MOLECULAR BIOLOGY

PRINCIPLES AND PRACTICE

SECOND EDITION

Michael M. Cox
Jennifer A. Doudna
Michael O'Donnell

Molecular Biology

This page intentionally left blank

Molecular Biology

Principles and Practice

Second Edition

Michael M. Cox
University of Wisconsin–Madison

Jennifer A. Doudna
University of California, Berkeley

Michael O'Donnell
The Rockefeller University

W. H. Freeman & Company
A Macmillan Education Imprint
New York

Publisher: Kate Ahr Parker
Senior Acquisitions Editor: Lauren Schultz
Developmental Editors: Anna Bristow, Erica Pantages Frost, Lisa Samols
Media Editors: Anna Bristow, Erica Pantages Frost
Assistant Editor: Tue Tran
Art Director: Diana Blume
Cover and text designer: Diana Andrews, DreamIt, Inc.
Senior Project Editor: Elizabeth Geller
Manuscript Editors: Linda Strange, Brook Soltvedt
Illustrations: H. Adam Steinberg, Dragonfly Media Group
Photo Editor: Jennifer Atkins
Photo Researcher: Teri Stratford
Illustration Coordinator: Janice Donnola
Production Coordinator: Paul Rohloff
Marketing Director: John Britch
Marketing Assistant: Bailey James
Composition: MPS Limited
Printing and Binding: Quad Versailles

Front cover image: Courtesy Illumina, Inc.

Throughout the text, a number of illustrations have been adapted from *Lehninger Principles of Biochemistry*, Sixth Edition, by David L. Nelson and Michael M. Cox.

Library of Congress Control Number: 2015930844

ISBN-13: 978-1-4641-2614-7
ISBN-10: 1-4641-2614-3

Printed in the United States of America

First Printing

W. H. Freeman and Company
41 Madison Avenue
New York, NY 10010
Houndmills, Basingstoke RG21 6XS, England

www.whfreeman.com

To our students, for the inspiration they provide every day,
and to our mentors, in gratitude for their guidance:

Tom Cech
Fred Grieman
Bill Jencks
Arthur Kornberg
Bob Lehman
Sharon Panasenko
David Sheppard
Jack Szostak
Hal White
Charles Williams

About the Authors

Courtesy Robin Davies/MediaLab-Biochemistry, UW-Madison

MICHAEL M. COX was born in Wilmington, Delaware. After graduating from the University of Delaware, he went to Brandeis University to do his doctoral work with William P. Jencks, and then to Stanford for postdoctoral study with I. Robert Lehman. He is currently Professor of Biochemistry at the University of Wisconsin–Madison. His research focuses on recombinational DNA repair processes. Cox has received awards for both teaching and research, including the 1989 Eli Lilly Award in Biological Chemistry from the American Chemical Society and two major teaching awards from the University of Wisconsin. He has coauthored five editions of *Lehninger Principles of Biochemistry.*

Photo by Sam Willard

JENNIFER A. DOUDNA grew up on the Big Island of Hawaii and became interested in chemistry and biochemistry in high school. She received her B.A. in biochemistry from Pomona College and her Ph.D. from Harvard University, working in the laboratory of Jack Szostak, with whom she also did postdoctoral research. She then went to the University of Colorado as a Lucille P. Markey scholar and postdoctoral fellow with Thomas Cech. Doudna is currently Professor of Molecular and Cell Biology and Professor of Chemistry at the University of California, Berkeley, and an Investigator of the Howard Hughes Medical Institute. She is a member of the National Academy of Sciences, the American Academy of Arts and Sciences, and the Institute of Medicine. She is also a Fellow of the American Association for the Advancement of Science.

Courtesy Rockefeller University

MICHAEL O'DONNELL grew up in a neighborhood on the banks of the Columbia River outside Vancouver, Washington. He had several inspirational teachers at Hudson Bay High School who led him into science. He received his B.A. in biochemistry from the University of Portland and his Ph.D. from the University of Michigan, where he worked under Charles Williams, Jr., on electron transfer in the flavoprotein thioredoxin reductase. He performed postdoctoral work on *E. coli* replication with Arthur Kornberg and then on herpes simplex virus replication with I. Robert Lehman in the Biochemistry Department at Stanford University. O'Donnell is currently Professor of Biochemistry and Structural Biology at The Rockefeller University and an Investigator of the Howard Hughes Medical Institute. He is a member of the National Academy of Sciences.

Contents in Brief

Contents

Preface

As teachers, we know that undergraduate science education is evolving. Simply conveying facts does not produce a scientifically literate student, a long-held perception now reinforced by numerous studies. Students of science need more: a better window on what science is and how it is done, a clear presentation of key concepts that rises above the recitation of details, an articulation of the philosophical underpinnings of the scientific discipline at hand, exercises that demand analysis of real data, and an appreciation for the contributions of science to the well-being of humans throughout the world. As undergraduate science educators rise to these challenges, we are faced with both higher numbers of students and declining resources. How can we all do more with less?

Textbooks are an important part of the equation. A good textbook must now be more than a guide to the information that defines a discipline. For instructors, a textbook must organize information, incorporate assessment tools, and provide resources to help bring a discipline to life. For students, a textbook must relate science to everyday experience, highlight the key concepts, and show each student the process that generated those key concepts.

This book had its genesis at a meeting of the authors in Napa Valley in January 2006. From the outset, we set ambitious goals designed to address the key challenges we face as teachers.

Students see science as a set of facts rather than an active human endeavor. Molecular biology has a wealth of important stories to tell. We wanted to convey the excitement that drives modern molecular biology, the creativity at the bench, and the genuine wonder that takes hold as the workings of a new biological process are revealed. This theme is set in the first chapter, dedicated in large measure to an introduction to the scientific process. Every chapter then begins with a *Moment of Discovery,* highlighting a researcher's own description of a memorable moment in his or her career. After Chapter 1, every chapter ends with a *How We Know* section, with stories relating the often circuitous path to a new insight. Additional anecdotes—scientists in action—are woven into the text and the accompanying *Highlights.* As students read the text, the laboratories and the people behind the discoveries will never be far away.


12 DNA Mutation and Repair

This second edition is an update, and much more. It has allowed us to refine the initial vision we had when we started this project and to augment that vision with unparalleled resources that will bring the subject to life for students and educators alike.

◀ MOMENT OF DISCOVERY

Scientific breakthroughs represent the exhilarating culmination of a lot of hard work. Each chapter opens with a description of a significant breakthrough in molecular biology, told by the scientist who made the discovery. The scientists featured in the Moments of Discovery are David Allis, Norm Arnheim, Bonnie Bassler, Steve Benner, James Berger, Carlos Bustamante,

Rose Byrne, Jamie Cate, Joe DeRisi, Roxana Georgescu, Lin He, Tracy Johnson, Melissa Jurica, Judith Kimble, Robert Lehman, Steve Mayo, Harry Noller, Smita Patel, Lorraine Symington, Jack Szostak, Robert Tjian, and Wei Yang.

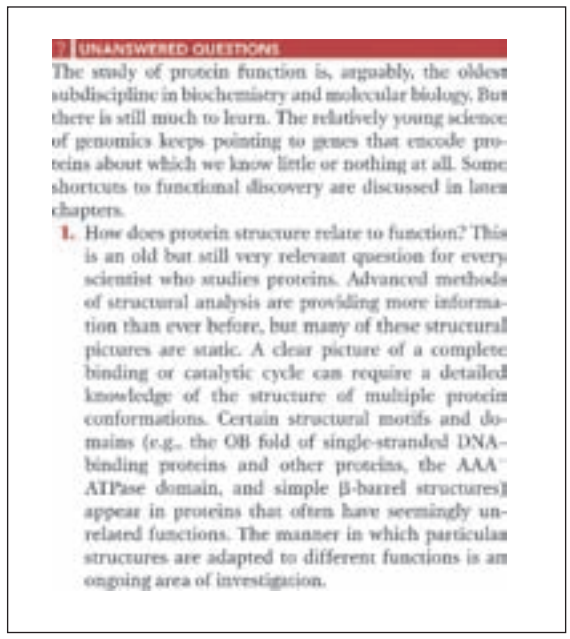
◀ HOW WE KNOW

Each chapter ends with a How We Know section that combines fascinating stories of research and researchers with experimental data for students to analyze.

Students often view science as a completed story. The reality is far different. Data can take a researcher in unexpected directions. An experiment designed to test one hypothesis can end up revealing something quite different. The analysis of real data is a fundamental skill to be honed by every student of science. We have tried to address this need aggressively. Each chapter in this text features a challenging set of problems, including at least one requiring the analysis of data from the scientific literature. Many of these are linked to the discoveries described in the How We Know sections. Each chapter also ends with some *Unanswered Questions*, providing just a sampling of the endless challenges that remain for those with the motivation to tackle them.

UNANSWERED QUESTIONS ▶

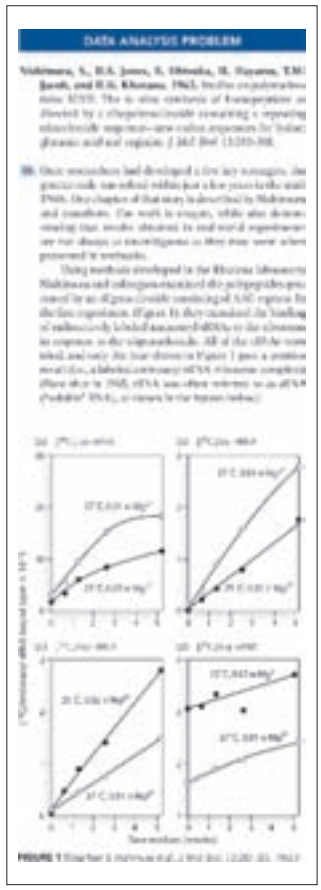
A short section at the end of each chapter describes important areas still open to discovery, showing students that even well-covered subjects, such as nucleic acid structure and DNA replication, are far from fully explored.



