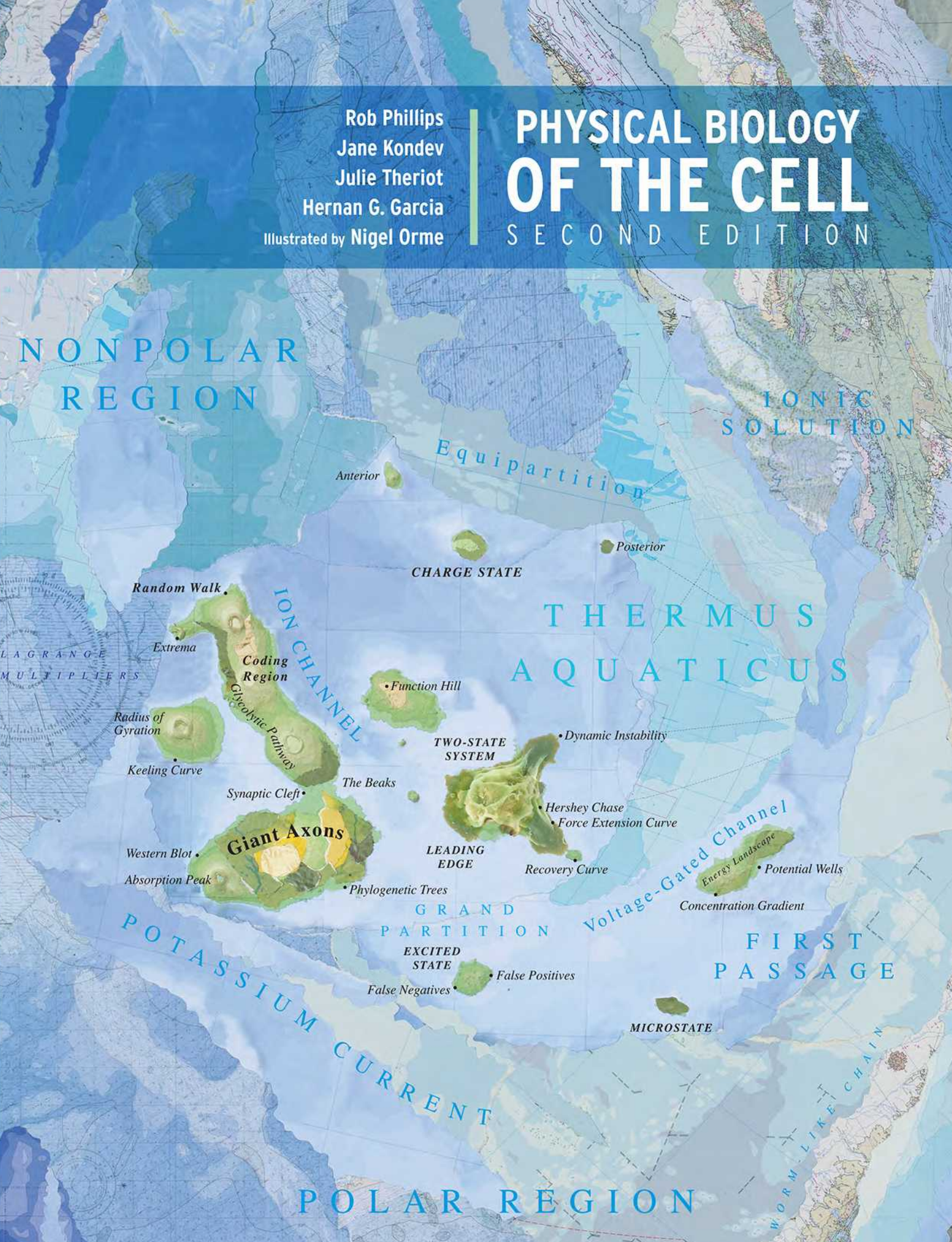

Rob Phillips
Jane Kondev
Julie Theriot
Hernan G. Garcia
Illustrated by Nigel Orme
PHYSICAL BIOLOGY
OF THE CELL
SECOND EDITION
NONPOLAR
REGION
IONIC
SOLUTION
Equipartition
Anterior
Posterior
CHARGE STATE
Random Walk
ION CHANNEL
THERMUS
AQUATICUS
LAGRANGE
MULTIPLIERS
Extrema
Coding
Region
Glycolytic Pathway
Function Hill
Radius of
Gyration
Keeling Curve
TWO-STATE
SYSTEM
Dynamic Instability
The Beaks
Synaptic Cleft
Hershey Chase
Force Extension Curve
Giant Axons
Western Blot
Absorption Peak
LEADING
EDGE
Recovery Curve
Energy Landscape
Potential Wells
Concentration Gradient
Voltage-Gated Channel
Phylogenetic Trees
GRAND
PARTITION
POTASSIUM CURRENT
EXCITED
STATE
False Positives
False Negatives
FIRST
PASSAGE
MICROSTATE
WORM-LIKE CHAIN
POLAR REGION

Physical Biology of the Cell

This page intentionally left blank
to match pagination of print book

Physical Biology of the Cell

Second Edition

Rob Phillips
Jane Kondev
Julie Theriot
Hernan G. Garcia

Illustrated by
Nigel Orme

Garland Science
Taylor & Francis Group
LONDON AND NEW YORK

Garland Science
Vice President: Denise Schanck
Editor: Summers Scholl
Senior Editorial Assistant: Kelly O'Connor
Cover design and illustrations: Nigel Orme
Production Editor: Natasha Wolfe
Copyeditor: Mac Clarke
Proofreader: Sally Huish
Typesetting: TechSet Composition India (P) Ltd.

Rob Phillips is the Fred and Nancy Morris Professor of Biophysics and Biology at the California Institute of Technology. He received a PhD in Physics from Washington University.

Jane Kondev is a Professor in the Department of Physics and within the Graduate Program in Quantitative Biology at Brandeis University. He attended the Mathematical High School in Belgrade, Serbia, received his Physics BS degree from the University of Belgrade, and his PhD from Cornell University.

Julie Theriot is a Professor in the Department of Biochemistry and the Department of Microbiology and Immunology at the Stanford University School of Medicine. She received concurrent BS degrees in Physics and Biology from the Massachusetts Institute of Technology, and a PhD in Cell Biology from the University of California at San Francisco.

Hernan G. Garcia is a Dicke Fellow in the Department of Physics at Princeton University. He received a BS in Physics from the University of Buenos Aires and a PhD in Physics from the California Institute of Technology.

ISBN 978-0-8153-4450-6

Library of Congress Cataloging-in-Publication Data

Phillips, Rob, 1960-
Physical biology of the cell. – Second edition / Rob Phillips, Jane Kondev, Julie Theriot, Hernan G. Garcia.
pages cm
ISBN 978-0-8153-4450-6 (pbk.)
1. Biophysics. 2. Cytology. I. Title.
QH505.P455 2013
571.6–dc23

2012030733

Published by Garland Science, Taylor & Francis Group, LLC, an informa business, 711 Third Avenue, New York, NY, 10017, USA, and 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

Garland Science
Taylor & Francis Group

Visit our web site at http://www.garlandscience.com

Dedicated to our friend Jon Widom

This page intentionally left blank
to match pagination of print book

Preface

"The map is not the territory."
Alfred Korzybski

The last 50 years in biology have seen an explosion of both data and understanding that rivals the fertile period between Tycho Brahe's definitive naked-eye investigations of the heavens and Newton's introduction of the "System of the World." One of the consequences of these stunning advances is the danger of becoming overwhelmed by the vast quantities of data coming at us from quarters ranging from next-generation sequencing to quantitative microscopy. For example, at the time of this writing, there are in excess of two million ribosomal RNA sequences deposited on publically accessible databases. But what does it all mean? A central role of scientific textbooks is to attempt to come to terms with broad areas of progress and to organize and distill the vast amounts of available information in a conceptually useful manner. In our view, an effective textbook can act as a map to help curious people discover unfamiliar territories. As with real maps, different purposes are served by different kinds of abstraction. Some maps show roads, some show topography, with each being useful in its own context.

A number of textbook writers have undertaken the formidable task of writing excellent, comprehensive surveys of cell and molecular biology, although each one of these books serves as a slightly different kind of map for the same overlapping territory. Although we cover some of the same material as a typical molecular and cell biology book, our goal in this book is fundamentally different. There is no single, correct way to construct a conceptually simplified map for a huge and complex field such as cell and molecular biology. Most modern biology textbooks organize ideas, facts, and experimental data based on their conceptual proximity for some particular biological function. In contrast, this book examines the same set of information from the distinct perspective of *physical biology*. We have therefore adopted an organization in which the proximity of topics is based on the physical concepts that unite a given set of biological phenomena, instead of the cell biology perspective. By analogy to a map of the United States, a cell biology textbook might describe the plains of Eastern Colorado in the same chapter as the mountains of Western Colorado, whereas our physical biology book would group Eastern Colorado with the rolling fields of Iowa, and Western Colorado with mountainous West Virginia.

This book does not assume extensive prior knowledge on the part of the reader, though a grounding in both calculus and elementary physics is essential. The material covered here is appropriate for a first course in physical biology or biophysics for either undergraduates or graduate students. It is also intended for any scientist interested in learning the basic principles and applications of physical modeling for research in biology, and aims to provide a novel perspective even to scientists who are already familiar with some of the material. Throughout the book, our organization of ideas and data based on proximity in physical biology space juxtaposes topics that are not obviously related in cell biology space. For example, DNA

wrapping around nucleosomes in the eukaryotic nucleus, DNA looping induced by the binding of transcriptional repressors in the context of bacterial gene regulation, and DNA packing into the narrow confines of bacteriophage capsids all appear in the same chapter because they are related by the mechanical rules governing the bending of DNA. Next, the physical and mathematical treatment we derive for DNA bending is directly applied to other kinds of long, thin, biological structures, including the filaments of the cytoskeleton. This organizational principle brings into focus the central thesis of this book, namely, that the appropriate application of a relatively small number of fundamental physical models can serve as the foundation of whole bodies of quantitative biological intuition, broadly useful across a wide range of apparently unrelated biological problems.

During the 12-year journey that led to this book, we benefited immeasurably from the generosity and enthusiasm of hundreds of scientific colleagues who graciously shared their data, ideas, and perspectives. Indeed, in many respects, we view our book as an exercise in *quantitative journalism*, based upon extensive "interviews" with these various scientists in a wide range of disciplines. We offer this book as a report from the front, to share some of the most interesting things that we have learned from our colleagues with any and all inquiring individuals who wish to think both deeply and broadly about the connections between biology and the physical sciences. Our imagined audience spans the range from 18-year-old mechanical engineering undergraduates curious about the application of their discipline to medicine, to 40-year-old string theorists wishing to apply their mathematical and physical talents to living matter, to 70-year-old renowned biologists wondering whether their insights into living systems might be improved by a mathematical treatment.

Although the claim that a handful of simple physical models can shed more than superficial light on complex biological processes might seem naive, the biological research literature is teeming with examples where important quantitative insight into questions of pressing interest has been gained by the application of such models. In every chapter, we have chosen specific examples from classic and current research papers where quantitative measurements on biological systems can be largely understood by recourse to simple, fundamental, physical ideas. In cases where the simplest possible physical models fail to fit the data, the specific quantitative nature of the disparities can often lead to testable new biological hypotheses. For example, a simple calculation estimating the amount of time it would take for a newly synthesized protein to diffuse from the cell body of a motor neuron in the spinal cord to the synapse formed by the same neuron in the foot proves that diffusion is far too slow to get the job done, and an active transport process must occur. Inevitably, researchers performing experiments on biological systems must have physical models explicitly or implicitly in mind, whether imagining how changes in the rate of transcription initiation for a particular gene will lead to changes in the overall amount of the gene product in the cell, or picturing the ways that signaling molecules move through cellular space to encounter their targets, or envisioning how cell movements during embryogenesis lead to the final three-dimensional structures of organs and limbs. In this book, we aim to provide a physical and mathematical toolkit so that people used to thinking deeply about biological problems can make this kind of quantitative intuition explicit; we also hope to provide a perspective on biology that may inspire people from a background more heavily based in physics or

mathematics to seek out new biological problems that are particularly appropriate for this kind of quantitative analysis.

Our general approach follows four steps. First, we introduce a biological phenomenon; second, we perform simple order-of-magnitude estimates to develop a "feeling for the numbers" involved in that process; third, we demonstrate the application of an extremely simple first-pass model; and finally, where possible, we present a refinement of the oversimplified model to better approximate biological reality. Our goal is to share the pleasure in seeing the extent to which simple models can be tailored to reveal the *complexity* of observed phenomena. For our examples, we have chosen particular biological cases that we believe to be worthy illustrations of the concepts at hand and that have captured our imaginations, often because of particularly elegant or clever experiments that were designed to generate intriguing sets of quantitative data. While we have been conscientious in our exploration of these facts and in our construction of simple models, it is inevitable that we will have made errors due to our ignorance and also due to the fact that, in many cases, new discoveries may change the particulars of our case studies. (A list of errors and their corrections will be posted on the book's website as well as the website of one of the authors (R.P.).) Nevertheless, because our goal is to demonstrate the power of applying simple models to complex systems, even when some details are wrong or missing, we hope that any particular lapses will not obscure the overall message. Furthermore, in many cases, we have described phenomena that are still awaiting a satisfying physical model. We hope that many of our readers will seize upon the holes and errors in our exploration of physical biology and take these as challenges and opportunities for launching exciting original work.

Our second edition builds upon the foundations laid in the previous edition, with the addition of two new chapters that focus on central themes of modern biology, namely, light and life and the emergence of patterns in living organisms. The new Chapter 18 focuses on several key ways in which light is central in biology. We begin with an analysis of photosynthesis that illustrates the quantum mechanical underpinnings of both the absorption of light and the transfer of energy and electrons through the photosynthetic apparatus. The second part of our story in that chapter considers the rich and beautiful subject of vision. The new Chapter 20 uses insights garnered throughout the book to ask how it is that organisms ranging from flies to plants can build up such exquisite patterns. Here we explore Turing's famed model of several interacting chemical species undergoing chemical reactions and diffusion and other more recent advances in thinking about problems ranging from somitogenesis to phyllotaxis.

The book is made up of four major parts. Part 1, *The Facts of Life*, largely focuses on introducing biological phenomena. For biology readers already familiar with this material, the hope is that the quantitative spin will be enlightening. For physics readers, the goal is to get a sense of the biological systems themselves. Part 2, *Life at Rest*, explores those problems in biology that can be attacked using quantitative models without any explicit reference to time. Part 3, *Life in Motion*, tackles head-on the enhanced complexity of time-dependent systems exhibiting dynamic behavior. Finally, Part 4, *The Meaning of Life*, addresses various kinds of information processing by biological systems.

Because our hope is that you, our readers, represent a broad diversity of backgrounds and interests, throughout the book we try as much as possible to introduce the origin of the facts and principles

that we exploit. We are reluctant to ever simply assert biological "facts" or physical "results," and would not expect you to blindly accept our assertions if we did. Therefore, we often describe classical observations by biologists over the past centuries as well as the most recent exciting results, and illustrate how current thinking about complex biological problems has been shaped by a progression of observations and insights. Extended discussions of this kind are separated from the main text in sections labeled *Experiments Behind the Facts*. In a complementary way, whenever we find it necessary to derive mathematical equations, we proceed step by step through the derivation and explain how each line leads to the next, so that readers lacking a strong background in mathematics can nevertheless follow every step of the logic and not be forced to take our word for any result. Specific sections labeled *The Math Behind the Models* and *The Tricks Behind the Math* provide summaries for the mathematical techniques that are used repeatedly throughout the book; many readers trained in physics will already be familiar with this material, but biologists may benefit from a brief refresher or introduction. In addition, we include sections labeled *Estimate* that help to develop a "feeling for the numbers" for particularly interesting cases.

Another critical new element in our second edition is a feature called *Computational Exploration*. The idea of these excursions is to show how simple computer analyses can help us attack problems that are otherwise inaccessible. In the first edition, we underemphasized "computation" because we wanted to combat the spurious idea that theory in biology is synonymous with computation. While we made this exaggeration to make a point, we did so at a price, because computation is not only useful, but downright indispensable in some problems. Further, one of the beauties of turning a model into a specific numerical computation is that to get a computer to produce a meaningful number, nothing can be left unspecified. The Computational Explorations are offered as a way for the reader to develop a particular habit of mind, and none of them should be viewed as illustrating the state of the art for making such calculations. Matlab and Mathematica code related to most of these explorations is provided on the book's website.

Although we review the basic information necessary to follow the exposition of each topic, you may also find it useful to have recourse to a textbook or reference book covering the details of scientific areas among biology, physics, chemistry, and mathematics, with which you consider yourself less familiar. Some references that are among our favorites in these fields are suggested at the end of each chapter. More generally, our references to the literature are treated in two distinct ways. Our suggestions for *Further Reading* reflect our own tastes. Often, the choices that appear at a chapter's end are chosen because of uniqueness of viewpoint or presentation. We make no attempt at completeness. The second class of *References* reflect work that has explicitly touched the content of each chapter, either through introducing us to a model, providing a figure, or constructing an argument.

At the end of each chapter, we include a series of problems that expand the material in the chapter or give the opportunity to attempt model-building for other case studies. In the second edition, we have considerably expanded the scope of the end-of-chapter problems. These problems can be used within formal courses or by individual readers. A complete *Solutions Manual*, covering all problems in the book, is available for instructors. There are several different types

of problems. Some, whose goal is to develop a "feeling for the numbers," are arithmetically simple, and primarily intended to develop a sense of order-of-magnitude biology. Others request difficult mathematical derivations that we could not include in the text. Still others, perhaps our favorites, invite the readers to apply quantitative model-building to provocative experimental data from the primary research literature. In each chapter, we have loosely identified the different problems with the aforementioned categories in order to assist the reader in choosing which one to attack depending on particular need. The book's website also includes *Hints for the Reader* for some of the more difficult problems.

Our book relies heavily on original data, both in the figures that appear throughout the book and in the various end-of-chapter problems. To make these data easily accessible to interested readers, the book's website includes the original experimental data used to make all the figures in the book that are based upon published measurements. Similarly, the data associated with the end-of-chapter problems are also provided on the book's website. It is our hope that you will use these data in order to perform your own calculations for fitting the many models introduced throughout the book to the relevant primary data, and perhaps refining the models in your own original work.

Student and Instructor Resources

Figures and PowerPoint® Presentations

The figures from the book are available in two convenient formats: PowerPoint and JPEG. There is one PowerPoint presentation for each chapter, and the JPEGs have been optimized for display on a computer.

Data Sets

The original data used to create both the figures and homework problems are available in Excel® spreadsheets. With this data, the reader can extend the theoretical tools developed in the book to fit experimental data for a wide range of problems. The data files contain explicit statement of all relevant units, and include references to the original sources.

Hints for Problems

This PDF provides both hints and strategies for attacking some of the more difficult end-of-chapter problems. In some cases, the hints provide intuition about how to set up the problem; in other cases, the hints provide explicit mathematical instructions on how to carry through more tricky manipulations.

Matlab® and Mathematica® Code

These files contain code for the Computational Explorations sidebars located throughout the book.

Movies

The movies complement the figures and discussion from the book by illustrating the rich dynamics exhibited by living organisms and the molecules that make them tick.

Solutions Manual

This PDF contains solutions to all problems in the book. It is available only to qualified instructors.

With the exception of the *Solutions Manual*, these resources are available on the *Physical Biology of the Cell*, 2nd Edition, media website:

http://microsite.garlandscience.com/pboc2

Access to the Solutions Manual is available to qualified instructors by emailing science@garland.com.

Acknowledgments

This book would not have been possible without a wide range of support from both people and institutions. We are grateful for the support of the Aspen Center for Physics, the Kavli Institute for Theoretical Physics at the University of California, Santa Barbara and the ESPCI in Paris, where some of the writing was done. Our funding during the course of this project was provided by the National Science Foundation, the National Institutes of Health, The Research Corporation, the Howard Hughes Medical Institute, and the MacArthur Foundation. We also particularly acknowledge the support of the NIH Director's Pioneer Award and La Fondation Pierre Gilles de Gennes granted to R.P and The Donna and Benjamin M. Rosen Center for Bioengineering at Caltech, all of which provided broad financial support for many facets of this project.

Our book would never have achieved its present incarnation without the close and expert collaboration of our gifted illustrator, Nigel Orme, who is responsible for the clarity and visual appeal of the hundreds of figures found in these pages, as well as the overall design of the book. We also had the pleasure of working with David Goodsell who produced many illustrations throughout the book showing detailed molecular structures. Genya Frenkel also provided assistance on the problems and their solutions. Amy Phillips assisted with editing, responding to reader comments, and obtaining permission for use of many of the previously published images in the figures. Maureen Storey (first edition) and Mac Clarke (second edition) improved our clarity and respectability with their expert copy editing. Our editors Mike Morales (first edition) and Summers Scholl (second edition) have offered great support through the entirety of the project. Simon Hill (first edition) and Natasha Wolfe's (second edition) expert assistance in the production process has been an impressive pleasure.

One of the most pleasurable parts of our experience of writing this book has been our interaction with generous friends and colleagues who have shared their insights, stories, prejudices, and likes and dislikes about biology, physics, chemistry, and their overlap. We are deeply grateful to all our colleagues who have contributed ideas directly or indirectly through these many enjoyable conversations over the past twelve years. Elio Schaechter told us the secret to maintaining a happy collaboration. Lubert Stryer inspired the overall section organization and section titles, and gave us much-needed practical advice on how to actually finish the book. Numerous others have helped us directly or indirectly through inspiration, extended lab visits, teaching us about whole fields, or just by influential interactions along the way. It is very important to note that in some cases these people explicitly disagreed with some of our particular conclusions, and deserve no blame for our mistakes and misjudgments. We specifically wish to thank: Gary Ackers, Bruce Alberts, Olaf Andersen, David Baltimore, Robert Bao, David Bensimon, Seymour Benzer, Howard Berg, Paul Berg, Maja Bialecka, Bill Bialek, Lacra Bintu, Pamela Bjorkman, Steve Block, Seth Blumberg, David Boal, James Boedicker, Rob Brewster, Robijn Bruinsma, Zev Bryant, Steve Burden, Carlos Bustamante, Anders Carlsson, Sherwood Casjens, Yi-Ju Chen, Kristina Dakos, Eric Davidson, Scott Delp, Micah Dembo, Michael Dickinson, Ken Dill, Marileen Dogterom, David Dunlap, Michael Elowitz, Evan Evans, Stan Falkow, Julio Fernandez, Jim Ferrell, Laura Finzi, Daniel Fisher, Dan Fletcher, Henrik Flyvbjerg, Seth Fraden, Scott Fraser, Ben Freund, Andrew J. Galambos, Ethan Garner, Bill Gelbart, Jeff Gelles, Kings Ghosh, Dan Gillespie, Yale Goldman, Bruce Goode, Paul Grayson, Thomas Gregor, Jim Haber, Mike Hagan, Randy Hampton, Lin Han, Pehr Harbury, Dan Herschlag, John Heuser, Joe Howard, KC Huang, Terry Hwa, Grant Jensen, Jack Johnson, Daniel Jones, Jason Kahn, Dale Kaiser, Suzanne Amador Kane, Sarah Keller, Doro Kern, Karla Kirkegaard, Marc Kirschner, Bill Klug, Chuck Knobler, Tolya Kolomeisky, Corinne Ladous, Jared Leadbetter, Heun Jin Lee, Henry Lester, Julian Lewis, Jennifer Lippincott-Schwartz, Sanjoy Mahajan, Jim Maher, Carmen Mannella, William Martin, Bob Meyer, Elliot Meyerowitz, Chris Miller, Ken Miller, Tim Mitchison, Alex Mogilner, Cathy Morris, Dyche Mullins, Richard Murray, Kees Murre, David Nelson, James Nelson, Phil Nelson, Keir Neuman, Dianne Newman, Lene Oddershede, Garry Odell, George Oster, Adrian Parsegian, Iva Perovic, Eduardo Perozo, Eric Peterson, Suzanne

Pfeffer, Tom Pollard, Dan Portnoy, Tom Powers, Ashok Prasad, Mark Ptashne, Prashant Purohit, Steve Quake, Sharad Ramanathan, Samuel Rauhala, Michael Reddy, Doug Rees, Dan Reeves, Joy Rimchala, Ellen Rothenberg, Michael Roukes, Dave Rutledge, Peter Sarnow, Klaus Schulten, Bob Schleif, Darren Segall, Udo Seifert, Paul Selvin, Lucy Shapiro, Boris Shraiman, Steve Small, Doug Smith, Steve Smith, Andy Spakowitz, Jim Spudich, Alasdair Steven, Sergei Sukharev, Christian Sulloway, Joel Swanson, Boo Shan Tseng, Tristan Ursell, Ron Vale, David Van Valen, Elizabeth Villa, Zhen-Gang Wang, Clare Waterman, Annemarie Weber, Jon Widom, Eric Wieschaus, Paul Wiggins, Ned Wingreen, Zeba Wunderlich, Ahmed Zewail, and Kai Zinn.

Finally, we are deeply grateful to the individuals who have given us critical feedback on the manuscript in its various stages, including the many students in our courses offered at Caltech, Brandeis, and Stanford over the last twelve years. They have all done their best to save us from error and any remaining mistakes are entirely our responsibility. We are indebted to all of them for their generosity with their time and expertise. A few hardy individuals read the entire first edition: Laila Ashegian, Andre Brown, Genya Frenkel, Steve Privitera, Alvaro Sanchez, and Sylvain Zorman. We thank them for their many insightful comments and remarkable stamina. For the second edition, we had the great fortune to have Howard Berg read every word of our book always providing pointed and thoughtful commentary. Similarly, Ron Milo has been a constant source of critical commentary, and encouragement throughout the process. Velocity Hughes and Madhav Mani were a tremendous help in reading the entire book in its near final form and providing critical comments at every turn. Justin Bois has also been a source of numerous critical insights. Niles Pierce has also provided his unflagging support throughout this project.

Many more people have given expert commentary on specific chapters, provided specific figures, advised on us end-of- chapter problems, or provided particular insights for either the first or second edition:

Chapter 1

Bill Gelbart (University of California, Los Angeles), Shura Grosberg (New York University), Randy Hampton (University of California, San Diego), Sanjoy Mahajan (Olin College), Ron Milo (Weizmann Institute of Science), Michael Rubinstein (University of North Carolina, Chapel Hill).

Chapter 2

John A. G. Briggs (European Molecular Biology Laboratory), James Boedicker (California Institute of Technology), James Brody (University of California, Irvine), Titus Brown (Michigan State University), Ian Chin-Sang (Queen's University), Avigdor Eldar (Tel Aviv University), Scott Fraser (California Institute of Technology), CT Lim (National University of Singapore), Dianne Newman (California Institute of Technology), Yitzhak Rabin (Bar-Ilan University), Manfred Radmacher (University of Bremen), Michael Rubinstein (University of North Carolina, Chapel Hill), Steve Small (New York University), Linda Song (Harvard University), Dave Tirrell (California Institute of Technology), Jon Widom (Northwestern University).

Chapter 3

Tom Cech (University of Colorado), Andreas Matouschek (Northwestern University), Yitzhak Rabin (Bar-Ilan University), Michael Reddy (University of Wisconsin, Milwaukee), Nitzan Rosenfeld (Rosetta Genomics), Michael Rubinstein (University of North Carolina, Chapel Hill), Antoine van Oijen (Rijksuniversiteit Groningen), Jon Widom (Northwestern University).

Chapter 4

Elaine Bearer (Brown University), Paul Jardine (University of Minnesota, Twin Cities), Michael Reddy (University of Wisconsin, Milwaukee), Michael Rubinstein (University of North Carolina, Chapel Hill).

Chapter 5

James Boedicker (California Institute of Technology), Ken Dill (Stony Brook University), Randy Hampton (University of California, San Diego), Rick James (University of Minnesota, Twin Cities), Heun Jin Lee (California Institute of Technology), Bill Klug (University of California, Los Angeles), Steve Quake (Stanford University), Elio Schaechter (San Diego State University).

Chapter 6

Ken Dill (Stony Brook University), Dan Herschlag (Stanford University), Terry Hwa (University of California, San Diego), Arbel Tadmor (California Institute of Technology).

Chapter 7

Gary Ackers (Washington University in St. Louis), Olaf Andersen (Cornell University), Ken Dill (Stony Brook University), Henry Lester (California Institute of Technology).

Chapter 8

Ken Dill (Stony Brook University), Shura Grosberg (New York University), Michael Rubinstein (University of North Carolina, Chapel Hill), Jeremy Schmit (Kansas State University), Andy Spakowitz (Stanford University), Paul Wiggins (University of Washington).

Chapter 9

Mike Hagan (Brandeis University), Thomas Record (University of Wisconsin, Madison), Bob Schleif (Johns Hopkins University), Pete von Hippel (University of Oregon).

Chapter 10

Zev Bryant (Stanford University), Carlos Bustamante (University of California, Berkeley), Hans-Günther Döbereiner (University of Bremen), Paul Forscher (Yale University), Ben Freund (Brown University), Bill Gelbart (University of California, Los Angeles), Paul Grayson (California Institute of Technology), Mandar Inamdar (Indian Institute of Technology, Bombay), Bill Klug (University of California, Los Angeles), Joy Rimchala (Massachusetts Institute of Technology), Doug Smith (University of California, San Diego), Megan Valentine (University of California, Santa Barbara), Jon Widom (Northwestern University), Paul Wiggins (University of Washington).

Chapter 11

Ashustosh Agrawal (University of Houston), Patricia Bassereau (Institut Curie), Hans-Günther Döbereiner (University of Bremen), Evan Evans (University of British Columbia), Dan Fletcher (University of California, Berkeley), Terry Frey (San Diego State University), Christoph Haselwandter (University of Southern California), KC Huang (Stanford University), Sarah Keller (University of Washington, Seattle), Bill Klug (University of California, Los Angeles), Carmen Mannella (State University of New York, Albany), Eva Schmid (University of California, Berkeley), Pierre Sens (ESPCI, Paris), Sergei Sukharev (University of Maryland, College Park), Tristan Ursell (Stanford University), Paul Wiggins (University of Washington).

Chapter 12

Howard Berg (Harvard University), Justin Bois (University of California, Los Angeles), Zev Bryant (Stanford University), Ray Goldstein (University of Cambridge), Jean-François Joanny (Institut Curie), Sanjoy Mahajan (Olin College), Tom Powers (Brown University), Todd Squires (University of California, Santa Barbara), Howard Stone (Harvard University).

Chapter 13

Howard Berg (Harvard University), Ariane Briegel (California Institute of Technology), Dan Gillespie, Jean-François Joanny (Institut Curie), Martin Linden (Stockholm University), Jennifer Lippincott-Schwartz (National Institutes of Health), Ralf Metzler (Technical University of Munich), Frosso Seitaridou (Emory University), Pierre Sens (ESPCI, Paris), Dave Wu (California Institute of Technology).

Chapter 14

Jean-François Joanny (Institut Curie), Randy Kamien (University of Pennsylvania), Martin Linden (Stockholm University), Ralf Metzler (Technical University of Munich), Pierre Sens (ESPCI, Paris), Arbel Tadmor (California Institute of Technology).

