPH has no effect on the electron-transfer properties of NADPH compared with NADH, being far away from the region involved in electron transfer (see Figure 2–36C). It does, however, give a molecule of NADPH a slightly different shape from that of NADH, making it possible for NADPH and NADH to bind as substrates to completely different sets of enzymes. Thus, the two types of carriers are used to transfer electrons (or hydride ions) between two different sets of molecules.

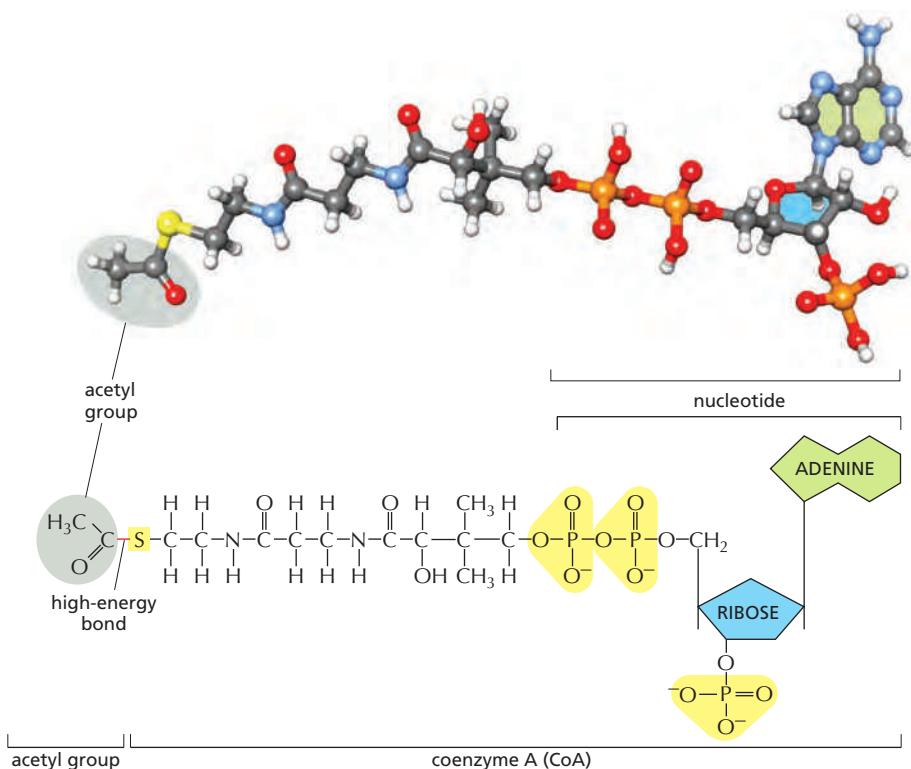
Why should there be this division of labor? The answer lies in the need to regulate two sets of electron-transfer reactions independently. NADPH operates chiefly with enzymes that catalyze anabolic reactions, supplying the high-energy electrons needed to synthesize energy-rich biological molecules. NADH, by contrast, has a special role as an intermediate in the catabolic system of reactions that generate ATP through the oxidation of food molecules, as we will discuss shortly. The genesis of NADH from NAD<sup>+</sup>, and of NADPH from NADP<sup>+</sup>, occur by different pathways and are independently regulated, so that the cell can adjust the supply of electrons for these two contrasting purposes. Inside the cell the ratio of NAD<sup>+</sup> to NADH is kept high, whereas the ratio of NADP<sup>+</sup> to NADPH is kept low. This provides plenty of NAD<sup>+</sup> to act as an oxidizing agent and plenty of NADPH to act as a reducing agent (Figure 2–37B)—as required for their special roles in catabolism and anabolism, respectively.

### There Are Many Other Activated Carrier Molecules in Cells

Other activated carriers also pick up and carry a chemical group in an easily transferred, high-energy linkage. For example, coenzyme A carries a readily transferable



**Figure 2–37 NADPH as a reducing agent.** (A) The final stage in the biosynthetic route leading to cholesterol. As in many other biosynthetic reactions, the reduction of the C=C bond is achieved by the transfer of a hydride ion from the carrier molecule NADPH, plus a proton (H<sup>+</sup>) from the solution. (B) Keeping NADPH levels high and NADH levels low alters their affinities for electrons (see Panel 14–1, p. 765). This causes NADPH to be a much stronger electron donor (reducing agent) than NADH, and NAD<sup>+</sup> therefore to be a much better electron acceptor (oxidizing agent) than NADP<sup>+</sup>, as indicated.

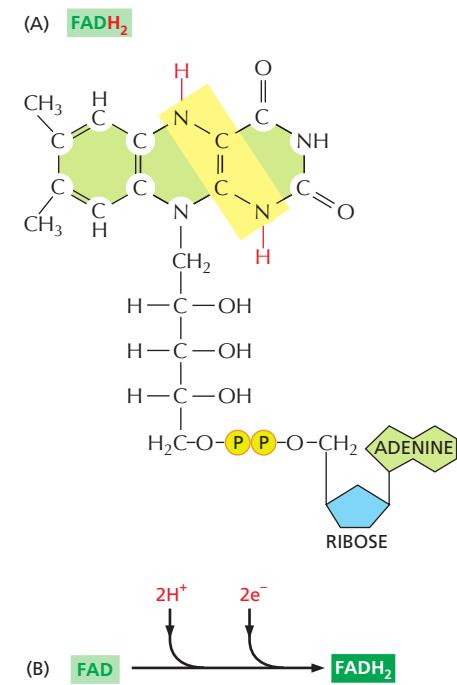


**Figure 2-38** The structure of the important activated carrier molecule acetyl CoA. A ball-and-stick model is shown above the structure. The sulfur atom (yellow) forms a thioester bond to acetate. Because this is a high-energy linkage, releasing a large amount of free energy when it is hydrolyzed, the acetate molecule can be readily transferred to other molecules.

acetyl group in a thioester linkage, and in this activated form is known as **acetyl CoA** (acetyl coenzyme A). Acetyl CoA (Figure 2-38) is used to add two carbon units in the biosynthesis of larger molecules.

In acetyl CoA, as in other carrier molecules, the transferable group makes up only a small part of the molecule. The rest consists of a large organic portion that serves as a convenient “handle,” facilitating the recognition of the carrier molecule by specific enzymes. As with acetyl CoA, this handle portion very often contains a nucleotide (usually adenosine), a curious fact that may be a relic from an early stage of evolution. It is currently thought that the main catalysts for early life-forms—before DNA or proteins—were RNA molecules (or their close relatives), as described in Chapter 6. It is tempting to speculate that many of the carrier molecules that we find today originated in this earlier RNA world, where their nucleotide portions could have been useful for binding them to RNA enzymes (ribozymes).

Thus, ATP transfers phosphate, NADPH transfers electrons and hydrogen, and acetyl CoA transfers two-carbon acetyl groups. **FADH<sub>2</sub>** (reduced flavin adenine dinucleotide) is used like NADH in electron and proton transfers (Figure 2-39). The reactions of other activated carrier molecules involve the transfer of a methyl, carboxyl, or glucose group for biosyntheses (Table 2-3). These activated carriers



**Figure 2-39** FADH<sub>2</sub> is a carrier of hydrogens and high-energy electrons, like NADH and NADPH. (A) Structure of FADH<sub>2</sub>, with its hydrogen-carrying atoms highlighted in yellow. (B) The formation of FADH<sub>2</sub> from FAD.

**TABLE 2-3** Some Activated Carrier Molecules Widely Used in Metabolism

Activated carrier	Group carried in high-energy linkage
ATP	Phosphate
NADH, NADPH, FADH <sub>2</sub>	Electrons and hydrogens
Acetyl CoA	Acetyl group
Carboxylated biotin	Carboxyl group
S-Adenosylmethionine	Methyl group
Uridine diphosphate glucose	Glucose

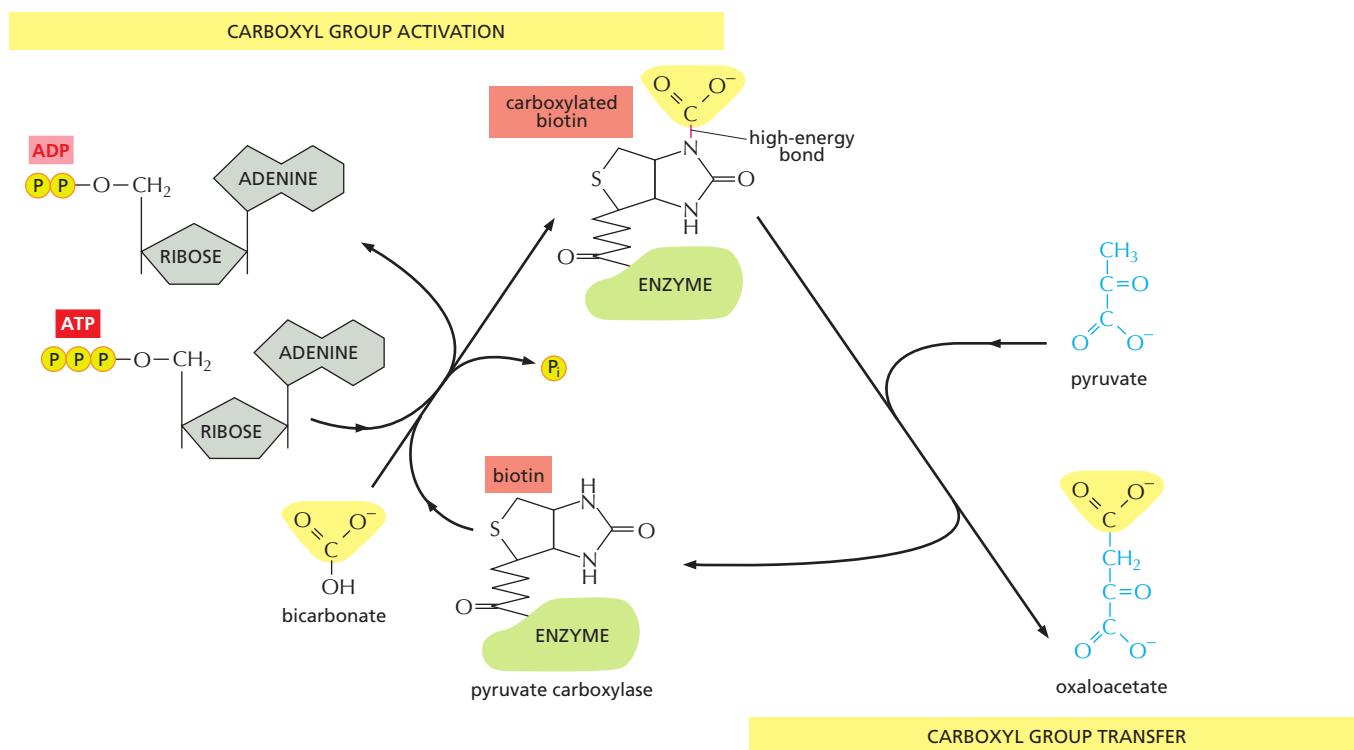
are generated in reactions that are coupled to ATP hydrolysis, as in the example in **Figure 2–40**. Therefore, the energy that enables their groups to be used for biosynthesis ultimately comes from the catabolic reactions that generate ATP. Similar processes occur in the synthesis of the very large molecules of the cell—the nucleic acids, proteins, and polysaccharides—that we discuss next.

### The Synthesis of Biological Polymers Is Driven by ATP Hydrolysis

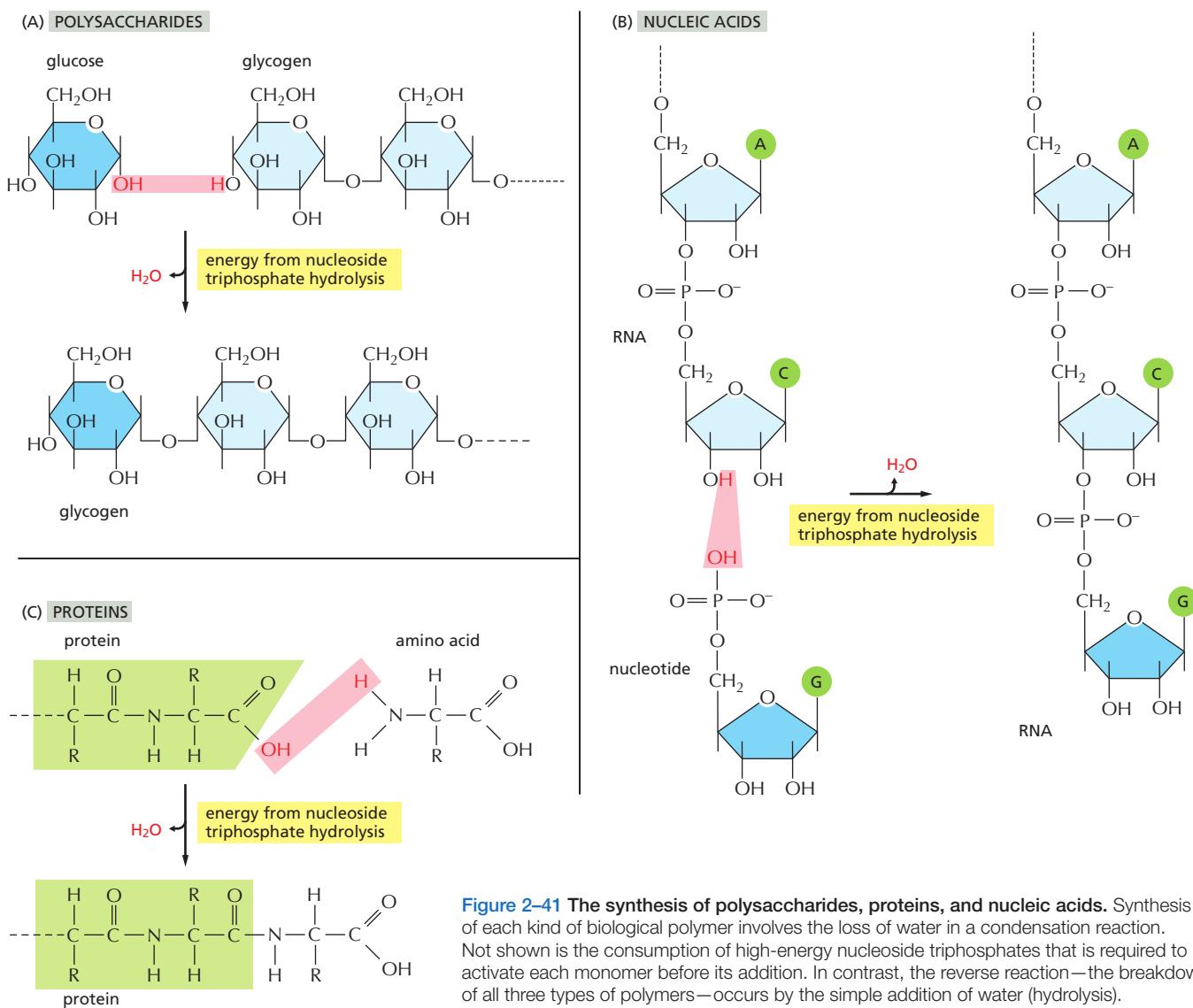
As discussed previously, the macromolecules of the cell constitute most of its dry mass (see Figure 2–7). These molecules are made from subunits (or monomers) that are linked together in a *condensation* reaction, in which the constituents of a water molecule (OH plus H) are removed from the two reactants. Consequently, the reverse reaction—the breakdown of all three types of polymers—occurs by the enzyme-catalyzed addition of water (*hydrolysis*). This hydrolysis reaction is energetically favorable, whereas the biosynthetic reactions require an energy input (see Figure 2–9).

The nucleic acids (DNA and RNA), proteins, and polysaccharides are all polymers that are produced by the repeated addition of a monomer onto one end of a growing chain. The synthesis reactions for these three types of macromolecules are outlined in **Figure 2–41**. As indicated, the condensation step in each case depends on energy from nucleoside triphosphate hydrolysis. And yet, except for the nucleic acids, there are no phosphate groups left in the final product molecules. How are the reactions that release the energy of ATP hydrolysis coupled to polymer synthesis?

For each type of macromolecule, an enzyme-catalyzed pathway exists which resembles that discussed previously for the synthesis of the amino acid glutamine (see Figure 2–35). The principle is exactly the same, in that the -OH group that will



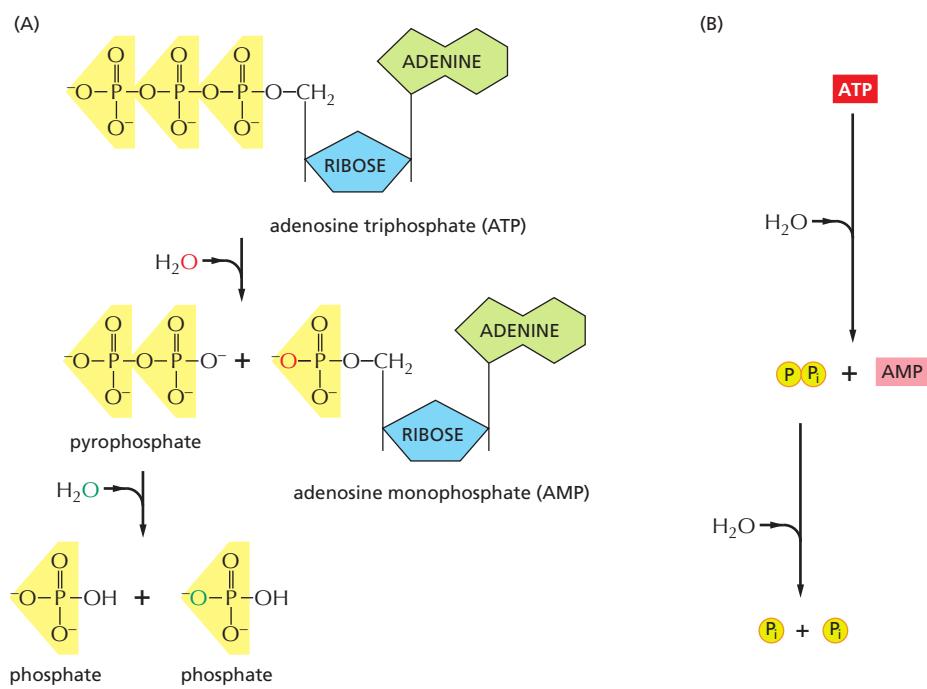
**Figure 2–40** A carboxyl group-transfer reaction using an activated carrier molecule. Carboxylated biotin is used by the enzyme *pyruvate carboxylase* to transfer a carboxyl group in the production of oxaloacetate, a molecule needed for the citric acid cycle. The acceptor molecule for this group-transfer reaction is pyruvate. Other enzymes use biotin, a B-complex vitamin, to transfer carboxyl groups to other acceptor molecules. Note that synthesis of carboxylated biotin requires energy that is derived from ATP—a general feature of many activated carriers.



**Figure 2–41** The synthesis of polysaccharides, proteins, and nucleic acids. Synthesis of each kind of biological polymer involves the loss of water in a condensation reaction. Not shown is the consumption of high-energy nucleoside triphosphates that is required to activate each monomer before its addition. In contrast, the reverse reaction—the breakdown of all three types of polymers—occurs by the simple addition of water (hydrolysis).

be removed in the condensation reaction is first activated by becoming involved in a high-energy linkage to a second molecule. However, the actual mechanisms used to link ATP hydrolysis to the synthesis of proteins and polysaccharides are more complex than that used for glutamine synthesis, since a series of high-energy intermediates is required to generate the final high-energy bond that is broken during the condensation step (discussed in Chapter 6 for protein synthesis).

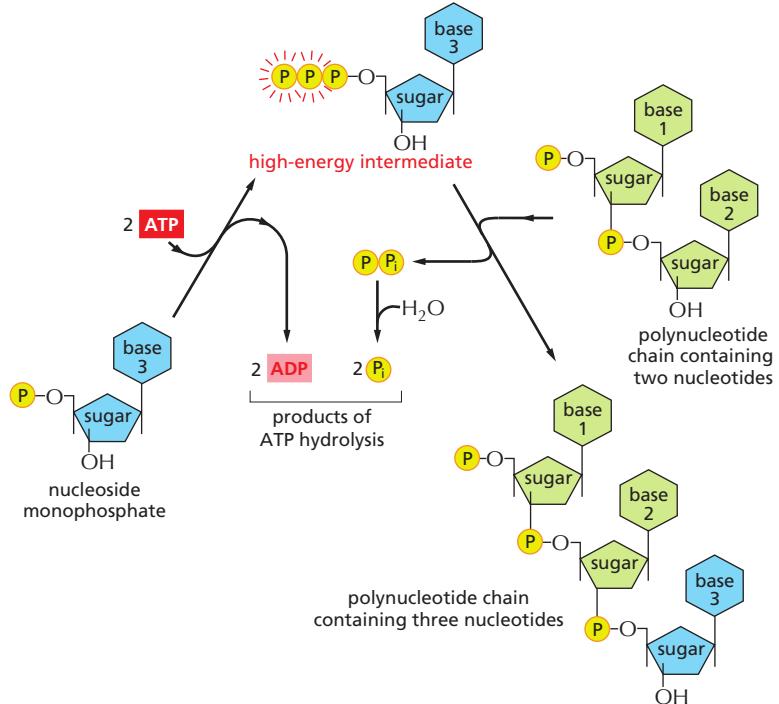
Each activated carrier has limits in its ability to drive a biosynthetic reaction. The  $\Delta G$  for the hydrolysis of ATP to ADP and inorganic phosphate ( $P_i$ ) depends on the concentrations of all of the reactants, but under the usual conditions in a cell it is between  $-46$  and  $-54$  kJ/mole. In principle, this hydrolysis reaction could drive an unfavorable reaction with a  $\Delta G$  of, perhaps,  $+40$  kJ/mole, provided that a suitable reaction path is available. For some biosynthetic reactions, however, even  $-50$  kJ/mole does not provide enough of a driving force. In these cases, the path of ATP hydrolysis can be altered so that it initially produces AMP and *pyrophosphate* ( $PP_i$ ), which is itself then hydrolyzed in a subsequent step (Figure 2–42). The whole process makes available a total free-energy change of about  $-100$  kJ/mole. An important type of biosynthetic reaction that is driven in this way is the



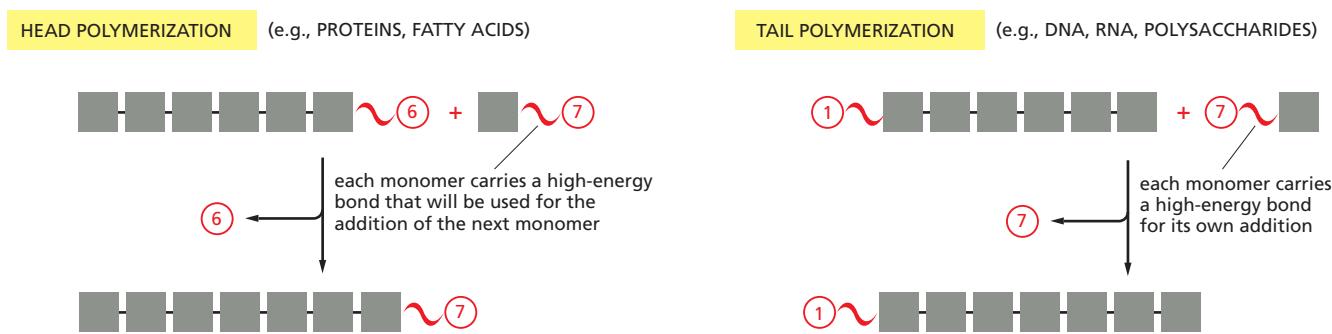
**Figure 2–42** An alternative pathway of ATP hydrolysis, in which pyrophosphate is first formed and then hydrolyzed. This route releases about twice as much free energy (approximately –100 kJ/mole) as the reaction shown earlier in Figure 2–33, and it forms AMP instead of ADP. (A) In the two successive hydrolysis reactions, oxygen atoms from the participating water molecules are retained in the products, as indicated, whereas the hydrogen atoms dissociate to form free hydrogen ions ( $H^+$ , not shown). (B) Summary of overall reaction.

synthesis of nucleic acids (polynucleotides) from nucleoside triphosphates, as illustrated on the right side of **Figure 2–43**.

Note that the repetitive condensation reactions that produce macromolecules can be oriented in one of two ways, giving rise to either the head polymerization or the tail polymerization of monomers. In so-called *head polymerization*, the reactive bond required for the condensation reaction is carried on the end of the



**Figure 2–43** Synthesis of a polynucleotide, RNA or DNA, is a multistep process driven by ATP hydrolysis. In the first step, a nucleoside monophosphate is activated by the sequential transfer of the terminal phosphate groups from two ATP molecules. The high-energy intermediate formed—a nucleoside triphosphate—exists free in solution until it reacts with the growing end of an RNA or a DNA chain with release of pyrophosphate. Hydrolysis of the latter to inorganic phosphate is highly favorable and helps to drive the overall reaction in the direction of polynucleotide synthesis. For details, see Chapter 5.



growing polymer, and it must therefore be regenerated each time that a monomer is added. In this case, each monomer brings with it the reactive bond that will be used in adding the *next* monomer in the series. In *tail polymerization*, the reactive bond carried by each monomer is instead used immediately for its own addition (Figure 2–44).

We shall see in later chapters that both of these types of polymerization are used. The synthesis of polynucleotides and some simple polysaccharides occurs by tail polymerization, for example, whereas the synthesis of proteins occurs by a head polymerization process.

**Figure 2–44** The orientation of the active intermediates in the repetitive condensation reactions that form biological polymers. The head growth of polymers is compared with its alternative, tail growth. As indicated, these two mechanisms are used to produce different types of biological macromolecules.

## Summary

*Living cells need to create and maintain order within themselves to survive and grow. This is thermodynamically possible only because of a continual input of energy, part of which must be released from the cells to their environment as heat that disorders the surroundings. The only chemical reactions possible are those that increase the total amount of disorder in the universe. The free-energy change for a reaction,  $\Delta G$ , measures this disorder, and it must be less than zero for a reaction to proceed spontaneously. This  $\Delta G$  depends both on the intrinsic properties of the reacting molecules and their concentrations, and it can be calculated from these concentrations if either the equilibrium constant ( $K$ ) for the reaction or its standard free-energy change,  $\Delta G^\circ$ , is known.*

*The energy needed for life comes ultimately from the electromagnetic radiation of the sun, which drives the formation of organic molecules in photosynthetic organisms such as green plants. Animals obtain their energy by eating organic molecules and oxidizing them in a series of enzyme-catalyzed reactions that are coupled to the formation of ATP—a common currency of energy in all cells.*

*To make possible the continual generation of order in cells, energetically favorable reactions, such as the hydrolysis of ATP, are coupled to energetically unfavorable reactions. In the biosynthesis of macromolecules, ATP is used to form reactive phosphorylated intermediates. Because the energetically unfavorable reaction of biosynthesis now becomes energetically favorable, ATP hydrolysis is said to drive the reaction. Polymeric molecules such as proteins, nucleic acids, and polysaccharides are assembled from small activated precursor molecules by repetitive condensation reactions that are driven in this way. Other reactive molecules, called either activated carriers or coenzymes, transfer other chemical groups in the course of biosynthesis: NADPH transfers hydrogen as a proton plus two electrons (a hydride ion), for example, whereas acetyl CoA transfers an acetyl group.*

## HOW CELLS OBTAIN ENERGY FROM FOOD

The constant supply of energy that cells need to generate and maintain the biological order that keeps them alive comes from the chemical-bond energy in food molecules.

The proteins, lipids, and polysaccharides that make up most of the food we eat must be broken down into smaller molecules before our cells can use them—either

as a source of energy or as building blocks for other molecules. Enzymatic digestion breaks down the large polymeric molecules in food into their monomer sub-units—proteins into amino acids, polysaccharides into sugars, and fats into fatty acids and glycerol. After digestion, the small organic molecules derived from food enter the cytosol of cells, where their gradual oxidation begins.

Sugars are particularly important fuel molecules, and they are oxidized in small controlled steps to carbon dioxide ( $\text{CO}_2$ ) and water (Figure 2–45). In this section, we trace the major steps in the breakdown, or catabolism, of sugars and show how they produce ATP, NADH, and other activated carrier molecules in animal cells. A very similar pathway also operates in plants, fungi, and many bacteria. As we shall see, the oxidation of fatty acids is equally important for cells. Other molecules, such as proteins, can also serve as energy sources when they are funneled through appropriate enzymatic pathways.

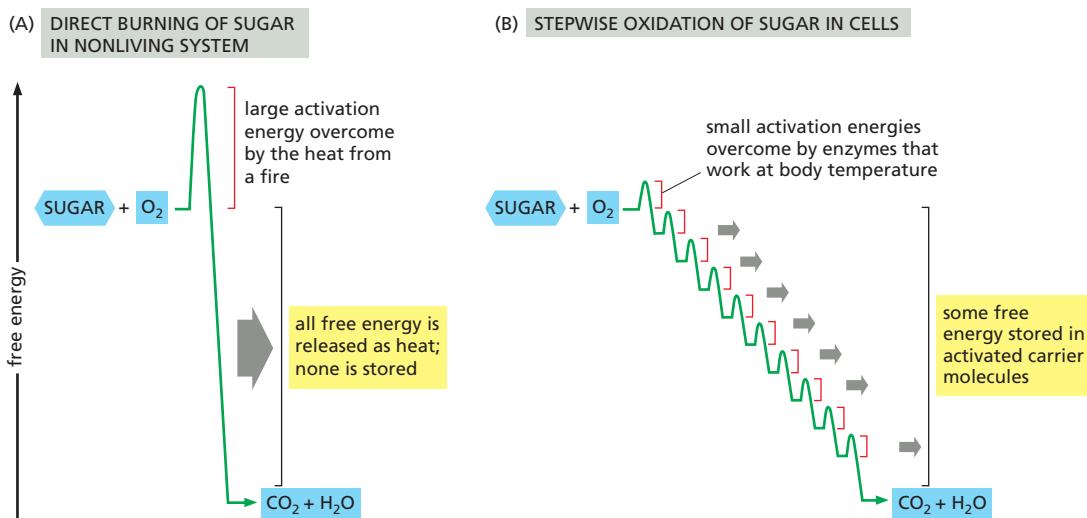
### Glycolysis Is a Central ATP-Producing Pathway

The major process for oxidizing sugars is the sequence of reactions known as **glycolysis**—from the Greek *glukos*, “sweet,” and *lisis*, “rupture.” Glycolysis produces ATP without the involvement of molecular oxygen ( $\text{O}_2$  gas). It occurs in the cytosol of most cells, including many anaerobic microorganisms. Glycolysis probably evolved early in the history of life, before photosynthetic organisms introduced oxygen into the atmosphere. During glycolysis, a glucose molecule with six carbon atoms is converted into two molecules of *pyruvate*, each of which contains three carbon atoms. For each glucose molecule, two molecules of ATP are hydrolyzed to provide energy to drive the early steps, but four molecules of ATP are produced in the later steps. At the end of glycolysis, there is consequently a net gain of two molecules of ATP for each glucose molecule broken down. Two molecules of the activated carrier NADH are also produced.

The glycolytic pathway is outlined in Figure 2–46 and shown in more detail in Panel 2–8 (pp. 104–105) and Movie 2.5. Glycolysis involves a sequence of 10 separate reactions, each producing a different sugar intermediate and each catalyzed by a different enzyme. Like most enzymes, these have names ending in *ase*—such as *isomerase* and *dehydrogenase*—to indicate the type of reaction they catalyze.

Although no molecular oxygen is used in glycolysis, oxidation occurs, in that electrons are removed by  $\text{NAD}^+$  (producing NADH) from some of the carbons derived from the glucose molecule. The stepwise nature of the process releases the energy of oxidation in small packets, so that much of it can be stored in activated carrier molecules rather than all of it being released as heat (see Figure 2–45). Thus, some of the energy released by oxidation drives the direct synthesis of ATP molecules from ADP and  $\text{P}_i$ , and some remains with the electrons in the electron carrier NADH.

**Figure 2–45** Schematic representation of the controlled stepwise oxidation of sugar in a cell, compared with ordinary burning. (A) If the sugar were oxidized to  $\text{CO}_2$  and  $\text{H}_2\text{O}$  in a single step, it would release an amount of energy much larger than could be captured for useful purposes. (B) In the cell, enzymes catalyze oxidation via a series of small steps in which free energy is transferred in conveniently sized packets to carrier molecules—most often ATP and NADH. At each step, an enzyme controls the reaction by reducing the activation-energy barrier that has to be surmounted before the specific reaction can occur. The total free energy released is exactly the same in (A) and (B).



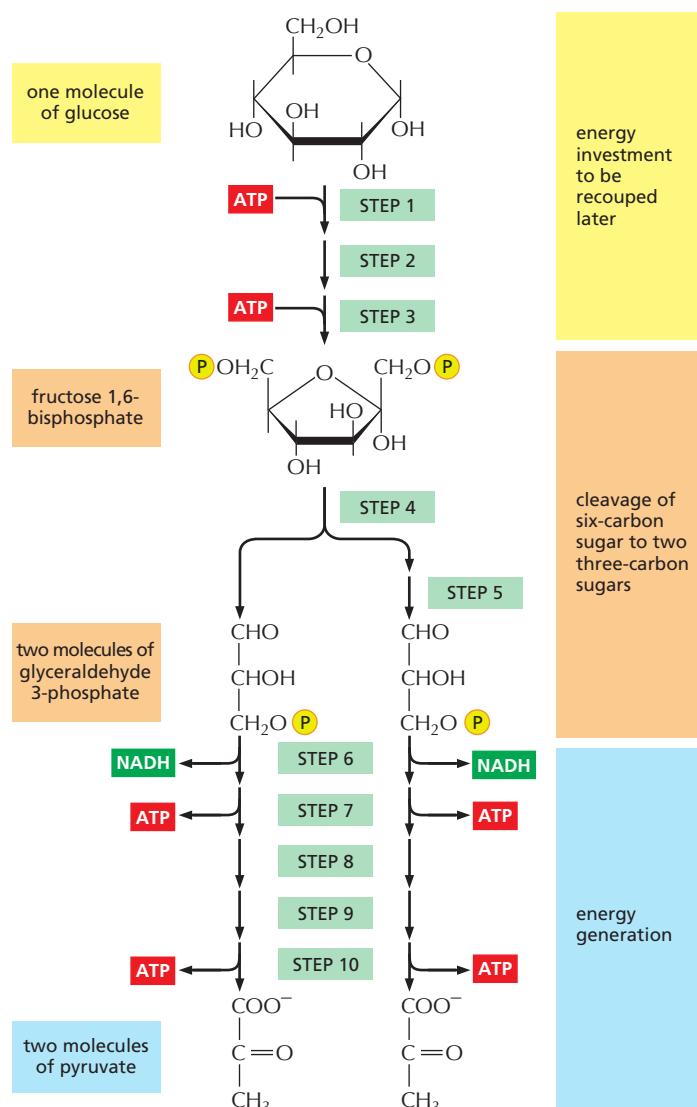
Two molecules of NADH are formed per molecule of glucose in the course of glycolysis. In aerobic organisms, these NADH molecules donate their electrons to the electron-transport chain described in Chapter 14, and the NAD<sup>+</sup> formed from the NADH is used again for glycolysis (see step 6 in Panel 2–8, pp. 104–105).

### Fermentations Produce ATP in the Absence of Oxygen

For most animal and plant cells, glycolysis is only a prelude to the final stage of the breakdown of food molecules. In these cells, the pyruvate formed by glycolysis is rapidly transported into the mitochondria, where it is converted into CO<sub>2</sub> plus acetyl CoA, whose acetyl group is then completely oxidized to CO<sub>2</sub> and H<sub>2</sub>O.

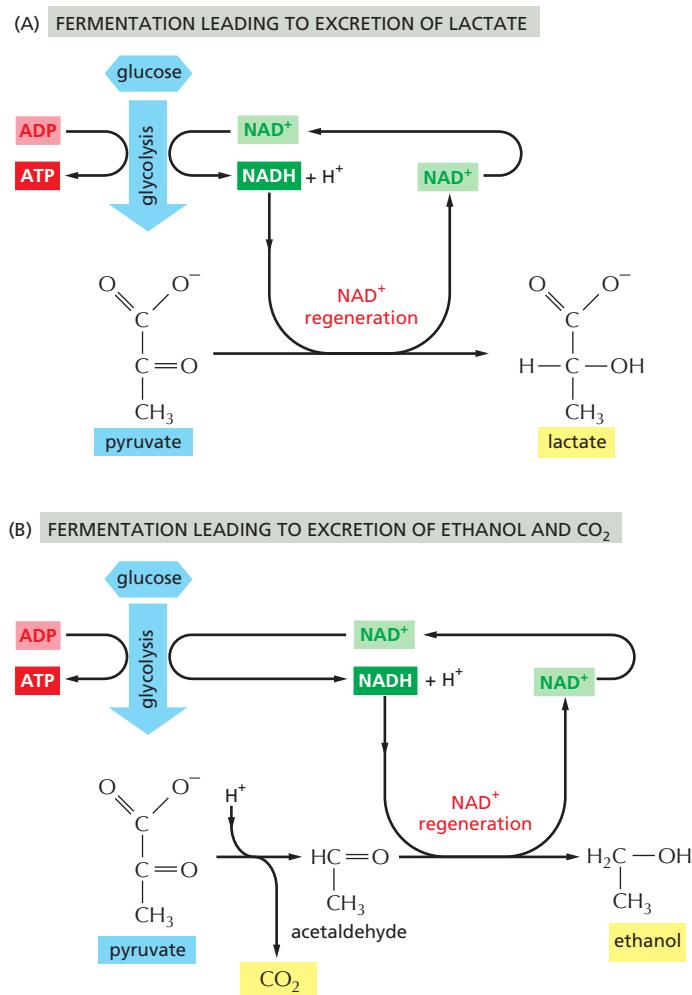
In contrast, for many anaerobic organisms—which do not utilize molecular oxygen and can grow and divide without it—glycolysis is the principal source of the cell's ATP. Certain animal tissues, such as skeletal muscle, can also continue to function when molecular oxygen is limited. In these anaerobic conditions, the pyruvate and the NADH electrons stay in the cytosol. The pyruvate is converted into products excreted from the cell—for example, into ethanol and CO<sub>2</sub> in the yeasts used in brewing and breadmaking, or into lactate in muscle. In this process, the NADH gives up its electrons and is converted back into NAD<sup>+</sup>. This regeneration of NAD<sup>+</sup> is required to maintain the reactions of glycolysis (Figure 2–47).

Energy-yielding pathways like these, in which organic molecules both donate and accept electrons (and which are often, as in these cases, anaerobic), are called



**Figure 2–46 An outline of glycolysis.**

Each of the 10 steps shown is catalyzed by a different enzyme. Note that step 4 cleaves a six-carbon sugar into two three-carbon sugars, so that the number of molecules at every stage after this doubles. As indicated, step 6 begins the energy-generation phase of glycolysis. Because two molecules of ATP are hydrolyzed in the early, energy-investment phase, glycolysis results in the net synthesis of 2 ATP and 2 NADH molecules per molecule of glucose (see also Panel 2–8).



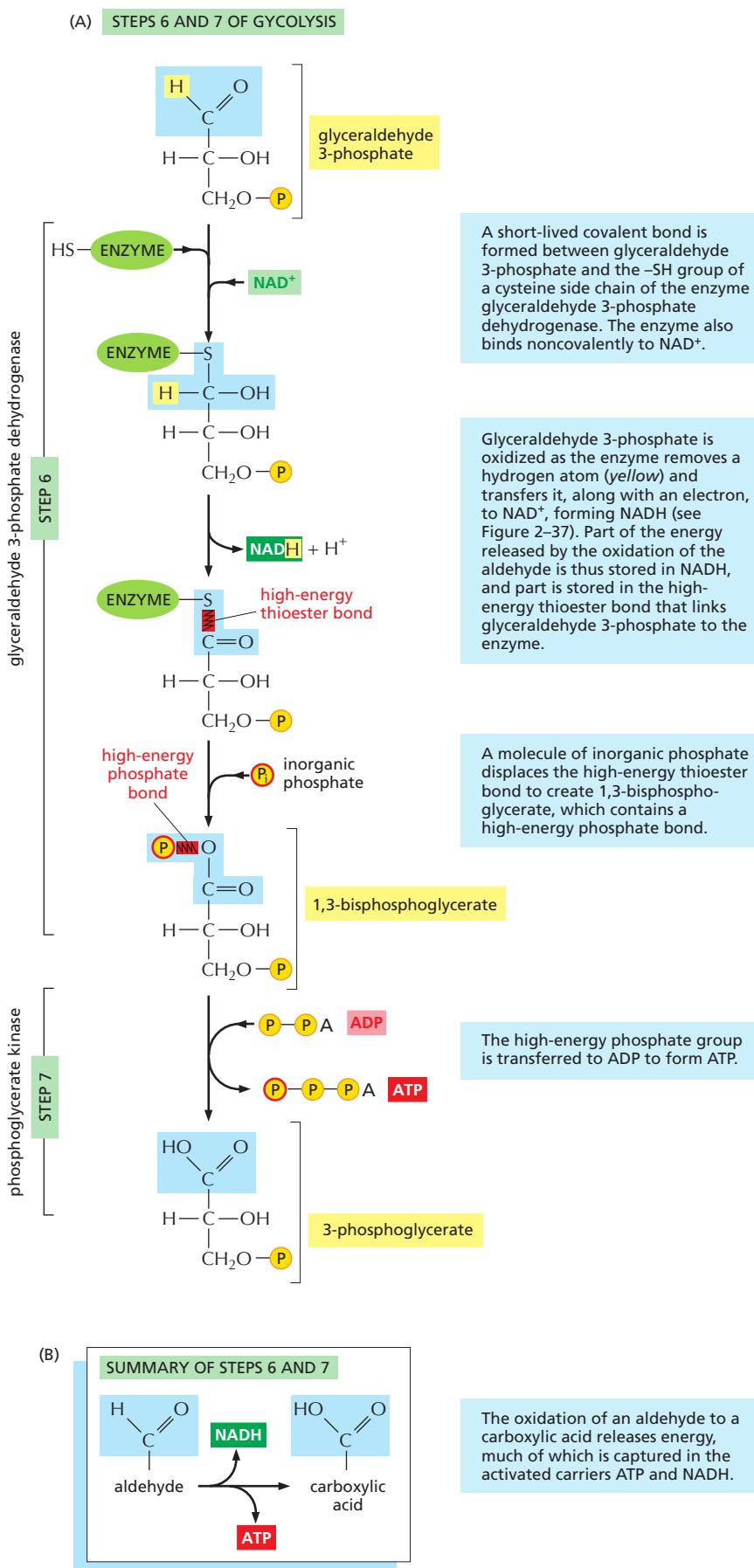
**Figure 2-47** Two pathways for the anaerobic breakdown of pyruvate. (A) When there is inadequate oxygen, for example, in a muscle cell undergoing vigorous contraction, the pyruvate produced by glycolysis is converted to lactate as shown. This reaction regenerates the NAD<sup>+</sup> consumed in step 6 of glycolysis, but the whole pathway yields much less energy overall than complete oxidation. (B) In some organisms that can grow anaerobically, such as yeasts, pyruvate is converted via acetaldehyde into carbon dioxide and ethanol. Again, this pathway regenerates NAD<sup>+</sup> from NADH, as required to enable glycolysis to continue. Both (A) and (B) are examples of *fermentations*.

**fermentations.** Studies of the commercially important fermentations carried out by yeasts inspired much of early biochemistry. Work in the nineteenth century led in 1896 to the then startling recognition that these processes could be studied outside living organisms, in cell extracts. This revolutionary discovery eventually made it possible to dissect out and study each of the individual reactions in the fermentation process. The piecing together of the complete glycolytic pathway in the 1930s was a major triumph of biochemistry, and it was quickly followed by the recognition of the central role of ATP in cell processes.

### Glycolysis Illustrates How Enzymes Couple Oxidation to Energy Storage

The formation of ATP during glycolysis provides a particularly clear demonstration of how enzymes couple energetically unfavorable reactions with favorable ones, thereby driving the many chemical reactions that make life possible. Two central reactions in glycolysis (steps 6 and 7) convert the three-carbon sugar intermediate glyceraldehyde 3-phosphate (an aldehyde) into 3-phosphoglycerate (a carboxylic acid; see Panel 2-8, pp. 104–105), thus oxidizing an aldehyde group to a carboxylic acid group. The overall reaction releases enough free energy to convert a molecule of ADP to ATP and to transfer two electrons (and a proton) from the aldehyde to NAD<sup>+</sup> to form NADH, while still liberating enough heat to the environment to make the overall reaction energetically favorable ( $\Delta G^\circ$  for the overall reaction is  $-12.5\text{ kJ/mole}$ ).

Figure 2-48 outlines this remarkable feat of energy harvesting. The chemical reactions are precisely guided by two enzymes to which the sugar intermediates



**Figure 2–48 Energy storage in steps 6 and 7 of glycolysis.** (A) In step 6, the enzyme glyceraldehyde 3-phosphate dehydrogenase couples the energetically favorable oxidation of an aldehyde to the energetically unfavorable formation of a high-energy phosphate bond. At the same time, it enables energy to be stored in NADH. The formation of the high-energy phosphate bond is driven by the oxidation reaction, and the enzyme thereby acts like the “paddle wheel” coupler in Figure 2–32B. In step 7, the newly formed high-energy phosphate bond in 1,3-bisphosphoglycerate is transferred to ADP, forming a molecule of ATP and leaving a free carboxylic acid group on the oxidized sugar. The part of the molecule that undergoes a change is shaded in blue; the rest of the molecule remains unchanged throughout all these reactions. (B) Summary of the overall chemical change produced by reactions 6 and 7.

are tightly bound. As detailed in Figure 2–48, the first enzyme (glyceraldehyde 3-phosphate dehydrogenase) forms a short-lived covalent bond to the aldehyde through a reactive -SH group on the enzyme, and catalyzes its oxidation by NAD<sup>+</sup> in this attached state. The reactive enzyme–substrate bond is then displaced by an inorganic phosphate ion to produce a high-energy phosphate intermediate, which is released from the enzyme. This intermediate binds to the second enzyme (phosphoglycerate kinase), which catalyzes the energetically favorable transfer of the high-energy phosphate just created to ADP, forming ATP and completing the process of oxidizing an aldehyde to a carboxylic acid. Note that the C–H bond oxidation energy in step 6 drives the formation of both NADH and a high-energy phosphate bond. The breakage of the high-energy bond then drives ATP formation.

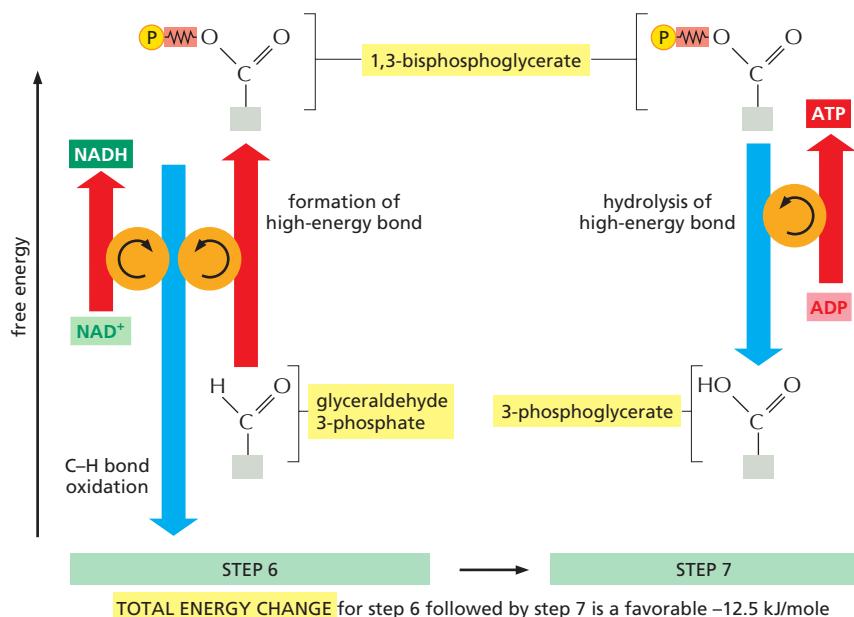
We have shown this particular oxidation process in some detail because it provides a clear example of enzyme-mediated energy storage through coupled reactions (Figure 2–49). Steps 6 and 7 are the only reactions in glycolysis that create a high-energy phosphate linkage directly from inorganic phosphate. As such, they account for the net yield of two ATP molecules and two NADH molecules per molecule of glucose (see Panel 2–8, pp. 104–105).

As we have just seen, ATP can be formed readily from ADP when a reaction intermediate is formed with a phosphate bond of higher energy than the terminal phosphate bond in ATP. Phosphate bonds can be ordered in energy by comparing the standard free-energy change ( $\Delta G^\circ$ ) for the breakage of each bond by hydrolysis. Figure 2–50 compares the high-energy phosphoanhydride bonds in ATP with the energy of some other phosphate bonds, several of which are generated during glycolysis.

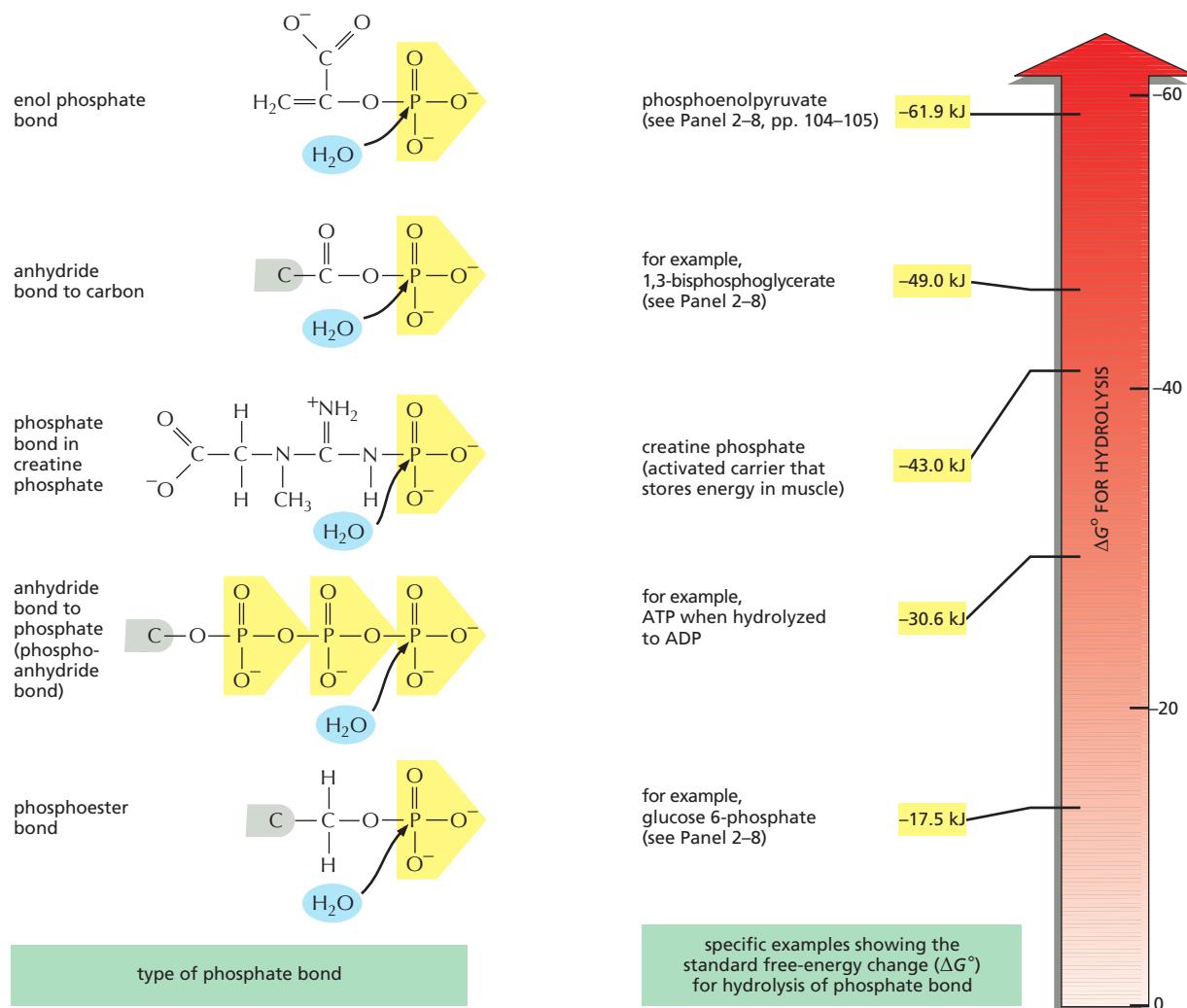
### Organisms Store Food Molecules in Special Reservoirs

All organisms need to maintain a high ATP/ADP ratio to maintain biological order in their cells. Yet animals have only periodic access to food, and plants need to survive overnight without sunlight, when they are unable to produce sugar from photosynthesis. For this reason, both plants and animals convert sugars and fats to special forms for storage (Figure 2–51).

To compensate for long periods of fasting, animals store fatty acids as fat droplets composed of water-insoluble *triacylglycerols* (also called triglycerides). The triacylglycerols in animals are mostly stored in the cytoplasm of specialized fat cells called adipocytes. For shorter-term storage, sugar is stored as glucose



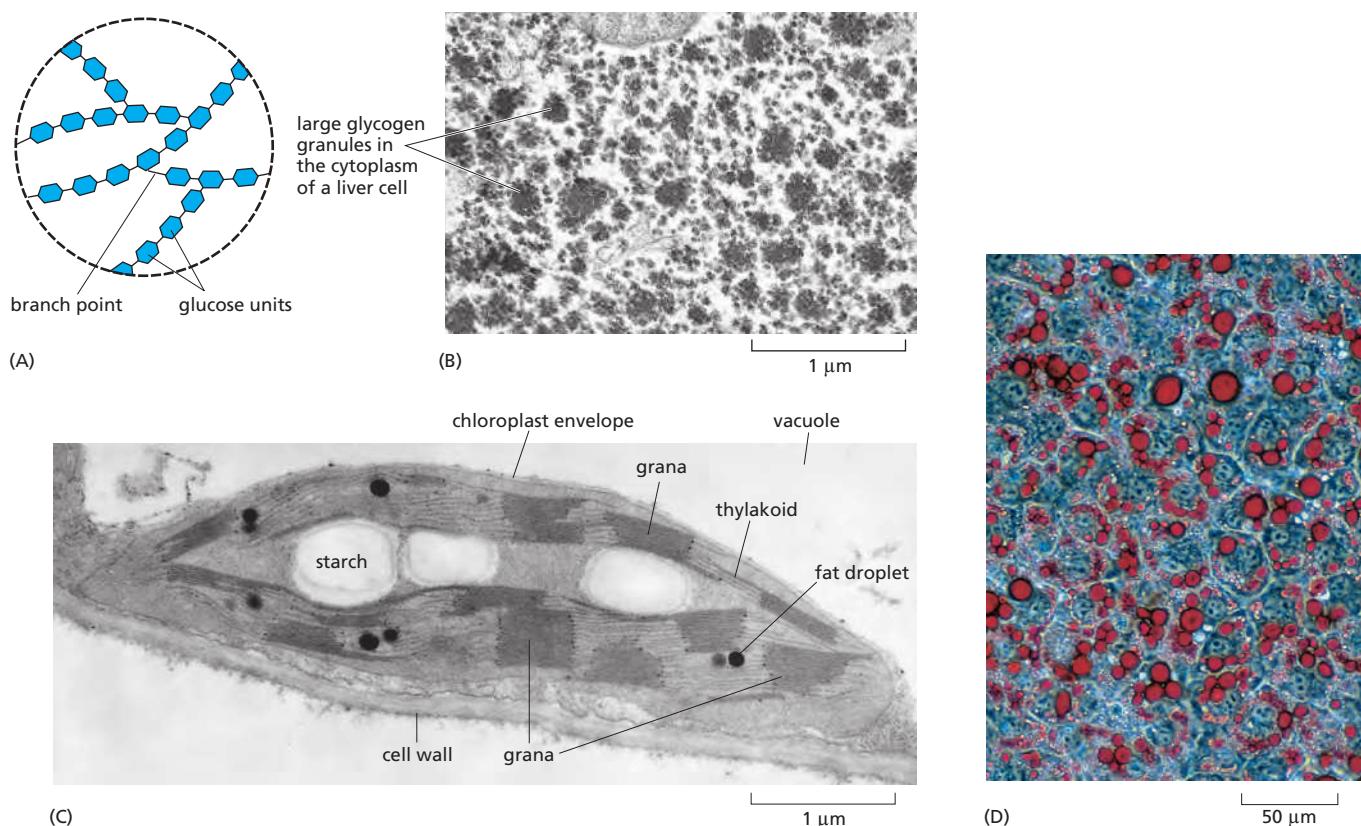
**Figure 2–49** Schematic view of the coupled reactions that form NADH and ATP in steps 6 and 7 of glycolysis. The C–H bond oxidation energy drives the formation of both NADH and a high-energy phosphate bond. The breakage of the high-energy bond then drives ATP formation.



**Figure 2-50 Phosphate bonds have different energies.** Examples of different types of phosphate bonds with their sites of hydrolysis are shown in the molecules depicted on the left. Those starting with a gray carbon atom show only part of a molecule. Examples of molecules containing such bonds are given on the right, with the standard free-energy change for hydrolysis in kilojoules. The transfer of a phosphate group from one molecule to another is energetically favorable if the free-energy change ( $\Delta G$ ) for hydrolysis of the phosphate bond of the first molecule is more negative than that for hydrolysis of the phosphate bond in the second. Thus, under standard conditions, a phosphate group is readily transferred from 1,3-bisphosphoglycerate to ADP to form ATP. (Standard conditions often do not pertain to living cells, where the relative concentrations of reactants and products will influence the actual change in free energy.) The hydrolysis reaction can be viewed as the transfer of the phosphate group to water.

subunits in the large branched polysaccharide **glycogen**, which is present as small granules in the cytoplasm of many cells, including liver and muscle. The synthesis and degradation of glycogen are rapidly regulated according to need. When cells need more ATP than they can generate from the food molecules taken in from the bloodstream, they break down glycogen in a reaction that produces glucose 1-phosphate, which is rapidly converted to glucose 6-phosphate for glycolysis (Figure 2-52).

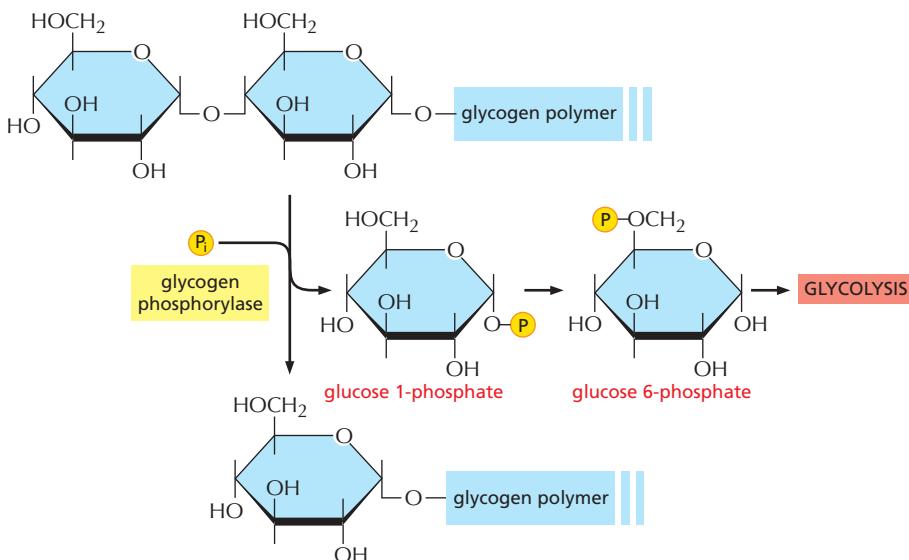
Quantitatively, **fat** is far more important than glycogen as an energy store for animals, presumably because it provides for more efficient storage. The oxidation of a gram of fat releases about twice as much energy as the oxidation of a gram of glycogen. Moreover, glycogen differs from fat in binding a great deal of water, producing a sixfold difference in the actual mass of glycogen required to store the same amount of energy as fat. An average adult human stores enough glycogen



**Figure 2–51** The storage of sugars and fats in animal and plant cells. (A) The structures of starch and glycogen, the storage form of sugars in plants and animals, respectively. Both are storage polymers of the sugar glucose and differ only in the frequency of branch points. There are many more branches in glycogen than in starch. (B) An electron micrograph of glycogen granules in the cytoplasm of a liver cell. (C) A thin section of a chloroplast from a plant cell, showing the starch granules and lipid (fat droplets) that have accumulated as a result of the biosyntheses occurring there. (D) Fat droplets (stained red) beginning to accumulate in developing fat cells of an animal. (B, courtesy of Robert Fletterick and Daniel S. Friend; C, courtesy of K. Plaskitt; D, courtesy of Ronald M. Evans and Peter Totonoz.)

for only about a day of normal activities, but enough fat to last for nearly a month. If our main fuel reservoir had to be carried as glycogen instead of fat, body weight would increase by an average of about 60 pounds.

The sugar and ATP needed by plant cells are largely produced in separate organelles: sugars in chloroplasts (the organelles specialized for photosynthesis),



**Figure 2–52** How sugars are produced from glycogen. Glucose subunits are released from glycogen by the enzyme glycogen phosphorylase. This produces glucose 1-phosphate, which is rapidly converted to glucose 6-phosphate for glycolysis.



**Figure 2–53** Some plant seeds that serve as important foods for humans. Corn, nuts, and peas all contain rich stores of starch and fat that provide the young plant embryo in the seed with energy and building blocks for biosynthesis. (Courtesy of the John Innes Foundation.)

and ATP in mitochondria. Although plants produce abundant amounts of both ATP and NADPH in their chloroplasts, this organelle is isolated from the rest of its plant cell by a membrane that is impermeable to both types of activated carrier molecules. Moreover, the plant contains many cells—such as those in the roots—that lack chloroplasts and therefore cannot produce their own sugars. Thus, sugars are exported from chloroplasts to the mitochondria present in all cells of the plant. Most of the ATP needed for general plant cell metabolism is synthesized in these mitochondria, using exactly the same pathways for the oxidative breakdown of sugars as in nonphotosynthetic organisms; this ATP is then passed to the rest of the cell (see Figure 14–42).

During periods of excess photosynthetic capacity during the day, chloroplasts convert some of the sugars that they make into fats and into **starch**, a polymer of glucose analogous to the glycogen of animals. The fats in plants are triacyl-glycerols (triglycerides), just like the fats in animals, and differ only in the types of fatty acids that predominate. Fat and starch are both stored inside the chloroplast until needed for energy-yielding oxidation during periods of darkness (see Figure 2–51C).

The embryos inside plant seeds must live on stored sources of energy for a prolonged period, until they germinate and produce leaves that can harvest the energy in sunlight. For this reason plant seeds often contain especially large amounts of fats and starch—which makes them a major food source for animals, including ourselves (**Figure 2–53**).

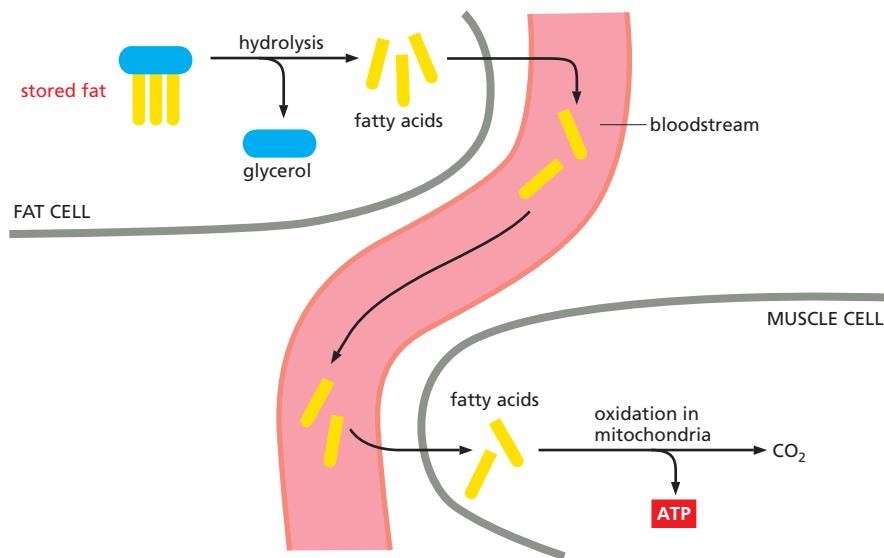
### Most Animal Cells Derive Their Energy from Fatty Acids Between Meals

After a meal, most of the energy that an animal needs is derived from sugars obtained from food. Excess sugars, if any, are used to replenish depleted glycogen stores, or to synthesize fats as a food store. But soon the fat stored in adipose tissue is called into play, and by the morning after an overnight fast, fatty acid oxidation generates most of the ATP we need.