7 UNANSWERED QUESTIONS

The study of protein function is, arguably, the oldest subdiscipline in biochemistry and molecular biology. But there is still much to learn. The relatively young science of genomics keeps pointing to genes that encode proteins about which we know little or nothing at all. Some shortcuts to functional discovery are discussed in later chapters.

1. How does protein structure relate to function? This is an old but still very relevant question for every scientist who studies proteins. Advanced methods of structural analysis are providing more information than ever before, but many of these structural pictures are static. A clear picture of a complete binding or catalytic cycle can require a detailed knowledge of the structure of multiple protein conformations. Certain structural motifs and domains (*e.g.*, the OB fold of single-stranded DNA-binding proteins and other proteins, the AAA$^+$ ATPase domain, and simple β-barrel structures) appear in proteins that often have seemingly unrelated functions. The manner in which particular structures are adapted to different functions is an ongoing area of investigation.

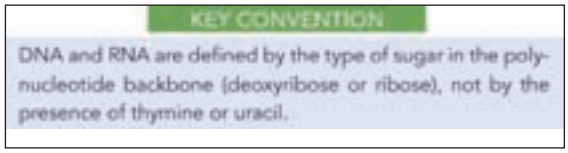


DATA ANALYSIS PROBLEM

◀ END-OF-CHAPTER PROBLEMS

Extensive problem sets at the end of each chapter give students the opportunity to think about and work with the chapter's key ideas. New problems have been added in each chapter for this second edition. Each problem set concludes with a Data Analysis Problem, giving students the critical experience of interpreting real research data. Solutions to all problems can be found at the back of the book.

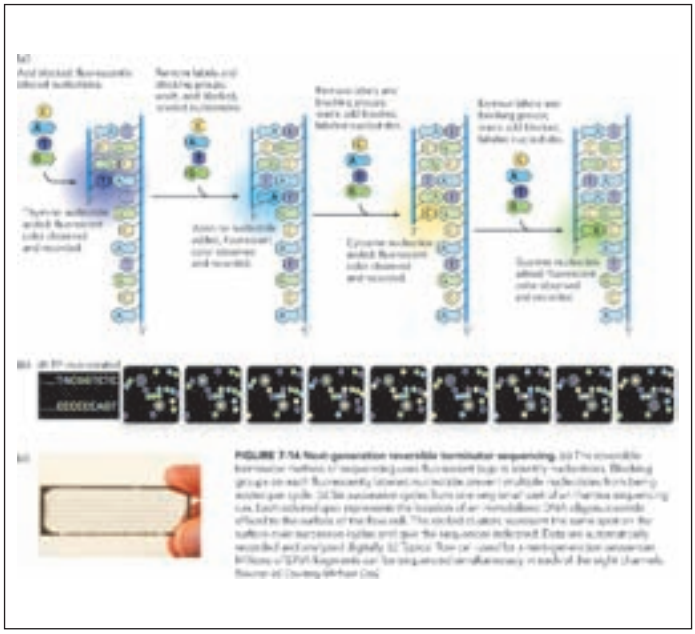
Students get lost in the details. Presenting the major concepts clearly, in the text as well as in the illustrations, is crucial to teaching students how science is done. We have worked to use straightforward language and a conversational writing style to draw students in to the material. We have collaborated closely with our illustrator, Adam Steinberg, to create clean, focused figures. Featured *Key Conventions* highlight the implicit but often unstated conventions used when sequences and structures are displayed and in naming biological molecules.


KEY CONVENTION

DNA and RNA are defined by the type of sugar in the polynucleotide backbone (deoxyribose or ribose), not by the presence of thymine or uracil.

◀ KEY CONVENTIONS

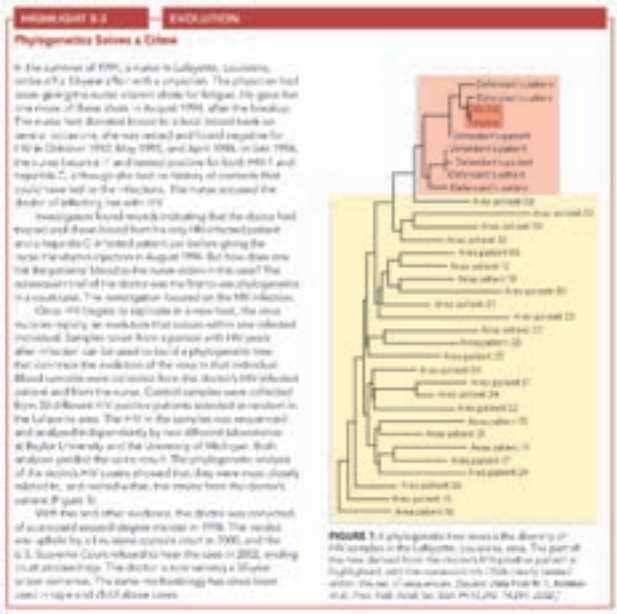
In brief paragraphs, the Key Conventions clearly lay out for students some fundamental principles often glossed over.



◀ ILLUSTRATIONS

Good figures should speak for themselves. We have worked to keep our figures simple and the figure legends as brief as possible. The illustrations in the text are the product of close collaboration with our colleague Adam Steinberg. Together with the talented artists at Dragonfly Media Group, Adam has helped to hone and implement our vision.

Students see evolution as an abstract theory. Every time a molecular biologist studies a developmental pathway in nematodes, identifies key parts of an enzyme active site by determining what parts are conserved among species, or searches for the gene underlying a human genetic disease, he or she is relying on evolutionary theory. *Evolution is a foundational concept, upon which every discipline in the biological sciences is built.* In this text, evolution is a theme that pervades every chapter, beginning with a major section in Chapter 1 and continuing as the topic of many Highlights and chapter segments.



◀ HIGHLIGHTS

These discussions are designed to enhance students' understanding and appreciation of the relevance of each chapter's material. There are four categories of Highlights:

- **Medicine** explores diseases that arise from defects in biochemical pathways, and how concepts uncovered in molecular biology have contributed to drug therapies and other treatments.
- **Technology** focuses on cutting-edge molecular biology methods.
- **Evolution** reveals the role of molecular biology research in understanding key biological processes and the connections among organisms.
- **A Closer Look** examines a wide variety of additional, intriguing topics.

EXPERIMENTAL TECHNIQUES

As researchers, we know that it is critical to understand the benefits and limitations of experimental techniques. We strive to give students a sense of how an experiment is designed and what makes the use of a particular technique or model organism appropriate. The techniques covered in this book are:

NEW AND UPDATED CONTENT

The second edition addresses recent discoveries and advances, corresponding to our ever-changing understanding of molecular biology. In addition to the text updates listed here, there are numerous new figures and photos, along with significantly updated figures in every chapter. There are also new end-of-chapter problems for every chapter and many new Unanswered Questions.

Chapter 1

- Updated discussions of evolution and the scientific method

Chapter 2
- Updated discussion of the central dogma
- Updated and expanded discussion of the types of RNA

Chapter 3
- New Moment of Discovery
- Expanded discussion of nucleosides
- Revised and expanded section: The Hydrophobic Effect Brings Together Nonpolar Molecules
- New section: Electronic Interactions between Bases in Nucleic Acids

Chapter 4
- Expanded section: Amino Acids Are Categorized by Chemical Properties
- Significantly expanded discussion of protein purification, including Highlight 4-1
- New section: Intrinsically Unstructured Proteins Have Versatile Binding Properties
- Expanded section on protein families
- Significantly expanded section on protein folding and computational biology

Chapter 5
- New Moment of Discovery

Chapter 6
- Expanded discussion of the instability of RNA
- New Highlight 6-1: DNA Nanotechnology
- New discussion of riboswitches

Chapter 7
- Expanded discussion on obtaining DNA fragments to clone
- Thoroughly updated section on next-gen and other modern DNA sequencing technologies.
- New section: Genomic Sequencing Is Aided by New Generations of DNA Sequencing Methods, incorporating the exciting new advances with programmable nucleases

Chapter 8
- Expanded Highlight 8-1, now including discussion of the microbiome
- Updated section on noncoding DNA
- Expanded section on mass spectrometry

Chapter 10
- New Moment of Discovery
- Significantly expanded discussion of histone modifications, including a new table

Chapter 11
- Expanded discussion of the β sliding clamp
- Expanded discussion of the Pol III holoenzyme
- Updated and expanded discussion of eukaryotic replication forks
- Updated and expanded section: Eukaryotic Origins “Fire” Only Once per Cell Cycle
- New section: Telomeres and Telomerase Solve the End Replication Problem in Eukaryotes
- New Highlight 11-2: Short Telomeres Portend Aging Diseases

Chapter 12
- New Moment of Discovery
- New table presenting overview of DNA repair processes

Chapter 13
- Updated and expanded sections on double-strand break repair and reconstruction of replication forks
- Updated section on meiotic recombination