Chapter 15

Anders Carlsson (Washington University in St. Louis), Marileen Dogterom (Institute for Atomic and Molecular Physics), Dan Fletcher (University of California, Berkeley), Dan Herschlag (Stanford University), Jean-François Joanny (Institut Curie), Tom Pollard (Yale University), Dimitrios Vavylonis (Lehigh University).

Chapter 16

Bill Gelbart (University of California, Los Angeles), Jean-François Joanny (Institut Curie), Tolya Kolomeisky (Rice University), Martin Linden (Stockholm University), Jens Michaelis (Ludwig-Maximilians University), George Oster (University of California, Berkeley), Megan Valentine (University of California, Santa Barbara), Jianhua Xing (Virginia Polytechnic Institute and State University).

Chapter 17

Olaf Andersen (Cornell University), Chris Gandhi (California Institute of Technology), Jean-François Joanny (Institut Curie), Stephanie Johnson (California Institute of Technology), Rod MacKinnon (Rockefeller University), Chris Miller (Brandeis University), Paul Miller (Brandeis University), Phil Nelson (University of Pennsylvania).

Chapter 18

Maja Bialecka-Fornal (California Institute of Technology), Bill Bialek (Princeton University), David Chandler (University of California, Berkeley), Anna Damjanovic (Johns Hopkins University), Govindjee (University of Illinois, Urbana-Champaign), Harry Gray (California Institute of Technology), Heun Jin Lee (California Institute of Technology), Rudy Marcus (California Institute of Technology), Tom Miller (California Institute of Technology), Ron Milo (Weizmann Institute of Science), Jose Onuchic (Rice University), Nipam Patel (University of California, Berkeley), Mark Ratner (Northwestern University), Mattias Rydenfelt (California Institute of Technology), Dave Savage (University of California, Berkeley), Klaus Schulten (University of Illinois, Urbana-Champaign), Kurt Warncke (Emory University), Jay Winkler (California Institute of Technology).

Chapter 19

James Boedicker (California Institute of Technology), Robert Brewster (California Institute of Technology), Titus Brown (Michigan State University), Nick Buchler (Duke University), Eric Davidson (California Institute of Technology), Avigdor Eldar (Tel Aviv University), Michael Elowitz (California Institute of Technology), Robert Endres (Imperial College London), Daniel Fisher (Stanford University), Scott Fraser (California Institute of Technology), Uli Gerland (Ludwig-Maximilians University), Ido Golding (Baylor College of Medicine), Mikko Haataja (Princeton University), Terry Hwa (University of California, San Diego), Daniel Jones (California Institute of Technology), Tom Kuhlman (University of Illinois, Urbana-Champaign), Wendell Lim (University of California, San Francisco), Chris Myers (Cornell University), Bob Schleif (Johns Hopkins University), Vivek Shenoy (Brown University), Steve Small (New York University), Peter Swain (McGill University), David Van Valen (California Institute of Technology), Ned Wingreen (Princeton University), Sunney Xie (Harvard University).

Chapter 20

Justin Bois (University of California, Los Angeles), Thomas Gregor (Princeton University), KC Huang (Stanford University), Frank Julicher (Max Planck Institute of Complex Systems, Dresden), Karsten Kruse (University of Saarlandes), Andy Oates (Max Planck Institute of Molecular Cell Biology and Genetics, Dresden), Jordi Garcia Ojalvo (Polytechnic University of Catalonia), George Oster (University of California, Berkeley), Andrew Rutenberg (Dalhousie University), David Sprinzak (Tel Aviv University), Carolina Tropini (Stanford University).

Chapter 21

Ralf Bundschuh (The Ohio State University), Uli Gerland (Ludwig-Maximilians University), Daniel Jones (California Institute of Technology), Justin Kinney (Cold Spring Harbor Laboratory), Chris Myers (Cornell University), Eric Peterson (California Institute of Technology), Frank Pugh (Pennsylvania State University), Jody Puglisi (Stanford University), Oliver Rando (University of Massachusetts Medical School), Tony Redondo (Los Alamos National Laboratory), Eran Segal (Weizmann Institute of Science), Boris Shraiman (University of California, Santa Barbara) Peter Swain (University of Edinburgh), Jon Widom (Northwestern University), Chris Wiggins (Columbia University).

Contents

Contents in Detail

Special Sections

There are five classes of special sections indicated with icons and colored bars throughout the text. They perform order of magnitude estimates, explore biological problems using computation, examine the experimental underpinnings of topics, and elaborate on mathematical details.

COMPUTATIONAL EXPLORATION

ESTIMATE

EXPERIMENTS

MATH

TRICKS

Map of the Maps

Part 1: Map of Alfred Russel Wallace's voyage with the black lines denoting Wallace's travel route and the red lines indicating chains of volcanoes. From *The Malay Archipelago* (1869) by Alfred Russel Wallace.

Chapter 1: Map of the world according to Eratosthenes (220 B.C.E.). Erastosthenes is known for, among many other things, his measurement of the circumference of the Earth, and is considered one of the founders of the subject of geography. From *Report on the Scientific Results of the Voyage of the H.M.S. Challenger During the Years 1872–76*, prepared under the superintendence of C. Wyville Thompson and John Murray (1895).

Chapter 2: Population density in Los Angeles County, as determined in the 2000 census. Darker colors represent denser populations (up to 100,000 people per square mile). From the United States Census Bureau.

Chapter 3: Sedimentary rock layers in the Grand Canyon. Geology and cross section by Peter J. Conley, artwork by Dick Beasley. From the United States National Park Service (1985).

Chapter 4: *Carta marina*, a map of Scandinavia, by Olaus Magnus. A translation of the Latin caption reads: *A Marine map and Description of the Northern Lands and of their Marvels, most carefully drawn up at Venice in the year 1539 through the generous assistance of the Most Honourable Lord Hieronymo Quirino*. This detail shows the sea monsters in the ocean between Norway and Iceland.

Part 2: Tourist map of Père Lachaise cemetery, Paris, France.

Chapter 5: Airplane routes around the nearly spherical Earth. Courtesy of OpenFlights.com.

Chapter 6: Josiah Willard Gibbs articulated the variational principle that shows how to find the equilibrium state of a system by maximizing the entropy. Gibbs spent his entire career in New Haven, Connecticut at Yale University. This 1886 map shows the university buildings during Gibbs' time. Source: Yale University Map Collection. Courtesy of the Yale University Map Collection.

Chapter 7: County map of Virginia and West Virginia, drawn by Samuel Augustus Mitchell Jr. in 1864, after the American Civil War.

Chapter 8: Aerial view of the hedge maze at Longleat Safari and Adventure Park, near Warminster, United Kingdom. Courtesy of Atlaspix/Alamy.

Chapter 9: Topographic map of the Great Salt Lake (Utah, United States) and surrounding region. From the United States Geological Survey (1970).

Chapter 10: Blueprint diagram of the Golden Gate Bridge, San Francisco, California, United States. Courtesy of EngineeringArtwork.com

Chapter 11: Digital elevation map of Mount Cotopaxi in the Andes Mountains, near Quito, Ecuador. Blue and green correspond to the lowest elevations in the image, while beige, orange, red, and white represent increasing elevations. Courtesy of the NASA Earth Observatory (2000).

Part 3: Migration tracks of the sooty shearwater, a small seabird, tracked with geolocating tags from two breeding colonies in New Zealand. Breeding season is shown in blue, northward migration in yellow, and wintering season and southward migration in orange. Over about 260 days, an individual animal travels about 64,000 km in a figure-8 pattern across the entire Pacific Ocean. From S. A. Shaffer et al., "Migratory shearwaters integrate oceanic resources across the Pacific Ocean in an endless summer," *Proceedings of the National Academy of Sciences USA*, **103**: 12799–12802, 2006.

Chapter 12: Worldwide distribution of ocean currents (warm in red, cold in green). Arrows indicate the direction of drift; the number of strokes on the arrow shafts denote the magnitude of the drift per hour. Sea ice is shown in purple. Prepared by the American Geographical Society for the United States Department of State in 1943.

Chapter 13: Temperature map of the sun's corona, recorded by the Extreme Ultraviolet Imaging Telescope at the Solar and Heliospheric Observatory on June 21, 2001. Courtesy of ESA/NASA.

Chapter 14: John Snow's map of the 1854 cholera outbreak in the Soho neighborhood of London. By interviewing residents of the neighborhood where nearly 500 people died of cholera in a ten-day period, Snow found that nearly all of the deaths occurred in homes close to the water pump in Broad Street, which he hypothesized was the source of the epidemic. Reproduced from *On the Mode of Communication of Cholera,* 2nd *Edition*, John Snow (1855).

Chapter 15: Positron emission tomography (PET scan) map of a healthy human brain, showing the rate of glucose utilization in various parts of the right hemisphere. Warmer colors indicate faster glucose uptake. Courtesy of Alzheimer's Disease Education and Referral Center, a service of the National Institute on Aging (United States National Institutes of Health).

Chapter 16: High speed train routes of France, mapped as a transit diagram. Courtesy of Cameron Booth.

Chapter 17: Nile River delta at night, as photographed by the crew in Expedition 25 on the International Space Station on October 28, 2010. Courtesy of Image Science & Analysis Laboratory, Johnson Space Center, Earth Observatory, NASA/GSFC SeaWiFS Project.

Chapter 18: Single-celled photosynthetic organisms such as the coccolithophore *Emiliana huxleyi* can form gigantic oceanic blooms visible from space. In this April 1998 image, the Aleutian Islands and the state of Alaska are visible next to the Bering Sea that harbors the algal bloom. Courtesy of NASA/GSFC SeaWiFS Project.

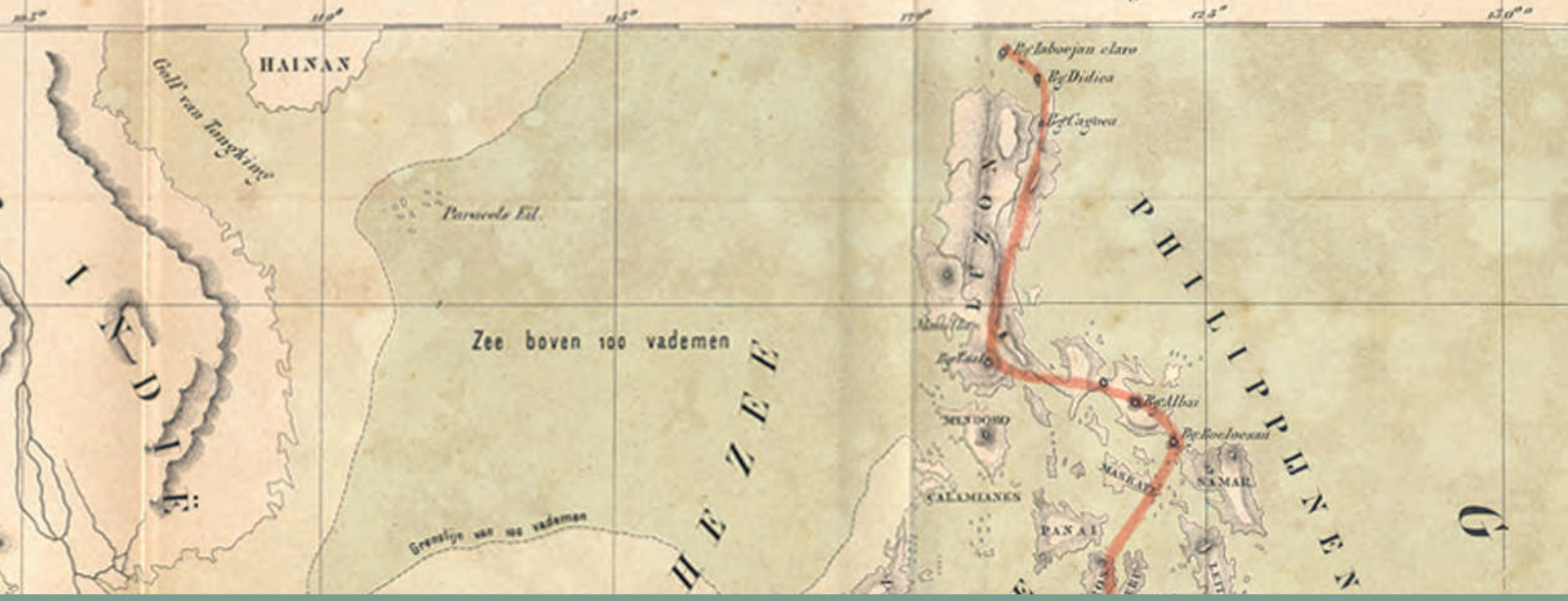
Part 4: A map of the infant universe, revealed by seven years of data from the Wilkinson Microwave Anisotropy Probe (WMAP). The image reveals 13.7 billion year old temperature fluctuations (the range of ±200 microKelvin is shown as color differences) that correspond to the seeds that grew to become the galaxies. Courtesy of NASA/WMAP Science Team.

Chapter 19: Map of the Internet, as of September, 1998, created by Bill Cheswick. *Courtesy of* Lumeta Corporation 2000–2011. Published in Wired Magazine, December 1998 (issue 6.12).

Chapter 20: The Sloan Great Wall measured by J. Richard Gott and Mario Juric shows a wall of galaxies spanning 1.37 billion light years. It stands in the Guinness Book of Records as the largest structure in the universe. Courtesy of Michael Blanton and the Sloan Digital Sky Survey Collaboration, www.sdss.org.

Chapter 21: This map shows the patterns of human migration as inferred from modern geographical distributions of marker sequences in the Y chromosome (blue), indicating patrilineal inheritance, and in the mitochondrial DNA (orange), indicating matrilineal inheritance. Courtesy of National Geographic Maps, Atlas of the Human Journey.

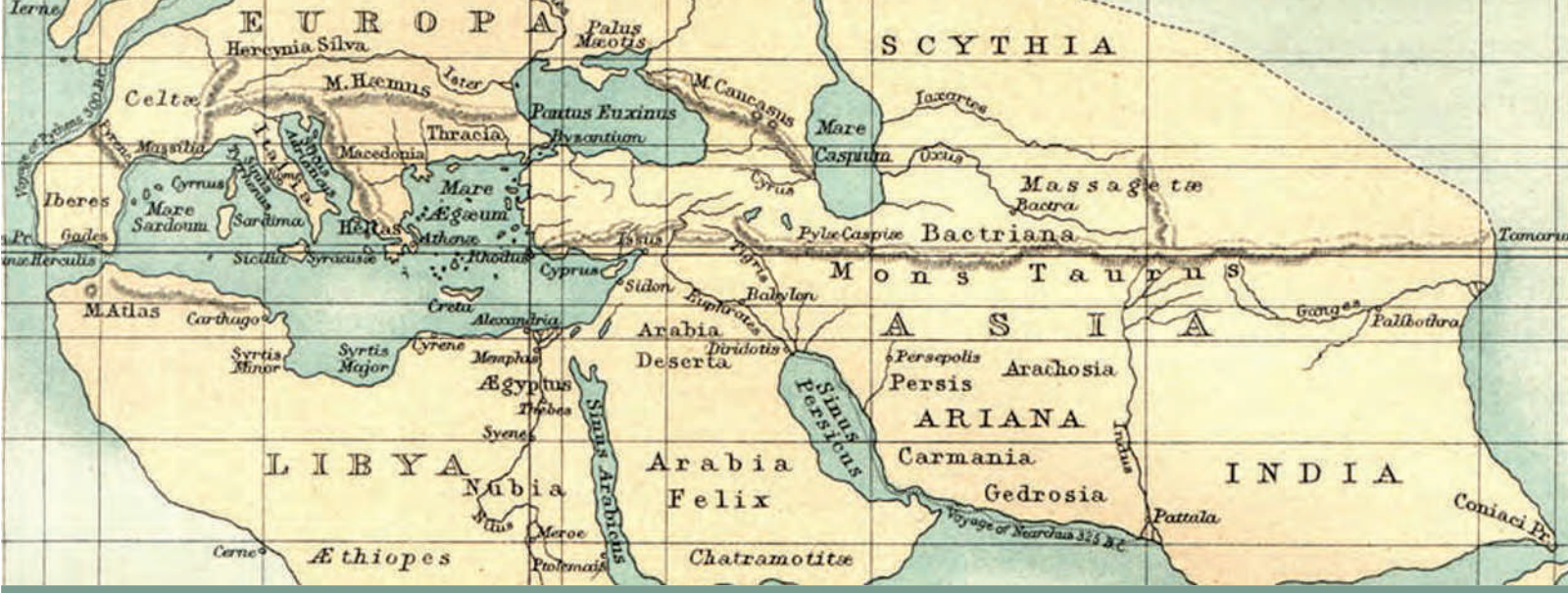
Chapter 22: "The Lands Beyond" drawn by Jules Feiffer for The Phantom Tollbooth (1961) by Norton Juster. Courtesy of Knopf Books for Young Readers, a division of Random House, Inc.



The Facts of Life

PART 1

This page intentionally left blank
to match pagination of print book



Why: Biology by the Numbers

1

Overview: In which the notion of biological numeracy is examined

Charles Darwin once said "All observation should be for or against some view if it is to be of any service," meaning that data are most meaningful against the backdrop of some conceptual or theoretical framework (Darwin, 1903, p. 194). The goal of this chapter is to develop a sense of the kinds of quantitative data that are being obtained in all fields of biology and the models that must be put forth to greet this data. Like with any useful map, we will argue that any good model has to overlook some of the full complexity and detail of a given biological problem in order to generate an abstraction that is simple enough to be easily grasped by the human mind, as an aid to developing intuition and insight. At the same time, it is critical that useful models make meaningful predictions, so they must include at least some of the realistic details of the biological system. The art of model building lies in striking the proper balance between too little detail and too much. Part of our emphasis on model building centers on the role of having a feeling for the numbers: sizes, shapes, times, and energies associated with biological processes. Here we introduce the style of making numerical estimates that will be used throughout the book.

> "To simplify is the greatest form of sophistication."
>
> **Attributed to Leonardo Da Vinci**

1.1 Biological Cartography

In one of the great stories in scientific history, Eratosthenes used the shadow of the noon-day sun at two different positions to work out a first, and surprisingly accurate, measurement of the radius of the Earth. Measurements like those of Eratosthenes made it possible to construct rudimentary maps of the world as it was then known to the Greeks with results such as that decorating the opening to this chapter. Such maps provide a distorted but recognizable view of the world.

In the chapters that follow, our aim is to similarly show how a few fundamental measurements and calculations suffice to provide the rough contours of what is clearly a much more nuanced biological landscape. The idea of such maps is to reveal enough detail to provide a conceptual framework that helps us navigate such biological landscapes, and serve as a starting point to orient ourselves for more in-depth exploration. In constructing such maps, we walk a perilous ridge between descriptions that are so distorted as to be uninformative and those that are so detailed that it is no longer correct to even think of them as maps. These ideas find their literary embodiment in one of Borges' shortest short stories, which serves as the battle cry for the chapters that follow.

> *On Exactitude in Science* . . .
> In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography. Suàrez Miranda: Viajes de varones prudentes, Libro Cuarto, cap. XLV, Lérida, 1658.
>
> Jorge Luis Borges and Adolfo Bioy Casares

1.2 Physical Biology of the Cell

With increasing regularity, the experimental methods used to query living systems and the data they produce are quantitative. For example, measurements of gene expression in living cells report how much of a given gene product there is at a given place at a given time. Measurements on signaling pathways examine the extent of a cellular response such as growth, DNA replication, or actin polymerization, as a function of the concentration of some upstream signaling molecule. Other measurements report the tendency of DNA to wrap into compact structures known as nucleosomes as a function of the underlying sequence. The list of examples goes on and on and is revealed by the many graphs depicting quantitative measurements on biological problems that grace the current research literature.

The overarching theme of the developing field of physical biology of the cell is that quantitative data demand quantitative models and, conversely, that quantitative models need to provide experimentally testable quantitative predictions about biological phenomena. What this means precisely is that in those cases where a biological problem has resulted in a quantitative measurement of how a particular biological output parameter varies as some input is changed (either because it is manipulated by the experimenter or because it varies as part of a natural biological process), the interpretation of that biological data should also be quantitative. In the chapters that follow, our aim is to provide a series of case studies involving some of the key players and processes in molecular and cell biology that demonstrate how physical model building can respond to this new era of quantitative data and sharpen the interplay between theory and experiment.

Model Building Requires a Substrate of Biological Facts and Physical (or Chemical) Principles

One of the first steps in building models of quantitative experiments like those described above is trying to decide which features of the problem are of central importance and which are not. However, even before this, we need to know facts about our system of interest. Throughout this book, we will assert a series of "facts" that fall into several different categories of certainty. We are most comfortable with statements describing observations that you as an individual observer can readily confirm, such as the statement that cells contain protein molecules. A second layer of facts with which we are nearly as comfortable are those arrived at and repeatedly confirmed by decades of experimentation using many different, independent experimental techniques. One fact of this kind is the statement that protein molecules within cells are synthesized by a large piece of machinery called the ribosome. A third category of "facts" that we will take care to mark as speculative are those that stand as compelling explanations for biological observations, but that may rest on unproven or controversial assumptions. In this category, we would include the commonly accepted proposition that modern-day mitochondria are the descendants of what was once a free-living bacterium that entered into a symbiotic relationship with an ancestral eukaryotic cell. In these cases, we will generally explain the observation and the interpretation that underlies our modeling approach. One of the most basic classes of facts that we will need throughout the remainder of the book concerns the nature of the molecules that make up cells and organisms. Because this will set up the common molecular language for the rest of our discussions, we will describe these molecules at some length in the next section.

Complementary to these biological observations, we embrace the proposition that biological entities cannot violate the laws of physics and chemistry. Many fundamental physical and chemical principles can be directly applied to understanding the behavior of biological systems, for example, the physical principle that the average energy of a molecule increases with increasing environmental temperature, and the chemical principle that oil and water do not mix. Along with biological facts described above, such principles will serve as the basis for the models to be described throughout the book.

1.3 The Stuff of Life

Scientists and other curious humans observe the natural world around them and try to make sense of its constituent parts and the ways that they interact. The particular subset of the natural world called life has always held special fascination. Any 3-year-old child knows that a rock is not alive but that a puppy, a tree, and indeed the child himself are alive. Yet the 3-year-old would be hard pressed to produce a rigorous definition of life and, indeed, scientists who have thought about this for decades can only approximate a description of this fascinating property. No single feature can reliably discriminate life from nonlife, but all living systems do share certain central properties. Living systems grow, consume energy from their environment, reproduce to form offspring that are rather similar to themselves, and die.

In the quest to understand the nature of life, scientists over hundreds of years have investigated the properties of living organisms

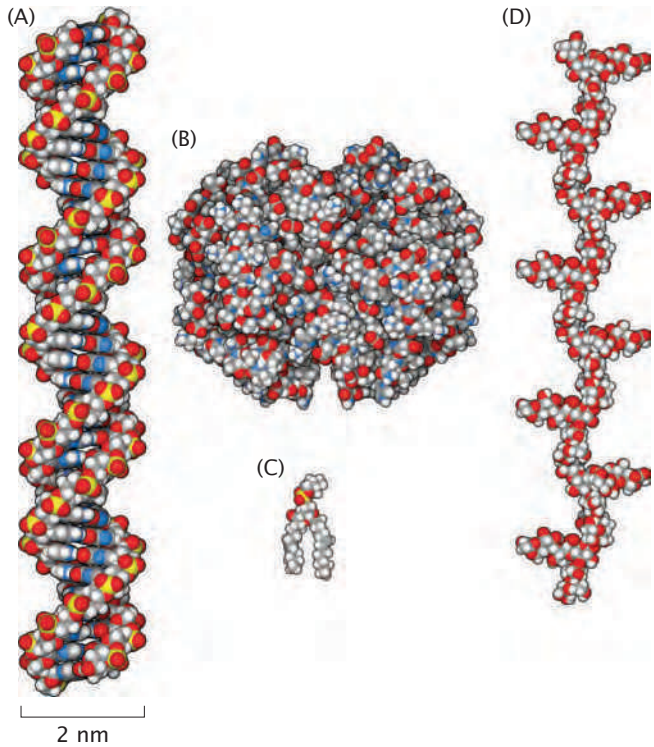


Figure 1.1: Atomic-level structural representation of members of each of the major classes of macromolecules, all drawn at the same scale. Nitrogen is colored in blue, oxygen in red, phosphorus in yellow, carbon in gray, and hydrogen in white. (A) Atomic structure of a small fragment of the nucleic acid DNA in the B form, (B) atomic structure of the oxygen-carrying protein hemoglobin (PDB 1hho), (C) phosphatidylcholine lipid molecule from a cell membrane, and (D) branched complex carbohydrate (M41 capsular polysaccharide) from the surface of the bacterium *Escherichia coli* (PDB 1cap). (Illustrations courtesy of D. Goodsell.)

and their constituent materials with an overarching goal of trying to understand why a puppy is different from a rock, even though they are both made of atoms that obey the same physical laws. Although this quest is still ongoing, some useful generalizations about the material nature of life are now widely accepted. For example, one observed feature of many of the molecules that make up living organisms is that they tend to be relatively large and structurally complex, and are hence called macromolecules. Living organisms also contain a large number of small, simple molecules that are critical to their function, ranging from water and metal ions to sugars such as glucose. Although these small molecules are important to life processes and, indeed, are used by the cell as building blocks to make the macromolecules, a distinction is usually drawn between small molecules that can arise through nonliving chemical processes and characteristic biological macromolecules that are found nowhere but in living organisms. Despite the diversity of macromolecules found in living organisms, the kinds of atoms found within them are surprisingly restricted, consisting mainly of carbon (C), oxygen (O), nitrogen (N), and hydrogen (H), with a smattering of sulfur (S) and phosphorus (P). It is largely the ways in which this humble suite of atoms are connected to each other that create the special properties of the macromolecules found in living systems. Of course, many other kinds of atoms and ions are critical for the biochemical processes of life, but they are generally not covalently incorporated in macromolecular structures.

Organisms Are Constructed from Four Great Classes of Macromolecules

If we take any organism, ranging from a small soil bacterium to a giant redwood tree or a massive whale, and separate out its constituent macromolecules, we will find a small number of distinct classes, chief among these being proteins, carbohydrates, lipids, and nucleic

acids. Figure 1.3 shows representative examples of each of these classes. Broadly speaking, these different classes of macromolecules perform different functions in living organisms. For example, proteins form structural elements within cells and are largely responsible for catalyzing specific chemical reactions necessary to life. Lipids form membrane barriers that separate the cell from the outside world and subdivide the cell into interior compartments called organelles. Carbohydrates are used for energy storage, for creating specific surface properties on the outside of cell membranes, and, sometimes, for building rigid structural units such as cell walls. Nucleic acids perform a critical function as the memory and operating instructions that enable cells to generate all the other kinds of macromolecules and to replicate themselves. Within each of these broad categories there is tremendous diversity. A single organism may have tens of thousands of structurally distinct proteins within it. Similarly, the membranes of cells can comprise hundreds of different lipid species.

Why are these particular classes of macromolecules the stuff of life? A complete understanding of the answer to this question would require knowledge of the specific chemical and physical conditions on the early Earth that gave rise to the first living cells. However, by studying living organisms on Earth today, it is seen that these molecules are wonderfully suited for the perpetuation of life for several reasons. First, each of these classes of molecules can be assembled by the cell from a small number of simpler subunit or precursor molecules. It is the combinatorial assembly of these simple subunits that gives rise to the tremendous structural diversity mentioned above. What this means is that a cell needs only a relatively restricted repertoire of chemical reactions to be able to synthesize these sets of subunits from the food in its environment. The fact that most cells on Earth make a living by consuming other living organisms is probably both a cause and a consequence of the fact that they all share very similar suites of these fundamental building blocks.

Nucleic Acids and Proteins Are Polymer Languages with Different Alphabets

The rest of this book will be largely devoted to exploring the special properties of biological macromolecules and the ways that they are used by cells, with the large-scale goal of understanding how the special properties of life may emerge from fundamental physical and chemical interactions. We will begin by discussing one specific and fundamentally important aspect of macromolecular behavior that is necessary for life as we know it. In particular, a great triumph of the study of molecular biology in the middle part of the twentieth century was the profound realization that the sequence of nucleic acid subunits in the cell's DNA is directly responsible for determining the sequence of amino acids in that same cell's proteins (that is, the discovery of the genetic code). The relationship between a DNA sequence and a protein sequence has been described by Francis Crick as comprising the "two great polymer languages" of cells.

As shown in Figure 1.2, both nucleic acids and proteins are built up from a limited alphabet of units. The nucleic acids are built up from an alphabet of four letters that each represent a chemically distinct "nucleotide" (A – adenine, G – guanine, T – thymine, and C – cytosine, for DNA). The chemical structure of these four nucleotides is shown in Figure 1.3. The protein language is constructed from an alphabet of 20

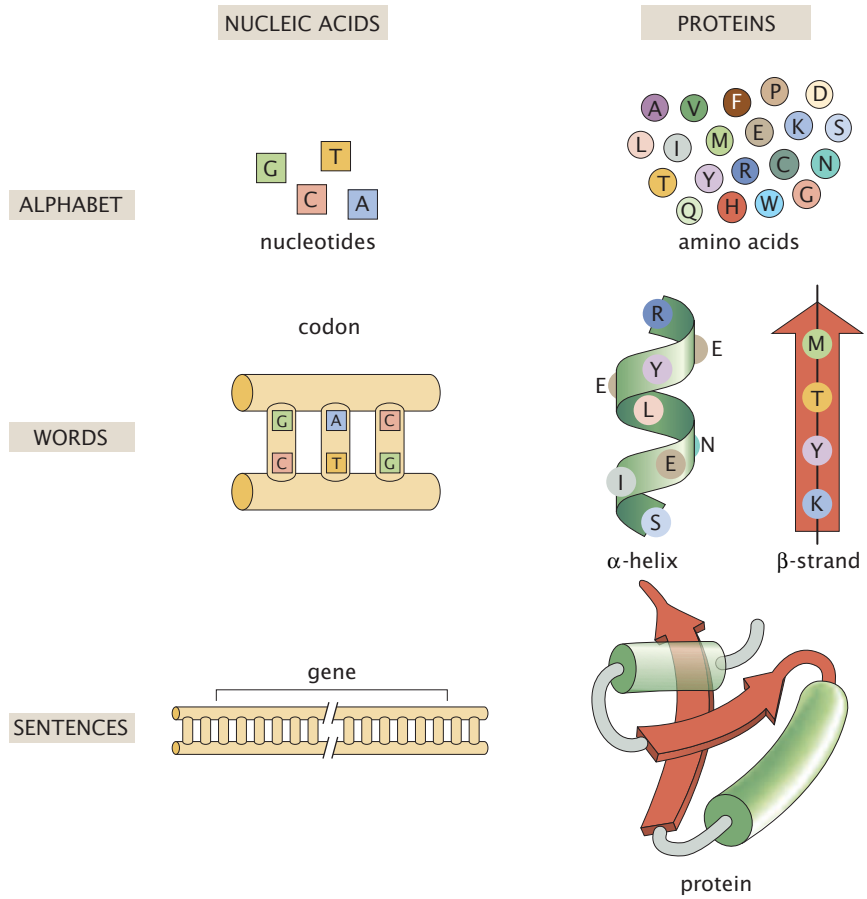


Figure 1.2: Illustration of Crick's "two great polymer languages." The left-hand column shows how nucleic acids can be thought of in terms of letters (nucleotides), words (codons), and sentences (genes). The right-hand column illustrates a similar idea for proteins, where the letters correspond to amino acids, the words to elements of secondary structure such as α helices and β strands, and the sentences to fully folded functional proteins.

distinct amino acids. How can just four nucleotide bases be translated into protein sequences containing 20 different amino acids? The key realization is that the DNA letters are effectively read as "words," called codons, made up of three sequential nucleotides, and each possible combination of three nucleotides encodes one amino acid (with some built-in redundancy in the code and a few exceptions such as the codon for "stop"). In the protein language, groups of amino acid "letters" also form "words," which give rise to the fundamental units of protein secondary structure, namely α helices and β strands, which we will discuss below. Finally, in the DNA language, sentences are formed by collections of words (that is, collections of codons) and correspond to genes, where we note that a given gene codes for a corresponding protein. An example of a complete sentence in the protein language is a fully folded, compact, biologically active enzyme that is capable of catalyzing a specific biochemical reaction within a living cell. Much of the rest of the book will be built around developing models of a host of biological processes in cells that appeal to the macromolecular underpinnings described here.