Low glucose levels in the blood trigger the breakdown of fats for energy production. As illustrated in **Figure 2–54**, the triacylglycerols stored in fat droplets in adipocytes are hydrolyzed to produce fatty acids and glycerol, and the fatty acids released are transferred to cells in the body through the bloodstream. While animals readily convert sugars to fats, they cannot convert fatty acids to sugars. Instead, the fatty acids are oxidized directly.

### Sugars and Fats Are Both Degraded to Acetyl CoA in Mitochondria

In aerobic metabolism, the pyruvate that was produced by glycolysis from sugars in the cytosol is transported into the *mitochondria* of eukaryotic cells. There, it is



**Figure 2–54** How stored fats are mobilized for energy production in animals. Low glucose levels in the blood trigger the hydrolysis of the triacylglycerol molecules in fat droplets to free fatty acids and glycerol. These fatty acids enter the bloodstream, where they bind to the abundant blood protein, serum albumin. Special fatty acid transporters in the plasma membrane of cells that oxidize fatty acids, such as muscle cells, then pass these fatty acids into the cytosol, from which they are moved into mitochondria for energy production.

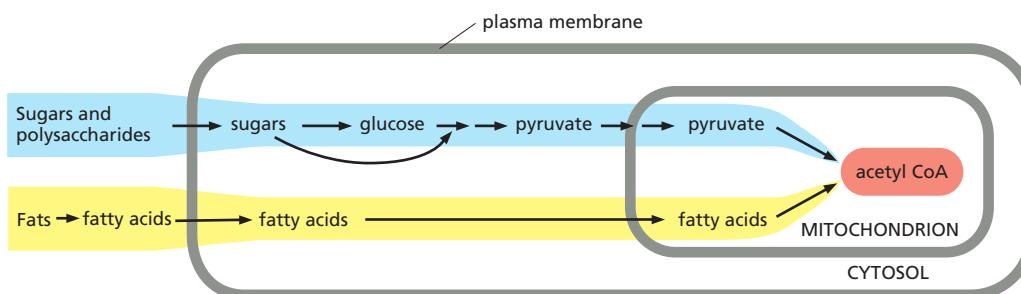
rapidly decarboxylated by a giant complex of three enzymes, called the *pyruvate dehydrogenase complex*. The products of pyruvate decarboxylation are a molecule of  $\text{CO}_2$  (a waste product), a molecule of NADH, and acetyl CoA (see Panel 2–9).

The fatty acids imported from the bloodstream are moved into mitochondria, where all of their oxidation takes place (Figure 2–55). Each molecule of fatty acid (as the activated molecule *fatty acyl CoA*) is broken down completely by a cycle of reactions that trims two carbons at a time from its carboxyl end, generating one molecule of acetyl CoA for each turn of the cycle. A molecule of NADH and a molecule of  $\text{FADH}_2$  are also produced in this process (Figure 2–56).

Sugars and fats are the major energy sources for most nonphotosynthetic organisms, including humans. However, most of the useful energy that can be extracted from the oxidation of both types of foodstuffs remains stored in the acetyl CoA molecules that are produced by the two types of reactions just described. The citric acid cycle of reactions, in which the acetyl group ( $-\text{COCH}_3$ ) in acetyl CoA is oxidized to  $\text{CO}_2$  and  $\text{H}_2\text{O}$ , is therefore central to the energy metabolism of aerobic organisms. In eukaryotes, these reactions all take place in mitochondria. We should therefore not be surprised to discover that the mitochondrion is the place where most of the ATP is produced in animal cells. In contrast, aerobic bacteria carry out all of their reactions, including the citric acid cycle, in a single compartment, the cytosol.

### The Citric Acid Cycle Generates NADH by Oxidizing Acetyl Groups to $\text{CO}_2$

In the nineteenth century, biologists noticed that in the absence of air cells produce lactic acid (for example, in muscle) or ethanol (for example, in yeast), while in its presence they consume  $\text{O}_2$  and produce  $\text{CO}_2$  and  $\text{H}_2\text{O}$ . Efforts to define the pathways of aerobic metabolism eventually focused on the oxidation of pyruvate and led in 1937 to the discovery of the **citric acid cycle**, also known as the



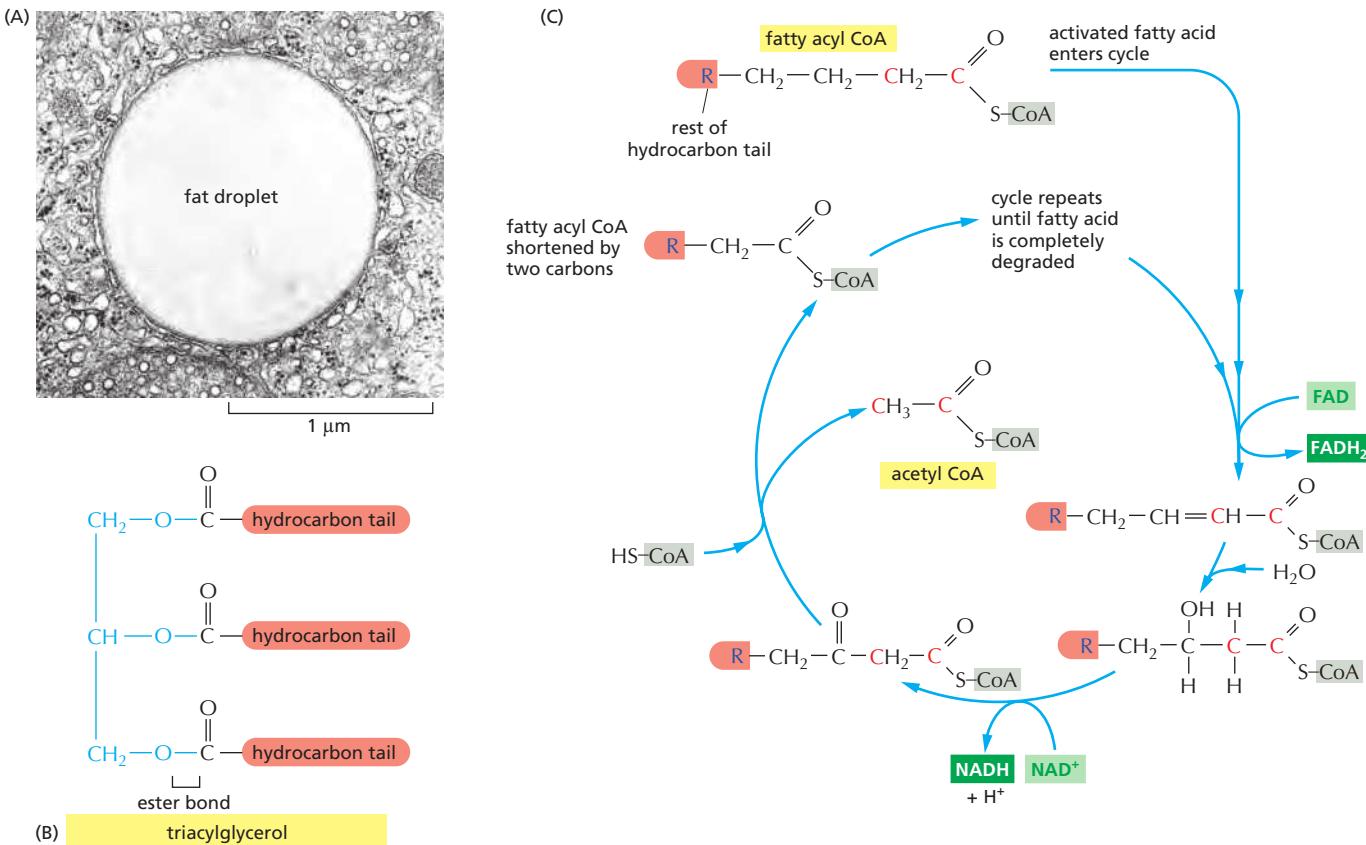
**Figure 2–55** Pathways for the production of acetyl CoA from sugars and fats. The mitochondrion in eukaryotic cells is where acetyl CoA is produced from both types of major food molecules. It is therefore the place where most of the cell's oxidation reactions occur and where most of its ATP is made. Amino acids (not shown) can also enter the mitochondria, to be converted there into acetyl CoA or another intermediate of the citric acid cycle. The structure and function of mitochondria are discussed in detail in Chapter 14.

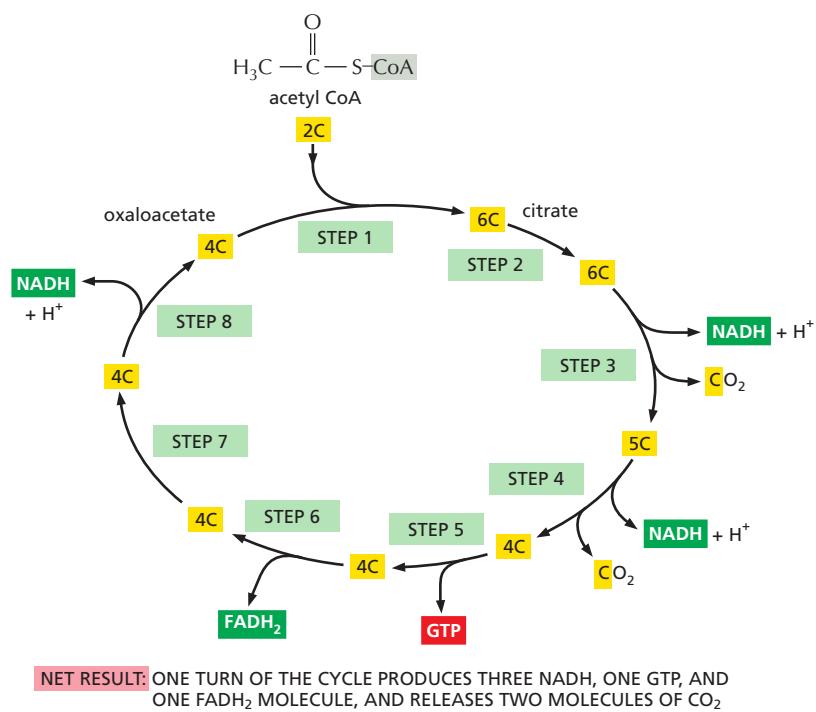
*tricarboxylic acid cycle* or the *Krebs cycle*. The citric acid cycle accounts for about two-thirds of the total oxidation of carbon compounds in most cells, and its major end products are  $\text{CO}_2$  and high-energy electrons in the form of NADH. The  $\text{CO}_2$  is released as a waste product, while the high-energy electrons from NADH are passed to a membrane-bound electron-transport chain (discussed in Chapter 14), eventually combining with  $\text{O}_2$  to produce  $\text{H}_2\text{O}$ . The citric acid cycle itself does not use gaseous  $\text{O}_2$  (it uses oxygen atoms from  $\text{H}_2\text{O}$ ). But the cycle does require  $\text{O}_2$  in subsequent reactions to keep it going. This is because there is no other efficient way for the NADH to get rid of its electrons and thus regenerate the  $\text{NAD}^+$  that is needed.

The citric acid cycle takes place inside mitochondria in eukaryotic cells. It results in the complete oxidation of the carbon atoms of the acetyl groups in acetyl CoA, converting them into  $\text{CO}_2$ . But the acetyl group is not oxidized directly. Instead, this group is transferred from acetyl CoA to a larger, four-carbon molecule, *oxaloacetate*, to form the six-carbon tricarboxylic acid, *citric acid*, for which the subsequent cycle of reactions is named. The citric acid molecule is then gradually oxidized, allowing the energy of this oxidation to be harnessed to produce energy-rich activated carrier molecules. The chain of eight reactions forms a cycle because at the end the oxaloacetate is regenerated and enters a new turn of the cycle, as shown in outline in **Figure 2–57**.

We have thus far discussed only one of the three types of activated carrier molecules that are produced by the citric acid cycle; NADH, the reduced form of the  $\text{NAD}^+/\text{NADH}$  electron carrier system (see **Figure 2–36**). In addition to three molecules of NADH, each turn of the cycle also produces one molecule of  $\text{FADH}_2$  (reduced flavin adenine dinucleotide) from FAD (see **Figure 2–39**), and one molecule of the ribonucleoside triphosphate GTP from GDP. The structure of GTP is illustrated in **Figure 2–58**. GTP is a close relative of ATP, and the transfer of its terminal phosphate group to ADP produces one ATP molecule in each cycle. As we discuss shortly, the energy that is stored in the readily transferred electrons of NADH and  $\text{FADH}_2$  will be utilized subsequently for ATP production through the

**Figure 2–56** The oxidation of fatty acids to acetyl CoA. (A) Electron micrograph of a lipid droplet in the cytoplasm. (B) The structure of fats. Fats are *triacylglycerols*. The glycerol portion, to which three fatty acids are linked through ester bonds, is shown in blue. Fats are insoluble in water and form large lipid droplets in the specialized fat cells (adipocytes) in which they are stored. (C) The fatty acid oxidation cycle. The cycle is catalyzed by a series of four enzymes in mitochondria. Each turn of the cycle shortens the fatty acid chain by two carbons (shown in red) and generates one molecule of acetyl CoA and one molecule each of NADH and  $\text{FADH}_2$ . (A, courtesy of Daniel S. Friend.)





**Figure 2–57** Simple overview of the citric acid cycle. The reaction of acetyl CoA with oxaloacetate starts the cycle by producing citrate (citric acid). In each turn of the cycle, two molecules of CO<sub>2</sub> are produced as waste products, plus three molecules of NADH, one molecule of GTP, and one molecule of FADH<sub>2</sub>. The number of carbon atoms in each intermediate is shown in a yellow box. For details, see Panel 2–9 (pp. 106–107).

process of *oxidative phosphorylation*, the only step in the oxidative catabolism of foodstuffs that directly requires gaseous oxygen (O<sub>2</sub>) from the atmosphere.

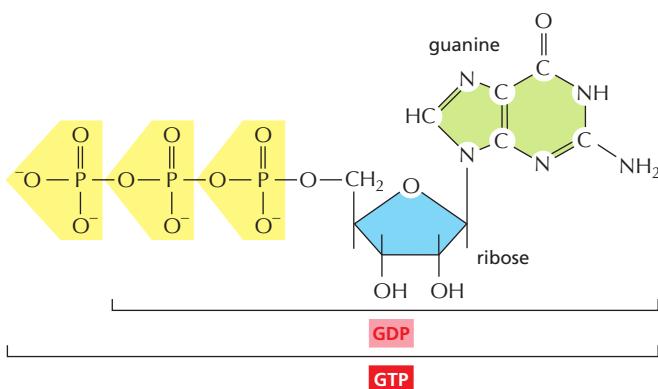
Panel 2–9 (pp. 106–107) and Movie 2.6 present the complete citric acid cycle. Water, rather than molecular oxygen, supplies the extra oxygen atoms required to make CO<sub>2</sub> from the acetyl groups entering the citric acid cycle. As illustrated in the panel, three molecules of water are split in each cycle, and the oxygen atoms of some of them are ultimately used to make CO<sub>2</sub>.

In addition to pyruvate and fatty acids, some amino acids pass from the cytosol into mitochondria, where they are also converted into acetyl CoA or one of the other intermediates of the citric acid cycle. Thus, in the eukaryotic cell, the mitochondrion is the center toward which all energy-yielding processes lead, whether they begin with sugars, fats, or proteins.

Both the citric acid cycle and glycolysis also function as starting points for important biosynthetic reactions by producing vital carbon-containing intermediates, such as *oxaloacetate* and *α-ketoglutarate*. Some of these substances produced by catabolism are transferred back from the mitochondria to the cytosol, where they serve in anabolic reactions as precursors for the synthesis of many essential molecules, such as amino acids (Figure 2–59).

### Electron Transport Drives the Synthesis of the Majority of the ATP in Most Cells

Most chemical energy is released in the last stage in the degradation of a food molecule. In this final process, NADH and FADH<sub>2</sub> transfer the electrons that they gained when oxidizing food-derived organic molecules to the **electron-transport chain**, which is embedded in the inner membrane of the mitochondrion (see Figure 14–10). As the electrons pass along this long chain of specialized electron acceptor and donor molecules, they fall to successively lower energy states. The energy that the electrons release in this process pumps H<sup>+</sup> ions (protons) across the membrane—from the innermost mitochondrial compartment (the matrix) to the intermembrane space (and then to the cytosol)—generating a gradient of H<sup>+</sup> ions (Figure 2–60). This gradient serves as a major source of energy for cells, being tapped like a battery to drive a variety of energy-requiring reactions. The most prominent of these reactions is the generation of ATP by the phosphorylation of ADP.



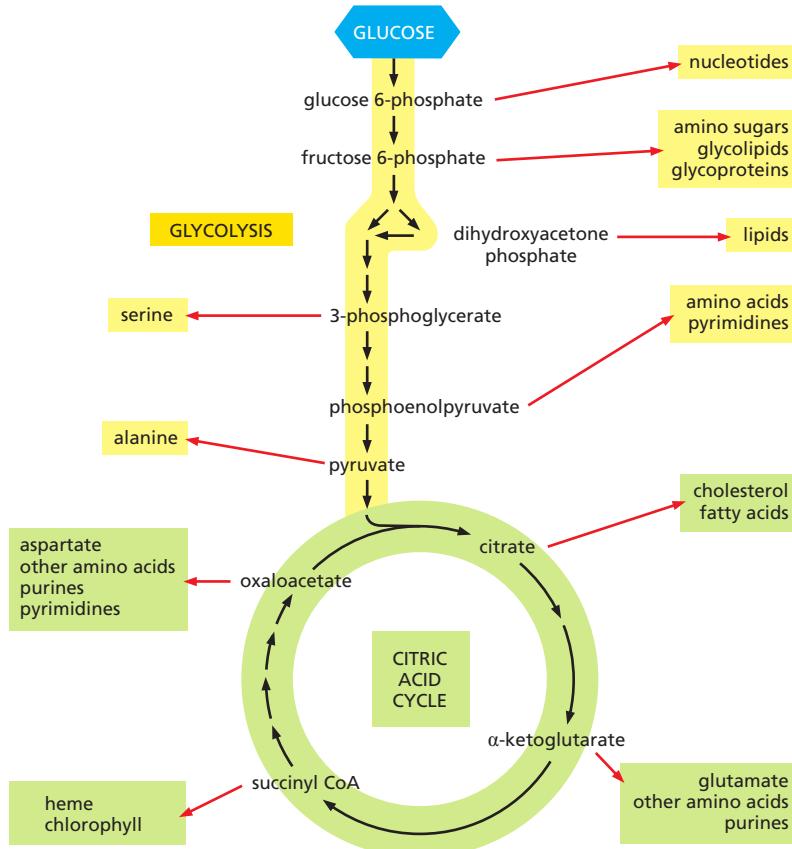
**Figure 2–58** The structure of GTP. GTP and GDP are close relatives of ATP and ADP, respectively.

At the end of this series of electron transfers, the electrons are passed to molecules of oxygen gas ( $O_2$ ) that have diffused into the mitochondrion, which simultaneously combine with protons ( $H^+$ ) from the surrounding solution to produce water. The electrons have now reached a low energy level, and all the available energy has been extracted from the oxidized food molecule. This process, termed **oxidative phosphorylation** (Figure 2–61), also occurs in the plasma membrane of bacteria. As one of the most remarkable achievements of cell evolution, it is a central topic of Chapter 14.

In total, the complete oxidation of a molecule of glucose to  $H_2O$  and  $CO_2$  is used by the cell to produce about 30 molecules of ATP. In contrast, only 2 molecules of ATP are produced per molecule of glucose by glycolysis alone.

### Amino Acids and Nucleotides Are Part of the Nitrogen Cycle

So far we have concentrated mainly on carbohydrate metabolism and have not yet considered the metabolism of nitrogen or sulfur. These two elements are important constituents of biological macromolecules. Nitrogen and sulfur atoms pass



**Figure 2–59** Glycolysis and the citric acid cycle provide the precursors needed to synthesize many important biological molecules. The amino acids, nucleotides, lipids, sugars, and other molecules—shown here as products—in turn serve as the precursors for the many macromolecules of the cell. Each black arrow in this diagram denotes a single enzyme-catalyzed reaction; the red arrows generally represent pathways with many steps that are required to produce the indicated products.

from compound to compound and between organisms and their environment in a series of reversible cycles.

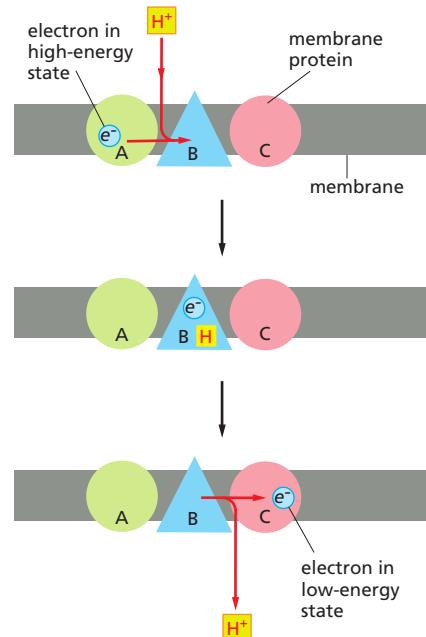
Although molecular nitrogen is abundant in the Earth's atmosphere, nitrogen is chemically unreactive as a gas. Only a few living species are able to incorporate it into organic molecules, a process called **nitrogen fixation**. Nitrogen fixation occurs in certain microorganisms and by some geophysical processes, such as lightning discharge. It is essential to the biosphere as a whole, for without it life could not exist on this planet. Only a small fraction of the nitrogenous compounds in today's organisms, however, is due to fresh products of nitrogen fixation from the atmosphere. Most organic nitrogen has been in circulation for some time, passing from one living organism to another. Thus, present-day nitrogen-fixing reactions can be said to perform a "topping-up" function for the total nitrogen supply.

Vertebrates receive virtually all of their nitrogen from their dietary intake of proteins and nucleic acids. In the body, these macromolecules are broken down to amino acids and the components of nucleotides, and the nitrogen they contain is used to produce new proteins and nucleic acids—or other molecules. About half of the 20 amino acids found in proteins are essential amino acids for vertebrates (Figure 2–62), which means that they cannot be synthesized from other ingredients of the diet. The other amino acids can be so synthesized, using a variety of raw materials, including intermediates of the citric acid cycle. The essential amino acids are made by plants and other organisms, usually by long and energetically expensive pathways that have been lost in the course of vertebrate evolution.

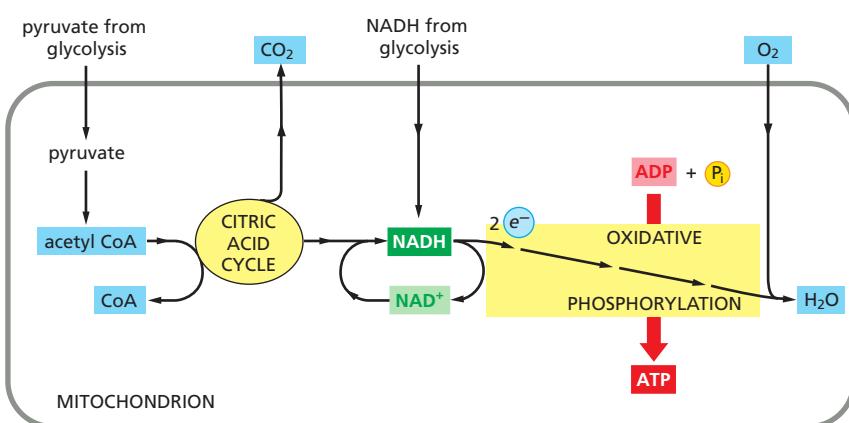
The nucleotides needed to make RNA and DNA can be synthesized using specialized biosynthetic pathways. All of the nitrogens in the purine and pyrimidine bases (as well as some of the carbons) are derived from the plentiful amino acids glutamine, aspartic acid, and glycine, whereas the ribose and deoxyribose sugars are derived from glucose. There are no "essential nucleotides" that must be provided in the diet.

Amino acids not used in biosynthesis can be oxidized to generate metabolic energy. Most of their carbon and hydrogen atoms eventually form  $\text{CO}_2$  or  $\text{H}_2\text{O}$ , whereas their nitrogen atoms are shuttled through various forms and eventually appear as urea, which is excreted. Each amino acid is processed differently, and a whole constellation of enzymatic reactions exists for their catabolism.

Sulfur is abundant on Earth in its most oxidized form, sulfate ( $\text{SO}_4^{2-}$ ). To be useful for life, sulfate must be reduced to sulfide ( $\text{S}^{2-}$ ), the oxidation state of sulfur required for the synthesis of essential biological molecules, including the amino acids methionine and cysteine, coenzyme A (see Figure 2–39), and the iron-sulfur centers essential for electron transport (see Figure 14–16). The sulfur-reduction process begins in bacteria, fungi, and plants, where a special group of enzymes use ATP and reducing power to create a sulfate assimilation pathway. Humans and other animals cannot reduce sulfate and must therefore acquire the sulfur they need for their metabolism in the food that they eat.



**Figure 2–60** The generation of an  $\text{H}^+$  gradient across a membrane by electron-transport reactions. An electron held in a high-energy state (derived, for example, from the oxidation of a metabolite) is passed sequentially by carriers A, B, and C to a lower energy state. In this diagram, carrier B is arranged in the membrane in such a way that it takes up  $\text{H}^+$  from one side and releases it to the other as the electron passes. The result is an  $\text{H}^+$  gradient. As discussed in Chapter 14, this gradient is an important form of energy that is harnessed by other membrane proteins to drive the formation of ATP (for an actual example, see Figure 14–21).



**Figure 2–61** The final stages of oxidation of food molecules. Molecules of NADH and FADH<sub>2</sub> (FADH<sub>2</sub> is not shown) are produced by the citric acid cycle. These activated carriers donate high-energy electrons that are eventually used to reduce oxygen gas to water. A major portion of the energy released during the transfer of these electrons along an electron-transfer chain in the mitochondrial inner membrane (or in the plasma membrane of bacteria) is harnessed to drive the synthesis of ATP—hence the name oxidative phosphorylation (discussed in Chapter 14).

## Metabolism Is Highly Organized and Regulated

One gets a sense of the intricacy of a cell as a chemical machine from the relation of glycolysis and the citric acid cycle to the other metabolic pathways sketched out in **Figure 2–63**. This chart represents only some of the enzymatic pathways in a human cell. It is obvious that our discussion of cell metabolism has dealt with only a tiny fraction of the broad field of cell chemistry.

All these reactions occur in a cell that is less than 0.1 mm in diameter, and each requires a different enzyme. As is clear from Figure 2–63, the same molecule can often be part of many different pathways. Pyruvate, for example, is a substrate for half a dozen or more different enzymes, each of which modifies it chemically in a different way. One enzyme converts pyruvate to acetyl CoA, another to oxaloacetate; a third enzyme changes pyruvate to the amino acid alanine, a fourth to lactate, and so on. All of these different pathways compete for the same pyruvate molecule, and similar competitions for thousands of other small molecules go on at the same time.

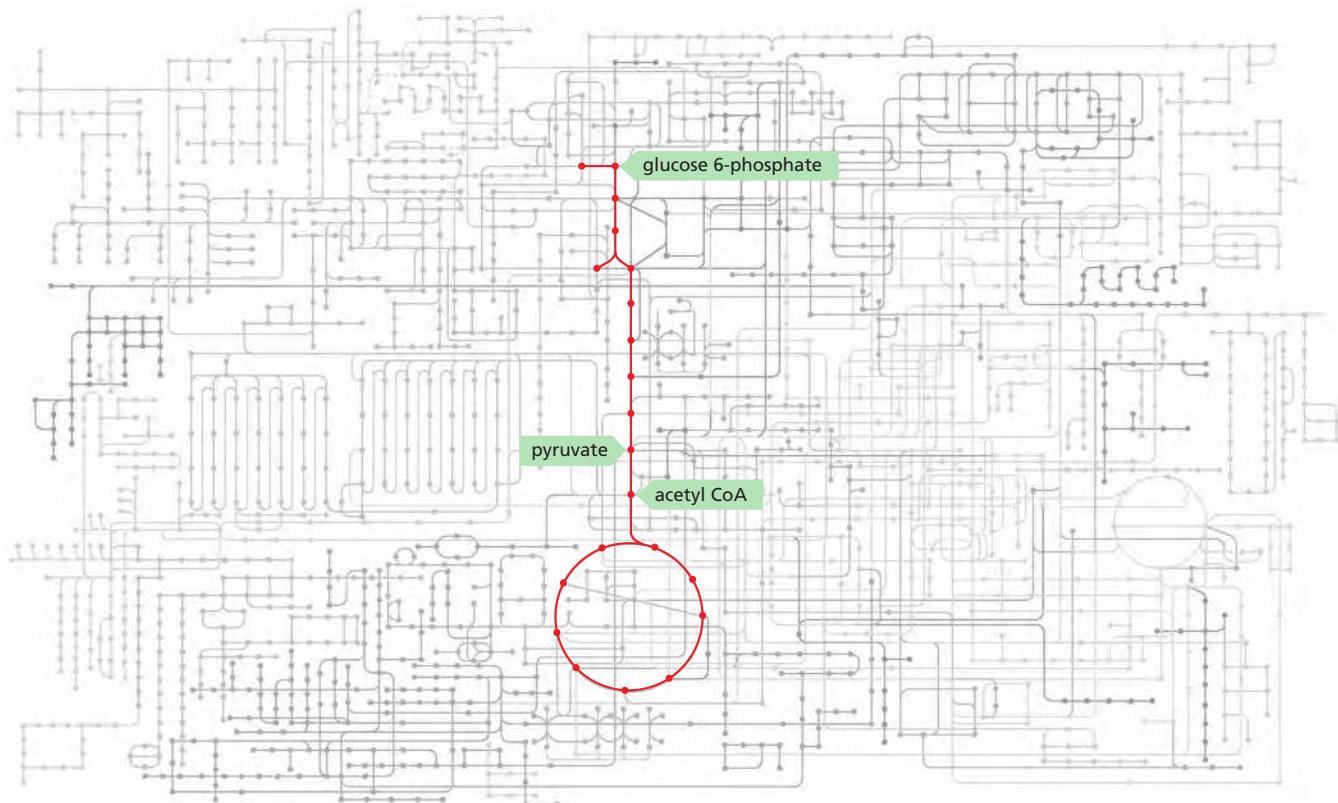
The situation is further complicated in a multicellular organism. Different cell types will in general require somewhat different sets of enzymes. And different tissues make distinct contributions to the chemistry of the organism as a whole. In addition to differences in specialized products such as hormones or antibodies, there are significant differences in the “common” metabolic pathways among various types of cells in the same organism.

Although virtually all cells contain the enzymes of glycolysis, the citric acid cycle, lipid synthesis and breakdown, and amino acid metabolism, the levels of these processes required in different tissues are not the same. For example, nerve cells, which are probably the most fastidious cells in the body, maintain almost no reserves of glycogen or fatty acids and rely almost entirely on a constant

## THE ESSENTIAL AMINO ACIDS

THREONINE
METHIONINE
LYSINE
VALINE
LEUCINE
ISOLEUCINE
HISTIDINE
PHENYLALANINE
TRYPTOPHAN

**Figure 2–62** The nine essential amino acids. These cannot be synthesized by human cells and so must be supplied in the diet.



**Figure 2–63** Glycolysis and the citric acid cycle are at the center of an elaborate set of metabolic pathways in human cells. Some 2000 metabolic reactions are shown schematically with the reactions of glycolysis and the citric acid cycle in red. Many other reactions either lead into these two central pathways—delivering small molecules to be catabolized with production of energy—or they lead outward and thereby supply carbon compounds for the purpose of biosynthesis. (Adapted with permission from Kanehisa Laboratories.)

supply of glucose from the bloodstream. In contrast, liver cells supply glucose to actively contracting muscle cells and recycle the lactic acid produced by muscle cells back into glucose. All types of cells have their distinctive metabolic traits, and they cooperate extensively in the normal state, as well as in response to stress and starvation. One might think that the whole system would need to be so finely balanced that any minor upset, such as a temporary change in dietary intake, would be disastrous.

In fact, the metabolic balance of a cell is amazingly stable. Whenever the balance is perturbed, the cell reacts so as to restore the initial state. The cell can adapt and continue to function during starvation or disease. Mutations of many kinds can damage or even eliminate particular reaction pathways, and yet—provided that certain minimum requirements are met—the cell survives. It does so because an elaborate network of *control mechanisms* regulates and coordinates the rates of all of its reactions. These controls rest, ultimately, on the remarkable abilities of proteins to change their shape and their chemistry in response to changes in their immediate environment. The principles that underlie how large molecules such as proteins are built and the chemistry behind their regulation will be our next concern.

### Summary

*Glucose and other food molecules are broken down by controlled stepwise oxidation to provide chemical energy in the form of ATP and NADH. There are three main sets of reactions that act in series, the products of each being the starting material for the next: glycolysis (which occurs in the cytosol), the citric acid cycle (in the mitochondrial matrix), and oxidative phosphorylation (on the inner mitochondrial membrane). The intermediate products of glycolysis and the citric acid cycle are used both as sources of metabolic energy and to produce many of the small molecules used as the raw materials for biosynthesis. Cells store sugar molecules as glycogen in animals and starch in plants; both plants and animals also use fats extensively as a food store. These storage materials in turn serve as a major source of food for humans, along with the proteins that comprise the majority of the dry mass of most of the cells in the foods we eat.*

### WHAT WE DON'T KNOW

- Did chemiosmosis precede fermentation as the source of biological energy, or did some form of fermentation come first, as had been assumed for many years?
- What is the minimum number of components required to make a living cell from scratch? How might we find out?
- Are other life chemistries possible besides the single one known on Earth (and described in this chapter)? When screening for life on other planets, what type of chemical signatures should we search for?
- Is the shared chemistry inside all living cells a clue for deciphering the environment on Earth where the first cells originated? For example, what might we conclude from the universally shared high  $K^+/Na^+$  ratio, neutral pH, and central role of phosphates?

## PROBLEMS

Which statements are true? Explain why or why not.

**2–1** A  $10^{-8}$  M solution of HCl has a pH of 8.

**2–2** Most of the interactions between macromolecules could be mediated just as well by covalent bonds as by noncovalent bonds.

**2–3** Animals and plants use oxidation to extract energy from food molecules.

**2–4** If an oxidation occurs in a reaction, it must be accompanied by a reduction.

**2–5** Linking the energetically unfavorable reaction A → B to a second, favorable reaction B → C will shift the equilibrium constant for the first reaction.

**2–6** The criterion for whether a reaction proceeds spontaneously is  $\Delta G$  not  $\Delta G^\circ$ , because  $\Delta G$  takes into account the concentrations of the substrates and products.

**2–7** The oxygen consumed during the oxidation of glucose in animal cells is returned as  $CO_2$  to the atmosphere.

Discuss the following problems.

**2–8** The organic chemistry of living cells is said to be special for two reasons: it occurs in an aqueous environment and it accomplishes some very complex reactions. But do you suppose it is really all that much different from the organic chemistry carried out in the top laboratories in the world? Why or why not?

**2–9** The molecular weight of ethanol ( $CH_3CH_2OH$ ) is 46 and its density is  $0.789\text{ g/cm}^3$ .

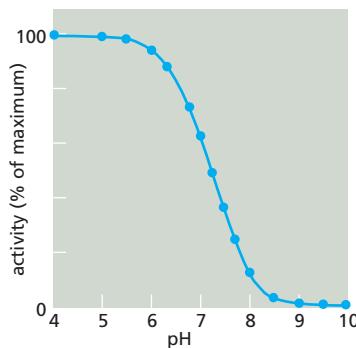
**A.** What is the molarity of ethanol in beer that is 5% ethanol by volume? [Alcohol content of beer varies from about 4% (lite beer) to 8% (stout beer).]

**B.** The legal limit for a driver's blood alcohol content varies, but 80 mg of ethanol per 100 mL of blood (usually referred to as a blood alcohol level of 0.08) is typical. What is the molarity of ethanol in a person at this legal limit?

**C.** How many 12-oz (355-mL) bottles of 5% beer could a 70-kg person drink and remain under the legal limit? A 70-kg person contains about 40 liters of water. Ignore the metabolism of ethanol, and assume that the water content of the person remains constant.

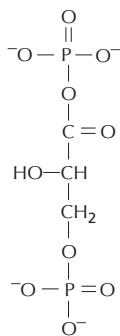
**D.** Ethanol is metabolized at a constant rate of about 120 mg per hour per kg body weight, regardless of its concentration. If a 70-kg person were at twice the legal limit (160 mg/100 mL), how long would it take for their blood alcohol level to fall below the legal limit?

**2-10** A histidine side chain is known to play an important role in the catalytic mechanism of an enzyme; however, it is not clear whether histidine is required in its protonated (charged) or unprotonated (uncharged) state. To answer this question you measure enzyme activity over a range of pH, with the results shown in **Figure Q2-1**. Which form of histidine is required for enzyme activity?



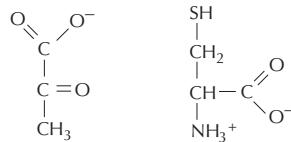
**Figure Q2-1** Enzyme activity as a function of pH (Problem 2-10).

**2-11** The three molecules in **Figure Q2-2** contain the seven most common reactive groups in biology. Most molecules in the cell are built from these functional groups. Indicate and name the functional groups in these molecules.

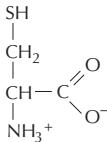


1,3-bisphosphoglycerate

**Figure Q2-2** Three molecules that illustrate the seven most common functional groups in biology (Problem 2-11). 1,3-Bisphosphoglycerate and pyruvate are intermediates in glycolysis and cysteine is an amino acid.



pyruvate



cysteine

**2-12** “Diffusion” sounds slow—and over everyday distances it is—but on the scale of a cell it is very fast. The average instantaneous velocity of a particle in solution—that is, the velocity between the very frequent collisions—is

$$v = (kT/m)^{1/2}$$

where  $k = 1.38 \times 10^{-16}$  g cm<sup>2</sup>/K sec<sup>2</sup>,  $T$  = temperature in K (37°C is 310 K), and  $m$  = mass in g/molecule.

Calculate the instantaneous velocity of a water molecule (molecular mass = 18 daltons), a glucose molecule (molecular mass = 180 daltons), and a myoglobin molecule (molecular mass = 15,000 daltons) at 37°C. Just for fun, convert these numbers into kilometers/hour.

Before you do any calculations, try to guess whether the molecules are moving at a slow crawl (<1 km/hr), an easy walk (5 km/hr), or a record-setting sprint (40 km/hr).

**2-13** Polymerization of tubulin subunits into microtubules occurs with an increase in the orderliness of the subunits. Yet tubulin polymerization occurs with an increase in entropy (decrease in order). How can that be?

**2-14** A 70-kg adult human (154 lb) could meet his or her entire energy needs for one day by eating 3 moles of glucose (540 g). (We do not recommend this.) Each molecule of glucose generates 30 molecules of ATP when it is oxidized to CO<sub>2</sub>. The concentration of ATP is maintained in cells at about 2 mM, and a 70-kg adult has about 25 liters of intracellular fluid. Given that the ATP concentration remains constant in cells, calculate how many times per day, on average, each ATP molecule in the body is hydrolyzed and resynthesized.

**2-15** Assuming that there are  $5 \times 10^{13}$  cells in the human body and that ATP is turning over at a rate of  $10^9$  ATP molecules per minute in each cell, how many watts is the human body consuming? (A watt is a joule per second.) Assume that hydrolysis of ATP yields 50 kJ/mole.

**2-16** Does a Snickers™ candy bar (65 g, 1360 kJ) provide enough energy to climb from Zermatt (elevation 1660 m) to the top of the Matterhorn (4478 m, **Figure Q2-3**), or might you need to stop at Hörnli Hut (3260 m) to eat another one? Imagine that you and your gear have a mass of 75 kg, and that all of your work is done against gravity (that is, you are just climbing straight up). Remember from your introductory physics course that

$$\text{work (J)} = \text{mass (kg)} \times g (\text{m/sec}^2) \times \text{height gained (m)}$$

where  $g$  is acceleration due to gravity (9.8 m/sec<sup>2</sup>). One joule is 1 kg m<sup>2</sup>/sec<sup>2</sup>.

What assumptions made here will greatly underestimate how much candy you need?

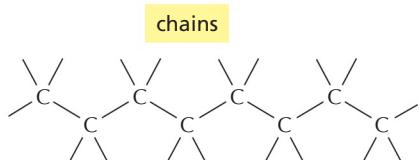


**Figure Q2-3** The Matterhorn (Problem 2-16). (Courtesy of Zermatt Tourism.)

**2-17** In the absence of oxygen, cells consume glucose at a high, steady rate. When oxygen is added, glucose consumption drops precipitously and is then maintained at the lower rate. Why is glucose consumed at a high rate in the absence of oxygen and at a low rate in its presence?

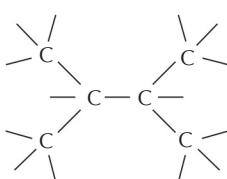
## CARBON SKELETONS

Carbon has a unique role in the cell because of its ability to form strong covalent bonds with other carbon atoms. Thus carbon atoms can join to form:



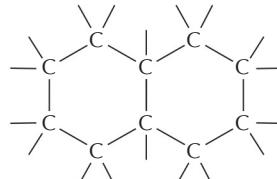
also written as

**branched trees**



also written as

**rings**

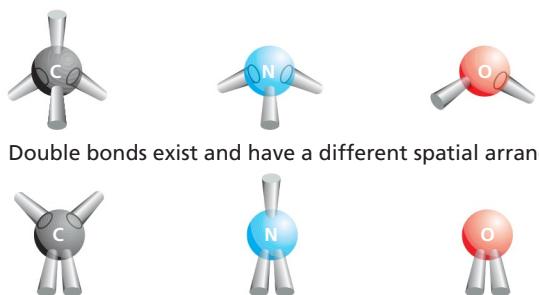


also written as

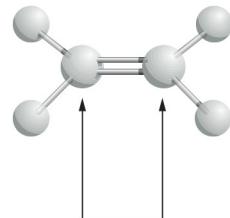
## COVALENT BONDS

A covalent bond forms when two atoms come very close together and share one or more of their electrons. In a single bond, one electron from each of the two atoms is shared; in a double bond, a total of four electrons are shared.

Each atom forms a fixed number of covalent bonds in a defined spatial arrangement. For example, carbon forms four single bonds arranged tetrahedrally, whereas nitrogen forms three single bonds and oxygen forms two single bonds arranged as shown below.



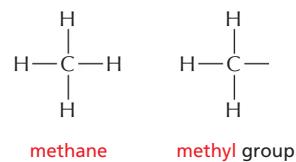
Double bonds exist and have a different spatial arrangement.



Atoms joined by two or more covalent bonds cannot rotate freely around the bond axis. This restriction is a major influence on the three-dimensional shape of many macromolecules.

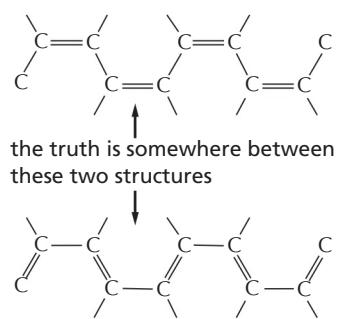
## HYDROCARBONS

Carbon and hydrogen combine together to make stable compounds (or chemical groups) called hydrocarbons. These are nonpolar, do not form hydrogen bonds, and are generally insoluble in water.

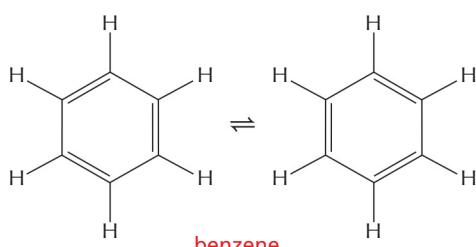


## ALTERNATING DOUBLE BONDS

The carbon chain can include double bonds. If these are on alternate carbon atoms, the bonding electrons move within the molecule, stabilizing the structure by a phenomenon called resonance.

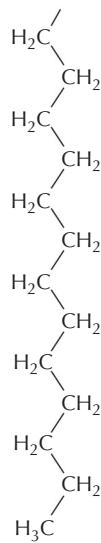


Alternating double bonds in a ring can generate a very stable structure.



benzene

often written as

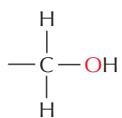


part of the hydrocarbon "tail" of a fatty acid molecule

## C-O CHEMICAL GROUPS

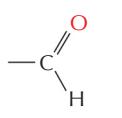
Many biological compounds contain a carbon bonded to an oxygen. For example,

alcohol



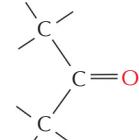
The  $-OH$  is called a **hydroxyl** group.

aldehyde



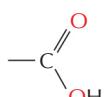
The  $=O$  is called a **carbonyl** group.

ketone



The  $=O$  is called a **carbonyl** group.

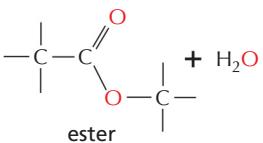
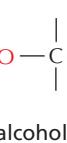
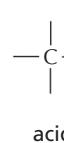
carboxylic acid



The  $-COOH$  is called a **carboxyl** group. In water this loses an  $H^+$  ion to become  $-COO^-$ .

esters

Esters are formed by a condensation reaction between an acid and an alcohol.



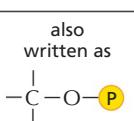
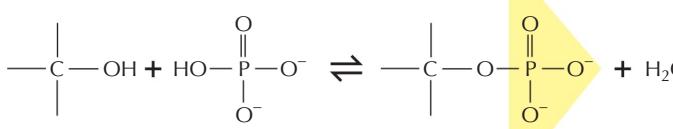
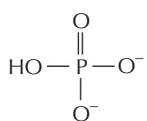
## SULFHYDRYL GROUP

The  $-\overset{|}{C}-SH$  is called a **sulphydryl** group. In the amino acid cysteine, the sulphydryl group may exist in the reduced form,  $-\overset{|}{C}-SH$  or more rarely in an oxidized, cross-bridging form,  $-\overset{|}{C}-S-S-\overset{|}{C}-$

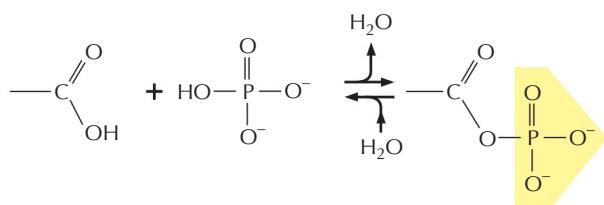
## PHOSPHATES

Inorganic phosphate is a stable ion formed from phosphoric acid,  $H_3PO_4$ . It is also written as  $\text{PO}_4^{3-}$ .

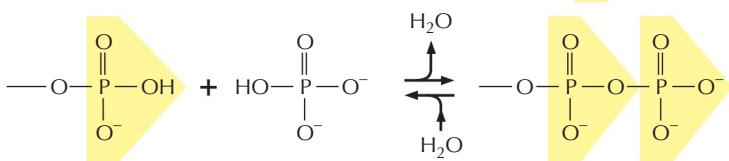
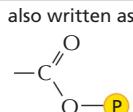
Phosphate esters can form between a phosphate and a free hydroxyl group. **Phosphate groups** are often attached to proteins in this way.



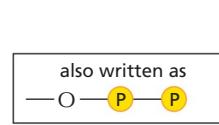
The combination of a phosphate and a carboxyl group, or two or more phosphate groups, gives an acid anhydride. Because compounds of this kind are easily hydrolysed in the cell, they are sometimes said to contain a "high-energy" bond.



high-energy acyl phosphate bond (carboxylic-phosphoric acid anhydride) found in some metabolites



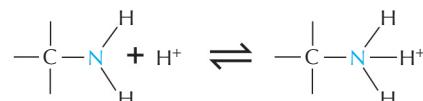
phosphoanhydride—a high-energy bond found in molecules such as ATP



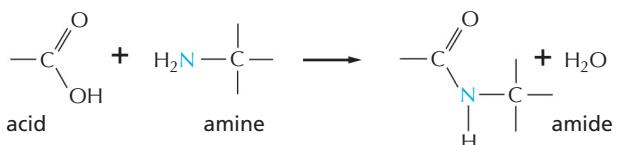
## C-N CHEMICAL GROUPS

Amines and amides are two important examples of compounds containing a carbon linked to a nitrogen.

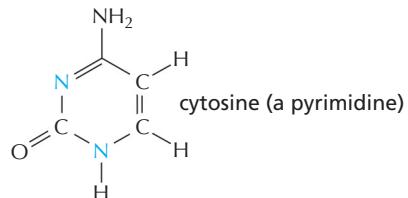
**Amines** in water combine with an  $H^+$  ion to become positively charged.



**Amides** are formed by combining an acid and an amine. Unlike amines, amides are uncharged in water. An example is the peptide bond that joins amino acids in a protein.

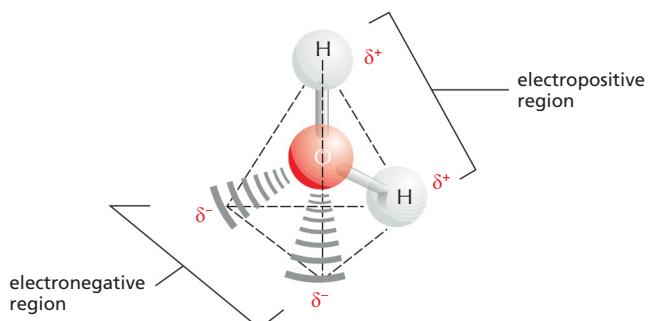


Nitrogen also occurs in several ring compounds, including important constituents of nucleic acids: purines and pyrimidines.



## WATER

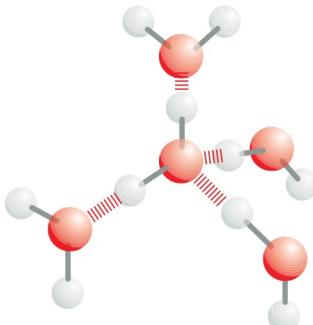
Two atoms, connected by a covalent bond, may exert different attractions for the electrons of the bond. In such cases the bond is **polar**, with one end slightly negatively charged ( $\delta^-$ ) and the other slightly positively charged ( $\delta^+$ ).



Although a water molecule has an overall neutral charge (having the same number of electrons and protons), the electrons are asymmetrically distributed, which makes the molecule polar. The oxygen nucleus draws electrons away from the hydrogen nuclei, leaving these nuclei with a small net positive charge. The excess of electron density on the oxygen atom creates weakly negative regions at the other two corners of an imaginary tetrahedron.

## WATER STRUCTURE

Molecules of water join together transiently in a hydrogen-bonded lattice. Even at 37°C, 15% of the water molecules are joined to four others in a short-lived assembly known as a “flickering cluster.”

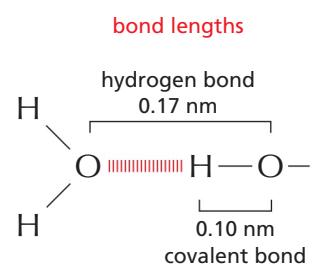
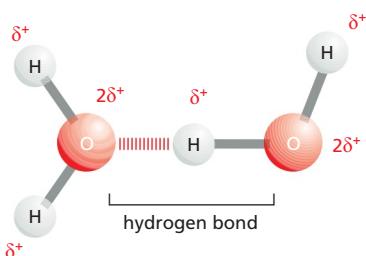


The cohesive nature of water is responsible for many of its unusual properties, such as high surface tension, specific heat, and heat of vaporization.

## HYDROGEN BONDS

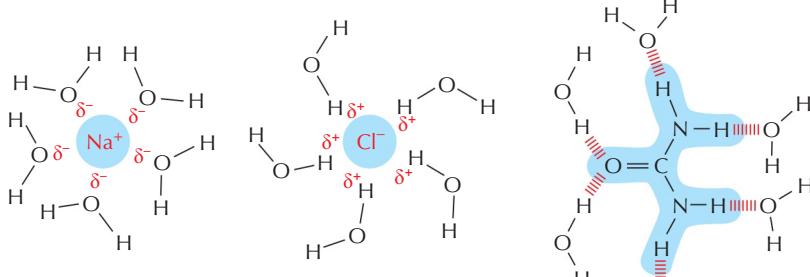
Because they are polarized, two adjacent H<sub>2</sub>O molecules can form a linkage known as a **hydrogen bond**. Hydrogen bonds have only about 1/20 the strength of a covalent bond.

Hydrogen bonds are strongest when the three atoms lie in a straight line.



## HYDROPHILIC MOLECULES

Substances that dissolve readily in water are termed **hydrophilic**. They are composed of ions or polar molecules that attract water molecules through electrical charge effects. Water molecules surround each ion or polar molecule on the surface of a solid substance and carry it into solution.

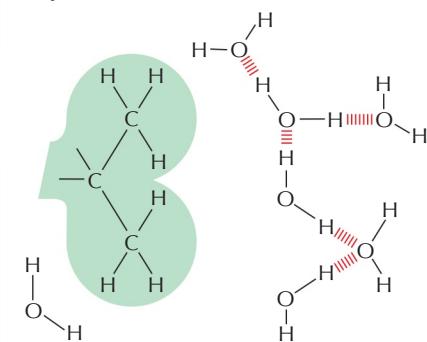


**Ionic** substances such as sodium chloride dissolve because water molecules are attracted to the positive (Na<sup>+</sup>) or negative (Cl<sup>-</sup>) charge of each ion.

**Polar** substances such as urea dissolve because their molecules form hydrogen bonds with the surrounding water molecules.

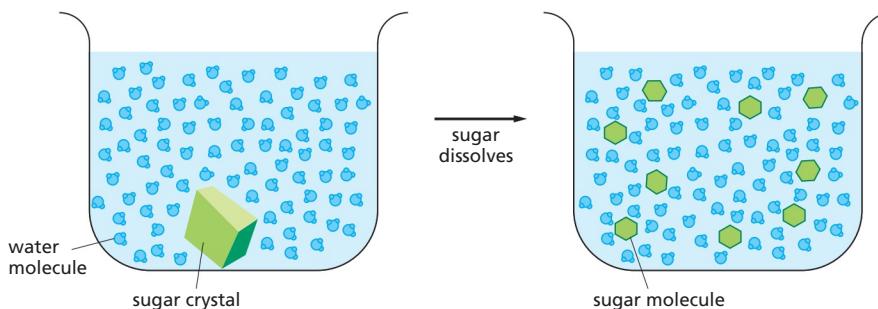
## HYDROPHOBIC MOLECULES

Molecules that contain a preponderance of nonpolar bonds are usually insoluble in water and are termed **hydrophobic**. This is true, especially, of hydrocarbons, which contain many C-H bonds. Water molecules are not attracted to such molecules and so have little tendency to surround them and carry them into solution.



## WATER AS A SOLVENT

Many substances, such as household sugar, **dissolve** in water. That is, their molecules separate from each other, each becoming surrounded by water molecules.



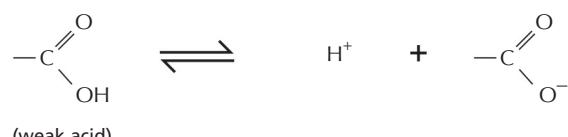
When a substance dissolves in a liquid, the mixture is termed a **solution**. The dissolved substance (in this case sugar) is the **solute**, and the liquid that does the dissolving (in this case water) is the **solvent**. Water is an excellent solvent for many substances because of its polar bonds.

## ACIDS

Substances that release hydrogen ions into solution are called **acids**.



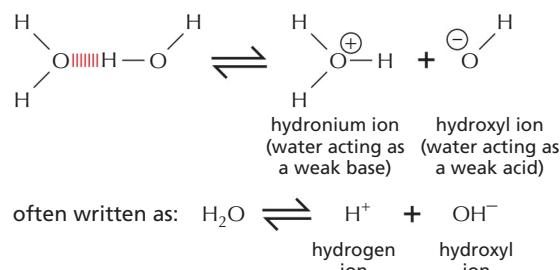
Many of the acids important in the cell are only partially dissociated, and they are therefore **weak acids**—for example, the carboxyl group ( $-\text{COOH}$ ), which dissociates to give a hydrogen ion in solution.



Note that this is a reversible reaction.

## HYDROGEN ION EXCHANGE

Positively charged hydrogen ions ( $\text{H}^+$ ) can spontaneously move from one water molecule to another, thereby creating two ionic species.



Since the process is rapidly reversible, hydrogen ions are continually shuttling between water molecules. Pure water contains a steady-state concentration of hydrogen ions and hydroxyl ions (both  $10^{-7} \text{ M}$ ).

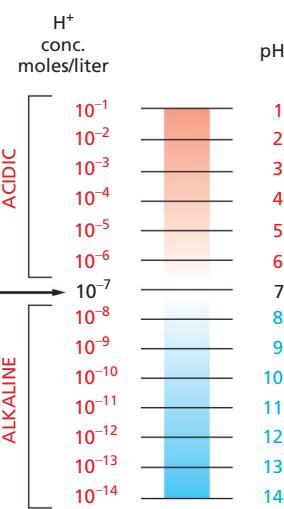
## pH

The acidity of a solution is defined by the concentration of  $\text{H}^+$  ions it possesses. For convenience we use the pH scale, where

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

For pure water

$$[\text{H}^+] = 10^{-7} \text{ moles/liter}$$



## BASES

Substances that reduce the number of hydrogen ions in solution are called **bases**. Some bases, such as ammonia, combine directly with hydrogen ions.



Other bases, such as sodium hydroxide, reduce the number of  $\text{H}^+$  ions indirectly, by making  $\text{OH}^-$  ions that then combine directly with  $\text{H}^+$  ions to make  $\text{H}_2\text{O}$ .

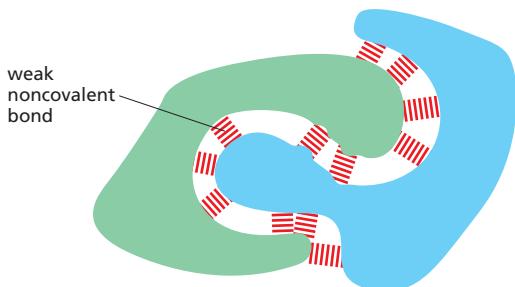


Many bases found in cells are partially associated with  $\text{H}^+$  ions and are termed **weak bases**. This is true of compounds that contain an amino group ( $-\text{NH}_2$ ), which has a weak tendency to reversibly accept an  $\text{H}^+$  ion from water, increasing the quantity of free  $\text{OH}^-$  ions.



### WEAK NONCOVALENT CHEMICAL BONDS

Organic molecules can interact with other molecules through three types of short-range attractive forces known as *noncovalent bonds*: van der Waals attractions, electrostatic attractions, and hydrogen bonds. The repulsion of hydrophobic groups from water is also important for the folding of biological macromolecules.



Weak noncovalent chemical bonds have less than 1/20 the strength of a strong covalent bond. They are strong enough to provide tight binding only when many of them are formed simultaneously.

### HYDROGEN BONDS

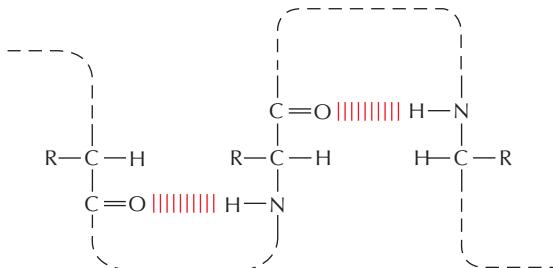
As already described for water (see Panel 2-2), **hydrogen bonds** form when a hydrogen atom is "sandwiched" between two electron-attracting atoms (usually oxygen or nitrogen).

Hydrogen bonds are strongest when the three atoms are in a straight line:

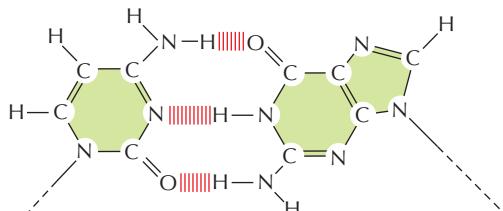


Examples in macromolecules:

Amino acids in a polypeptide chain can be hydrogen-bonded together. These stabilize the structure of folded proteins.



Two bases, G and C, are hydrogen-bonded in a DNA double helix.



### VAN DER WAALS ATTRACTIONS

If two atoms are too close together they repel each other very strongly. For this reason, an atom can often be treated as a sphere with a fixed radius. The characteristic "size" for each atom is specified by a unique **van der Waals radius**. The contact distance between any two noncovalently bonded atoms is the sum of their van der Waals radii.

0.12 nm radius	0.2 nm radius	0.15 nm radius	0.14 nm radius

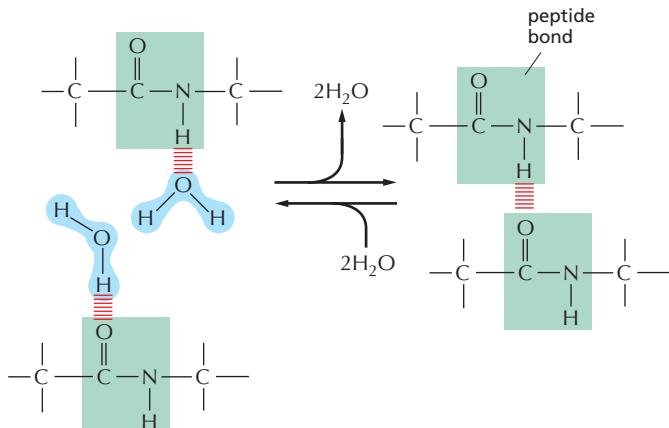
At very short distances any two atoms show a weak bonding interaction due to their fluctuating electrical charges. The two atoms will be attracted to each other in this way until the distance between their nuclei is approximately equal to the sum of their van der Waals radii. Although they are individually very weak, **van der Waals attractions** can become important when two macromolecular surfaces fit very close together, because many atoms are involved.

Note that when two atoms form a covalent bond, the centers of the two atoms (the two atomic nuclei) are much closer together than the sum of the two van der Waals radii. Thus,

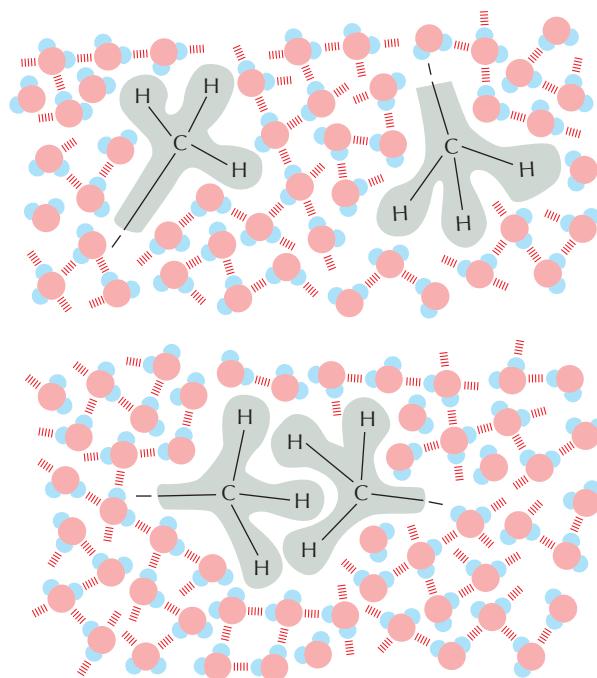


### HYDROGEN BONDS IN WATER

Any molecules that can form hydrogen bonds to each other can alternatively form hydrogen bonds to water molecules. Because of this competition with water molecules, the hydrogen bonds formed between two molecules dissolved in water are relatively weak.



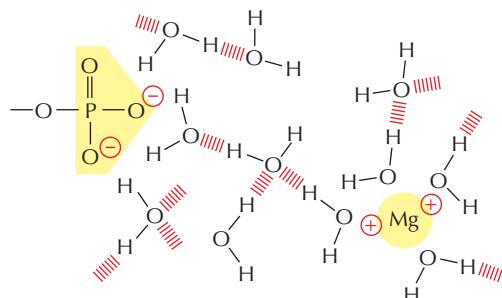
### HYDROPHOBIC FORCES



Water forces hydrophobic groups together, because doing so minimizes their disruptive effects on the hydrogen-bonded water network. Hydrophobic groups held together in this way are sometimes said to be held together by "hydrophobic bonds," even though the apparent attraction is actually caused by a repulsion from the water.