Chapter 14
- Updated and expanded introductory section on transposable elements and site-specific recombination
- Updated and expanded section: Precise DNA Rearrangements Are Promoted by Site-Specific Recombinases
- Reorganized section on the use of site-specific recombination systems in biotechnology
- Updated and expanded sections on transposition

Chapter 15
- Updated section on transcription elongation
- Updated and expanded discussion of the role of transcription factors
- Updated and expanded discussion of termination mechanisms among RNA polymerases

Chapter 16
- Streamlined chapter organization
- Expanded discussion of P bodies

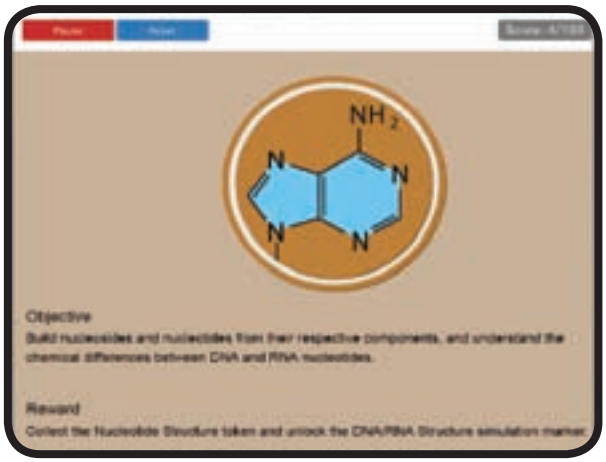
Chapter 18
- Streamlined chapter organization
- Updated discussion of protein release factors
- Updated discussion of nuclear export signals

Chapter 19
- Updated section: Gene Expression Is Regulated through Feedback Loops, now including inducer exclusion

Chapter 22
- Expanded section on alternative splicing, including ESEs and ESSs
- Updated section on RNA interference
- New section: RNAs Regulate a Wide Range of Cellular Processes
- Updated section on the developmental potential of stem cells

MEDIA

Simulations
One of our central goals in tackling the revision of this textbook was to provide special resources to engage students (and educators) in molecular biology. New to the second edition are simulations that cover core molecular biology concepts and techniques. Created using the art from the text, the simulations reinforce students' understanding by allowing them to interact with the structures and processes they have encountered. A game-like format guides students through the simulations, unlocking them in order, and multiple-choice questions after each simulation ensure that instructors can assess whether students have thoroughly understood each topic. These simulations are the product of many days of meetings among the authors, editors, and media developers. From storyboarding to the finished product, these simulations were one of the most challenging as well as stimulating efforts associated with preparing the second edition. We are excited to present this new approach to learning key concepts.



Nucleotide Structure (Chapter 3)
DNA/RNA Structure (Chapter 6)
PCR (Chapter 7)
Sanger Sequencing (Chapter 7)
CRISPR (Chapters 7 and 19)
DNA Replication (Chapter 11)
DNA Polymerase (Chapter 11)
Mutation and Repair (Chapter 12)
Transcription (Chapter 15)
mRNA Processing (Chapter 16)
Translation (Chapter 18)

Nature Articles with Assessment

These articles engage students in reading about primary research and encourage critical thinking. Specifically selected for both alignment with the text coverage and exploration of identified difficult topics, the *Nature* articles include assessment questions that can be automatically graded. Also included are open-ended questions that are suitable for use in flipped classrooms and active learning discussions either in class or online.

The simulations and *Nature* articles for *Molecular Biology: Principles and Practice* are available in our LaunchPad system, along with many additional resources.

Macmillan Education LaunchPad

This dynamic, fully integrated learning environment brings together all of our teaching and learning resources in one place. It also contains the fully interactive **e-Book** and other newly updated resources for students and instructors, including the following:

New Clicker Questions allow instructors to integrate active learning in the classroom and to assess students' understanding of key concepts during lectures.

Updated Test Bank contains at least 40 multiple-choice and short-answer questions for each chapter.

LearningCurve allows students to test their comprehension of the chapter concepts. The system adapts to students' individual level of preparedness by giving them questions at varying levels of difficulty, depending on whether they answer a question without help, or they need help but eventually get the question right, or they are unable to answer the question. Links to the appropriate e-Book section, hints, and feedback help students realize where they need more practice on a topic.

Key Term Flashcards allow students to review the definitions of all the glossary terms and quiz themselves.

Textbook Images and Tables are offered as high-resolution JPEG files. Each image has been fully optimized to increase type size and adjust color saturation. These images have been tested in a large lecture hall to ensure maximum clarity and visibility.

Acknowledgments

This text represents our best effort to synthesize a complex and ever-shifting field and to contribute to the broadening requirements of twenty-first-century education in molecular biology. We welcome your comments and suggestions. We thank our many colleagues whose input has helped shape this book:

Steven Ackerman, *University of Massachusetts*
Ravi Allada, *Northwestern University*
Rick Amasino, *University of Wisconsin–Madison*
Andrew Andres, *University of Nevada, Las Vegas*
Brian Ashburner, *University of Toledo*
Matthew Bahamonde, *Farmingdale State College*
Kenneth Belanger, *Colgate University*
Joel Belasco, *NYU School of Medicine*
Morgan Benowitz-Fredericks, *Bucknell University*
Bradford Berges, *Brigham Young University*
Xin Bi, *University of Rochester*
Robert Borgon, *University of Central Florida*
David Bourgaize, *Whittier College*
Nicole Bournias-Vardiabasis, *California State University, San Bernardino*
John Boyle, *Mississippi State University*
Jeremy Bruenn, *University at Buffalo*
Douglas Burks, *Wilmington College*
Aaron Cassill, *University of Texas, San Antonio*
Karl Chai, *University of Central Florida*
Davis Cheng, *California State University, Fresno*
William Cody, *University of Dallas*
Mary Connell, *Appalachian State University*
Scott Covey, *University of British Columbia*
Fred Cross, *Rockefeller University*
Cristina Cummings, *Providence College*
Rodney Dale, *Loyola University Chicago*
Susan DiBartolomeis, *Millersville University*
Dessislava Dimova, *Rutgers University, New Brunswick*
David Donze, *Louisiana State University, Baton Rouge*
Arri Eisen, *Emory University*
Danielle Ellis, *California State University, Pomona*
R. Paul Evans, *Brigham Young University*
Nicholas Ewing, *California State University, Sacramento*
Jason Fitzgerald, *Southeastern Illinois College*
Gerald Frenkel, *Rutgers University, Newark*
Louise Glass, *University of California, Berkeley*
Ann Grens, *Indiana University South Bend*
Theresa Grove, *Valdosta State University*
Nancy Guild, *University of Colorado Boulder*
Immo Hansen, *New Mexico State University*
Daniel Herman, *University of Wisconsin–Eau Claire*
Margaret Hollingsworth, *University at Buffalo*
Stan Ivey, *Delaware State University*
Russell Johnson, *Colby College*
Jason Kahn, *University of Maryland*
Mijung Kim, *Chicago State University*
Timothy Lane, *University of California, Los Angeles*
Curtis Loer, *University of San Diego*
Charles Mallery, *University of Miami, College of Arts and Sciences*
Kathryn McMenimen, *Mount Holyoke College*
Mitch McVey, *Tufts University*
Thomas Mennella, *Bay Path College*
Karl Miletti, *Delaware State University*
Yuko Miyamoto, *Elon University*
Evangelos Moudrianakis, *Johns Hopkins University*
Arunachalam Muthaiyan, *Chicago State University*
Hao Nguyen, *California State University, Sacramento*
Brent Nielsen, *Brigham Young University*
James Ntambi, *University of Wisconsin–Madison*
Greg Odorizzi, *University of Colorado*
James Olesen, *Ball State University*
Harold Olivey, *Indiana University Northwest*
Anthony Otsuka, *Illinois State University*
Rekha Patel, *University of South Carolina*
Bruce Patterson, *University of Arizona, Tucson*
Brian Poole, *Brigham Young University*
Megan Porter, *University of South Dakota*
Ted Powers, *University of California, Davis*
April Pyle, *University of California, Los Angeles*
Brian Ring, *Valdosta State University*
Herve Roy, *University of Central Florida*
Edmund Rucker, *University of Kentucky*
Ivan Sadowski, *University of British Columbia*
Steven Sandler, *University of Massachusetts Amherst*
Brian Sato, *University of California, Irvine*
Mary Schuler, *University of Illinois Urbana-Champaign*
William Scovell, *Bowling Green State University*
Andrei Seluanov, *University of Rochester*
Konstantin Severinov, *Rutgers University, New Brunswick*

Xueyan Shan, *Mississippi State University*
Elaine Sia, *University of Rochester*
Ron Siu, *University of California, Los Angeles, Extension*
Agnes Southgate, *College of Charleston*
Daniel Stoebel, *Harvey Mudd College*
Derek Tan, *Sloan-Kettering Institute for Cancer Research*
Ignatius Tan, *New York University*
Lloyd Turniten, *University of Wisconsin–Eau Claire*
Jill Wildonger, *University of Wisconsin–Madison*
Bruce Wolff, *University of Waterloo*
Michael Yu, *University at Buffalo*

This book would not have been possible without the support of our publishers at W. H. Freeman. A book of this sort is an undertaking measured not just in hours but in sleepless nights, almost-met deadlines, conference calls, and occasional levity. It is an enterprise in which teachers sometimes become students. The needed guidance has been provided by an exceptionally talented team of editors and copy editors. Kate Ahr Parker has overseen the effort from the beginning. Few human beings are as gifted in the art of articulating urgency with grace. Anna Bristow, Erica Frost, and Lisa Samols have been our development editors. Guided by their capable hands, first-draft chapters have been created, reworked, broken up, and sometimes merged. They provided encouragement and pointed out deficiencies. They have been our partners throughout, scrutinizing every word we produced. As the project progressed, the work was honed and the chapters integrated with the help of Brook Soltvedt and Linda Strange. Both Brook and Linda are longtime veterans of the *Lehninger Biochemistry* series, and their expertise added immeasurably to the final product in your hands. In the end, they have managed the impressive feat of merging three voices into one. We are extraordinarily grateful to all of the editors for their dedication to this project. We are fortunate to have had the benefit of their insights and expertise.