The sequences associated with nucleic acids and proteins are linked mechanistically through the ribosome, which takes nucleic acid sequences (in the form of messenger RNA (mRNA)) and converts them into amino acid sequences (in the form of proteins). These two polymer languages are linked informationally in the form of the genetic code. In particular, each three-letter collection from the nucleic acid language has a corresponding letter that it codes for in the protein language. The determination of this genetic code is one of the great chapters in the history of molecular biology and the results of that

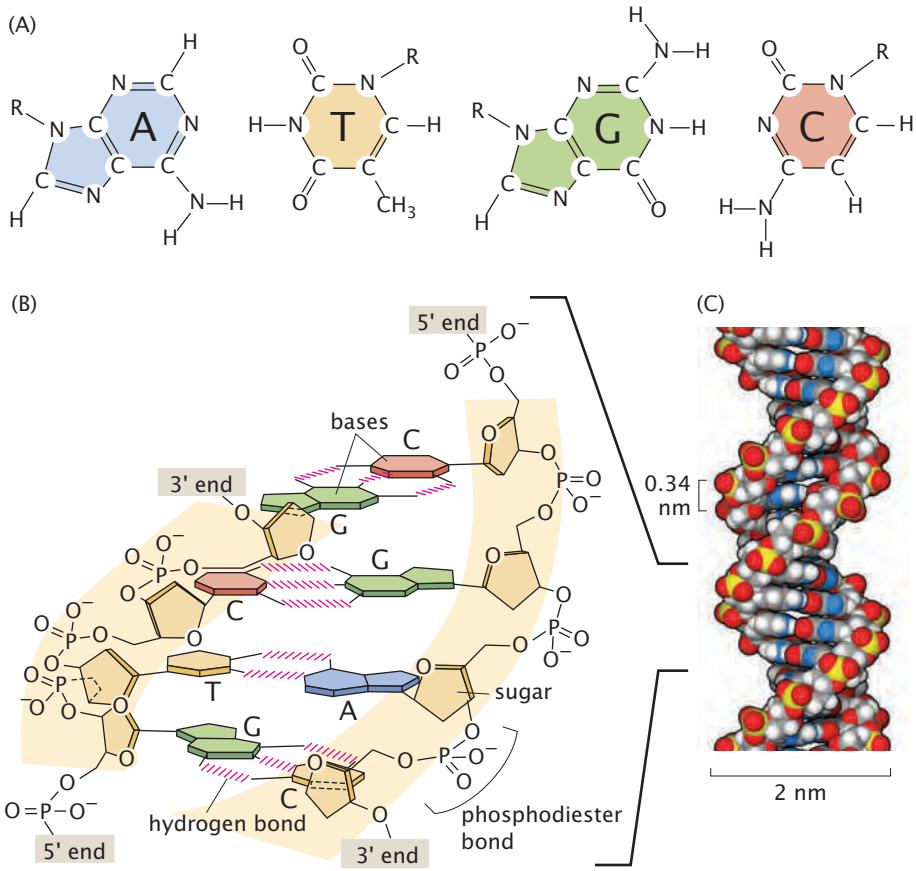


Figure 1.3: The chemical structure of nucleotides and DNA. (A) In DNA, the four distinct subunits or bases are abbreviated: A (adenine), T (thymine), G (guanine), and C (cytosine). In these diagrams, carbon is represented by C, oxygen by O, nitrogen by N and hydrogen by H. The letter R indicates attachment to a larger chemical group (the rest of the molecule); for nucleotides, the R group consists of the pentose sugar deoxyribose attached to phosphate. A single line connecting two atoms indicates a single covalent bond and a double line indicates a double covalent bond. The two large bases, A and G, are called purines and the two small bases, C and T, are called pyrimidines. (B) Illustration of how bases are assembled to form DNA, a double helix with two "backbones" made of the deoxyribose and phosphate groups. The four bases are able to form stable hydrogen bonds uniquely with one partner such that A pairs only with T and G pairs only with C. The structural complementarity of the bases enables the faithful copying of the nucleotide sequence when DNA is replicated or when RNA is transcribed. (C) Space-filling atomic model approximating the structure of DNA. The spacing between neighboring base pairs is 0.34 nm. (Adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed. Garland Science, 2008.)

quest are shown in Figure 1.4, which shows the information content recognized by the universal translating machine (the ribosome) as it converts the message contained in mRNA into proteins.

1.4 Model Building in Biology

1.4.1 Models as Idealizations

In trying to understand the nature of life, the approach that we take in this book is the same as that followed by scientists for generations. We will approach each system with the goal of gaining some useful insight into its behavior by abstracting and simplifying the highly complex materials of living organisms so that we can apply simple, analytical models that have some predictive value. By definition, we cannot retain a complete atomic description of each macromolecule. Instead, we aim to select only the relevant properties of the macromolecule that speak to the particular aspect of that molecule's behavior that we are trying to address. Therefore, it would be inaccurate to say that we have in mind "a" simple model for a complex macromolecule such as DNA. Instead, we will use a suite of simple models that can be thought of as projections of the complex, multifaceted reality of the DNA molecule into a variety of simpler DNA maps. In the same way, any useful map of a territory emphasizes only the features that will be most relevant to the person who will be using the map. A classic example of the extraction and elegant presentation of only the most useful features is Harry Beck's famous map of the London Underground system, which shows the

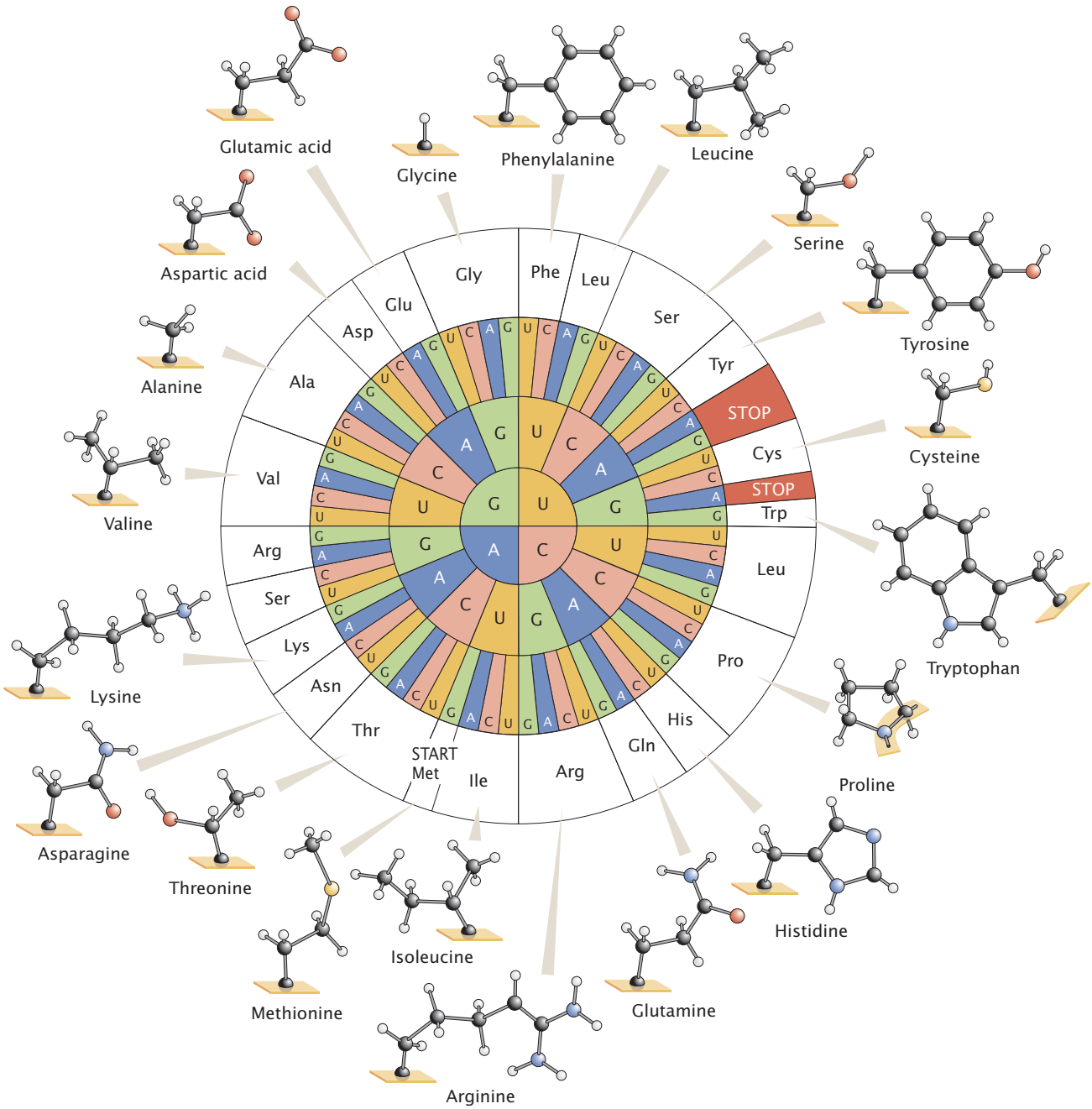


Figure 1.4: Genetic code. In this schematic representation, the first nucleotide in a coding triplet is shown at the center of the ring, the second nucleotide in the middle colored ring, and the third nucleotide in the outer colored ring. In this representation of the genetic code, the four bases are adenine (A), cytosine (C), guanine (G), and uracil (U). Uracil is structurally very similar to thymine (T), and is used instead of thymine in messenger RNA. The amino acids corresponding to each group of triplets are illustrated with their names (outer ring) and atomic structures. Two amino acids, tryptophan and methionine, are encoded by only a single triplet, whereas others, including serine, leucine, and arginine, are encoded by up to six. Three codons do not code for any amino acid and are recognized as stop signals. The unique codon for methionine, AUG, is typically used to initiate protein synthesis.

stations and connecting lines with an idealized geometry that does not accurately represent the complex and rather frustrating topography of London's streets. Although certain features of London's geography are deliberately ignored or misrepresented, such as the actual physical distances between stations, the London Underground map shows the user exactly enough information about the connectivity of the transport system to go anywhere he wishes. In this section, we will

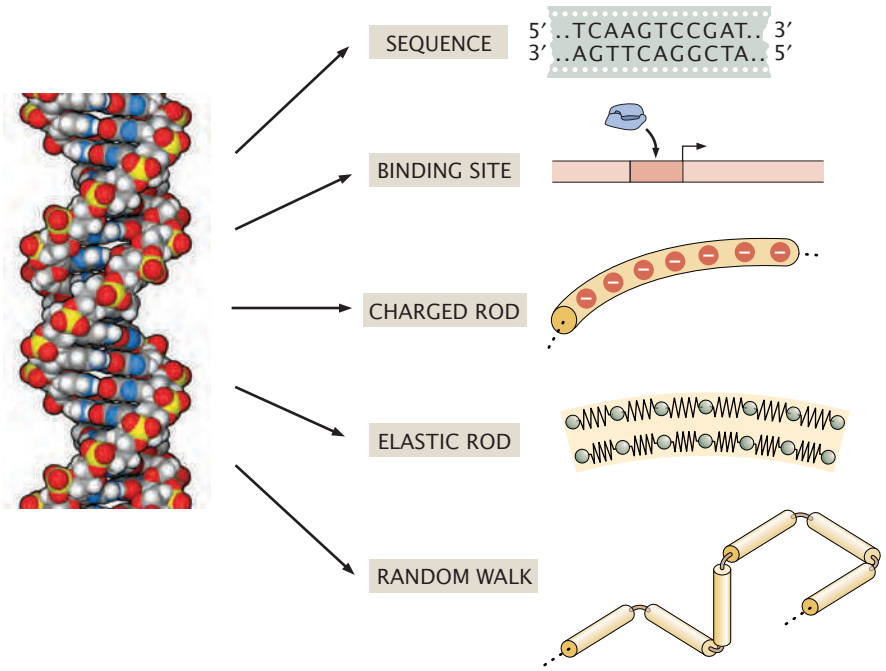


Figure 1.5: Idealizations of DNA. DNA can be thought of as a sequence of base pairs, as a series of binding sites, as a charged rod, as an elastic rod, or as a freely jointed polymer arranged in a random walk, depending upon the problem of interest.

summarize the kinds of physical properties of biological materials that we may wish to consider in isolation. In Section 1.5, we will list the kinds of fundamental physical models that will be applied throughout the book.

Biological Stuff Can Be Idealized Using Many Different Physical Models

Often, the essence of building a useful and enlightening model is figuring out what features of the problem can be set aside or ignored. Throughout the book, we will come to view molecules and cells from many different angles. An example of this diversity of outlook is shown in Figure 1.5. The nature of the simplified space, and therefore the nature of the projected DNA model, depends upon the specific question that we are asking. Figure 1.5 shows five different projections for the meaning of DNA that we will use repeatedly throughout the book.

DNA is one of the most important molecules, both because it is the carrier of genetic information and because of its iconic role in the history of molecular biology. There is a polarity to each strand of this helical polymeric molecule; that is, the two ends of the strand are structurally and chemically distinct, and are called the 5′ and 3′ ends. In addition, there is heterogeneity along their length, since the chemical identity of the bases from one nucleotide to the next can be different. It is often the informational content of the DNA molecule that garners attention. Indeed, the now-routine sequencing of whole genomes of organisms ranging from bacteria to humans permits us to forget the molecular details of the DNA molecule and to focus just on the sequence of A's, T's, G's, and C's that determines its information content. This representation will be most useful, for example, when we are trying to discern the history of mutational events that have given rise to diversification of species or when we are trying to calculate the information content of genomes. The second representation shown in Figure 1.5 abstracts variations in DNA sequence

in large blocks where one particular stretch of sequence (shown in the figure as the more darkly colored block) represents a portion of sequence that has particularly strong affinity for a protein binding partner, such as the RNA polymerase enzyme that copies the sequence encoded by the DNA into a molecule of mRNA. Both these first two representations draw attention to dissimilarities in detailed chemical structure within the DNA molecule. However, sometimes we will consider DNA simply as a physical entity within the cell where the specific nature of its sequence is not relevant. Instead, other properties of the molecule will be emphasized, such as its net distribution of charge, bending elasticity, or behavior as a long polymer chain subject to thermal undulations. Schematics of models emphasizing these physical properties of DNA are also shown in Figure 1.5.

Proteins are subject to many different idealizations as well. Just as nucleic acids can be thought of as sequences of nucleotides, we can think of proteins in terms of the linear sequence of amino acids of which they are made. The second great polymer language gives rise to proteins, which are considered to provide most of the important biological functions in a cell, including the ability to catalyze specific chemical reactions. The detailed atomic structure of proteins is substantially more complex than that of DNA, both because there are 20 rather than 4 fundamental subunits and because the amino acid chain can form a wider variety of compact three-dimensional structures than can the DNA double helix. Even so, we will extract abstract representations of proteins emphasizing only one feature or another as shown in Figure 1.6 in much the same way we did for DNA.

A protein can be represented as a simple linear sequence of amino acids. In cases where this representation contains too much detail, we will elect to simplify further by grouping together all hydrophobic amino acids with a common designation H and all polar amino acids with a common designation P. In these simple models, the physical differences between hydrophobic and polar amino acids (that is, those that are like oil versus those that are like water) affect the nature of their interactions with water molecules as well as with each other and thereby dictate the ways that the amino acid chain can fold up to make the functional protein. Similarly, we can think of the three-dimensional folded structure of proteins at different levels of abstraction depending on the question under consideration. A common simplification for protein structure is to imagine the protein as a series of cylinders (α-helices) and ribbons (β-strands) that are connected to one another in a defined way. This representation ignores much of the chemical complexity and atomic-level structure of the protein but is useful for considering the basic structural elements. An even more extreme representation that we will find to be useful for thinking about protein folding is a class of lattice models of compact polymers in which the amino acids are only permitted to occupy a regular array of positions in space.

Fundamentally, the amino acid sequence and folded structure of the protein serve a biological purpose expressed in protein function or activity, and we will also develop classes of models that refer only to the activity of the molecule. For example, the vast majority of proteins in the cell are capable of forming specific binding interactions with one another or with other biomolecules; these binding events are usually important to the protein's biological function. When modeling these kinds of interactions, we will envision proteins as "receptors" carrying binding sites that may be occupied or unoccupied by a binding partner, called a "ligand," and our description will make no further

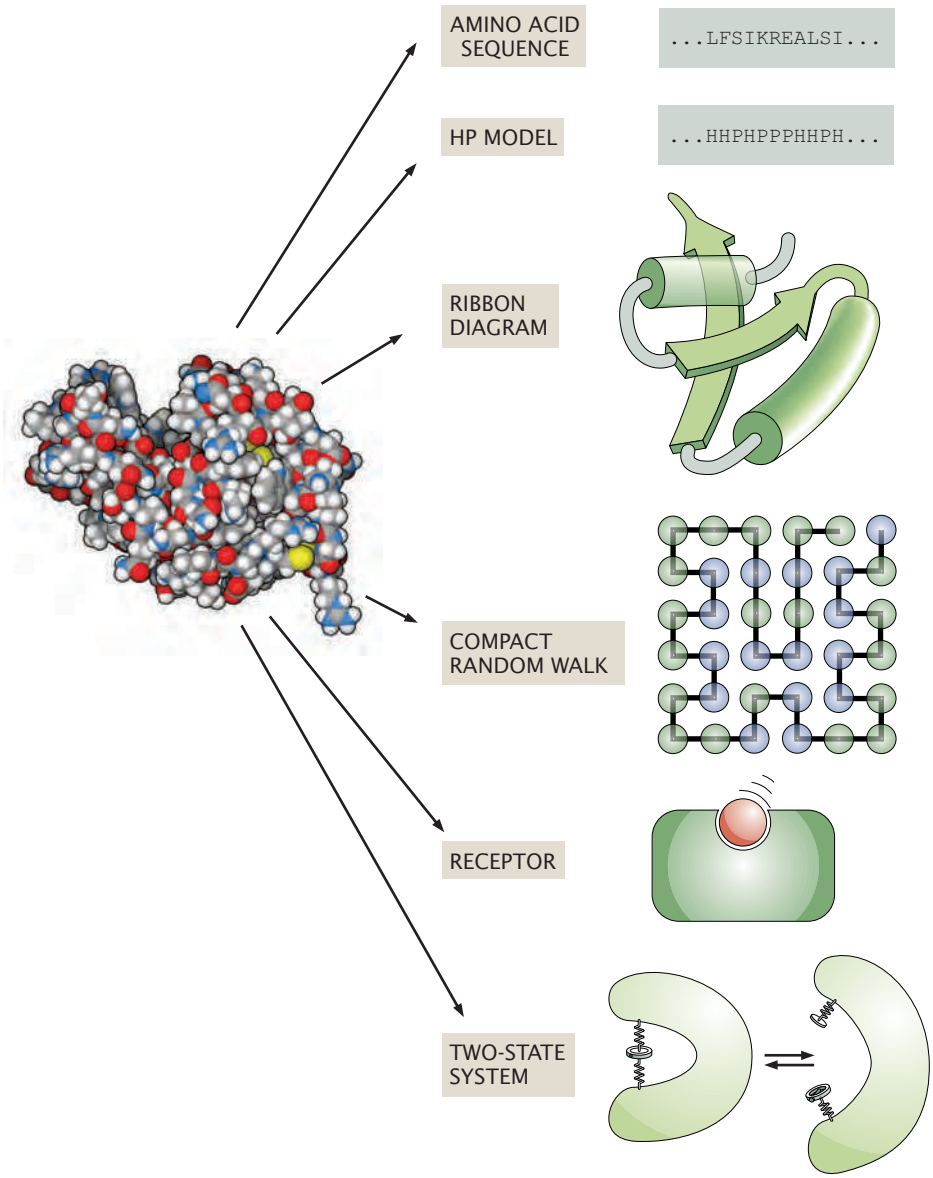


Figure 1.6: Idealizations of protein. Proteins can be thought of as a particular sequence of amino acids, as a simplified sequence reporting only the hydrophobic (H, oil-like) or polar (P, water-like) chemical character of the amino acids, as a collection of connected ribbons and cylinders, as a compact polymer on a lattice, as a binding platform for ligands, or as a two-state system capable of interconverting between different functional forms.

reference to the protein's complex internal structure. For some models, we will focus on changes in activity of proteins, considering them as simple two-state systems that can interconvert between two different functional forms. For example, an enzyme that catalyzes a particular biochemical reaction may exist in an "active" or "inactive" state, depending on the presence of ligands or other modifications to its structure.

By performing these simplifications and abstractions, we are not intending to deny the extraordinary complexity of these molecules and are fully cognizant that amino acid sequence, folding and compaction properties, and conformational changes between states dictate binding affinity, and it is an artificial simplification to consider these properties in isolation from one another. Nonetheless, it is generally useful to consider one single level of description at a time when trying to gain intuition from simple estimates and models, which will serve as our primary emphasis.

We will extend this approach beyond individual macromolecules to large-scale assemblies of macromolecules such as cell membranes

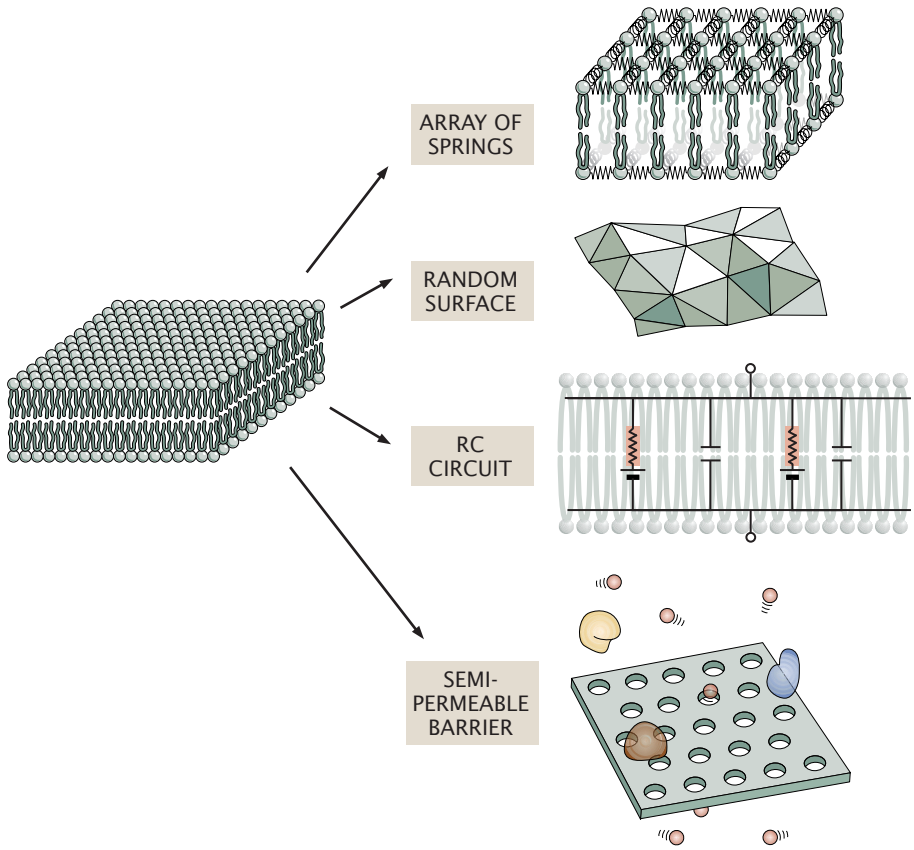


Figure 1.7: Idealization of membranes. A membrane can be modeled as an elastic object that deforms in response to force, as a random surface fluctuating as a result of collisions with the molecules in the surrounding medium, as an electrical circuit element, or as a barrier with selective permeability.

made up of millions of individual lipid and protein molecules. Some of the different ways in which we will think about biological membranes are shown in Figure 1.7. Just as with individual macromolecules, these large-scale assemblies can be usefully characterized by focusing on one aspect of their physical behavior at a time. For example, membranes can be thought of as bendable, springy elastic sheets, as random surfaces, or even as arrays of electrical elements such as resistors and capacitors. In their interactions with other molecules, membranes often serve as selective barriers that will enable some molecular species to cross while blocking others. Depending upon the particular question being asked, we will exploit different idealizations of the membrane.

Audaciously, we will bring these same idealization techniques to bear even on living cells, as shown in Figure 1.8. Although a living cell is exceedingly complex, we will nonetheless choose to extract one physical property of the cell at a time so that we can gain insight from our simple approach. For example, the bacterium *Escherichia coli*, a beloved experimental organism, can for some purposes be thought of as an object carrying an array of protein receptors on its surface. For other purposes, for example when we think of how a bacterium swims through water, we will think of it as an elastohydrodynamic object, that is, a mechanical object that can bend and can interact with the flow of water. The physical properties of the path it follows through water can be further abstracted by considering the cell's large-scale motion in the context of a biased random walk, ignoring the hydrodynamic details that actually enable that motion. Finally, because much biological research focuses on how cells alter the expression of their polymer languages in response to changing conditions, we

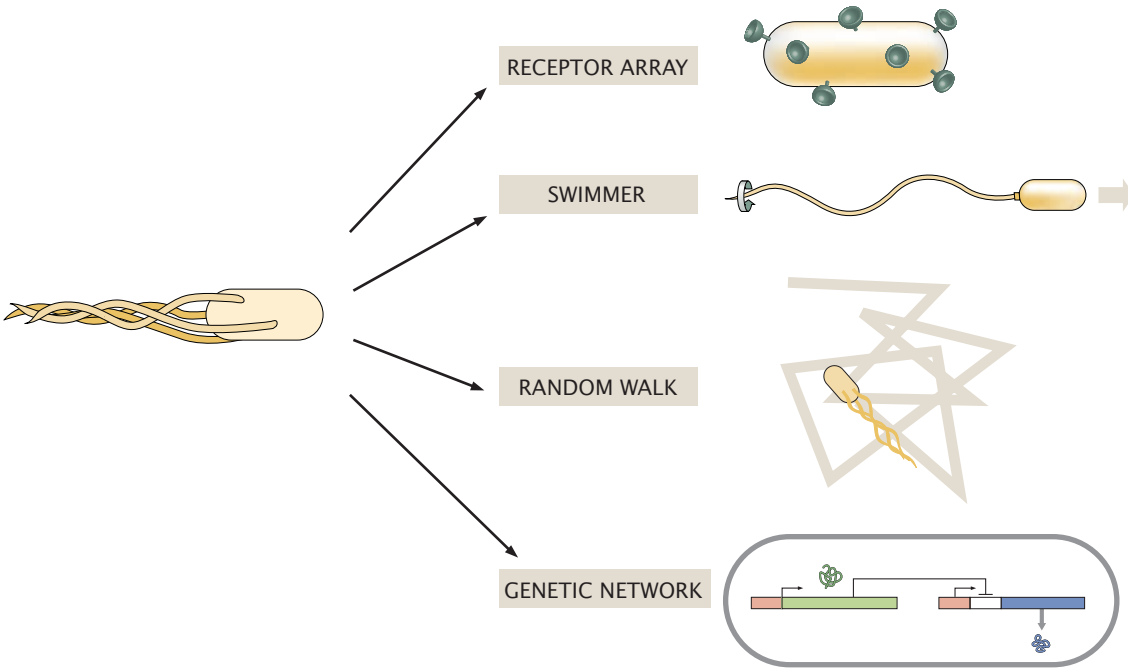


Figure 1.8: *E. coli* idealization. The cell can be modeled as an array of receptors for a ligand of interest, as an elastohydrodynamic object, as a biased random walker moving through water, or as an information processing device characterized by a series of genetic networks.

will frequently find it useful to consider the cell as an information processing device, for example as a network governing the flow of information in the form of gene expression. These descriptions are not mutually exclusive, nor are any of them comprehensive. They are again deliberate abstractions made for practical reasons.

Although so far we have emphasized the utility of abstract projections centering on a single physical property for descriptions of biological entities, the same approach is actually extremely useful in consideration of other systems such as solutions of charged ions in water. Various abstractions of the watery medium of life are shown in Figure 1.9. Sometimes we will pretend that a solution is a regular lattice. For example, if we are interested in ligands in solution and their tendency to bind to a receptor, we adopt a picture of the solution in which the ligands are only permitted to occupy specific discrete positions. Although this is an extreme approximation, it is nonetheless immensely valuable for calculations involving chemical equilibrium and remarkably gives accurate quantitative predictions. As the watery medium interacts with the living creatures it contains, sometimes its hydrodynamic flow properties (or viscous properties) are of most interest. When considering rates of chemical reactions taking place in water, we will find it useful to think of water as the seat of diffusive fluctuations. The macromolecules that exist in the watery medium of cells also interact with specific properties of water. For example, some of the chemical groups in proteins are polar or "hydrophilic" (water-loving), meaning that they are able to form hydrogen bonds with water, while other groups are oil-like and repel water. In the folded protein, the oil-like portions tend to cluster on the inside. In contrast, charged molecules on the surface can interact with the dielectric character of water molecules, which are neutral overall but do exhibit a slight charge separation.