### ELECTROSTATIC ATTRACtIONS IN AQUEOUS SOLUTIONS

Charged groups are shielded by their interactions with water molecules. Electrostatic attractions are therefore quite weak in water.



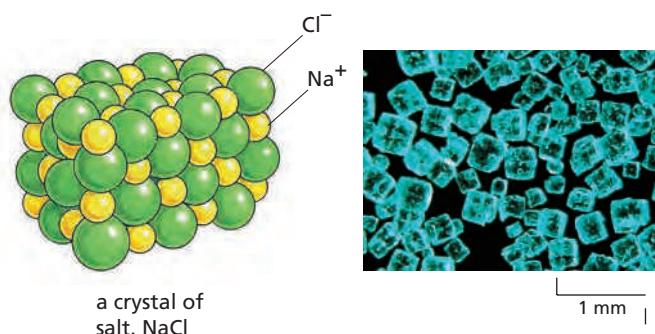
### ELECTROSTATIC ATTRACtIONS

Attractive forces occur both between fully charged groups (ionic bond) and between the partially charged groups on polar molecules.

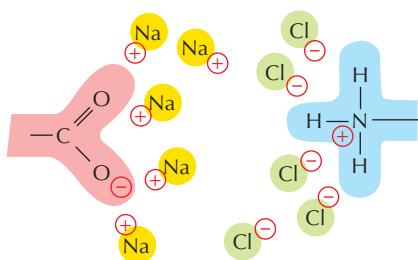


The force of attraction between the two charges,  $\delta^+$  and  $\delta^-$ , falls off rapidly as the distance between the charges increases.

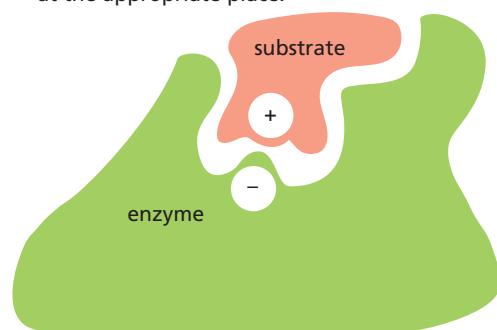
In the absence of water, electrostatic forces are very strong. They are responsible for the strength of such minerals as marble and agate, and for crystal formation in common table salt, NaCl.



Similarly, ions in solution can cluster around charged groups and further weaken these attractions.

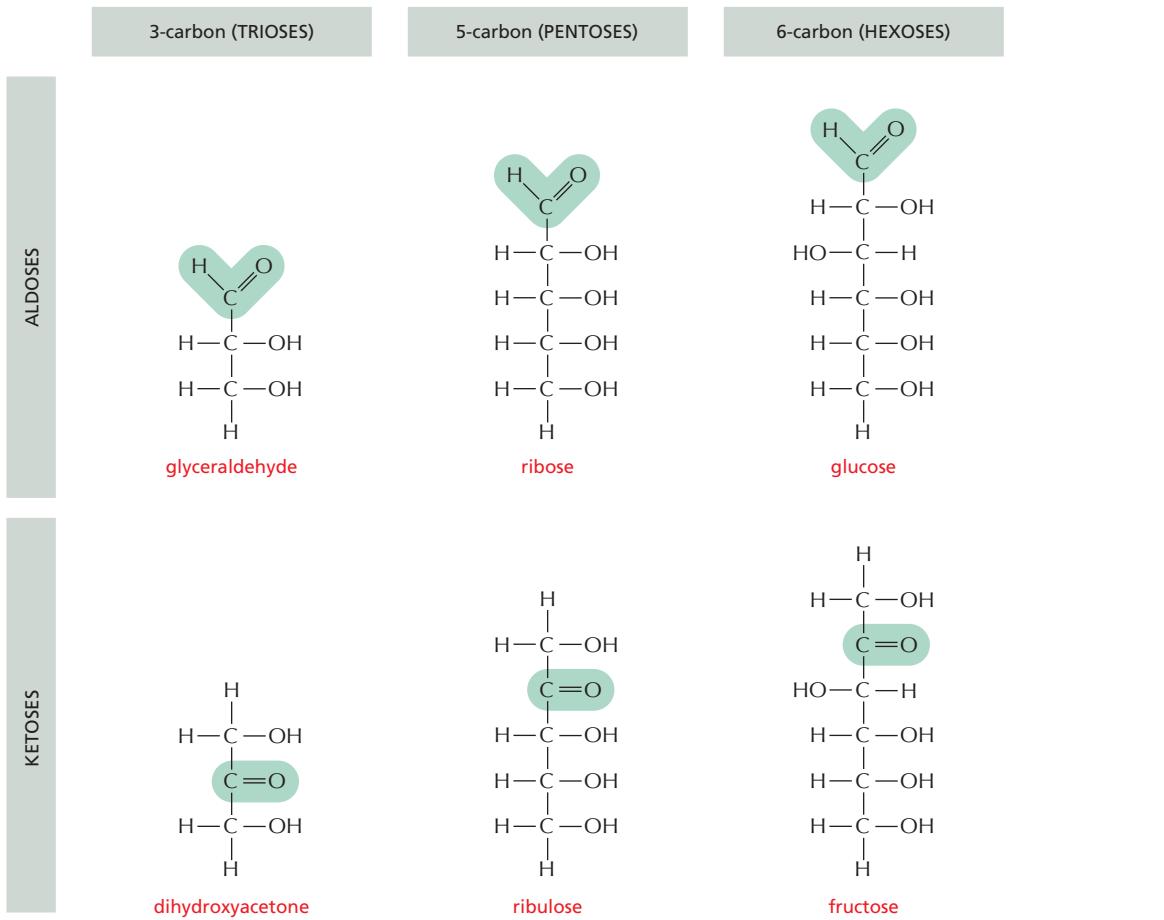


Despite being weakened by water and salt, electrostatic attractions are very important in biological systems. For example, an enzyme that binds a positively charged substrate will often have a negatively charged amino acid side chain at the appropriate place.



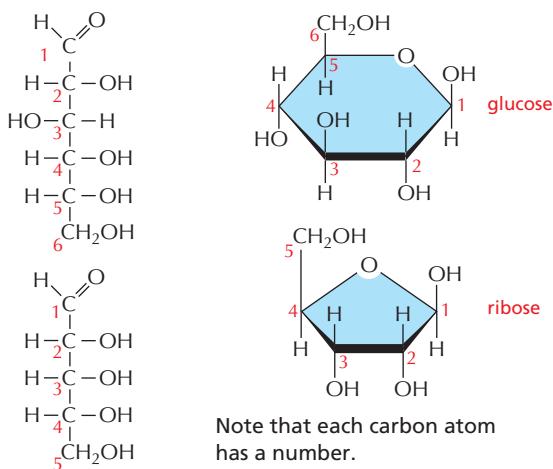
## MONOSACCHARIDES

Monosaccharides usually have the general formula  $(\text{CH}_2\text{O})_n$ , where  $n$  can be 3, 4, 5, 6, 7, or 8, and have two or more hydroxyl groups. They either contain an aldehyde group ( $-\text{C}\text{H}_2\text{OH}$ ) and are called aldoses, or a ketone group ( $\text{C}=\text{O}$ ) and are called ketoses.



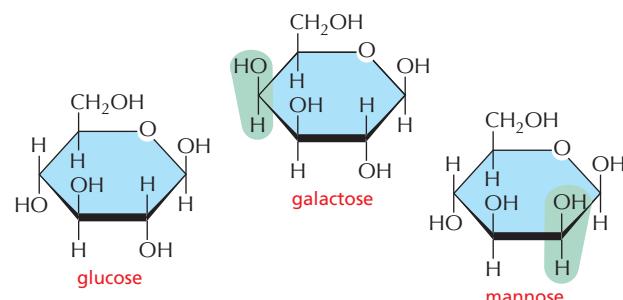
## RING FORMATION

In aqueous solution, the aldehyde or ketone group of a sugar molecule tends to react with a hydroxyl group of the same molecule, thereby closing the molecule into a ring.



## ISOMERS

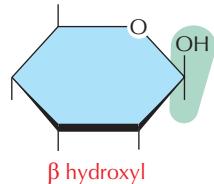
Many monosaccharides differ only in the spatial arrangement of atoms—that is, they are **isomers**. For example, glucose, galactose, and mannose have the same formula ( $\text{C}_6\text{H}_{12}\text{O}_6$ ) but differ in the arrangement of groups around one or two carbon atoms.



These small differences make only minor changes in the chemical properties of the sugars. But they are recognized by enzymes and other proteins and therefore can have major biological effects.

### $\alpha$ AND $\beta$ LINKS

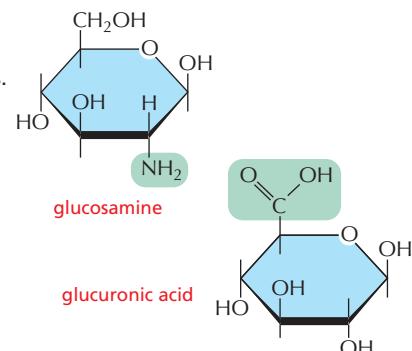
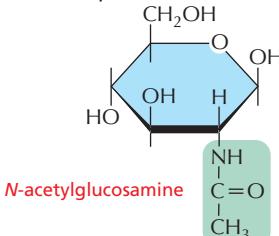
The hydroxyl group on the carbon that carries the aldehyde or ketone can rapidly change from one position to the other. These two positions are called  $\alpha$  and  $\beta$ .



As soon as one sugar is linked to another, the  $\alpha$  or  $\beta$  form is frozen.

### SUGAR DERIVATIVES

The hydroxyl groups of a simple monosaccharide such as glucose can be replaced by other groups. For example,



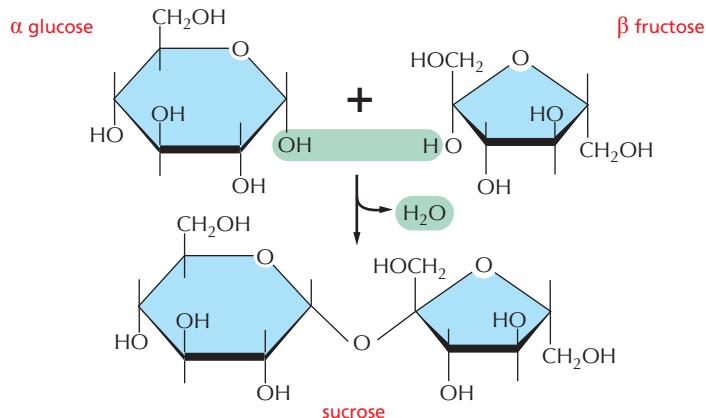
### DISACCHARIDES

The carbon that carries the aldehyde or the ketone can react with any hydroxyl group on a second sugar molecule to form a **disaccharide**. The linkage is called a glycosidic bond.

Three common disaccharides are

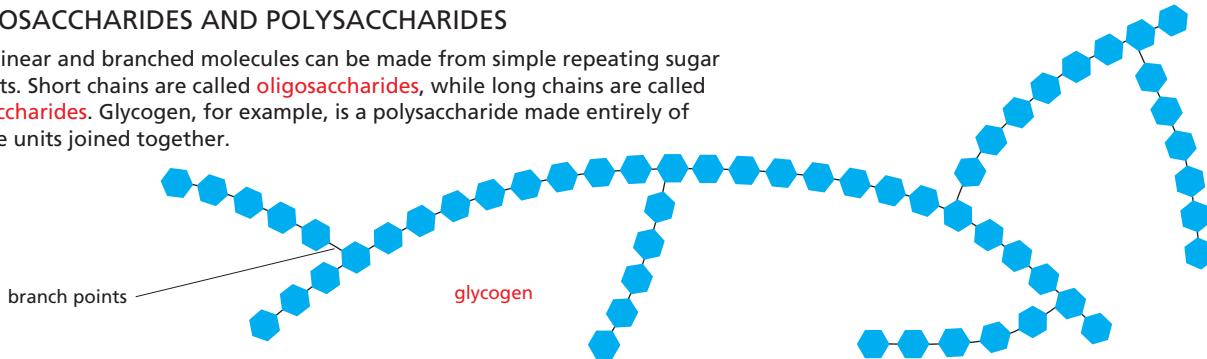
- maltose (glucose + glucose)
- lactose (galactose + glucose)
- sucrose (glucose + fructose)

The reaction forming sucrose is shown here.



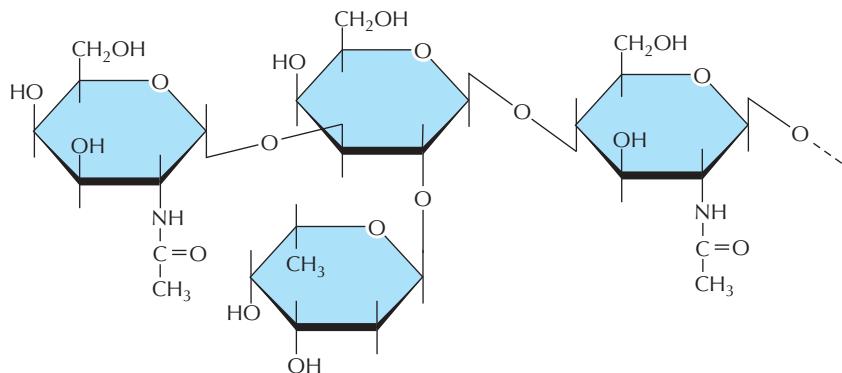
### OLIGOSACCHARIDES AND POLYSACCHARIDES

Large linear and branched molecules can be made from simple repeating sugar subunits. Short chains are called **oligosaccharides**, while long chains are called **polysaccharides**. Glycogen, for example, is a polysaccharide made entirely of glucose units joined together.



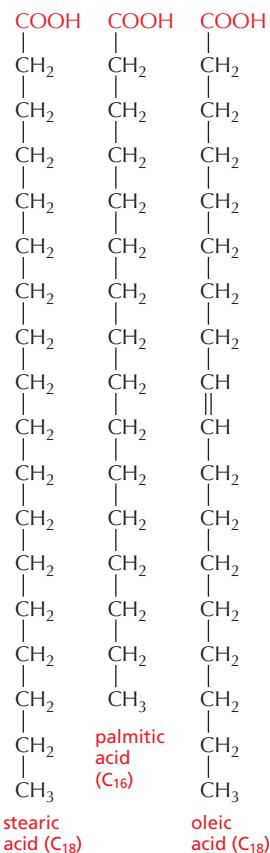
### COMPLEX OLIGOSACCHARIDES

In many cases a sugar sequence is nonrepetitive. Many different molecules are possible. Such complex oligosaccharides are usually linked to proteins or to lipids, as is this oligosaccharide, which is part of a cell-surface molecule that defines a particular blood group.

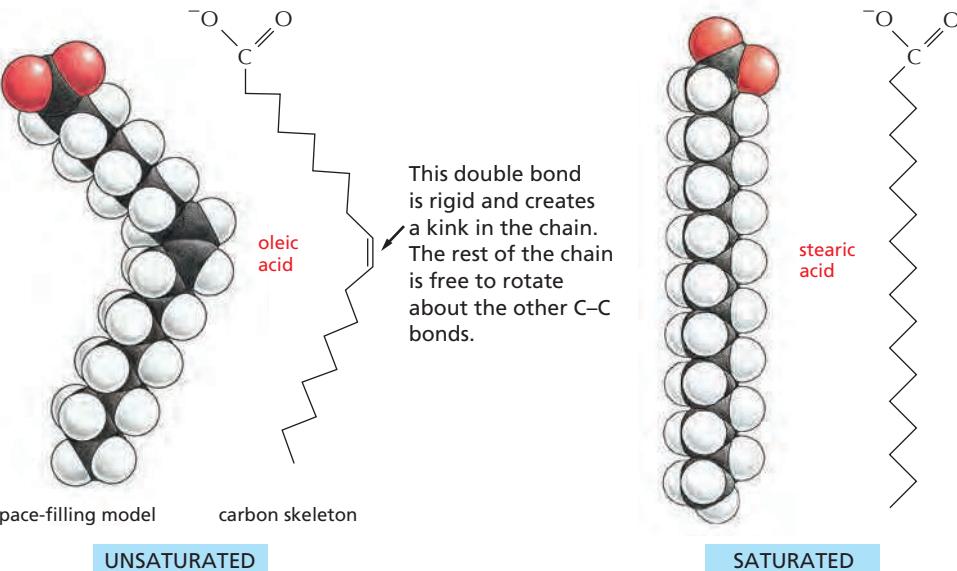


## COMMON FATTY ACIDS

These are carboxylic acids with long hydrocarbon tails.

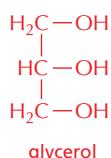
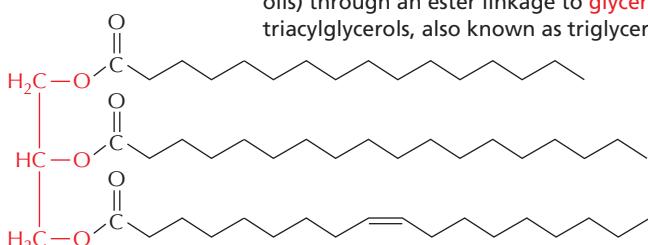


Hundreds of different kinds of fatty acids exist. Some have one or more double bonds in their hydrocarbon tail and are said to be **unsaturated**. Fatty acids with no double bonds are **saturated**.



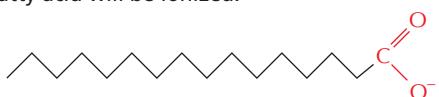
## TRIACYLGLYCEROLS

Fatty acids are stored as an energy reserve (fats and oils) through an ester linkage to **glycerol** to form triacylglycerols, also known as triglycerides.

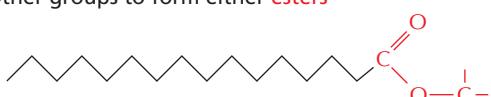


## CARBOXYL GROUP

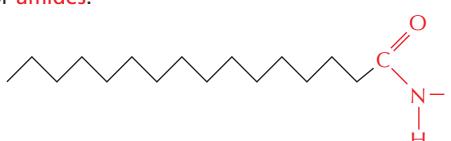
If free, the carboxyl group of a fatty acid will be ionized.



But more usually it is linked to other groups to form either **esters**

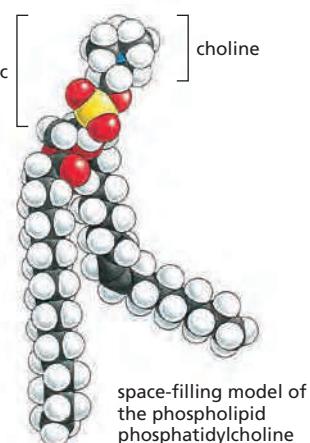
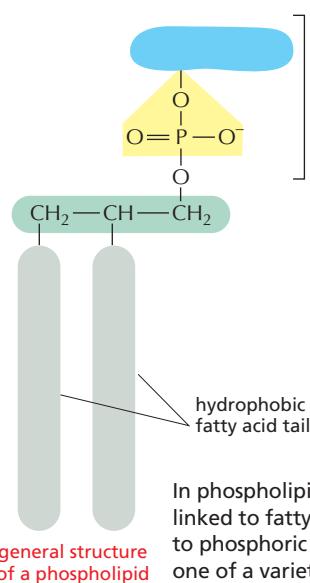


or amides.



## PHOSPHOLIPIDS

Phospholipids are the major constituents of cell membranes.

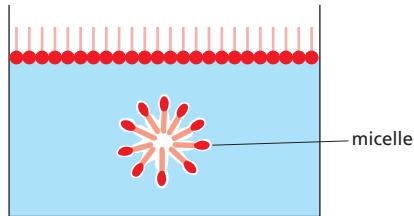


In phospholipids, two of the -OH groups in glycerol are linked to fatty acids, while the third -OH group is linked to phosphoric acid. The phosphate is further linked to one of a variety of small polar groups, such as choline.

## LIPID AGGREGATES

Fatty acids have a hydrophilic head and a hydrophobic tail.

In water they can form a surface film or form small micelles.

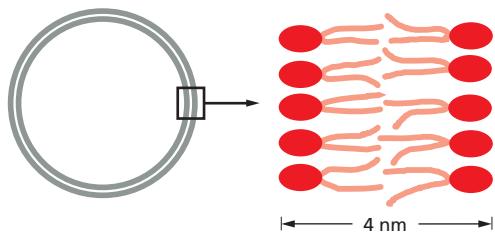


Their derivatives can form larger aggregates held together by hydrophobic forces:

**Triacylglycerols** (triglycerides) can form large spherical fat droplets in the cell cytoplasm

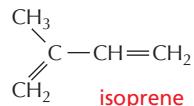


**Phospholipids** and **glycolipids** form self-sealing lipid bilayers that are the basis for all cell membranes.



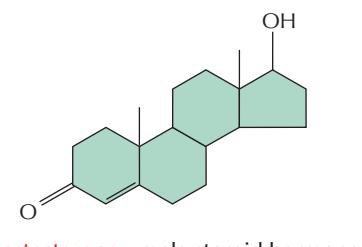
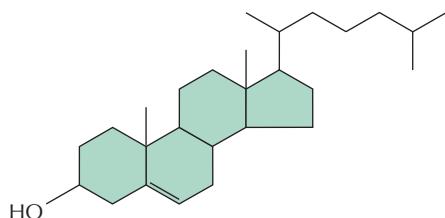
## OTHER LIPIDS

Lipids are defined as the water-insoluble molecules in cells that are soluble in organic solvents. Two other common types of lipids are steroids and polyisoprenoids. Both are made from isoprene units.



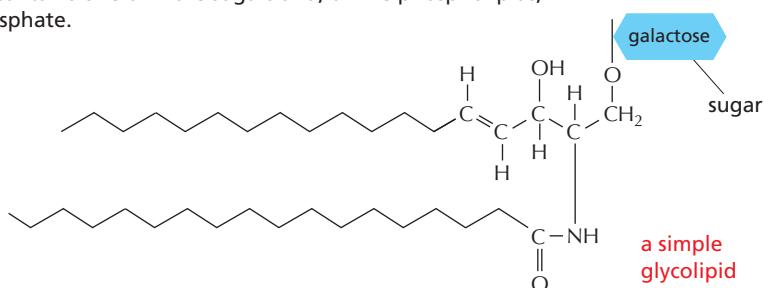
## STEROIDS

Steroids have a common multiple-ring structure.



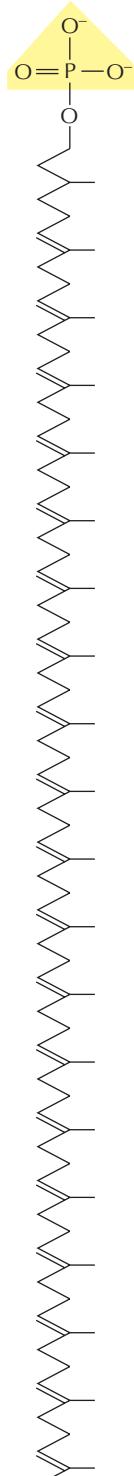
## GLYCOLIPIDS

Like phospholipids, these compounds are composed of a hydrophobic region, containing two long hydrocarbon tails and a polar region, which contains one or more sugars and, unlike phospholipids, no phosphate.

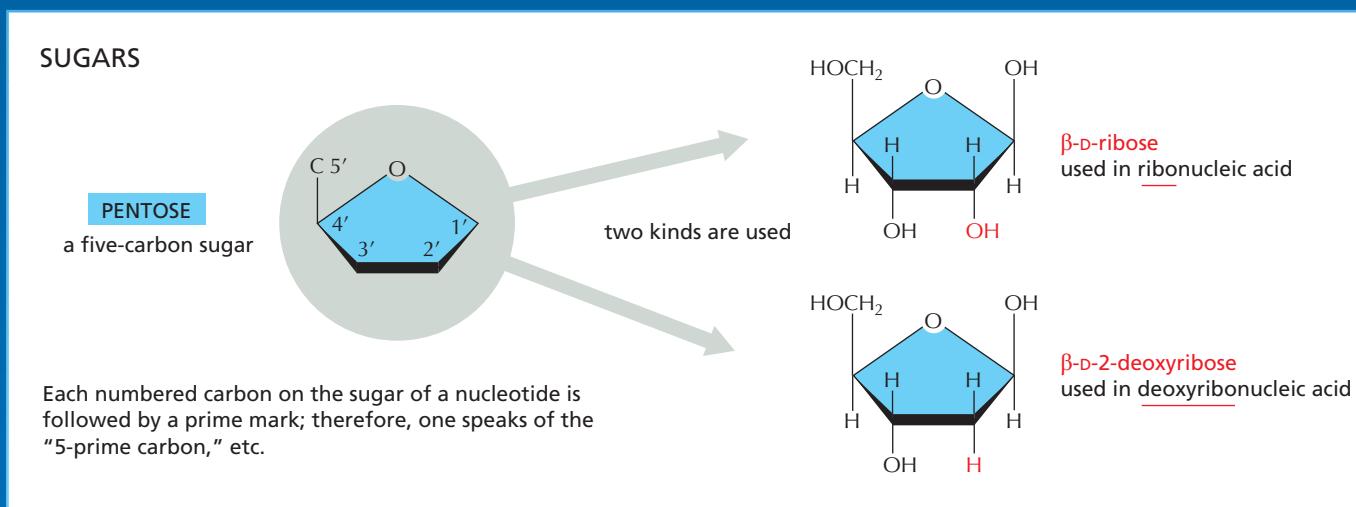
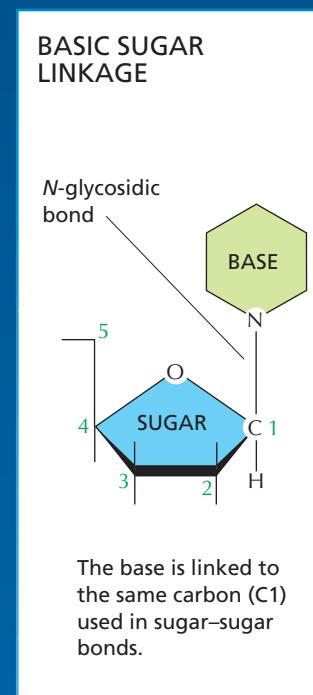
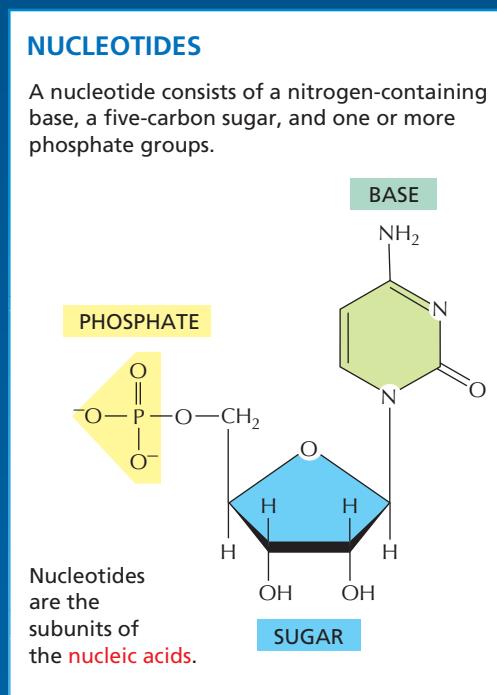
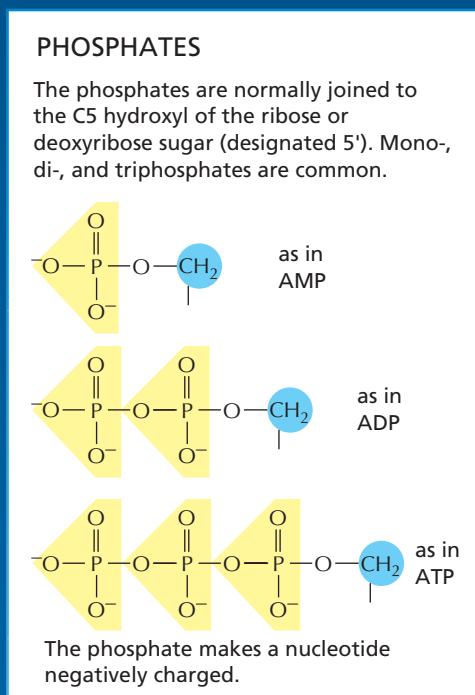
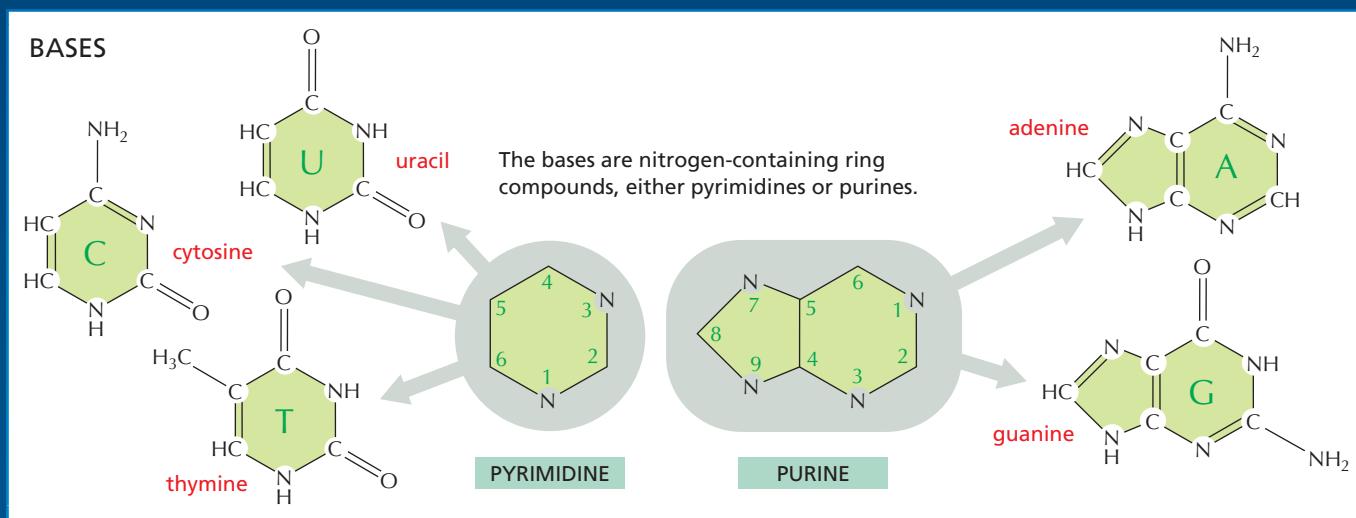


## POLYISOPRENOIDS

long-chain polymers of isoprene



**dolichol phosphate**—used to carry activated sugars in the membrane-associated synthesis of glycoproteins and some polysaccharides



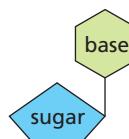
**NOMENCLATURE**

A nucleoside or nucleotide is named according to its nitrogenous base.

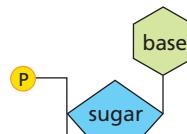
BASE	NUCLEOSIDE	ABBR.
adenine	adenosine	A
guanine	guanosine	G
cytosine	cytidine	C
uracil	uridine	U
thymine	thymidine	T

Single-letter abbreviations are used variously as shorthand for (1) the base alone, (2) the nucleoside, or (3) the whole nucleotide—the context will usually make clear which of the three entities is meant. When the context is not sufficient, we will add the terms "base", "nucleoside", "nucleotide", or—as in the examples below—use the full 3-letter nucleotide code.

AMP = adenosine monophosphate  
dAMP = deoxyadenosine monophosphate  
UDP = uridine diphosphate  
ATP = adenosine triphosphate



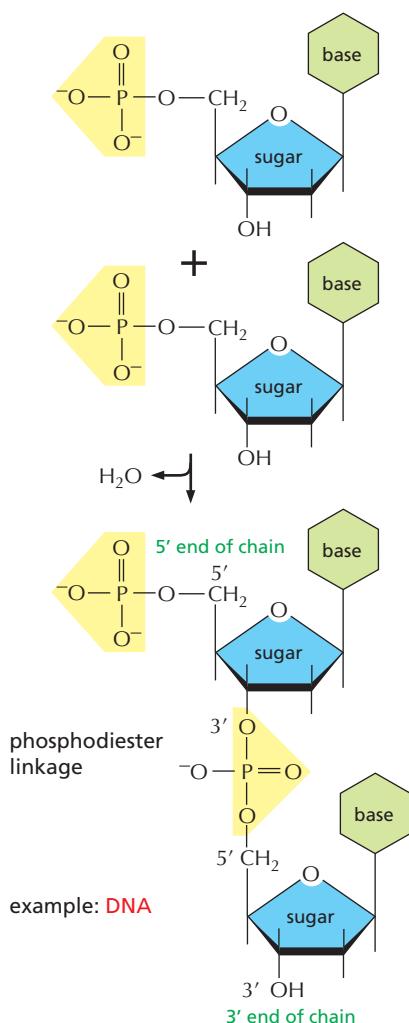
BASE + SUGAR = NUCLEOSIDE



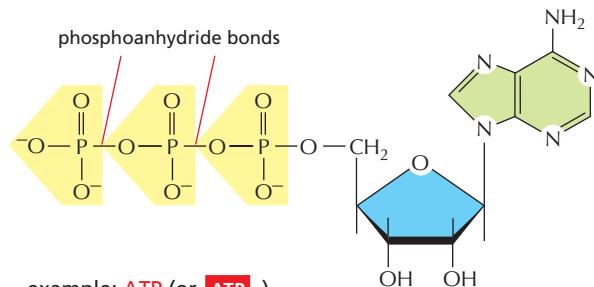
BASE + SUGAR + PHOSPHATE = NUCLEOTIDE

**NUCLEIC ACIDS**

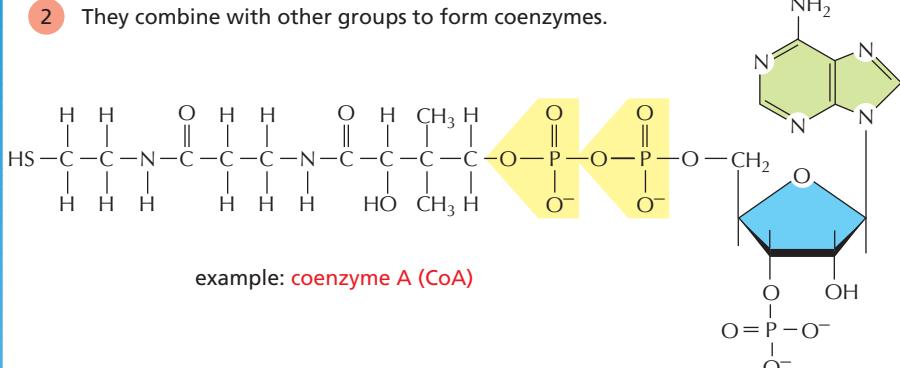
Nucleotides are joined together by a **phosphodiester linkage** between 5' and 3' carbon atoms to form nucleic acids. The linear sequence of nucleotides in a nucleic acid chain is commonly abbreviated by a one-letter code, such as A—G—C—T—T—A—C—A, with the 5' end of the chain at the left.

**NUCLEOTIDES HAVE MANY OTHER FUNCTIONS**

- 1 They carry chemical energy in their easily hydrolyzed phosphoanhydride bonds.

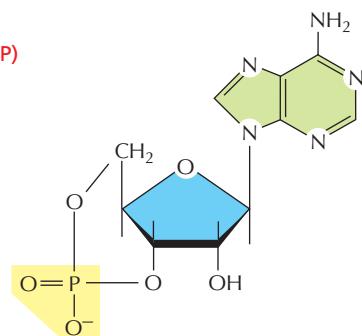


- 2 They combine with other groups to form coenzymes.



- 3 They are used as specific signaling molecules in the cell.

example: cyclic AMP (cAMP)

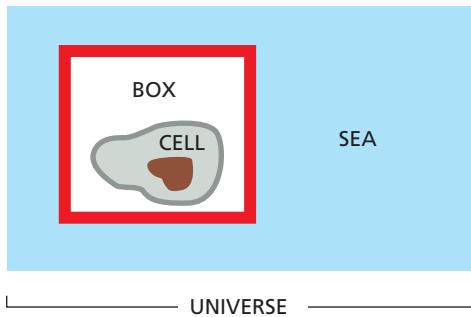


## THE IMPORTANCE OF FREE ENERGY FOR CELLS

Life is possible because of the complex network of interacting chemical reactions occurring in every cell. In viewing the metabolic pathways that comprise this network, one might suspect that the cell has had the ability to evolve an enzyme to carry out any reaction that it needs. But this is not so. Although enzymes are powerful catalysts, they can speed up only those reactions that are thermodynamically possible; other reactions proceed in cells only because they are *coupled* to very favorable reactions that drive them. The question of whether a reaction

can occur spontaneously, or instead needs to be coupled to another reaction, is central to cell biology. The answer is obtained by reference to a quantity called the *free energy*: the total change in free energy during a set of reactions determines whether or not the entire reaction sequence can occur. In this panel, we shall explain some of the fundamental ideas—derived from a special branch of chemistry and physics called *thermodynamics*—that are required for understanding what free energy is and why it is so important to cells.

## ENERGY RELEASED BY CHANGES IN CHEMICAL BONDING IS CONVERTED INTO HEAT



An *enclosed system* is defined as a collection of molecules that does not exchange matter with the rest of the universe (for example, the “cell in a box” shown above). Any such system will contain molecules with a total energy  $E$ . This energy will be distributed in a variety of ways: some as the translational energy of the molecules, some as their vibrational and rotational energies, but most as the bonding energies between the individual atoms that make up the molecules. Suppose that a reaction occurs in the system. The *first law of thermodynamics* places a constraint on what types of reactions are possible: it states that “*in any process, the total energy of the universe remains constant*.” For example, suppose that reaction  $A \rightarrow B$  occurs somewhere in the box and releases a great deal of chemical-bond energy. This energy will initially increase the intensity of molecular motions (translational, vibrational, and rotational) in the system, which is equivalent to raising its temperature. However, these increased motions will soon be transferred out of the system by a series

of molecular collisions that heat up first the walls of the box and then the outside world (represented by the sea in our example). In the end, the system returns to its initial temperature, by which time all the chemical-bond energy released in the box has been converted into heat energy and transferred out of the box to the surroundings. According to the first law, the change in the energy in the box ( $\Delta E_{\text{box}}$ , which we shall denote as  $\Delta E$ ) must be equal and opposite to the amount of heat energy transferred, which we shall designate as  $h$ : that is,  $\Delta E = -h$ . Thus, the energy in the box ( $E$ ) decreases when heat leaves the system.

$E$  also can change during a reaction as a result of work being done on the outside world. For example, suppose that there is a small increase in the volume ( $\Delta V$ ) of the box during a reaction. Since the walls of the box must push against the constant pressure ( $P$ ) in the surroundings in order to expand, this does work on the outside world and requires energy. The energy used is  $P(\Delta V)$ , which according to the first law must decrease the energy in the box ( $E$ ) by the same amount. In most reactions, chemical-bond energy is converted into both work and heat. *Enthalpy* ( $H$ ) is a composite function that includes both of these ( $H = E + PV$ ). To be rigorous, it is the change in enthalpy ( $\Delta H$ ) in an enclosed system, and not the change in energy, that is equal to the heat transferred to the outside world during a reaction. Reactions in which  $H$  decreases release heat to the surroundings and are said to be “exothermic,” while reactions in which  $H$  increases absorb heat from the surroundings and are said to be “endothermic.” Thus,  $-h = \Delta H$ . However, the volume change is negligible in most biological reactions, so to a good approximation

$$-h = \Delta H \approx \Delta E$$

## THE SECOND LAW OF THERMODYNAMICS

Consider a container in which 1000 coins are all lying heads up. If the container is shaken vigorously, subjecting the coins to the types of random motions that all molecules experience due to their frequent collisions with other molecules, one will end up with about half the coins oriented heads down. The reason for this reorientation is that there is only a single way in which the original orderly state of the coins can be reinstated (every coin must lie heads up), whereas there are many different ways (about  $10^{298}$ ) to achieve a disorderly state in which there is an equal mixture of heads and tails; in fact, there are more ways

to achieve a 50-50 state than to achieve any other state. Each state has a probability of occurrence that is proportional to the number of ways it can be realized. The *second law of thermodynamics* states that “*systems will change spontaneously from states of lower probability to states of higher probability*.” Since states of lower probability are more “ordered” than states of high probability, the second law can be restated: “*the universe constantly changes so as to become more disordered*.”

## THE ENTROPY, $S$

The second law (but not the first law) allows one to predict the *direction* of a particular reaction. But to make it useful for this purpose, one needs a convenient measure of the probability or, equivalently, the degree of disorder of a state. The entropy ( $S$ ) is such a measure. It is a logarithmic function of the probability such that the *change in entropy* ( $\Delta S$ ) that occurs when the reaction  $A \rightarrow B$  converts one mole of A into one mole of B is

$$\Delta S = R \ln p_B / p_A$$

where  $p_A$  and  $p_B$  are the probabilities of the two states A and B,  $R$  is the gas constant ( $8.31 \text{ J K}^{-1} \text{ mole}^{-1}$ ), and  $\Delta S$  is measured in entropy units (eu). In our initial example of 1000 coins, the relative probability of all heads (state A) versus half heads and half tails (state B) is equal to the ratio of the number of different ways that the two results can be obtained. One can calculate that  $p_A = 1$  and  $p_B = 1000!(500! \times 500!) = 10^{299}$ . Therefore, the entropy change for the reorientation of the coins when their

container is vigorously shaken and an equal mixture of heads and tails is obtained is  $R \ln (10^{298})$ , or about 1370 eu per mole of such containers ( $6 \times 10^{23}$  containers). We see that, because  $\Delta S$  defined above is positive for the transition from state A to state B ( $p_B/p_A > 1$ ), reactions with a large *increase* in  $S$  (that is, for which  $\Delta S > 0$ ) are favored and will occur spontaneously.

As discussed in Chapter 2, heat energy causes the random commotion of molecules. Because the transfer of heat from an enclosed system to its surroundings increases the number of different arrangements that the molecules in the outside world can have, it increases their entropy. It can be shown that the release of a fixed quantity of heat energy has a greater disordering effect at low temperature than at high temperature, and that the value of  $\Delta S$  for the surroundings, as defined above ( $\Delta S_{\text{sea}}$ ), is precisely equal to  $h$ , the amount of heat transferred to the surroundings from the system, divided by the absolute temperature ( $T$ ):

$$\Delta S_{\text{sea}} = h / T$$

## THE GIBBS FREE ENERGY, $G$

When dealing with an enclosed biological system, one would like to have a simple way of predicting whether a given reaction will or will not occur spontaneously in the system. We have seen that the crucial question is whether the entropy change for the universe is positive or negative when that reaction occurs. In our idealized system, the cell in a box, there are two separate components to the entropy change of the universe—the entropy change for the system enclosed in the box and the entropy change for the surrounding “sea”—and both must be added together before any prediction can be made. For example, it is possible for a reaction to absorb heat and thereby decrease the entropy of the sea ( $\Delta S_{\text{sea}} < 0$ ) and at the same time to cause such a large degree of disordering inside the box ( $\Delta S_{\text{box}} > 0$ ) that the total  $\Delta S_{\text{universe}} = \Delta S_{\text{sea}} + \Delta S_{\text{box}}$  is greater than 0. In this case, the reaction will occur spontaneously, even though the sea gives up heat to the box during the reaction. An example of such a reaction is the dissolving of sodium chloride in a beaker containing water (the “box”), which is a spontaneous process even though the temperature of the water drops as the salt goes into solution.

Chemists have found it useful to define a number of new “composite functions” that describe *combinations* of physical properties of a system. The properties that can be combined include the temperature ( $T$ ), pressure ( $P$ ), volume ( $V$ ), energy ( $E$ ), and entropy ( $S$ ). The enthalpy ( $H$ ) is one such composite function. But by far the most useful composite function for biologists is the *Gibbs free energy*,  $G$ . It serves as an accounting device that allows one to deduce the entropy change of the universe resulting from a chemical reaction in the box, while avoiding any separate consideration of the entropy change in the sea. The definition of  $G$  is

$$G = H - TS$$

where, for a box of volume  $V$ ,  $H$  is the enthalpy described above ( $E + PV$ ),  $T$  is the absolute temperature, and  $S$  is the entropy. Each of these quantities applies to the inside of the box only. The change in free energy during a reaction in the box (the  $G$  of the products minus the  $G$  of the starting materials) is denoted as  $\Delta G$  and, as we shall now demonstrate, it is a direct measure of the amount of disorder that is created in the universe when the reaction occurs.

At constant temperature the change in free energy ( $\Delta G$ ) during a reaction equals  $\Delta H - T\Delta S$ . Remembering that  $\Delta H = -h$ , the heat absorbed from the sea, we have

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

$$-\Delta G = h + T\Delta S, \text{ so } -\Delta G/T = h/T + \Delta S$$

But  $h/T$  is equal to the entropy change of the sea ( $\Delta S_{\text{sea}}$ ), and the  $\Delta S$  in the above equation is  $\Delta S_{\text{box}}$ . Therefore

$$-\Delta G/T = \Delta S_{\text{sea}} + \Delta S_{\text{box}} = \Delta S_{\text{universe}}$$

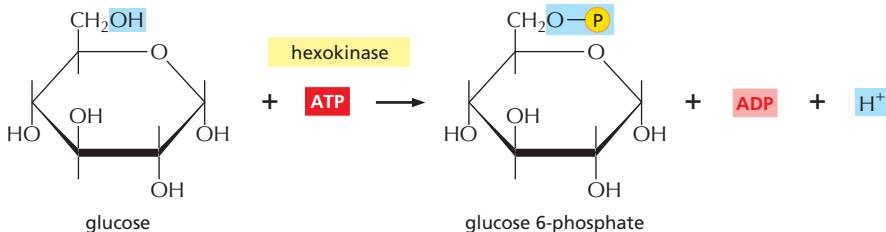
We conclude that the *free-energy change is a direct measure of the entropy change of the universe*. A reaction will proceed in the direction that causes the change in the free energy ( $\Delta G$ ) to be less than zero, because in this case there will be a positive entropy change in the universe when the reaction occurs.

For a complex set of coupled reactions involving many different molecules, the total free-energy change can be computed simply by adding up the free energies of all the different molecular species after the reaction and comparing this value with the sum of free energies before the reaction; for common substances the required free-energy values can be found from published tables. In this way, one can predict the direction of a reaction and thereby readily check the feasibility of any proposed mechanism. Thus, for example, from the observed values for the magnitude of the electrochemical proton gradient across the inner mitochondrial membrane and the  $\Delta G$  for ATP hydrolysis inside the mitochondrion, one can be certain that ATP synthase requires the passage of more than one proton for each molecule of ATP that it synthesizes.

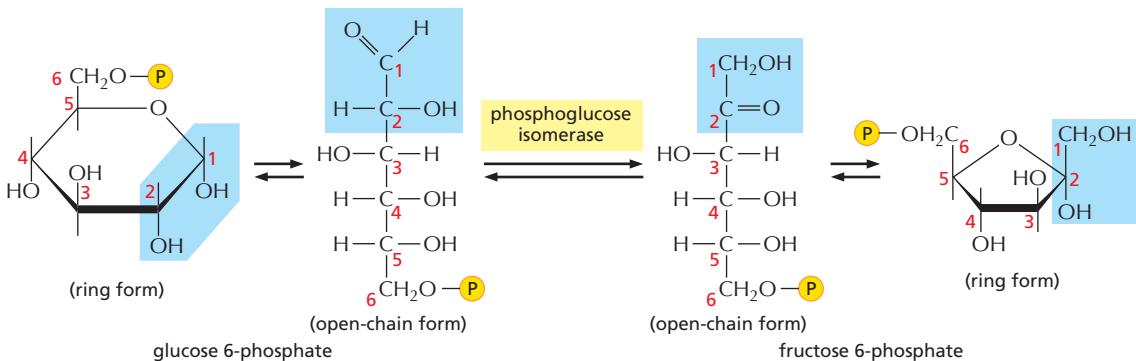
The value of  $\Delta G$  for a reaction is a direct measure of how far the reaction is from equilibrium. The large negative value for ATP hydrolysis in a cell merely reflects the fact that cells keep the ATP hydrolysis reaction as much as 10 orders of magnitude away from equilibrium. If a reaction reaches equilibrium,  $\Delta G = 0$ , the reaction then proceeds at precisely equal rates in the forward and backward direction. For ATP hydrolysis, equilibrium is reached when the vast majority of the ATP has been hydrolyzed, as occurs in a dead cell.

For each step, the part of the molecule that undergoes a change is shadowed in blue, and the name of the enzyme that catalyzes the reaction is in a yellow box.

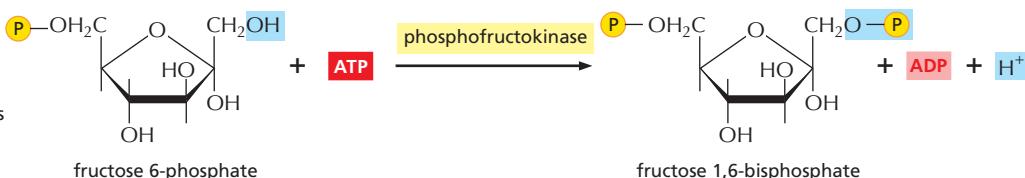
**Step 1** Glucose is phosphorylated by ATP to form a sugar phosphate. The negative charge of the phosphate prevents passage of the sugar phosphate through the plasma membrane, trapping glucose inside the cell.



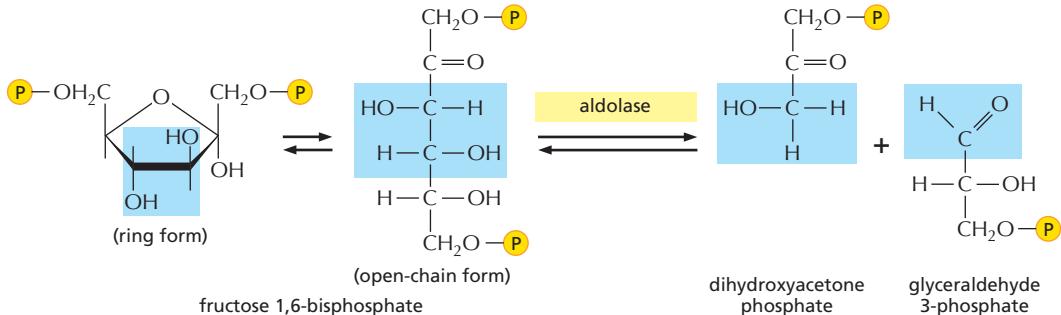
**Step 2** A readily reversible rearrangement of the chemical structure (isomerization) moves the carbonyl oxygen from carbon 1 to carbon 2, forming a ketose from an aldose sugar. (See Panel 2-3, pp. 70-71.)



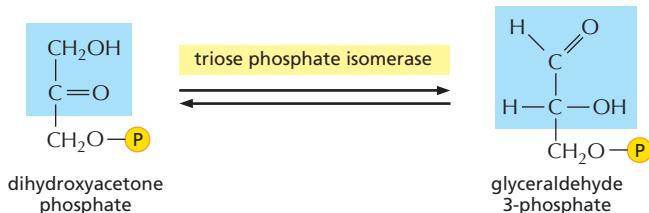
**Step 3** The new hydroxyl group on carbon 1 is phosphorylated by ATP, in preparation for the formation of two three-carbon sugar phosphates. The entry of sugars into glycolysis is controlled at this step, through regulation of the enzyme *phosphofructokinase*.



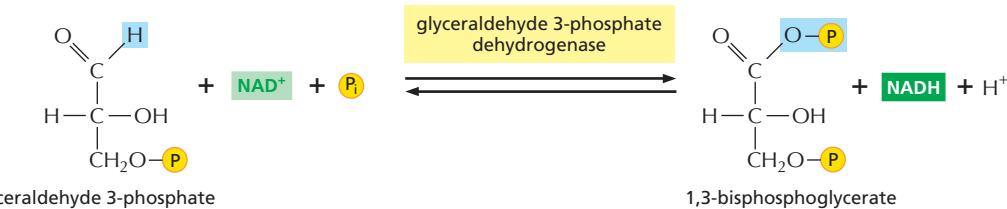
**Step 4** The six-carbon sugar is cleaved to produce two three-carbon molecules. Only the glyceraldehyde 3-phosphate can proceed immediately through glycolysis.



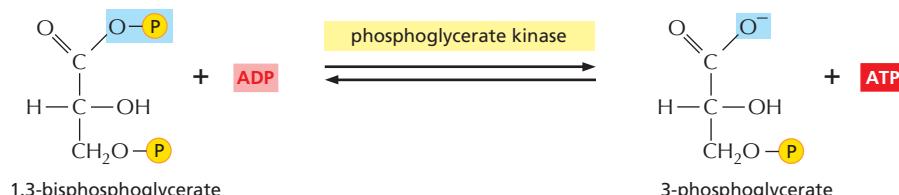
**Step 5** The other product of step 4, dihydroxyacetone phosphate, is isomerized to form glyceraldehyde 3-phosphate.



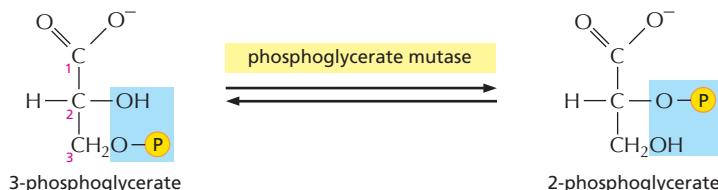
**Step 6** The two molecules of glyceraldehyde 3-phosphate are oxidized. The energy-generation phase of glycolysis begins, as NADH and a new high-energy anhydride linkage to phosphate are formed (see Figure 13–5).



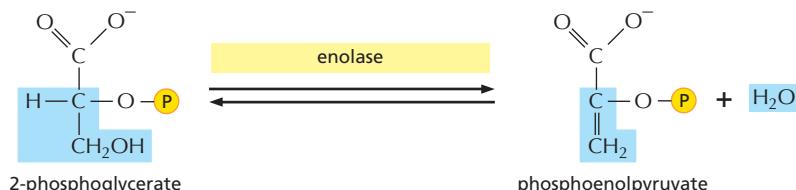
**Step 7** The transfer to ADP of the high-energy phosphate group that was generated in step 6 forms ATP.



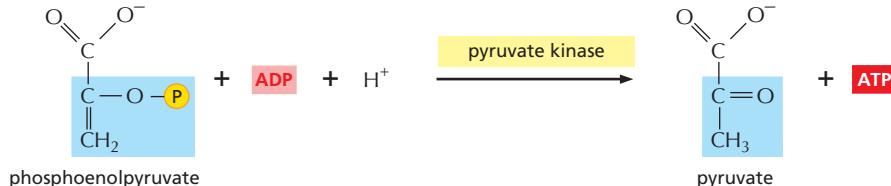
**Step 8** The remaining phosphate ester linkage in 3-phosphoglycerate, which has a relatively low free energy of hydrolysis, is moved from carbon 3 to carbon 2 to form 2-phosphoglycerate.



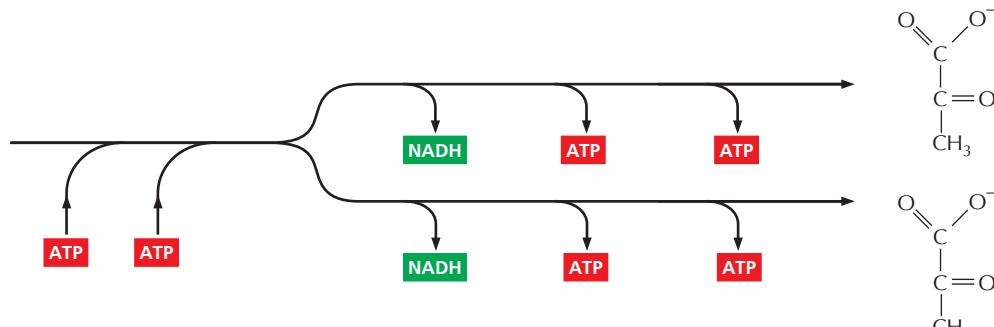
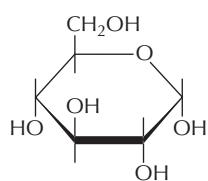
**Step 9** The removal of water from 2-phosphoglycerate creates a high-energy enol phosphate linkage.

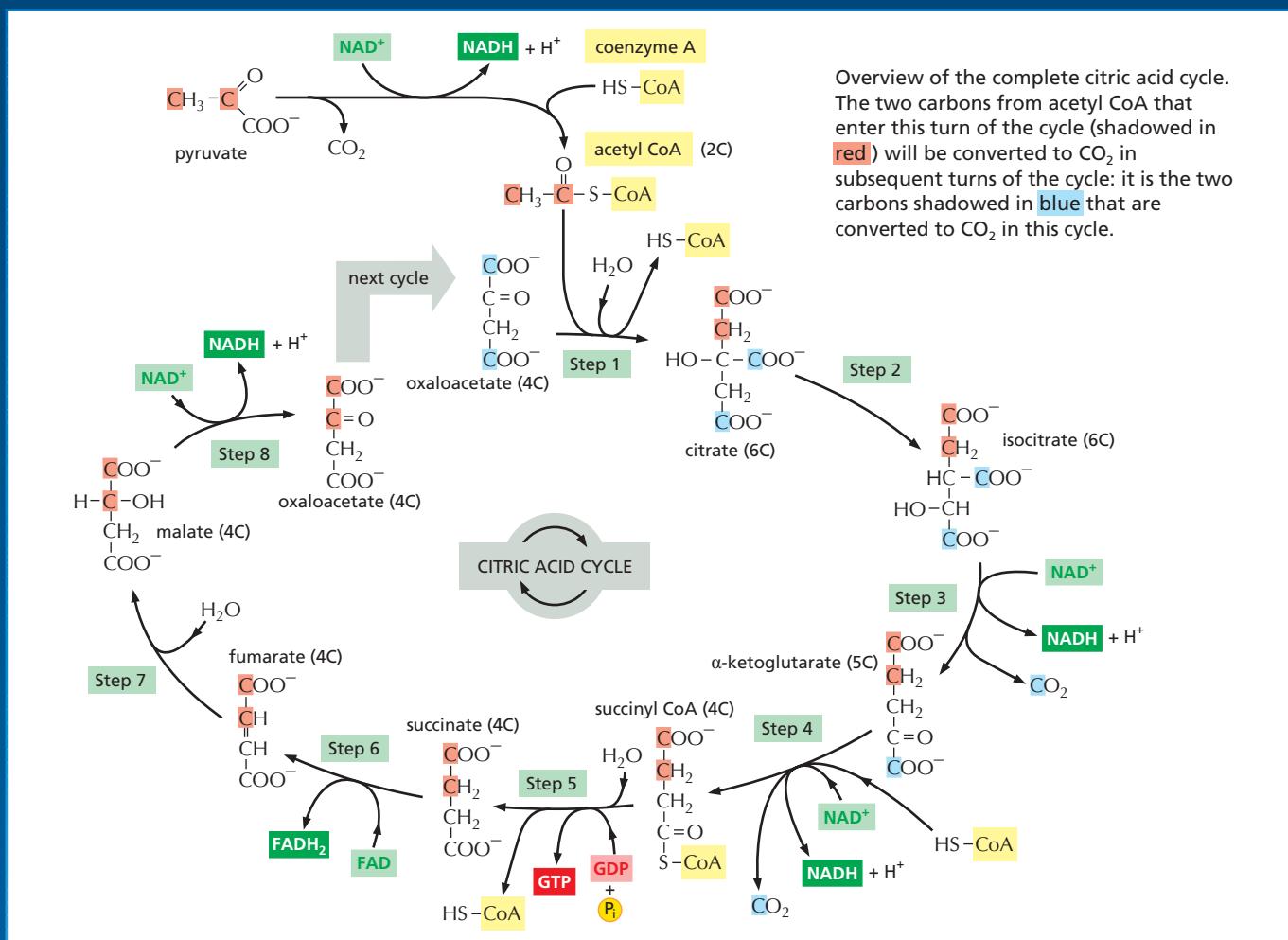


**Step 10** The transfer to ADP of the high-energy phosphate group that was generated in step 9 forms ATP, completing glycolysis.

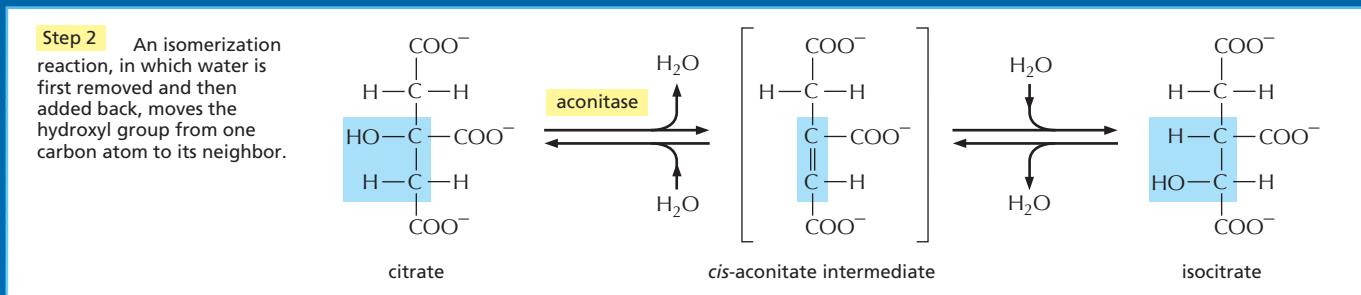
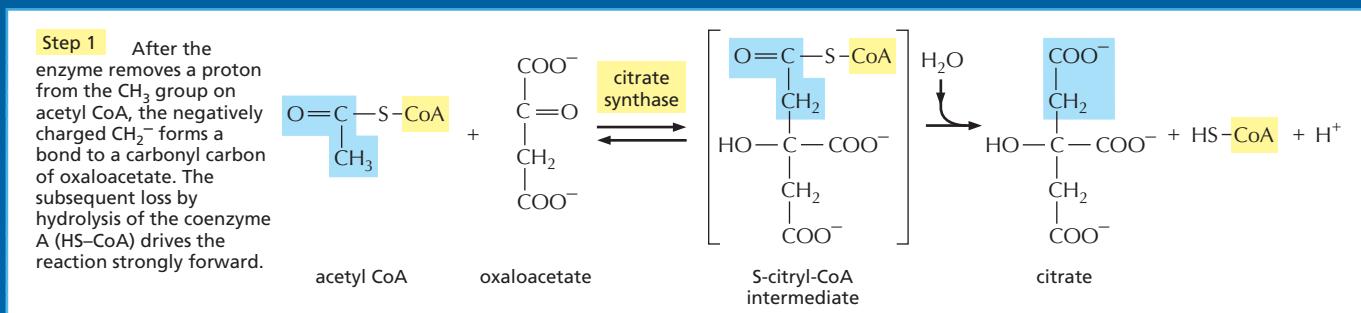


#### NET RESULT OF GLYCOLYSIS

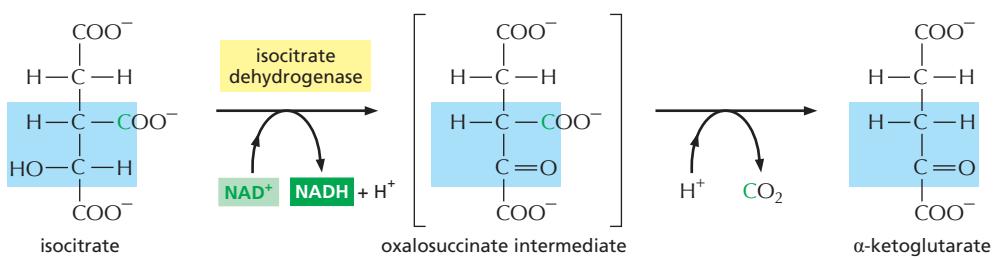




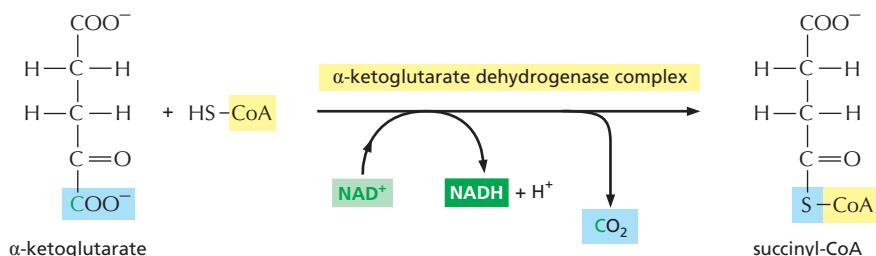
Details of these eight steps are shown below. In this part of the panel, for each step, the part of the molecule that undergoes a change is shadowed in blue, and the name of the enzyme that catalyzes the reaction is in a yellow box.