The artwork for this book was a labor of love handled by Adam Steinberg and the artists at Dragonfly Media. Adam is also a *Lehninger Principles of Biochemistry* veteran, and his experience and skill are evident on almost every page of this book. He worked hand in hand with the authors to create illustrations that convey concepts concisely and in a unified style.

Our thanks also go to the consummate professionals who ensured the high quality in the production of the book: art director Diana Blume, project editor Elizabeth Geller, production coordinator Paul Rohloff, illustration coordinator Janice Donnola, photo editor Jennifer Atkins, and photo researcher Teri Stratford. We greatly appreciated their flexibility and creativity in working with complex material and ever-shifting schedules.

We express our appreciation to our colleagues, friends, and families for their patience and support.

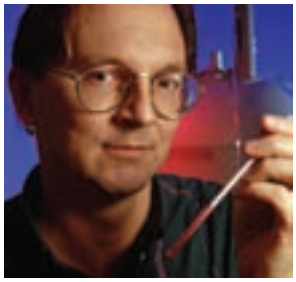
Last, but certainly not least, we are grateful to the Moment of Discovery authors, who shared some of their favorite scientific career moments with us. Each of them provided valuable time and effort for this project and helped us add a personal touch to every chapter.

Michael M. Cox
Jennifer A. Doudna
Michael O'Donnell
December 2014

Courtesy Mike Cox

Mike Cox, Jennifer Doudna, and Mike O'Donnell

1 Evolution, Science, and Molecular Biology



Jack Szostak *[Source: © Jim Sugar/Corbis.]*

MOMENT OF DISCOVERY

A big question in the origin of life concerns how primitive cells might have evolved. My own approach to this question involved lots of discussions with Irene Chen and others in my lab about how lipid vesicles containing RNA, which might mimic a simple self-replicating life form, could be capable of dividing. In other words, as the amount of genetic material (here, RNA) increased by making more copies of itself, *how would the increased RNA content affect the physical properties of the vesicle?* We envisioned that osmotic pressure might make vesicles grow by extracting lipids from neighboring vesicles, ultimately leading to division by rupture and resealing. This idea seemed pretty far out, though, until Irene began doing experiments with vesicles containing lipids bearing fluorescent dyes. We could encapsulate RNA inside the vesicles and watch the vesicles change in size (or not) under different conditions by following the level of fluorescence as a function of vesicle surface area. Irene found that empty vesicles or vesicles "swollen" with RNA were stable over time, but when she mixed them together, the swollen vesicles started to grow by stealing lipid molecules from neighboring empty vesicles! So the system worked exactly as we had imagined, demonstrating that vesicle growth and division is a process that can occur spontaneously.

More recently, we found that vesicles loaded with RNA can also take up nucleotides (the building blocks of RNA and DNA) from the environment, disproving an old idea that it would be hard for primitive cells to survive by taking up small molecules, including negatively charged nucleotides, from their surroundings. It has been very exciting to find that each potential roadblock to primitive cellular replication that we have explored so far can be overcome, often without requiring specialized catalysts or input energy.

—Jack Szostak, on his discovery of self-dividing vesicles that mimic growing cells

Online resources related to this chapter:

***Nature* exercise** Genome dynamics during experimental evolution

Born in the second half of the twentieth century, molecular biology has only recently come of age. Broadly speaking, **molecular biology** is the study of essential cellular macromolecules, including DNA, RNA, and proteins, and the biological pathways between them. Over the decades, molecular biology has become firmly associated with the structure, function, and regulation of information pathways at the molecular level. All of the processes required to reliably pass genetic information from one generation to another and from DNA to RNA to protein are included in this area of study. Of the requirements for life, it is the information in our genetic material that links all organisms to each other and documents their intertwined history. The biological information pathways that maintain, use, and transmit that information are the focus of this book.

Molecular biology may have a relatively short history, but its impact on the human experience is already considerable. Medicine, modern agriculture, forensic science, and many other endeavors rely on technologies developed by molecular biologists. Our current understanding of information pathways has given rise to diagnostic tests for genetic diseases, forensic DNA analysis, crops with improved yields and resistance to disease, new cancer therapies, an unprecedented ability to track pandemics, new wastewater treatment methods, new approaches to the generation of energy, and much more. Many of these advances are chronicled throughout this textbook.

This first chapter introduces three of the most important themes that link the book's topics. The first theme concerns the two key requirements for life: **biological information**, the genetic instructions that shape every living cell and virus, and **catalysis**, a capacity to accelerate critical molecular processes. Molecular biology deals with both, and much of the discipline focuses on the interplay between information-containing polymers (nucleic acids and proteins) and the enzymes that catalyze and regulate their synthesis, modification, function, and degradation.

The second theme is **evolution**. Many of the processes we will consider can be traced back billions of years, and a few can be traced to the last universal common ancestor. Genetic information is a kind of molecular clock that can help define ancestral relationships among species. Shared information pathways connect humans to every other living organism on Earth and to all the organisms that came before.

The third theme in this book is how we look at molecular biology as a scientific endeavor. Any scientific discipline is a construct not only of the knowledge it has generated but also of the human processes behind that knowledge. Molecular biology has both an inspirational history and a promising future, to be forged by contributors as yet unnamed. Breakthroughs rely on more than technology and ideas: they require an understanding of the scientific process and are informed by the struggles of the past.

1.1 THE EVOLUTION OF LIFE ON EARTH

All organisms on Earth are connected by an evolutionary journey spanning more than 3 billion years. The diversity of life we see around us is the sum of a limitless number of **mutations**, changes in genetic information that are usually subtle but sometimes dramatic. When Charles Darwin proposed that natural selection acts on variation in populations, he had no knowledge of the mechanisms that give rise to that variation. Such mechanisms lie at the heart of modern molecular biology.

What Is Life?

Almost anyone can distinguish a living organism from an inanimate object. However, a rigorous scientific description of life is harder to achieve. Life differs from nonlife in identifiable ways, as summarized in **Figure 1-1**. Organisms move, reproduce, grow, and alter their environment in ways that inanimate objects cannot. But such characteristics alone provide an unsatisfying definition of life, particularly when a few of them may be shared by inanimate substances. In 1994, the United States National Aeronautics and Space Administration (NASA) convened a panel to consider the question, "What is life?" A simple definition resulted: *Life is a chemical system capable of Darwinian evolution*. The importance of evolutionary theory to all biological sciences gains full expression in this concise statement.

Every living system we know about has several requirements for its existence. Two of these—raw materials and energy—are supplied by a home planet endowed with an abundance of both. Molecules in Earth's life forms are made up largely of the elements hydrogen, oxygen, nitrogen, and carbon. These are the smallest and most abundant atoms that can make, respectively, one, two, three, and four covalent bonds with other atoms. The molecules formed by these elements tend to be quite stable and can be very complex. The energy required for life is derived from the sun. Plants and photosynthetic microorganisms collect and store the energy derived from sunlight in the chemical bonds of complex biomolecules.

A third requirement for a living system is an envelope, creating a barrier between the living and inanimate worlds and establishing a means of selective interaction between a cell and its environment. The work of Jack Szostak, chronicled in this chapter's Moment of Discovery, may be replicating some key evolutionary moments that led to enveloped living systems (**Figure 1-2** on p. 4).