Each of these representations will resurface multiple times throughout the book in different contexts related to different biological questions. Although none of them can give a complete understanding

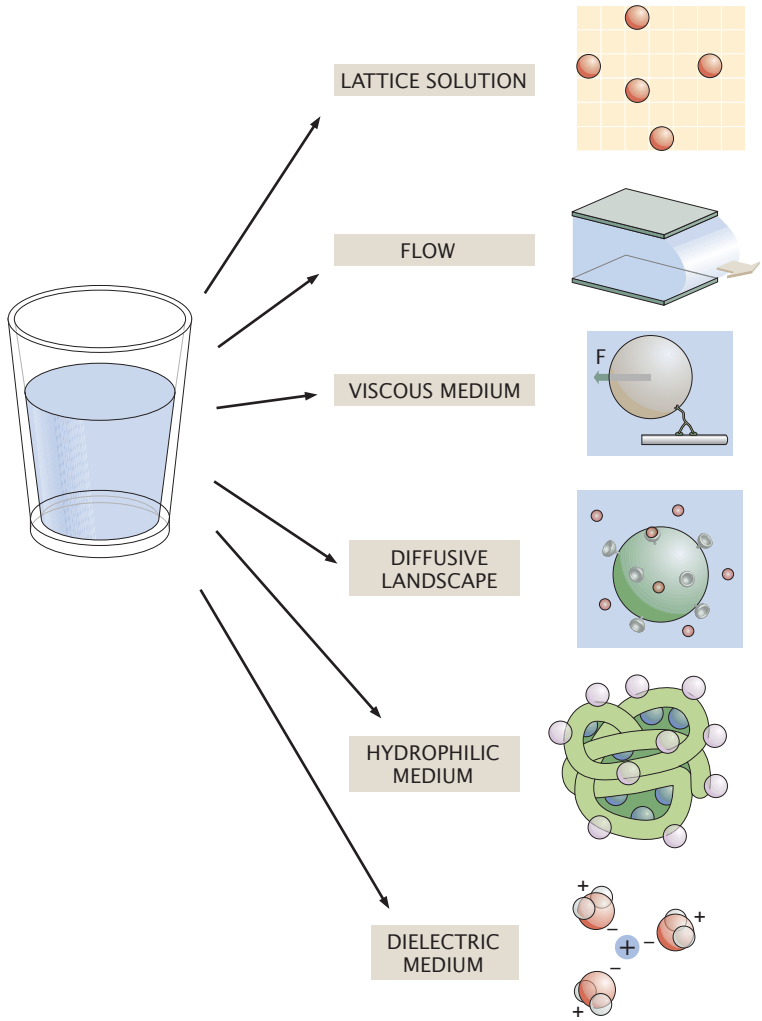


Figure 1.9: Idealization of a solution. A solution can be represented as a lattice of discrete positions where solutes might exist, as a fluid with a mean flow, as a viscous medium that exerts a drag force on objects moving through it, as a fluctuating environment that induces random motion of macromolecules and ligands, as a hydrophilic medium that readily dissolves polar molecules but repels oil-like molecules, or as a dielectric medium that is a poor electrical conductor but can support electrostatic fields.

of the behavior of living cells, each of them serves to provide quantitative insight into some aspect of life and, taken as a whole, they can begin to project a more realistic image of cells than any single abstraction can do alone.

1.4.2 Cartoons and Models

Biological Cartoons Select Those Features of the Problem Thought to Be Essential

We have argued that the art of model building ultimately reflects a tasteful separation of that which is essential for understanding a given phenomenon from that which is not. As is true of all sciences, the history of experimental progress in biology has always involved researchers constructing explicit conceptual models to help them make sense of their data. In some cases, the theoretical framework developed for biological systems is of precisely the mathematical character that will be presented throughout the book, such as that illustrated in the use of thermodynamics and statistical mechanics to understand biochemical reactions or in the development of the theory of the action potential. More commonly, much of the important modeling of biological systems has been in the form of visual schematics

that illustrate the most essential features in a given biological process and how they interact. In fact, the act of drawing the type of cartoon found in molecular biology textbooks or research papers reflects an important type of conceptual model building and either implicitly or explicitly reflects choices about which features of the problem are really important and which can be ignored.

As an example of the model building that takes place in constructing biological cartoons, consider the structure of one of the most beautiful and intriguing eukaryotic organelles, namely, the mitochondrion. Figures 1.10(A) and (B) show a planar section of a mitochondrion as obtained using electron microscopy and part of a three-dimensional structure resulting from cryo-electron microscopy. Figure 1.10(C) shows two cartoons that describe the structure of this important organelle. This pair of drawings emphasizes the central importance of the membranes which bound the organelle and which separate the interior into different, relatively isolated compartments. Though the three-dimensional model represents more detail regarding the precise geometry of the membranes, the essential *conceptual* elements are the same, namely, (a) the mitochondria are closed, membrane-bound organelles and (b) the inner membrane has a complex folded structure. Furthermore, the folds of the inner membrane greatly increase the total amount of surface area of this membrane and therefore increase the area available for the ATP-synthesizing machinery that lies within this membrane, as will be discussed below.

The other cartoons in this figure also represent the mitochondrion, but in very different ways, focusing on specific conceptual elements involved in different aspects of mitochondrial function and behavior. For example, Figure 1.10(D) reveals the surprising existence of mitochondrial DNA, which is a vestige of the microbial origins of this important organelle. Interestingly, these organelles contain their very own DNA genomes, completely distinct from the genome found in the cell's nucleus. As indicated in the diagram, the genome is a circular piece of DNA (like the genome found in bacterial cells, but very different from eukaryotic chromosomes, which are linear pieces of DNA). Each mitochondrion contains several copies of the genome, and they are distributed between the two daughters when the mitochondrion divides.

The diagram provided in Figure 1.10(E) shows a highly magnified segment of the mitochondrial inner membrane, illustrating the proteins that cross the membrane and that are involved in the generation of ATP. A central function of mitochondria in eukaryotic cells is to convert one form of chemical energy, high-energy electrons in the compound NADH, into a different form of chemical energy, a phosphodiester bond in ATP (these different forms of chemical energy and their significance will be discussed in more detail in Chapter 5). The process of performing this conversion requires the cooperative action of several different large protein complexes that span the inner mitochondrial membrane. Sequential transfer of electrons from one protein complex to another results in the transport of hydrogen ions from the mitochondrial matrix across the inner membrane into the intermembrane space, establishing a gradient with an excess of hydrogen ions on the outside. In the final step, the hydrogen ions are allowed to travel back down their concentration gradient into the mitochondrial matrix, by passing through a remarkable protein machine called ATP synthase, which we will revisit several times throughout the rest of the book. ATP synthase catalyzes the construction of a molecule of ATP by combining a molecule of ADP and an

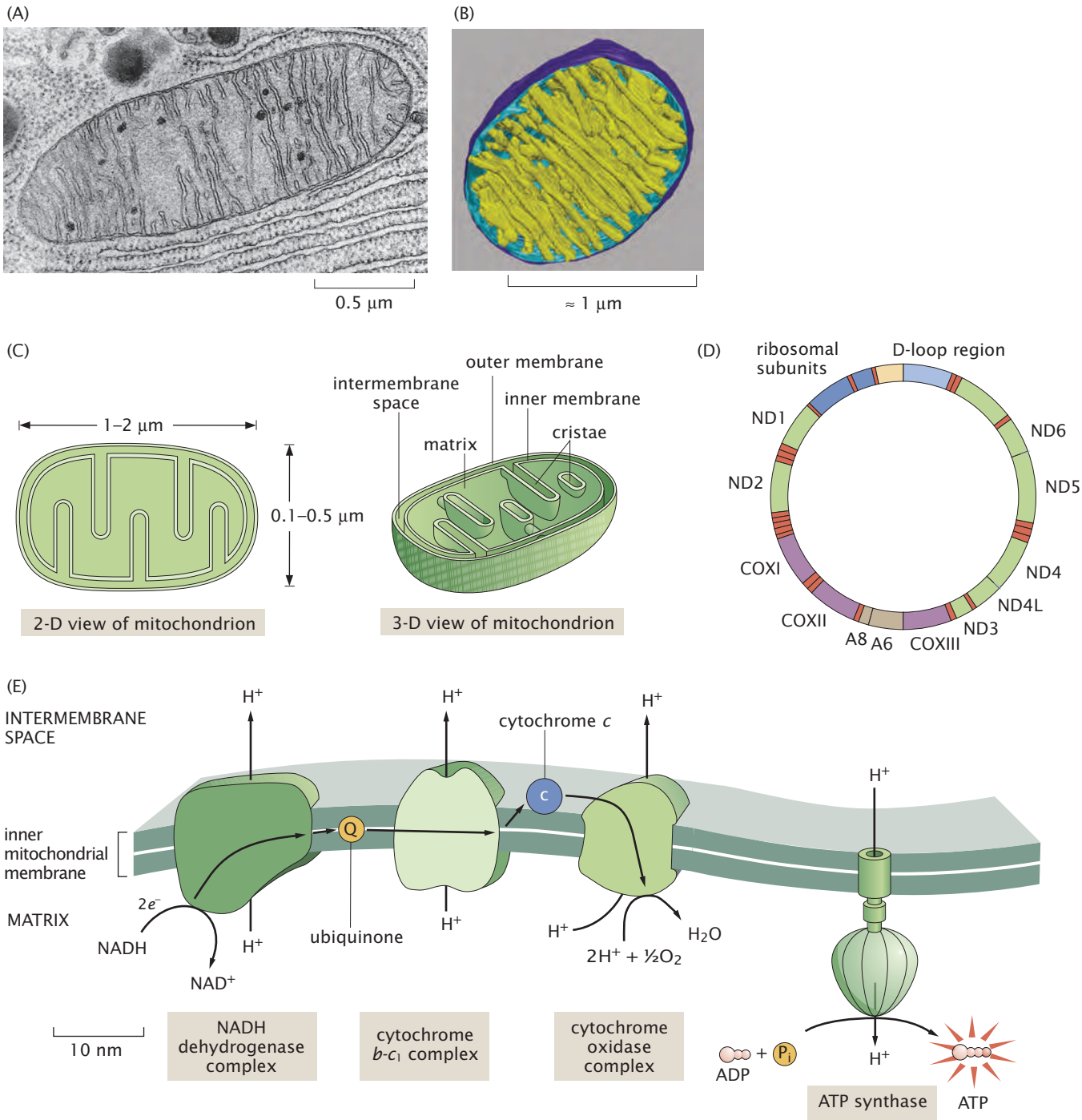


Figure 1.10: Several different ways of illustrating mitochondria. (A) Thin-section electron micrograph showing a mitochondrion found in a cell within the pancreas of a bat. (B) Cryo-electron microscopy reconstruction of the structure of a mitochondrion, shown cut in half to reveal the internal structure. (C) Diagrams showing the arrangement of membranes dividing the mitochondrion into distinct compartments. While the outer membrane forms a smooth capsule, the inner membrane is convoluted to form a series of cristae. Distinct sets of proteins are found in the matrix (inside the inner membrane) and in the intermembrane space (between the inner and outer membranes). (D) Schematic of mitochondrial DNA. (E) Schematic illustration of the major proteins involved in the electron transport chain and in ATP synthesis in the inner membrane of the mitochondrion. The overall purpose of the complex chemical reactions carried out by this series of proteins is to catalyze the creation of the important energy carrier molecule ATP. (A, from D. W. Fawcett, The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure, W. B. Saunders, 1966; B, courtesy of T. Frey; D, adapted from S. DiMauro and E. A. Schon, *N. Engl. J. Med.* 348:2656, 2003; E, adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed., Garland Science, 2008.)

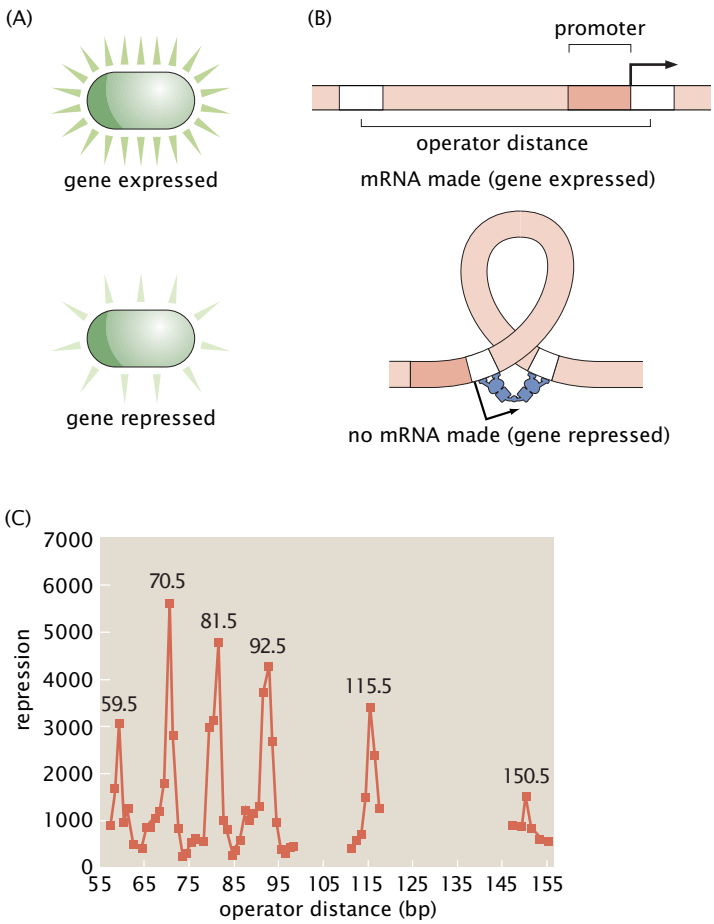


Figure 1.11: Qualitative and quantitative illustrations of gene repression. (A) Bacteria containing a gene whose expression level is regulated by the Lac repressor protein may exist in different states. When repression is low, the gene is expressed, and large amounts of the gene product are present. If the gene being regulated encodes a fluorescent protein, the bacteria will glow brightly when the gene is expressed. In contrast, when the gene is repressed, relatively little fluorescent protein will be produced, and the bacteria will glow dimly. (B) Schematic showing events at the level of the DNA during gene expression and repression. During repression, the Lac repressor protein binds to two distinct operator sites on the DNA (white boxes) and forces the DNA to form a loop. This prevents binding of the protein machine that copies the DNA into mRNA. When no mRNA is produced, the gene is repressed. (C) Quantitative measurements of the strength of gene repression as a function of the exact distance (expressed in base pairs) between the two operator sites. (C, data from J. Müller et al., *J. Mol. Biol.* 257:21, 1996.)

inorganic phosphate ion. The precise molecular details of this process are fascinating and complex; our intention here is merely to point out that this kind of cartoon represents an abstraction of the mitochondrion that emphasizes a completely different set of components and concepts than the membrane-focused cartoons shown above.

Each of these different kinds of cartoon is a valid and informative model of certain aspects of the mitochondrion, but they serve essentially nonoverlapping conceptual purposes. Furthermore, each cartoon represents the culmination of thousands of separate experiments covering decades of hard-won, detailed knowledge; experts have carefully sifted the raw data and made difficult decisions about which details are important and which are dispensable in considering each of these aspects of mitochondrial reality. Throughout the rest of this book, we will take advantage of the hard work of conceptualization that has already gone into constructing these kinds of diagrams. Our usual goal will be to take these existing biological models one step further, by rendering them into a mathematical form.

Quantitative Models Can Be Built by Mathematicizing the Cartoons

One of the key mantras of the book is that the emergence of quantitative data in biology generates a demand that the cartoons of molecular biology be mathematicized. In particular, in some cases only by constructing a mathematical model of a given biological problem can our

theoretical description of the problem be put on the same footing as the data itself, that is, quantitatively. A concrete example that will arise in Chapter 19 is the measurement of gene expression as a function of the distance between two binding sites on DNA as shown in Figure 1.11. The mechanistic basis of these data is that a certain protein (Lac repressor) binds simultaneously at two sites on the DNA and forms a loop of the DNA between the two binding sites. When the Lac repressor protein is bound to the DNA, it prevents the binding of the protein machine that copies the DNA into RNA. The strength of this effect (repression) can be measured quantitatively. The experimental data show that the amount of repression depends in a complex way on the precise distance between the two binding sites.

Our argument is that a cartoon-level model of this process, while instructive, provides no quantitative basis for responding to these data. For example, why are there clear periodic peaks in the data? Why does the height of the peaks first increase and then decrease? In Chapter 19, using the framework of statistical mechanics, we will derive a mathematical model for this process that attempts to predict the shapes and sizes of these features in the data while at the same time highlighting some features of DNA mechanics that remain unclear. Further, by constructing a physical model in mathematical terms of this process, new experiments are suggested that sharpen our understanding of transcriptional regulation. In nearly every chapter, we will repeat this same basic motif: some intriguing quantitative observation on a biological system is described first in terms of a cartoon, which is then recast in mathematical form. Once the mathematical version of the model is in hand, we use the model to examine previous data and to suggest new experiments.

As a result of examples like this (and many others to be seen later), one of the central arguments of the present chapter is that there are many cartoons in molecular biology and biochemistry that have served as conceptual models of a wide variety of phenomena and that reflect detailed understanding of these systems. To make these models quantitatively predictive, they need to be cast in mathematical form; this in keeping with our battle cry that quantitative data demand quantitative models. As a result, one way of viewing the chapters that follow is as an attempt to show how the cartoons of molecular biology and biochemistry have been and can be "translated into a mathematical form."

1.5 Quantitative Models and the Power of Idealization

Quantitative models of the natural world are at the very heart of physics and in the remainder of this book we will illustrate how pervasive these ideas have become in biology as well. This approach is characterized by a rich interplay between theory and experiment in which the consequences of a given model are explored experimentally. In return, novel experimental results beckon for the development of new theories. To illustrate the way in which particular key models are recycled again and again in surprisingly varied contexts, we consider one of the most fundamental physical models in all of science, namely, the simple spring (also known as the harmonic oscillator).

1.5.1 On the Springiness of Stuff

The concept of springiness arises from the confluence of the very important mathematical idea of a Taylor series (explained in detail in "The Math Behind the Models" on p. 215) and an allied physical idea known as Hooke's law (the same Hooke who ushered in the use of microscopy in biology and coined the term "cell"). These ideas will be developed in detail in Chapter 5 and for now we content ourselves with the conceptual framework. The fundamental mathematical idea shared by all "springs" is that the potential energy for almost any system subjected to small displacements from equilibrium is well approximated by a quadratic function of the displacement. Mathematically, we can write this as

$$\text{energy} = \tfrac{1}{2}kx^2. \tag{1.1}$$

What this equation states is that the potential energy increases as the square of the displacement x away from the equilibrium position. The "stiffness" k is a measure of how costly it is to move away from equilibrium and reflects the material properties of the spring itself.

Another way of thinking about springs is that they are characterized by a restoring force that is proportional to how far the spring has been displaced from its equilibrium position. Mathematically, this idea is embodied in the equation

$$F = -kx, \tag{1.2}$$

where F is the restoring force, x is the displacement of the spring from its equilibrium position, and k is the so-called spring constant (the stiffness). The minus sign signals that the force is a restoring one towards $x = 0$, which designates the equilibrium position. This result is known as Hooke's law. When stated in this way, this kind of physical model gives the impression of an abstract example of masses on frictionless tables with pulleys and springs. But empirically, scientists have found that precisely this mathematical model arises naturally in many different practical contexts, as shown in Figure 1.12.

In biology, the simple spring comes cloaked in many different disguises, as we will see throughout the book. From a technological perspective, many of the key single-molecule techniques used to study macromolecules and their assemblies invoke this description. Both optical tweezers and the atomic force microscope can be mapped precisely and unequivocally onto spring problems. However, as shown in Figure 1.12, there are many other unexpected examples in which the problem can be recast in a way that is mathematically equivalent to the simple spring problem introduced above. Simple spring models have been invoked when considering the bending of DNA, the beating of the flagellum on a swimming sperm, and even the fluctuations of membranes at the cell surface. As these are all mechanical processes, it is fairly straightforward to see how the idea of a spring can apply to them. But the real power of the simple spring model will be revealed when we see how it can be directly applied to nonmechanical problems. For example, when we discuss biochemical reactions in a cell, we will imagine molecules moving on an "energy landscape" where they accumulate in "potential wells." The mathematics of simple springs will help us understand the rates that govern these biochemical reactions. The ideas of harmonic oscillators and potential wells will also illuminate our exploration of protein conformational changes. Even more abstractly, we will see that changes in

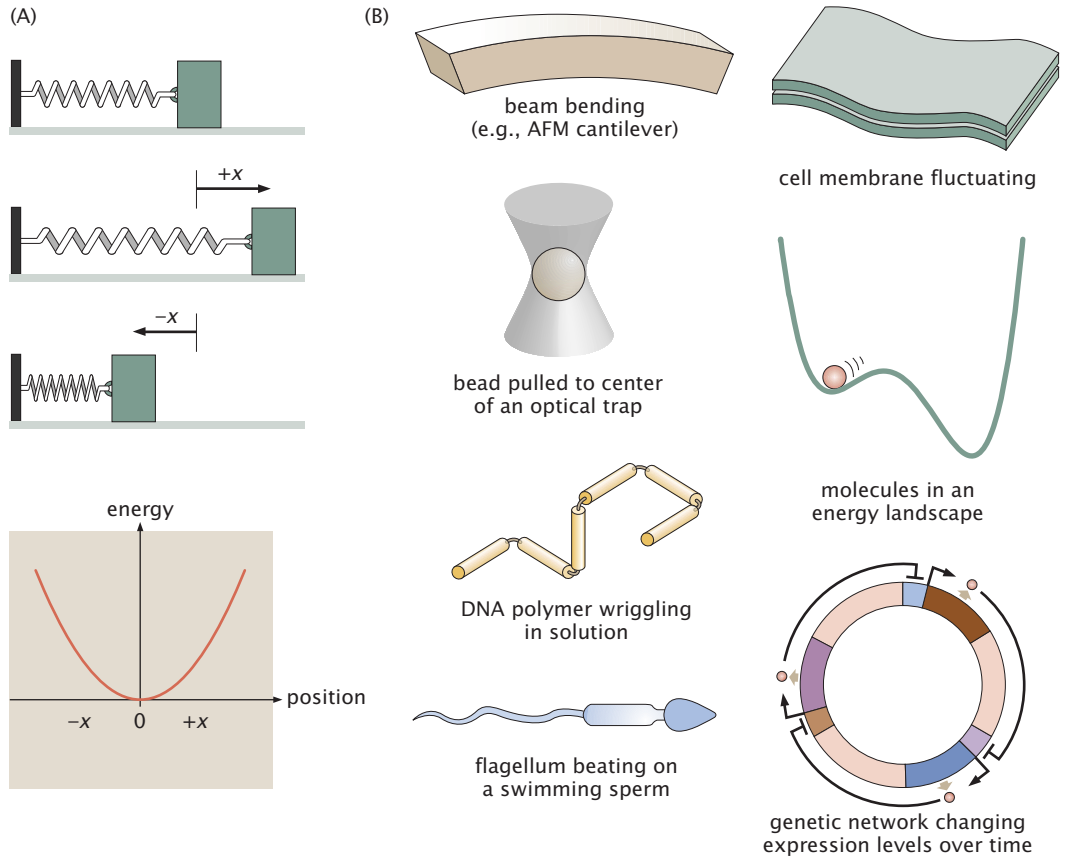


Figure 1.12: Gallery showing some of the different ways in which the idea of a spring is used in biological model building. (A) Classical physics representation of the harmonic oscillator. A simple spring is attached at one end to an immovable wall, and at the other end to a block-shaped mass that can slide on a frictionless table. The position of the mass changes as the spring extends and compresses. The graph at the bottom shows the energy of the spring as a function of the mass position. When the spring is relaxed at its resting length, the energy is at a minimum. (B) Multiple manifestations of the simple harmonic oscillator relevant to biological systems.

gene expression over time for certain interesting kinds of genetic networks can also be surprisingly well described using the mathematics of simple harmonic oscillators.

1.5.2 The Toolbox of Fundamental Physical Models

The overall strategy of this book will be to apply this kind of fundamental physical thinking to the widest possible range of biological problems. Through careful analysis of the quantitative assumptions that go into the analysis of biological data, it has become clear that a mere handful of fundamental physical models are often sufficient to provide a rigorous framework for interpretation of many kinds of quantitative biological data.

Beyond the harmonic oscillator introduced above, we believe that there are fewer than 10 other key physical concepts that require mastery. A scientist who understands and appreciates the applications of these concepts in biology is afforded a powerful framework for zeroing in on the most important, unexplained open questions. It is not our contention that this family of broad reaching physical models is sufficient to explain biological phenomena, far from it. Rather, we contend that any observations that can be completely quantitatively accounted for within the framework of one of these models require no further speculation of unidentified mechanism, while data that cannot be embraced within this physical framework cry out for more investigation and exploration.

Our choice of this set of concepts is purely utilitarian. We are not attempting to write a book surveying all of physics, or surveying all of

biology. Modern, cutting-edge research at the interdisciplinary interface between biology and physics demands that scientists trained in one discipline develop a working knowledge of the other. Students currently training in these fields have the opportunity to explore both in detail and to understand their complementarity and interconnectedness. In this book, we have attempted to address all three audiences: physicists curious about biology, biologists curious about physics, and broad-minded learners ready to delve into both. Our hope is that these ideas will provide a functional foundation of some of the key physical concepts most directly relevant to biological inquiry. At the same time, we hope to provide enough examples of specific biological phenomena to allow a physical scientist to identify further important unsolved problems in biology.

The bulk of this book comprises detailed dissections and case studies of these fundamental concepts and concludes with chapters showing applications of several of these concepts simultaneously. As noted above, many individual chapters each focus on a single class of physical model. The key broad physical foundations are as follows:

- simple harmonic oscillator (Chapter 5)
- ideal gas and ideal solution models (Chapter 6)
- two-level systems and the Ising model (Chapter 7)
- random walks, entropy, and macromolecular structure (Chapter 8)
- Poisson–Boltzmann model of charges in solution (Chapter 9)
- elastic theory of one-dimensional rods and two-dimensional sheets (Chapters 10 and 11)
- Newtonian fluid model and the Navier–Stokes equation (Chapter 12)
- diffusion and random walks (Chapter 13)
- rate equation models of chemical kinetics (Chapter 15)

1.5.3 The Unifying Ideas of Biology

One of the most pleasing and powerful aspects of science is its ability to link together seemingly unrelated phenomena. Few examples from the physical sciences are more potent at making this point than the surprising unification of the age-old phenomenon of magnetism known from lodestones and the early days of compasses and navigation to the diverse manifestations of electricity such as the lightning that attends summer thunderstorms to the many and varied appearances of light such as rainbows or the apparent bending of a spoon in a glass of water. All of these phenomena and many more were brought under the same roof in the form of the modern theory of electromagnetism.

The toolbox of fundamental physical ideas introduced above focused on the way in which certain key ideas from physics have served to buckle together a huge number of observations in a way that is not just descriptive, but, more importantly, is predictive. But such unifying concepts are certainly not unique to the physical sciences. Biology has far-reaching ideas of its own. Just as it is true that a mere handful of fundamental physical models are often sufficient to provide a rigorous framework for interpretation of many kinds of quantitative biological data, there is a similar handful of fundamental biological ideas that serve as an overarching foundation for all that we think about in biology. Indeed, many earlier authors

(see the references at the end of the chapter) have provided various interesting perspectives on what one might call the "great ideas of biology."

On the second page of the nearly 1400 pages of the *Feynman Lectures on Physics*, Feynman says "If, in some cataclysm, all of scientific knowledge were to be destroyed, and only one sentence passed on to the next generation of creatures, what statement would contain the most information in the fewest words?" Feynman chooses the idea that "all things are made of atoms". However, a survey of scientists would not be unanimous on this point, and some fraction of them would come down in favor of something along the lines of "all living organisms are related through descent from a common ancestor through the process of evolution by natural selection." Though perhaps it has become hackneyed through overuse, the words of Dobzhansky remain both timely and true: "Nothing in biology makes sense except in the light of evolution." Indeed, there is a strong argument that evolution is biology's greatest idea and will leave its indelible stamp on nearly everything we say in the chapters that follow.

While Darwin and Wallace labored to find an understanding of the process of evolution, Mendel was hard at work in establishing the early ideas on the underlying mechanisms of inheritance, resulting in discoveries that founded another one of biology's uniquely great ideas, namely, genetics. One of the most remarkable adventures of the human mind has been the quest to relate Mendel's abstract particles of inheritance to the molecules of the cell, with the successful insight being the realization that DNA is the carrier of genetic information. History is often characterized by various ages and it is no exaggeration to say that we are now living in the DNA Age.

Another of the key ideas that is now largely taken for granted, but is a unifying concept all the same and was hard won in the making, is the cell theory, the insight that all living organisms are made up of cells and that individual cells are themselves the quantum unit of living matter. The biochemical processes that take place within cells also reveal a remarkable unity, whether we think of the common metabolic pathways shared by so many organisms or the fact that the molecular machinery that implements our shared genetic heritage is based upon a few common macromolecular assemblies such as polymerases and the ribosome.

The key broad biological foundations that will serve as a point of departure for much of what follows in the remainder of the book are as follows:

- the theory of evolution (Chapters 3, 4, 18, 20, and 21, and many others less explicitly)
- genetics and the nature of inheritance (Chapters 2–5 and 21)
- the cell theory (Chapters 2, 3, 11, and 14)
- the unity of biochemistry (Chapters 2, 3, 5, 6, and 15)

One of the features we find especially intriguing about physical biology is that this perspective yields a different view of the proximity of different biological phenomena. As illustrated in Figure 1.13, the way we think about problems depends greatly upon both the kinds of questions we are asking and on how we go about answering them. The development of a frog and a fly both raise questions that a developmental biologist would recognize as related: what determines the timing and synchronization of the developmental process,

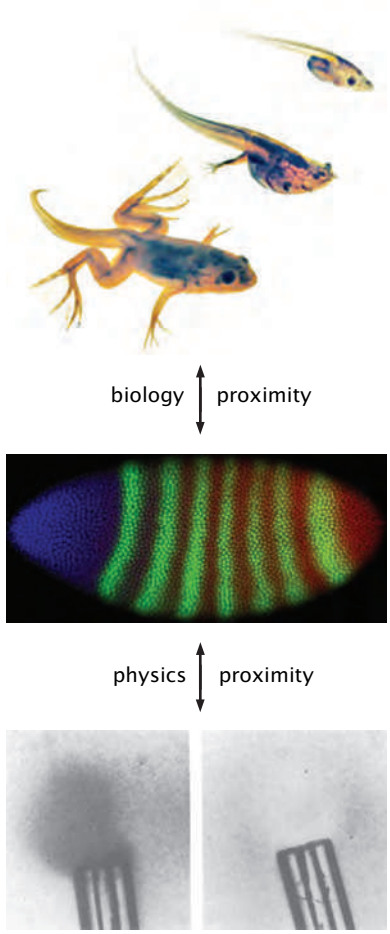


Figure 1.13: Proximity of ideas in cell biology and in physical biology. During the early development of the fruit fly *Drosophila* (middle), the eventual fate of the cells is determined by the patterned expression of a group of regulatory proteins (colored stripes). In a traditional organization of biological topics, the process of cell fate determination in *Drosophila* development might be discussed in the context of development of other animals such as the frog (top). However, the critical physical concepts that enable cell specification in the fly embryo involve processes such as diffusion and precise measurement of chemical concentrations processes that are also central to very different biological phenomena that have nothing to do with animal development. One of these phenomena is illustrated at the bottom, where on the left a group of bacteria is seen swarming toward the tip of a glass micropipette that is leaking a delicious amino acid (serine) into the medium. When the serine is switched for a toxin (phenol), as shown on the right, the bacteria quickly reverse course and swim away. In both the developing embryo and the swimming bacterium, cells measure the concentrations in their environment and make decisions as a result of similar physical measurements. (Top, courtesy of R. J. Denver, photo by David Bay; middle, adapted from E. Poustelnikova et al., *Bioinformatics* 20:2212, 2004; bottom, adapted from B. A. Rubik et al., *Proc. Natl Acad. Sci. USA* 75:2820, 1978.)

how is polarity established, etc.? On the other hand, pattern formation in the fly embryo and bacterial chemotaxis are conceptually linked when viewed as questions about the limits of chemical gradient detection. In the chapters that follow, we seek these kinds of unconventional statements of proximity with the ambition of using proximity in the physical biology sense to shed light on a host of different problems.