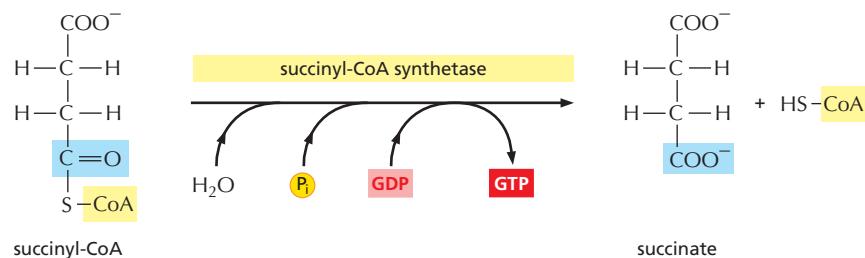
**Step 3** In the first of four oxidation steps in the cycle, the carbon carrying the hydroxyl group is converted to a carbonyl group. The immediate product is unstable, losing  $\text{CO}_2$  while still bound to the enzyme.



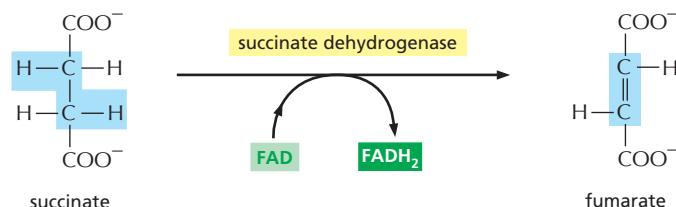
**Step 4** The  $\alpha$ -ketoglutarate dehydrogenase complex closely resembles the large enzyme complex that converts pyruvate to acetyl CoA, the pyruvate dehydrogenase complex in Figure 13–10. It likewise catalyzes an oxidation that produces NADH,  $\text{CO}_2$ , and a high-energy thioester bond to coenzyme A (CoA).



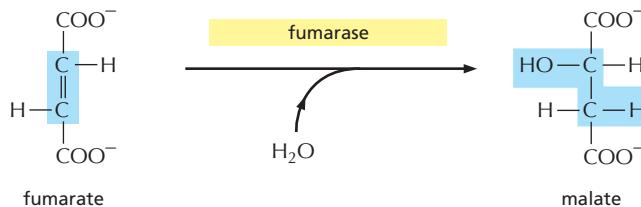
**Step 5** A phosphate molecule from solution displaces the CoA, forming a high-energy phosphate linkage to succinate. This phosphate is then passed to GDP to form GTP. (In bacteria and plants, ATP is formed instead.)



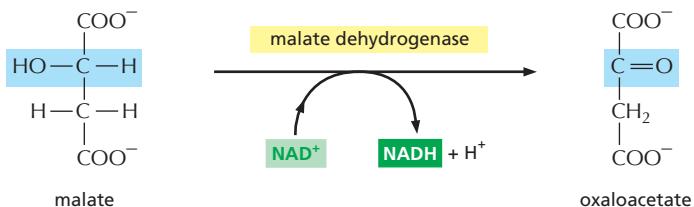
**Step 6** In the third oxidation step in the cycle, FAD accepts two hydrogen atoms from succinate.



**Step 7** The addition of water to fumarate places a hydroxyl group next to a carbonyl carbon.



**Step 8** In the last of four oxidation steps in the cycle, the carbon carrying the hydroxyl group is converted to a carbonyl group, regenerating the oxaloacetate needed for step 1.



## REFERENCES

### General

- Berg JM, Tymoczko JL & Stryer L (2011) Biochemistry, 7th ed. New York: WH Freeman.
- Garrett RH & Grisham CM (2012) Biochemistry, 5th ed. Philadelphia: Thomson Brooks/Cole.
- Moran LA, Horton HR, Scrimgeour G & Perry M (2011) Principles of Biochemistry, 5th ed. Upper Saddle River, NJ: Prentice Hall.
- Nelson DL & Cox MM (2012) Lehninger Principles of Biochemistry, 6th ed. New York: Worth.
- van Holde KE, Johnson WC & Ho PS (2005) Principles of Physical Biochemistry, 2nd ed. Upper Saddle River, NJ: Prentice Hall.
- Van Vranken D & Weiss G (2013) Introduction to Bioorganic Chemistry and Chemical Biology. New York: Garland Science.
- Voet D, Voet JG & Pratt CM (2012) Fundamentals of Biochemistry, 4th ed. New York: Wiley.

### The Chemical Components of a Cell

- Atkins PW (2003) Molecules, 2nd ed. New York: WH Freeman.
- Baldwin RL (2014) Dynamic hydration shell restores Kauzmann's 1959 explanation of how the hydrophobic factor drives protein folding. *Proc. Natl Acad. Sci. USA* 111, 13052–13056.
- Bloomfield VA, Crothers DM & Tinoco I (2000) Nucleic Acids: Structures, Properties, and Functions. Sausalito, CA: University Science Books.
- Branden C & Tooze J (1999) Introduction to Protein Structure, 2nd ed. New York: Garland Science.
- de Duve C (2005) Singularities: Landmarks on the Pathways of Life. Cambridge: Cambridge University Press.
- Dowhan W (1997) Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annu. Rev. Biochem.* 66, 199–232.
- Eisenberg D & Kauzmann W (1969) The Structure and Properties of Water. Oxford: Oxford University Press.
- Franks F (1993) Water. Cambridge: Royal Society of Chemistry.
- Henderson LJ (1927) The Fitness of the Environment, 1958 ed. Boston, MA: Beacon.
- Neidhardt FC, Ingraham JL & Schaechter M (1990) Physiology of the Bacterial Cell: A Molecular Approach. Sunderland, MA: Sinauer.
- Phillips R & Milo R (2009) A feeling for the numbers in biology. *Proc. Natl Acad. Sci. USA* 106, 21465–21471.
- Skinner JL (2010) Following the motions of water molecules in aqueous solutions. *Science* 328, 985–986.
- Taylor ME & Drickamer K (2011) Introduction to Glycobiology, 3rd ed. New York: Oxford University Press.
- Vance DE & Vance JE (2008) Biochemistry of Lipids, Lipoproteins and Membranes, 5th ed. Amsterdam: Elsevier.

### Catalysis and the Use of Energy by Cells

- Atkins PW (1994) The Second Law: Energy, Chaos and Form. New York: Scientific American Books.
- Atkins PW & De Paula JD (2011) Physical Chemistry for the Life Sciences, 2nd ed. Oxford: Oxford University Press.
- Baldwin JE & Krebs H (1981) The evolution of metabolic cycles. *Nature* 291, 381–382.
- Berg HC (1983) Random Walks in Biology. Princeton, NJ: Princeton University Press.
- Dill KA & Bromberg S (2010) Molecular Driving Forces: Statistical Thermodynamics in Biology, Chemistry, Physics, and Nanoscience, 2nd ed. New York: Garland Science.
- Einstein A (1956) Investigations on the Theory of the Brownian Movement. New York: Dover.
- Fruton JS (1999) Proteins, Enzymes, Genes: The Interplay of Chemistry and Biology. New Haven, CT: Yale University Press.

- Hohmann-Marriott MF & Blankenship RE (2011) Evolution of photosynthesis. *Annu. Rev. Plant Biol.* 62, 515–548.
- Karplus M & Petsko GA (1990) Molecular dynamics simulations in biology. *Nature* 347, 631–639.
- Kauzmann W (1967) Thermodynamics and Statistics: with Applications to Gases. In Thermal Properties of Matter, Vol 2. New York: WA Benjamin, Inc.
- Kornberg A (1989) For the Love of Enzymes. Cambridge, MA: Harvard University Press.
- Lipmann F (1941) Metabolic generation and utilization of phosphate bond energy. *Adv. Enzymol.* 1, 99–162.
- Lipmann F (1971) Wanderings of a Biochemist. New York: Wiley.
- Nisbet EG & Sleep NH (2001) The habitat and nature of early life. *Nature* 409, 1083–1091.
- Racker E (1980) From Pasteur to Mitchell: a hundred years of bioenergetics. *Fed. Proc.* 39, 210–215.
- Schrödinger E (1944 & 1958) What is Life? The Physical Aspect of the Living Cell and Mind and Matter, 1992 combined ed. Cambridge: Cambridge University Press.
- van Meer G, Voelker DR & Feigenson GW (2008) Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* 9, 112–124.
- Walsh C (2001) Enabling the chemistry of life. *Nature* 409, 226–231.
- Westheimer FH (1987) Why nature chose phosphates. *Science* 235, 1173–1178.
- ### How Cells Obtain Energy from Food
- Caetano-Anollés D, Kim KM, Mittenthal JE & Caetano-Anollés G (2011) Proteome evolution and the metabolic origins of translation and cellular life. *J. Mol. Evol.* 72, 14–33.
- Cramer WA & Knaff DB (1990) Energy Transduction in Biological Membranes. New York: Springer-Verlag.
- Dismukes GC, Klimov WV, Baranov SV et al. (2001) The origin of atmospheric oxygen on Earth: the innovation of oxygenic photosynthesis. *Proc. Natl Acad. Sci. USA* 98, 2170–2175.
- Fell D (1997) Understanding the Control of Metabolism. London: Portland Press.
- Fothergill-Gilmore LA (1986) The evolution of the glycolytic pathway. *Trends Biochem. Sci.* 11, 47–51.
- Friedmann HC (2004) From *Butyribacterium* to *E. coli*: an essay on unity in biochemistry. *Perspect. Biol. Med.* 47, 47–66.
- Heinrich R, Meléndez-Hevia E, Montero F et al. (1999) The structural design of glycolysis: an evolutionary approach. *Biochem. Soc. Trans.* 27, 294–298.
- Huyuen MA, Dandekar T & Bork P (1999) Variation and evolution of the citric-acid cycle: a genomic perspective. *Trends Microbiol.* 7, 281–291.
- Koonin EV (2014) The origins of cellular life. *Antonie van Leeuwenhoek* 106, 27–41.
- Kornberg HL (2000) Krebs and his trinity of cycles. *Nat. Rev. Mol. Cell Biol.* 1, 225–228.
- Krebs HA (1970) The history of the tricarboxylic acid cycle. *Perspect. Biol. Med.* 14, 154–170.
- Krebs HA & Martin A (1981) Reminiscences and Reflections. Oxford/New York: Clarendon Press/Oxford University Press.
- Lane N & Martin WF (2012) The origin of membrane bioenergetics. *Cell* 151, 1406–1416.
- Morowitz HJ (1993) Beginnings of Cellular Life: Metabolism Recapitulates Biogenesis. New Haven, CT: Yale University Press.
- Martijn J & Ettema TJG (2013) From archaeon to eukaryote: the evolutionary dark ages of the eukaryotic cell. *Biochem. Soc. Trans.* 41, 451–457.
- Mulkidjanian AY, Bychkov AY, Dibrova DV et al. (2012) Open questions on the origin of life at anoxic geothermal fields. *Orig. Life Evol. Biosph.* 42, 507–516.
- Zimmer C (2009) On the origin of eukaryotes. *Science* 325, 665–668.

## CHAPTER

## 3

# Proteins

When we look at a cell through a microscope or analyze its electrical or biochemical activity, we are, in essence, observing proteins. Proteins constitute most of a cell's dry mass. They are not only the cell's building blocks; they also execute the majority of the cell's functions. Thus, proteins that are enzymes provide the intricate molecular surfaces inside a cell that catalyze its many chemical reactions. Proteins embedded in the plasma membrane form channels and pumps that control the passage of small molecules into and out of the cell. Other proteins carry messages from one cell to another, or act as signal integrators that relay sets of signals inward from the plasma membrane to the cell nucleus. Yet others serve as tiny molecular machines with moving parts: *kinesin*, for example, propels organelles through the cytoplasm; *topoisomerase* can untangle knotted DNA molecules. Other specialized proteins act as antibodies, toxins, hormones, antifreeze molecules, elastic fibers, ropes, or sources of luminescence. Before we can hope to understand how genes work, how muscles contract, how nerves conduct electricity, how embryos develop, or how our bodies function, we must attain a deep understanding of proteins.

## THE SHAPE AND STRUCTURE OF PROTEINS

From a chemical point of view, proteins are by far the most structurally complex and functionally sophisticated molecules known. This is perhaps not surprising, once we realize that the structure and chemistry of each protein has been developed and fine-tuned over billions of years of evolutionary history. The theoretical calculations of population geneticists reveal that, over evolutionary time periods, a surprisingly small selective advantage is enough to cause a randomly altered protein sequence to spread through a population of organisms. Yet, even to experts, the remarkable versatility of proteins can seem truly amazing.

In this section, we consider how the location of each amino acid in the long string of amino acids that forms a protein determines its three-dimensional shape. Later in the chapter, we use this understanding of protein structure at the atomic level to describe how the precise shape of each protein molecule determines its function in a cell.

### The Shape of a Protein Is Specified by Its Amino Acid Sequence

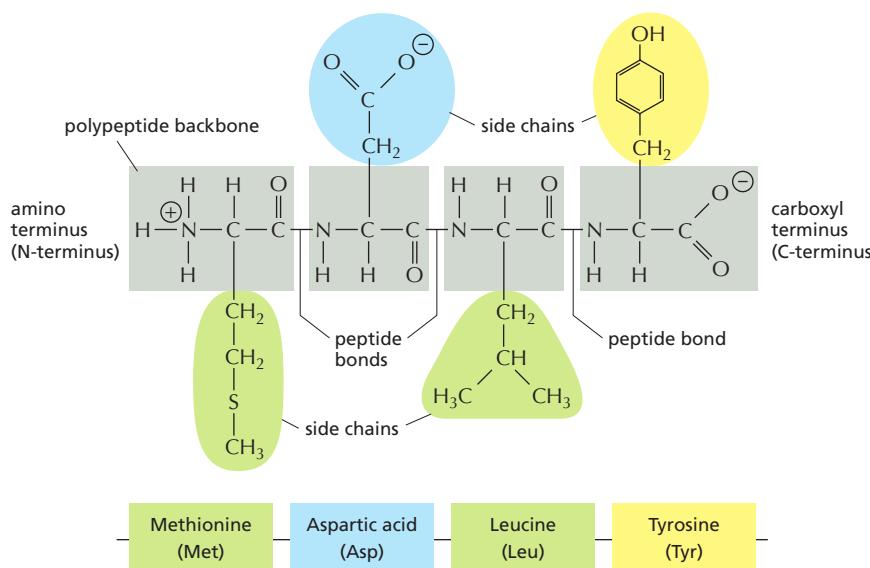
There are 20 different of amino acids in proteins that are coded for directly in an organism's DNA, each with different chemical properties. A **protein** molecule is made from a long unbranched chain of these amino acids, each linked to its neighbor through a covalent peptide bond. Proteins are therefore also known as *polypeptides*. Each type of protein has a unique sequence of amino acids, and there are many thousands of different proteins in a cell.

The repeating sequence of atoms along the core of the polypeptide chain is referred to as the **polypeptide backbone**. Attached to this repetitive chain are those portions of the amino acids that are not involved in making a peptide bond and that give each amino acid its unique properties: the 20 different amino acid side chains (**Figure 3-1**). Some of these side chains are nonpolar and hydrophobic

### IN THIS CHAPTER

#### THE SHAPE AND STRUCTURE OF PROTEINS

#### PROTEIN FUNCTION



**Figure 3–1** The components of a protein. A protein consists of a polypeptide backbone with attached side chains. Each type of protein differs in its sequence and number of amino acids; therefore, it is the sequence of the chemically different side chains that makes each protein distinct. The two ends of a polypeptide chain are chemically different: the end carrying the free amino group ( $\text{NH}_3^+$ , also written  $\text{NH}_2$ ) is the amino terminus, or N-terminus, and that carrying the free carboxyl group ( $\text{COO}^-$ , also written  $\text{COOH}$ ) is the carboxyl terminus or C-terminus. The amino acid sequence of a protein is always presented in the N-to-C direction, reading from left to right.

(“water-fearing”), others are negatively or positively charged, some readily form covalent bonds, and so on. **Panel 3–1** (pp. 112–113) shows their atomic structures and **Figure 3–2** lists their abbreviations.

As discussed in Chapter 2, atoms behave almost as if they were hard spheres with a definite radius (their *van der Waals radius*). The requirement that no two atoms overlap plus other constraints limit the possible bond angles in a polypeptide chain (**Figure 3–3**), severely restricting the possible three-dimensional arrangements (or *conformations*) of atoms. Nevertheless, a long flexible chain such as a protein can still fold in an enormous number of ways.

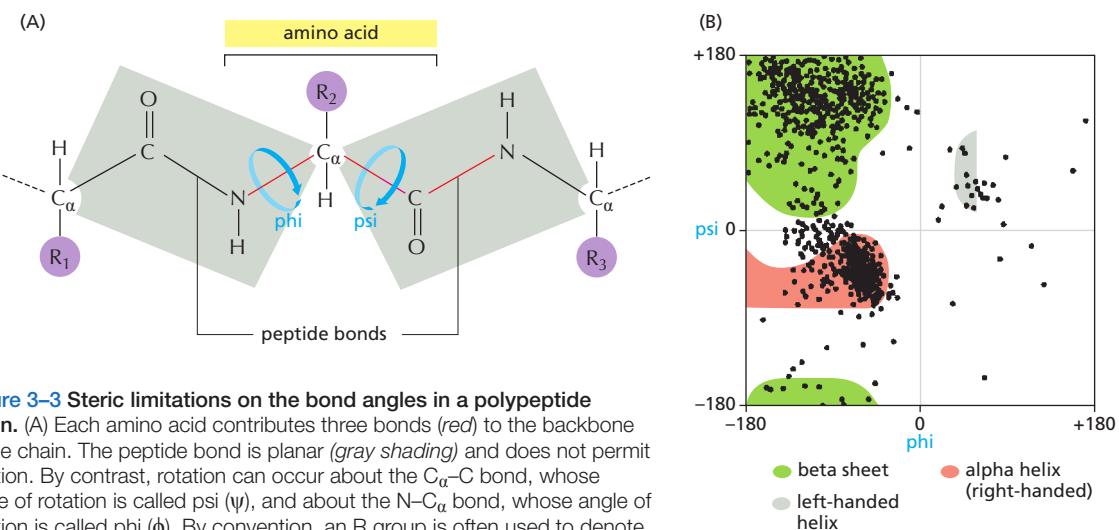
The folding of a protein chain is also determined by many different sets of weak *noncovalent bonds* that form between one part of the chain and another. These involve atoms in the polypeptide backbone, as well as atoms in the amino acid side chains. There are three types of these weak bonds: *hydrogen bonds*, *electrostatic attractions*, and *van der Waals attractions*, as explained in Chapter 2 (see p. 44). Individual noncovalent bonds are 30–300 times weaker than the typical covalent bonds that create biological molecules. But many weak bonds acting in parallel can hold two regions of a polypeptide chain tightly together. In this way, the combined strength of large numbers of such noncovalent bonds determines the stability of each folded shape (**Figure 3–4**).

AMINO ACID		SIDE CHAIN		AMINO ACID		SIDE CHAIN	
Aspartic acid	Asp	D	negative	Alanine	Ala	A	nonpolar
Glutamic acid	Glu	E	negative	Glycine	Gly	G	nonpolar
Arginine	Arg	R	positive	Valine	Val	V	nonpolar
Lysine	Lys	K	positive	Leucine	Leu	L	nonpolar
Histidine	His	H	positive	Isoleucine	Ile	I	nonpolar
Asparagine	Asn	N	uncharged polar	Proline	Pro	P	nonpolar
Glutamine	Gln	Q	uncharged polar	Phenylalanine	Phe	F	nonpolar
Serine	Ser	S	uncharged polar	Methionine	Met	M	nonpolar
Threonine	Thr	T	uncharged polar	Tryptophan	Trp	W	nonpolar
Tyrosine	Tyr	Y	uncharged polar	Cysteine	Cys	C	nonpolar

POLAR AMINO ACIDS

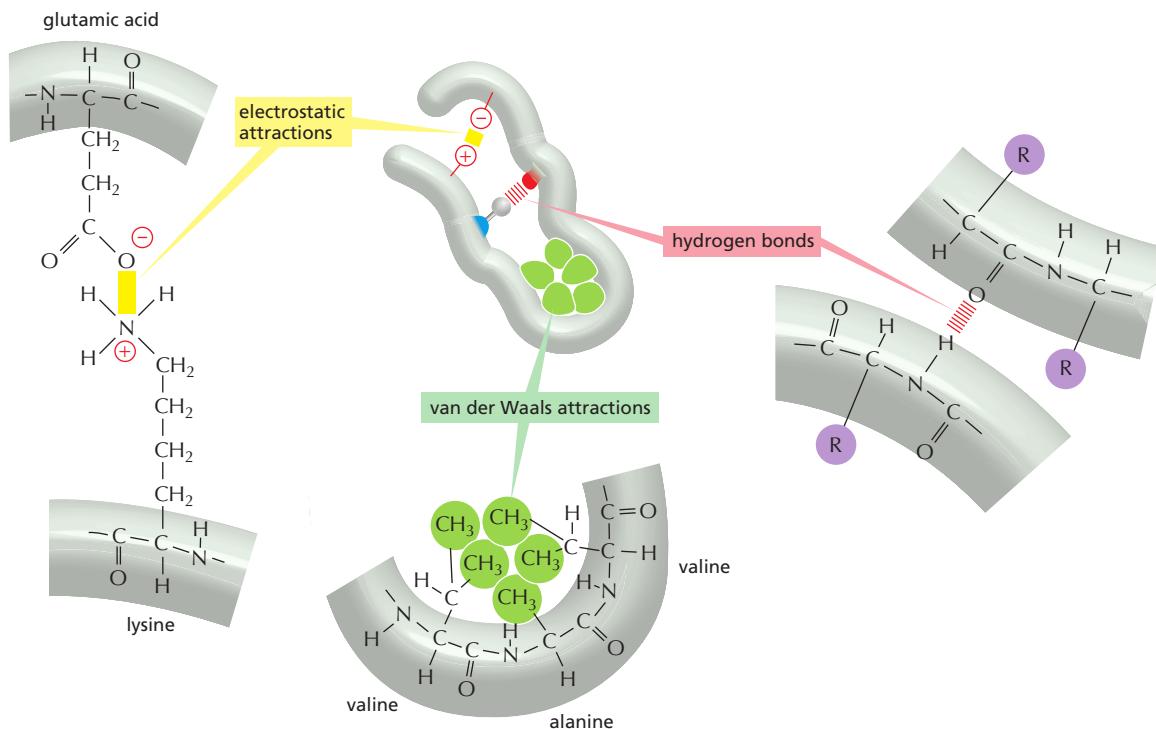
NONPOLAR AMINO ACIDS

**Figure 3–2** The 20 amino acids commonly found in proteins. Each amino acid has a three-letter and a one-letter abbreviation. There are equal numbers of polar and nonpolar side chains; however, some side chains listed here as polar are large enough to have some nonpolar properties (for example, Tyr, Thr, Arg, Lys). For atomic structures, see Panel 3–1 (pp. 112–113).



**Figure 3-3 Steric limitations on the bond angles in a polypeptide chain.** (A) Each amino acid contributes three bonds (red) to the backbone of the chain. The peptide bond is planar (gray shading) and does not permit rotation. By contrast, rotation can occur about the  $C_\alpha-C$  bond, whose angle of rotation is called  $\psi$  (psi), and about the  $N-C_\alpha$  bond, whose angle of rotation is called  $\phi$  (phi). By convention, an R group is often used to denote an amino acid side chain (purple circles). (B) The conformation of the main-chain atoms in a protein is determined by one pair of  $\phi$  and  $\psi$  angles for each amino acid; because of steric collisions between atoms within each amino acid, most of the possible pairs of  $\phi$  and  $\psi$  angles do not occur. In this so-called Ramachandran plot, each dot represents an observed pair of angles in a protein. The three differently shaded clusters of dots reflect three different “secondary structures” repeatedly found in proteins, as will be described in the text. (B, from J. Richardson, *Adv. Prot. Chem.* 34:174–175, 1981. © Academic Press.)

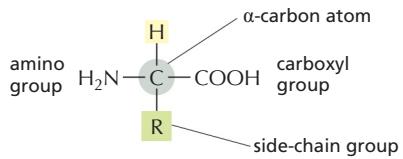
A fourth weak force—a hydrophobic clustering force—also has a central role in determining the shape of a protein. As described in Chapter 2, hydrophobic molecules, including the nonpolar side chains of particular amino acids, tend to be forced together in an aqueous environment in order to minimize their disruptive effect on the hydrogen-bonded network of water molecules (see Panel 2-2, pp. 92–93). Therefore, an important factor governing the folding of any protein is



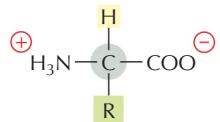
**Figure 3-4 Three types of noncovalent bonds help proteins fold.** Although a single one of these bonds is quite weak, many of them act together to create a strong bonding arrangement, as in the example shown. As in the previous figure, R is used as a general designation for an amino acid side chain.

**THE AMINO ACID**

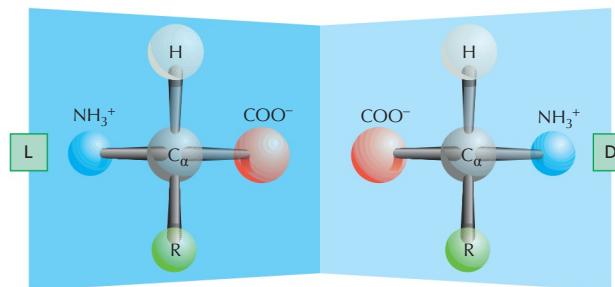
The general formula of an amino acid is



$\text{R}$  is commonly one of 20 different side chains. At pH 7 both the amino and carboxyl groups are ionized.

**OPTICAL ISOMERS**

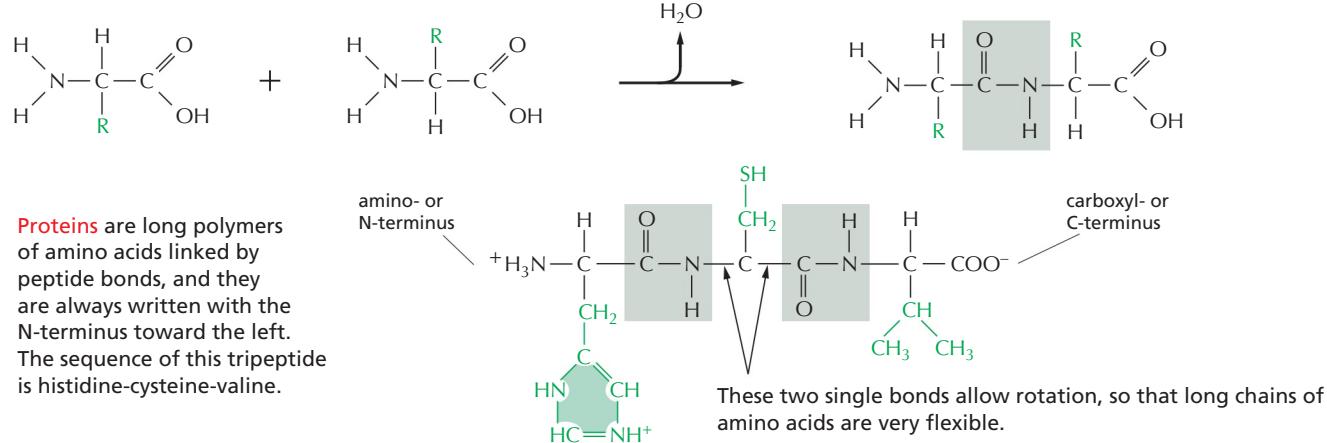
The  $\alpha$ -carbon atom is asymmetric, which allows for two mirror images (or stereoisomers), L and D.



Proteins consist exclusively of L-amino acids.

**PEPTIDE BONDS**

Amino acids are commonly joined together by an amide linkage, called a peptide bond.

**FAMILIES OF AMINO ACIDS**

The common amino acids are grouped according to whether their side chains are

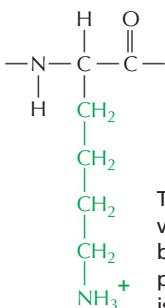
- acidic
- basic
- uncharged polar
- nonpolar

These 20 amino acids are given both three-letter and one-letter abbreviations.

Thus: alanine = Ala = A

**BASIC SIDE CHAINS****lysine**

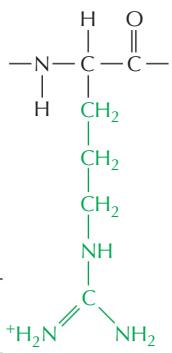
(Lys, or K)



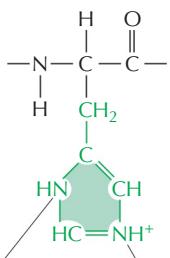
This group is very basic because its positive charge is stabilized by resonance.

**arginine**

(Arg, or R)

**histidine**

(His, or H)

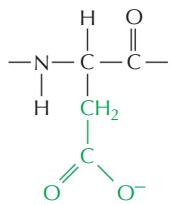


These nitrogens have a relatively weak affinity for an  $\text{H}^+$  and are only partly positive at neutral pH.

## ACIDIC SIDE CHAINS

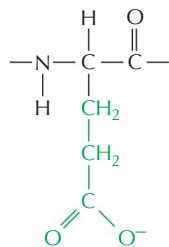
aspartic acid

(Asp, or D)



glutamic acid

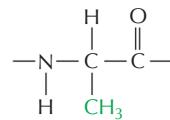
(Glu, or E)



## NONPOLAR SIDE CHAINS

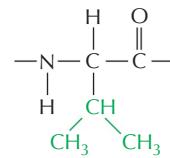
alanine

(Ala, or A)



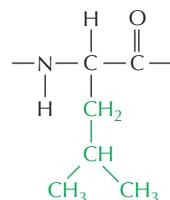
valine

(Val, or V)



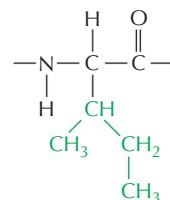
leucine

(Leu, or L)



isoleucine

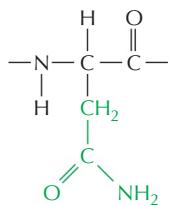
(Ile, or I)



## UNCHARGED POLAR SIDE CHAINS

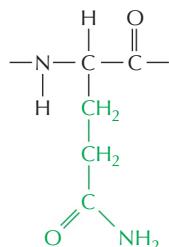
asparagine

(Asn, or N)



glutamine

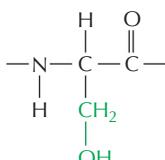
(Gln, or Q)



Although the amide N is not charged at neutral pH, it is polar.

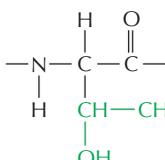
serine

(Ser, or S)



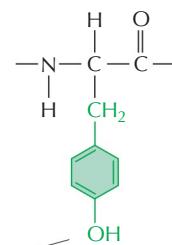
threonine

(Thr, or T)



tyrosine

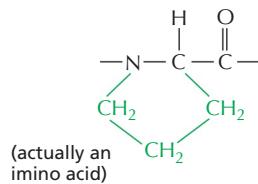
(Tyr, or Y)



The -OH group is polar.

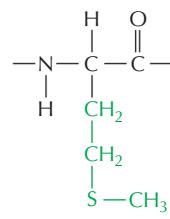
proline

(Pro, or P)



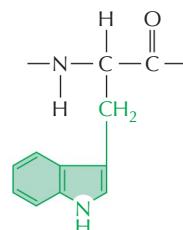
methionine

(Met, or M)



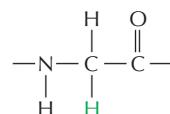
tryptophan

(Trp, or W)



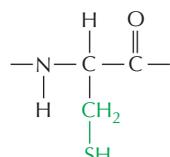
glycine

(Gly, or G)



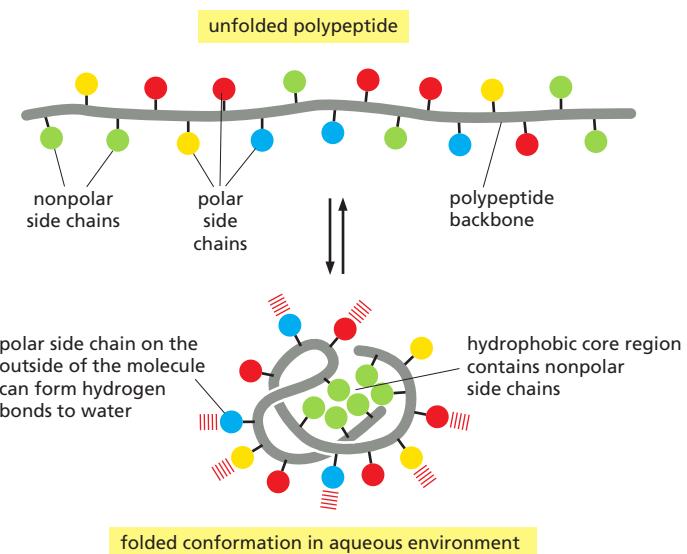
cysteine

(Cys, or C)



Disulfide bonds can form between two cysteine side chains in proteins.





**Figure 3–5** How a protein folds into a compact conformation. The polar amino acid side chains tend to lie on the outside of the protein, where they can interact with water; the nonpolar amino acid side chains are buried on the inside forming a tightly packed hydrophobic core of atoms that are hidden from water. In this schematic drawing, the protein contains only about 35 amino acids.

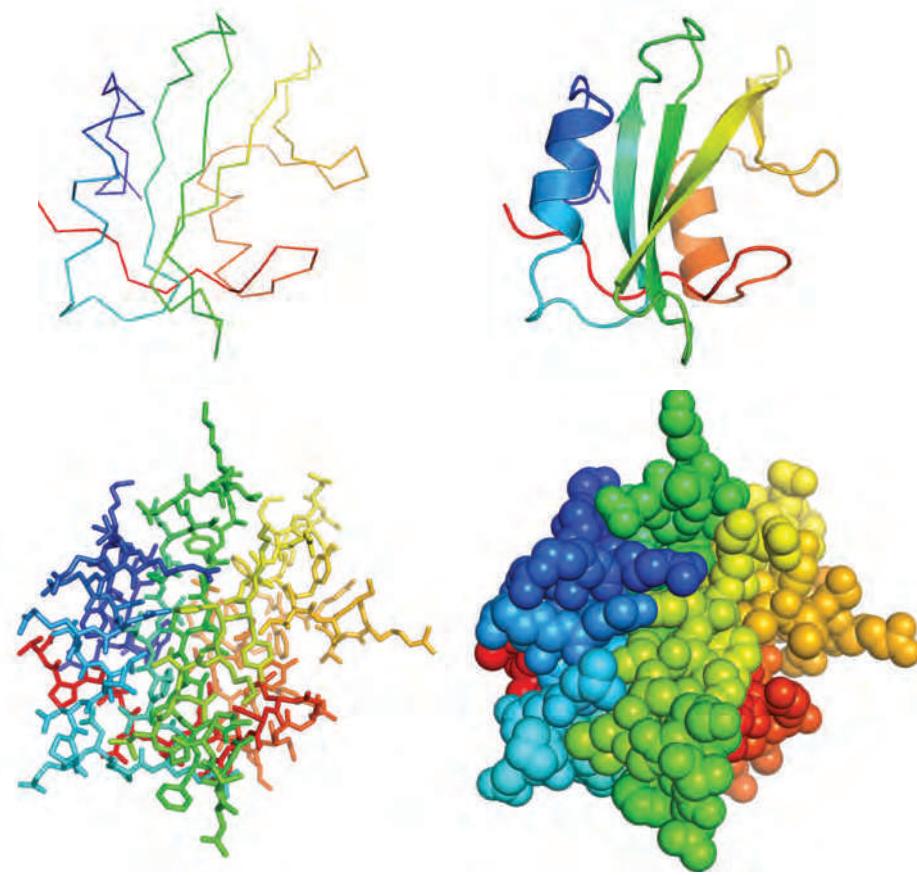
the distribution of its polar and nonpolar amino acids. The nonpolar (hydrophobic) side chains in a protein—belonging to such amino acids as phenylalanine, leucine, valine, and tryptophan—tend to cluster in the interior of the molecule (just as hydrophobic oil droplets coalesce in water to form one large droplet). This enables them to avoid contact with the water that surrounds them inside a cell. In contrast, polar groups—such as those belonging to arginine, glutamine, and histidine—tend to arrange themselves near the outside of the molecule, where they can form hydrogen bonds with water and with other polar molecules (Figure 3–5). Polar amino acids buried within the protein are usually hydrogen-bonded to other polar amino acids or to the polypeptide backbone.

### Proteins Fold into a Conformation of Lowest Energy

As a result of all of these interactions, most proteins have a particular three-dimensional structure, which is determined by the order of the amino acids in its chain. The final folded structure, or **conformation**, of any polypeptide chain is generally the one that minimizes its free energy. Biologists have studied protein folding in a test tube using highly purified proteins. Treatment with certain solvents, which disrupt the noncovalent interactions holding the folded chain together, unfolds, or *denatures*, a protein. This treatment converts the protein into a flexible polypeptide chain that has lost its natural shape. When the denaturing solvent is removed, the protein often refolds spontaneously, or *renatures*, into its original conformation. This indicates that the amino acid sequence contains all of the information needed for specifying the three-dimensional shape of a protein, a critical point for understanding cell biology.

Most proteins fold up into a single stable conformation. However, this conformation changes slightly when the protein interacts with other molecules in the cell. This change in shape is often crucial to the function of the protein, as we see later.

Although a protein chain can fold into its correct conformation without outside help, in a living cell special proteins called *molecular chaperones* often assist in protein folding. Molecular chaperones bind to partly folded polypeptide chains and help them progress along the most energetically favorable folding pathway. In the crowded conditions of the cytoplasm, chaperones are required to prevent the temporarily exposed hydrophobic regions in newly synthesized protein chains from associating with each other to form protein aggregates (see p. 355). However, the final three-dimensional shape of the protein is still specified by its amino acid sequence: chaperones simply make reaching the folded state more reliable.



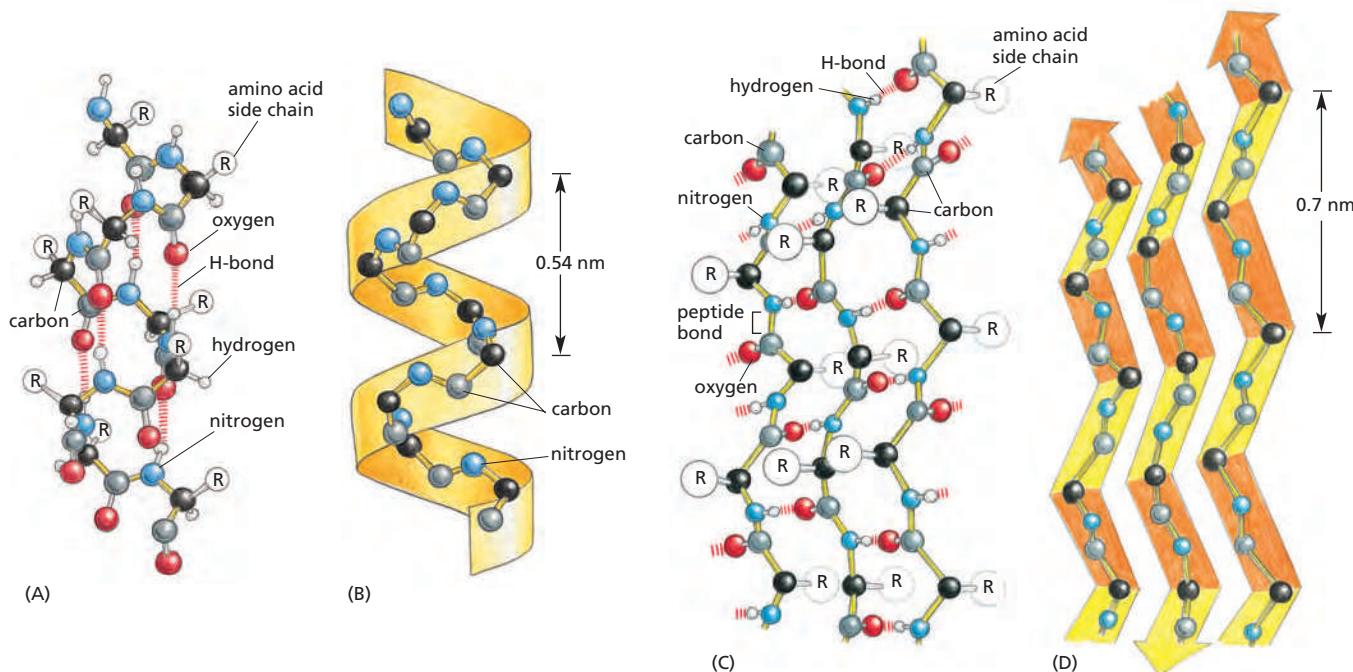
**Figure 3–6** Four representations describing the structure of a small protein domain. Constructed from a string of 100 amino acids, the SH2 domain is part of many different proteins (see, for example, Figure 3–61). Here, the structure of the SH2 domain is displayed as (A) a polypeptide backbone model, (B) a ribbon model, (C) a wire model that includes the amino acid side chains, and (D) a space-filling model ([Movie 3.1](#)). These images are colored in a way that allows the polypeptide chain to be followed from its N-terminus (purple) to its C-terminus (red) (PDB code: 1SHA).

Proteins come in a wide variety of shapes, and most are between 50 and 2000 amino acids long. Large proteins usually consist of several distinct *protein domains*—structural units that fold more or less independently of each other, as we discuss below. The structure of even a small domain is complex, and for clarity, several different representations are conventionally used, each of which emphasizes distinct features. As an example, **Figure 3–6** presents four representations of a protein domain called SH2, a structure present in many different proteins in eukaryotic cells and involved in cell signaling (see Figure 15–46).

Descriptions of protein structures are aided by the fact that proteins are built up from combinations of several common structural motifs, as we discuss next.

### The $\alpha$ Helix and the $\beta$ Sheet Are Common Folding Patterns

When we compare the three-dimensional structures of many different protein molecules, it becomes clear that, although the overall conformation of each protein is unique, two regular folding patterns are often found within them. Both patterns were discovered more than 60 years ago from studies of hair and silk. The first folding pattern to be discovered, called the  $\alpha$  helix, was found in the protein  *$\alpha$ -keratin*, which is abundant in skin and its derivatives—such as hair, nails, and horns. Within a year of the discovery of the  $\alpha$  helix, a second folded structure, called a  $\beta$  sheet, was found in the protein *fibroin*, the major constituent of silk. These two patterns are particularly common because they result from hydrogen-bonding between the N-H and C=O groups in the polypeptide backbone, without involving the side chains of the amino acids. Thus, although incompatible with some amino acid side chains, many different amino acid sequences can form them. In each case, the protein chain adopts a regular, repeating conformation. **Figure 3–7** illustrates the detailed structures of these two important conformations, which in ribbon models of proteins are represented by a helical ribbon and by a set of aligned arrows, respectively.



**Figure 3-7** The regular conformation of the polypeptide backbone in the  $\alpha$  helix and the  $\beta$  sheet. The  $\alpha$  helix is shown in (A) and (B). The N-H of every peptide bond is hydrogen-bonded to the C=O of a neighboring peptide bond located four peptide bonds away in the same chain. Note that all of the N-H groups point up in this diagram and that all of the C=O groups point down (toward the C-terminus); this gives a polarity to the helix, with the C-terminus having a partial negative and the N-terminus a partial positive charge (Movie 3.2). The  $\beta$  sheet is shown in (C) and (D). In this example, adjacent peptide chains run in opposite (antiparallel) directions. Hydrogen-bonding between peptide bonds in different strands holds the individual polypeptide chains (strands) together in a  $\beta$  sheet, and the amino acid side chains in each strand alternately project above and below the plane of the sheet (Movie 3.3). (A) and (C) show all the atoms in the polypeptide backbone, but the amino acid side chains are truncated and denoted by R. In contrast, (B) and (D) show only the carbon and nitrogen backbone atoms.

The cores of many proteins contain extensive regions of  $\beta$  sheet. As shown in Figure 3-8, these  $\beta$  sheets can form either from neighboring segments of the polypeptide backbone that run in the same orientation (parallel chains) or from a polypeptide backbone that folds back and forth upon itself, with each section of the chain running in the direction opposite to that of its immediate neighbors (antiparallel chains). Both types of  $\beta$  sheet produce a very rigid structure, held together by hydrogen bonds that connect the peptide bonds in neighboring chains (see Figure 3-7C).

An  $\alpha$  helix is generated when a single polypeptide chain twists around on itself to form a rigid cylinder. A hydrogen bond forms between every fourth peptide bond, linking the C=O of one peptide bond to the N-H of another (see Figure 3-7A). This gives rise to a regular helix with a complete turn every 3.6 amino acids. The SH2 protein domain illustrated in Figure 3-6 contains two  $\alpha$  helices, as well as a three-stranded antiparallel  $\beta$  sheet.

Regions of  $\alpha$  helix are abundant in proteins located in cell membranes, such as transport proteins and receptors. As we discuss in Chapter 10, those portions of a transmembrane protein that cross the lipid bilayer usually cross as  $\alpha$  helices composed largely of amino acids with nonpolar side chains. The polypeptide backbone, which is hydrophilic, is hydrogen-bonded to itself in the  $\alpha$  helix and shielded from the hydrophobic lipid environment of the membrane by its protruding nonpolar side chains (see also Figure 3-7A).

In other proteins,  $\alpha$  helices wrap around each other to form a particularly stable structure, known as a **coiled-coil**. This structure can form when the two (or in some cases, three or four)  $\alpha$  helices have most of their nonpolar (hydrophobic) side chains on one side, so that they can twist around each other with these side chains facing inward (Figure 3-9). Long rodlike coiled-coils provide the structural

framework for many elongated proteins. Examples are  $\alpha$ -keratin, which forms the intracellular fibers that reinforce the outer layer of the skin and its appendages, and the myosin molecules responsible for muscle contraction.

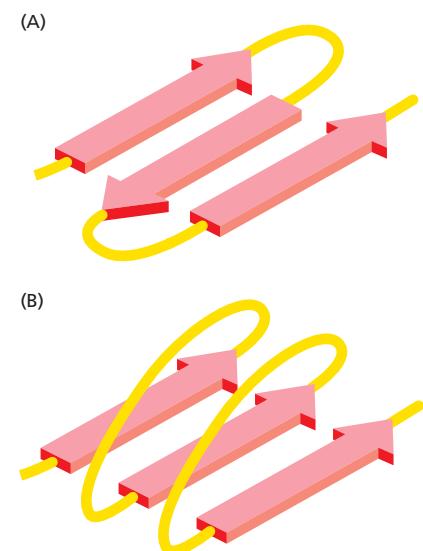
### Protein Domains Are Modular Units from Which Larger Proteins Are Built

Even a small protein molecule is built from thousands of atoms linked together by precisely oriented covalent and noncovalent bonds. Biologists are aided in visualizing these extremely complicated structures by various graphic and computer-based three-dimensional displays. The student resource site that accompanies this book contains computer-generated images of selected proteins, displayed and rotated on the screen in a variety of formats.

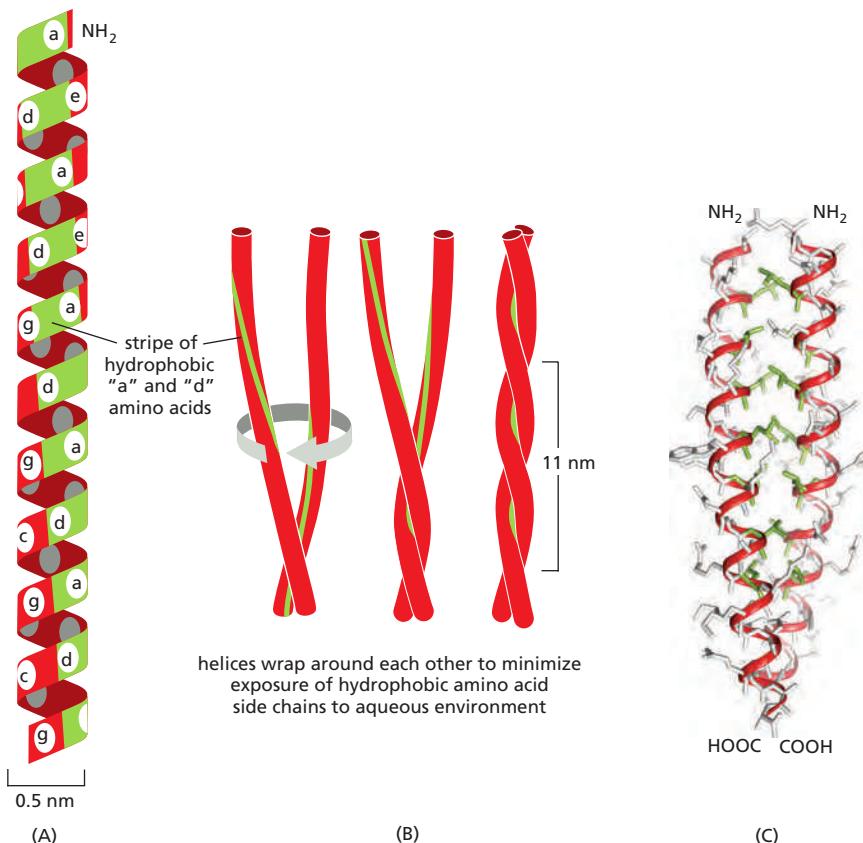
Scientists distinguish four levels of organization in the structure of a protein. The amino acid sequence is known as the **primary structure**. Stretches of polypeptide chain that form  $\alpha$  helices and  $\beta$  sheets constitute the protein's **secondary structure**. The full three-dimensional organization of a polypeptide chain is sometimes referred to as the **tertiary structure**, and if a particular protein molecule is formed as a complex of more than one polypeptide chain, the complete structure is designated as the **quaternary structure**.

Studies of the conformation, function, and evolution of proteins have also revealed the central importance of a unit of organization distinct from these four. This is the **protein domain**, a substructure produced by any contiguous part of a polypeptide chain that can fold independently of the rest of the protein into a compact, stable structure. A domain usually contains between 40 and 350 amino acids, and it is the modular unit from which many larger proteins are constructed.

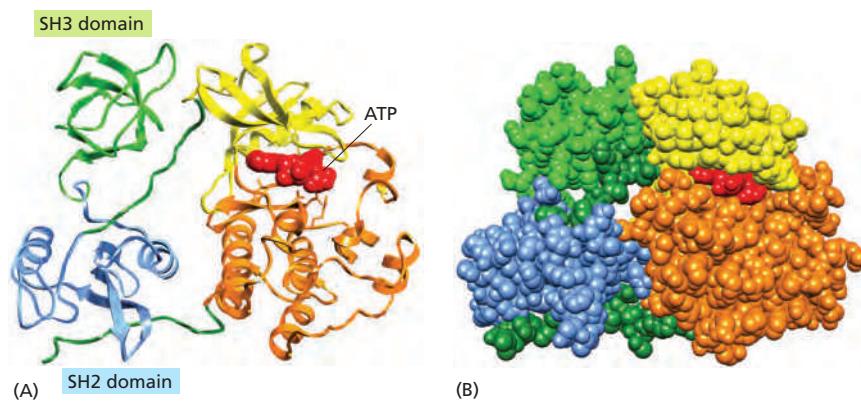
The different domains of a protein are often associated with different functions. **Figure 3–10** shows an example—the Src protein kinase, which functions in signaling pathways inside vertebrate cells (Src is pronounced “sarc”). This protein



**Figure 3–8** Two types of  $\beta$  sheet structures. (A) An antiparallel  $\beta$  sheet (see Figure 3–7C). (B) A parallel  $\beta$  sheet. Both of these structures are common in proteins.



**Figure 3–9** A coiled-coil. (A) A single  $\alpha$  helix, with successive amino acid side chains labeled in a sevenfold sequence, “abcdefg” (from bottom to top). Amino acids “a” and “d” in such a sequence lie close together on the cylinder surface, forming a “stripe” (green) that winds slowly around the  $\alpha$  helix. Proteins that form coiled-coils typically have nonpolar amino acids at positions “a” and “d.” Consequently, as shown in (B), the two  $\alpha$  helices can wrap around each other with the nonpolar side chains of one  $\alpha$  helix interacting with the nonpolar side chains of the other. (C) The atomic structure of a coiled-coil determined by x-ray crystallography. The alpha helical backbone is shown in red and the nonpolar side chains in green, while the more hydrophilic amino acid side chains, shown in gray, are left exposed to the aqueous environment (**Movie 3.4**). (PDB code: 3NMD.)



**Figure 3–10** A protein formed from multiple domains. In the Src protein shown, a C-terminal domain with two lobes (yellow and orange) forms a protein kinase enzyme, while the SH2 and SH3 domains perform regulatory functions. (A) A ribbon model, with ATP substrate in red. (B) A space-filling model, with ATP substrate in red. Note that the site that binds ATP is positioned at the interface of the two lobes that form the kinase. The structure of the SH2 domain was illustrated in Figure 3–6. (PDB code: 2SRC.)

is considered to have three domains: the SH2 and SH3 domains have regulatory roles, while the C-terminal domain is responsible for the kinase catalytic activity. Later in the chapter, we shall return to this protein, in order to explain how proteins can form molecular switches that transmit information throughout cells.

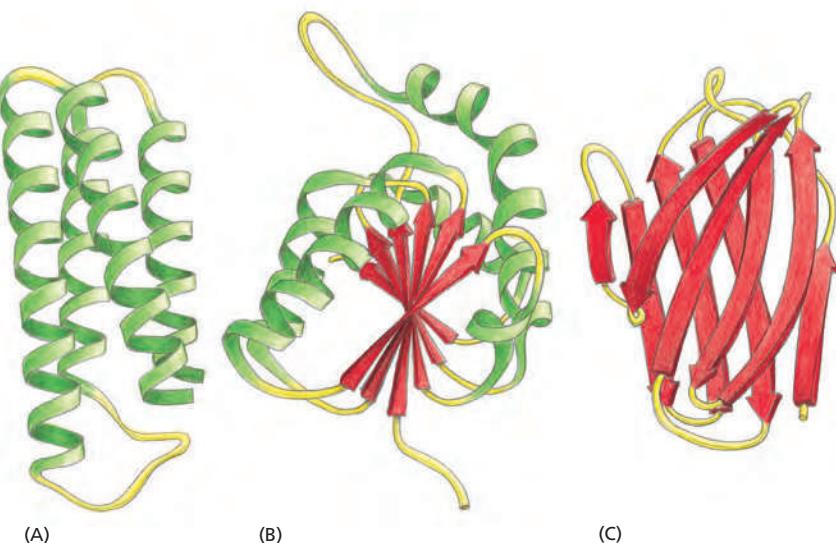
**Figure 3–11** presents ribbon models of three differently organized protein domains. As these examples illustrate, the central core of a domain can be constructed from  $\alpha$  helices, from  $\beta$  sheets, or from various combinations of these two fundamental folding elements.

The smallest protein molecules contain only a single domain, whereas larger proteins can contain several dozen domains, often connected to each other by short, relatively unstructured lengths of polypeptide chain that can act as flexible hinges between domains.

### Few of the Many Possible Polypeptide Chains Will Be Useful to Cells

Since each of the 20 amino acids is chemically distinct and each can, in principle, occur at any position in a protein chain, there are  $20 \times 20 \times 20 \times 20 = 160,000$  different possible polypeptide chains four amino acids long, or  $20^n$  different possible polypeptide chains  $n$  amino acids long. For a typical protein length of about 300 amino acids, a cell could theoretically make more than  $10^{390}$  ( $20^{300}$ ) different polypeptide chains. This is such an enormous number that to produce just one molecule of each kind would require many more atoms than exist in the universe.

Only a very small fraction of this vast set of conceivable polypeptide chains would adopt a stable three-dimensional conformation—by some estimates, less



**Figure 3–11** Ribbon models of three different protein domains. (A) Cytochrome  $b_{562}$ , a single-domain protein involved in electron transport in mitochondria. This protein is composed almost entirely of  $\alpha$  helices. (B) The NAD-binding domain of the enzyme lactic dehydrogenase, which is composed of a mixture of  $\alpha$  helices and parallel  $\beta$  sheets. (C) The variable domain of an immunoglobulin (antibody) light chain, composed of a sandwich of two antiparallel  $\beta$  sheets. In these examples, the  $\alpha$  helices are shown in green, while strands organized as  $\beta$  sheets are denoted by red arrows. Note how the polypeptide chain generally traverses back and forth across the entire domain, making sharp turns only at the protein surface (Movie 3.5). It is the protruding loop regions (yellow) that often form the binding sites for other molecules. (Adapted from drawings courtesy of Jane Richardson.)

than one in a billion. And yet the majority of proteins present in cells do adopt unique and stable conformations. How is this possible? The answer lies in natural selection. A protein with an unpredictably variable structure and biochemical activity is unlikely to help the survival of a cell that contains it. Such proteins would therefore have been eliminated by natural selection through the enormously long trial-and-error process that underlies biological evolution.

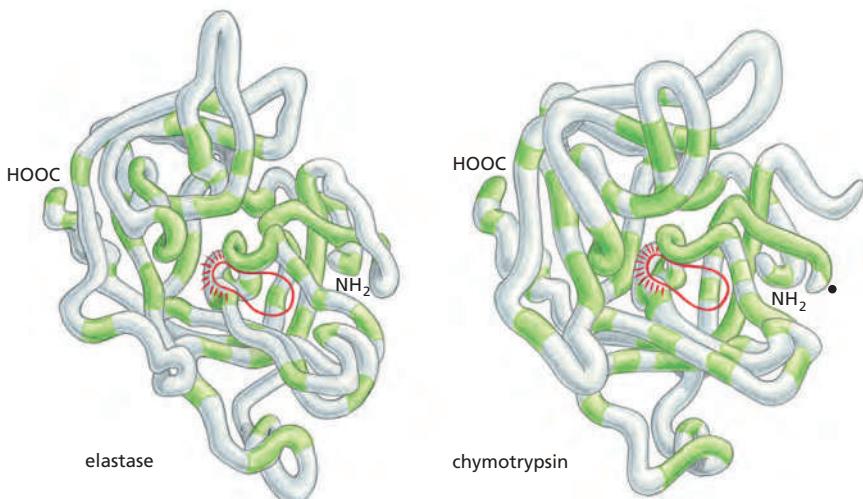
Because evolution has selected for protein function in living organisms, the amino acid sequence of most present-day proteins is such that a single conformation is stable. In addition, this conformation has its chemical properties finely tuned to enable the protein to perform a particular catalytic or structural function in the cell. Proteins are so precisely built that the change of even a few atoms in one amino acid can sometimes disrupt the structure of the whole molecule so severely that all function is lost. And, as discussed later in this chapter, when certain rare protein misfolding accidents occur, the results can be disastrous for the organisms that contain them.

### Proteins Can Be Classified into Many Families

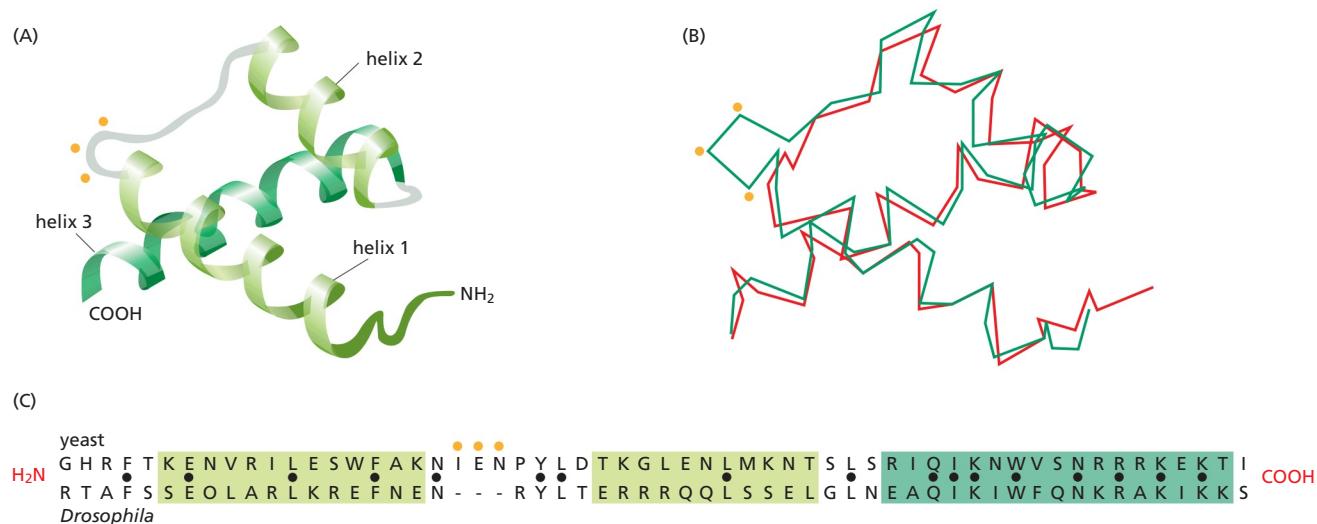
Once a protein had evolved that folded up into a stable conformation with useful properties, its structure could be modified during evolution to enable it to perform new functions. This process has been greatly accelerated by genetic mechanisms that occasionally duplicate genes, allowing one gene copy to evolve independently to perform a new function (discussed in Chapter 4). This type of event has occurred very often in the past; as a result, many present-day proteins can be grouped into protein families, each family member having an amino acid sequence and a three-dimensional conformation that resemble those of the other family members.

Consider, for example, the *serine proteases*, a large family of protein-cleaving (proteolytic) enzymes that includes the digestive enzymes chymotrypsin, trypsin, and elastase, and several proteases involved in blood clotting. When the protease portions of any two of these enzymes are compared, parts of their amino acid sequences are found to match. The similarity of their three-dimensional conformations is even more striking: most of the detailed twists and turns in their polypeptide chains, which are several hundred amino acids long, are virtually identical (**Figure 3–12**). The many different serine proteases nevertheless have distinct enzymatic activities, each cleaving different proteins or the peptide bonds between different types of amino acids. Each therefore performs a distinct function in an organism.

The story we have told for the serine proteases could be repeated for hundreds of other protein families. In general, the structure of the different members of a



**Figure 3–12** A comparison of the conformations of two serine proteases. The backbone conformations of elastase and chymotrypsin. Although only those amino acids in the polypeptide chain shaded in green are the same in the two proteins, the two conformations are very similar nearly everywhere. The active site of each enzyme is circled in red; this is where the peptide bonds of the proteins that serve as substrates are bound and cleaved by hydrolysis. The serine proteases derive their name from the amino acid serine, whose side chain is part of the active site of each enzyme and directly participates in the cleavage reaction. The two dots on the right side of the chymotrypsin molecule mark the new ends created when this enzyme cuts its own backbone.



**Figure 3–13** A comparison of a class of DNA-binding domains, called homeodomains, in a pair of proteins from two organisms separated by more than a billion years of evolution. (A) A ribbon model of the structure common to both proteins. (B) A trace of the  $\alpha$ -carbon positions. The three-dimensional structures shown were determined by x-ray crystallography for the yeast  $\alpha 2$  protein (green) and the *Drosophila* engrailed protein (red). (C) A comparison of amino acid sequences for the region of the proteins shown in (A) and (B). Black dots mark sites with identical amino acids. Orange dots indicate the position of a three-amino-acid insert in the  $\alpha 2$  protein. (Adapted from C. Wolberger et al., *Cell* 67:517–528, 1991. With permission from Elsevier.)

protein family has been more highly conserved than has the amino acid sequence. In many cases, the amino acid sequences have diverged so far that we cannot be certain of a family relationship between two proteins without determining their three-dimensional structures. The yeast  $\alpha 2$  protein and the *Drosophila* engrailed protein, for example, are both gene regulatory proteins in the homeodomain family (discussed in Chapter 7). Because they are identical in only 17 of their 60 amino acid residues, their relationship became certain only by comparing their three-dimensional structures (Figure 3–13). Many similar examples show that two proteins with more than 25% identity in their amino acid sequences usually share the same overall structure.

The various members of a large protein family often have distinct functions. Some of the amino acid changes that make family members different were no doubt selected in the course of evolution because they resulted in useful changes in biological activity, giving the individual family members the different functional properties they have today. But many other amino acid changes are effectively “neutral,” having neither a beneficial nor a damaging effect on the basic structure and function of the protein. In addition, since mutation is a random process, there must also have been many deleterious changes that altered the three-dimensional structure of these proteins sufficiently to harm them. Such faulty proteins would have been lost whenever the individual organisms making them were at enough of a disadvantage to be eliminated by natural selection.