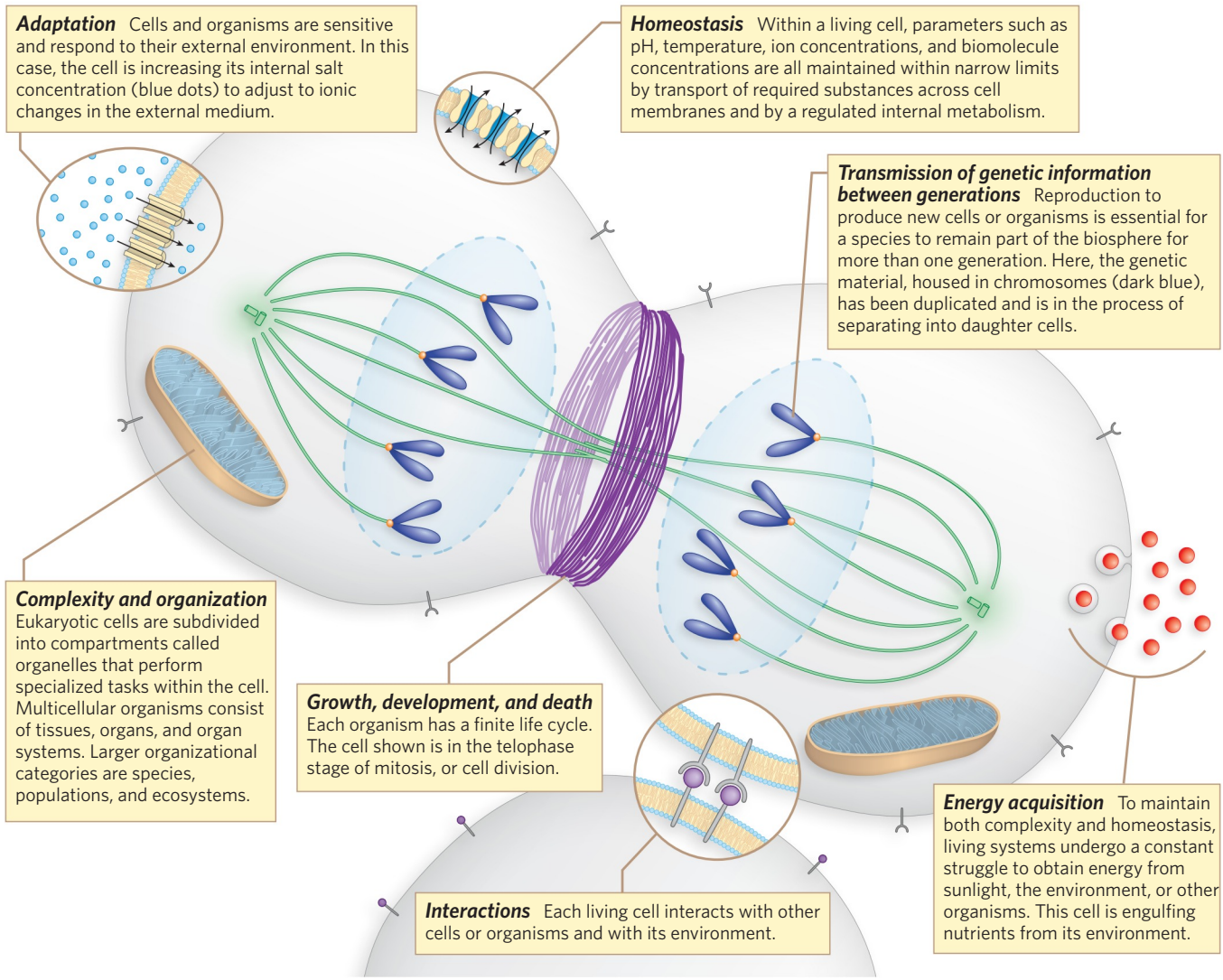


FIGURE 1-1 Characteristics of living systems. Each characteristic distinguishes living organisms from inanimate matter.

The final two requirements—catalysis and biological information—are particularly important, truly distinguishing a living organism from an inanimate object. These requirements are the domain of molecular biology. The energy transactions that support homeostasis (the maintenance of parameters such as pH and biomolecule concentrations within the narrow range needed to support life) and enable the transmission of genetic information from one generation to the next are initiated by powerful catalysts called **enzymes**. Enzymes are highly specific, and each enzyme accelerates only one or a small number of chemical reactions. Most enzymes are proteins, although a few catalytic RNA molecules play important roles in cells. The catalysts that a particular organism possesses define which reactions can occur in that organism. Enzymes determine what a cell takes in for nourishment, how fast the cell grows, how it discards wastes, how it constructs its cellular membranes, how it responds to other cells, and how it reproduces.

The presence of enzymes in a cell depends on the faithful transmission of the genetic information that encodes them from one generation to the next. Enzymes, as well as the myriad other proteins and RNA molecules that regulate their synthesis and function, are the actual molecular targets of evolution. When a cell acquires a new function, it generally reflects the presence of a new enzyme or set of enzymes, or an alteration in the regulation or function of an existing enzyme or process. The new functions arise through changes in genes—changes that are shaped by evolutionary processes. In the biosphere of today, DNA is the standard macromolecule for the long-term storage and transmission of biological information. It is exquisitely adapted to that function (**Figure 1-3** on p. 5). However, as we shall see, there were probably stages in the evolution of life when DNA did not serve as the primary genetic library in living systems.

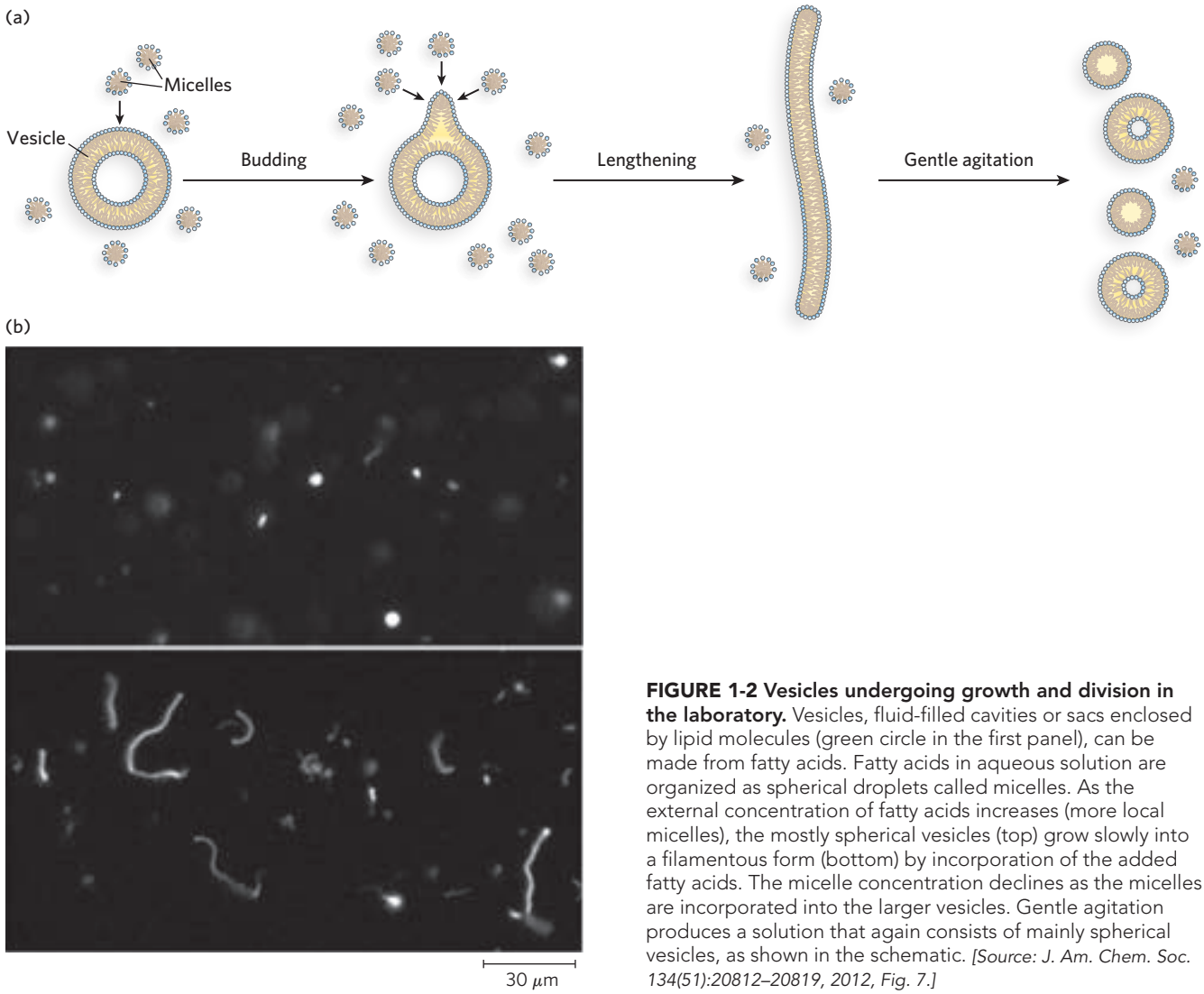


FIGURE 1-2 Vesicles undergoing growth and division in the laboratory. Vesicles, fluid-filled cavities or sacs enclosed by lipid molecules (green circle in the first panel), can be made from fatty acids. Fatty acids in aqueous solution are organized as spherical droplets called micelles. As the external concentration of fatty acids increases (more local micelles), the mostly spherical vesicles (top) grow slowly into a filamentous form (bottom) by incorporation of the added fatty acids. The micelle concentration declines as the micelles are incorporated into the larger vesicles. Gentle agitation produces a solution that again consists of mainly spherical vesicles, as shown in the schematic. *[Source: J. Am. Chem. Soc. 134(51):20812–20819, 2012, Fig. 7.]*

Evolution Underpins Molecular Biology

In 1973, the geneticist Theodosius Dobzhansky published an article in the professional journal *The American Biology Teacher* entitled "Nothing in Biology Makes Sense Except in the Light of Evolution." This sentiment has special meaning in molecular biology, because the pathways and processes in living systems give rise to the genetic variation on which natural selection acts (**Figure 1-4**). They also inform the ongoing investigations into how life arose on Earth.