1.5.4 Mathematical Toolkit

One of the most surprising aspects of science is the way in which the same ideas show up again and again. For example, the random walk model does duty not only as a way to think about the diffusive motions of molecules in solution such as described in Figure 1.13 but also as a tool for describing the conformations of polymers. Thus far, we have argued for such unity and pervasiveness in the context of physics and biology. But as will be seen throughout the book, certain overarching mathematical ideas reveal themselves again and again as well. Indeed, this observation has been canonized in much the same way as Dhobzhansky's remark about evolution, in the title of another famous essay, this time by Eugene Wigner and entitled "The Unreasonable Effectiveness of Mathematics in the Natural Sciences."

One example of this unreasonable effectiveness has to do with our ability to assign numbers to our degree of belief when faced with

questions where there are many outcomes, such as whether a receptor will be bound by a ligand or not. Probability theory will serve as one of the mathematical cornerstones of what we have to say in coming chapters. In particular, we will repeatedly and often appeal to four key distributions: the binomial distribution, the Gaussian distribution, the Poisson distribution, and the exponential distribution, each of which will help us sharpen the way we think about many different biological questions.

However, we will need more than just the powerful calculus of uncertainty provided by the modern theory of probability and statistics. Ideas from the differential and integral calculus are also foundational in much of what we will say, especially through our use of differential equations. Often, we will want to know how the quantity of some particular species, such as the number of microtubules, varies in time. With our calculus toolkit, one of our favorite tools will concern the mathematics of superlatives, how do we find the largest or smallest values of some function or functional? These ideas will be explored in detail in Section 5.3.1 (p. 209).

Sometimes, the push for a strictly analytical treatment of the mathematics that arises in thinking about our models will be either too tedious or demanding. In such cases, we will turn to the amplification in personal mathematical power that is available to us all now as a result of computers. Indeed, the exploitation of such computation will serve as one of the centerpieces through the remainder of the book in the form of "Computational Explorations" in which we use Matlab to numerically explore our models.

1.5.5 The Role of Estimates

In this era of high-speed computation, the temptation to confront new problems in their full mathematical complexity is seductive. Nevertheless, the core philosophy of the present book advocates a spirit that is embodied in a story, perhaps apocryphal, of the demise of Archimedes. Plutarch tells us that Archimedes was deeply engaged in a calculation, which he was performing with a stick in the sand, at the time when the Roman army invaded Syracuse. When confronted by a Roman soldier and asked to identify himself, Archimedes told the soldier to wait until he was done with his calculation. The soldier replied by impaling Archimedes, bringing the life of one of the great scientists of history to an early and unjust end. Our reason for recounting this story is as a reminder of the mathematical character of many of the models and estimates that we will report on in this book (but hopefully with less dire personal consequences). In particular, we like to imagine that a model is of the proper level of sophistication to help build intuition and insight if, indeed, its consequences can be queried by calculations carried out with a stick in the sand, without recourse to sophisticated computers or detailed references.

Throughout the book, we will repeatedly make use of estimates to develop a feeling for the numbers associated with biological structures and biological processes. This philosophy follows directly from the biological tradition articulated by the pioneering geneticist Barbara McClintock, who emphasized the necessity of developing a "feeling for the organism," in her case the crop plant maize. We believe that a quantitative intuition is a critical component of having a feeling for the organism. Indeed, we think of model building as proceeding through a series of increasingly sophisticated steps starting with

simple estimates to develop a feel for the relevant magnitudes, followed by the development of toy models that can (usually) be solved analytically, followed in turn by models with an even higher degree of realism that might require numerical resolution.

The estimates that we advocate are all meant to be sufficiently simple that they can be done on the back of an envelope or written with a stick in the sand on a desert island. In keeping with this purpose, we will rarely use more than one significant digit in our estimates. One useful rule of thumb that we will adopt is the geometric mean rule (see *Guesstimation* by Weinstein and Adam for a more complete description). The idea is that when making an estimate where we are uncertain of the magnitudes, begin by attempting to guess a lower and upper bound for the magnitude of interest. With these two guesses in hand, multiply them together and take the square root of the result as your best guess for that poorly known quantity. This strategy is illustrated in Figure 1.14. What this scheme really amounts to is averaging the exponents rather than the quantities themselves.

Naturally, this introduces somewhat low precision, for example, an estimate may yield an answer of 100 molecules in a system when the precisely measured value is actually 42, but has the benefit of greater simplicity and clarity. In general we will consider an estimate to be successful if the result is within an order of magnitude of the measured value. Indeed, as shown in Figure 1.15, we will often resort to simple arithmetic rules in which we are intentionally numerically imprecise in the name of rapidly getting to an order-of-magnitude result. In cases where our estimates differ by more than an order of magnitude, it will be clear that we have made an incorrect assumption and we will use this as an opportunity to learn something more about the problem of interest.

The precept that animates much of the book is that we are closest to understanding a phenomenon, not when we unleash the full weight of a powerful numerical simulation of the problem at hand, but rather when we are engaged in developing a feel for the scaling of the quantities of interest and their associated numerical values.

In many cases, one of the most powerful tools that can be unleashed in order to understand the relevant numerical magnitudes in a given problem is known as *dimensional analysis*. Indeed, in many examples throughout the book, whether when talking about fluid flow using the Navier–Stokes equations or the nature of the genetic switch, we will write our equations in dimensionless form. One of the reasons that these dimensionless forms are so powerful is that they permit us to write the equations describing the phenomenon of interest in the "natural variables" of the problem. For example, when thinking about a circuit with a resistor and a capacitor, the natural time unit is given by $\tau = RC$, the product of the resistance and the capacitance. In thinking about the establishment of morphogen gradients during anterior–posterior patterning in flies, the fundamental length scale is given by $\lambda = \sqrt{D\tau}$, where D is the diffusion constant of the morphogen and τ is its degradation time. Further, when we are trying to estimate some quantity of interest, a powerful trick is to simply use the relevant constants describing that problem to construct a quantity with the dimensions of the quantity we are trying to estimate.

In the morphogen gradient example, if we assume that the length scale λ over which the pattern changes significantly is determined by the diffusion constant of the morphogen and its mean lifetime, then the dimensions of these two quantities determine the formula for the length scale λ up to a numerical constant. To be specific, the diffusion

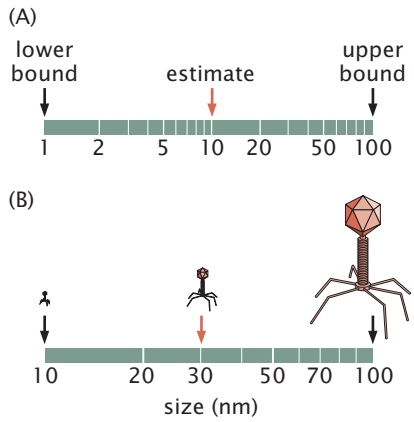


Figure 1.14: Geometric mean rule for estimates. (A) For cases in which we don't know the relevant magnitude, a powerful scheme is to guess an upper and lower bound and to take their geometric mean $\sqrt{x_{\text{lower}} x_{\text{upper}}}$ as our best guess. (B) Application of the geometric mean rule to guess the size of bacteriophages that are larger than 10 nm and smaller than 100 nm. The geometric mean guess is for a typical size of roughly 30 nm.

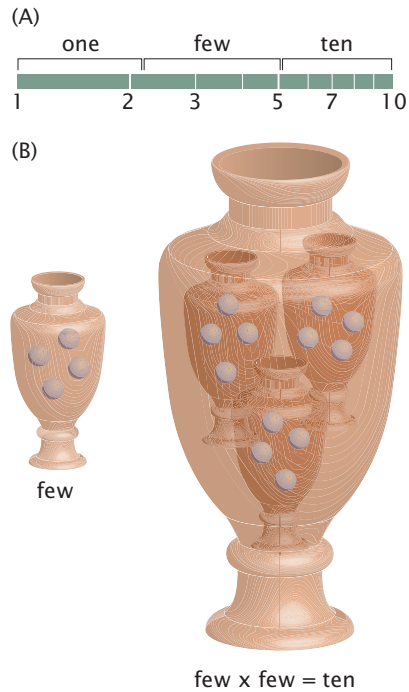


Figure 1.15: Few times one rule for multiplication. (A) Scheme for assigning numbers either to "one," "few," or "ten (concept from "Street Fighting Mathematics" by Sanjoy Mahajan)." (B) Demonstration that few $\times$ few $\approx$ 10.

constant has dimensions of Length^2/Time, while the mean lifetime has dimensions of Time. The only way these two quantities can be combined into one that has dimensions of Length is by multiplying them and then taking a square root, leading to the formula for λ given above. The particular power of dimensional analysis comes from comparing the estimate it yields with the measured value of the relevant quantity. In the morphogen example, a large discrepancy of the measured λ and the one estimated from the formula obtained by dimensional analysis would indicate that our determination of the morphogen diffusion constant or its lifetime is wrong. If more careful measurements of these two parameters still lead to a discrepancy, then we are led to an even more exciting conclusion, namely, that there is something fundamentally missing in our understanding of how the morphogen gradient is established, something that would introduce yet another parameter that determines λ.

Taylor and the Cover of Life *Magazine* One of the greatest stories of the power and subtlety of estimation and, particularly, of dimensional analysis is associated with G. I. Taylor. During the early days of the atomic bomb, the energetic yield of the first explosion was classified as a military secret. This did not keep the people who make such decisions from releasing several time lapse photos of the explosion that showed the radius of the fireball as a function of time. On this meager information, shown on the cover of *Life* magazine, Taylor was able to use the principles of fluid mechanics to *estimate* the yield of the explosion, an estimate that was exceedingly close to the secret value of the actual yield.

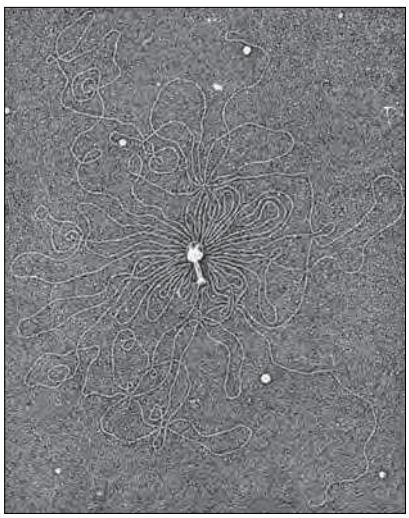


Figure 1.16: Electron microscopy image of a bacteriophage genome that has escaped its capsid. Simple arguments from polymer physics can be used to estimate the genomic size of the DNA by examining the physical size of the randomly spread DNA. We will perform these kinds of calculations in Chapter 8. (Adapted from G. Stent, Molecular Biology of Bacterial Viruses. W. H. Freeman, 1963.)

Interestingly, this famous episode from the cultural history of the physical sciences could have happened just as well in biology. Pictures like that shown in Figure 1.16 were the thrilling outcomes of the early days of the use of electron microscopy in biology. Here too, by observing the physical region occupied by the DNA molecule from the ruptured bacterial virus (bacteriophage), it is possible to make a highly simplified estimate of the *genomic* size (that is, the length in base pairs) of the bacteriophage's DNA.

Why do numbers matter? Why is it useful to know the number of copies of a given molecule that is controlling the expression of a gene of interest within an *E. coli* cell? What merit is there in estimating the rate of actin polymerization at the leading edge of a motile cell? What is the value in knowing the time it takes a given protein to diffuse a millimeter? First, simple estimates serve as a reality check to help us see whether or not our impressions of how a system works are reasonable at all. Second, a sense of the numbers can tell us what kinds of physical constraints are in play. For example, by knowing the size of a bacterium and its swimming speed, we can construct a characteristic number (the Reynolds number) that tells us about the importance of acceleration in the dynamical problem of swimming. Similarly, a knowledge of the number of the specific DNA-binding proteins known as transcription factors that control the expression of a gene of interest can tell us about the importance of fluctuations in the gene expression process.

As will be seen in Chapters 3 and 19, biological oscillators are one of the centerpieces of the regulatory architecture of almost all living organisms. Oscillatory behavior is seen in examples ranging from the clocks that control the cell cycle in developing embryos to the circadian rhythms that make jet lag the unwanted but constant companion

of the weary traveler. As we explore mathematical models of regulatory biology, we will see that minimal oscillators can be constructed from genetic networks involving just two proteins, namely, an activator and a repressor. But what we will also see in contemplating these important networks is that for many of the possible parameters we might choose for the protein production and degradation rates, they don't yield oscillatory behavior at all, but rather stable and boring solutions in which the protein concentrations relax after some transient period into stationary values. Hence, even a clear picture of the wiring of some network cannot tell us what that network does in the normal life of a cell and calls for a quantitative examination of how that network behaves as a function of its governing parameters.

1.5.6 On Being Wrong

When performing simple quantitative estimates, or when attempting to write simple mathematical models to describe some aspect of a biological process, we are easily tempted to worry about whether our estimates and models are "right" or "wrong." There are at least four different ways in which a model might justly be called "wrong," but we believe that committing each of these kinds of wrongness can be a useful and worthwhile experience.

One category of wrong models might fail because they use an inappropriate type of fundamental physical model to abstract an aspect of a particular biological problem. This kind of wrongness might be illustrated by the question: how many hours can fit in the back of a pickup truck? The investigator will quickly learn that the predictions of such a model will simply have no relevance to the actual experiment or biological phenomenon being considered. It is inevitable that any person learning to conceptualize the kinds of complicated issues we explore in biology will make some of these kinds of mistakes at first. With time and experience, these errors will become less frequent.

A second category of wrong models might fail because they do not include enough detail about the system. For example, when we try to predict how fast *E. coli* bacterial cells can swim through water, we might be able to write an equation that correctly predicts the order of magnitude of the average speed of the cells, but fails to provide any basis for predicting how the speed might fluctuate over time. This kind of error, we believe, is not so much wrongness as incompleteness. As we have argued, it is actually extremely important that physical models be incomplete in order to be simple enough to help in building intuition. A model that is nearly as complex as the system itself may be able to make very accurate predictions, but will not be useful for promoting understanding.

A third category of wrong models includes those where inappropriate assumptions are made about the relative importance of different physical considerations. Although these models may be mathematically correct, they will fail to make accurate predictions for the behavior of cells in the real world. For example, if the swimming *E. coli* cell stops rotating its flagellum, then its future trajectory will be governed by a combination of inertia, viscous drag, and Brownian motion. A modeler whose primary experience lies in hydrodynamics for large objects such as submarines or fish may be inclined to ignore the contribution of Brownian motion, but for the bacterium this is the primary determinant of its behavior. Familiarity with the biological system and the opportunity to directly compare model predictions

with experimental data for iteratively refined generations of models will usually allow this problem to be overcome.

The fourth category of wrong models is by far our favorite. These are the wrong models that drive breakthrough experiments. When a person who is familiar with a particular biological system writes down a quantitative model for that system incorporating the correct basic physical models and uses the correct order-of-magnitude estimates for relevant parameters nevertheless finds that it makes dramatically incorrect predictions, this is almost always a golden opportunity to learn something fundamentally new about the system. Science advances when people notice that something does not quite make sense, that something is puzzling, that something seems to be missing. It is our hope that readers of this book will develop sufficient confidence in their ability to create appropriate models that when they find a model that seems to be wrong in this way they will seize the opportunity to make a new scientific discovery.

1.5.7 Rules of Thumb: Biology by the Numbers

A mastery of any field requires knowledge of a few key facts. For the physicist who uses quantum mechanics, this means knowing the ground state energy of the hydrogen atom, the charge on an electron, Planck's constant, and several other key physical parameters. In physical biology, there are similarly a small set of numbers that are worth carrying around as the basis of simple estimates of the kind one might carry out with a stick in the sand. In that spirit, the key numbers given in Table 1.1 are those that we find useful when thinking about physical biology of the cell. In the chapters that follow, these key numbers will permit us to form the habit of making quantitative estimates as a preliminary step when confronting new problems.

Throughout the remainder of the book, we will repeatedly make reference to certain numerical rules of thumb such as that the volume of an *E. coli* cell is roughly $1\,\mu\text{m}^3$ or that the natural energy scale of physical biology is $k_\text{B}T$ (that is, Boltzmann's constant multiplied by room temperature in kelvins ≈ 4 pN nm (piconewton nanometers)). It is convenient to express this natural energy scale in terms of piconewtons and nanometers when we are considering the forces generated by protein molecular motors moving along DNA, for example, but this natural energy scale can also be expressed in different units depending on the type of calculation we are doing. When considering processes involving charge and electricity, it is convenient to recall that $k_\text{B}T$ is approximately equal to 25 meV (millielectron volts). When considering biochemical reactions, it is more useful to express $k_\text{B}T$ in biochemical units, as 0.6 kcal/mol (kilocalories per mole) or 2.5 kJ/mol (kilojoules per mole).

The quantitative rules of thumb in Table 1.1 will serve as the basis of our rough numerical estimates that could be carried out using a stick in the sand without reference to books, papers, or tables of data. Where do these numbers come from? Each comes from the results of a long series of experimental measurements of many different kinds. Right now we will simply assert their values, but later in the book we will provide numerous examples showing how these kinds of measurements have been made and also noting that in many cases, there remains uncertainty about the precise values of the parameters that emerge from such measurements. How precisely must we know these numbers? As stated earlier, our approach is to be generally satisfied

Table 1.1: Rules of thumb for biological estimates.

	Quantity of interest	Symbol	Rule of thumb
E. coli			
	Cell volume	$V_{E.\,coli}$	$\approx 1\,\mu m^3$
	Cell mass	$m_{E.\,coli}$	≈ 1 pg
	Cell cycle time	$t_{E.\,coli}$	≈ 3000 s
	Cell surface area	$A_{E.\,coli}$	$\approx 6\,\mu m^2$
	Macromolecule concentration in cytoplasm	$c_{E.\,coli}^{\text{macromol}}$	≈ 300 mg/mL
	Genome length	$N_{\text{bp}}^{E.\,coli}$	$\approx 5 \times 10^6$ bp
	Swimming speed	$v_{E.\,coli}$	$\approx 20\,\mu$m/s
Yeast			
	Volume of cell	V_{yeast}	$\approx 60\,\mu m^3$
	Mass of cell	m_{yeast}	≈ 60 pg
	Diameter of cell	d_{yeast}	$\approx 5\,\mu$m
	Cell cycle time	t_{yeast}	≈ 200 min
	Genome length	$N_{\text{bp}}^{\text{yeast}}$	$\approx 10^7$ bp
Organelles			
	Diameter of nucleus	d_{nucleus}	$\approx 5\,\mu$m
	Length of mitochondrion	l_{mito}	$\approx 2\,\mu$m
	Diameter of transport vesicles	d_{vesicle}	≈ 50 nm
Water			
	Volume of molecule	V_{H_2O}	$\approx 10^{-2}$ nm^3
	Density of water	ρ	1 g/cm^3
	Viscosity of water	η	≈ 1 centipoise (10^{-2} g/(cm s))
	Hydrophobic embedding energy	$\approx E_{\text{hydr}}$	2500 cal/(mol nm^2)
DNA			
	Length per base pair	l_{bp}	$\approx 1/3$ nm
	Volume per base pair	V_{bp}	≈ 1 nm^3
	Charge density	λ_{DNA}	2 e/0.34 nm
	Persistence length	ξ_{p}	50 nm
Amino acids and proteins			
	Radius of "average" protein	r_{protein}	≈ 2 nm
	Volume of "average" protein	V_{protein}	≈ 25 nm^3
	Mass of "average" amino acid	M_{aa}	≈ 100 Da
	Mass of "average" protein	M_{protein}	$\approx 30{,}000$ Da
	Protein concentration in cytoplasm	c_{protein}	≈ 150 mg/mL
	Characteristic force of protein motor	F_{motor}	≈ 5 pN
	Characteristic speed of protein motor	v_{motor}	≈ 200 nm/s
	Diffusion constant of "average" protein in cytoplasm	D_{protein}	$\approx 10\,\mu m^2/s$
Lipid bilayers			
	Thickness of lipid bilayer	d	≈ 5 nm
	Area per molecule	A_{lipid}	$\approx \frac{1}{2}$ nm^2
	Mass of lipid molecule	m_{lipid}	≈ 800 Da

in those cases where our numerical estimates are of the right order of magnitude. To that end, we make certain gross simplifications. For example, when thinking about proteins and the amino acids that make them up, we will often adopt a minimalistic perspective in which the mass of amino acids is *approximated* to be 100 Da (daltons, or grams per mole). In actuality, real amino acids have masses that range over a wider set of values from 75 to 204 Da. Likewise, the rest of the numbers in Table 1.1 represent only rough average values to use as a starting point for quantitative estimates. As we progress through the examples in the rest of the book, these values will become increasingly familiar. However, our book does not make a systematic attempt at the important task of collecting values for a wide range of biologically important numbers. For this, the interested reader is encouraged to visit the "BioNumbers" website to search out what is known about the quantitative status of a given biological parameter and to be on the lookout for those cases in which we cite such BioNumbers by their BioNumber ID (BNID).

1.6 Summary and Conclusions

When applying physical and quantitative thinking to biological problems, two different skills are critically important. First, the biological system must be described in terms of fundamental physical models that can help to shed light on the behavior of the system. Fewer than 10 truly basic physical models are able to provide broad explanatory frameworks for many problems of interest in biology. Second, appropriate quantitative estimates to give order-of-magnitude predictions of sizes, numbers, time scales, and energies involved in a biological process can serve as a reality check that our conception of the system is appropriate and can help predict what kinds of experimental measurements are likely to provide useful information. Armed with these skills, an investigator can begin to systematize the large amounts of detailed quantitative data generated by biological experimentation and determine where new scientific breakthroughs might be sought.

1.7 Further Reading

Alberts, B, Johnson, A, Lewis, J, et al. (2008) Molecular Biology of the Cell, 5th ed., Garland Science. This book provides a discussion of much of the biological backdrop for our book.

Schaechter, M, Ingraham, JL, & Neidhardt, FC (2006) Microbe, ASM Press. This book is full of insights into the workings of microscopic organisms.

Stryer, L (1995) Biochemistry, 5th ed., W. H. Freeman. We like to imagine our readers with this book at their side. Stryer is full of interesting insights presented in a clear fashion.

Schleif, R (1993) Genetics and Molecular Biology, 2nd ed., The Johns Hopkins University Press. Schleif's book discusses many of the same topics found throughout our book from a deeper biological perspective, but with a useful and interesting quantitative spin.

Vogel, S (2003) Comparative Biomechanics—Life's Physical World, Princeton University Press. Though largely aimed at describing problems at a different scale than those described in our book, Vogel's book (as well as his many others) presents an amazing merger of biology and physical reasoning.

Nurse, P (2003) The Great Ideas of Biology, *Clinical Medicine* **3**, 560. This article describes Nurse's views on some of the greatest ideas in biology.

Bonner, JT (2002) The Ideas of Biology, Dover Publications. This book espouses Bonner's view on the overarching themes of modern biology.

The spirit of physical biology we are interested in developing here has already been described in a number of other books. The following are some of our favorites:

Dill, K, & Bromberg, S (2011) Molecular Driving Forces, 2nd ed., Garland Science. Despite a deceptive absence of difficult mathematics, their book is full of insights, subtlety, and tasteful modeling.

Boal, D (2012) Mechanics of the Cell, Cambridge University Press. Boal's book illustrates the way in which physical

reasoning can be brought to bear on problems of biological significance. We have found that many of the exercises at the back of each chapter are the jumping off point for the analysis of fascinating biological problems.

Howard, J (2001) Mechanics of Motor Proteins and the Cytoskeleton, Sinauer Associates. Howard's book is a treasure trove of information and concepts.

Nelson, P (2004) Biological Physics: Energy, Information, Life, W. H. Freeman. This book discusses a wide range of topics at the interface between physics and biology.

Benedek, GB, & Villars, FMH (2000) Physics with Illustrative Examples from Medicine and Biology, 2nd ed., Springer-Verlag. This three-volume series should be more widely known. It is full of useful insights into everything ranging from the analysis of the Luria–Delbrück experiment to the action potential.

Sneppen, K, & Zocchi, G (2005) Physics in Molecular Biology, Cambridge University Press. This book presents a series of vignettes that illustrate how physical and mathematical reasoning can provide insights into biological problems.

Jackson, MB (2006) Molecular and Cellular Biophysics, Cambridge University Press. Jackson's book covers many themes in common with our own book and is especially strong on ion channels and electrical properties of cells.

Waigh, TA (2007) Applied Biophysics, Wiley. This interesting book covers many of the same topics covered here.

Harte, J (1988) Consider a Spherical Cow, University Science Books. This book does a beautiful job of showing how order of magnitude estimates can be used to examine complex problems. Those that are intrigued by this book might consider dipping into his Consider a Cylindrical Cow (University Science Books, 2001) as well.

Weinstein, L, & Adam, JA (2008) Guesstimation: Solving the World's Problems on the Back of a Cocktail Napkin, Princeton University Press. This wonderful little book shows by example how to do the kind of estimates advocated here and also is the basis for the rule captured in Figure 1.14.

Goldreich, P, Mahajan, S, & Phinney, S, Order-of-Magnitude Physics: Understanding the World with Dimensional Analysis, Educated Guesswork, and White Lies, online book, available at www.inference.phy.cam.ac.uk/sanjoy/oom/. This excellent work embodies the science of estimation so helpful to biology as well as physics.

Cohen, J (2004) Mathematics is biology's next microscope, only better; Biology is Mathematics' next physics, only better, *PLoS Biology* **2**, e439. This article argues for the way that quantitative reasoning will enrich biology and for the ways that biology will enrich mathematics.

Mahajan, S (2010) Street-Fighting Mathematics, The MIT Press. Much of our mathematics will be of the street-fighting kind and there is nowhere better to get a feeling for this than from Mahajan's book.

1.8 References

Bray, D (2001) Reasoning for results, *Nature* **412**, 863. Source of Darwin's quote at the beginning of the chapter.

Darwin, F (ed.) (1903) More Letters of Charles Darwin: A Record of His Work in a Series of Hitherto Unpublished Letters, Volume I, John Murray.

DiMauro, S, & Schon, EA (2003) Mitochondrial respiratory-chain diseases, *N. Engl. J. Med.* **348**, 2656.

Fawcett, DW (1996) The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure, W. B. Saunders. Fawcett's atlas is full of wonderful electron microscopy images of cells.

Milo, R, Jorgensen, P, Moran, U, et al. (2010) BioNumbers—the database of key numbers in molecular and cell biology, *Nucl. Acids Res*. **38**, D750.

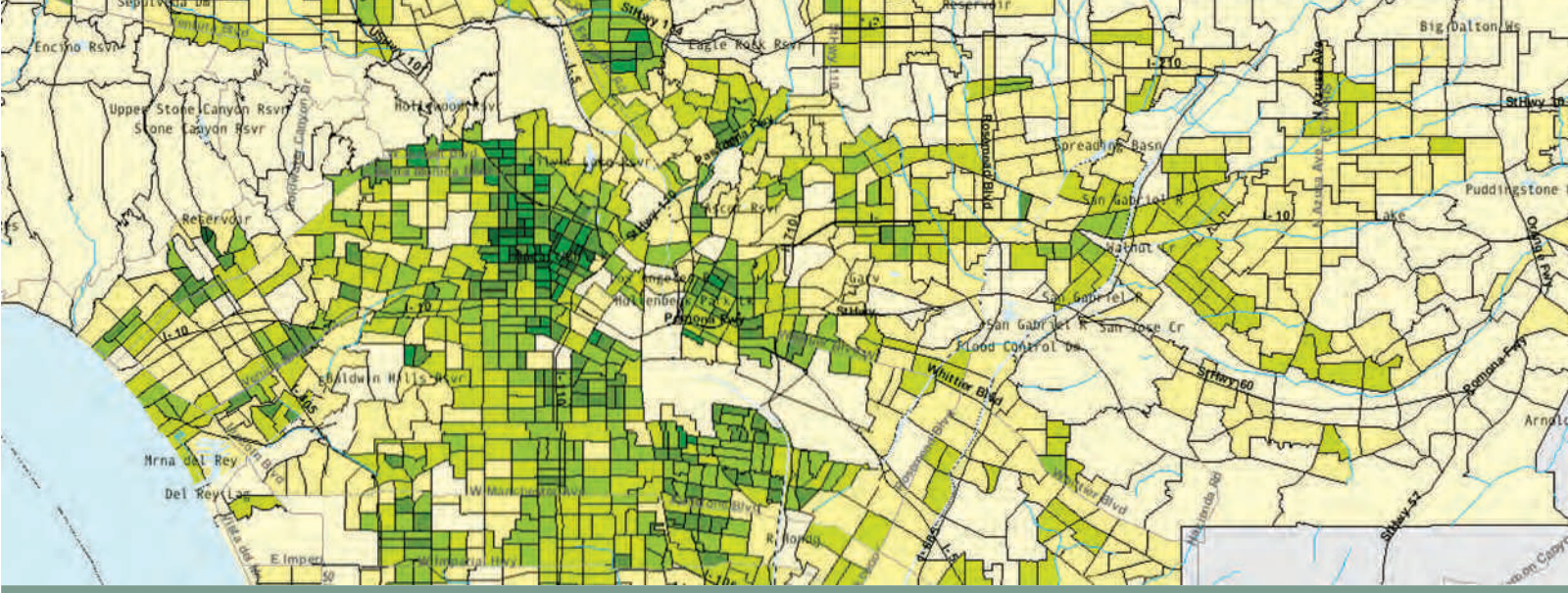
Müller, J, Oehler, S, & Müller-Hill, B (1996) Repression of *lac* promoter as a function of distance, phase and quality of an auxiliary *lac* operator, *J. Mol. Biol.* **257**, 21.

Poustelnikova, E, Pisarev, A, Blagov, M, et al. (2004) A database for management of gene expression data *in situ, Bioinformatics* **20**, 2212.

Rubik, BA, & Koshland, DE, Jr. (1978) Potentiation, desensitization, and inversion of response in bacterial sensing of chemical stimuli, *Proc. Natl. Acad. Sci*., **75**, 2820.

Stent, G (1963) Molecular Biology of Bacterial Viruses, W. H. Freeman.

This page intentionally left blank
to match pagination of print book



What and Where: Construction Plans for Cells and Organisms

2

Overview: In which we consider the size of cells and the nature of their contents

Cells come in a dazzling variety of shapes and sizes. Even so, their molecular inventories share many common features, reflecting the underlying biochemical unity of life. In this chapter, we introduce the bacterium *Escherichia coli* (commonly abbreviated as *E. coli*) as our biological standard ruler. This cell serves as the basis for a first examination of the inventory of cells and will permit us to get a sense of the size of cells and the nature of their contents. Using simple estimates, we will take stock of the genome size, numbers of lipids and proteins, and the ribosome content of bacteria. With the understanding revealed by *E. coli* in hand, we then take a powers-of-10 journey down and up from the scale of individual cells. Our downward journey will examine organelles within cells, macromolecular assemblies ranging from ribosomes to viruses, and then the macromolecules that are the engines of cellular life. Our upward journey from the scale of individual cells will examine the biological structures resulting from multicellularity, this time with an emphasis on how cells act together in contexts ranging from bacterial biofilms to the networks of neurons in the brain.