Protein families are readily recognized when the genome of any organism is sequenced; for example, the determination of the DNA sequence for the entire human genome has revealed that we contain about 21,000 protein-coding genes. (Note, however, that as a result of alternative RNA splicing, human cells can produce much more than 21,000 different proteins, as will be explained in Chapter 6.) Through sequence comparisons, we can assign the products of at least 40% of our protein-coding genes to known protein structures, belonging to more than 500 different protein families. Most of the proteins in each family have evolved to perform somewhat different functions, as for the enzymes elastase and chymotrypsin illustrated previously in Figure 3–12. As explained in Chapter 1 (see Figure 1–21), these are sometimes called *paralogs* to distinguish them from the many corresponding proteins in different organisms (*orthologs*, such as mouse and human elastase).

As described in Chapter 8, because of the powerful techniques of x-ray crystallography and nuclear magnetic resonance (NMR), we now know the three-dimensional shapes, or conformations, of more than 100,000 proteins. By carefully comparing the conformations of these proteins, structural biologists (that is, experts on the structure of biological molecules) have concluded that there are a limited number of ways in which protein domains fold up in nature—maybe as few as 2000, if we consider all organisms. For most of these so-called *protein folds*, representative structures have been determined.

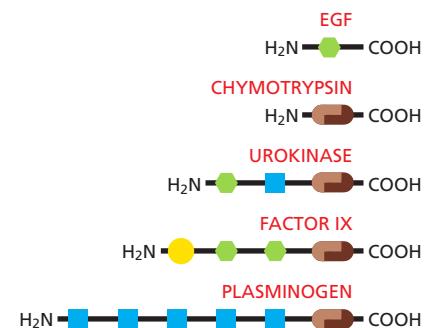
The present database of known protein sequences contains more than twenty million entries, and it is growing very rapidly as more and more genomes are sequenced—revealing huge numbers of new genes that encode proteins. The encoded polypeptides range widely in size, from 6 amino acids to a gigantic protein of 33,000 amino acids. Protein comparisons are important because related structures often imply related functions. Many years of experimentation can be saved by discovering that a new protein has an amino acid sequence similarity with a protein of known function. Such sequence relationships, for example, first indicated that certain genes that cause mammalian cells to become cancerous encode protein kinases (discussed in Chapter 20).

### Some Protein Domains Are Found in Many Different Proteins

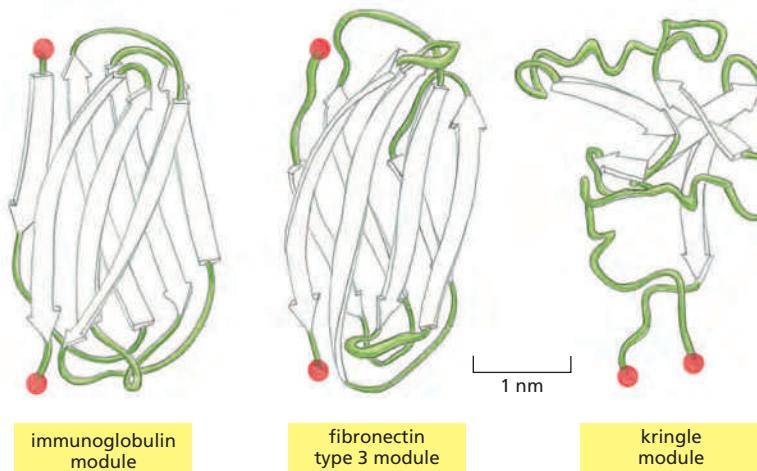
As previously stated, most proteins are composed of a series of protein domains, in which different regions of the polypeptide chain fold independently to form compact structures. Such multidomain proteins are believed to have originated from the accidental joining of the DNA sequences that encode each domain, creating a new gene. In an evolutionary process called *domain shuffling*, many large proteins have evolved through the joining of preexisting domains in new combinations (Figure 3–14). Novel binding surfaces have often been created at the juxtaposition of domains, and many of the functional sites where proteins bind to small molecules are found to be located there.

A subset of protein domains has been especially mobile during evolution; these seem to have particularly versatile structures and are sometimes referred to as *protein modules*. The structure of one, the SH2 domain, was illustrated in Figure 3–6. Three other abundant protein domains are illustrated in Figure 3–15.

Each of the domains shown has a stable core structure formed from strands of  $\beta$  sheets, from which less-ordered loops of polypeptide chain protrude. The loops are ideally situated to form binding sites for other molecules, as most clearly demonstrated for the immunoglobulin fold, which forms the basis for antibody molecules. Such  $\beta$ -sheet-based domains may have achieved their evolutionary success because they provide a convenient framework for the generation of new binding sites for ligands, requiring only small changes to their protruding loops (see Figure 3–42).



**Figure 3–14 Domain shuffling.** An extensive shuffling of blocks of protein sequence (protein domains) has occurred during protein evolution. Those portions of a protein denoted by the same shape and color in this diagram are evolutionarily related. Serine proteases like chymotrypsin are formed from two domains (brown). In the three other proteases shown, which are highly regulated and more specialized, these two protease domains are connected to one or more domains that are similar to domains found in epidermal growth factor (EGF; green), to a calcium-binding protein (yellow), or to a “kringle” domain (blue). Chymotrypsin is illustrated in Figure 3–12.



**Figure 3–15 The three-dimensional structures of three commonly used protein domains.** In these ribbon diagrams,  $\beta$ -sheet strands are shown as arrows, and the N- and C-termini are indicated by red spheres. Many more such “modules” exist in nature. (Adapted from M. Baron, D.G. Norman and I.D. Campbell, *Trends Biochem. Sci.* 16:13–17, 1991, with permission from Elsevier, and D.J. Leahy et al., *Science* 258:987–991, 1992, with permission from AAAS.)

**Figure 3–16 An extended structure formed from a series of protein domains.** Four fibronectin type 3 domains (see Figure 3–15) from the extracellular matrix molecule fibronectin are illustrated in (A) ribbon and (B) space-filling models. (Adapted from D.J. Leahy, I. Aukhil and H.P. Erickson, *Cell* 84:155–164, 1996. With permission from Elsevier.)

A second feature of these protein domains that explains their utility is the ease with which they can be integrated into other proteins. Two of the three domains illustrated in Figure 3–15 have their N- and C-terminal ends at opposite poles of the domain. When the DNA encoding such a domain undergoes tandem duplication, which is not unusual in the evolution of genomes (discussed in Chapter 4), the duplicated domains with this “in-line” arrangement can be readily linked in series to form extended structures—either with themselves or with other in-line domains (**Figure 3–16**). Stiff extended structures composed of a series of domains are especially common in extracellular matrix molecules and in the extracellular portions of cell-surface receptor proteins. Other frequently used domains, including the kringle domain illustrated in Figure 3–15 and the SH2 domain, are of a “plug-in” type, with their N- and C-termini close together. After genomic rearrangements, such domains are usually accommodated as an insertion into a loop region of a second protein.

A comparison of the relative frequency of domain utilization in different eukaryotes reveals that, for many common domains, such as protein kinases, this frequency is similar in organisms as diverse as yeast, plants, worms, flies, and humans. But there are some notable exceptions, such as the Major Histocompatibility Complex (MHC) antigen-recognition domain (see Figure 24–36) that is present in 57 copies in humans, but absent in the other four organisms just mentioned. Domains such as these have specialized functions that are not shared with the other eukaryotes; they are assumed to have been strongly selected for during recent evolution to produce the multiple copies observed. Similarly, the SH2 domain shows an unusual increase in its numbers in higher eukaryotes; such domains might be assumed to be especially useful for multicellularity.

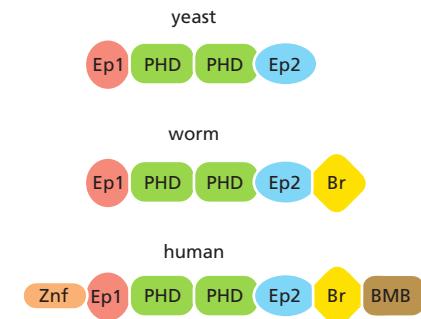
### Certain Pairs of Domains Are Found Together in Many Proteins

We can construct a large table displaying domain usage for each organism whose genome sequence is known. For example, the human genome contains the DNA sequences for about 1000 immunoglobulin domains, 500 protein kinase domains, 250 DNA-binding homeodomains, 300 SH3 domains, and 120 SH2 domains. In addition, we find that more than two-thirds of all proteins consist of two or more domains, and that the same pairs of domains occur repeatedly in the same relative arrangement in a protein. Although half of all domain families are common to archaea, bacteria, and eukaryotes, only about 5% of the two-domain combinations are similarly shared. This pattern suggests that most proteins containing especially useful two-domain combinations arose through domain shuffling relatively late in evolution.

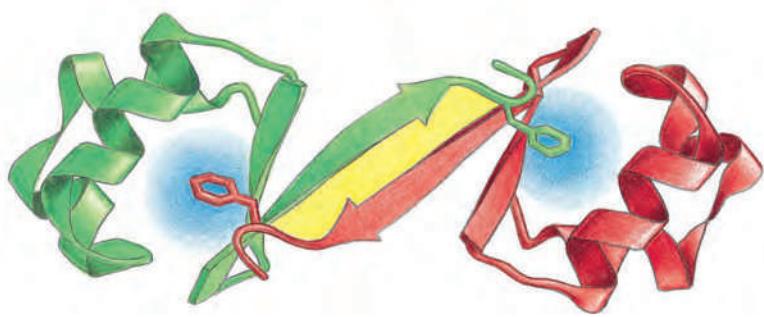
### The Human Genome Encodes a Complex Set of Proteins, Revealing That Much Remains Unknown

The result of sequencing the human genome has been surprising, because it reveals that our chromosomes contain only about 21,000 protein-coding genes. Based on this number alone, we would appear to be no more complex than the tiny mustard weed, *Arabidopsis*, and only about 1.3-fold more complex than a nematode worm. The genome sequences also reveal that vertebrates have inherited nearly all of their protein domains from invertebrates—with only 7% of identified human domains being vertebrate-specific.

Each of our proteins is on average more complicated, however (**Figure 3–17**). Domain shuffling during vertebrate evolution has given rise to many novel



**Figure 3–17 Domain structure of a group of evolutionarily related proteins that are thought to have a similar function.** In general, there is a tendency for the proteins in more complex organisms, such as humans, to contain additional domains—as is the case for the DNA-binding protein compared here.



**Figure 3–18** Two identical protein subunits binding together to form a symmetric protein dimer. The Cro repressor protein from bacteriophage lambda binds to DNA to turn off a specific subset of viral genes. Its two identical subunits bind head-to-head, held together by a combination of hydrophobic forces (blue) and a set of hydrogen bonds (yellow region). (Adapted from D.H. Ohlendorf, D.E. Tronrud and B.W. Matthews, *J. Mol. Biol.* 280:129–136, 1998. With permission from Academic Press.)

combinations of protein domains, with the result that there are nearly twice as many combinations of domains found in human proteins as in a worm or a fly. Thus, for example, the trypsinlike serine protease domain is linked to at least 18 other types of protein domains in human proteins, whereas it is found covalently joined to only 5 different domains in the worm. This extra variety in our proteins greatly increases the range of protein–protein interactions possible (see Figure 3–79), but how it contributes to making us human is not known.

The complexity of living organisms is staggering, and it is quite sobering to note that we currently lack even the tiniest hint of what the function might be for more than 10,000 of the proteins that have thus far been identified through examining the human genome. There are certainly enormous challenges ahead for the next generation of cell biologists, with no shortage of fascinating mysteries to solve.

### Larger Protein Molecules Often Contain More Than One Polypeptide Chain

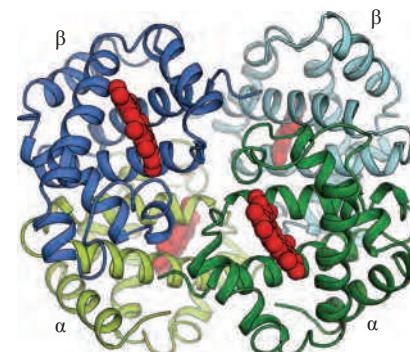
The same weak noncovalent bonds that enable a protein chain to fold into a specific conformation also allow proteins to bind to each other to produce larger structures in the cell. Any region of a protein’s surface that can interact with another molecule through sets of noncovalent bonds is called a **binding site**. A protein can contain binding sites for various large and small molecules. If a binding site recognizes the surface of a second protein, the tight binding of two folded polypeptide chains at this site creates a larger protein molecule with a precisely defined geometry. Each polypeptide chain in such a protein is called a **protein subunit**.

In the simplest case, two identical folded polypeptide chains bind to each other in a “head-to-head” arrangement, forming a symmetric complex of two protein subunits (a *dimer*) held together by interactions between two identical binding sites. The *Cro repressor protein*—a viral gene regulatory protein that binds to DNA to turn specific viral genes off in an infected bacterial cell—provides an example (Figure 3–18). Cells contain many other types of symmetric protein complexes, formed from multiple copies of a single polypeptide chain (for example, see Figure 3–20 below).

Many of the proteins in cells contain two or more types of polypeptide chains. *Hemoglobin*, the protein that carries oxygen in red blood cells, contains two identical  $\alpha$ -globin subunits and two identical  $\beta$ -globin subunits, symmetrically arranged (Figure 3–19). Such multisubunit proteins are very common in cells, and they can be very large (Movie 3.6).

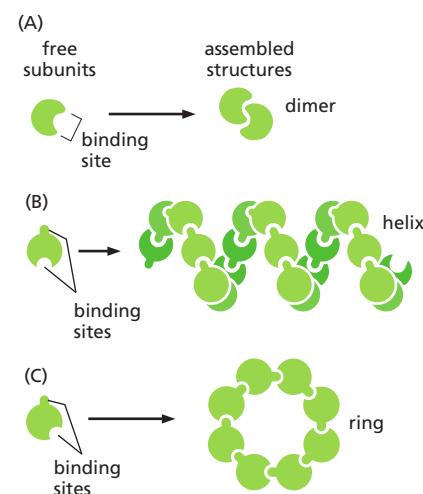
### Some Globular Proteins Form Long Helical Filaments

Most of the proteins that we have discussed so far are *globular proteins*, in which the polypeptide chain folds up into a compact shape like a ball with an irregular surface. Some of these protein molecules can nevertheless assemble to form filaments that may span the entire length of a cell. Most simply, a long chain of identical protein molecules can be constructed if each molecule has a binding



**Figure 3–19** A protein formed as a symmetric assembly using two each of two different subunits. Hemoglobin is an abundant protein in red blood cells that contains two copies of  $\alpha$ -globin (green) and two copies of  $\beta$ -globin (blue). Each of these four polypeptide chains contains a heme molecule (red), which is the site that binds oxygen ( $O_2$ ). Thus, each molecule of hemoglobin in the blood carries four molecules of oxygen. (PDB code: 2DHB.)

**Figure 3–20 Protein assemblies.** (A) A protein with just one binding site can form a dimer with another identical protein. (B) Identical proteins with two different binding sites often form a long helical filament. (C) If the two binding sites are disposed appropriately in relation to each other, the protein subunits may form a closed ring instead of a helix. (For an example of A, see Figure 3–18; for an example of B, see Figure 3–21; for examples of C, see Figures 5–14 and 14–31.)



site complementary to another region of the surface of the same molecule (**Figure 3–20**). An actin filament, for example, is a long helical structure produced from many molecules of the protein *actin* (**Figure 3–21**). Actin is a globular protein that is very abundant in eukaryotic cells, where it forms one of the major filament systems of the cytoskeleton (discussed in Chapter 16).

We will encounter many helical structures in this book. Why is a helix such a common structure in biology? As we have seen, biological structures are often formed by linking similar subunits into long, repetitive chains. If all the subunits are identical, the neighboring subunits in the chain can often fit together in only one way, adjusting their relative positions to minimize the free energy of the contact between them. As a result, each subunit is positioned in exactly the same way in relation to the next, so that subunit 3 fits onto subunit 2 in the same way that subunit 2 fits onto subunit 1, and so on. Because it is very rare for subunits to join up in a straight line, this arrangement generally results in a helix—a regular structure that resembles a spiral staircase, as illustrated in **Figure 3–22**. Depending on the twist of the staircase, a helix is said to be either right-handed or left-handed (see **Figure 3–22E**). Handedness is not affected by turning the helix upside down, but it is reversed if the helix is reflected in the mirror.

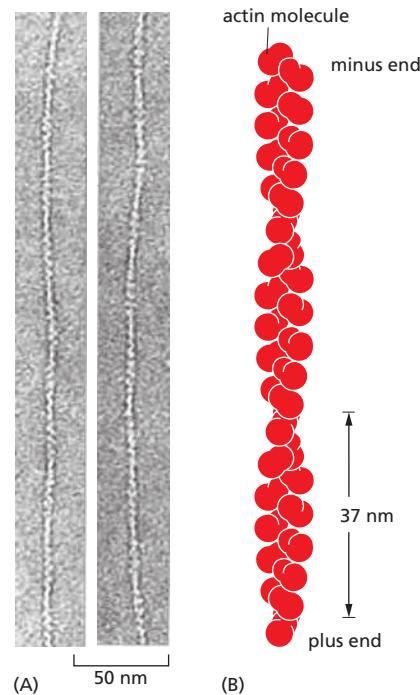
The observation that helices occur commonly in biological structures holds true whether the subunits are small molecules linked together by covalent bonds (for example, the amino acids in an  $\alpha$  helix) or large protein molecules that are linked by noncovalent forces (for example, the actin molecules in actin filaments). This is not surprising. A helix is an unexceptional structure, and it is generated simply by placing many similar subunits next to each other, each in the same strictly repeated relationship to the one before—that is, with a fixed rotation followed by a fixed translation along the helix axis, as in a spiral staircase.

### Many Protein Molecules Have Elongated, Fibrous Shapes

Enzymes tend to be globular proteins: even though many are large and complicated, with multiple subunits, most have an overall rounded shape. In **Figure 3–21**, we saw that a globular protein can also associate to form long filaments. But there are also functions that require each individual protein molecule to span a large distance. These proteins generally have a relatively simple, elongated three-dimensional structure and are commonly referred to as *fibrous proteins*.

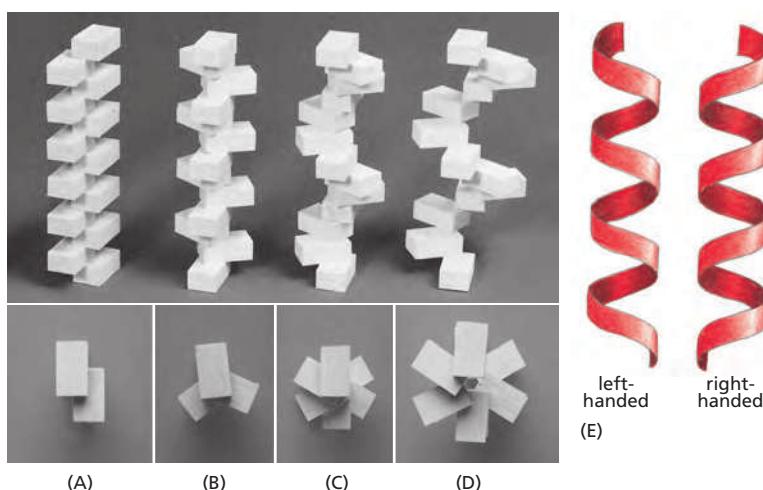
One large family of intracellular fibrous proteins consists of  $\alpha$ -keratin, introduced when we presented the  $\alpha$  helix, and its relatives. Keratin filaments are extremely stable and are the main component in long-lived structures such as hair, horn, and nails. An  $\alpha$ -keratin molecule is a dimer of two identical subunits, with the long  $\alpha$  helices of each subunit forming a coiled-coil (see **Figure 3–9**). The coiled-coil regions are capped at each end by globular domains containing binding sites. This enables this class of protein to assemble into ropelike *intermediate filaments*—an important component of the cytoskeleton that creates the cell's internal structural framework (see **Figure 16–67**).

Fibrous proteins are especially abundant outside the cell, where they are a main component of the gel-like *extracellular matrix* that helps to bind collections of cells together to form tissues. Cells secrete extracellular matrix proteins into their surroundings, where they often assemble into sheets or long fibrils. *Collagen* is the most abundant of these proteins in animal tissues. A collagen molecule consists of three long polypeptide chains, each containing the nonpolar amino



**Figure 3–21** Actin filaments.

(A) Transmission electron micrographs of negatively stained actin filaments. (B) The helical arrangement of actin molecules in an actin filament. (A, courtesy of Roger Craig.)

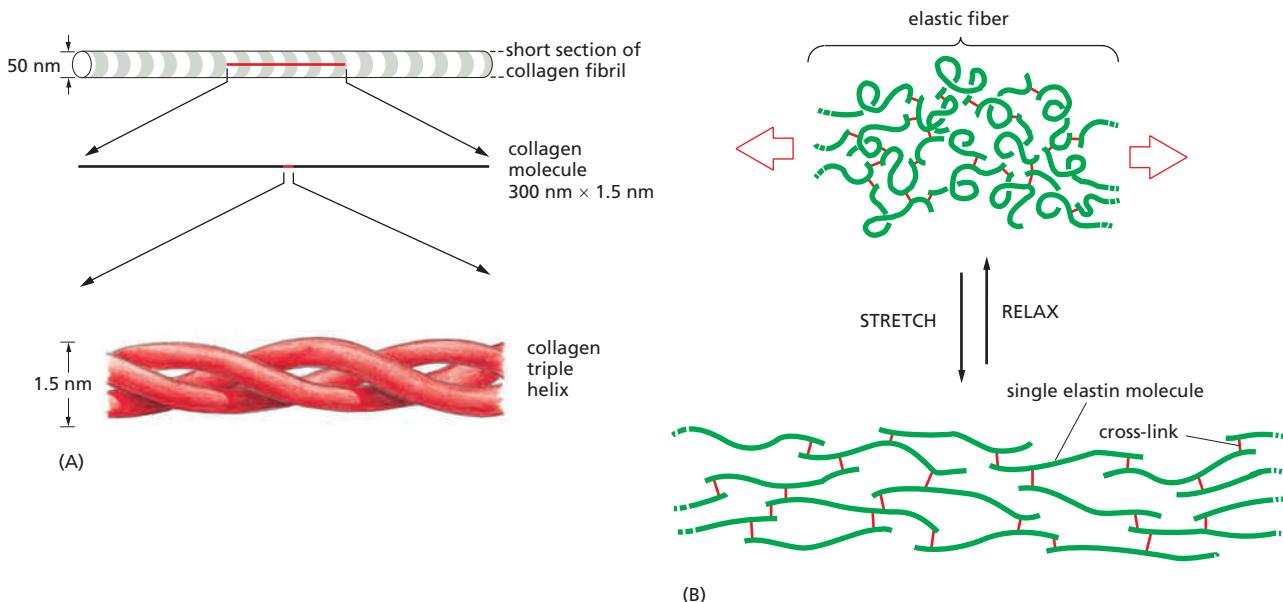
**Figure 3–22** Some properties of a helix.

(A–D) A helix forms when a series of subunits bind to each other in a regular way. At the bottom, each of these helices is viewed from directly above the helix and seen to have two (A), three (B), and six (C and D) subunits per helical turn. Note that the helix in (D) has a wider path than that in (C), but the same number of subunits per turn. (E) As discussed in the text, a helix can be either right-handed or left-handed. As a reference, it is useful to remember that standard metal screws, which insert when turned clockwise, are right-handed. Note that a helix retains the same handedness when it is turned upside down. (PDB code: 2DHB.)

acid glycine at every third position. This regular structure allows the chains to wind around one another to generate a long regular triple helix (Figure 3–23A). Many collagen molecules then bind to one another side-by-side and end-to-end to create long overlapping arrays—thereby generating the extremely tough collagen fibrils that give connective tissues their tensile strength, as described in Chapter 19.

### Proteins Contain a Surprisingly Large Amount of Intrinsically Disordered Polypeptide Chain

It has been well known for a long time that, in complete contrast to collagen, another abundant protein in the extracellular matrix, *elastin*, is formed as a highly disordered polypeptide. This disorder is essential for elastin's function. Its relatively loose and unstructured polypeptide chains are covalently cross-linked to



**Figure 3–23** Collagen and elastin. (A) Collagen is a triple helix formed by three extended protein chains that wrap around one another (bottom). Many rodlike collagen molecules are cross-linked together in the extracellular space to form unextendable collagen fibrils (top) that have the tensile strength of steel. The striping on the collagen fibril is caused by the regular repeating arrangement of the collagen molecules within the fibril. (B) Elastin polypeptide chains are cross-linked together in the extracellular space to form rubberlike, elastic fibers. Each elastin molecule uncoils into a more extended conformation when the fiber is stretched and recoils spontaneously as soon as the stretching force is relaxed. The cross-linking in the extracellular space mentioned creates covalent linkages between lysine side chains, but the chemistry is different for collagen and elastin.

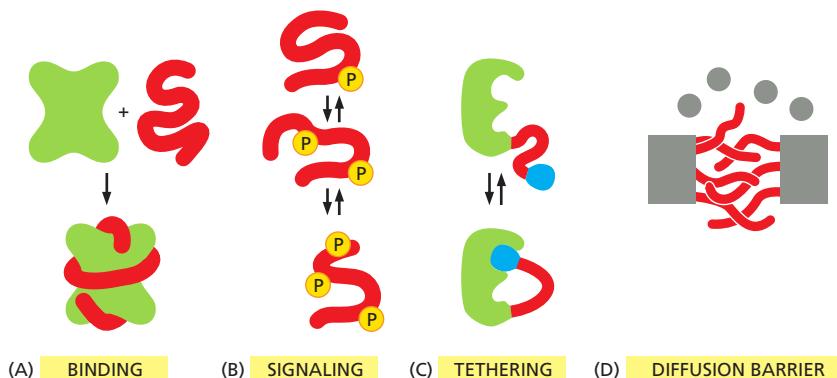
produce a rubberlike, elastic meshwork that can be reversibly pulled from one conformation to another, as illustrated in Figure 3–23B. The elastic fibers that result enable skin and other tissues, such as arteries and lungs, to stretch and recoil without tearing.

Intrinsically disordered regions of proteins are frequent in nature, and they have important functions in the interior of cells. As we have already seen, proteins often have loops of polypeptide chain that protrude from the core region of a protein domain to bind other molecules. Some of these loops remain largely unstructured until they bind to a target molecule, adopting a specific folded conformation only when this other molecule is bound. Many proteins were also known to have intrinsically disordered tails at one or the other end of a structured domain (see, for example, the histones in Figure 4–24). But the extent of such disordered structure only became clear when genomes were sequenced. This allowed bioinformatic methods to be used to analyze the amino acid sequences that genes encode, searching for disordered regions based on their unusually low hydrophobicity and relatively high net charge. Combining these results with other data, it is now thought that perhaps a quarter of all eukaryotic proteins can adopt structures that are mostly disordered, fluctuating rapidly between many different conformations. Many such intrinsically disordered regions contain repeated sequences of amino acids. What do these disordered regions do?

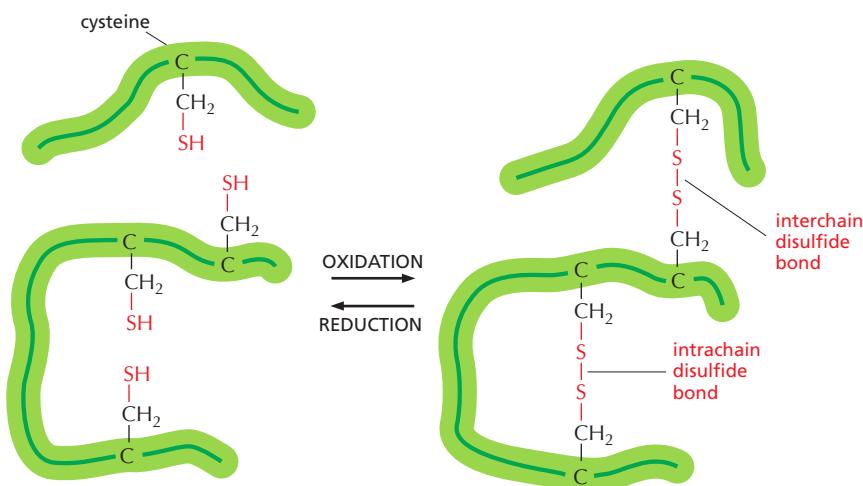
Some known functions are illustrated in **Figure 3–24**. One predominant function is to form specific binding sites for other protein molecules that are of high specificity, but readily altered by protein phosphorylation, protein dephosphorylation, or any of the other covalent modifications that are triggered by cell signaling events (Figure 3–24A and B). We shall see, for example, that the eukaryotic RNA polymerase enzyme that produces mRNAs contains a long, unstructured C-terminal tail that is covalently modified as its RNA synthesis proceeds, thereby attracting specific other proteins to the transcription complex at different times (see Figure 6–22). And this unstructured tail interacts with a different type of low complexity domain when the RNA polymerase is recruited to the specific sites on the DNA where it begins synthesis.

As illustrated in Figure 3–24C, an unstructured region can also serve as a “tether” to hold two protein domains in close proximity to facilitate their interaction. For example, it is this tethering function that allows substrates to move between active sites in large multienzyme complexes (see Figure 3–54). A similar tethering function allows large *scaffold proteins* with multiple protein-binding sites to concentrate sets of interacting proteins, both increasing reaction rates and confining their reaction to a particular site in a cell (see Figure 3–78).

Like elastin, other proteins have a function that directly requires that they remain largely unstructured. Thus, large numbers of disordered protein chains in close proximity can create micro-regions of gel-like consistency inside the cell that restrict diffusion. For example, the abundant nucleoporins that coat the inner surface of the nuclear pore complex form a random coil meshwork (Figure 3–24) that is critical for selective nuclear transport (see Figure 12–8).



**Figure 3–24** Some important functions for intrinsically disordered protein sequences. (A) Unstructured regions of polypeptide chain often form binding sites for other proteins. Although these binding events are of high specificity, they are often of low affinity due to the free-energy cost of folding the normally unfolded partner (and they are thus readily reversible). (B) Unstructured regions can be easily modified covalently to change their binding preferences, and they are therefore frequently involved in cell signaling processes. In this schematic, multiple sites of protein phosphorylation are indicated. (C) Unstructured regions frequently create “tethers” that hold interacting protein domains in close proximity. (D) A dense network of unstructured proteins can form a diffusion barrier, as the nucleoporins do for the nuclear pore.



**Figure 3–25 Disulfide bonds.** Covalent disulfide bonds form between adjacent cysteine side chains. These cross-linkages can join either two parts of the same polypeptide chain or two different polypeptide chains. Since the energy required to break one covalent bond is much larger than the energy required to break even a whole set of noncovalent bonds (see Table 2–1, p. 45), a disulfide bond can have a major stabilizing effect on a protein ([Movie 3.7](#)).

### Covalent Cross-Linkages Stabilize Extracellular Proteins

Many protein molecules are either attached to the outside of a cell's plasma membrane or secreted as part of the extracellular matrix. All such proteins are directly exposed to extracellular conditions. To help maintain their structures, the polypeptide chains in such proteins are often stabilized by covalent cross-linkages. These linkages can either tie together two amino acids in the same protein, or connect different polypeptide chains in a multisubunit protein. Although many other types exist, the most common cross-linkages in proteins are covalent sulfur-sulfur bonds. These *disulfide bonds* (also called *S-S bonds*) form as cells prepare newly synthesized proteins for export. As described in Chapter 12, their formation is catalyzed in the endoplasmic reticulum by an enzyme that links together two pairs of -SH groups of cysteine side chains that are adjacent in the folded protein ([Figure 3–25](#)). Disulfide bonds do not change the conformation of a protein but instead act as atomic staples to reinforce its most favored conformation. For example, lysozyme—an enzyme in tears that dissolves bacterial cell walls—retains its antibacterial activity for a long time because it is stabilized by such cross-linkages.

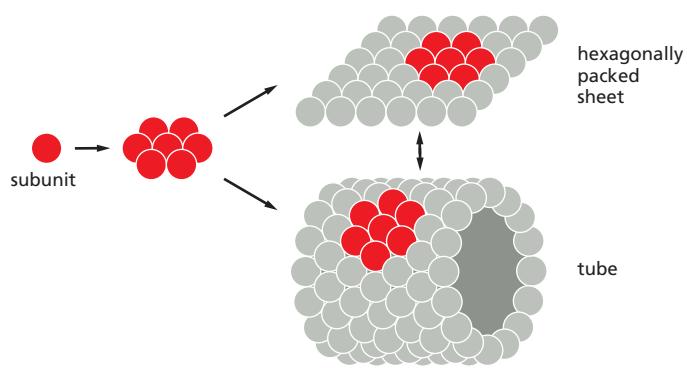
Disulfide bonds generally fail to form in the cytosol, where a high concentration of reducing agents converts S-S bonds back to cysteine -SH groups. Apparently, proteins do not require this type of reinforcement in the relatively mild environment inside the cell.

### Protein Molecules Often Serve as Subunits for the Assembly of Large Structures

The same principles that enable a protein molecule to associate with itself to form rings or a long filament also operate to generate much larger structures formed from a set of different macromolecules, such as enzyme complexes, ribosomes, viruses, and membranes. These large objects are not made as single, giant, covalently linked molecules. Instead they are formed by the noncovalent assembly of many separately manufactured molecules, which serve as the subunits of the final structure.

The use of smaller subunits to build larger structures has several advantages:

1. A large structure built from one or a few repeating smaller subunits requires only a small amount of genetic information.
2. Both assembly and disassembly can be readily controlled reversible processes, because the subunits associate through multiple bonds of relatively low energy.
3. Errors in the synthesis of the structure can be more easily avoided, since correction mechanisms can operate during the course of assembly to exclude malformed subunits.



**Figure 3–26** Single protein subunits form protein assemblies that feature multiple protein–protein contacts. Hexagonally packed globular protein subunits are shown here forming either flat sheets or tubes. Generally, such large structures are not considered to be single “molecules.” Instead, like the actin filament described previously, they are viewed as assemblies formed of many different molecules.

Some protein subunits assemble into flat sheets in which the subunits are arranged in hexagonal patterns. Specialized membrane proteins are sometimes arranged this way in lipid bilayers. With a slight change in the geometry of the individual subunits, a hexagonal sheet can be converted into a tube (Figure 3–26) or, with more changes, into a hollow sphere. Protein tubes and spheres that bind specific RNA and DNA molecules in their interior form the coats of viruses.

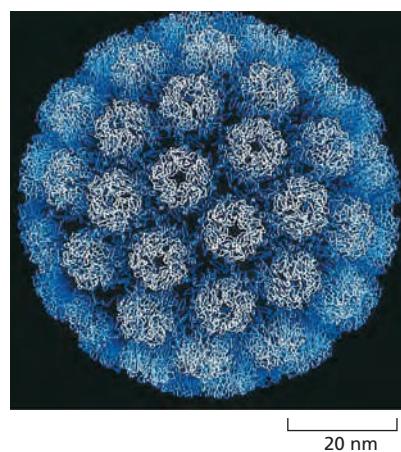
The formation of closed structures, such as rings, tubes, or spheres, provides additional stability because it increases the number of bonds between the protein subunits. Moreover, because such a structure is created by mutually dependent, cooperative interactions between subunits, a relatively small change that affects each subunit individually can cause the structure to assemble or disassemble. These principles are dramatically illustrated in the protein coat or *capsid* of many simple viruses, which takes the form of a hollow sphere based on an icosahedron (Figure 3–27). Capsids are often made of hundreds of identical protein subunits that enclose and protect the viral nucleic acid (Figure 3–28). The protein in such a capsid must have a particularly adaptable structure: not only must it make several different kinds of contacts to create the sphere, it must also change this arrangement to let the nucleic acid out to initiate viral replication once the virus has entered a cell.

### Many Structures in Cells Are Capable of Self-Assembly

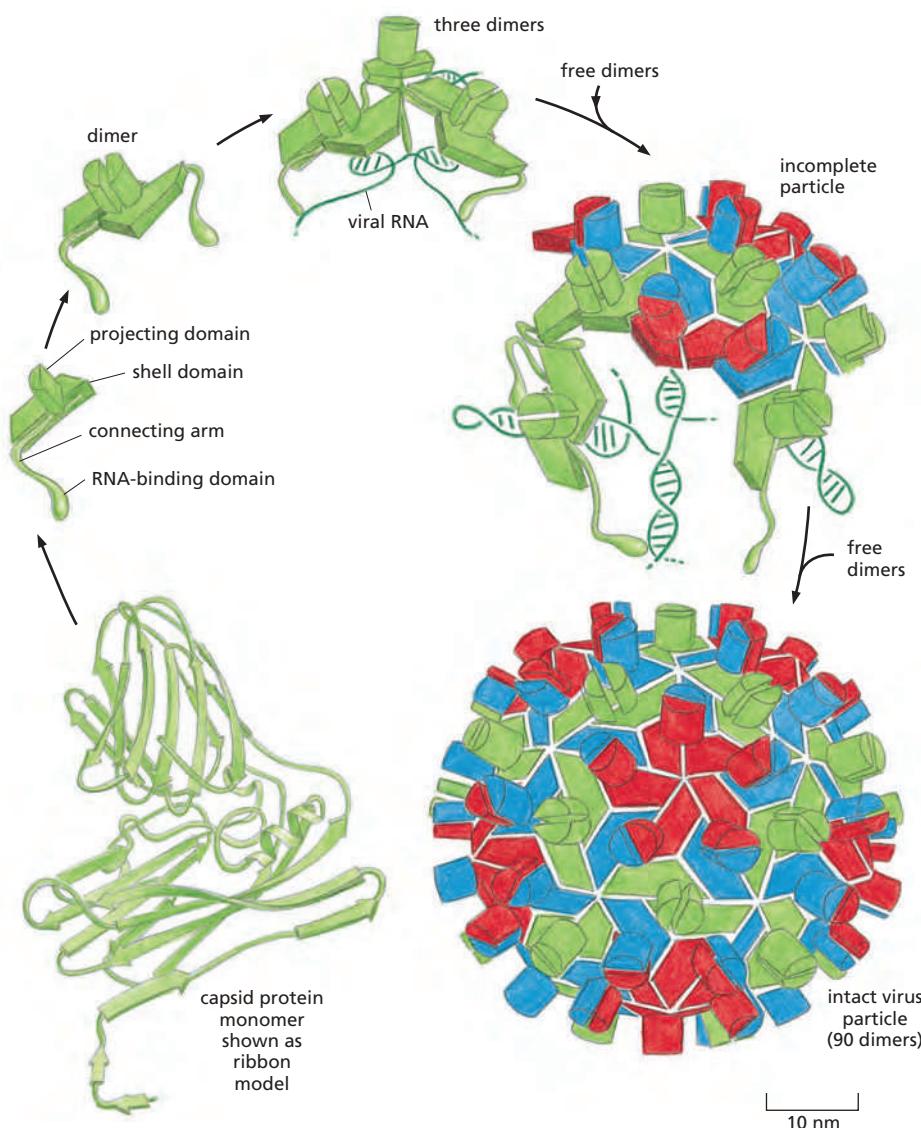
The information for forming many of the complex assemblies of macromolecules in cells must be contained in the subunits themselves, because purified subunits can spontaneously assemble into the final structure under the appropriate conditions. The first large macromolecular aggregate shown to be capable of self-assembly from its component parts was *tobacco mosaic virus* (TMV). This virus is a long rod in which a cylinder of protein is arranged around a helical RNA core (Figure 3–29). If the dissociated RNA and protein subunits are mixed together in solution, they recombine to form fully active viral particles. The assembly process is unexpectedly complex and includes the formation of double rings of protein, which serve as intermediates that add to the growing viral coat.

Another complex macromolecular aggregate that can reassemble from its component parts is the bacterial ribosome. This structure is composed of about 55 different protein molecules and 3 different rRNA molecules. Incubating a mixture of the individual components under appropriate conditions in a test tube causes them to spontaneously re-form the original structure. Most importantly, such reconstituted ribosomes are able to catalyze protein synthesis. As might be expected, the reassembly of ribosomes follows a specific pathway: after certain proteins have bound to the RNA, this complex is then recognized by other proteins, and so on, until the structure is complete.

It is still not clear how some of the more elaborate self-assembly processes are regulated. Many structures in the cell, for example, seem to have a precisely defined length that is many times greater than that of their component macromolecules. How such length determination is achieved is in many cases a mystery. In

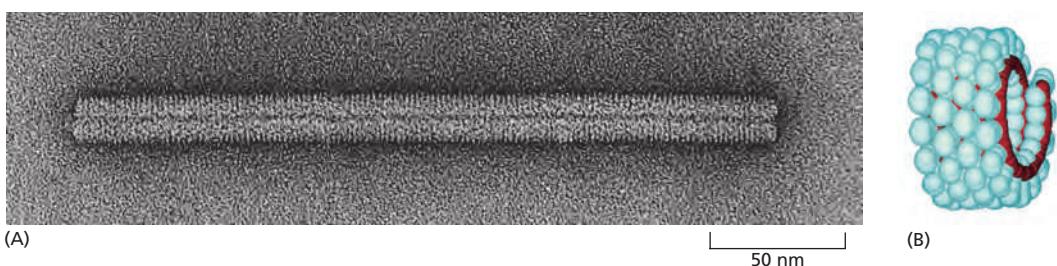


**Figure 3–27** The protein capsid of a virus. The structure of the simian virus SV40 capsid has been determined by x-ray crystallography and, as for the capsids of many other viruses, it is known in atomic detail. (Courtesy of Robert Grant, Stephan Crainic, and James M. Hogle.)



**Figure 3–28 The structure of a spherical virus.** In viruses, many copies of a single protein subunit often pack together to create a spherical shell (a capsid). This capsid encloses the viral genome, composed of either RNA or DNA (see also Figure 3–27). For geometric reasons, no more than 60 identical subunits can pack together in a precisely symmetric way. If slight irregularities are allowed, however, more subunits can be used to produce a larger capsid that retains icosahedral symmetry. The tomato bushy stunt virus (TBSV) shown here, for example, is a spherical virus about 33 nm in diameter formed from 180 identical copies of a 386-amino-acid capsid protein plus an RNA genome of 4500 nucleotides. To construct such a large capsid, the protein must be able to fit into three somewhat different environments. This requires three slightly different conformations, each of which is differently colored in the virus particle shown here. The postulated pathway of assembly is shown; the precise three-dimensional structure has been determined by x-ray diffraction. (Courtesy of Steve Harrison.)

In the simplest case, a long core protein or other macromolecule provides a scaffold that determines the extent of the final assembly. This is the mechanism that determines the length of the TMV particle, where the RNA chain provides the core. Similarly, a core protein interacting with actin is thought to determine the length of the thin filaments in muscle.



**Figure 3–29 The structure of tobacco mosaic virus (TMV).** (A) An electron micrograph of the viral particle, which consists of a single long RNA molecule enclosed in a cylindrical protein coat composed of identical protein subunits. (B) A model showing part of the structure of TMV. A single-stranded RNA molecule of 6395 nucleotides is packaged in a helical coat constructed from 2130 copies of a coat protein 158 amino acids long. Fully infective viral particles can self-assemble in a test tube from purified RNA and protein molecules. (A, courtesy of Robley Williams; B, courtesy of Richard J. Feldmann.)

**Figure 3–30 Proteolytic cleavage in insulin assembly.** The polypeptide hormone insulin cannot spontaneously re-form efficiently if its disulfide bonds are disrupted. It is synthesized as a larger protein (*proinsulin*) that is cleaved by a proteolytic enzyme after the protein chain has folded into a specific shape. Excision of part of the proinsulin polypeptide chain removes some of the information needed for the protein to fold spontaneously into its normal conformation. Once insulin has been denatured and its two polypeptide chains have separated, its ability to reassemble is lost.

### Assembly Factors Often Aid the Formation of Complex Biological Structures

Not all cellular structures held together by noncovalent bonds self-assemble. A cilium, or a myofibril of a muscle cell, for example, cannot form spontaneously from a solution of its component macromolecules. In these cases, part of the assembly information is provided by special enzymes and other proteins that perform the function of templates, serving as *assembly factors* that guide construction but take no part in the final assembled structure.

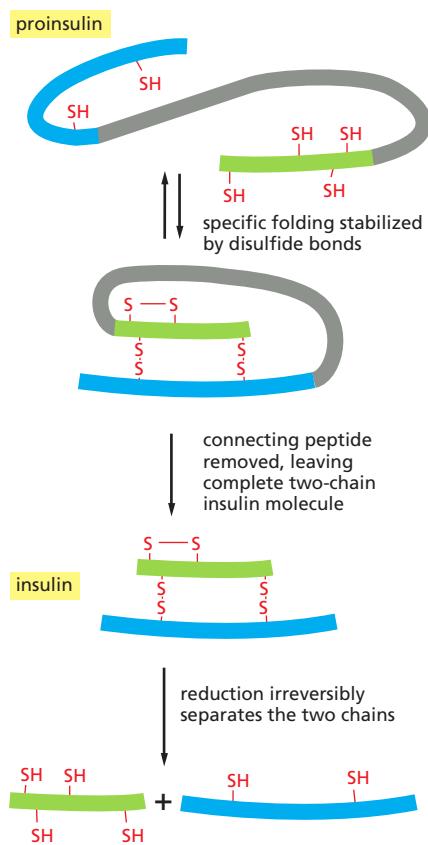
Even relatively simple structures may lack some of the ingredients necessary for their own assembly. In the formation of certain bacterial viruses, for example, the head, which is composed of many copies of a single protein subunit, is assembled on a temporary scaffold composed of a second protein that is produced by the virus. Because the second protein is absent from the final viral particle, the head structure cannot spontaneously reassemble once it has been taken apart. Other examples are known in which proteolytic cleavage is an essential and irreversible step in the normal assembly process. This is even the case for some small protein assemblies, including the structural protein collagen and the hormone insulin (Figure 3–30). From these relatively simple examples, it seems certain that the assembly of a structure as complex as a cilium will involve a temporal and spatial ordering that is imparted by numerous other components.

### Amyloid Fibrils Can Form from Many Proteins

A special class of protein structures, utilized for some normal cell functions, can also contribute to human diseases when not controlled. These are self-propagating, stable  $\beta$ -sheet aggregates called **amyloid fibrils**. These fibrils are built from a series of identical polypeptide chains that become layered one over the other to create a continuous stack of  $\beta$  sheets, with the  $\beta$  strands oriented perpendicular to the fibril axis to form a *cross-beta filament* (Figure 3–31). Typically, hundreds of monomers will aggregate to form an unbranched fibrous structure that is several micrometers long and 5 to 15 nm in width. A surprisingly large fraction of proteins have the potential to form such structures, because the short segment of the polypeptide chain that forms the spine of the fibril can have a variety of different sequences and follow one of several different paths (Figure 3–32). However, very few proteins will actually form this structure inside cells.

In normal humans, the quality control mechanisms governing proteins gradually decline with age, occasionally permitting normal proteins to form pathological aggregates. The protein aggregates may be released from dead cells and accumulate as amyloid in the extracellular matrix. In extreme cases, the accumulation of such amyloid fibrils in the cell interior can kill the cells and damage tissues. Because the brain is composed of a highly organized collection of nerve cells that cannot regenerate, the brain is especially vulnerable to this sort of cumulative damage. Thus, although amyloid fibrils may form in different tissues, and are known to cause pathologies in several sites in the body, the most severe amyloid pathologies are neurodegenerative diseases. For example, the abnormal formation of highly stable amyloid fibrils is thought to play a central causative role in both Alzheimer's and Parkinson's diseases.

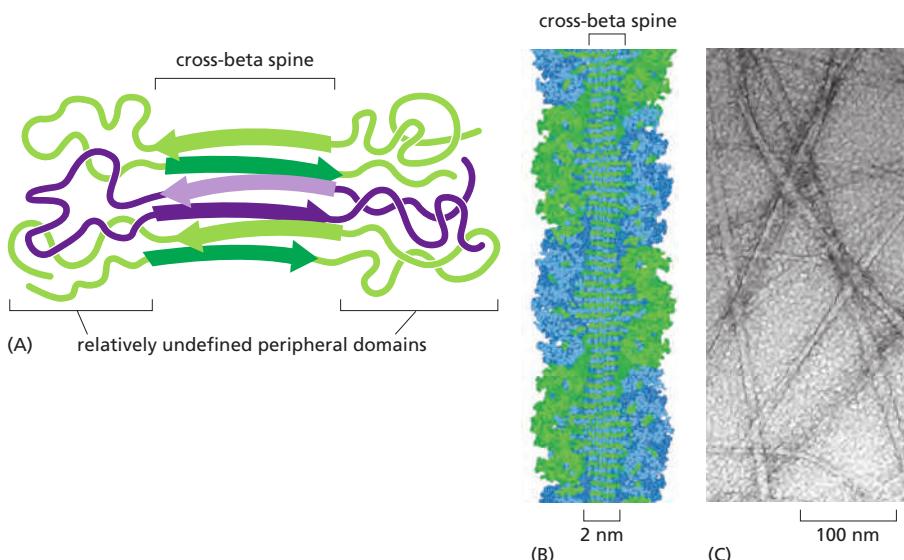
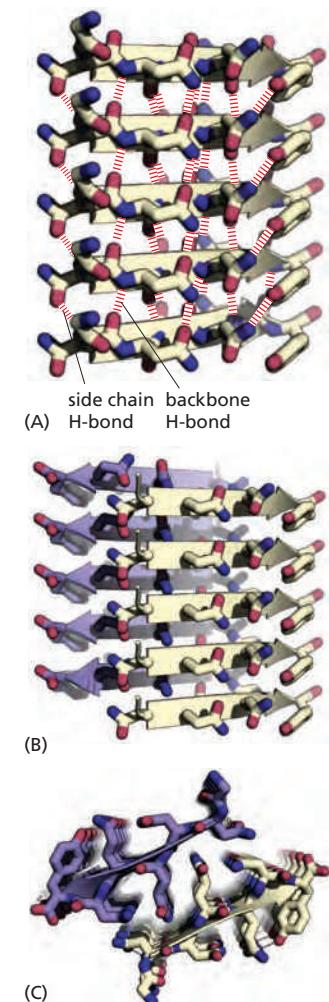
**Prion diseases** are a special type of these pathologies. They have attained special notoriety because, unlike Parkinson's or Alzheimer's, prion diseases can spread from one organism to another, providing that the second organism eats a



**Figure 3–31 Detailed structure of the core of an amyloid fibril.** Illustrated here is the cross-beta spine of the amyloid fibril that is formed by a peptide of seven amino acids from the protein Sup35, an extensively studied yeast prion. Consisting of the sequence glycine-asparagine-asparagine-glutamine-glutamine-asparagine-tyrosine (GNNQQNY), its structure was determined by X-ray crystallography. Although the cross-beta spines of other amyloids are similar, being composed of two long  $\beta$  sheets held together by a “steric zipper,” different detailed structures are observed depending on the short peptide sequence involved. (A) One half of the spine is illustrated. Here, a standard parallel  $\beta$ -sheet structure (see p. 116) is held together by a set of hydrogen bonds between two side chains plus hydrogen bonds between two backbone atoms, as illustrated (oxygen atoms red and nitrogen atoms blue). Note that in this example, the adjacent peptides are exactly in register. Although only five layers are shown (each layer depicted as an arrow), the actual structure would extend for many tens of thousands of layers in the plane of the paper. (B) The complete cross-beta spine. A second, identical  $\beta$ -sheet is paired with the first one to form a two-sheet motif that runs the entire length of the fibril. (C) View of the complete spine in (B) from the top. The closely interdigitated side chains form a tight, water-free junction known as a steric zipper. (Courtesy of David Eisenberg and Michael Sawaya, UCLA; based on R. Nelson et al., *Nature* 435:773–778, 2005. With permission from Macmillan Publishers Ltd.)

tissue containing the protein aggregate. A set of closely related diseases—scrapie in sheep, Creutzfeldt-Jakob disease (CJD) in humans, Kuru in humans, and bovine spongiform encephalopathy (BSE) in cattle—are caused by a misfolded, aggregated form of a particular protein called PrP (for prion protein). PrP is normally located on the outer surface of the plasma membrane, most prominently in neurons, and it has the unfortunate property of forming amyloid fibrils that are “infectious” because they convert normally folded molecules of PrP to the same pathological form (Figure 3–33). This property creates a positive feedback loop that propagates the abnormal form of PrP, called PrP\*, and allows the pathological conformation to spread rapidly from cell to cell in the brain, eventually causing death. It can be dangerous to eat the tissues of animals that contain PrP\*, as witnessed by the spread of BSE (commonly referred to as “mad cow disease”) from cattle to humans. Fortunately, in the absence of PrP\*, PrP is extraordinarily difficult to convert to its abnormal form.

A closely related “protein-only inheritance” has been observed in yeast cells. The ability to study infectious proteins in yeast has clarified another remarkable feature of prions. These protein molecules can form several distinctively different types of amyloid fibrils from the same polypeptide chain. Moreover, each type of aggregate can be infectious, forcing normal protein molecules to adopt the same type of abnormal structure. Thus, several different “strains” of infectious particles can arise from the same polypeptide chain.



**Figure 3–32 The structure of an amyloid fibril.** (A) Schematic diagram of the structure of a amyloid fibril that is formed by the aggregation of a protein. Only the cross-beta spine of an amyloid fibril resembles the structure shown in Figure 3–31. (B) A cut-away view of a structure proposed for the amyloid fibril that can be formed in a test tube by the enzyme ribonuclease A, illustrating how the core of the fibril—formed by a short segment—relates to the rest of the structure. (C) Electron micrograph of amyloid fibrils. (A, from L. Esposito, C. Pedone and L. Vitagliano, *Proc. Natl Acad. Sci. USA* 103:11533–11538, 2006; B, from S. Sambashivan et al., *Nature* 437:266–269, 2005; C, courtesy of David Eisenberg.)

**Figure 3–33 The special protein aggregates that cause prion diseases.**

(A) Schematic illustration of the type of conformational change in the PrP protein (prion protein) that produces material for an amyloid fibril. (B) The self-infectious nature of the protein aggregation that is central to prion diseases. PrP is highly unusual because the misfolded version of the protein, called PrP\*, induces the normal PrP protein it contacts to change its conformation, as shown.

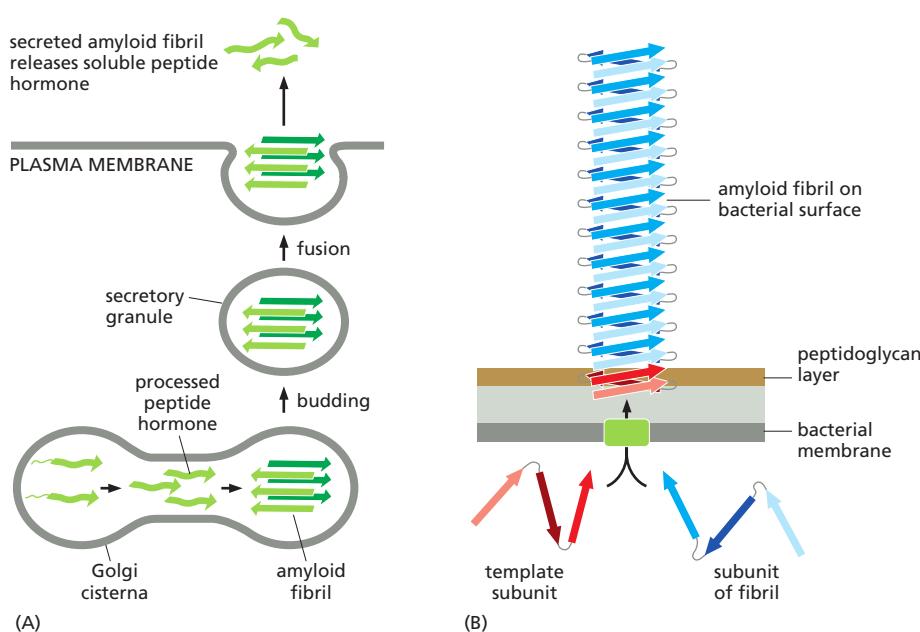
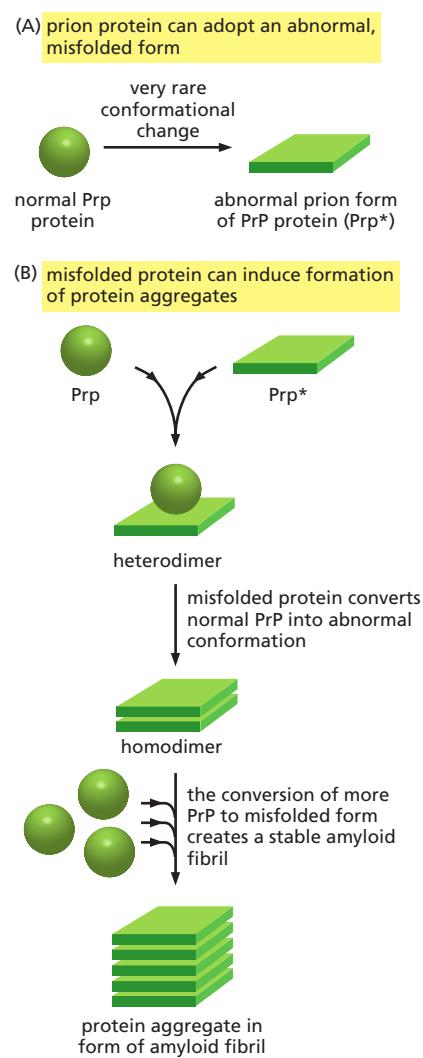
### Amyloid Structures Can Perform Useful Functions in Cells

Amyloid fibrils were initially studied because they cause disease. But the same type of structure is now known to be exploited by cells for useful purposes. Eukaryotic cells, for example, store many different peptide and protein hormones that they will secrete in specialized “secretory granules,” which package a high concentration of their cargo in dense cores with a regular structure (see Figure 13–65). We now know that these structured cores consist of amyloid fibrils, which in this case have a structure that causes them to dissolve to release soluble cargo after being secreted by exocytosis to the cell exterior (Figure 3–34A). Many bacteria use the amyloid structure in a very different way, secreting proteins that form long amyloid fibrils projecting from the cell exterior that help to bind bacterial neighbors into biofilms (Figure 3–34B). Because these biofilms help bacteria to survive in adverse environments (including in humans treated with antibiotics), new drugs that specifically disrupt the fibrous networks formed by bacterial amyloids have promise for treating human infections.

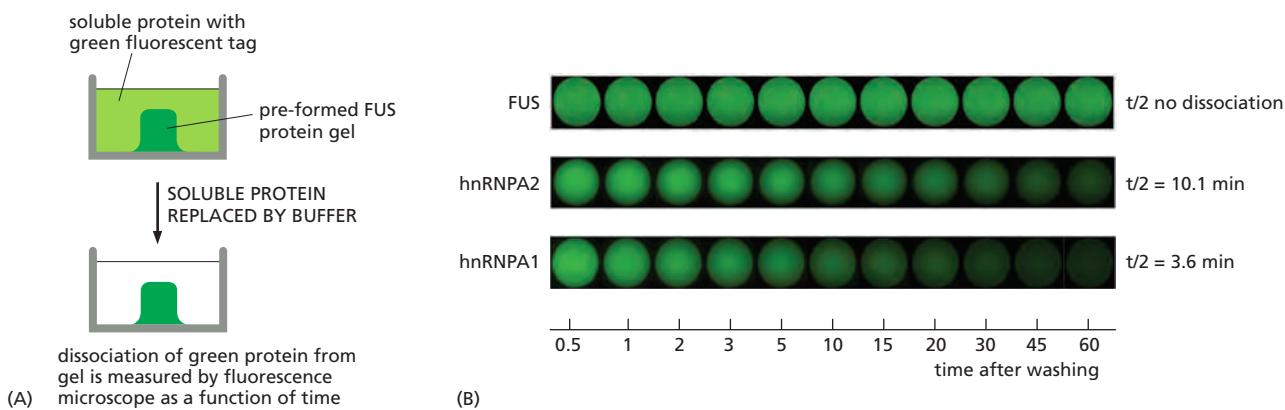
### Many Proteins Contain Low-complexity Domains that Can Form “Reversible Amyloids”

Until recently, those amyloids with useful functions were thought to be either confined to the interior of specialized vesicles or expressed on the exterior of cells, as in Figure 3–34. However, new experiments reveal that a large set of *low complexity domains* can form amyloid fibers that have functional roles in both the cell nucleus and the cell cytoplasm. These domains are normally unstructured and consist of stretches of amino acid sequence that can span hundreds of amino acids, while containing only a small subset of the 20 different amino acids. In contrast to the disease-associated amyloid in Figure 3–33, these newly discovered structures are held together by weaker noncovalent bonds and readily dissociate in response to signals—hence their name *reversible amyloids*.

Many proteins with such domains also contain a different set of domains that bind to specific other protein or RNA molecules. Thus, their controlled aggregation



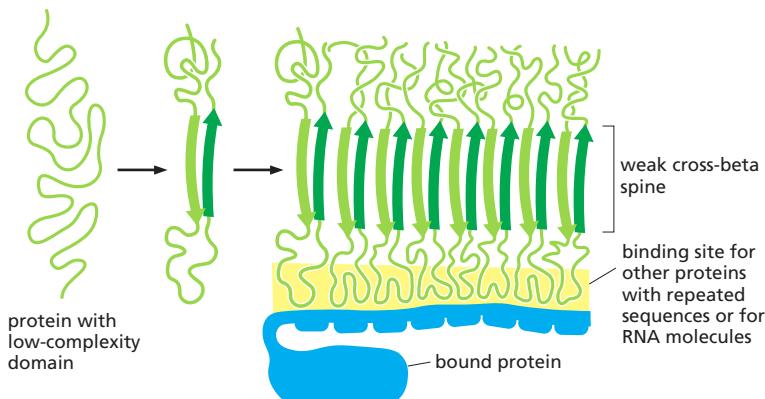
**Figure 3–34 Two normal functions for amyloid fibrils.** (A) In eukaryotic cells, protein cargo can be packed very densely in secretory vesicles and stored until signals cause a release of this cargo by exocytosis. For example, proteins and peptide hormones of the endocrine system, such as glucagon and calcitonin, are efficiently stored as short amyloid fibrils, which dissociate when they reach the cell exterior. (B) Bacteria produce amyloid fibrils on their surface by secreting the precursor proteins; these fibrils then create biofilms that link together, and help to protect, large numbers of individual bacteria.



**Figure 3–35 Measuring the association between “reversible amyloids.”** (A) Experimental setup. The fiber-forming domains from proteins that contain a low-complexity domain were produced in large quantities by cloning the DNA sequence that encodes them into an *E. coli* plasmid so as to allow overproduction of that domain (see p. 483). After these protein domains were purified by affinity chromatography, a tiny droplet of concentrated solution of one of the domains (here the FUS low-complexity domain) was deposited onto a microscope dish and allowed to gel. The gel was then soaked in a dilute solution of a fluorescently labeled low-complexity domain from the same or a different protein, making the gel fluorescent. After replacing the dilute protein solution with buffer, the relative strength of binding of the various domains to each other could then be measured by the decay of fluorescence, as indicated. (B) Results. The low-complexity domain from the FUS protein binds more tightly to itself than it does to the low-complexity domains from the proteins hnRNPA1 or hnRNPA2. A separate experiment reveals that these three different RNA binding proteins associate by forming mixed amyloid fibrils. (Adapted from M.Kato et al., *Cell* 149: 753-767, 2012.)

in the cell can form a hydrogel that pulls these and other molecules into punctate structures called *intracellular bodies*, or *granules*. Specific mRNAs can be sequestered in such granules, where they are stored until made available by a controlled disassembly of the core amyloid structure that holds them together.

Consider the FUS protein, an essential nuclear protein with roles in the transcription, processing, and transport of specific mRNA molecules. Over 80 percent of its C-terminal domain of two hundred amino acids is composed of only four amino acids: glycine, serine, glutamine, and tyrosine. This low complexity domain is attached to several other domains that bind to RNA molecules. At high enough concentrations in a test tube, it forms a hydrogel that will associate with either itself or with the low complexity domains from other proteins. As illustrated by the experiment in **Figure 3–35**, although different low complexity domains bind to each other, homotypic interactions appear to be of greatest affinity (thus, the FUS low complexity domain binds most tightly to itself). Further experiments reveal that both the homotypic and the heterotypic bindings are mediated through a  $\beta$ -sheet core structure forming amyloid fibrils, and that these structures bind to other types of repeat sequences in the manner indicated in **Figure 3–36**. Many of these interactions appear to be controlled by the phosphorylation of serine side chains in the one or both of the interacting partners. However, a great deal remains to be learned concerning these newly discovered structures and the varied roles that they play in the cell biology of eukaryotic cells.