Evolution relies on spontaneous and generally random changes in an organism's genomic material, called mutations. In spite of the elaborate cellular mechanisms we consider in this book, all of which help ensure accurate transmission of genetic information from one generation to the next, mutations regularly occur. Mutations can be as simple as a change in a single base pair of DNA or base of RNA or as substantial as the inversion, deletion, or insertion of large segments of genetic material. As we will be discussing in detail, errors can arise during replication (Chapter 11), and DNA damage can lead to permanent mutation when repair systems (Chapter 12) go awry. Larger chromosomal changes can arise from recombination (Chapter 13) or transposition (Chapter 14). Some mutations affect genes directly; others affect the ways in which DNA is transcribed into RNA or RNA is processed or translated (Chapters 15–18). Relatively minor changes in genes involved in regulatory processes (Chapters 19–22) can give rise to dramatic changes in the organism; this realization has created a new field, essentially a modern merger of the fields of evolutionary and developmental biology, dubbed "evo-devo" (described in Chapter 22). All the processes that contribute to information transfer are highly, but not perfectly, accurate, and the slow accumulation of alterations is inevitable. Many organisms even have mechanisms to speed up the pace of mutational change, which they draw upon in times of stress.

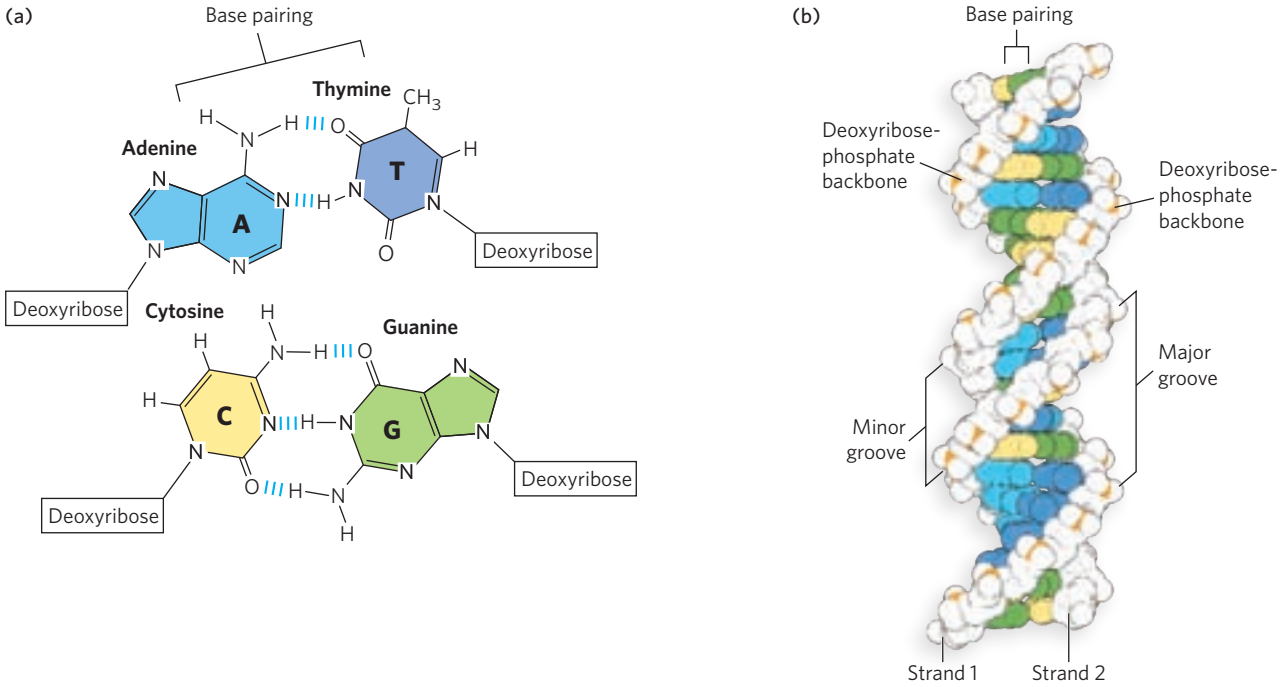


FIGURE 1-3 DNA structure. Because of its structural properties, DNA is well suited for long-term information storage. Genomic DNA almost always consists of two complementary strands of deoxyribonucleic acid. Each strand has a backbone consisting of deoxyribose residues connected by phosphate groups, and a base is attached to each ribose. Strand complementarity is enforced by specific interactions between the bases in each strand. The interactions create base pairs. (a) The G ≡ C and A = T base pairs are similarly sized, giving the DNA double helix a uniform width and allowing base pairs, in any sequence, to stack. Complementary base pairing facilitates replication and transmission from one generation to the next. (b) The double-helical structure and base stacking confer stability. Major and minor helical grooves in the structure provide access to genetic information for a wide range of DNA-binding proteins. The uniform structure of the DNA backbone allows the synthesis of very long polymers.

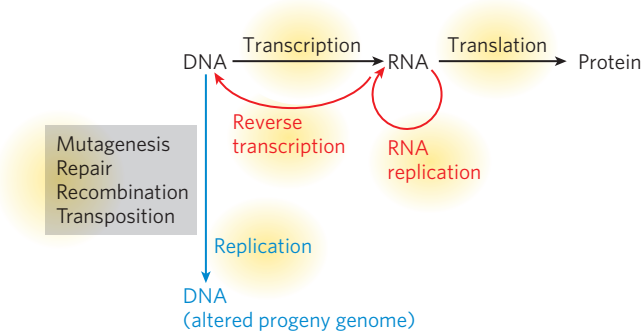


FIGURE 1-4 Pathways of biological information flow. In almost all living systems, information is stored in DNA, then transcribed into RNA, which is processed and translated into protein. DNA is replicated to prepare for cell division. The transfer and maintenance of genetic information are regulated at each of these stages. Exceptions to this general pattern are found in certain viruses (RNA viruses and retroviruses) that store their genetic information in RNA. Viruses with RNA genomes make use of additional pathways (denoted by the red arrows)—RNA replication and reverse transcription (creation of DNA from RNA, instead of the other way around)—to maintain their genomes. The yellow highlighting represents points of regulation. Processes in the gray shaded box, along with occasional errors in replication, reverse transcription, and RNA replication, give rise to genomic alterations (mutations) that fuel evolution.

An understanding of these processes has also given us insights into the origins of life and the process of evolution. Continuing explorations of RNA structure (Chapter 6) and metabolism (Chapters 15 and 16) have informed new theories of prebiotic evolution. The genetic code (Chapter 17) provides a particularly vivid look at the shared history of every organism on Earth.

Molecular biology has provided the enzymes that make most of the methods of biotechnology possible (Chapter 7). These increasingly powerful methods for studying the genes of many different organisms allow us to trace their evolution. Through modern genomics (Chapter 8), molecular biology is opening a window onto evolution that Charles Darwin would marvel at.

The interrelationship of molecular biology and evolution is of more than academic interest. Human beings exist in a world where every organism continues to evolve. Microorganisms, with their short life cycles, evolve most rapidly (**Highlight 1-1**). Of special concern are human pathogens, as well as the microorganisms, fungi, insects, and other organisms that affect our food crops, livestock, and water supply. Molecular biology

HIGHLIGHT 1-1 **EVOLUTION**

Observing Evolution in the Laboratory

The bacterium *Deinococcus radiodurans* has a remarkable capacity to survive the effects of ionizing radiation (IR, or gamma rays). A human being would be killed by exposure to 2 Gy (1 Gy (gray) = 100 rads) of IR, but cultures of *Deinococcus* routinely survive 5,000 Gy with no lethality. *Deinococcus* is a desert dweller, and this characteristic reflects its adaptation to the effects of desiccation. In dry conditions, the bacterium cannot grow and its cellular metabolism shuts down. Spontaneous damage to the cellular DNA accumulates, including strand breaks. DNA repair processes, which require ATP generated by cellular metabolism, do not take place. However, the bacterium can repair its genome quickly when conditions favorable for growth return. Like desiccation, IR also generates numerous DNA strand breaks, and that same extraordinary capacity for DNA repair is put to use after exposure to IR.

How long does it take for a bacterium to evolve extreme resistance to IR? A recent study demonstrated that *Escherichia coli*, the common laboratory bacterium, can acquire this resistance by directed evolution. Twenty cycles of exposure to enough IR to kill more than 99% of the cells, with each cycle followed by the outgrowth of survivors, produced an *E. coli* population with a radiation resistance approaching that of *Deinococcus*. The entire selection process can be achieved in less than a month. Complete genomic sequencing of cells isolated from the evolved populations typically reveals 40 to 80 mutations. The answer to survival varies from cell to cell, with different cells displaying different arrays of mutations, even when they come from the same evolved population. In just a single, small bacterial culture, evolution can take many paths, and a variety of solutions are found that lead to acquisition of a new trait.

This is just one of many experiments demonstrating that dramatic changes in microorganisms can be readily generated and observed in the laboratory within short periods of time. The same kind of evolutionary processes are occurring constantly in microorganisms in our environment, including human pathogens. When AIDS appeared as a new threat to human health in the early 1980s, the power of evolutionary theory was quickly on display. The causative agent, HIV, was soon isolated and its genomic sequence determined. Characterizing this novel and very dangerous virus from scratch would have delayed treatments for years. But scientists had a shortcut at hand. A deep reservoir of information about viruses and their evolutionary relationships had already been built up over decades of research. The small HIV genome thus held all the clues that science needed for a rapid understanding of its infection cycle and the development of a medical response. Its genome revealed that HIV is a type of RNA virus called a retrovirus, with clear evolutionary relationships to other viruses that were already known and understood (Figure 1). It was immediately evident which HIV genes encode the enzymes essential to the virus life cycle, and these enzymes rapidly became drug targets. One result was the development of highly effective treatments at an unprecedented rate, ranging from AZT to protease inhibitors (see Highlights 5-2 and 14-3 for more detailed descriptions of the retrovirus life cycle). Millions of lives have been saved, in large measure because all biological and medical research is carried out in the context of evolutionary theory.