> “By the help of Microscopes there is nothing so small as to escape our inquiry; hence there is a new visible World discovered.”
>
> **Robert Hooke**

2.1 An Ode to *E. coli*

Scientific observers of the natural world have been intrigued by the processes of life for many thousands of years, as evidenced by early written records from Aristotle, for example. Early thinkers wondered about the nature of life and its “indivisible” units in much the same way that they mused about the fundamental units of matter. Just as physical scientists arrived at a consensus that the fundamental unit of matter is the atom (at least for chemical transactions), likewise, observers of living organisms have agreed that the indivisible unit of life is the cell. Nothing smaller than a cell can be shown to

be alive in a sense that is generally agreed upon. At the same time, there are no currently known reasons to attribute some higher "living" status to multicellular organisms, in contrast to single-celled organisms.

Cells are able to consume energy from their environments and use that energy to create ordered structures. They can also harness energy and materials from the environment to create new cells. A standard definition of life merges the features of metabolism (that is, consumption and use of energy from the environment) and replication (generating offspring that resemble the original organism). Stated simply, the cell is the smallest unit of replication (though viruses are also replicative units, but depend upon their infected host cell to provide much of the machinery making this replication possible).

The recognition that the cell is the fundamental unit of biological organization originated in the seventeenth century with the microscopic observations of Hooke and van Leeuwenhoek. This idea was put forth as the modern cell theory by Schwann, Schleiden, and Virchow in the mid-nineteenth century and was confirmed unequivocally by Pasteur shortly thereafter and repeatedly in the time since. Biologists agree that all forms of life share cells as the basis of their organization. It is also generally agreed that all living organisms on Earth shared a common ancestor several billion years ago (or, more accurately, a community of common ancestors) that would be recognized as cellular by any modern biologist.

In terms of understanding the basic rules governing metabolism, replication, and life more generally, one cell type should be as good as any other as the basis of experimental investigations of these mechanisms. For practical reasons, however, biologists have focused on a few particular types of cell to try to illuminate these general issues. Among these, the human intestinal inhabitant *E. coli* stands unchallenged as the most useful and important representative of the living world in the biologist's laboratory.

Several properties of *E. coli* have contributed to its great utility and have made it a source of repeated discoveries. First, it is easy to isolate because it is present in great abundance in human fecal matter. Unlike most other bacteria that populate the human colon, *E. coli* is able to grow well in the presence of oxygen. In the laboratory, it replicates rapidly and can easily adjust to changes in its environment, including changes in nutrients. Second, it is nearly routine to deliberately manipulate these cells (using insults such as radiation, chemicals, or genetic tricks, for example) to produce mutants. Mutant organisms are those that differ from their parents and from other members of their species found in the wild because of specific changes in DNA sequence that give rise to biologically significant differences. For example, *E. coli* is normally able to synthesize purines to make DNA and RNA from sugar as a nutrient source. However, particular mutants of *E. coli* with enzymatic deficiencies in these pathways have lost the ability to make their own purines and become reliant on being fed precursors for these molecules. A more familiar and frightening example of the consequence of mutations is the way in which bacteria acquire antibiotic resistance. Throughout the book, we will be using specific examples of biological phenomena to illustrate general physical principles that are relevant to life. Often, we will have recourse to *E. coli* because of particular experiments that have been performed on this organism. Further, even when we speak of experiments on other cells or organisms, often *E. coli* will be behind the scenes coloring our thinking.

2.1.1 The Bacterial Standard Ruler

The Bacterium *E. coli* Will Serve as Our Standard Ruler

Throughout this book, we will discuss many different cells, which all share with *E. coli* the fundamental biological directive to convert energy from the environment into structural order and to perpetuate their species. On Earth, it is observed that there are certain minimal requirements for the perpetuation of cellular life. These are not necessarily absolute physical requirements, but in the competitive environment of our planet, all surviving cells share these features in common. These include a DNA-based genome, mechanisms to transcribe DNA into RNA, and, subsequently, translation mechanisms using ribosomes to convert information in RNA sequences into protein sequence and protein structure. Within those individual cells, there are many substructures with interesting functions. Larger than the cell, there are also structures of biological interest that arise because of cooperative interactions among many cells or even among different species. In this chapter, we will begin with the cell as the fundamental unit of biological organization, using *E. coli* as the standard reference and standard ruler. We will then look at smaller structures within cells and, finally, larger multicellular structures, zooming in and out from our fundamental cell reference frame.

Figure 2.1 shows several experimental pictures of an *E. coli* cell and its schematization into our standard ruler. The AFM image and the electron micrograph in Figures 2.1(A) and (B) show that these bacteria have a rod-like morphology with a typical length between 1 and 2 μm and a diameter between 0.5 and 1 μm. The reader is invited to use light microscopy images of these bacteria and to determine the size of an *E. coli* cell for him or herself in the "Computational Exploration" at the end of this section.

The length unit of 1 μm, or micrometer, is so useful for the discussion of cell biology that it has a nickname, the "micron." To put this standard ruler in perspective, we note that, with its characteristic length scale of 1 μm, it would take roughly 50 such cells lined up end to end in order to measure out the width of a human hair. On the other hand, we would need to divide the cell into roughly 500 slices of equal width in order to measure out the diameter of a DNA molecule. Using these insights, the question of how many bacteria can dance on the head of a pin is answered unequivocally in Figure 2.1(D).

Note that the average size of these cells depends on the nutrients with which they are provided, with those growing in richer media having a larger mass. An extremely elegant experiment that explores the connection between growth rate and mass is shown in Figure 2.2. Our reference growth condition throughout the book will be a chemically defined solution referred to by microbiologists as "minimal medium" that is a mixture of salts along with glucose as the sole carbon source. In the laboratory, bacteria are often grown in "rich media," which are poorly defined but nutrient-rich mixtures of extracts from organic materials such as yeast cultures or cow brains. Although microorganisms can grow very rapidly in rich media, they are rarely used for biochemical studies because their exact contents are not known. For consistency, we will therefore refer primarily to experimental results for bacterial growth in minimal media.

Because of its central role as the quantitative standard in the remainder of the book, it is useful to further characterize the geometry of *E. coli*. One example in which we will need a better sense of the

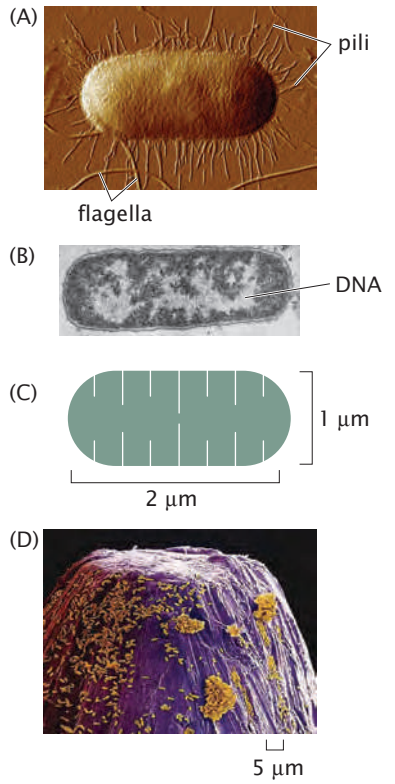


Figure 2.1: *E. coli* as a standard ruler for characterizing spatial scales. (A) Atomic-force microscopy (AFM) image of an *E. coli* cell, (B) electron micrograph of a sectioned *E. coli* bacterium, and (C) the *E. coli* ruler. (D) Bacteria on the head of a pin giving an impression of how our standard ruler compares with the dimensions of everyday objects. (A, courtesy of Ang Li; D, courtesy of Tony Brain.)

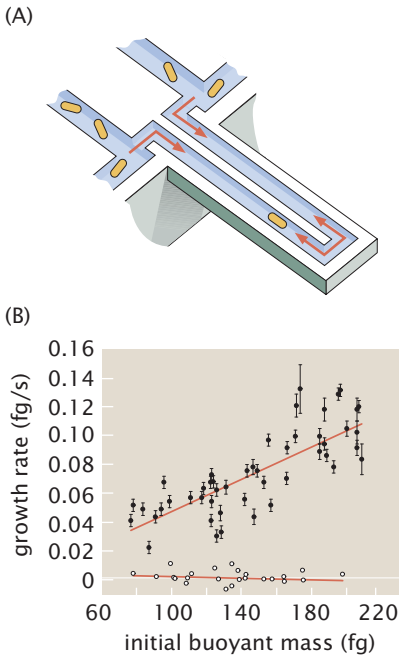


Figure 2.2: *E. coli* mass and growth rate as measured using a cantilever assay. (A) Schematic of the geometry of the device used to characterize the buoyant mass of bacterial cells by dynamically trapping them within the hollow cantilever. The hollowed out cantilever oscillates with a slightly different frequency when there is a cell present in its interior. (B) Relation between the buoyant mass at the start of the experiment and the growth rate. *E. coli* K12 cells grown at 37 °C. Filled circles correspond to cells that are growing normally and open circles correspond to fixed cells. (Figures adapted from M. Godin et al., *Nat. Meth.* 7: 387, 2010.)

geometry of cells and their internal compartments is in the context of reconciling experiments performed *in vitro* (that is, in test tubes) and *in vivo* (that is, in living cells). Results from experiments done *in vitro* are based upon the free concentrations of different molecular species. On the other hand, in *in vivo* situations, we might know the number of copies of a given molecule such as a transcription factor inside the cell (transcription factors are proteins that regulate expression of genes by binding to DNA). To reconcile these two pictures, we will need the cellular volume to make the translation between molecular counts and concentrations. Similarly, when examining the distribution of membrane proteins on the cell surface, we will need a sense of the cell surface area to estimate the mean spacing between these proteins, which will tell us about the extent of interaction between them. For most cases of interest in this book, it suffices to attribute a volume $V_{E.\ coli} \approx 1\,\mu\text{m}^3 = 1\,\text{fL}$ to *E. coli* and an area of roughly $A_{E.\ coli} \approx 6\,\mu\text{m}^2$ (see the problems for examples of how to work out these numbers from known cellular dimensions).

Computational Exploration: Sizing Up *E. coli* As already revealed in Figure 2.1, the simplest way for us to figure out how big bacteria are is to stick them under a microscope and to look at them. Of course, to do this in reality requires some way to relate the cells we see magnified in our microscope and their actual size. In modern terms, what this really means is figuring out the relationship between the size of the pixels produced by our camera and real world length units such as microns.

In this Computational Exploration, the reader is invited to take the images shown in Figures 2.3(A) and (C) and to use them to determine how many nanometers are in a pixel. To that end, the idea is to read the image file from Figure 2.3(A) into Matlab and to produce a graph of the intensity for one row of the image as shown in Figure 2.3(B). Each one of the lines in the graticule shown in Figure 2.3(A) is separated by 10 μm. With this information in hand, the reader can work out the size of our pixels and hence install the missing scale bar on Figure 2.3(C), which is the ultimate objective of this Computational Exploration.

COMPUTATIONAL EXPLORATION

2.1.2 Taking the Molecular Census

In the remainder of this section, we will proceed through a variety of estimates to try to get a grip on the number of molecules of different kinds that are in an *E. coli* cell. Why should we care about these numbers? First, a realistic physical picture of any biological phenomenon demands a quantitative understanding of the individual particles involved (for biological phenomena, this usually means molecules) and the spatial dimensions over which they have the freedom to act. One of the most immediate outcomes of our cellular census will be the realization of just how crowded the cellular interior really is, a subject explored in detail in Chapter 14. Our census will paint a very different picture of the cellular interior as the seat of biochemical reactions than is suggested by the dilute and homogeneous environment of the biochemical test tube. Indeed, we will see that the mean spacing between protein molecules within a typical cell is less than 10 nm. As we will see below this distance is comparable to the size of the proteins themselves!

Taking the molecular census is also important because we will use our molecular counts in Chapter 3 to estimate the rates of macromolecular synthesis during the cell cycle. How fast is a genome replicated? What is the average rate of protein synthesis during the cell cycle and, given what we know about ribosomes, how do they maintain this rate of synthesis? A prerequisite to beginning to answer these questions is the macromolecular census itself.

Ultimately, to understand many experiments in biology, it is important to realize that most experimentation is comparative. That is, we compare "normal" behavior to perturbed behavior to see if some measurable property has increased or decreased. To make these statements meaningful, we need first to understand the quantitative baseline relative to which such increases and decreases are compared. There is another sense in which numbers of molecules are particularly meaningful which will be explored in detail in subsequent chapters that has to do with whether we can describe a cell as having "a lot" or "a few" copies of some specific molecule. If a cell has a lot of some particular molecule, then it is appropriate to describe the concentration of that molecule as the basis for predicting cellular function. However, when a cell has only a few copies of a particular molecule, then we need to consider the influence of random chance (or stochasticity) on its function. In many cases, cells have an interesting intermediate number of molecules where it is not immediately clear which perspective is appropriate. However, knowing the absolute numbers always gives us a reality check for subsequent assumptions and approximations for modeling biological processes.

Because of these considerations, much effort among biological scientists has been focused on the development of quantitative techniques for measuring the molecular census of living cells (both bacteria and eukaryotes). In this chapter, we will rely primarily on order-of-magnitude estimates based on simple assumptions. These estimates are validated by comparison with measurements. In subsequent chapters, these estimates will be refined through explicit model building and direct comparison with quantitative experiments.

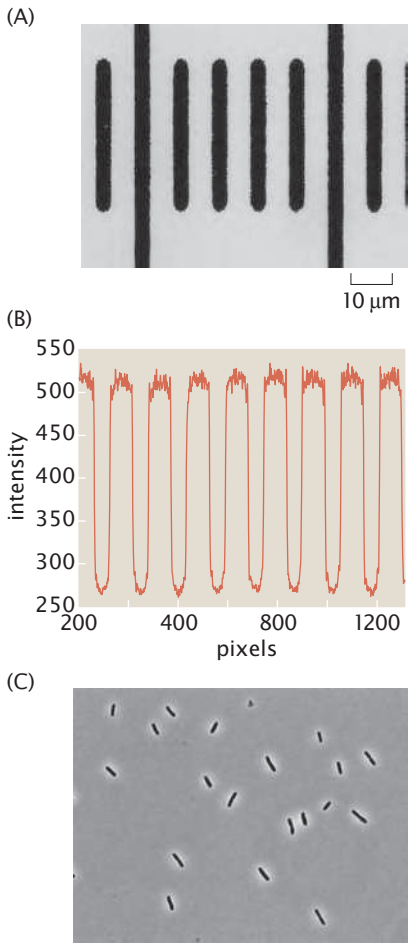


Figure 2.3: Sizing up *E. coli*. (A) Image of a graticule at 100× magnification. (B) Matlab plot of the intensity for a horizontal cut through the image. (C) Phase contrast image of a field of view with several *E. coli* cells taken at the same magnification as in (A).

Estimate: Sizing Up *E. coli* As already noted in the previous chapter, cells are made up of an array of different macromolecules as well as small molecules and ions. To estimate the number of proteins in an *E. coli* cell, we begin by noting that, with its 1 fL volume, the mass of such a cell is roughly 1 pg, where we have assumed that the density of the cell is that of water, which is 1 g/mL, though clearly Figure 2.2 shows that this assumption is not true. Measurements reveal that the dry weight of the cell is roughly 30% of its total and half of that mass is protein. As a result, the total protein mass within the cell is roughly 0.15 pg. We can also estimate the number of carbon atoms in a bacterium by considering the chemical composition of the macromolecules of the cell. This implies that roughly half the dry mass comes from the carbon content of these cells, a figure that reveals of the order of 10^{10} carbon atoms per cell. Two of the key sources that have served as a jumping off point for these estimates are Pedersen et al. (1978) and Zimmerman and Trach (1991), who describe the result of a molecular census of a bacterium.

ESTIMATE

As a first step toward revealing the extent of crowding within a bacterium, we can estimate the number of proteins

by assuming an average protein of 300 amino acids with each amino acid having a characteristic mass of 100 Da. These assumptions are further examined in the problems at the end of the chapter. Using these rules of thumb, we find that the mean protein has a mass of 30,000 Da. Using the conversion factor that $1\,\text{Da} \approx 1.6 \times 10^{-24}\,\text{g}$, we have that our typical protein has a mass of $5 \times 10^{-20}\,\text{g}$. The number of proteins per *E. coli* cell is estimated as

$$N_{\text{protein}} = \frac{\text{total protein mass}}{\text{mass per protein}} \approx \frac{15 \times 10^{-14}\,\text{g}}{5 \times 10^{-20}\,\text{g}} \approx 3 \times 10^{6}. \quad (2.1)$$

If we invoke the rough estimate that one-third of the proteins coded for in a typical genome correspond to membrane proteins, this implies that the number of cytoplasmic proteins is of the order of 2×10^6 and the number of membrane proteins is 10^6, although we note that not all of these membrane-associated proteins are strictly transmembrane proteins.

Another interesting use of this estimate is to get a rough impression of the number of ribosomes—the cellular machines that synthesize proteins. We can estimate the total number of ribosomes by first estimating the total mass of the ribosomes in the cell and then dividing by the mass per ribosome. To be concrete, we need one other fact, which is that roughly 20% of the protein complement of a cell is ribosomal protein. If we assume that all of this protein is tied up in assembled ribosomes, then we can estimate the number of ribosomes by noting that (a) the mass of an individual ribosome is roughly 2.5 MDa and (b) an individual ribosome is roughly one-third by mass protein and two-thirds by mass RNA, facts that can be directly confirmed by the reader by inspecting the structural biology of ribosomes. As a result, we have

$$\begin{aligned} N_{\text{ribosome}} &= \frac{0.2 \times 0.15 \times 10^{-12}\,\text{g}}{830{,}000\,\text{Da}} \times \frac{1\,\text{Da}}{1.6 \times 10^{-24}\,\text{g}} \\ &\approx 20{,}000 \text{ ribosomes.} \end{aligned} \quad (2.2)$$

The numerator of the first fraction has 0.2 as the fraction of protein that is ribosomal, 0.15 as the fraction of the total cell mass that is protein, and 1 pg as the cell mass. Our estimate for that part of the ribosomal mass that is protein is 830,000 Da. The size of a ribosome is roughly 20 nm (in "diameter") and hence the total volume taken up by these 20,000 ribosomes is roughly $10^8\,\text{nm}^3$. This is 10% of the total cell volume.

Idealizing an *E. coli* cell as a cube, sphere, or spherocylinder yields (see the problems) that the surface area of such cells is $A_{E.\,coli} \approx 6\,\mu\text{m}^2$. This number may be used in turn to estimate the number of lipid molecules associated with the inner and outer membranes of these cells as

$$N_{\text{lipid}} \approx \frac{4 \times 0.5 \times A_{E.\,coli}}{A_{\text{lipid}}} \approx \frac{4 \times 0.5 \times (6 \times 10^6\,\text{nm}^2)}{0.5\,\text{nm}^2} \approx 2 \times 10^7, \quad (2.3)$$

where the factor of 4 comes from the fact that the inner and outer membranes are each *bilayers*, implying that the lipids effectively cover the cell surface area four times. A lipid bilayer consists of two sheets of lipids with their tails pointing toward

each other. The factor of 0.5 is based on the crude estimate that roughly half of the surface area is covered by membrane proteins rather than lipids themselves. We have made the similarly crude estimate that the area per lipid is $0.5\,\mathrm{nm}^2$. The measured number of lipids is of the order of 2×10^7 as well.

In terms of sheer numbers, water molecules are by far the majority constituent of the cellular interior. One of the reasons this fact is intriguing is that during the process of cell division, a bacterium such as *E. coli* has to take on a very large number of new water molecules each second. The estimate we obtain here will be used to examine this transport problem in the next chapter. To estimate the number of water molecules, we exploit the fact that roughly 70% of the cellular mass (or volume) is water. As a result, the total mass of water is 0.7 pg. We can find the approximate number of water molecules as

$$\begin{aligned} N_{\mathrm{H_2O}} &\approx \frac{0.7 \times 10^{-12}\,\mathrm{g}}{18\ \mathrm{g/mol}} \times 6 \times 10^{23}\ \text{molecules/mol} \\ &\approx 2 \times 10^{10}\ \text{water molecules.} \end{aligned} \tag{2.4}$$

It is also of interest to gain an impression of the content of inorganic ions in a typical bacterial cell. To that end, we assume that a typical concentration of positively charged ions such as K^+ is 100 mM, resulting in the estimate

$$\begin{aligned} N_{\mathrm{ions}} &\approx \frac{(100 \times 10^{-3}\ \mathrm{mol}) \times (6 \times 10^{23}\ \text{molecules/mol})}{10^{15}\,\mu\mathrm{m}^3} \times 1\,\mu\mathrm{m}^3 \\ &= 6 \times 10^7. \end{aligned} \tag{2.5}$$

Here we use the fact that $1\,\mathrm{L} = 10^{15}\,\mu\mathrm{m}^3$. This result could have been obtained even more easily by noting yet another simple rule of thumb, namely, that one molecule per *E. coli* cell corresponds roughly to a concentration of 2 nM.

The outcome of our attempt to size up *E. coli* is illustrated schematically in summary form in Figure 2.4. A more complete census of an *E. coli* bacterium can be found in Neidhardt et al. (1990). The outcome of experimental investigations of the molecular census of an *E. coli* cell is summarized (for the purposes of comparing with our estimates) in Table 2.1.

How is the census of a cell taken experimentally? This is a question we will return to a number of different times, but will give a first answer here. For the case of *E. coli*, one important tool has been the use of gels like that shown in Figure 2.5. Such experiments work by breaking open cells and keeping only their protein components. The complex protein mixture is then separated into individual molecular species using a polyacrylamide gel matrix. First, the protein mixture is distributed through a tube-shaped polyacrylamide gel that has been polymerized to contain a stable pH gradient, and then an electric field is applied across the gel. The net charge on each protein depends on the pH and on the number and type of charged (protonatable) amino acid side chains that it contains. For example, the carboxylic acid group on aspartate will be negatively charged at high pH, but

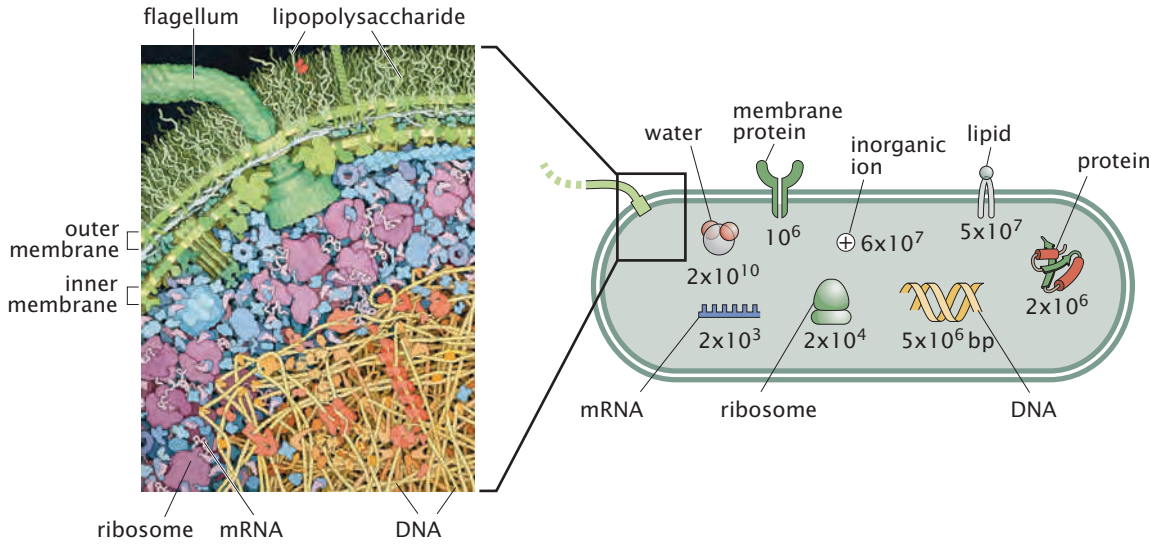


Figure 2.4: Molecular contents of the bacterium *E. coli*. The illustration on the left shows the crowded cytoplasm of the bacterial cell. The cartoon on the right shows an order-of-magnitude molecular census of the *E. coli* bacterium with the approximate number of different molecules in *E. coli*. (Illustration of the cellular interior courtesy of D. Goodsell.)

Table 2.1: Observed macromolecular census of an *E. coli* cell. (Data from F. C. Neidhardt et al., Physiology of the Bacterial Cell, Sinauer Associates, 1990 and M. Schaechter et al., Microbe, ASM Press, 2006.)

Substance	% of total dry weight	Number of molecules
Macromolecules		
Protein	55.0	2.4×10^6
RNA	20.4	
23S RNA	10.6	19,000
16S RNA	5.5	19,000
5S RNA	0.4	19,000
Transfer RNA (4S)	2.9	200,000
Messenger RNA	0.8	1,400
Phospholipid	9.1	22×10^6
Lipopolysaccharide (outer membrane)	3.4	1.2×10^6
DNA	3.1	2
Murein (cell wall)	2.5	1
Glycogen (sugar storage)	2.5	4,360
Total macromolecules	**96.1**	
Small molecules		
Metabolites, building blocks, etc.	2.9	
Inorganic ions	1.0	
Total small molecules	**3.9**	

will pick up a hydrogen ion and will be neutral at low pH. Conversely, the amine group on lysine will be neutral at high pH, but will pick up a hydrogen ion and will be positively charged at low pH. The pH where a protein's charge is net neutral is called its "isoelectric point."

When a protein finds itself in a region of the gel where the pH is below its isoelectric point, there will be an excess of positive charge associated with the protein, and it will move toward the cathode when the electric field is applied. When a protein finds itself in a region of the gel where the pH is above its isoelectric point, it will have a net negative charge, and will therefore move toward the anode. When

all the proteins in the mixture have moved to the location of the pH where each is neutral, then all movement stops. At this point, "isoelectric focusing" is complete, and all of the proteins are arrayed along the tube-shaped gel in positions according to their net charge. Next, a charged detergent is added that binds to all proteins so the total number of detergent molecules associated with an individual protein is roughly proportional to the protein's overall size. The detergent-soaked isolectric focusing gel tube is placed at one end of a flat, square gel that also contains detergent, and an electric field is applied in a direction perpendicular to the first field. Because the net charge on the detergent molecules is much larger than the original net charge of the protein, the rate of migration in the second direction through the gel is determined by the protein size.

After the procedure described above, the individual protein species in the original mixture have been resolved into a series of spots on the gel, with large, negatively charged proteins at the upper left-hand corner and small, positively charged proteins at the lower right-hand corner for the gel shown in Figure 2.5. The proteins can then be stained with a nonspecific dye so that their locations within the gel can be directly observed. The intensity of the spots on such a gel can then be used as a basis for quantifying each species. The identity of the protein that congregates in each spot can be determined by physically cutting each spot out of the gel, eluting the protein, and determining its size and amino acid content using mass spectrometry. Similar tricks are used to characterize the amount of RNA and lipids, for example, resulting in a total census like that shown in Table 2.1.

More recently, several other methods have been brought to bear on the molecular census of cells. As shown in Figure 2.6, two of these methods are the use of mass spectrometry and fluorescence microscopy. In mass spectrometry, fragments of the many proteins contained within the cell are run through the mass spectrometer and their absolute abundances determined. The results of this technique applied to *E. coli* are shown in Figure 2.6(A). Alternatively, in fluorescence microscopy, a library of cells is created where each strain in this library expresses a particular protein from the *E. coli* proteome fused to a fluorescent protein. By calibrating the fluorescence corresponding to an individual fluorescent protein, one can measure the amount

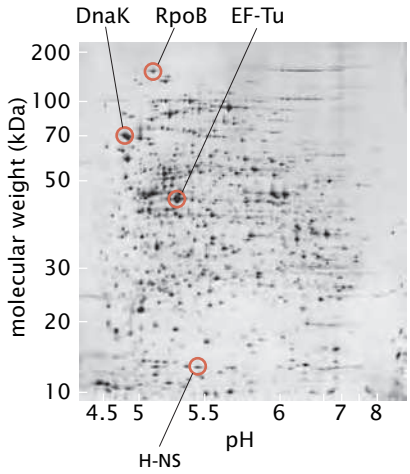


Figure 2.5: Protein census of the cell. Measurement of protein census of *E. coli* using two-dimensional polyacrylamide gel electrophoresis. Each spot represents an individual protein with a unique size and charge distribution. The spots arising from several well-known bacterial proteins are labeled. (Copyright Swiss Institute of Bioinformatics, Geneva, Switzerland.)

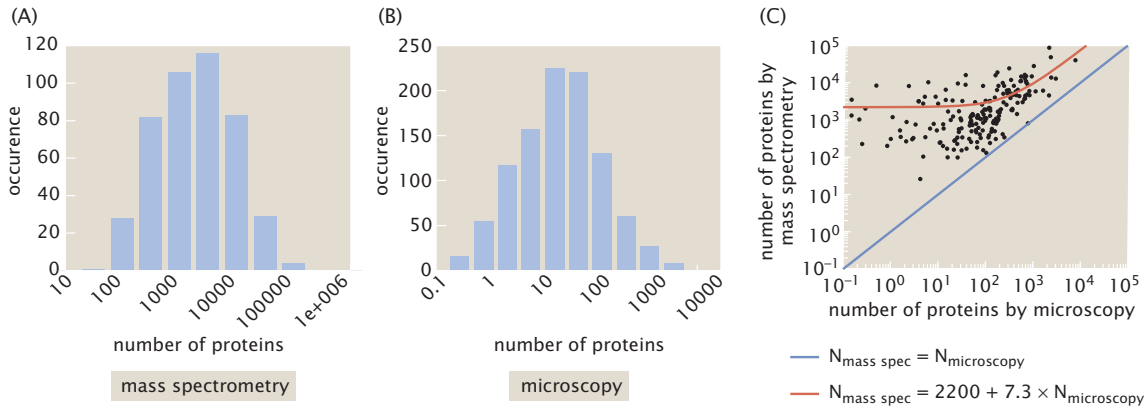


Figure 2.6: Protein census of *E. coli* using several techniques. A histogram of protein number in *E. coli* is shown from results using (A) mass spectrometry and (B) fluorescence microscopy with protein fusions to fluorescence proteins. (C) Comparison of mass spectrometry and fluorescence methods for the same collection of proteins showing a discrepancy between the two techniques as revealed by the fact that the best fit is not given by a line of slope one. (A, adapted from P. Lu et al., *Nat. Biotechnol.* 25:117, 2007; B, adapted from Y. Taniguchi et al., *Science* 329:533, 2010.)