**Figure 3–36 One type of complex that is formed by reversible amyloids.** The structure shown is based on the observed interaction of RNA polymerase with a low-complexity domain of a protein that regulates DNA transcription. (Adapted from I. Kwon et al., *Cell* 155:1049–1060, 2013.)

## Summary

A protein molecule's amino acid sequence determines its three-dimensional conformation. Noncovalent interactions between different parts of the polypeptide chain stabilize its folded structure. The amino acids with hydrophobic side chains tend to cluster in the interior of the molecule, and local hydrogen-bond interactions between neighboring peptide bonds give rise to  $\alpha$  helices and  $\beta$  sheets.

Regions of amino acid sequence known as domains are the modular units from which many proteins are constructed. Such domains generally contain 40–350 amino acids, often folded into a globular shape. Small proteins typically consist of only a single domain, while large proteins are formed from multiple domains linked together by various lengths of polypeptide chain, some of which can be relatively disordered. As proteins have evolved, domains have been modified and combined with other domains to construct large numbers of new proteins.

Proteins are brought together into larger structures by the same noncovalent forces that determine protein folding. Proteins with binding sites for their own surface can assemble into dimers, closed rings, spherical shells, or helical polymers. The amyloid fibril is a long unbranched structure assembled through a repeating aggregate of  $\beta$  sheets. Although some mixtures of proteins and nucleic acids can assemble spontaneously into complex structures in a test tube, not all structures in the cell are capable of spontaneous reassembly after they have been dissociated into their component parts, because many biological assembly processes involve assembly factors that are not present in the final structure.

## PROTEIN FUNCTION

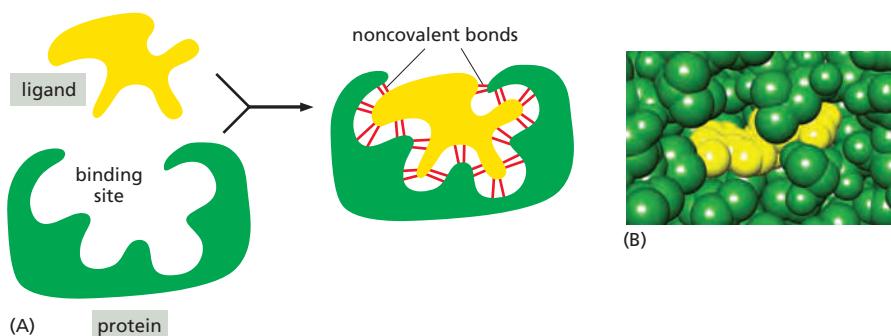
We have seen that each type of protein consists of a precise sequence of amino acids that allows it to fold up into a particular three-dimensional shape, or conformation. But proteins are not rigid lumps of material. They often have precisely engineered moving parts whose mechanical actions are coupled to chemical events. It is this coupling of chemistry and movement that gives proteins the extraordinary capabilities that underlie the dynamic processes in living cells.

In this section, we explain how proteins bind to other selected molecules and how a protein's activity depends on such binding. We show that the ability to bind to other molecules enables proteins to act as catalysts, signal receptors, switches, motors, or tiny pumps. The examples we discuss in this chapter by no means exhaust the vast functional repertoire of proteins. You will encounter the specialized functions of many other proteins elsewhere in this book, based on similar principles.

### All Proteins Bind to Other Molecules

A protein molecule's physical interaction with other molecules determines its biological properties. Thus, antibodies attach to viruses or bacteria to mark them for destruction, the enzyme hexokinase binds glucose and ATP so as to catalyze a reaction between them, actin molecules bind to each other to assemble into actin filaments, and so on. Indeed, all proteins stick, or *bind*, to other molecules. In some cases, this binding is very tight; in others it is weak and short-lived. But the binding always shows great *specificity*, in the sense that each protein molecule can usually bind just one or a few molecules out of the many thousands of different types it encounters. The substance that is bound by the protein—whether it is an ion, a small molecule, or a macromolecule such as another protein—is referred to as a **ligand** for that protein (from the Latin word *ligare*, meaning “to bind”).

The ability of a protein to bind selectively and with high affinity to a ligand depends on the formation of a set of weak noncovalent bonds—hydrogen bonds, electrostatic attractions, and van der Waals attractions—plus favorable hydrophobic interactions (see Panel 2–3, pp. 94–95). Because each individual bond is weak, effective binding occurs only when many of these bonds form simultaneously. Such binding is possible only if the surface contours of the ligand molecule fit very closely to the protein, matching it like a hand in a glove ([Figure 3–37](#)).



**Figure 3–37** The selective binding of a protein to another molecule. Many weak bonds are needed to enable a protein to bind tightly to a second molecule, or *ligand*. A ligand must therefore fit precisely into a protein's binding site, like a hand into a glove, so that a large number of noncovalent bonds form between the protein and the ligand. (A) Schematic; (B) space-filling model. (PDB code: 1G6N.)

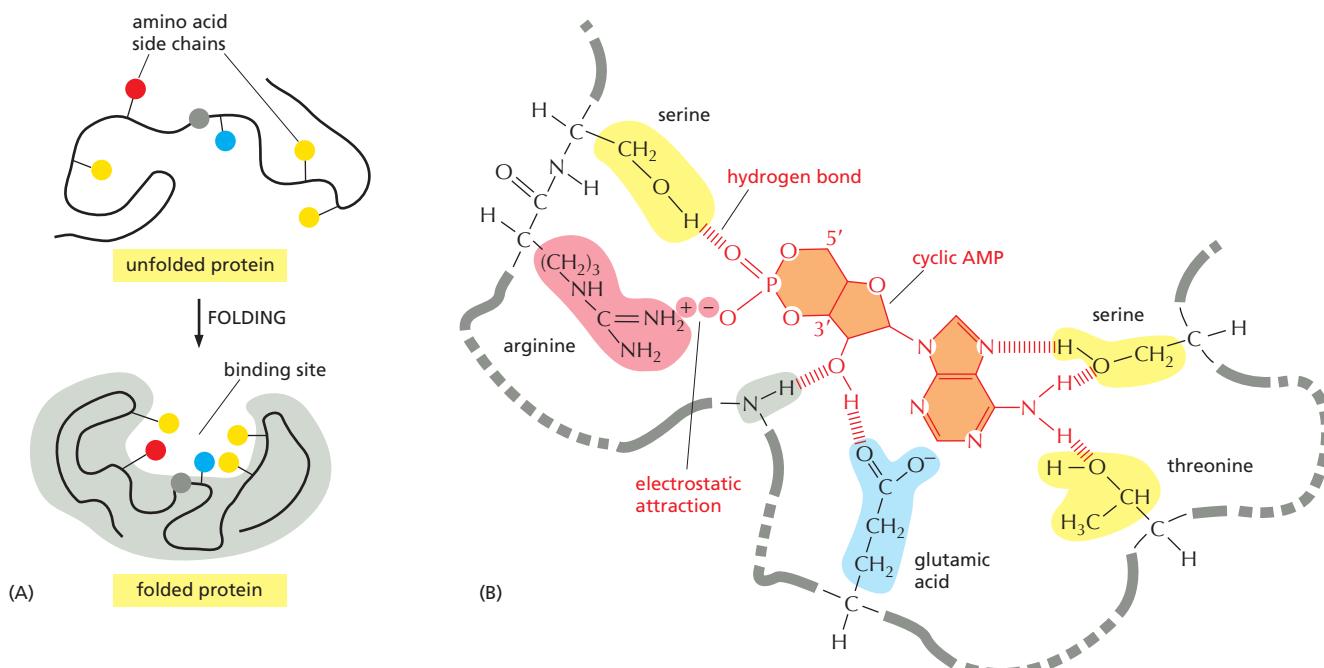
The region of a protein that associates with a ligand, known as the ligand's *binding site*, usually consists of a cavity in the protein surface formed by a particular arrangement of amino acids. These amino acids can belong to different portions of the polypeptide chain that are brought together when the protein folds (Figure 3–38). Separate regions of the protein surface generally provide binding sites for different ligands, allowing the protein's activity to be regulated, as we shall see later. And other parts of the protein act as a handle to position the protein in the cell—an example is the SH2 domain discussed previously, which often moves a protein containing it to particular intracellular sites in response to signals.

Although the atoms buried in the interior of the protein have no direct contact with the ligand, they form the framework that gives the surface its contours and its chemical and mechanical properties. Even small changes to the amino acids in the interior of a protein molecule can change its three-dimensional shape enough to destroy a binding site on the surface.

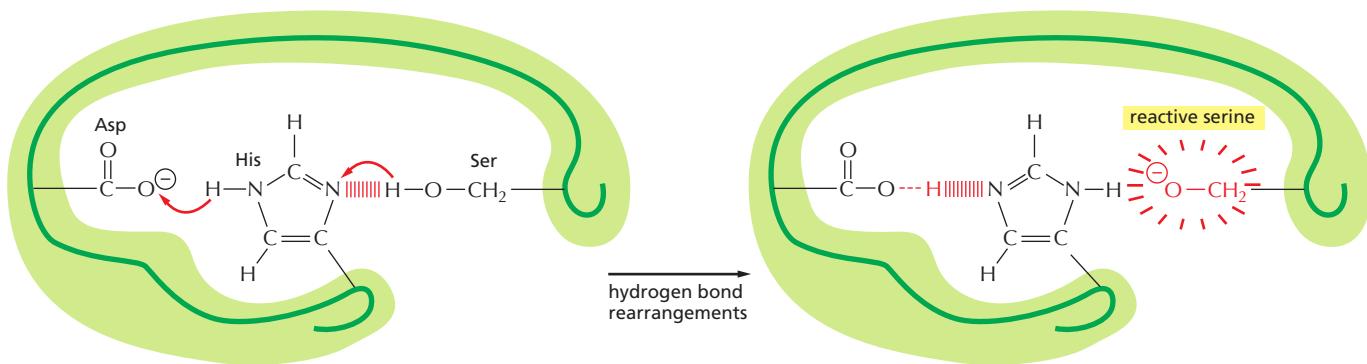
### The Surface Conformation of a Protein Determines Its Chemistry

The impressive chemical capabilities of proteins often require that the chemical groups on their surface interact in ways that enhance the chemical reactivity of one or more amino acid side chains. These interactions fall into two main categories.

First, the interaction of neighboring parts of the polypeptide chain may restrict the access of water molecules to that protein's ligand-binding sites. Because water molecules readily form hydrogen bonds that can compete with ligands for sites



**Figure 3–38** The binding site of a protein. (A) The folding of the polypeptide chain typically creates a crevice or cavity on the protein surface. This crevice contains a set of amino acid side chains disposed in such a way that they can form noncovalent bonds only with certain ligands. (B) A close-up of an actual binding site, showing the hydrogen bonds and electrostatic interactions formed between a protein and its ligand. In this example, a molecule of cyclic AMP is the bound ligand.



on the protein surface, a ligand will form tighter hydrogen bonds (and electrostatic interactions) with a protein if water molecules are kept away. It might be hard to imagine a mechanism that would exclude a molecule as small as water from a protein surface without affecting the access of the ligand itself. However, because of the strong tendency of water molecules to form water–water hydrogen bonds, water molecules exist in a large hydrogen-bonded network (see Panel 2–2, pp. 92–93). In effect, a protein can keep a ligand-binding site dry, increasing that site's reactivity, because it is energetically unfavorable for individual water molecules to break away from this network—as they must do to reach into a crevice on a protein's surface.

Second, the clustering of neighboring polar amino acid side chains can alter their reactivity. If protein folding forces together a number of negatively charged side chains against their mutual repulsion, for example, the affinity of the site for a positively charged ion is greatly increased. In addition, when amino acid side chains interact with one another through hydrogen bonds, normally unreactive groups (such as the  $-\text{CH}_2\text{OH}$  on the serine shown in Figure 3-39) can become reactive, enabling them to be used to make or break selected covalent bonds.

The surface of each protein molecule therefore has a unique chemical reactivity that depends not only on which amino acid side chains are exposed, but also on their exact orientation relative to one another. For this reason, two slightly different conformations of the same protein molecule can differ greatly in their chemistry.

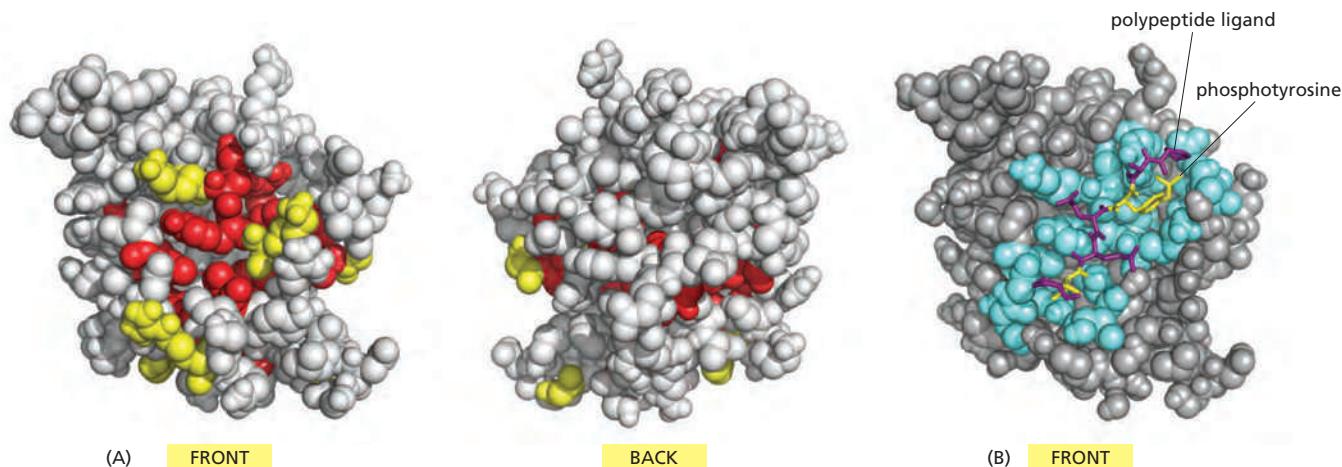
### Sequence Comparisons Between Protein Family Members Highlight Crucial Ligand-Binding Sites

As we have described previously, genome sequences allow us to group many of the domains in proteins into families that show clear evidence of their evolution from a common ancestor. The three-dimensional structures of members of the same domain family are remarkably similar. For example, even when the amino acid sequence identity falls to 25%, the backbone atoms in a domain can follow a common protein fold within 0.2 nanometers (2 Å).

We can use a method called *evolutionary tracing* to identify those sites in a protein domain that are the most crucial to the domain's function. Those sites that bind to other molecules are the most likely to be maintained, unchanged as organisms evolve. Thus, in this method, those amino acids that are unchanged, or nearly unchanged, in all of the known protein family members are mapped onto a model of the three-dimensional structure of one family member. When this is done, the most invariant positions often form one or more clusters on the protein surface, as illustrated in Figure 3-40A for the SH2 domain described previously (see Figure 3-6). These clusters generally correspond to ligand-binding sites.

The SH2 domain functions to link two proteins together. It binds the protein containing it to a second protein that contains a phosphorylated tyrosine side chain in a specific amino acid sequence context, as shown in Figure 3-40B. The amino acids located at the binding site for the phosphorylated polypeptide have been the slowest to change during the long evolutionary process that produced

**Figure 3-39** An unusually reactive amino acid at the active site of an enzyme. This example is the “catalytic triad” Asp-His-Ser found in chymotrypsin, elastase, and other serine proteases (see Figure 3-12). The aspartic acid side chain (Asp) induces the histidine (His) to remove the proton from a particular serine (Ser). This activates the serine and enables it to form a covalent bond with an enzyme substrate, hydrolyzing a peptide bond. The many convolutions of the polypeptide chain are omitted here.



the large SH2 family of peptide recognition domains. Mutation is a random process; survival is not. Thus, natural selection (random mutation followed by non-random survival) produces the sequence conservation by preferentially eliminating organisms whose SH2 domains become altered in a way that inactivates the SH2 binding site, destroying SH2 function.

Genome sequencing has revealed huge numbers of proteins whose functions are unknown. Once a three-dimensional structure has been determined for one member of a protein family, evolutionary tracing allows biologists to determine binding sites for the members of that family, providing a useful start in deciphering protein function.

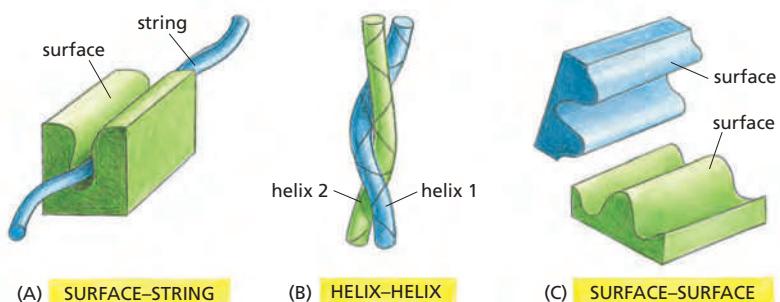
### Proteins Bind to Other Proteins Through Several Types of Interfaces

Proteins can bind to other proteins in multiple ways. In many cases, a portion of the surface of one protein contacts an extended loop of polypeptide chain (a “string”) on a second protein (Figure 3-41A). Such a surface–string interaction, for example, allows the SH2 domain to recognize a phosphorylated polypeptide loop on a second protein, as just described, and it also enables a protein kinase to recognize the proteins that it will phosphorylate (see below).

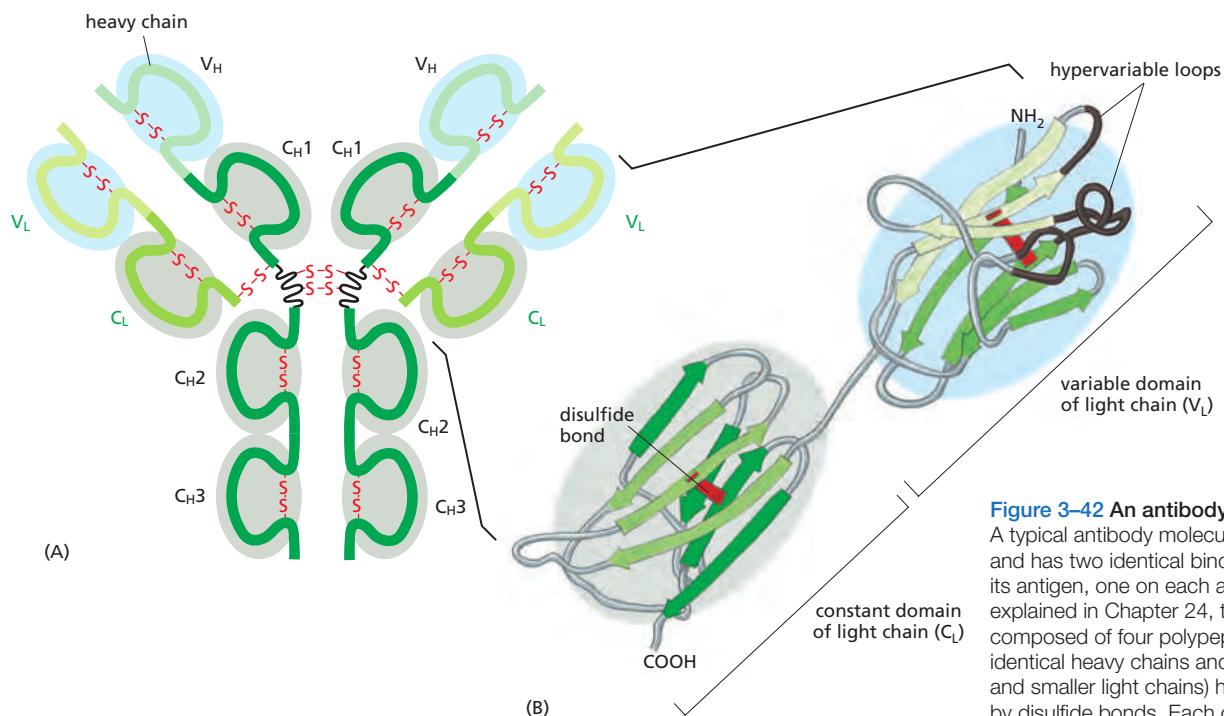
A second type of protein–protein interface forms when two  $\alpha$  helices, one from each protein, pair together to form a coiled-coil (Figure 3-41B). This type of protein interface is found in several families of gene regulatory proteins, as discussed in Chapter 7.

The most common way for proteins to interact, however, is by the precise matching of one rigid surface with that of another (Figure 3-41C). Such interactions can be very tight, since a large number of weak bonds can form between two surfaces that match well. For the same reason, such surface–surface interactions can be extremely specific, enabling a protein to select just one partner from the many thousands of different proteins found in a cell.

**Figure 3-40** The evolutionary trace method applied to the SH2 domain. (A) Front and back views of a space-filling model of the SH2 domain, with evolutionarily conserved amino acids on the protein surface colored yellow, and those more toward the protein interior colored red. (B) The structure of one specific SH2 domain with its bound polypeptide. Here, those amino acids located within 0.4 nm of the bound ligand are colored blue. The two key amino acids of the ligand are yellow, and the others are purple. Note the high degree of correspondence between (A) and (B). (Adapted from O. Lichtarge, H.R. Bourne and F.E. Cohen, *J. Mol. Biol.* 257:342–358, 1996. With permission from Elsevier; PDB codes: 1SPR, 1SPS.)



**Figure 3-41** Three ways in which two proteins can bind to each other. Only the interacting parts of the two proteins are shown. (A) A rigid surface on one protein can bind to an extended loop of polypeptide chain (a “string”) on a second protein. (B) Two  $\alpha$  helices can bind together to form a coiled-coil. (C) Two complementary rigid surfaces often link two proteins together. Binding interactions can also involve the pairing of  $\beta$  strands (see, for example, Figure 3-18).

**Figure 3–42** An antibody molecule.

A typical antibody molecule is Y-shaped and has two identical binding sites for its antigen, one on each arm of the Y. As explained in Chapter 24, the protein is composed of four polypeptide chains (two identical heavy chains and two identical and smaller light chains) held together by disulfide bonds. Each chain is made up of several different immunoglobulin domains, here shaded either blue or gray. The antigen-binding site is formed where a heavy-chain variable domain ( $V_H$ ) and a light-chain variable domain ( $V_L$ ) come close together. These are the domains that differ most in their sequence and structure in different antibodies. At the end of each of the two arms of the antibody molecule, these two domains form loops that bind to the antigen (see Movie 24.5).

### Antibody Binding Sites Are Especially Versatile

All proteins must bind to particular ligands to carry out their various functions. The antibody family is notable for its capacity for tight, highly selective binding (discussed in detail in Chapter 24).

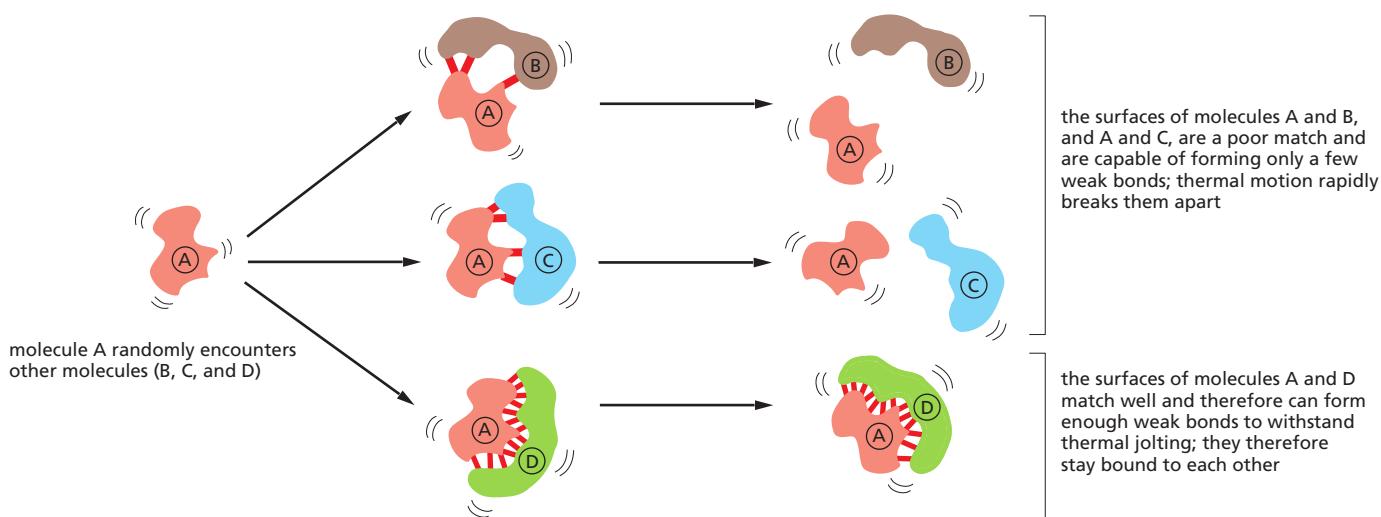
**Antibodies**, or immunoglobulins, are proteins produced by the immune system in response to foreign molecules, such as those on the surface of an invading microorganism. Each antibody binds tightly to a particular target molecule, thereby either inactivating the target molecule directly or marking it for destruction. An antibody recognizes its target (called an **antigen**) with remarkable specificity. Because there are potentially billions of different antigens that humans might encounter, we have to be able to produce billions of different antibodies.

Antibodies are Y-shaped molecules with two identical binding sites that are complementary to a small portion of the surface of the antigen molecule. A detailed examination of the antigen-binding sites of antibodies reveals that they are formed from several loops of polypeptide chain that protrude from the ends of a pair of closely juxtaposed protein domains (Figure 3–42). Different antibodies generate an enormous diversity of antigen-binding sites by changing only the length and amino acid sequence of these loops, without altering the basic protein structure.

Loops of this kind are ideal for grasping other molecules. They allow a large number of chemical groups to surround a ligand so that the protein can link to it with many weak bonds. For this reason, loops often form the ligand-binding sites in proteins.

### The Equilibrium Constant Measures Binding Strength

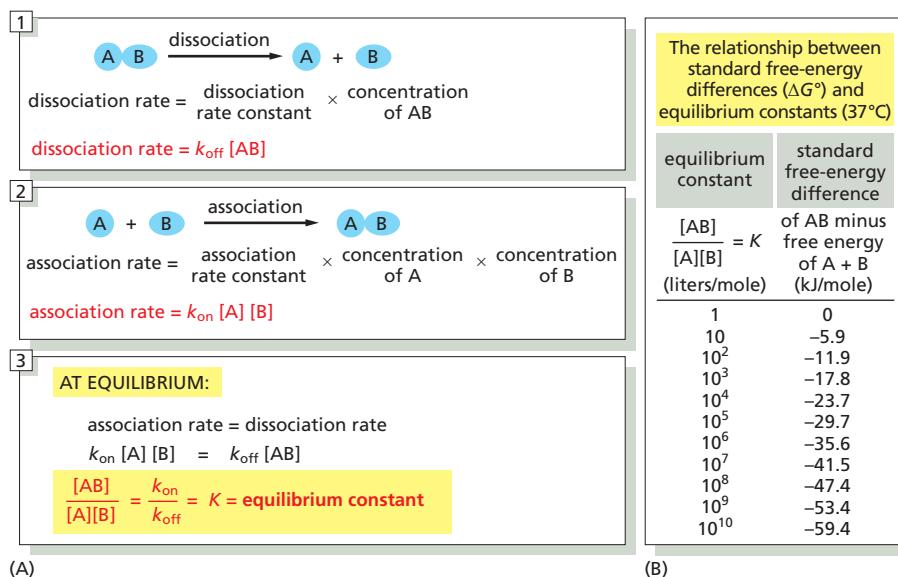
Molecules in the cell encounter each other very frequently because of their continual random thermal movements. Colliding molecules with poorly matching surfaces form few noncovalent bonds with one another, and the two molecules dissociate as rapidly as they come together. At the other extreme, when many noncovalent bonds form between two colliding molecules, the association can persist for a very long time (Figure 3–43). Strong interactions occur in cells whenever a biological function requires that molecules remain associated for a long time—for example, when a group of RNA and protein molecules come together to make a subcellular structure such as a ribosome.



We can measure the strength with which any two molecules bind to each other. As an example, consider a population of identical antibody molecules that suddenly encounters a population of ligand molecules diffusing in the fluid surrounding them. At frequent intervals, one of the ligand molecules will bump into the binding site of an antibody and form an antibody-ligand complex. The population of antibody-ligand complexes will therefore increase, but not without limit: over time, a second process, in which individual complexes break apart because of thermally induced motion, will become increasingly important. Eventually, any population of antibody molecules and ligands will reach a steady state, or equilibrium, in which the number of binding (association) events per second is precisely equal to the number of “unbinding” (dissociation) events (see Figure 2–30).

From the concentrations of the ligand, antibody, and antibody-ligand complex at equilibrium, we can calculate a convenient measure of the strength of binding—the **equilibrium constant ( $K$ )**—(Figure 3–44A). This constant was described in detail in Chapter 2, where its connection to free energy differences was derived (see p. 62). The equilibrium constant for a reaction in which two molecules (A and B) bind to each other to form a complex (AB) has units of liters/mole, and half of the binding sites will be occupied by ligand when that ligand’s concentration (in moles/liter) reaches a value that is equal to  $1/K$ . This equilibrium constant is larger the greater the binding strength, and it is a direct measure of the free-energy difference between the bound and free states (Figure 3–44B). Even a change

**Figure 3–43** How noncovalent bonds mediate interactions between macromolecules (see Movie 2.1).



(A)

(B)

**Figure 3–44** Relating standard free-energy difference ( $\Delta G^\circ$ ) to the equilibrium constant ( $K$ ). (A) The equilibrium between molecules A and B and the complex AB is maintained by a balance between the two opposing reactions shown in panels 1 and 2. Molecules A and B must collide if they are to react, and the association rate is therefore proportional to the product of their individual concentrations  $[A] \times [B]$ . (Square brackets indicate concentration.) As shown in panel 3, the ratio of the rate constants for the association and the dissociation reactions is equal to the equilibrium constant ( $K$ ) for the reaction (see also p. 63). (B) The equilibrium constant in panel 3 is that for the reaction  $A + B \leftrightarrow AB$ , and the larger its value, the stronger the binding between A and B. Note that for every 5.91 kJ/mole decrease in standard free energy, the equilibrium constant increases by a factor of 10 at 37°C.

The equilibrium constant here has units of liters/mole; for simple binding interactions it is also called the *affinity constant* or *association constant*, denoted  $K_a$ . The reciprocal of  $K_a$  is called the *dissociation constant*,  $K_d$  (in units of moles/liter).

of a few noncovalent bonds can have a striking effect on a binding interaction, as shown by the example in **Figure 3–45**. (Note that the equilibrium constant, as defined here, is also known as the association or affinity constant,  $K_a$ .)

We have used the case of an antibody binding to its ligand to illustrate the effect of binding strength on the equilibrium state, but the same principles apply to any molecule and its ligand. Many proteins are enzymes, which, as we now discuss, first bind to their ligands and then catalyze the breakage or formation of covalent bonds in these molecules.

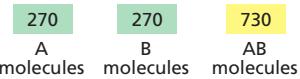
### Enzymes Are Powerful and Highly Specific Catalysts

Many proteins can perform their function simply by binding to another molecule. An actin molecule, for example, need only associate with other actin molecules to form a filament. There are other proteins, however, for which ligand binding is only a necessary first step in their function. This is the case for the large and very important class of proteins called **enzymes**. As described in Chapter 2, enzymes are remarkable molecules that cause the chemical transformations that make and break covalent bonds in cells. They bind to one or more ligands, called **substrates**, and convert them into one or more chemically modified *products*, doing this over and over again with amazing rapidity. Enzymes speed up reactions, often by a factor of a million or more, without themselves being changed—that is, they act as **catalysts** that permit cells to make or break covalent bonds in a controlled way. It is the catalysis of organized sets of chemical reactions by enzymes that creates and maintains the cell, making life possible.

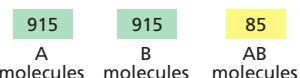
We can group enzymes into functional classes that perform similar chemical reactions (**Table 3–1**). Each type of enzyme within such a class is highly specific,

Consider 1000 molecules of A and 1000 molecules of B in a eukaryotic cell. The concentration of both will be about  $10^{-9}$  M.

If the equilibrium constant ( $K$ ) for  $A + B \rightleftharpoons AB$  is  $10^{10}$ , then one can calculate that at equilibrium there will be



If the equilibrium constant is a little weaker at  $10^8$ , which represents a loss of 11.9 kilojoule/mole of binding energy from the example above, or 2–3 fewer hydrogen bonds, then there will be

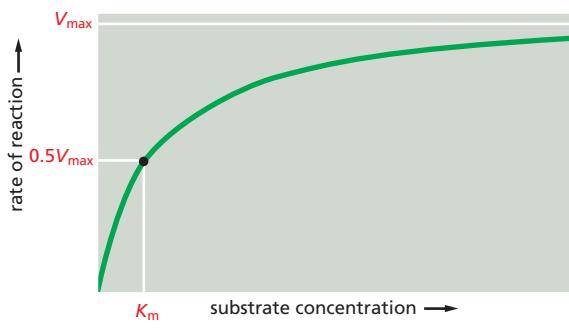


**Figure 3–45** Small changes in the number of weak bonds can have drastic effects on a binding interaction. This example illustrates the dramatic effect of the presence or absence of a few weak noncovalent bonds in a biological context.

**TABLE 3–1** Some Common Types of Enzymes

Enzyme	Reaction catalyzed
Hydrolases	General term for enzymes that catalyze a hydrolytic cleavage reaction; <i>nucleases</i> and <i>proteases</i> are more specific names for subclasses of these enzymes
Nucleases	Break down nucleic acids by hydrolyzing bonds between nucleotides. <i>Endo-</i> and <i>exonucleases</i> cleave nucleic acids <i>within</i> and <i>from the ends of</i> the polynucleotide chains, respectively
Proteases	Break down proteins by hydrolyzing bonds between amino acids
Synthases	Synthesize molecules in anabolic reactions by condensing two smaller molecules together
Ligases	Join together (ligate) two molecules in an energy-dependent process. DNA ligase, for example, joins two DNA molecules together end-to-end through phosphodiester bonds
Isomerases	Catalyze the rearrangement of bonds within a single molecule
Polymerases	Catalyze polymerization reactions such as the synthesis of DNA and RNA
Kinases	Catalyze the addition of phosphate groups to molecules. Protein kinases are an important group of kinases that attach phosphate groups to proteins
Phosphatases	Catalyze the hydrolytic removal of a phosphate group from a molecule
Oxido-Reductases	General name for enzymes that catalyze reactions in which one molecule is oxidized while the other is reduced. Enzymes of this type are often more specifically named <i>oxidases</i> , <i>reductases</i> , or <i>dehydrogenases</i>
ATPases	Hydrolyze ATP. Many proteins with a wide range of roles have an energy-harnessing ATPase activity as part of their function; for example, motor proteins such as <i>myosin</i> and membrane transport proteins such as the <i>sodium-potassium pump</i>
GTPases	Hydrolyze GTP. A large family of GTP-binding proteins are GTPases with central roles in the regulation of cell processes

Enzyme names typically end in “-ase,” with the exception of some enzymes, such as pepsin, trypsin, thrombin, and lysozyme, that were discovered and named before the convention became generally accepted at the end of the nineteenth century. The common name of an enzyme usually indicates the substrate or product and the nature of the reaction catalyzed. For example, citrate synthase catalyzes the synthesis of citrate by a reaction between acetyl CoA and oxaloacetate.



**Figure 3–46 Enzyme kinetics.** The rate of an enzyme reaction ( $V$ ) increases as the substrate concentration increases until a maximum value ( $V_{max}$ ) is reached. At this point all substrate-binding sites on the enzyme molecules are fully occupied, and the rate of reaction is limited by the rate of the catalytic process on the enzyme surface. For most enzymes, the concentration of substrate at which the reaction rate is half-maximal ( $K_m$ ) is a measure of how tightly the substrate is bound, with a large value of  $K_m$  corresponding to weak binding.

catalyzing only a single type of reaction. Thus, *hexokinase* adds a phosphate group to D-glucose but ignores its optical isomer L-glucose; the blood-clotting enzyme *thrombin* cuts one type of blood protein between a particular arginine and its adjacent glycine and nowhere else, and so on. As discussed in detail in Chapter 2, enzymes work in teams, with the product of one enzyme becoming the substrate for the next. The result is an elaborate network of metabolic pathways that provides the cell with energy and generates the many large and small molecules that the cell needs (see Figure 2–63).

### Substrate Binding Is the First Step in Enzyme Catalysis

For a protein that catalyzes a chemical reaction (an enzyme), the binding of each substrate molecule to the protein is an essential prelude. In the simplest case, if we denote the enzyme by E, the substrate by S, and the product by P, the basic reaction path is  $E + S \rightarrow ES \rightarrow EP \rightarrow E + P$ . There is a limit to the amount of substrate that a single enzyme molecule can process in a given time. Although an increase in the concentration of substrate increases the rate at which product is formed, this rate eventually reaches a maximum value (Figure 3–46). At that point the enzyme molecule is saturated with substrate, and the rate of reaction ( $V_{max}$ ) depends only on how rapidly the enzyme can process the substrate molecule. This maximum rate divided by the enzyme concentration is called the *turnover number*. Turnover numbers are often about 1000 substrate molecules processed per second per enzyme molecule, although turnover numbers between 1 and 10,000 are known.

The other kinetic parameter frequently used to characterize an enzyme is its  $K_m$ , the concentration of substrate that allows the reaction to proceed at one-half its maximum rate ( $0.5 V_{max}$ ) (see Figure 3–46). A *low  $K_m$*  value means that the enzyme reaches its maximum catalytic rate at a *low concentration* of substrate and generally indicates that the enzyme binds to its substrate very tightly, whereas a *high  $K_m$*  value corresponds to weak binding. The methods used to characterize enzymes in this way are explained in Panel 3–2 (pp. 142–143).

### Enzymes Speed Reactions by Selectively Stabilizing Transition States

Enzymes achieve extremely high rates of chemical reaction—rates that are far higher than for any synthetic catalysts. There are several reasons for this efficiency. First, when two molecules need to react, the enzyme greatly increases the local concentration of both of these substrate molecules at the catalytic site, holding them in the correct orientation for the reaction that is to follow. More importantly, however, some of the binding energy contributes directly to the catalysis. Substrate molecules must pass through a series of intermediate states of altered geometry and electron distribution before they form the ultimate products of the reaction. The free energy required to attain the most unstable intermediate state, called the **transition state**, is known as the *activation energy* for the reaction, and it is the major determinant of the reaction rate. Enzymes have a much higher affinity for the transition state of the substrate than they have for the stable form.

## WHY ANALYZE THE KINETICS OF ENZYMES?

Enzymes are the most selective and powerful catalysts known. An understanding of their detailed mechanisms provides a critical tool for the discovery of new drugs, for the large-scale industrial synthesis of useful chemicals, and for appreciating the chemistry of cells and organisms. A detailed study of the rates of the chemical reactions that are catalyzed by a purified enzyme—more specifically how these rates change with changes in conditions such as the concentrations of substrates, products, inhibitors, and regulatory ligands—allows

biochemists to figure out exactly how each enzyme works. For example, this is the way that the ATP-producing reactions of glycolysis, shown previously in Figure 2–48, were deciphered—allowing us to appreciate the rationale for this critical enzymatic pathway.

In this Panel, we introduce the important field of **enzyme kinetics**, which has been indispensable for deriving much of the detailed knowledge that we now have about cell chemistry.

## STEADY-STATE ENZYME KINETICS

Many enzymes have only one substrate, which they bind and then process to produce products according to the scheme outlined in Figure 3–50A. In this case, the reaction is written as

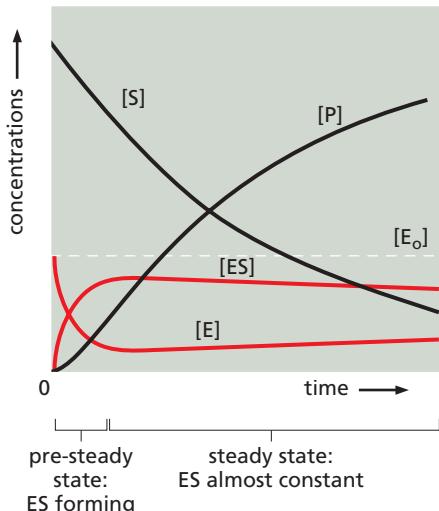


Here we have assumed that the reverse reaction, in which  $E + P$  recombine to form  $EP$  and then  $ES$ , occurs so rarely that we can ignore it. In this case,  $EP$  need not be represented, and we can express the rate of the reaction—known as its velocity,  $V$ , as

$$V = k_{\text{cat}} [ES]$$

where  $[ES]$  is the concentration of the enzyme–substrate complex, and  $k_{\text{cat}}$  is the **turnover number**, a rate constant that has a value equal to the number of substrate molecules processed per enzyme molecule each second.

But how does the value of  $[ES]$  relate to the concentrations that we know directly, which are the total concentration of the enzyme,  $[E_o]$ , and the concentration of the substrate,  $[S]$ ? When enzyme and substrate are first mixed, the concentration  $[ES]$  will rise rapidly from zero to a so-called **steady-state** level, as illustrated below.



At this steady state,  $[ES]$  is nearly constant, so that

$$\frac{\text{rate of } ES \text{ breakdown}}{k_{-1} [ES] + k_{\text{cat}} [ES]} = \frac{\text{rate of } ES \text{ formation}}{k_1 [E][S]}$$

or, since the concentration of the free enzyme,  $[E]$ , is equal to  $[E_o] - [ES]$ ,

$$[ES] = \left( \frac{k_1}{k_{-1} + k_{\text{cat}}} \right) [E][S] = \left( \frac{k_1}{k_{-1} + k_{\text{cat}}} \right) ([E_o] - [ES])[S]$$

Rearranging, and defining the constant  $K_m$  as

$$\frac{k_{-1} + k_{\text{cat}}}{k_1}$$

we get

$$[ES] = \frac{[E_o][S]}{K_m + [S]}$$

or, remembering that  $V = k_{\text{cat}} [ES]$ , we obtain the famous Michaelis–Menten equation

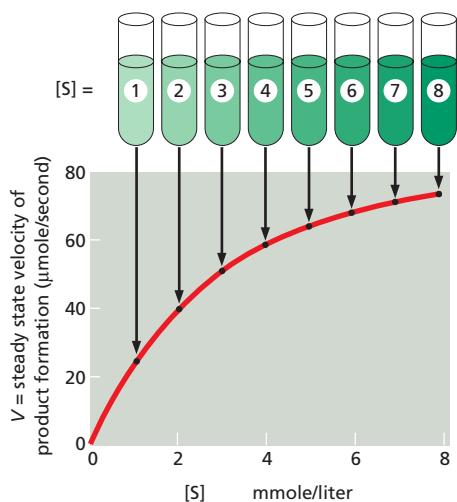
$$V = \frac{k_{\text{cat}} [E_o][S]}{K_m + [S]}$$

As  $[S]$  is increased to higher and higher levels, essentially all of the enzyme will be bound to substrate at steady state; at this point, a maximum rate of reaction,  $V_{\text{max}}$ , will be reached where  $V = V_{\text{max}} = k_{\text{cat}} [E_o]$ . Thus, it is convenient to rewrite the Michaelis–Menten equation as

$$V = \frac{V_{\text{max}} [S]}{K_m + [S]}$$

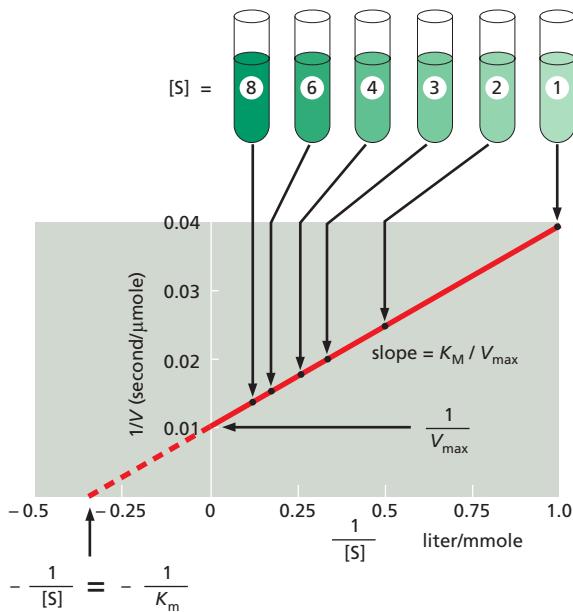
### THE DOUBLE-RECIPROCAL PLOT

A typical plot of  $V$  versus  $[S]$  for an enzyme that follows Michaelis–Menten kinetics is shown below. From this plot, neither the value of  $V_{\max}$  nor of  $K_m$  is immediately clear.



To obtain  $V_{\max}$  and  $K_m$  from such data, a double-reciprocal plot is often used, in which the Michaelis–Menten equation has merely been rearranged, so that  $1/V$  can be plotted versus  $1/[S]$ .

$$\frac{1}{V} = \left( \frac{K_m}{V_{\max}} \right) \left( \frac{1}{[S]} \right) + \frac{1}{V_{\max}}$$



### THE SIGNIFICANCE OF $K_m$ , $k_{\text{cat}}$ , and $k_{\text{cat}}/K_m$

As described in the text,  $K_m$  is an approximate measure of substrate affinity for the enzyme: it is numerically equal to the concentration of  $[S]$  at  $V = 0.5 V_{\max}$ . In general, a lower value of  $K_m$  means tighter substrate binding. In fact, for those cases where  $k_{\text{cat}}$  is much smaller than  $k_{-1}$ , the  $K_m$  will be equal to  $K_d$ , the dissociation constant for substrate binding to the enzyme ( $K_d = 1/K_a$ ; see Figure 3–44).

We have seen that  $k_{\text{cat}}$  is the turnover number for the enzyme. At very low substrate concentrations, where  $[S] \ll K_m$ , most of the enzyme is free. Thus we can think of  $[E] = [E_0]$ , so that the Michaelis–Menten equation becomes  $V = k_{\text{cat}}/K_m [E][S]$ . Thus, the ratio  $k_{\text{cat}}/K_m$  is equivalent to the rate constant for the reaction between free enzyme and free substrate.

A comparison of  $k_{\text{cat}}/K_m$  for the same enzyme with different substrates, or for two enzymes with their different substrates, is widely used as a measure of enzyme effectiveness.

For simplicity, in this Panel we have discussed enzymes that have only one substrate, such as the lysozyme enzyme described in the text (see p. 144). Most enzymes have two substrates, one of which is often an active carrier molecule—such as NADH or ATP.

A similar, but more complex, analysis is used to determine the kinetics of such enzymes—allowing the order of substrate binding and the presence of covalent intermediates along the pathway to be revealed.

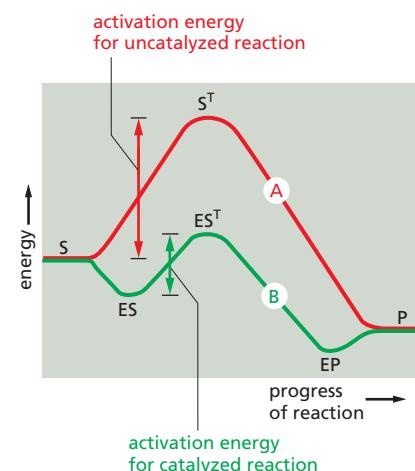
### SOME ENZYMES ARE DIFFUSION LIMITED

The values of  $k_{\text{cat}}$ ,  $K_m$ , and  $k_{\text{cat}}/K_m$  for some selected enzymes are given below:

enzyme	substrate	$k_{\text{cat}}$ (sec $^{-1}$ )	$K_m$ (M)	$k_{\text{cat}}/K_m$ (sec $^{-1}$ M $^{-1}$ )
acetylcholinesterase	acetylcholine	$1.4 \times 10^4$	$9 \times 10^{-5}$	$1.6 \times 10^8$
catalase	$H_2O_2$	$4 \times 10^7$	1	$4 \times 10^7$
fumarase	fumarate	$8 \times 10^2$	$5 \times 10^{-6}$	$1.6 \times 10^8$

Because an enzyme and its substrate must collide before they can react,  $k_{\text{cat}}/K_m$  has a maximum possible value that is limited by collision rates. If every collision forms an enzyme–substrate complex, one can calculate from diffusion theory that  $k_{\text{cat}}/K_m$  will be between  $10^8$  and  $10^9$  sec $^{-1}$ M $^{-1}$ , in the case where all subsequent steps proceed immediately. Thus, it is claimed that enzymes like acetylcholinesterase and fumarase are “perfect enzymes,” each enzyme having evolved to the point where nearly every collision with its substrate converts the substrate to a product.

**Figure 3–47 Enzymatic acceleration of chemical reactions by decreasing the activation energy.** There is a single transition state in this example. However, often both the uncatalyzed reaction (A) and the enzyme-catalyzed reaction (B) go through a series of transition states. In that case, it is the transition state with the highest energy ( $S^T$  and  $ES^T$ ) that determines the activation energy and limits the rate of the reaction. (S = substrate; P = product of the reaction; ES = enzyme–substrate complex; EP = enzyme–product complex.)



Because this tight binding greatly lowers the energy of the transition state, the enzyme greatly accelerates a particular reaction by lowering the activation energy that is required (Figure 3–47).

### Enzymes Can Use Simultaneous Acid and Base Catalysis

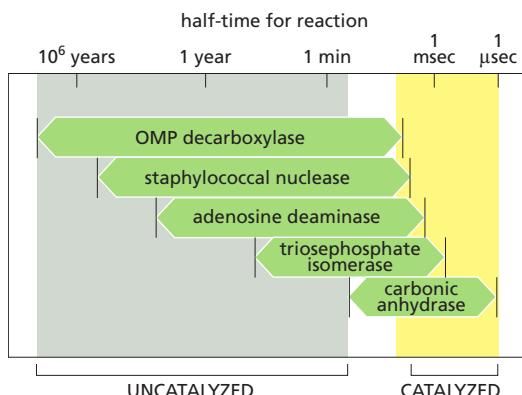
**Figure 3–48** compares the spontaneous reaction rates and the corresponding enzyme-catalyzed rates for five enzymes. Rate accelerations range from  $10^9$  to  $10^{23}$ . Enzymes not only bind tightly to a transition state, they also contain precisely positioned atoms that alter the electron distributions in the atoms that participate directly in the making and breaking of covalent bonds. Peptide bonds, for example, can be hydrolyzed in the absence of an enzyme by exposing a polypeptide to either a strong acid or a strong base. Enzymes are unique, however, in being able to use acid and base catalysis simultaneously, because the rigid framework of the protein constrains the acidic and basic residues and prevents them from combining with each other, as they would do in solution (Figure 3–49).

The fit between an enzyme and its substrate needs to be precise. A small change introduced by genetic engineering in the active site of an enzyme can therefore have a profound effect. Replacing a glutamic acid with an aspartic acid in one enzyme, for example, shifts the position of the catalytic carboxylate ion by only 1 Å (about the radius of a hydrogen atom); yet this is enough to decrease the activity of the enzyme a thousandfold.

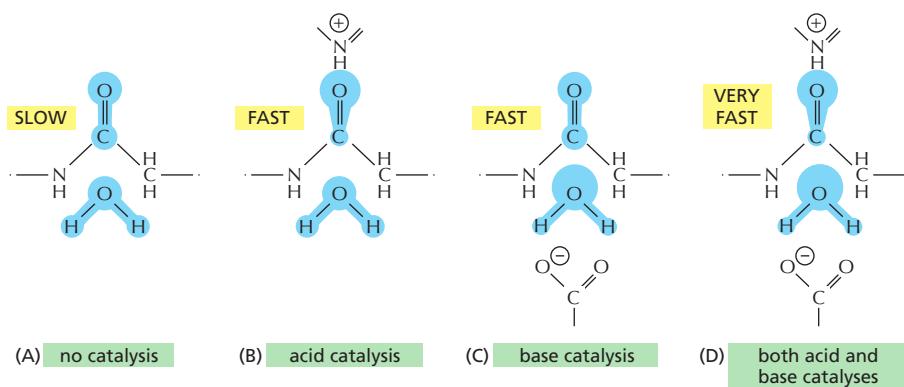
### Lysozyme Illustrates How an Enzyme Works

To demonstrate how enzymes catalyze chemical reactions, we examine an enzyme that acts as a natural antibiotic in egg white, saliva, tears, and other secretions. Lysozyme catalyzes the cutting of polysaccharide chains in the cell walls of bacteria. The bacterial cell is under pressure from osmotic forces, and cutting even a small number of these chains causes the cell wall to rupture and the cell to burst. A relatively small and stable protein that can be easily isolated in large quantities, lysozyme was the first enzyme to have its structure worked out in atomic detail by x-ray crystallography (in the mid-1960s).

The reaction that lysozyme catalyzes is a hydrolysis: it adds a molecule of water to a single bond between two adjacent sugar groups in the polysaccharide chain, thereby causing the bond to break (see Figure 2–9). The reaction is energetically favorable because the free energy of the severed polysaccharide chain is lower



**Figure 3–48** The rate accelerations caused by five different enzymes.  
(Adapted from A. Radzicka and R. Wolfenden, *Science* 267:90–93, 1995.)



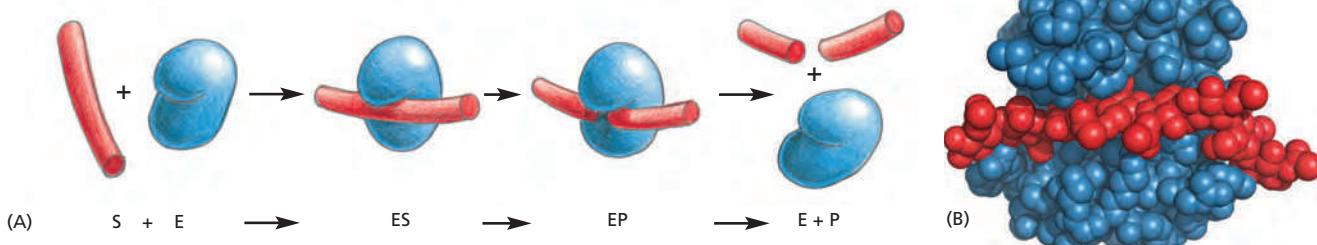
than the free energy of the intact chain. However, there is an energy barrier to the reaction, and a colliding water molecule can break a bond linking two sugars only if the polysaccharide molecule is distorted into a particular shape—the transition state—in which the atoms around the bond have an altered geometry and electron distribution. Because of this requirement, random collisions must supply a very large activation energy for the reaction to take place. In an aqueous solution at room temperature, the energy of collisions almost never exceeds the activation energy. The pure polysaccharide can therefore remain for years in water without being hydrolyzed to any detectable degree.

This situation changes drastically when the polysaccharide binds to lysozyme. The active site of lysozyme, because its substrate is a polymer, is a long groove that holds six linked sugars at the same time. As soon as the polysaccharide binds to form an enzyme–substrate complex, the enzyme cuts the polysaccharide by adding a water molecule across one of its sugar-sugar bonds. The product chains are then quickly released, freeing the enzyme for further cycles of reaction (Figure 3-50).

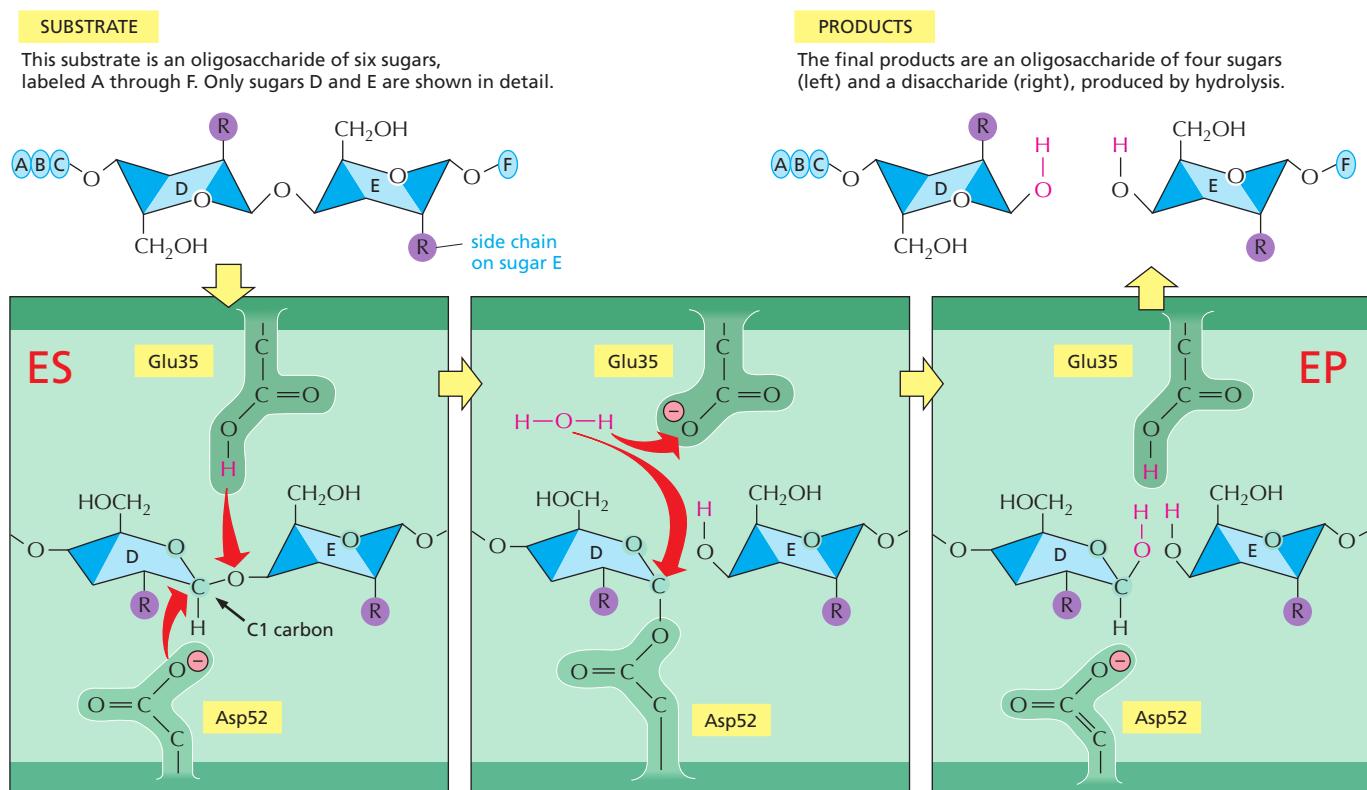
An impressive increase in hydrolysis rate is possible because conditions are created in the microenvironment of the lysozyme active site that greatly reduce the activation energy necessary for the hydrolysis to take place. In particular, lysozyme distorts one of the two sugars connected by the bond to be broken from its normal, most stable conformation. The bond to be broken is also held close to two amino acids with acidic side chains (a glutamic acid and an aspartic acid) that participate directly in the reaction. Figure 3-51 shows the three central steps in this enzymatically catalyzed reaction, which occurs millions of times faster than uncatalyzed hydrolysis.

Other enzymes use similar mechanisms to lower activation energies and speed up the reactions they catalyze. In reactions involving two or more reactants, the active site also acts like a template, or mold, that brings the substrates together in the proper orientation for a reaction to occur between them (Figure 3-52A). As we saw for lysozyme, the active site of an enzyme contains precisely positioned

**Figure 3-49 Acid catalysis and base catalysis.** (A) The start of the uncatalyzed reaction that hydrolyzes a peptide bond, with blue shading used to indicate electron distribution in the water and carbonyl bonds. (B) An acid likes to donate a proton ( $H^+$ ) to other atoms. By pairing with the carbonyl oxygen, an acid causes electrons to move away from the carbonyl carbon, making this atom much more attractive to the electronegative oxygen of an attacking water molecule. (C) A base likes to take up  $H^+$ . By pairing with a hydrogen of the attacking water molecule, a base causes electrons to move toward the water oxygen, making it a better attacking group for the carbonyl carbon. (D) By having appropriately positioned atoms on its surface, an enzyme can perform both acid catalysis and base catalysis at the same time.



**Figure 3-50 The reaction catalyzed by lysozyme.** (A) The enzyme lysozyme (E) catalyzes the cutting of a polysaccharide chain, which is its substrate (S). The enzyme first binds to the chain to form an enzyme–substrate complex (ES) and then catalyzes the cleavage of a specific covalent bond in the backbone of the polysaccharide, forming an enzyme–product complex (EP) that rapidly dissociates. Release of the severed chain (the products P) leaves the enzyme free to act on another substrate molecule. (B) A space-filling model of the lysozyme molecule bound to a short length of polysaccharide chain before cleavage (Movie 3.8). (B, courtesy of Richard J. Feldmann; PDB code: 3AB6.)



In the enzyme–substrate complex (ES), the enzyme forces sugar D into a strained conformation. The Glu35 in the enzyme is positioned to serve as an acid that attacks the adjacent sugar–sugar bond by donating a proton ( $H^+$ ) to sugar E; Asp52 is poised to attack the C1 carbon atom.

The Asp52 has formed a covalent bond between the enzyme and the C1 carbon atom of sugar D. The Glu35 then polarizes a water molecule (red), so that its oxygen can readily attack the C1 carbon atom and displace Asp52.

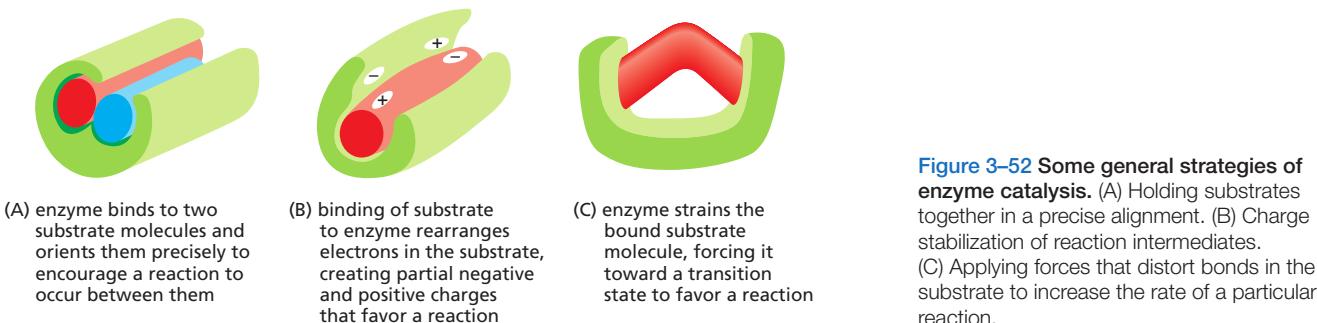
The reaction of the water molecule (red) completes the hydrolysis and returns the enzyme to its initial state, forming the final enzyme–product complex (EP).

atoms that speed up a reaction by using charged groups to alter the distribution of electrons in the substrates (Figure 3–52B). And as we have also seen, when a substrate binds to an enzyme, bonds in the substrate are often distorted, changing the substrate shape. These changes, along with mechanical forces, drive a substrate toward a particular transition state (Figure 3–52C). Finally, like lysozyme, many enzymes participate intimately in the reaction by transiently forming a covalent bond between the substrate and a side chain of the enzyme. Subsequent steps in the reaction restore the side chain to its original state, so that the enzyme remains unchanged after the reaction (see also Figure 2–48).

### Tightly Bound Small Molecules Add Extra Functions to Proteins

Although we have emphasized the versatility of enzymes—and proteins in general—as chains of amino acids that perform remarkable functions, there are many instances in which the amino acids by themselves are not enough. Just as humans

**Figure 3–51 Events at the active site of lysozyme.** The top left and top right drawings show the free substrate and the free products, respectively, whereas the other three drawings show the sequential events at the enzyme active site. Note the change in the conformation of sugar D in the enzyme–substrate complex; this shape change stabilizes the oxocarbenium ion-like transition states required for formation and hydrolysis of the covalent intermediate shown in the middle panel. It is also possible that a carbonium ion intermediate forms in step 2, but the covalent intermediate shown in the middle panel has been detected with a synthetic substrate (Movie 3.9). (See D.J. Vocablo et al., *Nature* 412:835–838, 2001.)