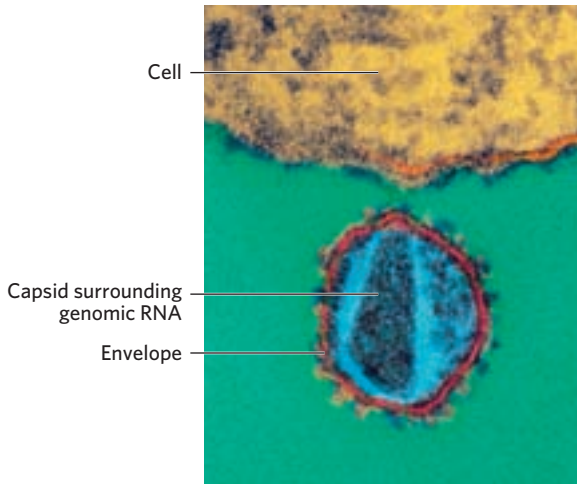


FIGURE 1 HIV is a retrovirus. Like other retroviruses, it has an RNA genome condensed within a proteinaceous capsid. The capsid is surrounded by a spherical lipid envelope derived from its host cell's cytoplasmic (plasma) membrane. Its relationship to other retroviruses is not just structural but embedded in definable ways in its chromosome. *[Source: Hans Gelderblom / Getty Images.]*

provides essential tools for use in tracking pandemics, investigating new microbial pathogens, identifying the genes underlying human genetic diseases, solving crimes, tracing the origin of diseases, treating cancer, and engineering microorganisms for new purposes in bioremediation and bioenergy. All of these efforts rely heavily on the concepts of evolutionary biology. Indeed, modern society relies on countless innovations in medicine and agriculture that would not exist but for Darwin's great insight.

Life on Earth Probably Began with RNA

About 4.6 billion years ago, the sun and Earth and the other planets and asteroids of our solar system were formed. Within the first billion years of our planet's existence, life appeared on its surface. How did this happen, and how likely is it that this has happened on other, similar worlds? Modern geologists, paleontologists, and molecular biologists are slowly piecing together the history of life on Earth from the rich trove of clues in the

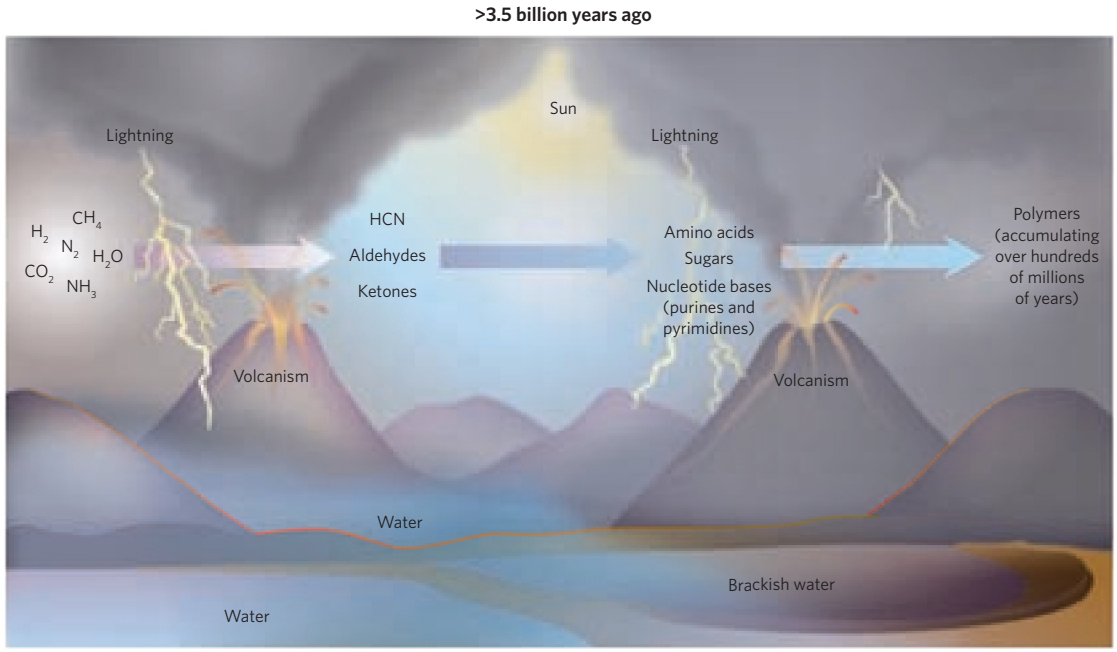


FIGURE 1-5 Prebiotic chemistry. Over hundreds of millions of years, and with constant energy input from solar radiation, volcanism, and other sources, the molecular constituents of Earth's early atmosphere were converted from simple molecules such as water, methane, ammonia, hydrogen, nitrogen, and carbon dioxide into a range of more complex organic molecules and polymers. The resulting tarry substance may have coated the planet's surface and turned bodies of water into concentrated and complex solutions.

geologic, fossil, and genomic records. A plausible sequence emerges, providing a wide range of hypotheses that can be tested using modern chemical and physical methods.

The first few hundred million years were a time of prebiotic chemistry (**Figure 1-5**). No life was present, but chemical reactions were happening everywhere. The atmosphere contained primarily water, methane, ammonia, hydrogen, nitrogen, and carbon dioxide. Reactions driven by the constant stream of energy coming from the sun were slowly yielding more complex molecules such as simple sugars, amino acids, and nucleotide bases. And the accumulation of organic material was supplemented by materials from a multitude of collisions between early Earth and meteors laden with organic matter. Prebiotic chemistry is being studied by a large community of researchers. A small sampling of their work is presented in the How We Know section at the end of this chapter.

Over a period of millions of years, the accumulation of reaction products yielded a soup containing molecules and polymers. As they grew increasingly complex, particular polymers acquired the capacity to duplicate themselves. The first self-replicating polymer, possessing two of the key requirements for life—catalysis and biological information—might be considered the first life form.

We do not know what this first "living" polymer was. However, modern molecular biology has given us many reasons to think that RNA either was the first self-replicator or arose as a much-improved descendant of that first self-replicator. RNA differs from DNA only in that it uses ribose instead of deoxyribose in its backbone. That single additional hydroxyl group in each monomeric unit of the polymer allows RNA to take up a plethora of complex structures that are inaccessible to DNA. The structural malleability of RNA gives it a capacity for both catalysis and information storage that has made it indispensable for life, from its beginnings to the present time.

The **RNA world hypothesis** was first proposed as a stage in evolution by molecular biologists Carl Woese, Francis Crick, and Leslie Orgel, in separate papers published in the late 1960s. The hypothesis describes a living system (or set of living systems) based on RNA. In this system, a variety of RNA enzymes could catalyze all of the reactions needed to synthesize the molecules required for life from simpler molecules available in the environment. The RNA enzymes would include replicators to duplicate all of the RNA catalysts. The "RNA organism," out of equilibrium with its surroundings, would have to be defined by a boundary. The experiments of Szostak and colleagues show one way in which lipid-enclosed RNA systems can arise (see the How We Know section at the end of this chapter).

Four more-recent lines of evidence have added much breadth and depth to the RNA world proposal. The first was the discovery by Thomas Cech and Sidney Altman, in the early 1980s, of **catalytic RNAs**, or **ribozymes**—enzymes that are made of RNA instead of protein. Thus

we learned that some extant RNA molecules catalyze reactions and so possess both of the key conditions for life—biological information and catalysis. In modern organisms, ribozymes catalyze a relatively narrow range of reactions, such as the cleavage and ligation of other RNA molecules—a range insufficient to support an RNA world.

What is the real catalytic potential of RNA? The second line of supportive research demonstrated that RNA molecules generated in the laboratory can catalyze almost any imaginable reaction needed in a living system—certainly a range of reactions much broader than those attributable to ribozymes existing today. Early RNA molecules could clearly have catalyzed all of the reactions required to set up a primordial cellular metabolism.

The third and fourth discoveries have further broadened our perspectives on RNA function. We now know that in ribosomes, the large ribonucleoprotein complexes that translate RNA into protein, the RNA is the active component with the capacity to catalyze protein synthesis (**Figure 1-6**; see also the Moment of Discovery for Chapter 18). Finally, and most recently, RNA sequences capable of simple forms of self-replication have been discovered (discussed in Chapter 16).

Ongoing research thus makes it possible to visualize a highly plausible sequence of events unfolding on the pathway from prebiotic soup to living systems. Arising from a myriad random primordial polymers, an RNA world came into being and gradually became more complex. An RNA capable of reliable self-replication may have been the first living entity. Self-replicators would have diversified to synthesize other ribozymes, leading to an RNA-based metabolism capable of providing a greater supply of needed RNA precursors. Ribozyme groupings became enclosed within lipid membranes. Particular groupings were successful, resulting in the first cells and a capacity to maintain a metabolic state out of equilibrium with the surroundings. As the RNA molecules in those cells increased in size and structural complexity, a need for stabilization and auxiliary functions arose. Peptides (proteins) were synthesized to neutralize the negative charges of the phosphates in the RNA backbone, to stabilize RNA structure in other ways, and to augment early metabolism. As more peptides were synthesized, some with catalytic activities arose. Proteins gradually supplanted RNA as catalysts, because the greater catalytic potential of proteins yielded an advantage. The protein world emerged, but not without retaining important vestiges of the RNA world (ribosomes and some other RNA catalysts), as we find them today.