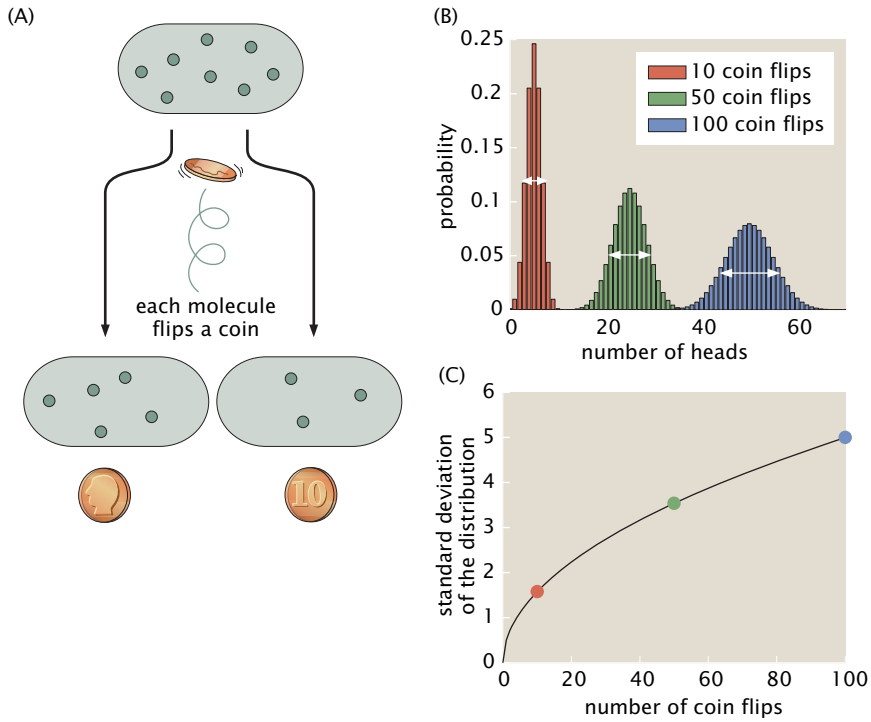


Figure 2.7: Schematic of the random partitioning of molecules during the process of cell division. (A) When the cell divides, each of the molecules chooses a daughter cell via a coin flip. (B) Probability distribution for the number of heads for different choices of the total number of coin flips. (C) The width of the distribution as a function of the number of coin flips.

of protein fusion in each strain of the library by comparing the total fluorescence in the cell with the single-molecule standard. Such a measurement for *E. coli* results in the histogram showed in Figure 2.6(B). It is important to point out, however, that these two methods are not in agreement, as shown in Figure 2.6(C). It is seen that, for a given protein, fluorescence microscopy tends to undercount proteins with respect to mass spectrometry (or mass spectrometry tends to overcount with respect to fluorescence microscopy). There are various possible sources for this discrepancy, ranging from systematic errors in the two experimental techniques to the fact that the experiments were done under different growth conditions leading to very different cell cycle times and cell sizes. One of the key features revealed by fluorescence measurements is the importance of cell-to-cell variability, which we can estimate using simple ideas from probability theory.

Estimate: Cell-To-Cell Variability in the Cellular Census

ESTIMATE

When cells divide, how much variability should we expect in the partitioning of the mRNA and protein contents of the different daughter cells? Conceptually, we can think of the passive molecular partitioning process as a series of coin flips in which each molecule of interest flips a coin to decide which of the two daughters it will go to. This idea is illustrated schematically in Figure 2.7.

To make a simple estimate of the variability, we turn to one of the most important probability distributions in all of science, namely, the binomial distribution. Despite the apparently contrived nature of coin flips, they are precisely the statistical experiment we need when thinking about cellular concentrations of molecules that are passively partitioned during cell division. The idea of the binomial distribution is that we carry out N trials of our coin flip process with the outcomes being heads and tails. However, a more biologically relevant

way to think of these outcomes is that with each cell division, the macromolecule of interest partitions either to daughter 1 or to daughter 2, with the probability of going to daughter 1 given by p and the probability of going to daughter 2 given by $q = 1 - p$. If we are dealing with a fair coin, or if the sizes of the two daughters are equal and there are no active segregation mechanisms in play, then both outcomes are equally likely and have probability $p = q = 1/2$.

To be precise, the probability of having n_1 of our N molecules partition to daughter cell 1 is given by

$$p(n_1, N) = \frac{N!}{n_1!(N - n_1)!} p^{n_1} q^{N-n_1}. \tag{2.6}$$

The factor $N!/[n_1!(N - n_1)!]$ reflects the fact that there are many different ways of flipping n_1 heads (h) and $N - n_1$ tails (t) and is given by the famed binomial coefficients (see Figure 2.8). For example, if $N = 3$ and $n_1 = 2$, then our three trials could turn out in three ways as *thh*, *hth*, and *hht*, exactly what we would have found by constructing $3!/(2!1!)$. Note also that throughout this book we will see that, in certain limits, the binomial distribution can be approximated by the Gaussian distribution (large-N case) or the Poisson distribution ($p \ll 1$ case).

If the number of copies of our mRNA or protein of interest is 10, then on division each daughter will get 5 of these molecules on average. More formally, this can be stated as

$$\langle n_1 \rangle = Np. \tag{2.7}$$

For the case of a fair coin (i.e., $p = 1/2$), this reduces to precisely the result we expect. This intuitive result can be appreciated quantitatively in the distributions shown in Figure 2.7(B). Here, the average number of heads obtained for a number of coin flips is given by the center of the corresponding distribution. The width of these distributions then gives a sense for how often there will be a result that deviates from the mean. As a result, to assess the cell-to-cell variability, we need to find a way to measure the width of the distribution. One way to characterize this width is by the standard deviation, which is computed from the function given in Equation 2.6 as follows:

$$\langle n_1^2 \rangle - \langle n_1 \rangle^2 = Npq, \tag{2.8}$$

a result the reader is invited to work out in the problems at the end of the chapter.

With these results in hand, we can now turn to quantifying the fluctuations that arise during partitioning. There is no one way to characterize the fluctuations, but a very convenient one is to construct the ratio

$$\frac{\sqrt{\langle n_1^2 \rangle - \langle n_1 \rangle^2}}{\langle n_1 \rangle} = \frac{1}{\sqrt{N}} \tag{2.9}$$

This result tells us that when the number of molecules is small (i.e., <100), the cell-to-cell variation is itself a significant fraction of the mean itself.

Such claims are not a mere academic curiosity and have been the subject of careful experimental investigation. On multiple occasions in the remainder of this book, we will examine partitioning data (for mRNA, for proteins, for carboxysomes,

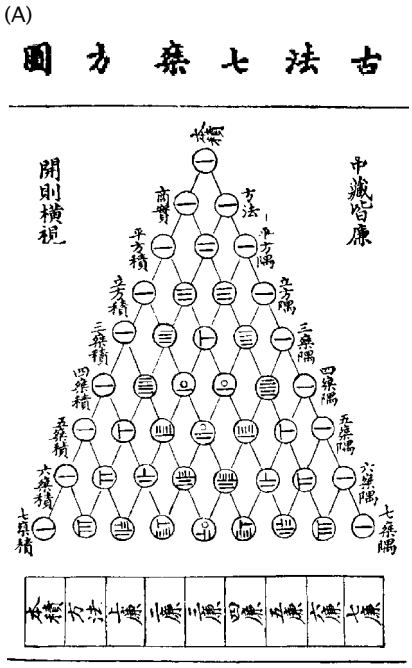


Figure 2.8: Binomial coefficients. (A) Though commonly known as Pascal's triangle, Chinese mathematicians had also earlier appreciated the symmetry of this collection of numbers. (B) The binomial coefficients that arise when raising $p + q$ to the power n, where n characterizes which row we are talking about, with the top "row" denoted by $n = 0$.

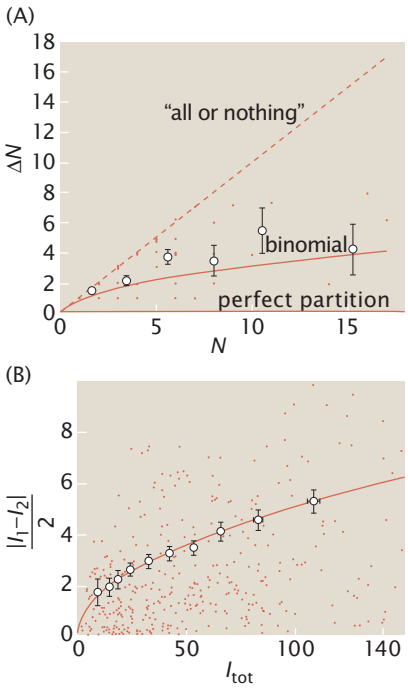


Figure 2.9: Binomial partitioning of mRNA and proteins in *E. coli* during cell division. (A) Difference ΔN in the number of mRNAs between the two daughters given that the mother has N mRNAs. The curves show three possible partitioning mechanisms involving "all or nothing" in which one daughter takes all of the mRNAs, binomial partitioning, and "perfect partitioning" in which each daughter gets exactly half of the proteins from the mother cell. (B) Difference in the fluorescence intensity of the two daughter cells for a fluorescent fusion to a repressor protein as a function of the fluorescence intensity of the mother cell. The line corresponds to binomial partitioning model. (A, adapted from I. Golding et al., *Cell* 123:1025, 2005; B, adapted from N. Rosenfeld et al., *Science* 307:1962, 2005.)

etc.) in which cell lineages are followed and the partitioning of the macromolecules or complexes of interest are measured directly. An example of such data for the case of mRNA and proteins is shown in Figure 2.9. As a result of the ability to fluorescently label individual mRNAs as they are produced, it is possible to characterize the number of such mRNAs in the mother cell before division and then to measure how they partition during the division process. Using fusions of fluorescent proteins to repressor molecules, similar measurements were taken in the context of proteins as shown in Figure 2.9(B). In both of these cases, the essential idea is to count the number of molecules in the mother cell and then to count the number in the two daughters and to see how different they are. Similar data is shown in Figure 18.7 (p. 724) for the case of the organelles in cyanobacteria known as carboxysomes, which contain the all-important enzyme Rubisco so critical to carbon fixation during the process of photosynthesis.

Computational Exploration: Counting mRNA and Proteins by Dilution As will be highlighted throughout this book, many of our key results depend critically upon the number of molecular actors involved in a given process. As shown in Figure 18.7 (p. 724), for example, fluorescence microscopy has made it possible to track the dynamics and localization of key cellular structures such as the carboxysomes. One of the main conclusions of that study was that the partitioning of the carboxysomes is not a random process since the *statistics* of the partitioning of these organelles between the two daughter cells does not follow the binomial distribution, which would be characteristic of random partitioning. The goal of this Computational Exploration is to explore the statistics of random partitioning from the perspective of simulations, with reference not only to the carboxysome example, but also to clever new ways of taking the molecular census. Using these simulations, we will show how the fluctuations in the partitioning of a fluorescent protein between daughter cells is related to the cells' intensity as schematized in Figure 2.10. Data from this type of experiment are shown in Figure 19.15 (p. 816).

The advent of fluorescent protein fusions now makes it possible to assess the relative quantities of a given protein of interest on the basis of the fluorescence of its fusion partner. However, for the purposes of the strict comparison between theory and experiment that we are often interested in, it is necessary to have a calibration that permits us to convert between the arbitrary fluorescence units as measured in a microscope and the more interesting molecular counts necessary for plugging into our various formulae or that we use in dissecting some molecular mechanism. To effect this calibration, we exploit the idea of a strict linear relation between the observed intensity and the number of fusion proteins as indicated by the relation

$$I_{\text{tot}} = \alpha N_{\text{tot}}, \tag{2.10}$$

where α is the unknown calibration factor linking fluorescence and the number of fluorescent molecules (N_{tot}).

As indicated in Figure 2.10, a series of recent experiments all using the same fluctuation method make it possible to

COMPUTATIONAL EXPLORATION

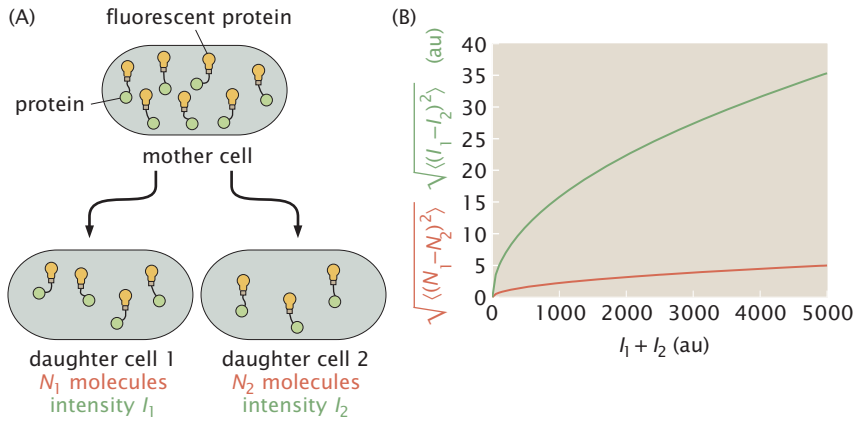


Figure 2.10: Intensity of the daughter cells as a result of binomial partitioning. (A) The mother cell starts out with a certain number of fluorescent proteins (represented by light bulbs) and upon division, these proteins are partitioned according to the binomial distribution. The number of fluorescent proteins and their corresponding total fluorescent intensity are related by the calibration factor α. (B) Plot of the intensity difference between the two daughters as a function of the intensity of the mother cell and the corresponding difference in partitioning of fluorescent molecules as a function of the total number of molecules in the mother cell.

actually determine N_{tot} by determining the unknown calibration factor. The idea of the experiment is to exploit the natural fluctuations resulting from imperfect partitioning of the fluorescent proteins between the two daughter cells. The key point is that for the case in which the partitioning between the two daughters is random (i.e., analogous to getting heads or tails in a coin flip), the size of those fluctuations depends in turn upon the total number of proteins being segregated between the two daughters. The larger the number of proteins, the larger the fluctuations. In particular, as shown in the problems at the end of the chapter, we can derive the simple and elegant result that the average difference in intensity between the two daughter cells is given by

$$\langle (I_1 - I_2)^2 \rangle = \alpha I_{tot}, \tag{2.11}$$

where I_1 and I_2 are the intensities of daughters 1 and 2, respectively, and I_{tot} is the total fluorescence intensity of the mother cell ($I_{tot} = I_1 + I_2$).

The simple analytic result of Equation 2.11 can perhaps be best appreciated through an appeal to a stochastic simulation. The idea is to "simulate" the dilution of fluorescent proteins during repeated cell divisions where all the mother cells are presumed to start with N copies of the fluorescent protein and the partitioning of those proteins to the two daughters is determined strictly by a simple binary random process (i.e., heads or tails of a coin flip of an honest coin). Conceptually, the algorithm for performing this simulation can be summarized as follows:

1. Choose the number of fluorescent molecules N in the mother cell and compute the intensity $I_{tot} = \alpha N$.
2. Generate N random numbers between 0 and 1. If a given random number is between 0 and 0.5, assign the protein to daughter 1 and otherwise assign the protein to daughter 2.
3. Compute the "intensity" of the daughters by multiplying N_1 and N_2 by α.
4. Perform the same operation for 100 cells (for example).
5. Choose a new N and repeat the process.

By considering the same division process again and again, for different choices of N, we can generate a curve that captures

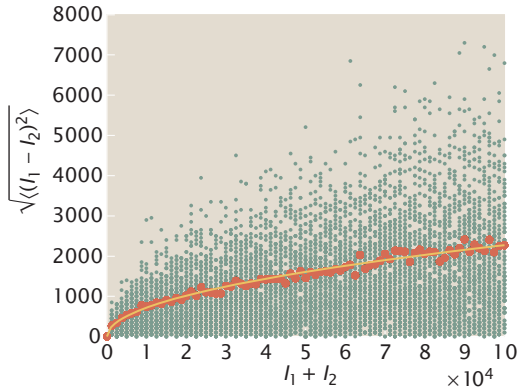


Figure 2.11: Simulation results of fluorescent protein dilution and counting experiment. For a given total number of proteins, N, in the mother cell, we simulate 100 different outcomes of partitioning by cellular division and plot their corresponding partitioning error (green dots). For each value of N, we can calculate the root mean square in partitioning (red dots) and fit them to the expected functional form given by the square root of Equation 2.11.

the statistics set forth above. In particular, for each and every N, 100 separate "divisions" were simulated and the results of each such division process are shown as the data points. For a given value of N, we can calculate the root mean square error in partitioning and its corresponding intensity, given by $\sqrt{\langle (I_1 - I_2)^2 \rangle}$. To see an example of the outcome of such a simulation, examine Figure 2.11.

Examples of Matlab code that could be used to perform this 'Computational Exploration' can be found on the book's, website.

The Cellular Interior Is Highly Crowded, with Mean Spacings Between Molecules That Are Comparable to Molecular Dimensions

One of the most intriguing implications of our census of the molecular parts of a bacterium is the extent to which the cellular interior is crowded. Because of experiments and associated estimates on the contents of *E. coli*, it is now possible to construct illustrations to depict the cellular interior in a way that respects the molecular census. The crowded environs of the interior of such a cell are shown in Figure 2.4. This figure gives a view of the crowding associated with any *in vivo* process. In Chapter 14, we will see that this crowding effect will force us to call in question our simplest models of chemical potentials, the properties of water, and the nature of diffusion. The generic conclusion is that the mean spacing of proteins and their assemblies is comparable to the dimensions of these macromolecules themselves. The cell is a very crowded place!

The quantitative significance of Figure 2.4 can be further appreciated by converting these numbers into concentrations. To do so, we recall that the volume of an *E. coli* cell is roughly 1 fL. The rule of thumb that emerges from this analysis is that a concentration of 2 nM implies roughly one molecule per bacterium. A concentration of 2 μM implies roughly 1000 copies of that molecule per cell. Concentration in terms of our standard ruler is shown in Figure 2.12. This figure shows the number of copies of the molecule of interest in such a cell as a function of the concentration.

We can use these concentrations directly to compute the mean spacing between molecules. That is, given a certain concentration, there is a corresponding average distance between the molecules. Having a sense of this distance can serve as a guide to thinking about the likelihood of diffusive encounters and reactions between various molecular constituents. If we imagine the molecules at a given concentration arranged on a cubic lattice of points, then the mean spacing between

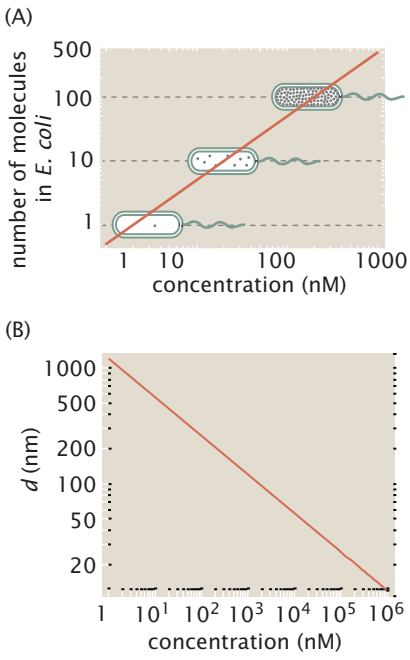


Figure 2.12: Physical interpretation of concentration. (A) Concentration in *E. coli* units: number of copies of a given molecule in a volume the size of an *E. coli* cell as a function of the concentration. (B) Concentration expressed in units of typical distance d between neighboring molecules measured in nanometers.

those points is given by

$$d = c^{-1/3}, \tag{2.12}$$

where c is the concentration of interest (measured in units of number of molecules per unit volume). Larger concentrations imply smaller intermolecular spacings. This idea is formalized in Figure 2.12(B), which shows the relation between the mean spacing measured in nanometers and the concentration in nanomolar units.

2.1.3 Looking Inside Cells

With our reference bacterium in mind, the remainder of the chapter focuses on the various structures that make up cells and organisms. To talk about these structures, it is helpful to have a sense of how we know what we know about them. Further, model building requires facts. To that end, we periodically take stock of the experimental basis for our models. For this chapter, the "Experiments Behind the Facts" focuses on how biological structures are explored and measured.

EXPERIMENTS

Experiments Behind the Facts: Probing Biological Structure To size up cells and their organelles, we need to extract "typical" structural parameters from a variety of experimental studies. Though we leave a description of the design and setup of such experiments to more specialized texts, the goal is to provide at least enough detail so that the reader can see where some of the key structural facts that we will use throughout the book originated. We emphasize two broad categories of experiments: (i) those in which some form of radiation interacts with the structure of interest and (ii) those in which forces are applied to the structure of interest.

Figure 2.13 shows three distinct experimental strategies that feed into our estimates, all of which reveal different facets of biological structure. One of the mainstays of structural analysis is light microscopy. Figure 2.13(A) shows a schematic of the way in which light can excite fluorescence in samples that have some distribution of fluorescent molecules within them. In particular, this example shows a schematic of a microtubule that has some distribution of fluorophores along its length. Incident photons of one wavelength are absorbed by the fluorophores and this excitation leads them to emit light of a different wavelength, which is then detected. As a result of selective labeling of only the microtubules with fluorophores, it is only these structures that are observed when the cell is examined in the microscope. These experiments permit a determination of the size of various structures of interest, how many of them there are, and where they are localized. By calibrating the intensity from single fluorophores, it has become possible to take a single-molecule census for many of the important proteins in cells. For an example of this strategy, see Wu and Pollard (2005). A totally different window on the structure of the cell and its components is provided by tools such as the atomic-force microscope (AFM). As will be explained in Chapter 10, the AFM is a cantilever beam with a sharp tip on its end. The tip is brought very close to the surface where the structure of interest is present and is then scanned in the

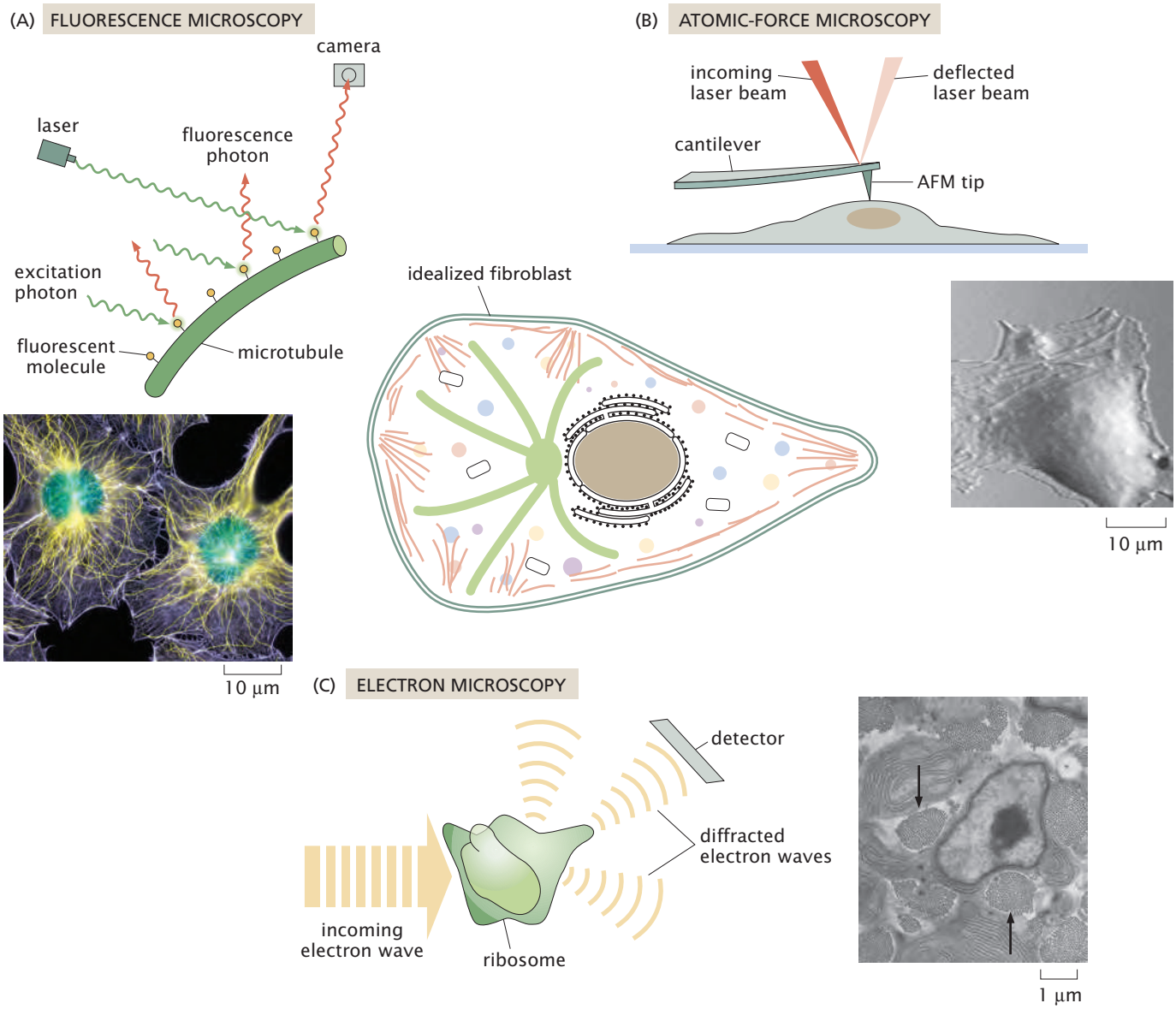


Figure 2.13: Experimental techniques that have revealed the structure of both cells and their organelles. (A) Fluorescence microscopy and the associated image of a fibroblast with labeled microtubules (yellow) and DNA (green). (B) Atomic-force microscopy schematic and the associated image of the surface topography of a fibroblast. (C) Electron microscopy schematic and image of cross-section through a fibroblast in an animal tissue. Arrows indicate bundles of collagen fibers. (A, courtesy of Torsten Wittman; B, adapted from M. Radmacher, *Meth. Cell Biol.* 83:347, 2007; C, adapted from D. E. Birk and R. L. Trelstad, *J. Cell Biol.* 103:231, 1986.)

plane of the sample. One way to operate the instrument is to move the cantilever up and down so that the force applied on the tip remains constant. Effectively, this demands a continual adjustment of the height as a function of the x–y position of the tip. The nonuniform pattern of cantilever displacements can be used to map out the topography of the structure of interest. Figure 2.13(B) shows a schematic of an AFM scanning a typical fibroblast cell as well as a corresponding image of the cell.

Figure 2.13(C) gives a schematic of the way in which X-rays or electrons are scattered off a biological sample. The schematic shows an incident plane wave of radiation that interacts with the biological specimen and results in the emergence

of radiation with the same wavelength but a new propagation direction. Each point within the sample can be thought of as a source of radiation, and the observed intensity at the detector reflects the interference from all of these different sources. By observing the pattern of intensity, it is possible to deduce something about the structure that did the scattering. This same basic idea is applicable to a wide variety of radiation sources, including X-rays, neutrons, and electrons.

An important variation on the theme of measuring the scattered intensity from irradiated samples is cryo-electron tomography. This technique is one of the centerpieces of structural biology and is built around uniting electron microscopy with sample preparation techniques that rapidly freeze the sample. The use of tomographic methods has made it possible to go beyond the planar sections seen in conventional electron microscopy images. The basis of the technique is indicated schematically in Figure 2.14, and relies on rotating the sample over a wide range of orientations and then assembling a corresponding three-dimensional reconstruction on the basis of the entirety of these images. This technique has already revolutionized our understanding of particular organelles and is now being used to image entire cells.

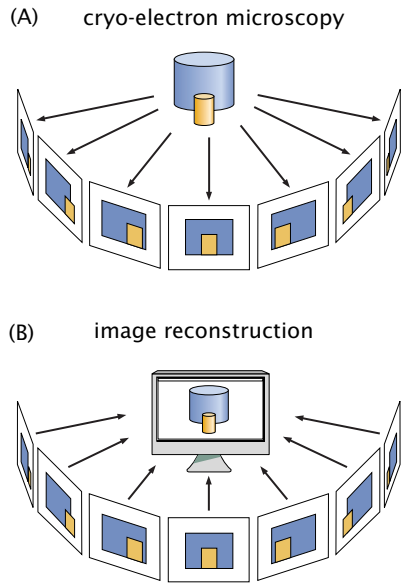


Figure 2.14: Schematic of tomographic reconstruction. (A) The three-dimensional sample is rotated in the electron microscope and imaged repeatedly to generate a series of two-dimensional images representing the pattern of radiation scattered at slightly different orientations. (B) Information from the image series is recombined computationally to generate a three-dimensional rendering of the original object.

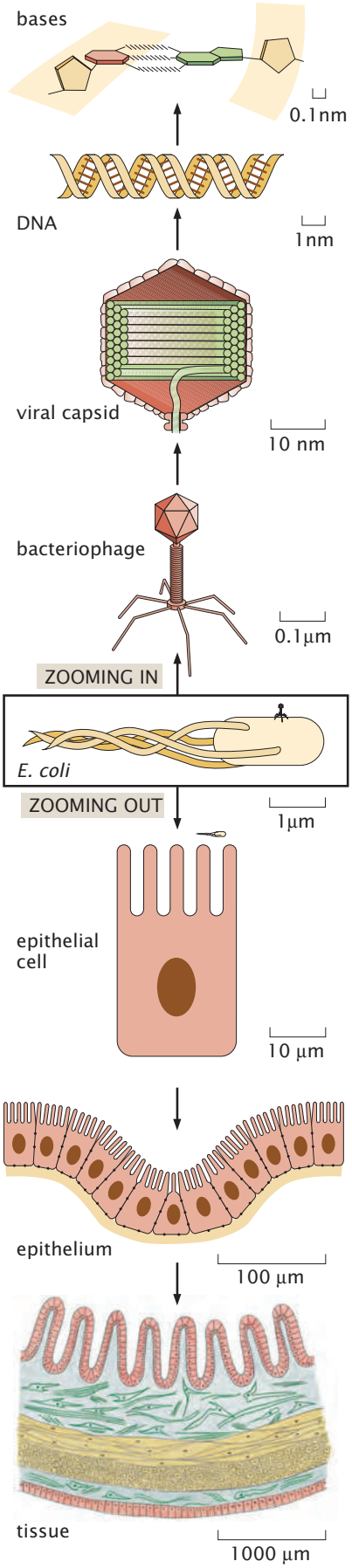
2.1.4 Where Does *E. coli* Fit?

Biological Structures Exist Over a Huge Range of Scales

The spatial scales associated with biological structures run from the nanometer scale of individual molecules all the way to the scale of the Earth itself. Where does *E. coli* fit into this hierachy of structures? Figure 2.15 shows the different structures that can be seen as we scale in and out from an *E. coli* cell. A roughly 10-fold increase in magnification relative to an individual bacterium reveals the viruses that attack bacteria. These viruses, known as bacteriophages, have a characteristic scale of roughly 100 nm. They are made up of a protein shell (the capsid) that is filled with the viral genome. Continuing our downward descent using yet higher magnification, we see the ordered packing of the viral genome within its capsid. These structures are intriguing because they involve the ordered arrangement of more than 10 μm of DNA in a capsid that is less than 100 nm across. Another rough factor-of-10 increase in resolution reveals the structure of the DNA molecule itself, with a characteristic cross-sectional radius of roughly 1 nm and a length of 3.4 nm per helical repeat.

A similar scaling out strategy reveals new classes of structures. As shown in Figure 2.15, a 10-fold increase in spatial scale brings us to the realm of eukaryotic cells in general, and specifically, to the scale of the epithelial cells that line the human intestine. We use this example because bacteria such as *E. coli* are central players as part of our intestinal ecosystem. Another 10-fold increase in spatial scale reveals one of the most important inventions of evolution, namely, multicellularity. In this case, the cartoon depicts the formation of planar sheets of epithelial cells. These planar sheets are themselves the building blocks of yet higher-order structures such as tissues and organs. Scaling out to larger scales would bring us to multicellular organisms and the structures they build.