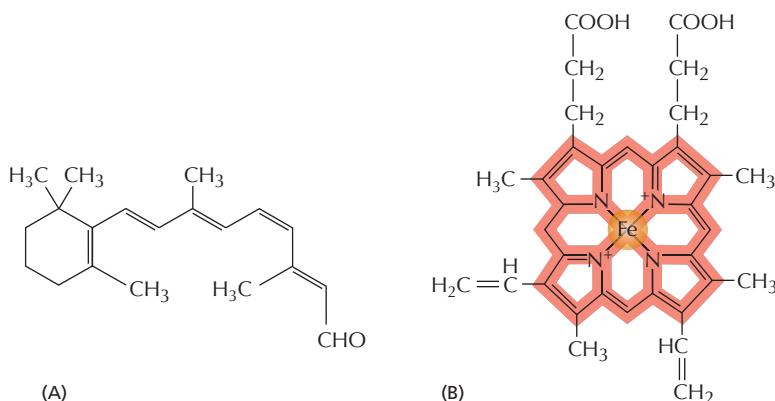


**TABLE 3–2** Many Vitamin Derivatives Are Critical Coenzymes for Human Cells

Vitamin	Coenzyme	Enzyme-catalyzed reactions requiring these coenzymes
Thiamine (vitamin B <sub>1</sub> )	Thiamine pyrophosphate	Activation and transfer of aldehydes
Riboflavin (vitamin B <sub>2</sub> )	FADH	Oxidation–reduction
Niacin	NADH, NADPH	Oxidation–reduction
Pantothenic acid	Coenzyme A	Acyl group activation and transfer
Pyridoxine	Pyridoxal phosphate	Amino acid activation; also glycogen phosphorylase
Biotin	Biotin	CO <sub>2</sub> activation and transfer
Lipoic acid	Lipoamide	Acyl group activation; oxidation–reduction
Folic acid	Tetrahydrofolate	Activation and transfer of single carbon groups
Vitamin B <sub>12</sub>	Cobalamin coenzymes	Isomerization and methyl group transfers

employ tools to enhance and extend the capabilities of their hands, enzymes and other proteins often use small nonprotein molecules to perform functions that would be difficult or impossible to do with amino acids alone. Thus, enzymes frequently have a small molecule or metal atom tightly associated with their active site that assists with their catalytic function. *Carboxypeptidase*, for example, an enzyme that cuts polypeptide chains, carries a tightly bound zinc ion in its active site. During the cleavage of a peptide bond by carboxypeptidase, the zinc ion forms a transient bond with one of the substrate atoms, thereby assisting the hydrolysis reaction. In other enzymes, a small organic molecule serves a similar purpose. Such organic molecules are often referred to as **coenzymes**. An example is *biotin*, which is found in enzymes that transfer a carboxylate group ( $-COO^-$ ) from one molecule to another (see Figure 2–40). Biotin participates in these reactions by forming a transient covalent bond to the  $-COO^-$  group to be transferred, being better suited to this function than any of the amino acids used to make proteins. Because it cannot be synthesized by humans, and must therefore be supplied in small quantities in our diet, biotin is a *vitamin*. Many other coenzymes are either vitamins or derivatives of vitamins (Table 3–2).

Other proteins also frequently require specific small-molecule adjuncts to function properly. Thus, the signal receptor protein *rhodopsin*, which is made by the photoreceptor cells in the retina, detects light by means of a small molecule, *retinal*, embedded in the protein (Figure 3–53A). Retinal, which is derived from vitamin A, changes its shape when it absorbs a photon of light, and this change causes the protein to trigger a cascade of enzymatic reactions that eventually lead to an electrical signal being carried to the brain.



**Figure 3–53** Retinal and heme. (A) The structure of retinal, the light-sensitive molecule attached to rhodopsin in the eye. The structure shown isomerizes when it absorbs light. (B) The structure of a heme group. The carbon-containing heme ring is red and the iron atom at its center is orange. A heme group is tightly bound to each of the four polypeptide chains in hemoglobin, the oxygen-carrying protein whose structure is shown in Figure 3–19.

Another example of a protein with a nonprotein portion is hemoglobin (see Figure 3–19). Each molecule of hemoglobin carries four *heme* groups, ring-shaped molecules each with a single central iron atom (Figure 3–53B). Heme gives hemoglobin (and blood) its red color. By binding reversibly to oxygen gas through its iron atom, heme enables hemoglobin to pick up oxygen in the lungs and release it in the tissues.

Sometimes these small molecules are attached covalently and permanently to their protein, thereby becoming an integral part of the protein molecule itself. We shall see in Chapter 10 that proteins are often anchored to cell membranes through covalently attached lipid molecules. And membrane proteins exposed on the surface of the cell, as well as proteins secreted outside the cell, are often modified by the covalent addition of sugars and oligosaccharides.

### Multienzyme Complexes Help to Increase the Rate of Cell Metabolism

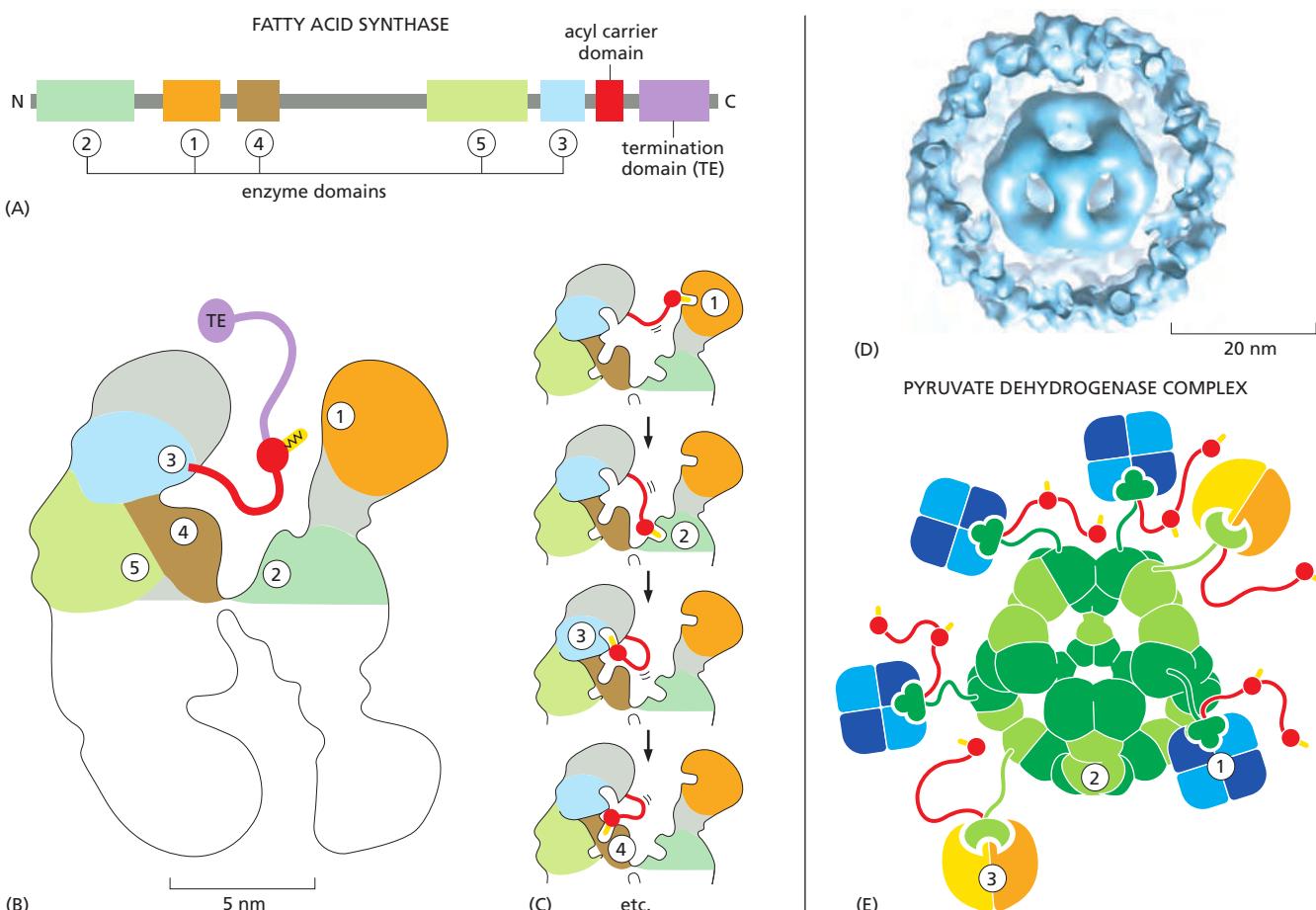
The efficiency of enzymes in accelerating chemical reactions is crucial to the maintenance of life. Cells, in effect, must race against the unavoidable processes of decay, which—if left unattended—cause macromolecules to run downhill toward greater and greater disorder. If the rates of desirable reactions were not greater than the rates of competing side reactions, a cell would soon die. We can get some idea of the rate at which cell metabolism proceeds by measuring the rate of ATP utilization. A typical mammalian cell “turns over” (i.e., hydrolyzes and restores by phosphorylation) its entire ATP pool once every 1 or 2 minutes. For each cell, this turnover represents the utilization of roughly  $10^7$  molecules of ATP per second (or, for the human body, about 1 gram of ATP every minute).

The rates of reactions in cells are rapid because enzyme catalysis is so effective. Some enzymes have become so efficient that there is no possibility of further useful improvement. The factor that limits the reaction rate is no longer the enzyme’s intrinsic speed of action; rather, it is the frequency with which the enzyme collides with its substrate. Such a reaction is said to be *diffusion-limited* (see Panel 3–2, pp. 142–143).

The amount of product produced by an enzyme will depend on the concentration of both the enzyme and its substrate. If a sequence of reactions is to occur extremely rapidly, each metabolic intermediate and enzyme involved must be present in high concentration. However, given the enormous number of different reactions performed by a cell, there are limits to the concentrations that can be achieved. In fact, most metabolites are present in micromolar ( $10^{-6}$  M) concentrations, and most enzyme concentrations are much lower. How is it possible, therefore, to maintain very fast metabolic rates?

The answer lies in the spatial organization of cell components. The cell can increase reaction rates without raising substrate concentrations by bringing the various enzymes involved in a reaction sequence together to form a large protein assembly known as a *multienzyme complex* (Figure 3–54). Because this assembly is organized in a way that allows the product of enzyme A to be passed directly to enzyme B, and so on, diffusion rates need not be limiting, even when the concentrations of the substrates in the cell as a whole are very low. It is perhaps not surprising, therefore, that such enzyme complexes are very common, and they are involved in nearly all aspects of metabolism—including the central genetic processes of DNA, RNA, and protein synthesis. In fact, few enzymes in eukaryotic cells diffuse freely in solution; instead, most seem to have evolved binding sites that concentrate them with other proteins of related function in particular regions of the cell, thereby increasing the rate and efficiency of the reactions that they catalyze (see p. 331).

Eukaryotic cells have yet another way of increasing the rate of metabolic reactions: using their intracellular membrane systems. These membranes can segregate particular substrates and the enzymes that act on them into the same membrane-enclosed compartment, such as the endoplasmic reticulum or the cell nucleus. If, for example, a compartment occupies a total of 10% of the volume of



**Figure 3–54** How unstructured regions of polypeptide chain serving as tethers allow reaction intermediates to be passed from one active site to another in large multienzyme complexes. (A–C) The fatty acid synthase in mammals. (A) The location of seven protein domains with different activities in this 270 kilodalton protein. The numbers refer to the order in which each enzyme domain must function to complete each two-carbon addition step. After multiple cycles of two-carbon addition, the termination domain releases the final product once the desired length of fatty acid has been synthesized. (B) The structure of the dimeric enzyme, with the location of the five active sites in one monomer indicated. (C) How a flexible tether allows the substrate that remains linked to the acyl carrier domain (red) to be passed from one active site to another in each monomer, sequentially elongating and modifying the bound fatty acid intermediate (yellow). The five steps are repeated until the final length of fatty acid chain has been synthesized. (Only steps 1 through 4 are illustrated here.)

(D) Multiple tethered subunits in the giant pyruvate dehydrogenase complex (9500 kilodaltons, larger than a ribosome) that catalyzes the conversion of pyruvate to acetyl CoA. As in (C), a covalently bound substrate held on a flexible tether (red balls with yellow substrate) is serially passed through active sites on subunits (here labeled 1 through 3) to produce the final products. Here, subunit 1 catalyzes the decarboxylation of pyruvate accompanied by the reductive acetylation of a lipoyl group linked to one of the red balls. Subunit 2 transfers this acetyl group to CoA, forming acetyl CoA, and subunit 3 reoxidizes the lipoyl group to prepare it for the next cycle. Only one-tenth of the subunits labeled 1 and 3, attached to the core formed by subunit 2, are illustrated here. This important reaction takes place in the mammalian mitochondrion, as part of the pathway that oxidizes sugars to  $\text{CO}_2$  and  $\text{H}_2\text{O}$  (see page 82). (A–C, adapted from T. Maier et al., *Quart. Rev. Biophys.* 43:373–422, 2010; D, from J.L.S. Milne et al., *J. Biol. Chem.* 281:4364–4370, 2006.)

the cell, the concentration of reactants in that compartment may be increased by 10 times compared with a cell with the same number of enzyme and substrate molecules, but no compartmentalization. Reactions limited by the speed of diffusion can thereby be speeded up by a factor of 10.

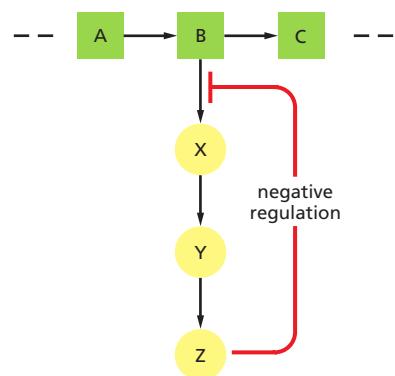
### The Cell Regulates the Catalytic Activities of Its Enzymes

A living cell contains thousands of enzymes, many of which operate at the same time and in the same small volume of the cytosol. By their catalytic action, these enzymes generate a complex web of metabolic pathways, each composed of chains of chemical reactions in which the product of one enzyme becomes the substrate of the next. In this maze of pathways, there are many branch points (nodes) where different enzymes compete for the same substrate. The system is

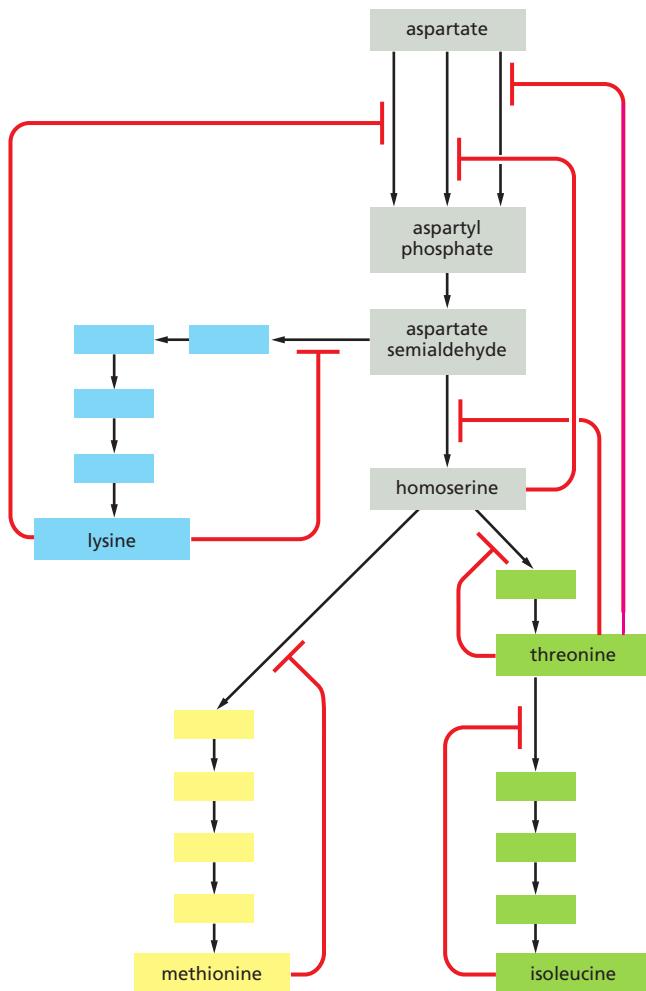
complex (see Figure 2–63), and elaborate controls are required to regulate when and how rapidly each reaction occurs.

Regulation occurs at many levels. At one level, the cell controls how many molecules of each enzyme it makes by regulating the expression of the gene that encodes that enzyme (discussed in Chapter 7). The cell also controls enzymatic activities by confining sets of enzymes to particular subcellular compartments, whether by enclosing them in a distinct membrane-bounded compartment (discussed in Chapters 12 and 14) or by concentrating them on a protein scaffold (see Figure 3–77). As will be explained later in this chapter, enzymes are also covalently modified to control their activity. The rate of protein destruction by targeted proteolysis represents yet another important regulatory mechanism (see Figure 6–86). But the most general process that adjusts reaction rates operates through a direct, reversible change in the activity of an enzyme in response to the specific small molecules that it binds.

The most common type of control occurs when an enzyme binds a molecule that is not a substrate to a special regulatory site outside the active site, thereby altering the rate at which the enzyme converts its substrates to products. For example, in **feedback inhibition**, a product produced late in a reaction pathway inhibits an enzyme that acts earlier in the pathway. Thus, whenever large quantities of the final product begin to accumulate, this product binds to the enzyme and slows down its catalytic action, thereby limiting the further entry of substrates into that reaction pathway (Figure 3–55). Where pathways branch or intersect, there are usually multiple points of control by different final products, each of which works to regulate its own synthesis (Figure 3–56). Feedback inhibition can work almost instantaneously, and it is rapidly reversed when the level of the product falls.



**Figure 3–55** Feedback inhibition of a single biosynthetic pathway. The end product Z inhibits the first enzyme that is unique to its synthesis and thereby controls its own level in the cell. This is an example of negative regulation.



**Figure 3–56** Multiple feedback inhibition. In this example, which shows the biosynthetic pathways for four different amino acids in bacteria, the red lines indicate positions at which products feed back to inhibit enzymes. Each amino acid controls the first enzyme specific to its own synthesis, thereby controlling its own levels and avoiding a wasteful, or even dangerous, buildup of intermediates. The products can also separately inhibit the initial set of reactions common to all the syntheses; in this case, three different enzymes catalyze the initial reaction, each inhibited by a different product.

Feedback inhibition is *negative regulation*: it prevents an enzyme from acting. Enzymes can also be subject to *positive regulation*, in which a regulatory molecule stimulates the enzyme's activity rather than shutting the enzyme down. Positive regulation occurs when a product in one branch of the metabolic network stimulates the activity of an enzyme in another pathway. As one example, the accumulation of ADP activates several enzymes involved in the oxidation of sugar molecules, thereby stimulating the cell to convert more ADP to ATP.

### Allosteric Enzymes Have Two or More Binding Sites That Interact

A striking feature of both positive and negative feedback regulation is that the regulatory molecule often has a shape totally different from the shape of the substrate of the enzyme. This is why the effect on a protein is termed **allostery** (from the Greek words *allos*, meaning “other,” and *stereos*, meaning “solid” or “three-dimensional”). As biologists learned more about feedback regulation, they recognized that the enzymes involved must have at least two different binding sites on their surface—an **active site** that recognizes the substrates, and a **regulatory site** that recognizes a regulatory molecule. These two sites must somehow communicate so that the catalytic events at the active site can be influenced by the binding of the regulatory molecule at its separate site on the protein’s surface.

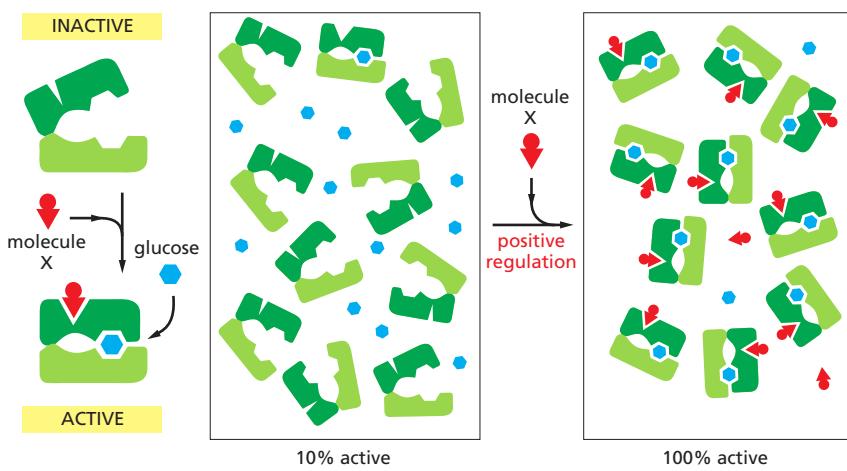
The interaction between separated sites on a protein molecule is now known to depend on a *conformational change* in the protein: binding at one of the sites causes a shift from one folded shape to a slightly different folded shape. During feedback inhibition, for example, the binding of an inhibitor at one site on the protein causes the protein to shift to a conformation that incapacitates its active site located elsewhere in the protein.

It is thought that most protein molecules are allosteric. They can adopt two or more slightly different conformations, and a shift from one to another caused by the binding of a ligand can alter their activity. This is true not only for enzymes but also for many other proteins, including receptors, structural proteins, and motor proteins. In all instances of allosteric regulation, each conformation of the protein has somewhat different surface contours, and the protein’s binding sites for ligands are altered when the protein changes shape. Moreover, as we discuss next, each ligand will stabilize the conformation that it binds to most strongly, and thus—at high enough concentrations—will tend to “switch” the protein toward the conformation that the ligand prefers.

### Two Ligands Whose Binding Sites Are Coupled Must Reciprocally Affect Each Other’s Binding

The effects of ligand binding on a protein follow from a fundamental chemical principle known as **linkage**. Suppose, for example, that a protein that binds glucose also binds another molecule, X, at a distant site on the protein’s surface. If the binding site for X changes shape as part of the conformational change in the protein induced by glucose binding, the binding sites for X and for glucose are said to be *coupled*. Whenever two ligands prefer to bind to the *same* conformation of an allosteric protein, it follows from basic thermodynamic principles that each ligand must increase the affinity of the protein for the other. For example, if the shift of a protein to a conformation that binds glucose best also causes the binding site for X to fit X better, then the protein will bind glucose more tightly when X is present than when X is absent. In other words, X will positively regulate the protein’s binding of glucose ([Figure 3–57](#)).

Conversely, linkage operates in a negative way if two ligands prefer to bind to *different* conformations of the same protein. In this case, the binding of the first ligand discourages the binding of the second ligand. Thus, if a shape change caused by glucose binding decreases the affinity of a protein for molecule X, the binding of X must also decrease the protein’s affinity for glucose ([Figure 3–58](#)). The linkage relationship is quantitatively reciprocal, so that, for example, if glucose has a very large effect on the binding of X, X has a very large effect on the binding of glucose.

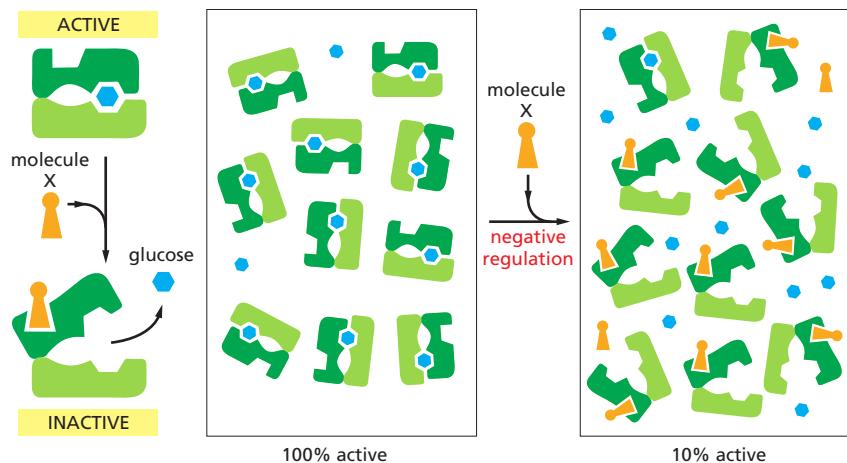


**Figure 3–57** Positive regulation caused by conformational coupling between two separate binding sites. In this example, both glucose and molecule X bind best to the *closed* conformation of a protein with two domains. Because both glucose and molecule X drive the protein toward its closed conformation, each ligand helps the other to bind. Glucose and molecule X are therefore said to bind cooperatively to the protein.

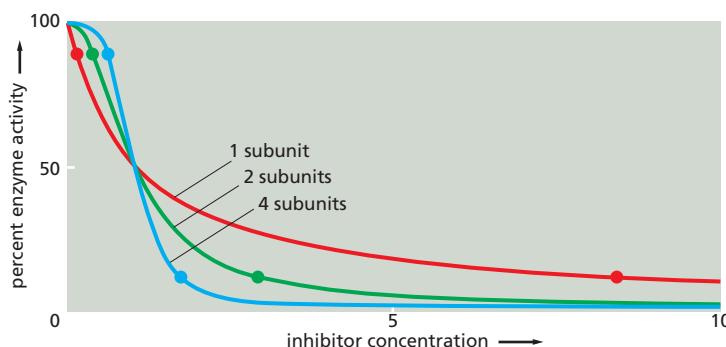
The relationships shown in Figures 3–57 and 3–58 apply to all proteins, and they underlie all of cell biology. The principle seems so obvious in retrospect that we now take it for granted. But the discovery of linkage in studies of a few enzymes in the 1950s, followed by an extensive analysis of allosteric mechanisms in proteins in the early 1960s, had a revolutionary effect on our understanding of biology. Since molecule X in these examples binds at a site on the enzyme that is distinct from the site where catalysis occurs, it need not have any chemical relationship to the substrate that binds at the active site. Moreover, as we have just seen, for enzymes that are regulated in this way, molecule X can either turn the enzyme on (positive regulation) or turn it off (negative regulation). By such a mechanism, **allosteric proteins** serve as general switches that, in principle, can allow one molecule in a cell to affect the fate of any other.

### Symmetric Protein Assemblies Produce Cooperative Allosteric Transitions

A single-subunit enzyme that is regulated by negative feedback can at most decrease from 90% to about 10% activity in response to a 100-fold increase in the concentration of an inhibitory ligand that it binds (Figure 3–59, red line). Responses of this type are apparently not sharp enough for optimal cell regulation, and most enzymes that are turned on or off by ligand binding consist of symmetric assemblies of identical subunits. With this arrangement, the binding of a molecule of ligand to a single site on one subunit can promote an allosteric change in the entire assembly that helps the neighboring subunits bind the same ligand. As a result, a *cooperative allosteric transition* occurs (Figure 3–59, blue line), allowing



**Figure 3–58** Negative regulation caused by conformational coupling between two separate binding sites. The scheme here resembles that in the previous figure, but here molecule X prefers the *open* conformation, while glucose prefers the *closed* conformation. Because glucose and molecule X drive the protein toward opposite conformations (closed and open, respectively), the presence of either ligand interferes with the binding of the other.



a relatively small change in ligand concentration in the cell to switch the whole assembly from an almost fully active to an almost fully inactive conformation (or vice versa).

The principles involved in a cooperative “all-or-none” transition are the same for all proteins, whether or not they are enzymes. Thus, for example, they are critical for the efficient uptake and release of O<sub>2</sub> by hemoglobin in our blood. But they are perhaps easiest to visualize for an enzyme that forms a symmetric dimer. In the example shown in Figure 3–60, the first molecule of an inhibitory ligand binds with great difficulty since its binding disrupts an energetically favorable interaction between the two identical monomers in the dimer. A second molecule of inhibitory ligand now binds more easily, however, because its binding restores the energetically favorable monomer-monomer contacts of a symmetric dimer (this also completely inactivates the enzyme).

As an alternative to this *induced fit* model for a cooperative allosteric transition, we can view such a symmetric enzyme as having only two possible conformations, corresponding to the “enzyme on” and “enzyme off” structures in Figure 3–60. In this view, ligand binding perturbs an all-or-none equilibrium between these two states, thereby changing the proportion of active molecules. Both models represent true and useful concepts.

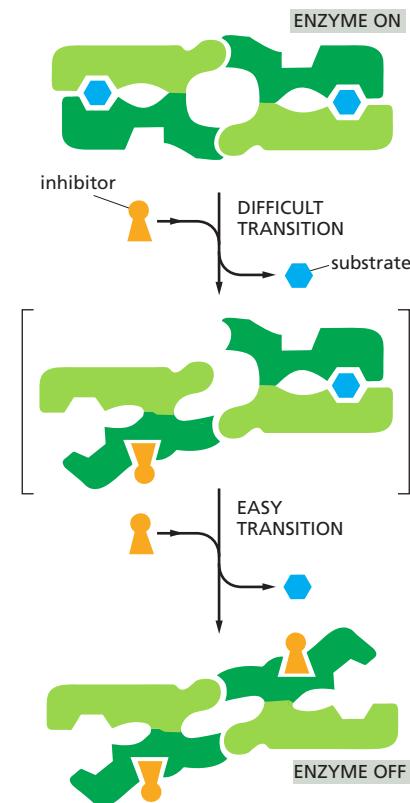
### Many Changes in Proteins Are Driven by Protein Phosphorylation

Proteins are regulated by more than the reversible binding of other molecules. A second method that eukaryotic cells use extensively to regulate a protein’s function is the covalent addition of a smaller molecule to one or more of its amino acid side chains. The most common such regulatory modification in higher eukaryotes is the addition of a phosphate group. We shall therefore use protein phosphorylation to illustrate some of the general principles involved in the control of protein function through the modification of amino acid side chains.

A phosphorylation event can affect the protein that is modified in three important ways. First, because each phosphate group carries two negative charges, the enzyme-catalyzed addition of a phosphate group to a protein can cause a major conformational change in the protein by, for example, attracting a cluster of positively charged amino acid side chains. This can, in turn, affect the binding of ligands elsewhere on the protein surface, dramatically changing the

**Figure 3–59** Enzyme activity versus the concentration of inhibitory ligand for single-subunit and multisubunit allosteric enzymes.

For an enzyme with a single subunit (red line), a drop from 90% enzyme activity to 10% activity (indicated by the two dots on the curve) requires a 100-fold increase in the concentration of inhibitor. The enzyme activity is calculated from the simple equilibrium relationship  $K = [IP]/[I][IP]$ , where P is active protein, I is inhibitor, and IP is the inactive protein bound to inhibitor. An identical curve applies to any simple binding interaction between two molecules, A and B. In contrast, a multisubunit allosteric enzyme can respond in a switchlike manner to a change in ligand concentration: the steep response is caused by a cooperative binding of the ligand molecules, as explained in Figure 3–60. Here, the green line represents the idealized result expected for the cooperative binding of two inhibitory ligand molecules to an allosteric enzyme with two subunits, and the blue line shows the idealized response of an enzyme with four subunits. As indicated by the two dots on each of these curves, the more complex enzymes drop from 90% to 10% activity over a much narrower range of inhibitor concentration than does the enzyme composed of a single subunit.



**Figure 3–60** A cooperative allosteric transition in an enzyme composed of two identical subunits. This diagram illustrates how the conformation of one subunit can influence that of its neighbor. The binding of a single molecule of an inhibitory ligand (yellow) to one subunit of the enzyme occurs with difficulty because it changes the conformation of this subunit and thereby disrupts the symmetry of the enzyme. Once this conformational change has occurred, however, the energy gained by restoring the symmetric pairing interaction between the two subunits makes it especially easy for the second subunit to bind the inhibitory ligand and undergo the same conformational change. Because the binding of the first molecule of ligand increases the affinity with which the other subunit binds the same ligand, the response of the enzyme to changes in the concentration of the ligand is much steeper than the response of an enzyme with only one subunit (see Figure 3–59 and Movie 3.10).

protein's activity. When a second enzyme removes the phosphate group, the protein returns to its original conformation and restores its initial activity.

Second, an attached phosphate group can form part of a structure that the binding sites of other proteins recognize. As previously discussed, the SH2 domain binds to a short peptide sequence containing a phosphorylated tyrosine side chain (see Figure 3–40B). More than ten other common domains provide binding sites for attaching their protein to phosphorylated peptides in other protein molecules, each recognizing a phosphorylated amino acid side chain in a different protein context. Third, the addition of a phosphate group can mask a binding site that otherwise holds two proteins together, and thereby disrupt protein–protein interactions. As a result, protein phosphorylation and dephosphorylation very often drive the regulated assembly and disassembly of protein complexes (see, for example, Figure 15–11).

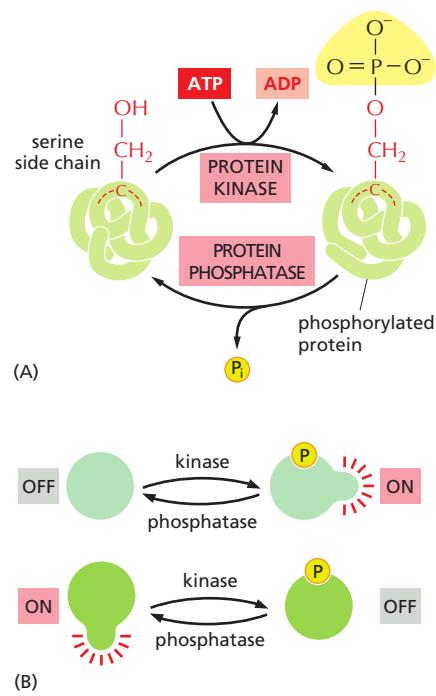
Reversible protein phosphorylation controls the activity, structure, and cellular localization of both enzymes and many other types of proteins in eukaryotic cells. In fact, this regulation is so extensive that more than one-third of the 10,000 or so proteins in a typical mammalian cell are thought to be phosphorylated at any given time—many with more than one phosphate. As might be expected, the addition and removal of phosphate groups from specific proteins often occur in response to signals that specify some change in a cell's state. For example, the complicated series of events that takes place as a eukaryotic cell divides is largely timed in this way (discussed in Chapter 17), and many of the signals mediating cell–cell interactions are relayed from the plasma membrane to the nucleus by a cascade of protein phosphorylation events (discussed in Chapter 15).

### A Eukaryotic Cell Contains a Large Collection of Protein Kinases and Protein Phosphatases

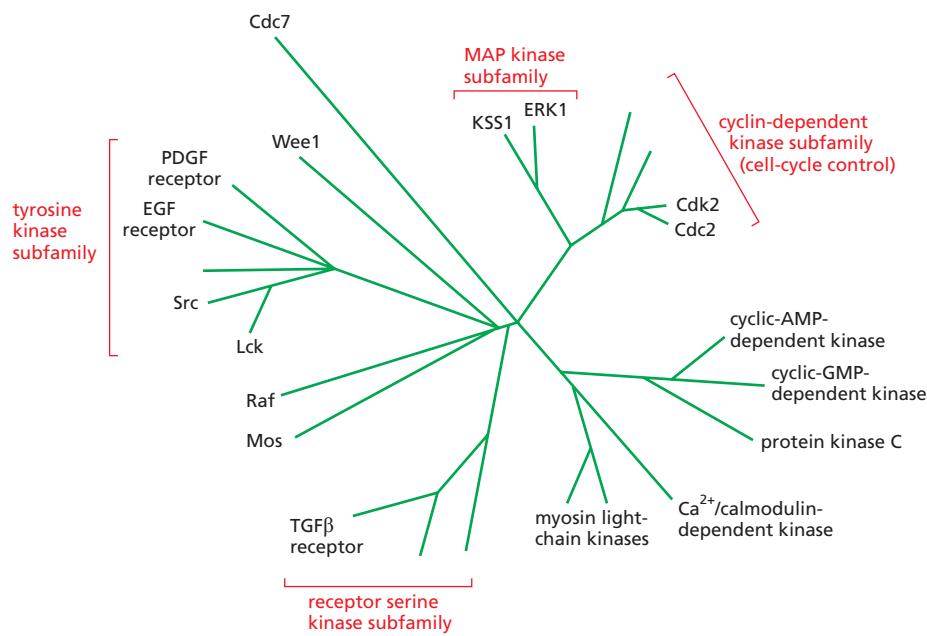
Protein phosphorylation involves the enzyme-catalyzed transfer of the terminal phosphate group of an ATP molecule to the hydroxyl group on a serine, threonine, or tyrosine side chain of the protein (Figure 3–61). A **protein kinase** catalyzes this reaction, and the reaction is essentially unidirectional because of the large amount of free energy released when the phosphate–phosphate bond in ATP is broken to produce ADP (discussed in Chapter 2). A **protein phosphatase** catalyzes the reverse reaction of phosphate removal, or *dephosphorylation*. Cells contain hundreds of different protein kinases, each responsible for phosphorylating a different protein or set of proteins. There are also many different protein phosphatases; some are highly specific and remove phosphate groups from only one or a few proteins, whereas others act on a broad range of proteins and are targeted to specific substrates by regulatory subunits. The state of phosphorylation of a protein at any moment, and thus its activity, depends on the relative activities of the protein kinases and phosphatases that modify it.

The protein kinases that phosphorylate proteins in eukaryotic cells belong to a very large family of enzymes that share a catalytic (kinase) sequence of about 290 amino acids. The various family members contain different amino acid sequences on either end of the kinase sequence (for example, see Figure 3–10), and often have short amino acid sequences inserted into loops within it. Some of these additional amino acid sequences enable each kinase to recognize the specific set of proteins it phosphorylates, or to bind to structures that localize it in specific regions of the cell. Other parts of the protein regulate the activity of each kinase, so it can be turned on and off in response to different specific signals, as described below.

By comparing the number of amino acid sequence differences between the various members of a protein family, we can construct an “evolutionary tree” that is thought to reflect the pattern of gene duplication and divergence that gave rise to the family. Figure 3–62 shows an evolutionary tree of protein kinases. Kinases with related functions are often located on nearby branches of the tree: the protein kinases involved in cell signaling that phosphorylate tyrosine side chains, for example, are all clustered in the top left corner of the tree. The other kinases shown



**Figure 3–61 Protein phosphorylation.** Many thousands of proteins in a typical eukaryotic cell are modified by the covalent addition of a phosphate group. (A) The general reaction transfers a phosphate group from ATP to an amino acid side chain of the target protein by a protein kinase. Removal of the phosphate group is catalyzed by a second enzyme, a protein phosphatase. In this example, the phosphate is added to a serine side chain; in other cases, the phosphate is instead linked to the –OH group of a threonine or a tyrosine in the protein. (B) The phosphorylation of a protein by a protein kinase can either increase or decrease the protein's activity, depending on the site of phosphorylation and the structure of the protein.



**Figure 3–62** An evolutionary tree of selected protein kinases. A higher eukaryotic cell contains hundreds of such enzymes, and the human genome codes for more than 500. Note that only some of these, those discussed in this book, are shown.

phosphorylate either a serine or a threonine side chain, and many are organized into clusters that seem to reflect their function—in transmembrane signal transduction, intracellular signal amplification, cell-cycle control, and so on.

As a result of the combined activities of protein kinases and protein phosphatases, the phosphate groups on proteins are continually turning over—being added and then rapidly removed. Such phosphorylation cycles may seem wasteful, but they are important in allowing the phosphorylated proteins to switch rapidly from one state to another: the more rapid the cycle, the faster a population of protein molecules can change its state of phosphorylation in response to a sudden change in its phosphorylation rate (see Figure 15–14). The energy required to drive this phosphorylation cycle is derived from the free energy of ATP hydrolysis, one molecule of which is consumed for each phosphorylation event.

### The Regulation of the Src Protein Kinase Reveals How a Protein Can Function as a Microprocessor

The hundreds of different protein kinases in a eukaryotic cell are organized into complex networks of signaling pathways that help to coordinate the cell's activities, drive the cell cycle, and relay signals into the cell from the cell's environment. Many of the extracellular signals involved need to be both integrated and amplified by the cell. Individual protein kinases (and other signaling proteins) serve as input-output devices, or “microprocessors,” in the integration process. An important part of the input to these signal-processing proteins comes from the control that is exerted by phosphates added and removed from them by protein kinases and protein phosphatases, respectively.

The Src family of protein kinases (see Figure 3–10) exhibits such behavior. The *Src protein* (pronounced “sarc” and named for the type of tumor, a sarcoma, that its deregulation can cause) was the first tyrosine kinase to be discovered. It is now known to be part of a subfamily of nine very similar protein kinases, which are found only in multicellular animals. As indicated by the evolutionary tree in Figure 3–62, sequence comparisons suggest that tyrosine kinases as a group were a relatively late innovation that branched off from the serine/threonine kinases, with the Src subfamily being only one subgroup of the tyrosine kinases created in this way.

The Src protein and its relatives contain a short N-terminal region that becomes covalently linked to a strongly hydrophobic fatty acid, which anchors the kinase at the cytoplasmic face of the plasma membrane. Next along the linear sequence of



**Figure 3–63** The domain structure of the Src family of protein kinases, mapped along the amino acid sequence. For the three-dimensional structure of Src, see Figure 3–13.

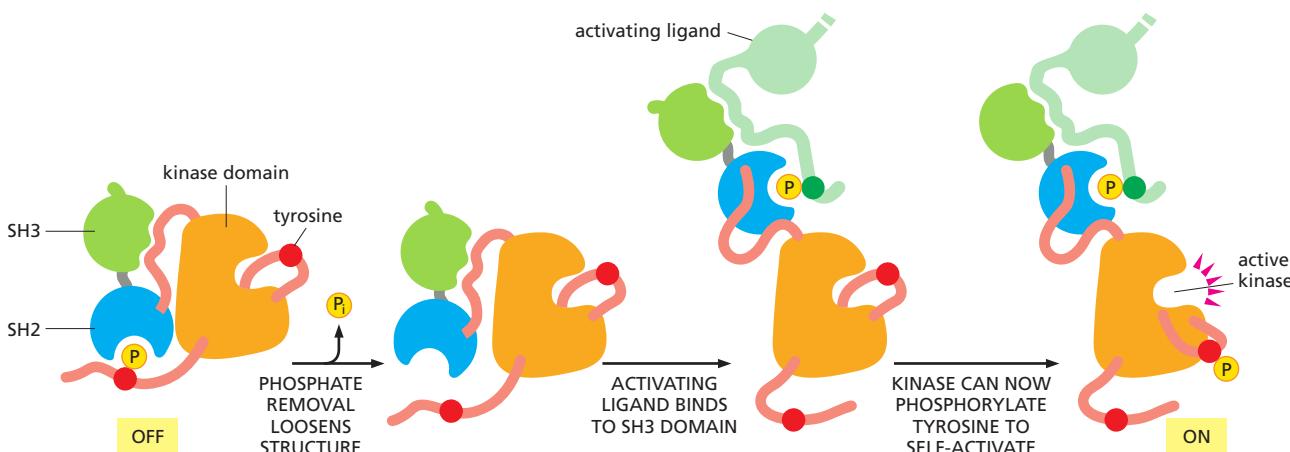
amino acids come two peptide-binding domains, a Src homology 3 (SH3) domain and an SH2 domain, followed by the kinase catalytic domain (Figure 3–63). These kinases normally exist in an inactive conformation, in which a phosphorylated tyrosine near the C-terminus is bound to the SH2 domain, and the SH3 domain is bound to an internal peptide in a way that distorts the active site of the enzyme and helps to render it inactive.

As shown in Figure 3–64, turning the kinase on involves at least two specific inputs: removal of the C-terminal phosphate and the binding of the SH3 domain by a specific activating protein. In this way, the activation of the Src kinase signals the completion of a particular set of separate upstream events (Figure 3–65). Thus, the Src family of protein kinases serves as specific signal integrators, contributing to the web of information-processing events that enable the cell to compute useful responses to a complex set of different conditions.

### Proteins That Bind and Hydrolyze GTP Are Ubiquitous Cell Regulators

We have described how the addition or removal of phosphate groups on a protein can be used by a cell to control the protein's activity. In the example just discussed, a kinase transfers a phosphate from an ATP molecule to an amino acid side chain of a target protein. Eukaryotic cells also have another way to control protein activity by phosphate addition and removal. In this case, the phosphate is not attached directly to the protein; instead, it is a part of the guanine nucleotide GTP, which binds very tightly to a class of proteins known as *GTP-binding proteins*. In general, proteins regulated in this way are in their active conformations with GTP bound. The loss of a phosphate group occurs when the bound GTP is hydrolyzed to GDP in a reaction catalyzed by the protein itself, and in its GDP-bound state the protein is inactive. In this way, GTP-binding proteins act as on-off switches whose activity is determined by the presence or absence of an additional phosphate on a bound GDP molecule (Figure 3–66).

**GTP-binding proteins** (also called **GTPases** because of the GTP hydrolysis they catalyze) comprise a large family of proteins that all contain variations on the same GTP-binding globular domain. When a tightly bound GTP is hydrolyzed by the GTP-binding protein to GDP, this domain undergoes a conformational



**Figure 3–64** The activation of a Src-type protein kinase by two sequential events. As described in the text, the requirement for multiple upstream events to trigger these processes allows the kinase to serve as a signal integrator (Movie 3.11). (Adapted from S.C. Harrison et al., *Cell* 112:737–740, 2003. With permission from Elsevier.)

**Figure 3–65** How a Src-type protein kinase acts as a signal-integrating device. A disruption of the inhibitory interaction illustrated for the SH3 domain (green) occurs when its binding to the indicated orange linker region is replaced with its higher-affinity binding to an activating ligand.

change that inactivates the protein. The three-dimensional structure of a prototypical member of this family, the monomeric GTPase called Ras, is shown in **Figure 3–67**.

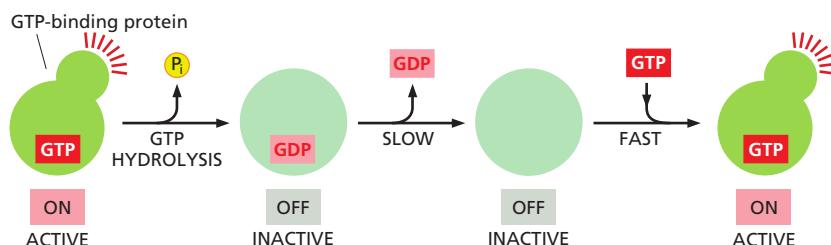
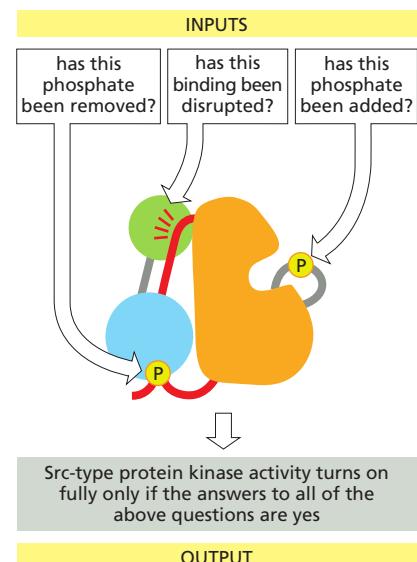
The *Ras* protein has an important role in cell signaling (discussed in Chapter 15). In its GTP-bound form, it is active and stimulates a cascade of protein phosphorylations in the cell. Most of the time, however, the protein is in its inactive, GDP-bound form. It becomes active when it exchanges its GDP for a GTP molecule in response to extracellular signals, such as growth factors, that bind to receptors in the plasma membrane (see Figure 15–47).

### Regulatory Proteins GAP and GEF Control the Activity of GTP-Binding Proteins by Determining Whether GTP or GDP Is Bound

GTP-binding proteins are controlled by regulatory proteins that determine whether GTP or GDP is bound, just as phosphorylated proteins are turned on and off by protein kinases and protein phosphatases. Thus, Ras is inactivated by a *GTPase-activating protein* (*GAP*), which binds to the Ras protein and induces Ras to hydrolyze its bound GTP molecule to GDP—which remains tightly bound—and inorganic phosphate ( $P_i$ ), which is rapidly released. The Ras protein stays in its inactive, GDP-bound conformation until it encounters a *guanine nucleotide exchange factor* (*GEF*), which binds to GDP-Ras and causes Ras to release its GDP. Because the empty nucleotide-binding site is immediately filled by a GTP molecule (GTP is present in large excess over GDP in cells), the GEF activates Ras by *indirectly* adding back the phosphate removed by GTP hydrolysis. Thus, in a sense, the roles of GAP and GEF are analogous to those of a protein phosphatase and a protein kinase, respectively (**Figure 3–68**).

### Proteins Can Be Regulated by the Covalent Addition of Other Proteins

Cells contain a special family of small proteins whose members are covalently attached to many other proteins to determine the activity or fate of the second protein. In each case, the carboxyl end of the small protein becomes linked to the amino group of a lysine side chain of a “target” protein through an isopeptide bond. The first such protein discovered, and the most abundantly used, is **ubiquitin** (**Figure 3–69A**). Ubiquitin can be covalently attached to target proteins in a variety of ways, each of which has a different meaning for cells. The major form of ubiquitin addition produces *polyubiquitin* chains in which—once the first ubiquitin molecule is attached to the target—each subsequent ubiquitin molecule links to Lys48 of the previous ubiquitin, creating a chain of Lys48-linked ubiquitins that are attached to a single lysine side chain of the target protein. This form of polyubiquitin directs the target protein to the interior of a proteasome, where it is digested to small peptides (see Figure 6–84). In other circumstances, only single molecules of ubiquitin are added to proteins. In addition, some target proteins are



**Figure 3–66** GTP-binding proteins as molecular switches. The activity of a GTP-binding protein (also called a GTPase) generally requires the presence of a tightly bound GTP molecule (switch “on”). Hydrolysis of this GTP molecule by the GTP-binding protein produces GDP and inorganic phosphate ( $P_i$ ), and it causes the protein to convert to a different, usually inactive, conformation (switch “off”). Resetting the switch requires that the tightly bound GDP dissociates. This is a slow step that is greatly accelerated by specific signals; once the GDP has dissociated, a molecule of GTP is quickly rebound.

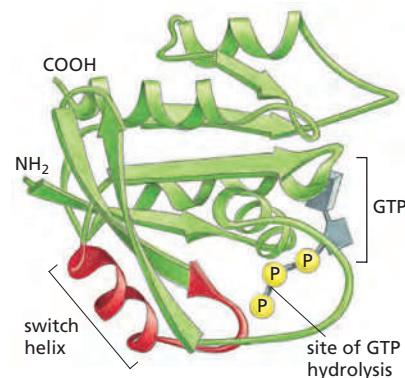
modified with a different type of polyubiquitin chain. These modifications have different functional consequences for the protein that is targeted (Figure 3–69B).

Related structures are created when a different member of the ubiquitin family, such as SUMO (small ubiquitin-related modifier), is covalently attached to a lysine side chain of target proteins. Not surprisingly, all such modifications are reversible. Cells contain sets of ubiquitylating and deubiquitylating (and sumoylating and desumoylating) enzymes that manipulate these covalent adducts, thereby playing roles analogous to the protein kinases and phosphatases that add and remove phosphates from protein side chains.

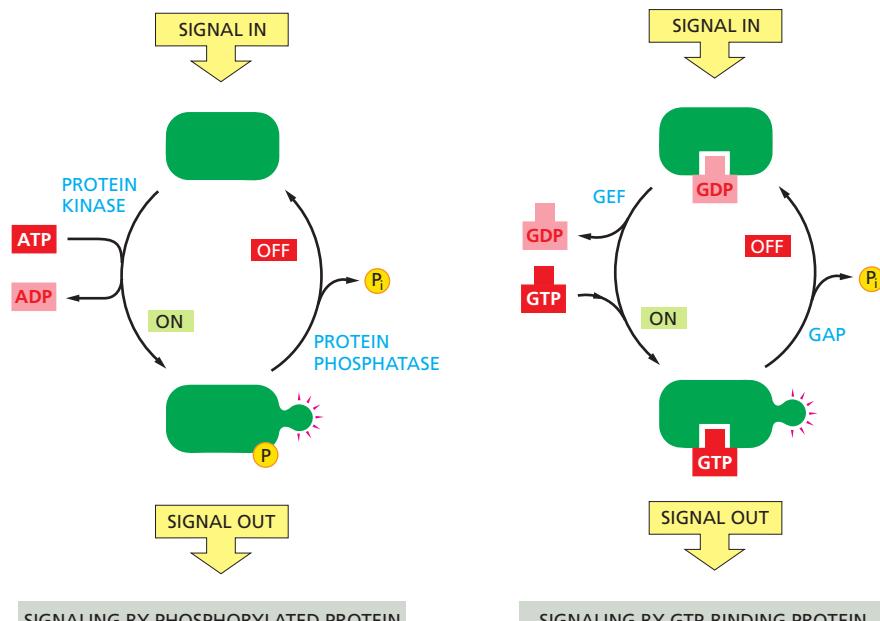
### An Elaborate Ubiquitin-Conjugating System Is Used to Mark Proteins

How do cells select target proteins for ubiquitin addition? As an initial step, the carboxyl end of ubiquitin needs to be activated. This activation is accomplished when a protein called a *ubiquitin-activating enzyme* (E1) uses ATP hydrolysis energy to attach ubiquitin to itself through a high-energy covalent bond (a thioester). E1 then passes this activated ubiquitin to one of a set of *ubiquitin-conjugating* (E2) enzymes, each of which acts in conjunction with a set of accessory (E3) proteins called *ubiquitin ligases*. There are roughly 30 structurally similar but distinct E2 enzymes in mammals, and hundreds of different E3 proteins that form complexes with specific E2 enzymes.

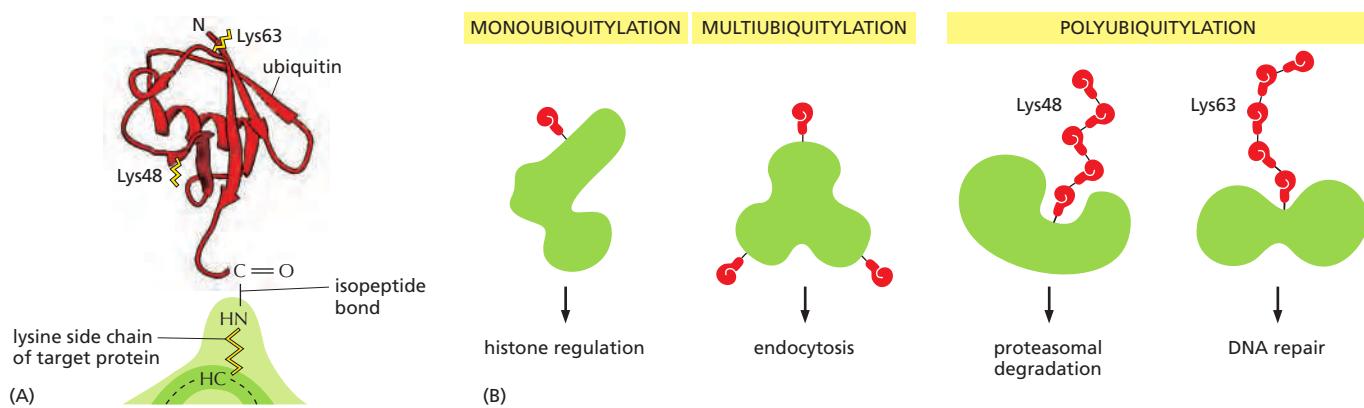
**Figure 3–70** illustrates the process used to mark proteins for proteasomal degradation. [Similar mechanisms are used to attach ubiquitin (and SUMO) to other types of target proteins.] Here, the ubiquitin ligase binds to specific degradation signals, called *degrons*, in protein substrates, thereby helping E2 to form a *polyubiquitin* chain linked to a lysine of the substrate protein. This polyubiquitin chain on a target protein will then be recognized by a specific receptor in the proteasome, causing the target protein to be destroyed. Distinct ubiquitin ligases recognize different degradation signals, thereby targeting distinct subsets of intracellular proteins for destruction, often in response to specific signals (see Figure 6–86).



**Figure 3–67** The structure of the Ras protein in its GTP-bound form. This monomeric GTPase illustrates the structure of a GTP-binding domain, which is present in a large family of GTP-binding proteins. The red regions change their conformation when the GTP molecule is hydrolyzed to GDP and inorganic phosphate by the protein; the GDP remains bound to the protein, while the inorganic phosphate is released. The special role of the “switch helix” in proteins related to Ras is explained in the text (see Figure 3–72 and Movie 15.7).



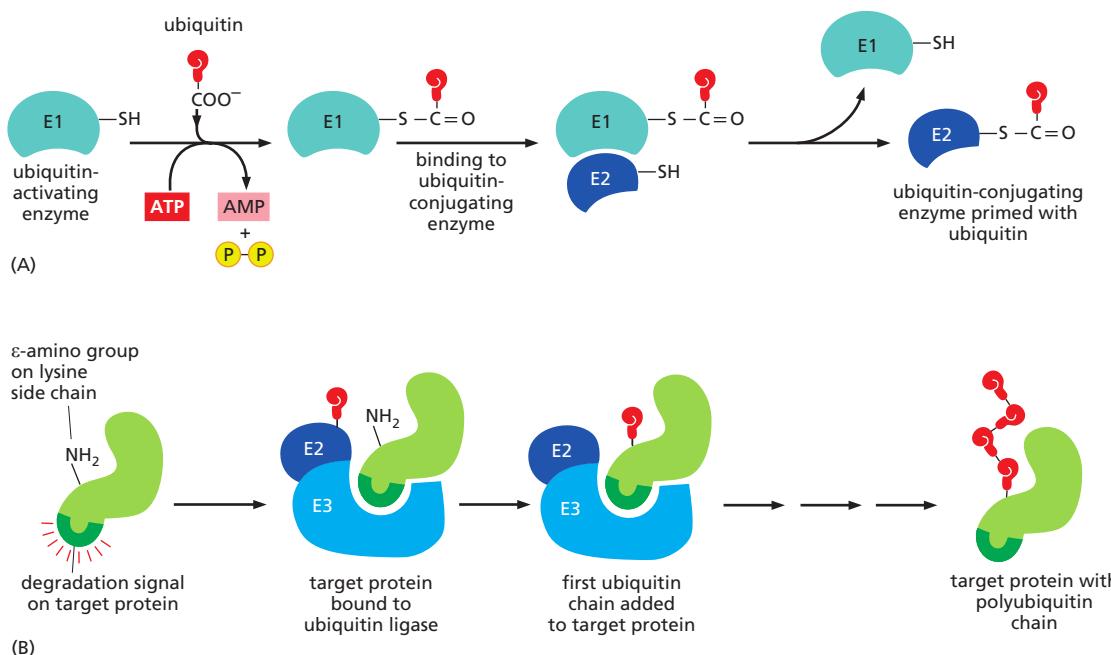
**Figure 3–68** A comparison of two major intracellular signaling mechanisms in eukaryotic cells. In both cases, a signaling protein is activated by the addition of a phosphate group and inactivated by the removal of this phosphate. Note that the addition of a phosphate to a protein can also be inhibitory. (Adapted from E.R. Kantrowitz and W.N. Lipscomb, *Trends Biochem. Sci.* 15:53–59, 1990.)



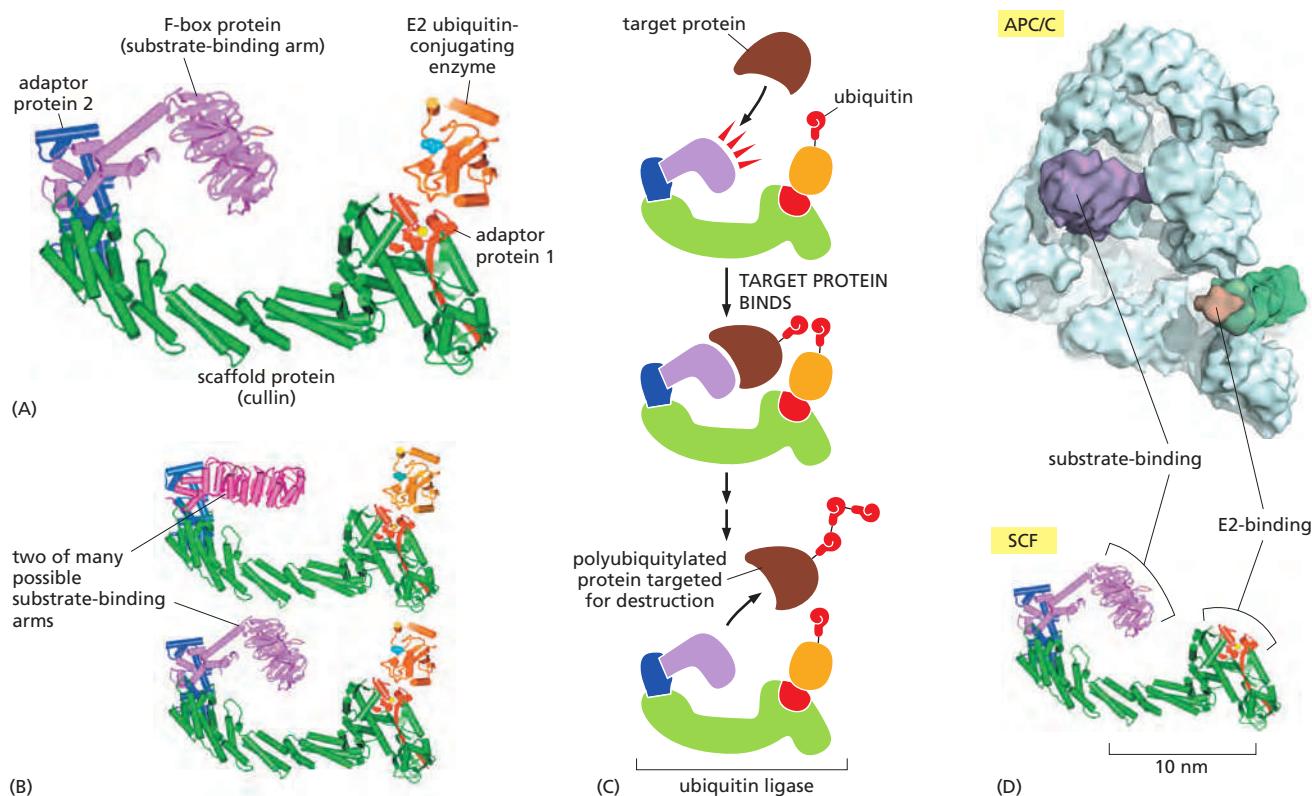
### Protein Complexes with Interchangeable Parts Make Efficient Use of Genetic Information

The *SCF ubiquitin ligase* is a protein complex that binds different “target proteins” at different times in the cell cycle, covalently adding polyubiquitin polypeptide chains to these targets. Its C-shaped structure is formed from five protein subunits, the largest of which serves as a scaffold on which the rest of the complex is built. The structure underlies a remarkable mechanism (Figure 3–71). At one end of the C is an E2 ubiquitin-conjugating enzyme. At the other end is a substrate-binding arm, a subunit known as an *F-box protein*. These two subunits are separated by a gap of about 5 nm. When this protein complex is activated, the F-box protein binds to a specific site on a target protein, positioning the protein in the gap so that some of its lysine side chains contact the ubiquitin-conjugating enzyme. The enzyme can then catalyze repeated additions of ubiquitin polypeptide to these lysines (see Figure 3–71C), producing polyubiquitin chains that mark the target proteins for rapid destruction in a proteasome.

**Figure 3–69** The marking of proteins by ubiquitin. (A) The three-dimensional structure of ubiquitin, a small protein of 76 amino acids. A family of special enzymes couples its carboxyl end to the amino group of a lysine side chain in a target protein molecule, forming an isopeptide bond. (B) Some modification patterns that have specific meanings to the cell. Note that the two types of polyubiquitylation differ in the way the ubiquitin molecules are linked together. Linkage through Lys48 signifies degradation by the proteasome (see Figure 6–84), whereas that through Lys63 has other meanings. Ubiquitin markings are “read” by proteins that specifically recognize each type of modification.



**Figure 3–70** The marking of proteins with ubiquitin. (A) The C-terminus of ubiquitin is initially activated by being linked via a high-energy thioester bond to a cysteine side chain on the E1 protein. This reaction requires ATP, and it proceeds via a covalent AMP-ubiquitin intermediate. The activated ubiquitin on E1, also known as the ubiquitin-activating enzyme, is then transferred to the cysteine on an E2 molecule. (B) The addition of a polyubiquitin chain to a target protein. In a mammalian cell, there are several hundred distinct E2-E3 complexes. The E2s are called ubiquitin-conjugating enzymes. The E3s are referred to as ubiquitin ligases. (Adapted from D.R. Knighton et al., *Science* 253:407–414, 1991.)