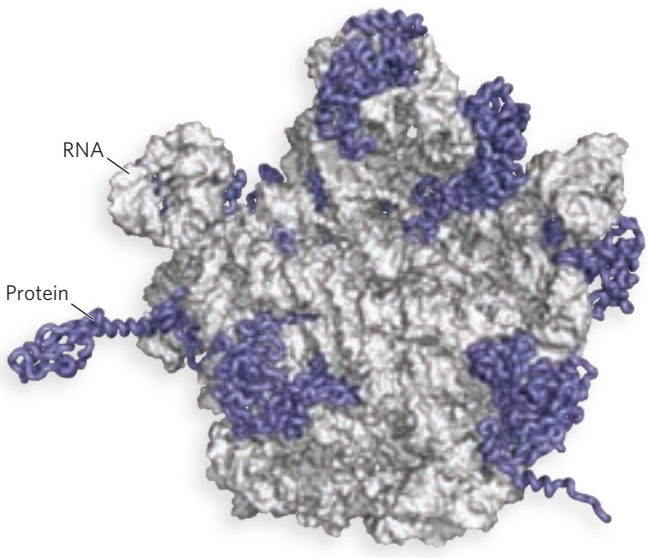


FIGURE 1-6 The 50S subunit of a bacterial ribosome. The gray parts of the subunit are RNA and the blue parts are protein. The structure is a huge ribozyme that evolved for the synthesis of protein. *[Source: PDB ID 1VSA.]*

The Last Universal Common Ancestor Is the Root of the Tree of Life

Countless nascent life forms probably arose from the primordial soup, along with many biological advances that improved their fitness. Successful combinations of RNA catalysts gave way to systems based on protein catalysts. Improvements in catalytic efficiency appeared, along with systematized genetic codes to link genetic information in RNA and DNA to protein sequences. Additional changes facilitated cellular metabolism and reproduction. Protein synthesis was systematized through the evolution of an efficient ribosome machine. RNA became more specialized for information storage and transmission. Cell membranes became more structured and specialized, eventually including mechanisms to selectively transport materials into and out of the cell as needed. And some processes became regulated. In this way, a variety of primitive cells may have evolved—each of them a viable living system. Organisms living today exhibit shared properties, telling us that one of these early experimental cells won out over the others. This cell, sometimes called **LUCA (last universal common ancestor)** (**Figure 1-7**), ultimately gave rise to all life now present on Earth.

LUCA is a special source of fascination for molecular biologists. Although LUCA probably lived more than 3 billion years ago, our speculation about what this cell was like is informed by experiment. One approach is to determine the minimum protein and genetic requirements for life. Attempts to create a minimal life form, either by reconstituting basic components or by taking bacteria and stripping them of all unnecessary parts, are underway in laboratories around the world. These experiments are not only defining properties that must have been present in LUCA; they are also setting the stage for the laboratory generation of engineered bacterial cells that can be used to manufacture chemicals for bioenergy, agriculture, and medicine.

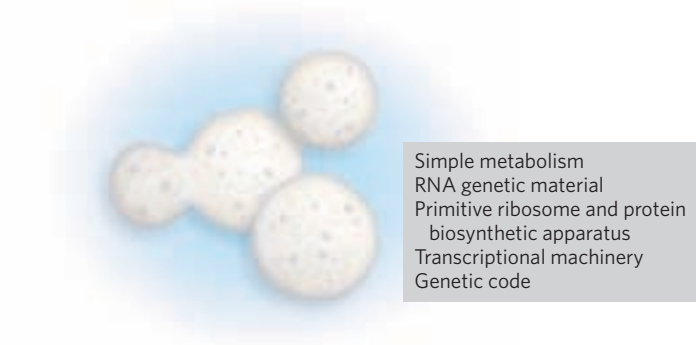


FIGURE 1-7 The last universal common ancestor. LUCA and its immediate descendants probably had a simple metabolism and a form of transcriptional machinery to replicate their RNA genome. A primitive ribosome and protein-biosynthetic apparatus would have used the same universal genetic code found in all modern organisms.

Another approach to understanding LUCA is to survey all types of living systems on Earth to determine which genes or characteristics are universal. The only genes that are truly universal in living systems are those encoding the cellular machinery for protein synthesis and some components of RNA transcription. All organisms also share (with very minor modifications discussed in Chapter 17) the same genetic code. That same code must have been present in LUCA. To support protein synthesis and RNA synthesis, a simple metabolism must have been present that allowed the uptake of chemical energy and its use to synthesize amino acids, nucleotides, and whatever lipids existed in the cell membrane from precursors available in the environment. The study of LUCA is described in more detail in Chapter 8.

The appearance of LUCA signaled the beginning of biological evolution on Earth. New types of cells gradually appeared, and new environments were exploited. The first cells were capable of taking up organic molecules from their surroundings and converting them to the molecules needed to support protein and RNA synthesis. Cellular complexity resulted in ever-increasing requirements for cellular genomic information. DNA, with a more uniform structure and some stability advantages relative to RNA, may first have appeared in viruses. It then gradually supplanted RNA as the most stable platform for the long-term storage and transmission of genetic information, and DNA replication and systems for the segregation of replicated DNA chromosomes into daughter cells evolved.

The early single-celled organisms derived from LUCA diversified to inhabit all niches in the ecosystem of this early Earth. The diversification eventually generated the three major groups of organisms that we recognize today: **bacteria**, **archaea**, and **eukaryotes** (**Figure 1-8**).

Many additional events helped shape the life we see around us. Notably, photosynthesis appeared about 2.5 billion years ago, as evidenced by the sudden rise in the concentration of atmospheric oxygen documented in the geologic record. As cells engulfed other cells, some endosymbiotic relationships developed and became permanent. The engulfed cells became organelles within their hosts more than 1 billion years ago, and we see these organelles today as chloroplasts and mitochondria. Loose

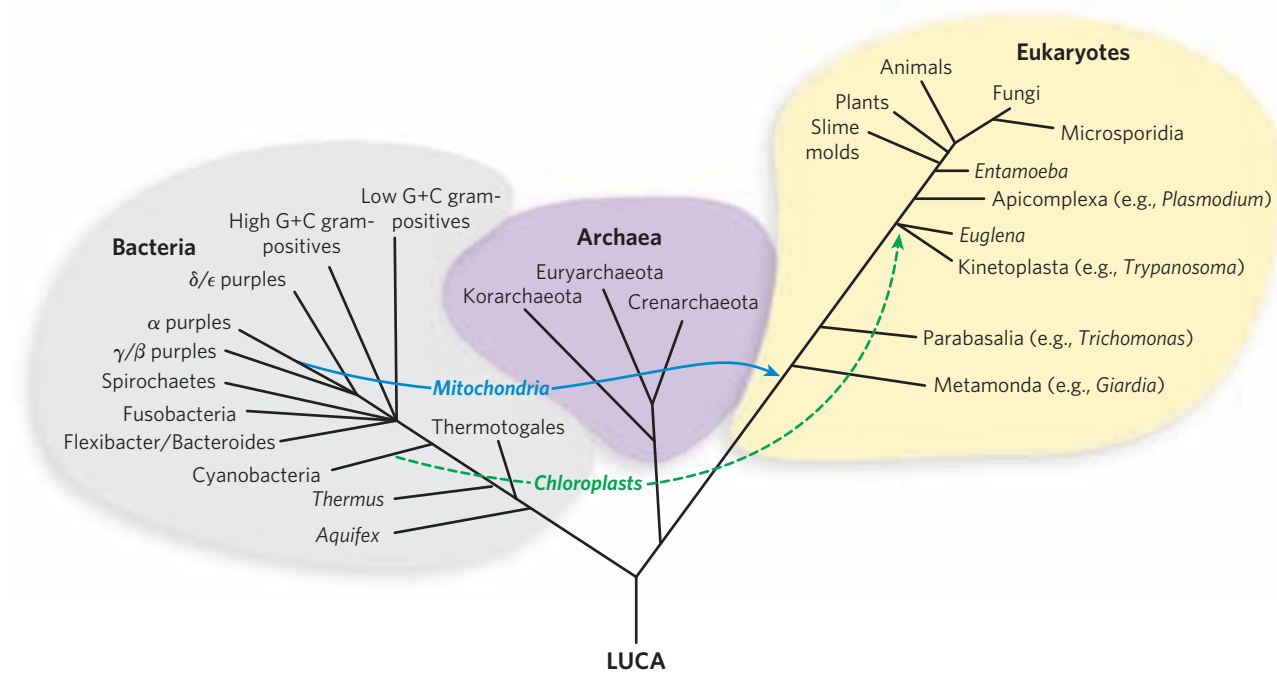


FIGURE 1-8 The universal tree of life. A current version of the tree is shown here, with branches for the three main groups of known organisms: bacteria, archaea, and eukaryotes. Particular types of bacteria, engulfed by other cells, gave rise to mitochondria and chloroplasts. *[Source: Data from J. R. Brown, "Universal tree of life," in* Encyclopedia of Life Sciences, *Wiley InterScience (online), 2005.]*