In this manner, specific proteins are targeted for rapid destruction in response to specific signals, thereby helping to drive the cell cycle (discussed in Chapter 17). The timing of the destruction often involves creating a specific pattern of phosphorylation on the target protein that is required for its recognition by the F-box subunit. It also requires the activation of an SCF ubiquitin ligase that carries the appropriate substrate-binding arm. Many of these arms (the F-box subunits) are interchangeable in the protein complex (see Figure 3-71B), and there are more than 70 human genes that encode them.

As emphasized previously, once a successful protein has evolved, its genetic information tends to be duplicated to produce a family of related proteins. Thus, for example, not only are there many F-box proteins—making possible the recognition of different sets of target proteins—but there is also a family of scaffolds (known as cullins) that give rise to a family of SCF-like ubiquitin ligases.

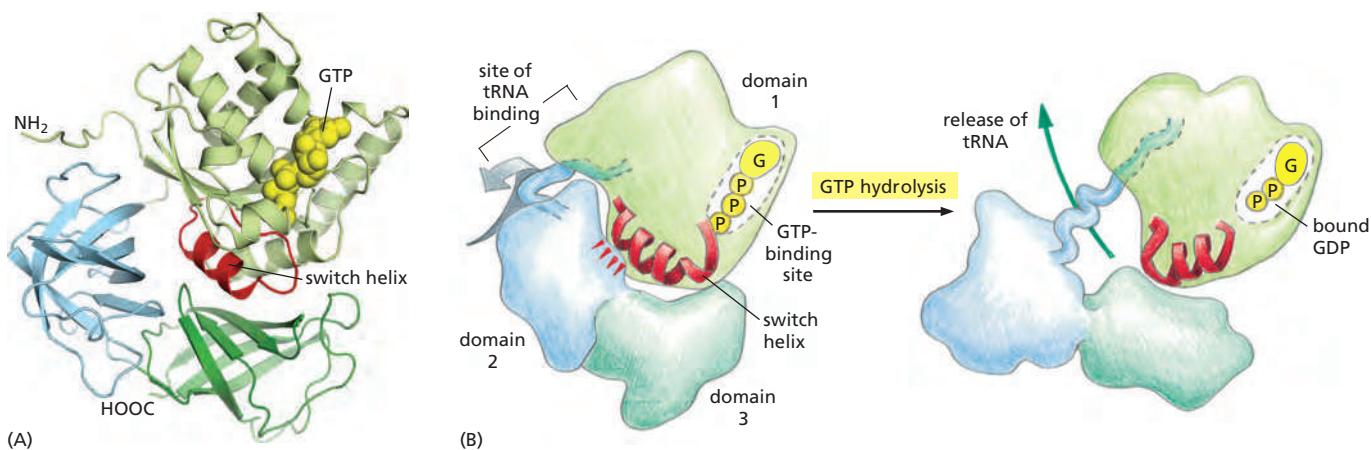
A protein machine like the SCF ubiquitin ligase, with its interchangeable parts, makes economical use of the genetic information in cells. It also creates opportunities for “rapid” evolution, inasmuch as new functions can evolve for the entire complex simply by producing an alternative version of one of its subunits.

Ubiquitin ligases form a diverse family of protein complexes. Some of these complexes are far larger and more complicated than SCF, but their underlying enzymatic function remains the same (Figure 3-71D).

### A GTP-Binding Protein Shows How Large Protein Movements Can Be Generated

Detailed structures obtained for one of the GTP-binding protein family members, the *EF-Tu* protein, provide a good example of how allosteric changes in protein conformations can produce large movements by amplifying a small, local conformational change. As will be discussed in Chapter 6, EF-Tu is an abundant molecule that serves as an elongation factor (hence the EF) in protein synthesis, loading each aminoacyl-tRNA molecule onto the ribosome. EF-Tu contains a Ras-like domain (see Figure 3-67), and the tRNA molecule forms a tight complex with its GTP-bound form. This tRNA molecule can transfer its amino acid to the growing

**Figure 3-71** The structure and mode of action of an SCF ubiquitin ligase. (A) The structure of the five-protein ubiquitin ligase complex that includes an E2 ubiquitin-conjugating enzyme. Four proteins form the E3 portion. The protein denoted here as adaptor protein 1 is the Rbx1/Hrt1 protein, adaptor protein 2 is the Skp1 protein, and the cullin is the Cul1 protein. One of the many different F-box proteins completes the complex. (B) Comparison of the same complex with two different substrate-binding arms, the F-box proteins Skp2 (top) and β-trCP1 (bottom), respectively. (C) The binding and ubiquitylation of a target protein by the SCF ubiquitin ligase. If, as indicated, a chain of ubiquitin molecules is added to the same lysine of the target protein, that protein is marked for rapid destruction by the proteasome. (D) Comparison of SCF (bottom) with a low-resolution electron microscopy structure of a ubiquitin ligase called the anaphase-promoting complex (APC/C; top) at the same scale. The APC/C is a large, 15-protein complex. As discussed in Chapter 17, its ubiquitylations control the late stages of mitosis. It is distantly related to SCF and contains a cullin subunit (green) that lies along the side of the complex at right, only partly visible in this view. E2 proteins are not shown here, but their binding sites are indicated in orange, along with substrate-binding sites in purple. (A and B, adapted from G. Wu et al., Mol. Cell 11:1445–1456, 2003. With permission from Elsevier; D, adapted from P. da Fonseca et al., Nature 470:274–278, 2011. With permission from Macmillan Publishers Ltd.)



**Figure 3–72** The large conformational change in EF-Tu caused by GTP hydrolysis. (A and B) The three-dimensional structure of EF-Tu with GTP bound. The domain at the top has a structure similar to the Ras protein, and its red  $\alpha$  helix is the switch helix, which moves after GTP hydrolysis. (C) The change in the conformation of the switch helix in domain 1 allows domains 2 and 3 to rotate as a single unit by about 90 degrees toward the viewer, which releases the tRNA that was bound to this structure (see also Figure 3–73). (A, adapted from H. Berchtold et al., *Nature* 365:126–132, 1993. With permission from Macmillan Publishers Ltd. B, courtesy of Mathias Sprinzl and Rolf Hilgenfeld. PDB code: 1EFT.)

polypeptide chain only after the GTP bound to EF-Tu is hydrolyzed, dissociating the EF-Tu. Since this GTP hydrolysis is triggered by a proper fit of the tRNA to the mRNA molecule on the ribosome, the EF-Tu serves as a factor that discriminates between correct and incorrect mRNA-tRNA pairings (see Figure 6–65).

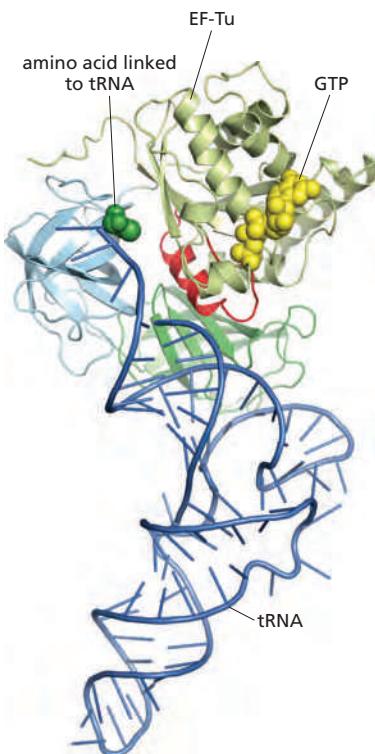
By comparing the three-dimensional structure of EF-Tu in its GTP-bound and GDP-bound forms, we can see how the repositioning of the tRNA occurs. The dissociation of the inorganic phosphate group ( $P_i$ ), which follows the reaction  $GTP \rightarrow GDP + P_i$ , causes a shift of a few tenths of a nanometer at the GTP-binding site, just as it does in the Ras protein. This tiny movement, equivalent to a few times the diameter of a hydrogen atom, causes a conformational change to propagate along a crucial piece of  $\alpha$  helix, called the *switch helix*, in the Ras-like domain of the protein. The switch helix seems to serve as a latch that adheres to a specific site in another domain of the molecule, holding the protein in a “shut” conformation. The conformational change triggered by GTP hydrolysis causes the switch helix to detach, allowing separate domains of the protein to swing apart, through a distance of about 4 nm (Figure 3–72). This releases the bound tRNA molecule, allowing its attached amino acid to be used (Figure 3–73).

Notice in this example how cells have exploited a simple chemical change that occurs on the surface of a small protein domain to create a movement 50 times larger. Dramatic shape changes of this type also cause the very large movements that occur in motor proteins, as we discuss next.

### Motor Proteins Produce Large Movements in Cells

We have seen that conformational changes in proteins have a central role in enzyme regulation and cell signaling. We now discuss proteins whose major function is to move other molecules. These **motor proteins** generate the forces responsible for muscle contraction and the crawling and swimming of cells. Motor proteins also power smaller-scale intracellular movements: they help to move chromosomes to opposite ends of the cell during mitosis (discussed in Chapter 17), to move organelles along molecular tracks within the cell (discussed in Chapter 16), and to move enzymes along a DNA strand during the synthesis of a new DNA molecule (discussed in Chapter 5). All these fundamental processes depend on proteins with moving parts that operate as force-generating machines.

How do these machines work? In other words, how do cells use shape changes in proteins to generate directed movements? If, for example, a protein is required to walk along a narrow thread such as a DNA molecule, it can do this by undergoing a series of conformational changes, such as those shown in Figure 3–74. But with nothing to drive these changes in an orderly sequence, they are perfectly



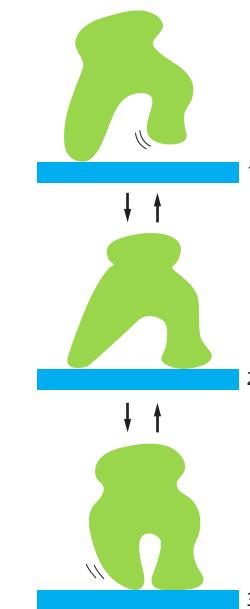
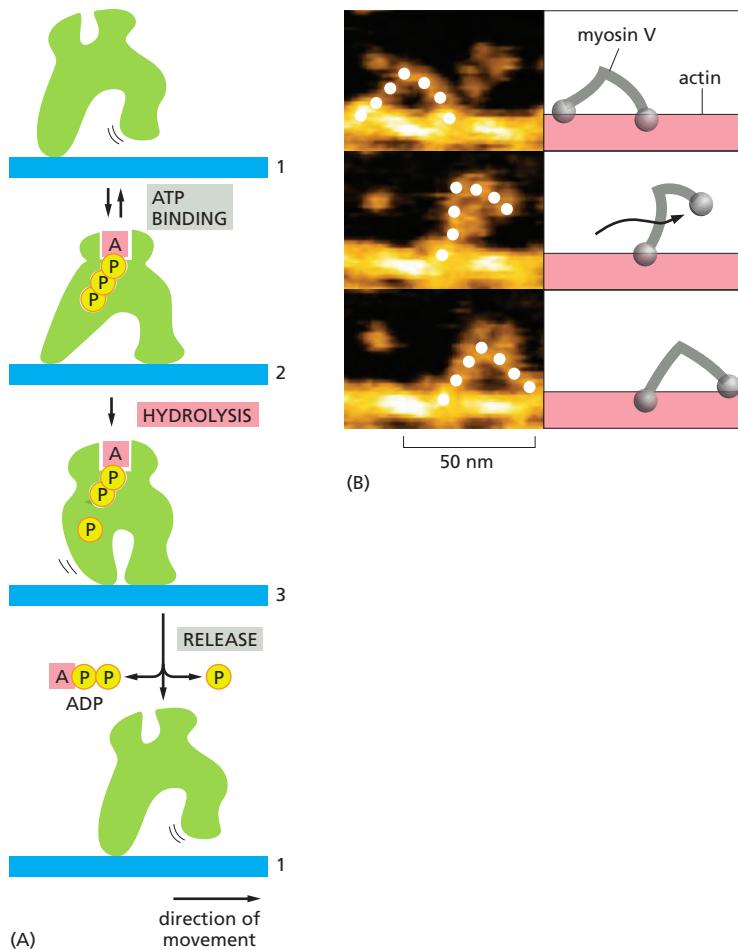
**Figure 3–73** An aminoacyl tRNA molecule bound to EF-Tu. Note how the bound protein blocks the use of the tRNA-linked amino acid (green) for protein synthesis until GTP hydrolysis triggers the conformational changes shown in Figure 3–72C, dissociating the protein-tRNA complex. EF-Tu is a bacterial protein; however, a very similar protein exists in eukaryotes, where it is called EF-1 (Movie 3.12). (Coordinates determined by P. Nissen et al., *Science* 270:1464–1472, 1995. PDB code: 1B23.)

reversible, and the protein can only wander randomly back and forth along the thread. We can look at this situation in another way. Since the directional movement of a protein does work, the laws of thermodynamics (discussed in Chapter 2) demand that such movement use free energy from some other source (otherwise the protein could be used to make a perpetual motion machine). Therefore, without an input of energy, the protein molecule can only wander aimlessly.

How can the cell make such a series of conformational changes unidirectional? To force the entire cycle to proceed in one direction, it is enough to make any one of the changes in shape irreversible. Most proteins that are able to walk in one direction for long distances achieve this motion by coupling one of the conformational changes to the hydrolysis of an ATP molecule that is tightly bound to the protein. The mechanism is similar to the one just discussed that drives allosteric protein shape changes by GTP hydrolysis. Because ATP (or GTP) hydrolysis releases a great deal of free energy, it is very unlikely that the nucleotide-binding protein will undergo the reverse shape change needed for moving backward—since this would require that it also reverse the ATP hydrolysis by adding a phosphate molecule to ADP to form ATP.

In the model shown in **Figure 3–75A**, ATP binding shifts a motor protein from conformation 1 to conformation 2. The bound ATP is then hydrolyzed to produce ADP and inorganic phosphate ( $P_i$ ), causing a change from conformation 2 to conformation 3. Finally, the release of the bound ADP and  $P_i$  drives the protein back to conformation 1. Because the energy provided by ATP hydrolysis drives the transition  $2 \rightarrow 3$ , this series of conformational changes is effectively irreversible. Thus, the entire cycle goes in only one direction, causing the protein molecule to walk continuously to the right in this example.

Many motor proteins generate directional movement through the use of a similar unidirectional ratchet, including the muscle motor protein *myosin*,



**Figure 3–74** An allosteric “walking” protein. Although its three different conformations allow it to wander randomly back and forth while bound to a thread or a filament, the protein cannot move uniformly in a single direction.

**Figure 3–75** How a protein can walk in one direction. (A) An allosteric motor protein driven by ATP hydrolysis. The transition between three different conformations includes a step driven by the hydrolysis of a bound ATP molecule, creating a “unidirectional ratchet” that makes the entire cycle essentially irreversible. By repeated cycles, the protein therefore moves continuously to the right along the thread. (B) Direct visualization of a walking myosin motor protein by high-speed atomic force microscopy; the elapsed time between steps was less than 0.5 sec (see Movie 16.3). (B, modified from N. Kodera et al., *Nature* 468:72–76, 2010. With permission from Macmillan Publishers Ltd.)

which walks along actin filaments (Figure 3–75B), and the *kinesin* proteins that walk along microtubules (both discussed in Chapter 16). These movements can be rapid: some of the motor proteins involved in DNA replication (the DNA helicases) propel themselves along a DNA strand at rates as high as 1000 nucleotides per second.

### Membrane-Bound Transporters Harness Energy to Pump Molecules Through Membranes

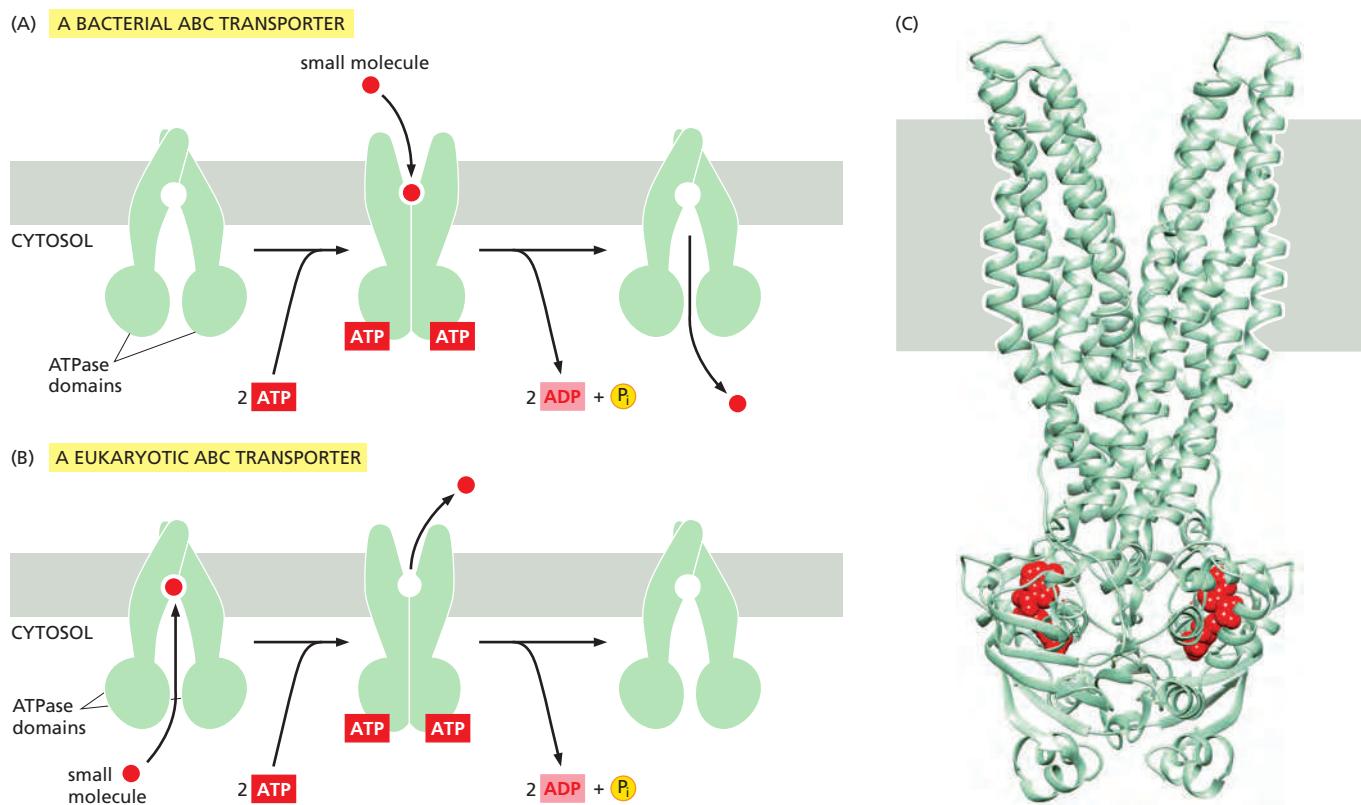
We have thus far seen how proteins that undergo allosteric shape changes can act as microprocessors (Src family kinases), as assembly factors (EF-Tu), and as generators of mechanical force and motion (motor proteins). Allosteric proteins can also harness energy derived from ATP hydrolysis, ion gradients, or electron-transport processes to pump specific ions or small molecules across a membrane. We consider one example here that will be discussed in more detail in Chapter 11.

The ABC transporters (ATP-binding cassette transporters) constitute an important class of membrane-bound pump proteins. In humans, at least 48 different genes encode them. These transporters mostly function to export hydrophobic molecules from the cytoplasm, serving to remove toxic molecules at the mucosal surface of the intestinal tract, for example, or at the blood-brain barrier. The study of ABC transporters is of intense interest in clinical medicine, because the overproduction of proteins in this class contributes to the resistance of tumor cells to chemotherapeutic drugs. In bacteria, the same types of proteins primarily function to import essential nutrients into the cell.

A typical ABC transporter contains a pair of membrane-spanning subunits linked to a pair of ATP-binding subunits located just below the plasma membrane. As in other examples we have discussed, the hydrolysis of the bound ATP molecules drives conformational changes in the protein, transmitting forces that cause the membrane-spanning subunits to move their bound molecules across the lipid bilayer (Figure 3–76).

Humans have invented many different types of mechanical pumps, and it should not be surprising that cells also contain membrane-bound pumps that

**Figure 3–76** The ABC (ATP-binding cassette) transporter, a protein machine that pumps molecules through a membrane. (A) How this large family of transporters pumps molecules into the cell in bacteria. As indicated, the binding of two molecules of ATP causes the two ATP-binding domains to clamp together tightly, opening a channel to the cell exterior. The binding of a substrate molecule to the extracellular face of the protein complex then triggers ATP hydrolysis followed by ADP release, which opens the cytoplasmic gate; the pump is then reset for another cycle. (B) As discussed in Chapter 11, in eukaryotes an opposite process occurs, causing selected substrate molecules to be pumped out of the cell. (C) The structure of a bacterial ABC transporter (see Movie 11.5). (C, from R.J. Dawson and K.P. Locher, *Nature* 443:180–185, 2006. With permission from Macmillan Publishers Ltd; PDB code: 2HYD).



function in other ways. Among the most notable are the rotary pumps that couple the hydrolysis of ATP to the transport of H<sup>+</sup> ions (protons). These pumps resemble miniature turbines, and they are used to acidify the interior of lysosomes and other eukaryotic organelles. Like other ion pumps that create ion gradients, they can function in reverse to catalyze the reaction ADP + P<sub>i</sub> → ATP, if the gradient across their membrane of the ion that they transport is steep enough.

One such pump, the ATP synthase, harnesses a gradient of proton concentration produced by electron-transport processes to produce most of the ATP used in the living world. This ubiquitous pump has a central role in energy conversion, and we shall discuss its three-dimensional structure and mechanism in Chapter 14.

### Proteins Often Form Large Complexes That Function as Protein Machines

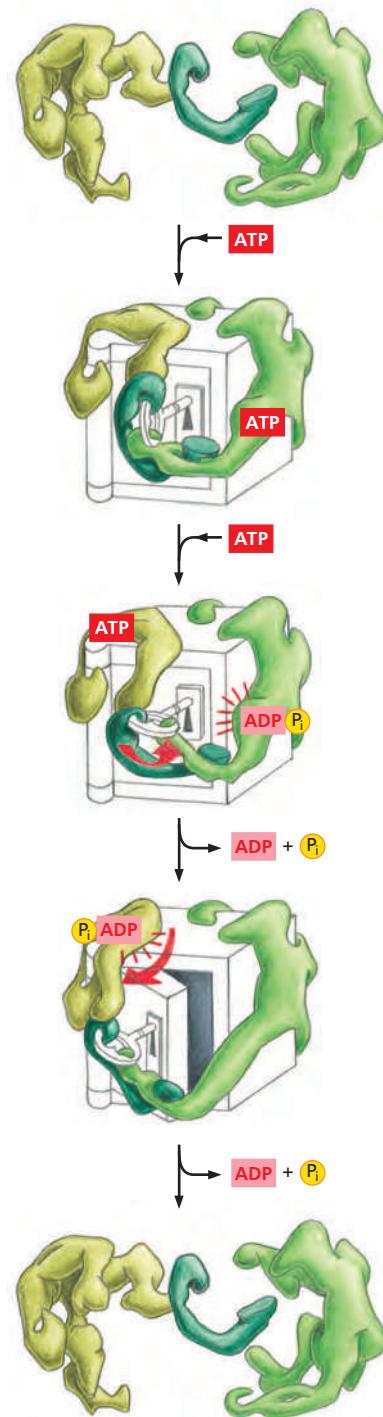
Large proteins formed from many domains are able to perform more elaborate functions than small, single-domain proteins. But large protein assemblies formed from many protein molecules linked together by noncovalent bonds perform the most impressive tasks. Now that it is possible to reconstruct most biological processes in cell-free systems in the laboratory, it is clear that each of the central processes in a cell—such as DNA replication, protein synthesis, vesicle budding, or transmembrane signaling—is catalyzed by a highly coordinated, linked set of 10 or more proteins. In most such *protein machines*, an energetically favorable reaction such as the hydrolysis of bound nucleoside triphosphates (ATP or GTP) drives an ordered series of conformational changes in one or more of the individual protein subunits, enabling the ensemble of proteins to move coordinately. In this way, each enzyme can be moved directly into position, as the machine catalyzes successive reactions in a series (Figure 3–77). This is what occurs, for example, in protein synthesis on a ribosome (discussed in Chapter 6)—or in DNA replication, where a large multiprotein complex moves rapidly along the DNA (discussed in Chapter 5).

Cells have evolved protein machines for the same reason that humans have invented mechanical and electronic machines. For accomplishing almost any task, manipulations that are spatially and temporally coordinated through linked processes are much more efficient than the use of many separate tools.

### Scaffolds Concentrate Sets of Interacting Proteins

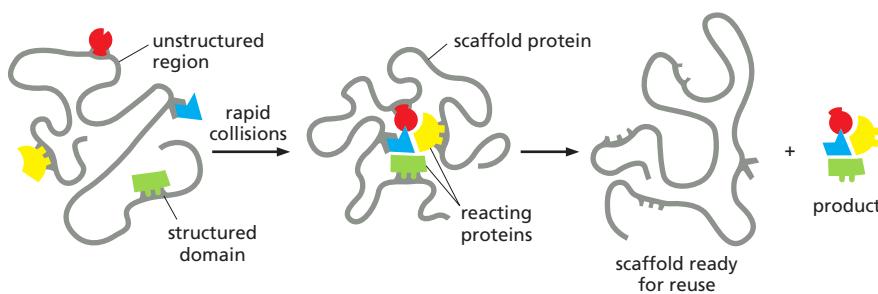
As scientists have learned more of the details of cell biology, they have recognized an increasing degree of sophistication in cell chemistry. Thus, not only do we now know that protein machines play a predominant role, but it has also become clear that they are very often localized to specific sites in the cell, being assembled and activated only where and when they are needed. As one example, when extracellular signaling molecules bind to receptor proteins in the plasma membrane, the activated receptors often recruit a set of other proteins to the inside surface of the plasma membrane to form a large protein complex that passes the signal on (discussed in Chapter 15).

The mechanisms frequently involve **scaffold proteins**. These are proteins with binding sites for multiple other proteins, and they serve both to link together specific sets of interacting proteins and to position them at specific locations inside a cell. At one extreme are rigid scaffolds, such as the cullin in SCF ubiquitin ligase (see Figure 3–71). At the other extreme are the large, flexible scaffold proteins that often underlie regions of specialized plasma membrane. These include the



**Figure 3–77** How “protein machines” carry out complex functions.

These machines are made of individual proteins that collaborate to perform a specific task (**Movie 3.13**). The movement of these proteins is often coordinated by the hydrolysis of a bound nucleotide such as ATP or GTP. Directional allosteric conformational changes of proteins that are driven in this way often occur in a large protein assembly in which the activities of several different protein molecules are coordinated by such movements within the complex.



*Discs-large protein* (Dlg), a protein of about 900 amino acids that is concentrated in special regions beneath the plasma membrane in epithelial cells and at synapses. Dlg contains binding sites for at least seven other proteins, interspersed with regions of more flexible polypeptide chain. An ancient protein, conserved in organisms as diverse as sponges, worms, flies, and humans, Dlg derives its name from the mutant phenotype of the organism in which it was first discovered; the cells in the imaginal discs of a *Drosophila* embryo with a mutation in the *Dlg* gene fail to stop proliferating when they should, and they produce unusually large discs whose epithelial cells can form tumors.

Although incompletely studied, Dlg and a large number of similar scaffold proteins are thought to function like the protein that is schematically illustrated in **Figure 3–78**. By binding a specific set of interacting proteins, these scaffolds can enhance the rate of critical reactions, while also confining them to the particular region of the cell that contains the scaffold. For similar reasons, cells also make extensive use of *scaffold RNA molecules*, as discussed in Chapter 7.

### Many Proteins Are Controlled by Covalent Modifications That Direct Them to Specific Sites Inside the Cell

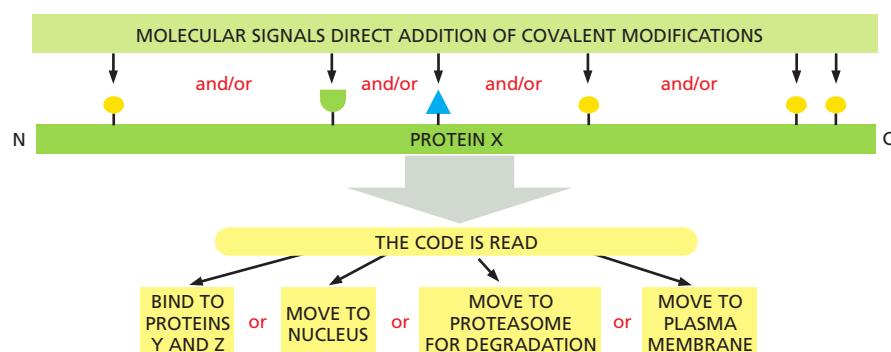
We have thus far described only a few ways in which proteins are post-translationally modified. A large number of other such modifications also occur, more than 200 distinct types being known. To give a sense of the variety, **Table 3–3** presents

**Figure 3–78** How the proximity created by scaffold proteins can greatly speed reactions in a cell. In this example, long unstructured regions of polypeptide chain in a large scaffold protein connect a series of structured domains that bind a set of reacting proteins. The unstructured regions serve as flexible “tethers” that greatly speed reaction rates by causing a rapid, random collision of all of the proteins that are bound to the scaffold. (For specific examples of protein tethering, see Figure 3–54 and Figure 16–18; for scaffold RNA molecules, see Figure 7–49B.)

**TABLE 3–3** Some Molecules Covalently Attached to Proteins Regulate Protein Function

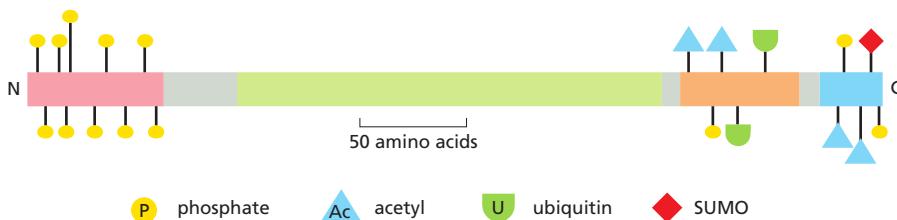
Modifying group	Some prominent functions
Phosphate on Ser, Thr, or Tyr	Drives the assembly of a protein into larger complexes (see Figure 15–11)
Methyl on Lys	Helps to create distinct regions in chromatin through forming either mono-, di-, or trimethyl lysine in histones (see Figure 4–36)
Acetyl on Lys	Helps to activate genes in chromatin by modifying histones (see Figure 4–33)
Palmityl group on Cys	This fatty acid addition drives protein association with membranes (see Figure 10–18)
N-acetylglucosamine on Ser or Thr	Controls enzyme activity and gene expression in glucose homeostasis
Ubiquitin on Lys	Monoubiquitin addition regulates the transport of membrane proteins in vesicles (see Figure 13–50) A polyubiquitin chain targets a protein for degradation (see Figure 3–70)
Ubiquitin is a 76-amino-acid polypeptide; there are at least 10 other ubiquitin-related proteins in mammalian cells.	

## (A) A SPECTRUM OF COVALENT MODIFICATIONS PRODUCES A REGULATORY PROTEIN CODE



**Figure 3–79** Multisite protein modification and its effects. (A) A protein that carries a post-translational addition to more than one of its amino acid side chains can be considered to carry a combinatorial regulatory code. Multisite modifications are added to (and removed from) a protein through signaling networks, and the resulting combinatorial regulatory code on the protein is read to alter its behavior in the cell. (B) The pattern of some covalent modifications to the protein p53.

## (B) SOME KNOWN MODIFICATIONS OF PROTEIN p53



a few of the modifying groups with known regulatory roles. As in phosphate and ubiquitin additions described previously, these groups are added and then removed from proteins according to the needs of the cell.

A large number of proteins are now known to be modified on more than one amino acid side chain, with different regulatory events producing a different pattern of such modifications. A striking example is the protein p53, which plays a central part in controlling a cell's response to adverse circumstances (see Figure 17–62). Through one of four different types of molecular additions, this protein can be modified at 20 different sites. Because an enormous number of different combinations of these 20 modifications are possible, the protein's behavior can in principle be altered in a huge number of ways. Such modifications will often create a site on the modified protein that binds it to a scaffold protein in a specific region of the cell, thereby connecting it—via the scaffold—to the other proteins required for a reaction at that site.

One can view each protein's set of covalent modifications as a *combinatorial regulatory code*. Specific modifying groups are added to or removed from a protein in response to signals, and the code then alters protein behavior—changing the activity or stability of the protein, its binding partners, and/or its specific location within the cell (Figure 3–79). As a result, the cell is able to respond rapidly and with great versatility to changes in its condition or environment.

### A Complex Network of Protein Interactions Underlies Cell Function

There are many challenges facing cell biologists in this information-rich era when a large number of complete genome sequences are known. One is the need to dissect and reconstruct each one of the thousands of protein machines that exist in an organism such as ourselves. To understand these remarkable protein complexes, each will need to be reconstituted from its purified protein parts, so that we can study its detailed mode of operation under controlled conditions in a test tube, free from all other cell components. This alone is a massive task. But we now know that each of these subcomponents of a cell also interacts with other sets of macromolecules, creating a large network of protein–protein and protein–nucleic acid interactions throughout the cell. To understand the cell, therefore, we will need to analyze most of these other interactions as well.

We can gain some idea of the complexity of intracellular protein networks from a particularly well-studied example described in Chapter 16: the many dozens of proteins that interact with the actin cytoskeleton to control actin filament behavior (see Panel 16–3, p. 905).

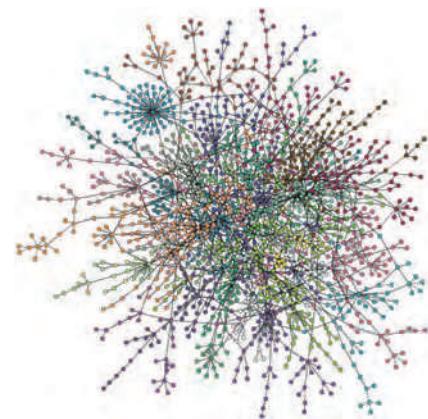
The extent of such protein–protein interactions can also be estimated more generally. An enormous amount of valuable information is now freely available in protein databases on the Internet: tens of thousands of three-dimensional protein structures plus tens of millions of protein sequences derived from the nucleotide sequences of genes. Scientists have been developing new methods for mining this great resource to increase our understanding of cells. In particular, computer-based bioinformatics tools are being combined with robotics and other technologies to allow thousands of proteins to be investigated in a single set of experiments. **Proteomics** is a term that is often used to describe such research focused on the analysis of large sets of proteins, analogous to the term *genomics* describing the large-scale analysis of DNA sequences and genes.

A biochemical method based on affinity tagging and mass spectroscopy has proven especially powerful for determining the direct binding interactions between the many different proteins in a cell (discussed in Chapter 8). The results are being tabulated and organized in Internet databases. This allows a cell biologist studying a small set of proteins to readily discover which other proteins in the same cell are likely to bind to, and thus interact with, that set of proteins. When displayed graphically as a *protein interaction map*, each protein is represented by a box or dot in a two-dimensional network, with a straight line connecting those proteins that have been found to bind to each other.

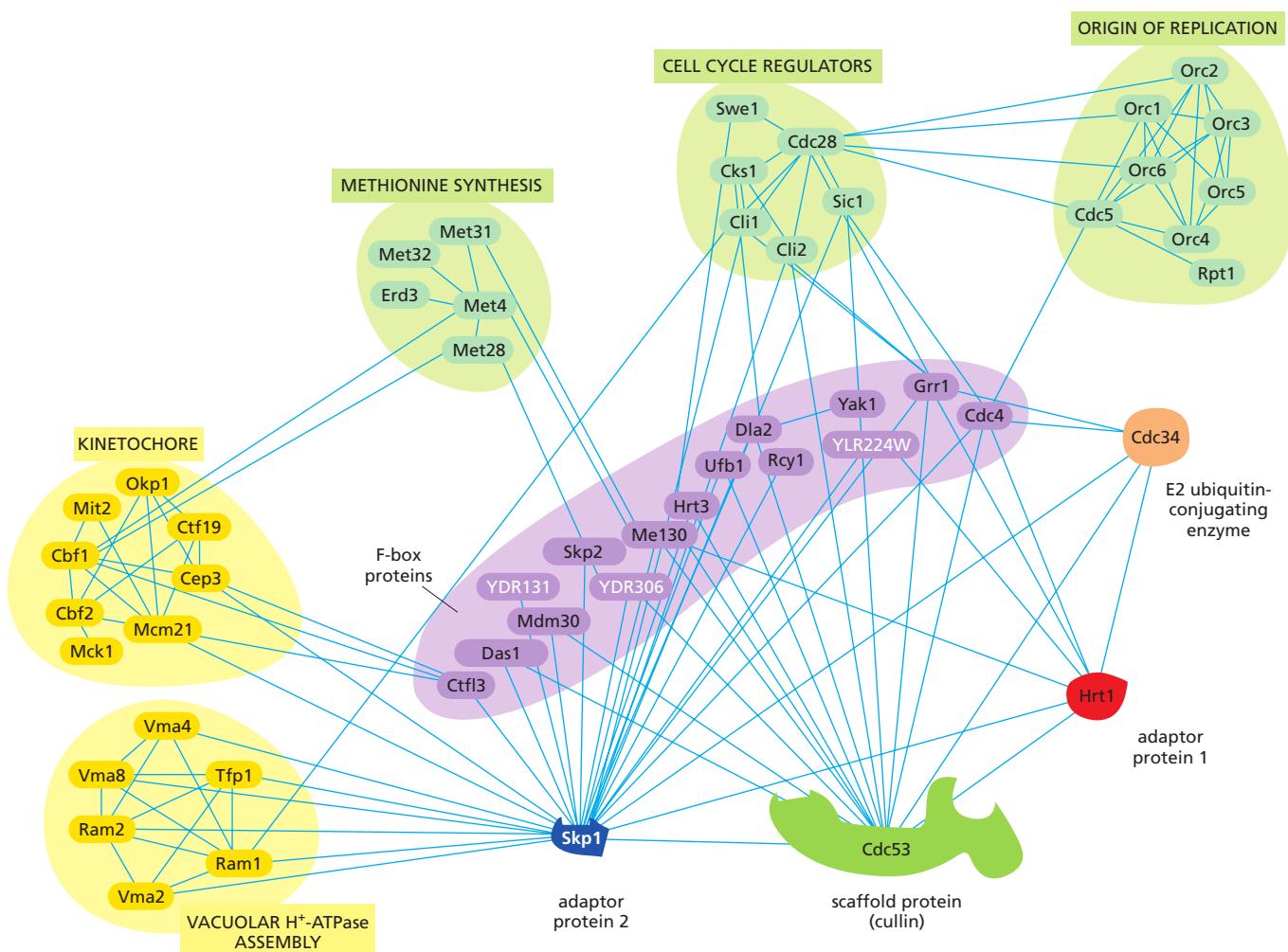
When hundreds or thousands of proteins are displayed on the same map, the network diagram becomes bewilderingly complicated, serving to illustrate the enormous challenges that face scientists attempting to understand the cell (**Figure 3–80**). Much more useful are small subsections of these maps, centered on a few proteins of interest.

We have previously described the structure and mode of action of the SCF ubiquitin ligase, using it to illustrate how protein complexes are constructed from interchangeable parts (see Figure 3–71). **Figure 3–81** shows a network of protein–protein interactions for the five proteins that form this protein complex in a yeast cell. Four of the subunits of this ligase are located at the bottom right of this figure. The remaining subunit, the F-box protein that serves as its substrate-binding arm, appears as a set of 15 different gene products that bind to adaptor protein 2 (the Skp1 protein). Along the top and left of the figure are sets of additional protein interactions marked with *yellow* and *green* shading: as indicated, these protein sets function at the origin of DNA replication, in cell cycle regulation, in methionine synthesis, in the kinetochore, and in vacuolar H<sup>+</sup>-ATPase assembly. We shall use this figure to explain how such protein interaction maps are used, and what they do and do not mean.

1. Protein interaction maps are useful for identifying the likely function of previously uncharacterized proteins. Examples are the products of the genes that have thus far only been inferred to exist from the yeast genome sequence, which are the three proteins in the figure that lack a simple three-letter abbreviation (*white letters* beginning with Y). The three in this diagram are F-box proteins that bind to Skp1; these are therefore likely to function as part of the ubiquitin ligase, serving as substrate-binding arms that recognize different target proteins. However, as we discuss next, neither assignment can be considered certain without additional data.
2. Protein interaction networks need to be interpreted with caution because, as a result of evolution making efficient use of each organism's genetic information, the same protein can be used as part of different protein complexes that have different types of functions. Thus, although protein A binds to protein B and protein B binds to protein C, proteins A and C need not function in the same process. For example, we know from detailed biochemical studies that the functions of Skp1 in the kinetochore and in



**Figure 3–80** A network of protein-binding interactions in a yeast cell. Each line connecting a pair of dots (proteins) indicates a protein–protein interaction. (From A. Guimerá and M. Sales-Pardo, *Mol. Syst. Biol.* 2:42, 2006. With permission from Macmillan Publishers Ltd.)



**Figure 3–81** A map of some protein–protein interactions of the SCF ubiquitin ligase and other proteins in the yeast *S. cerevisiae*. The symbols and/or colors used for the five proteins of the ligase are those in Figure 3–71. Note that 15 different F-box proteins are shown (purple); those with white lettering (beginning with Y) are known from the genome sequence as open reading frames. For additional details, see text. (Courtesy of Peter Bowers and David Eisenberg, UCLA-DOE Institute for Genomics and Proteomics, UCLA.)

vacuolar H<sup>+</sup>-ATPase assembly (*yellow shading*) are separate from its function in the SCF ubiquitin ligase. In fact, only the remaining three functions of Skp1 illustrated in the diagram—methionine synthesis, cell cycle regulation, and origin of replication (*green shading*)—involve ubiquitylation.

3. In cross-species comparisons, those proteins displaying similar patterns of interactions in the two protein interaction maps are likely to have the same function in the cell. Thus, as scientists generate more and more highly detailed maps for multiple organisms, the results will become increasingly useful for inferring protein function. These map comparisons will be a particularly powerful tool for deciphering the functions of human proteins, because a vast amount of direct information about protein function can be obtained from genetic engineering, mutational, and genetic analyses in experimental organisms—such as yeast, worms, and flies—that are not feasible in humans.

What does the future hold? There are likely to be on the order of 10,000 different proteins in a typical human cell, each of which interacts with 5 to 10 different partners. Despite the enormous progress made in recent years, we cannot yet claim to understand even the simplest known cells, such as the small *Mycoplasma* bacterium formed from only about 500 gene products (see Figure 1–10). How then

can we hope to understand a human? Clearly, a great deal of new biochemistry will be essential, in which each protein in a particular interacting set is purified so that its chemistry and interactions can be dissected in a test tube. But in addition, more powerful ways of analyzing networks will be needed based on mathematical and computational tools not yet invented, as we shall emphasize in Chapter 8. Clearly, there are many wonderful challenges that remain for future generations of cell biologists.

## Summary

*Proteins can form enormously sophisticated chemical devices, whose functions largely depend on the detailed chemical properties of their surfaces. Binding sites for ligands are formed as surface cavities in which precisely positioned amino acid side chains are brought together by protein folding. In this way, normally unreactive amino acid side chains can be activated to make and break covalent bonds. Enzymes are catalytic proteins that greatly speed up reaction rates by binding the high-energy transition states for a specific reaction path; they also can perform acid catalysis and base catalysis simultaneously. The rates of enzyme reactions are often so fast that they are limited only by diffusion. Rates can be further increased only if enzymes that act sequentially on a substrate are joined into a single multienzyme complex, or if the enzymes and their substrates are attached to protein scaffolds, or otherwise confined to the same part of the cell.*

*Proteins reversibly change their shape when ligands bind to their surface. The allosteric changes in protein conformation produced by one ligand affect the binding of a second ligand, and this linkage between two ligand-binding sites provides a crucial mechanism for regulating cell processes. Metabolic pathways, for example, are controlled by feedback regulation: some small molecules inhibit and other small molecules activate enzymes early in a pathway. Enzymes controlled in this way generally form symmetric assemblies, allowing cooperative conformational changes to create a steep response to changes in the concentrations of the ligands that regulate them.*

*The expenditure of chemical energy can drive unidirectional changes in protein shape. By coupling allosteric shape changes to ATP hydrolysis, for example, proteins can do useful work, such as generating a mechanical force or moving for long distances in a single direction. The three-dimensional structures of proteins have revealed how a small local change caused by nucleoside triphosphate hydrolysis is amplified to create major changes elsewhere in the protein. By such means, these proteins can serve as input-output devices that transmit information, as assembly factors, as motors, or as membrane-bound pumps. Highly efficient protein machines are formed by incorporating many different protein molecules into larger assemblies that coordinate the allosteric movements of the individual components. Such machines perform most of the important reactions in cells.*

*Proteins are subjected to many reversible, post-translational modifications, such as the covalent addition of a phosphate or an acetyl group to a specific amino acid side chain. The addition of these modifying groups is used to regulate the activity of a protein, changing its conformation, its binding to other proteins, and its location inside the cell. A typical protein in a cell will interact with more than five different partners. Through proteomics, biologists can analyze thousands of proteins in one set of experiments. One important result is the production of detailed protein interaction maps, which aim at describing all of the binding interactions between the thousands of distinct proteins in a cell. However, understanding these networks will require new biochemistry, through which small sets of interacting proteins can be purified and their chemistry dissected in detail. In addition, new computational techniques will be required to deal with the enormous complexity.*

## WHAT WE DON'T KNOW

- What are the functions of the surprisingly large amount of unfolded polypeptide chain found in proteins?
- How many types of protein functions remain to be discovered? What are the most promising approaches for discovering them?
- When will scientists be able to take any amino acid sequence and accurately predict both that protein's three-dimensional conformations and its chemical properties? What breakthroughs will be needed to accomplish this important goal?
- Are there ways to reveal the detailed workings of a protein machine that do not require the purification of each of its component parts in large amounts, so that the machine's functions can be reconstituted and dissected using chemical techniques in a test tube?
- What are the roles of the dozens of different types of covalent modifications of proteins that have been found in addition to those listed in Table 3–3? Which ones are critical for cell function and why?
- Why is amyloid toxic to cells and how does it contribute to neurodegenerative diseases such as Alzheimer's disease?

## PROBLEMS

Which statements are true? Explain why or why not.

**3–1** Each strand in a  $\beta$  sheet is a helix with two amino acids per turn.

**3–2** Intrinsically disordered regions of proteins can be identified using bioinformatic methods to search genes for encoded amino acid sequences that possess high hydrophobicity and low net charge.

**3–3** Loops of polypeptide that protrude from the surface of a protein often form the binding sites for other molecules.

**3–4** An enzyme reaches a maximum rate at high substrate concentration because it has a fixed number of active sites where substrate binds.

**3–5** Higher concentrations of enzyme give rise to a higher turnover number.

**3–6** Enzymes that undergo cooperative allosteric transitions invariably consist of symmetric assemblies of multiple subunits.

**3–7** Continual addition and removal of phosphates by protein kinases and protein phosphatases is wasteful of energy—since their combined action consumes ATP—but it is a necessary consequence of effective regulation by phosphorylation.

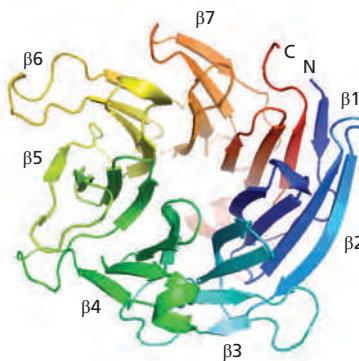
Discuss the following problems.

**3–8** Consider the following statement. “To produce one molecule of each possible kind of polypeptide chain, 300 amino acids in length, would require more atoms than exist in the universe.” Given the size of the universe, do you suppose this statement could possibly be correct? Since counting atoms is a tricky business, consider the problem from the standpoint of mass. The mass of the observable universe is estimated to be about  $10^{80}$  grams, give or take an order of magnitude or so. Assuming that the average mass of an amino acid is 110 daltons, what would be the mass of one molecule of each possible kind of polypeptide chain 300 amino acids in length? Is this greater than the mass of the universe?

**3–9** A common strategy for identifying distantly related homologous proteins is to search the database using a short signature sequence indicative of the particular protein function. Why is it better to search with a short sequence than with a long sequence? Do you not have more chances for a “hit” in the database with a long sequence?

**3–10** The so-called kelch motif consists of a four-stranded  $\beta$  sheet, which forms what is known as a  $\beta$  propeller. It is usually found to be repeated four to seven times, forming a kelch repeat domain in a multidomain protein.

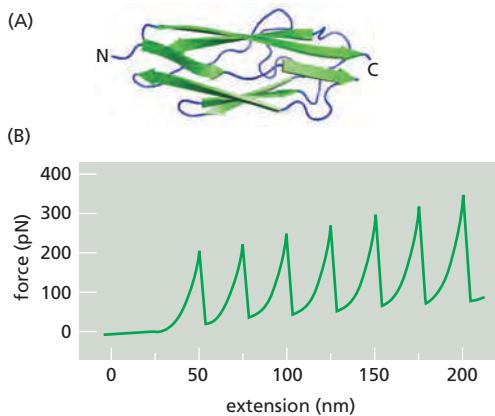
One such kelch repeat domain is shown in **Figure Q3–1**. Would you classify this domain as an “in-line” or “plug-in” type domain?



**Figure Q3–1** The kelch repeat domain of galactose oxidase from *D. dendroides* (Problem 3–10). The seven individual  $\beta$  propellers are color coded and labeled. The N- and C-termini are indicated by N and C.

**3–11** Titin, which has a molecular weight of about  $3 \times 10^6$ , is the largest polypeptide yet described. Titin molecules extend from muscle thick filaments to the Z disc; they are thought to act as springs to keep the thick filaments centered in the sarcomere. Titin is composed of a large number of repeated immunoglobulin (Ig) sequences of 89 amino acids, each of which is folded into a domain about 4 nm in length (**Figure Q3–2A**).

You suspect that the springlike behavior of titin is caused by the sequential unfolding (and refolding) of individual Ig domains. You test this hypothesis using the atomic force microscope, which allows you to pick up one end of a protein molecule and pull with an accurately measured force. For a fragment of titin containing seven repeats of the Ig domain, this experiment gives the sawtooth force-versus-extension curve shown in **Figure Q3–2B**. If the experiment is repeated in a solution of 8 M urea (a protein denaturant), the peaks disappear and the measured extension becomes much longer for a given force. If the experiment is repeated after the protein has been cross-linked by treatment with glutaraldehyde, once again the peaks disappear but the extension becomes much smaller for a given force.



**Figure Q3–2** Springlike behavior of titin (Problem 3–11). (A) The structure of an individual Ig domain. (B) Force in piconewtons versus extension in nanometers obtained by atomic force microscopy.

**A.** Are the data consistent with your hypothesis that titin's springlike behavior is due to the sequential unfolding of individual Ig domains? Explain your reasoning.

**B.** Is the extension for each putative domain-unfolding event the magnitude you would expect? (In an extended polypeptide chain, amino acids are spaced at intervals of 0.34 nm.)

**C.** Why is each successive peak in Figure Q3-2B a little higher than the one before?

**D.** Why does the force collapse so abruptly after each peak?

**3-12** Rous sarcoma virus (RSV) carries an oncogene called *Src*, which encodes a continuously active protein tyrosine kinase that leads to unchecked cell proliferation. Normally, Src carries an attached fatty acid (myristoylate) group that allows it to bind to the cytoplasmic side of the plasma membrane. A mutant version of Src that does not allow attachment of myristoylate does not bind to the membrane. Infection of cells with RSV encoding either the normal or the mutant form of Src leads to the same high level of protein tyrosine kinase activity, but the mutant Src does not cause cell proliferation.

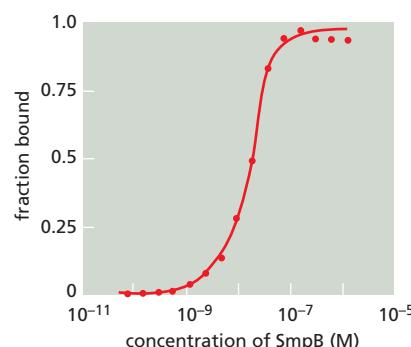
**A.** Assuming that the normal Src is all bound to the plasma membrane and that the mutant Src is distributed throughout the cytoplasm, calculate their relative concentrations in the neighborhood of the plasma membrane. For the purposes of this calculation, assume that the cell is a sphere with a radius ( $r$ ) of 10  $\mu\text{m}$  and that the mutant Src is distributed throughout the cell, whereas the normal Src is confined to a 4-nm-thick layer immediately beneath the membrane. [For this problem, assume that the membrane has no thickness. The volume of a sphere is  $(4/3)\pi r^3$ .]

**B.** The target (X) for phosphorylation by Src resides in the membrane. Explain why the mutant Src does not cause cell proliferation.

**3-13** An antibody binds to another protein with an equilibrium constant,  $K$ , of  $5 \times 10^9 \text{ M}^{-1}$ . When it binds to a second, related protein, it forms three fewer hydrogen bonds, reducing its binding affinity by 11.9 kJ/mole. What is the  $K$  for its binding to the second protein? (Free-energy change is related to the equilibrium constant by the equation  $\Delta G^\circ = -2.3 RT \log K$ , where  $R$  is  $8.3 \times 10^{-3} \text{ kJ}/(\text{mole K})$  and  $T$  is 310 K.)

**3-14** The protein SmpB binds to a special species of tRNA, tmRNA, to eliminate the incomplete proteins made from truncated mRNAs in bacteria. If the binding of SmpB to tmRNA is plotted as fraction tmRNA bound versus SmpB concentration, one obtains a symmetrical S-shaped curve as shown in Figure Q3-3. This curve is a visual display of a very useful relationship between  $K_d$  and concentration, which has broad applicability. The general expression for fraction of ligand bound is derived from the equation for  $K_d$  ( $K_d = [\text{Pr-L}] / [\text{Pr-L}]_{\text{TOT}}$ ) by substituting  $([\text{L}]_{\text{TOT}} - [\text{L}])$  for  $[\text{Pr-L}]$  and rearranging. Because the total concentration of ligand ( $[\text{L}]_{\text{TOT}}$ ) is equal to the free ligand ( $[\text{L}]$ ) plus bound ligand ( $[\text{Pr-L}]$ ),

$$\text{fraction bound} = [\text{Pr-L}] / [\text{L}]_{\text{TOT}} = [\text{Pr}] / ([\text{Pr}] + K_d)$$



**Figure Q3-3** Fraction of tmRNA bound versus SmpB concentration (Problem 3-14).

For SmpB and tmRNA, the fraction bound =  $[\text{SmpB-tmRNA}]_{\text{TOT}} / [\text{tmRNA}]_{\text{TOT}} = [\text{SmpB}] / ([\text{SmpB}] + K_d)$ . Using this relationship, calculate the fraction of tmRNA bound for SmpB concentrations equal to  $10^4 K_d$ ,  $10^3 K_d$ ,  $10^2 K_d$ ,  $10^1 K_d$ ,  $K_d$ ,  $10^{-1} K_d$ ,  $10^{-2} K_d$ ,  $10^{-3} K_d$ , and  $10^{-4} K_d$ .

**3-15** Many enzymes obey simple Michaelis-Menten kinetics, which are summarized by the equation

$$\text{rate} = V_{\text{max}} [\text{S}] / ([\text{S}] + K_m)$$

where  $V_{\text{max}}$  = maximum velocity,  $[\text{S}]$  = concentration of substrate, and  $K_m$  = the Michaelis constant.

It is instructive to plug a few values of  $[\text{S}]$  into the equation to see how rate is affected. What are the rates for  $[\text{S}]$  equal to zero, equal to  $K_m$ , and equal to infinite concentration?

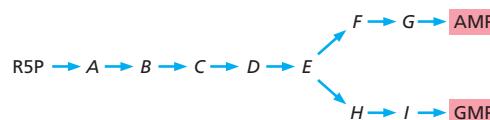
**3-16** The enzyme hexokinase adds a phosphate to D-glucose but ignores its mirror image, L-glucose. Suppose that you were able to synthesize hexokinase entirely from D-amino acids, which are the mirror image of the normal L-amino acids.

**A.** Assuming that the “D” enzyme would fold to a stable conformation, what relationship would you expect it to bear to the normal “L” enzyme?

**B.** Do you suppose the “D” enzyme would add a phosphate to L-glucose, and ignore D-glucose?

**3-17** How do you suppose that a molecule of hemoglobin is able to bind oxygen efficiently in the lungs, and yet release it efficiently in the tissues?

**3-18** Synthesis of the purine nucleotides AMP and GMP proceeds by a branched pathway starting with ribose 5-phosphate (R5P), as shown schematically in Figure Q3-4. Using the principles of feedback inhibition, propose a regulatory strategy for this pathway that ensures an adequate supply of both AMP and GMP and minimizes the buildup of the intermediates ( $A-I$ ) when supplies of AMP and GMP are adequate.



**Figure Q3-4** Schematic diagram of the metabolic pathway for synthesis of AMP and GMP from R5P (Problem 3-18).

## REFERENCES

### General

- Berg JM, Tymoczko JL & Stryer L (2011) Biochemistry, 7th ed. New York: WH Freeman.
- Branden C & Tooze J (1999) Introduction to Protein Structure, 2nd ed. New York: Garland Science.
- Dickerson RE (2005) Present at the Flood: How Structural Molecular Biology Came About. Sunderland, MA: Sinauer.
- Kuriyan J, Konforti B & Wemmer D (2013) The Molecules of Life: Physical and Chemical Principles. New York: Garland Science.
- Perutz M (1992) Protein Structure: New Approaches to Disease and Therapy. New York: WH Freeman.
- Petsko GA & Ringe D (2004) Protein Structure and Function. London: New Science Press.
- Williamson M (2011) How Proteins Work. New York: Garland Science.

### The Shape and Structure of Proteins

- Anfinsen CB (1973) Principles that govern the folding of protein chains. *Science* 181, 223–230.
- Caspar DLD & Klug A (1962) Physical principles in the construction of regular viruses. *Cold Spring Harb. Symp. Quant. Biol.* 27, 1–24.
- Dill KA & MacCallum JL (2012) The protein-folding problem, 50 years on. *Science* 338, 1042–1046.
- Eisenberg D (2003) The discovery of the alpha-helix and beta-sheet, the principal structural features of proteins. *Proc. Natl. Acad. Sci. USA* 100, 11207–11210.
- Fraenkel-Conrat H & Williams RC (1955) Reconstitution of active tobacco mosaic virus from its inactive protein and nucleic acid components. *Proc. Natl. Acad. Sci. USA* 41, 690–698.
- Goodsell DS & Olson AJ (2000) Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* 29, 105–153.
- Greenwald J & Riek R (2010) Biology of amyloid: structure, function, and regulation. *Structure* 18, 1244–1260.
- Harrison SC (1992) Viruses. *Curr. Opin. Struct. Biol.* 2, 293–299.
- Hudder A, Nathanson L & Deutscher MP (2003) Organization of mammalian cytoplasm. *Mol. Cell. Biol.* 23, 9318–9326.
- Kato M, Han TW, Xie S et al. (2012) Cell-free formation of RNA granules: low complexity sequence domains form dynamic fibers within hydrogels. *Cell* 149, 753–767.
- Koga N, Tatsumi-Koga R, Liu G et al. (2012) Principles for designing ideal protein structures. *Nature* 491, 222–227.
- Li P, Banjade S, Cheng H-C et al. (2012) Phase transitions in the assembly of multivalent signalling proteins. *Nature* 483, 336–340.
- Lindquist SL & Kelly JW (2011) Chemical and biological approaches for adapting proteostasis to ameliorate protein misfolding and aggregation diseases—progress and prognosis. *Cold Spring Harb. Perspect. Biol.* 3, a004507.
- Maji SK, Perrin MH, Sawaya MR et al. (2009) Functional amyloids as natural storage of peptide hormones in pituitary secretory granules. *Science* 325, 328–332.
- Nelson R, Sawaya MR, Balbirnie M et al. (2005) Structure of the cross- $\beta$  spine of amyloid-like fibrils. *Nature* 435, 773–778.
- Nomura M (1973) Assembly of bacterial ribosomes. *Science* 179, 864–873.
- Oldfield CJ & Dunker AK (2014) Intrinsically disordered proteins and intrinsically disordered protein regions. *Annu. Rev. Biochem.* 83, 553–584.
- Orengo CA & Thornton JM (2005) Protein families and their evolution—a structural perspective. *Annu. Rev. Biochem.* 74, 867–900.
- Pauling L & Corey RB (1951) Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. *Proc. Natl. Acad. Sci. USA* 37, 729–740.
- Pauling L, Corey RB & Branson HR (1951) The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. USA* 37, 205–211.
- Prusiner SB (1998) Prions. *Proc. Natl. Acad. Sci. USA* 95, 13363–13383.

- Toyama BH & Weissman JS (2011) Amyloid structure: conformational diversity and consequences. *Annu. Rev. Biochem.* 80, 557–585.
- Zhang C & Kim SH (2003) Overview of structural genomics: from structure to function. *Curr. Opin. Chem. Biol.* 7, 28–32.

### Protein Function

- Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92, 291–294.
- Benkovic SJ (1992) Catalytic antibodies. *Annu. Rev. Biochem.* 61, 29–54.
- Berg OG & von Hippel PH (1985) Diffusion-controlled macromolecular interactions. *Annu. Rev. Biophys. Biomol. Struct.* 14, 131–160.
- Bourne HR (1995) GTPases: a family of molecular switches and clocks. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 349, 283–289.
- Costanzo M, Baryshnikova A, Bellay J et al. (2010) The genetic landscape of a cell. *Science* 327, 425–431.
- Deshaires RJ & Joazeiro CAP (2009) RING domain E3 ubiquitin ligases. *Annu. Rev. Biochem.* 78, 399–434.
- Dickerson RE & Geis I (1983) Hemoglobin: Structure, Function, Evolution, and Pathology. Menlo Park, CA: Benjamin Cummings.
- Fersht AR (1999) Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding. New York: WH Freeman.
- Haucke V, Neher E & Sigrist SJ (2011) Protein scaffolds in the coupling of synaptic exocytosis and endocytosis. *Nat. Rev. Neurosci.* 12, 127–138.
- Hua Z & Vierstra RD (2011) The cullin-RING ubiquitin-protein ligases. *Annu. Rev. Plant Biol.* 62, 299–334.
- Hunter T (2012) Why nature chose phosphate to modify proteins. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2513–2516.
- Johnson LN & Lewis RJ (2001) Structural basis for control by phosphorylation. *Chem. Rev.* 101, 2209–2242.
- Kantrowitz ER & Lipscomb WN (1988) *Escherichia coli* aspartate transcarbamylase: the relation between structure and function. *Science* 241, 669–674.
- Kerscher O, Felberbaum R & Hochstrasser M (2006) Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu. Rev. Cell Dev. Biol.* 22, 159–180.
- Kim E & Sheng M (2004) PDZ domain proteins of synapses. *Nat. Rev. Neurosci.* 5, 771–781.
- Koshland DE Jr (1984) Control of enzyme activity and metabolic pathways. *Trends Biochem. Sci.* 9, 155–159.
- Krogan NJ, Cagney G, Yu H et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 440, 637–643.
- Lichtarge O, Bourne HR & Cohen FE (1996) An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* 257, 342–358.
- Maier T, Leibundgut M & Ban N (2008) The crystal structure of a mammalian fatty acid synthase. *Science* 321, 1315–1322.
- Monod J, Changeux JP & Jacob F (1963) Allosteric proteins and cellular control systems. *J. Mol. Biol.* 6, 306–329.
- Perutz M (1990) Mechanisms of Cooperativity and Allosteric Regulation in Proteins. Cambridge: Cambridge University Press.
- Radzicka A & Wolfenden R (1995) A proficient enzyme. *Science* 267, 90–93.
- Schramm VL (2011) Enzymatic transition states, transition-state analogs, dynamics, thermodynamics, and lifetimes. *Annu. Rev. Biochem.* 80, 703–732.
- Scott JD & Pawson T (2009) Cell signaling in space and time: where proteins come together and when they're apart. *Science* 326, 1220–1224.
- Taylor SS, Keshwani MM, Steichen JM & Kornev AP (2012) Evolution of the eukaryotic protein kinases as dynamic molecular switches. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2517–2528.
- Vale RD & Milligan RA (2000) The way things move: looking under the hood of molecular motor proteins. *Science* 288, 88–95.
- Wilson MZ & Gitai Z (2013) Beyond the cytoskeleton: mesoscale assemblies and their function in spatial organization. *Curr. Opin. Microbiol.* 16, 177–183.