The remainder of the chapter takes stock of the structures at each of these scales and provides a feeling for the molecular building blocks



that make up these different structures. Our strategy will be to build upon our cell-centered view and to first descend in length scale from that of cells to the molecules of which they are made. Once this structural descent is complete, we will embark on an analysis of biological structure in which we zoom out from the scale of individual cells to collections of cells.

2.2 Cells and Structures within Them

2.2.1 Cells: A Rogue's Gallery

All living organisms are based on cells as the indivisible unit of biological organization. However, within this general rule there is tremendous diversity among living cells. Several billion years ago, our last common ancestor gave rise to three different lineages of cells now commonly called Bacteria, Archaea, and Eukarya, a classification suggested by similarities and differences in ribosomal RNA sequences. Every living organism on Earth is a member of one of these groups. Most bacteria and archaea are small (3 μm or less) and extremely diverse in their preferred habitats and associated lifestyles ranging from geothermal vents at the bottom of the ocean to permafrost in Antarctica. Bacteria and Archaea look similar to one another and it has only been within the last few decades that molecular analysis has revealed that they are completely distinct lineages that are no more closely related to each other than the two are to Eukarya.

The organisms that we most often encounter in our everyday life and can see with the naked eye are members of Eukarya (individuals are called eukaryotes). These include all animals, all plants ranging from trees to moss, and also all fungi, such as mushrooms and mold. Thus far, we have focused on *E. coli* as a representative cell, although we must acknowledge that *E. coli*, as a member of the bacterial group, is in some ways very different from a eukaryotic or archaeal cell. The traditional definition of a eukaryotic cell is one that contains its DNA genome within a membrane-bound nucleus. Most bacteria and archaea lack this feature and also lack other elaborate intracellular membrane-bound structures such as the endoplasmic reticulum and the Golgi apparatus that are characteristic of the larger and more complex eukaryotic cells.

Cells Come in a Wide Variety of Shapes and Sizes and with a Huge Range of Functions

Cells come in such a wide variety of shapes, sizes, and lifestyles that choosing one representative cell type to tell their structural story is misleading. In Figure 2.16, we show a rogue's gallery illustrating a small segment of the variety of cell sizes and shapes found in the eukaryotic group, all referenced to the *E. coli* standard ruler. This gallery is by no means complete. There is much more variety than we can illustrate, but this covers a reasonable range of eukaryotic cell

Figure 2.15: Powers-of-10 representation of biological length scales. The hierarchy of scales is built around the *E. coli* standard ruler. Starting with *E. coli*, Section 2.2 considers a succession of 10-fold increases in resolution as are shown in the figure. Section 2.3 zooms out from the scale of an *E. coli* cell.

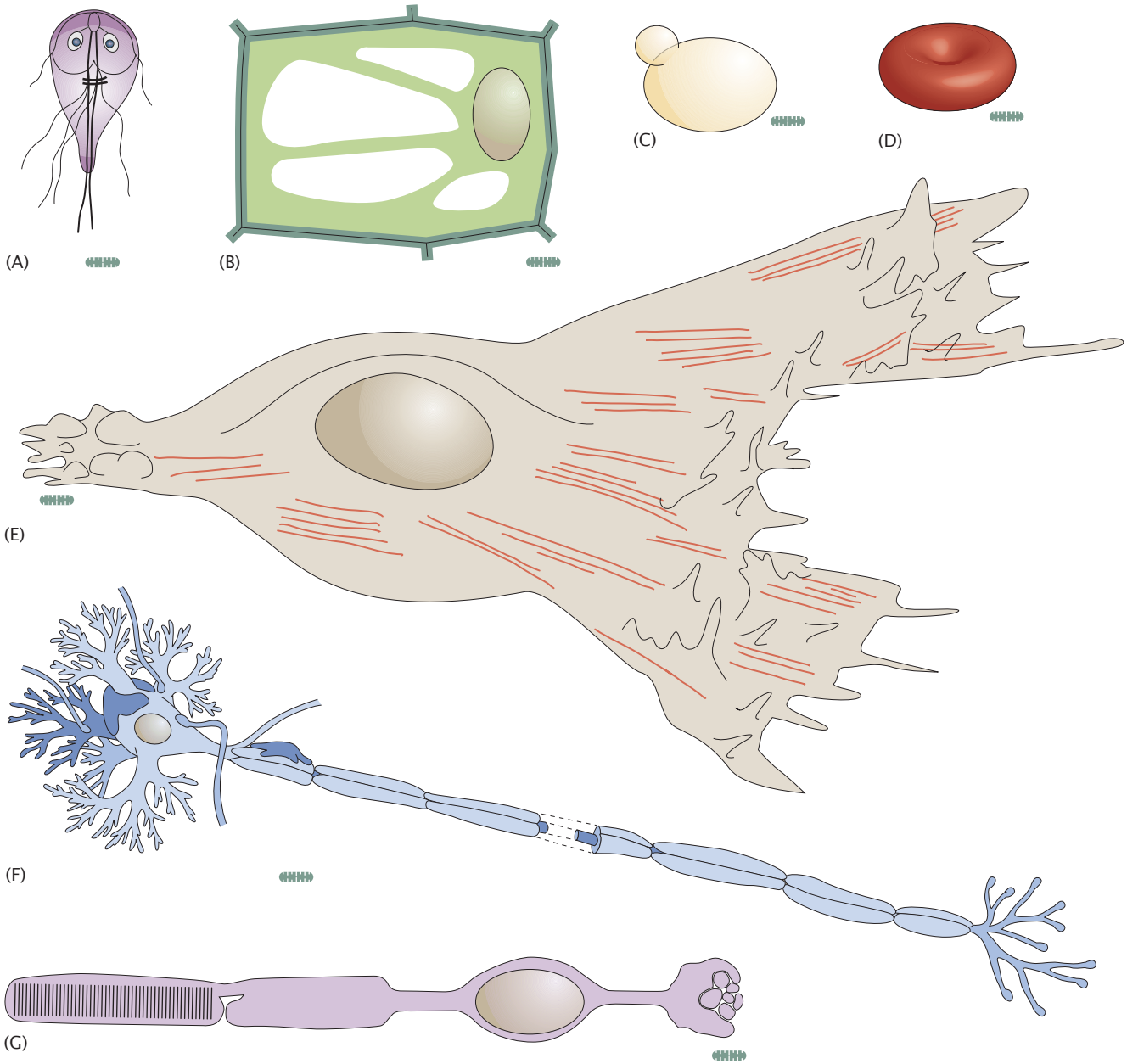


Figure 2.16: Cartoons of several different types of cells all referenced to the standard *E. coli* ruler. (A) The protist *Giardia lamblia*, (B) a plant cell, (C) a budding yeast cell, (D) a red blood cell, (E) a fibroblast cell, (F) a eukaryotic nerve cell, and (G) a retinal rod cell.

types that have been well studied by biologists. In this figure, we have chosen a variety of examples that represent experimental bias among biologists where more than half of the examples are human cells and the others represent the rest of the eukaryotic group.

The vast majority of eukaryotes are members of a group called protists, as shown in Figure 2.17. This poorly defined group encompasses all eukaryotes that are neither plants nor animals nor fungi. Protists are extremely diverse in their appearance and lifestyles, but they are all small (ranging from 0.002 mm to 2 mm). Some examples of protists are marine plankton such as *Emiliana huxleyi*, soil amoebae such as *Dictyostelium discoideum*, and the lovely creature *Paramecium* seen in any sample of pond water and familiar from many high-school biology

Figure 2.17: Protist diversity. This figure illustrates the morphological diversity of free-living protists. The various organisms are drawn to scale relative to the head of a pin. (Adapted from B. J. Finlay, *Science* 296:1061, 2002.)

classes. Another notable protist is the pathogen that causes malaria, called *Plasmodium falciparum*. Figure 2.16(A) shows the intriguing protist *Giardia lamblia*, a parasite known to hikers as a source of water contamination that causes diarrhea.

Although protists constitute the vast majority of eukaryotic cells on the planet, biologists are often inclined to study cells more closely related to or directly useful to us. This includes the plant kingdom, which is obviously important as a source of food and flowers. Plant cells like that shown in Figure 2.16(B) are characterized by a rigid cell wall, often giving them angular structures like that shown in the figure. The typical length scale associated with these cells is often tens of microns. One of the distinctive features of plant cells is their large vacuoles within the intracellular space that hold water and contribute to the mechanical properties of plant stems. These large vacuoles are very distinct from animal cells, where most of the intracellular space is filled with cytoplasm. Consequently, in comparing a plant and an animal cell of similar overall size, the plant cell will have typically 10-fold less cytoplasmic volume because most of its intracellular space is filled with vacuoles. Hydrostatic forces matter much more to plants than animals. For example, a wilting flower can be revived simply by application of water since this allows the vacuoles to fill and stiffen the plant stem.

Among the eukaryotes, the group most closely related to the animals (as proved by ribosomal RNA similarity and many other lines of evidence) is, surprisingly, the fungi. The representative fungus shown in Figure 2.16 is the budding yeast *Saccharomyces cerevisiae* (which we will refer to as *S. cerevisiae*). *S. cerevisiae* was domesticated by humans several thousand years ago and continues to serve as a treasured microbial friend that makes our bread rise and provides alcohol in our fermented beverages such as wine. Just as *E. coli* often serves as a key model prokaryotic system, the yeast cell often serves as the model single-celled eukaryotic organism. Besides the fact that humans are fond of *S. cerevisiae* for its own intrinsic properties, it is also useful to biologists as a representative fungus. Of all the other organisms on Earth, fungi are closest to animals in terms of evolutionary descent and similarity of protein functions. Although there are no single-celled animals, there are some single-celled fungi, including

S. cerevisiae. Therefore, many laboratory experiments relying on rapid replication of single cells are most easily performed on this organism and its relatives *Candida albicans* and *Schizosaccharomyces pombe* (a "fission" yeast that divides in the middle). Figure 2.18 shows a scanning electron microscope image of a yeast cell engaged in budding. As this image shows, the geometry of yeast is relatively simple compared with many other eukaryotic cells and it is also a fairly small member of this group, with a characteristic diameter of roughly 5 μm. Nonetheless, it possesses all the important structural hallmarks of the eukaryotes, including, in particular, a membrane-bound nucleus, separating the DNA genome from the cytoplasmic machinery that performs most metabolic functions.

Earlier, we estimated the molecular census of an *E. coli* cell. It will now be informative to compare those estimates with the corresponding model eukaryotic cell that will continue to serve as a comparative basis for all our eukaryotic estimates.

Figure 2.18: Microscopy images of yeast and their organelles. (A) Scanning electron micrograph of the yeast *Candida albicans* revealing the overall size scale of budding yeast. (B) Electron microscopy image of a cross-section through a budding *Candida albicans* yeast cell. (C) Confocal microscopy images of the mitochondria of *S. cerevisiae* under different growth conditions. (A, courtesy of Ira Herskowitz and Eric Schabtach; B, adapted from G. M. Walker, Yeast, Physiology and Biotechnology, John Wiley, 1998; C, adapted from A. Egner et al., *Proc. Natl Acad. Sci. USA* 99:3370, 2002.)

Estimate: Sizing Up Yeast The volume of a yeast cell can be computed in *E. coli* volume units, $V_{E.\ coli}$. In particular, if we recall that $V_{E.\ coli} \approx 1.0\,\mu m^3 = 1.0$ fL and think of yeast as a sphere of diameter 5 μm, then we have the relation $V_{yeast} \approx 60\,V_{E.\ coli}$; that is, roughly 60 *E. coli* cells would fit inside of a yeast cell. This result is consistent with experimental observations of the yeast cell size distribution such as shown in Figure 2.19. The surface area of a yeast cell can be estimated using a radius of $r_{yeast} \approx 2.5\,\mu m$, which yields $A_{yeast} \approx 80\,\mu m^2$. If we treat the yeast nucleus as a sphere with a diameter of roughly 2.0 μm, its volume is roughly $4\,\mu m^3$. Within this nucleus is housed the 1.2×10^7 bp of the yeast genome, which is divided among 16 chromosomes. The DNA in yeast is packed into higher-order structures mediated by protein assemblies known as histones. In particular, the DNA is wrapped around a series of cylindrical cores made up of eight such histone proteins each, with roughly 150 bp wrapped around each histone octamer, and approximately a 50 bp spacer between. As a result, we can estimate the number of nucleosomes (the histone–DNA complex) as

$$N_{nucleosome} \approx \frac{12 \times 10^6\,bp}{200\,bp/nucleosome} \approx 60{,}000. \qquad (2.13)$$

Experimentally, the measured number appears to be closer to 80,000, with a mean spacing between nucleosomes of the order

ESTIMATE

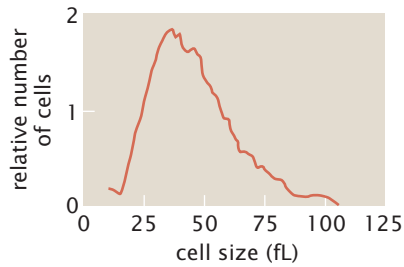


Figure 2.19: Yeast cell size distribution. Distribution of cell volumes measured for wild-type yeast cells. (Adapted from P. Jorgensen et al., *Science* 297:395, 2002.)

of 170 bp. The total volume taken up by the histones is roughly $230\,\text{nm}^3$ per histone (thinking of each histone octamer as a cylindrical disk of radius 3.5 nm and height 6 nm), resulting in a total volume of $14 \times 10^6\,\text{nm}^3$ taken up by the histones. The volume taken up by the genome itself is comparable at $1.2 \times 10^7\,\text{nm}^3$, where we have used the rule of thumb that the volume per base pair is $1\,\text{nm}^3$. The packing fraction (defined as the ratio of the volume taken up by the genome to the volume of the nucleus) associated with the yeast genomic DNA can be estimated by evaluating the ratio

$$\rho_{\text{pack}} \approx \frac{(1.2 \times 10^7\,\text{bp}) \times (1\,\text{nm}^3/\text{bp})}{4 \times 10^9\,\text{nm}^3} \approx 3 \times 10^{-3}. \tag{2.14}$$

Note that we have used the fact that the yeast genome is 1.2×10^7 bp in length and is packed in the nucleus, which has a volume $\approx 4\,\mu\text{m}^3$.

These geometric estimates may be used to make corresponding molecular estimates, such as the number of lipids and proteins in a typical yeast cell. The number of proteins can be estimated in several ways—perhaps the simplest is just to assume that the fractional occupancy of yeast cytoplasm is identical to that of *E. coli*, with the result that there will be 60 times as many proteins in yeast as in *E. coli* based strictly on scaling up the cytoplasmic volume. This simple estimate is obtained by *assuming* that the composition of the yeast interior is more or less the same as that of an *E. coli* cell. This strategy results in

$$N_{\text{protein}}^{\text{yeast}} \approx 60 \times N_{\text{protein}}^{\textit{E. coli}} \approx 2 \times 10^8. \tag{2.15}$$

The number of lipids associated with the plasma membrane of the yeast cell can be obtained as

$$N_{\text{lipid}} \approx \frac{2 \times 0.5 \times A_{\text{yeast}}}{A_{\text{lipid}}} \approx \frac{2 \times 0.5 \times (80 \times 10^6\,\text{nm}^2)}{0.5\,\text{nm}^2} \approx 2 \times 10^8, \tag{2.16}$$

where the factor of 0.5 is based on the idea that roughly half of the surface area is covered by membrane proteins rather than lipids themselves and the factor of 2 accounts for the fact that the membrane is a bilayer. This should be contrasted with the situation in *E. coli*, which has a double membrane surrounding the cytoplasm.

Another interesting estimate suggested by Figure 2.18(C) is associated with the organellar content of these cells. In particular, this figure shows the mitochondria of yeast that are being grown in two different media. These pictures suggest several interesting questions such as what fraction of the cellular volume is occupied by mitochondria and what is the surface area tied up with the mitochondrial outer membranes? Using the image of the cells grown on glucose in the left of Figure 2.18(C), we see that the mitochondria have a reticular form made up of several long tubes with a radius of roughly $r = 100$ nm. These tubes are nearly circular like the inner tubes of a bicycle with

a radius of curvature of approximately $R = 1\,\mu$m, meaning that each such tube has a length of roughly $L = 2\pi R \approx 6\,\mu$m, for a total tube length of 12 μm given that there appear to be two such tubes. Hence, the volume taken up by the mitochondria is $V = \pi r^2(2L) \approx 0.5\ \mu\text{m}^3$, with an area of $A = 2\pi r(2L) \approx 6\ \mu\text{m}^2$, meaning that about 1% of the volume of the cells grown on glucose is taken up by their mitochondria and an area is tied up in the mitochondrial outer membrane that is roughly 10% of the total plasma membrane area. In other cell types, the fraction of the area taken up by the mitochondria can be much larger, as will be seen in Chapter 11 (p. 427).

Our estimates are brought into sharpest focus when they are juxtaposed with actual measurements. The census of yeast cells has been performed in several distinct and fascinating ways. For a series of recent studies, the key idea is to generate thousands of different yeast strains, a strain library, each of which has a tag on a different one of the yeast gene products. Each yeast protein can then be specifically recognized by antibodies in the specific strain where it has been tagged. A second scheme is to construct protein fusions in which the protein of interest is attached to a fluorescent protein such as the green fluorescent protein (GFP). Then, by querying each and every yeast strain in the library either by examining the extent of antibody binding or fluorescence, it is possible to count up the numbers of each type of protein. Figure 2.20(B) shows a histogram of the number of proteins that occur with a given protein copy number in yeast. Although the average protein appears to be present in the cell with a few thousand copies, this histogram shows that some proteins are present with fewer than 50 copies and others with more than a million copies. Further, similarly quantifying mRNA as shown in Figure 2.20(C) reveals that many genes, including essential genes, are expressed with an average of less than one molecule of RNA per cell. By adding up the total number of proteins on the basis of this census, we estimate there are 50×10^6 proteins in a yeast cell, somewhat less than suggested by our crude estimate given above. Now that we have completed our first introduction to the important yeast as a representative eukaryote, we return to our tour of cell types in Figure 2.16.

Cells from Humans Have a Huge Diversity of Structure and Function

The remainder of the cells in Figure 2.16 are all human cells and show another interesting aspect of cellular diversity. To a first approximation, every cell in the human body contains the same DNA genome. And yet, individual human cells differ significantly with respect to their sizes (with sizes varying from roughly 5 μm for red blood cells to 1 m for the largest neurons), shapes, and functions. For example, rod cells in the retina are specialized to detect incoming light and transmit that information to the neural system so that we can see. Red blood cells are primarily specialized as carriers of oxygen and, in fact, are dramatically different from almost all other cells in having dispensed with their nucleus as part of their developmental process. As we will discuss extensively throughout the book, other cells have other unique classes of specialization.

Figure 2.16(D) is a schematic of the structure of a red blood cell. Note that the shapes of these cells are decidedly not spherical, raising interesting questions about the mechanisms of cell-shape maintenance. Despite their characteristic size of order 5 μm, these

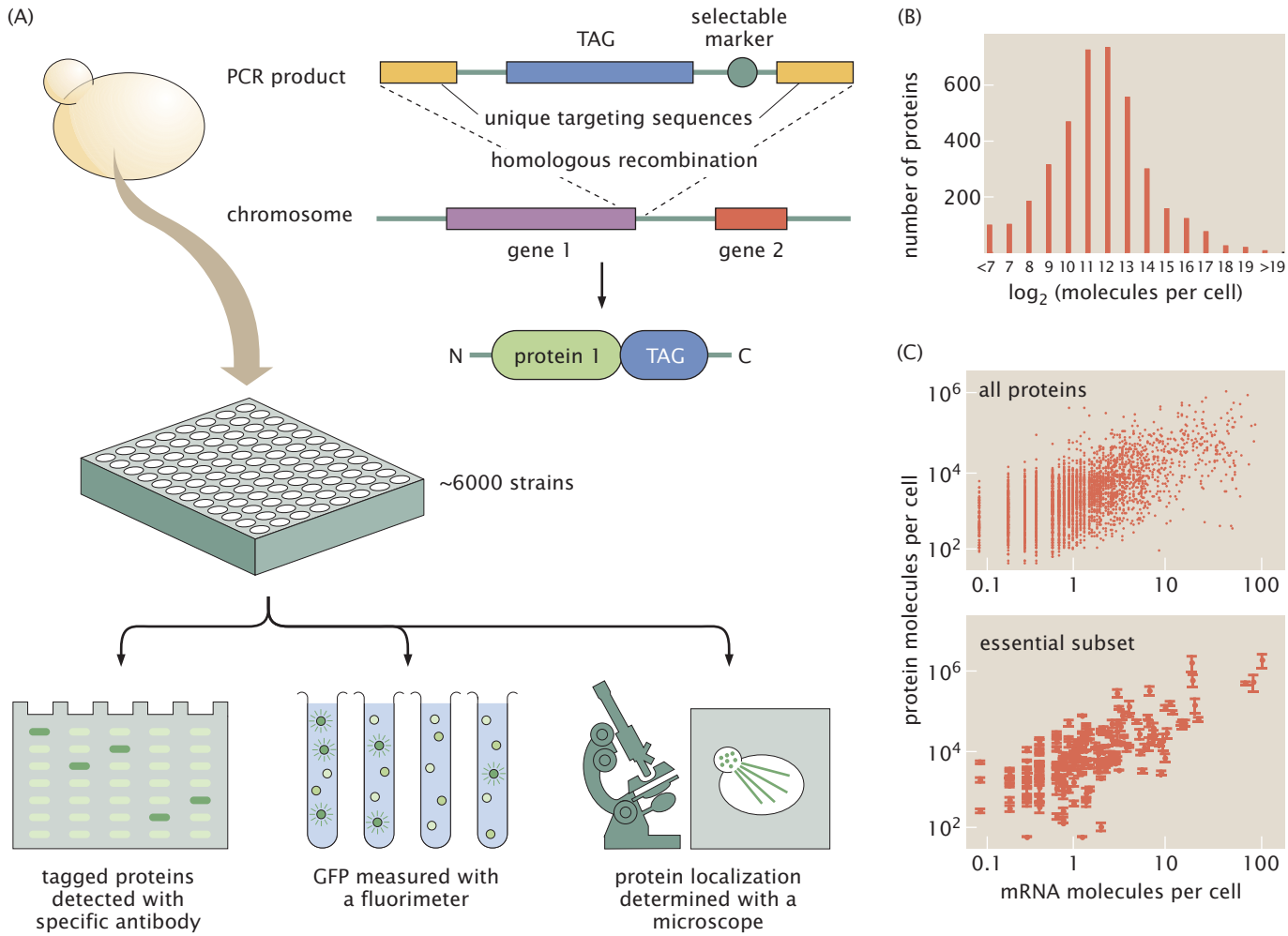


Figure 2.20: Protein copy numbers in yeast. (A) Schematic of constructs used to measure the protein census and several different methods for quantifying protein in labeled cells. The "tag" attached to each gene may encode a fluorescent protein or a site for antibody recognition, depending on the experiment. (B) Result of antibody detection of various proteins in yeast showing the number of proteins that have a given copy number. The number of copies of the protein per cell is expressed in powers of 2 as 2^N. (C) Abundances of various proteins as a function of their associated mRNA copy numbers. The bottom plot shows this result for essential soluble proteins. (A–C, adapted from S. Ghaemmaghami et al. *Nature* 425:737, 2003.)

cells easily pass through capillaries with less than half their diameter as shown in Figure 2.21, implying that their shape is altered significantly as part of their normal life cycle. While in capillaries (either artificial or *in vivo*), the red blood cell is severely deformed to pass through the narrow passage. In their role as the transport vessels for oxygen-rich hemoglobin, these cells will serve as an inspiration for our discussion of the statistical mechanics of cooperative binding. Red blood cells are a target of one of the most common infectious diseases suffered by humans, malaria, caused by the invasion of a protozoan. Malaria-infected red blood cells are much stiffer than normal cells and cannot deform to enter small capillaries. Consequently, people suffering from malaria experience severe pain and damage to tissues because of the inability of their red blood cells to enter those tissues and deliver oxygen.

One of the most commonly studied eukaryotic cells from multicellular organisms is the fibroblast, which is shown schematically in Figure 2.16(E) and in an AFM image in Figure 2.22. These cells will serve as a centerpiece for much of what we will have to say about

"typical" eukaryotic cells in the remainder of the book. Fibroblasts are associated with animal connective tissue and are notable for secreting the macromolecules of the extracellular matrix.

Cells in multicellular organisms can be even more exotic. For example, nerve cells (Figure 2.16F) and rod cells (Figure 2.16G) reveal a great deal more complexity than the examples highlighted above. In these cases, the cell shape is intimately related to its function. In the case of nerve cells, their sinewy appearance is tied to the fact that the various branches (also called "processes") known as dendrites and axons convey electrical signals that permit communication between distant parts of an animal's nervous system. Despite having nuclei with typical eukaryotic dimensions, the cells themselves can extend processes with characteristic lengths of up to tens of centimeters. The structural complexity of rod cells is tied to their primary function of light detection in the retina of the eye. These cells are highly specialized to perform transduction of light energy into chemical energy that can be used to communicate with other cells in the body and, in particular, with brain cells that permit us to be conscious of perceiving images. Rod cells accomplish this task using large stacks of membranes which are the antennas participating in light detection. Figure 2.16 only scratches the surface of the range of cellular size and shape, but at least conveys an impression of cell sizes relative to our standard ruler.

2.2.2 The Cellular Interior: Organelles

As we descend from the scale of the cell itself, a host of new structures known as organelles come into view. The presence of these membrane-bound organelles is one of the defining characteristics that distinguishes eukaryotes from bacteria and archaea. Figure 2.23 shows a schematic of a eukaryotic cell and an associated electron microscopy image revealing some of the key organelles. These organelles serve as the specialized apparatus of cell function in capacities ranging from genome management (the nucleus) to energy generation (mitochondria and chloroplasts) to protein synthesis and modification (endoplasmic reticulum and Golgi apparatus) and beyond. The compartments that are bounded by organellar membranes can have completely different protein and ion compositions. In addition, the membranes of each of these different membrane systems are characterized by distinct lipid and protein compositions.

A characteristic feature of many organelles is that they are compartmentalized structures that are separated from the rest of the cell by membranes. The nucleus is one of the most familiar examples since it is often easily visible using standard light microscopy. If we use

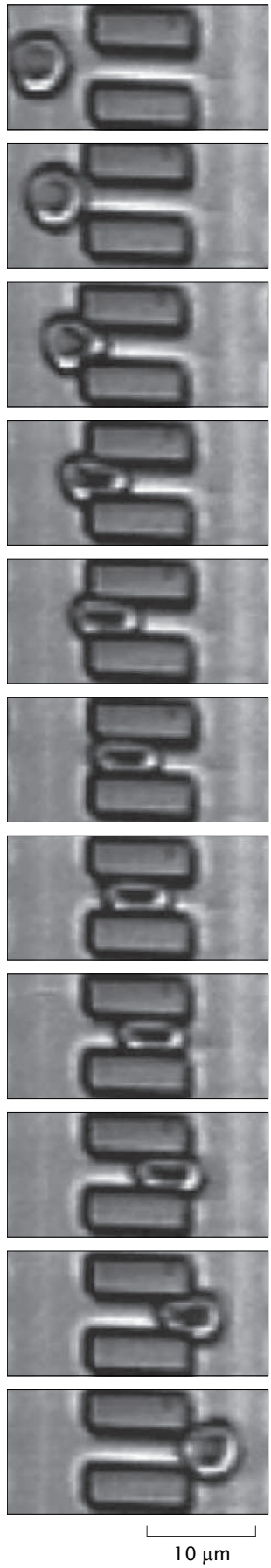


Figure 2.21: Deformability of red blood cells. To measure the deformability of human red blood cells, an array of blocks was fabricated in silicon, each block was $4\,\mu m \times 4\,\mu m \times 12\,\mu m$. The blocks were spaced by $4\,\mu m$ in one direction and $13\,\mu m$ in the other. A glass coverslip covered the top of this array of blocks. A dilute suspension of red blood cells in a saline buffer was introduced to the system. A slight pressure applied at one end of the array of blocks provided bulk liquid flow, from left to right in the figure. This liquid flow carried the red blood cells through the narrow passages. Video microscopy captured the results. The figure shows consecutive video fields with the total elapsed time just over one-third of a second. (Courtesy of J. Brody.)

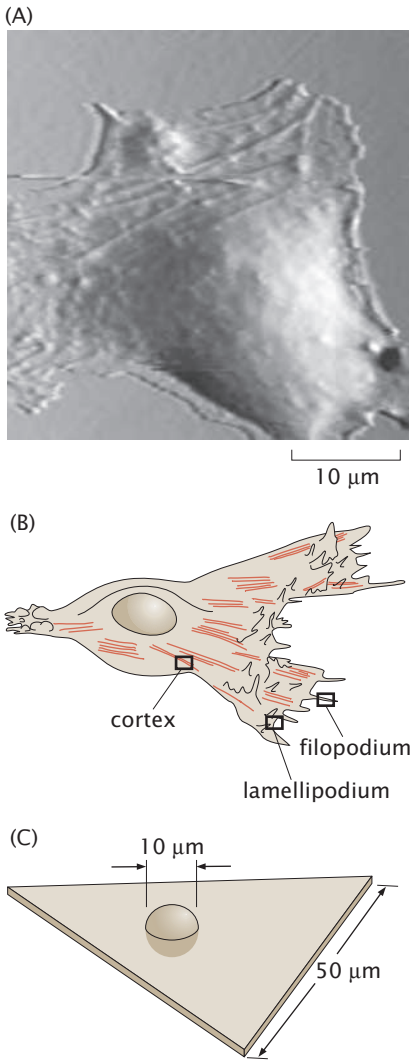


Figure 2.22: Structure of a fibroblast. (A) Atomic-force microscopy image of a fibroblast, (B) cartoon of the external morphology of a fibroblast, and (C) characteristic dimensions of a "typical" fibroblast. (A, adapted from M. Radmacher, *Meth. Cell Biol.* 83:347, 2007.)

the fibroblast as an example, then the cell itself has dimensions of roughly 50 μm, while the nucleus has a characteristic linear dimension of roughly 10 μm, as shown schematically in Figure 2.22. From a functional perspective, the nucleus is much more complex than simply serving as a storehouse for the genetic material. Chromosomes are organized within the nucleus, forming specific domains as will be discussed in more detail in Chapter 8. Transcription as well as several kinds of RNA processing occur in the nucleus. There is a flux of molecules such as transcription factors moving in and completed RNA molecules moving out through elaborate gateways in the nuclear membrane known as nuclear pores. Portions of the genome involved in synthesis of ribosomal RNA are clustered together, forming striking spots that can be seen in the light microscope and are called nucleoli.

Moving outward from the nucleus, the next membranous organelle we encounter is often the endoplasmic reticulum. Indeed, the membrane of the endoplasmic reticulum is contiguous with the membrane of the nuclear envelope. In some cells such as the pancreatic cell shown in Figure 2.24, the endoplasmic reticulum takes up the bulk of the cell interior. This elaborate organelle is the site of lipid synthesis and also the site of synthesis of proteins that are destined to be secreted or incorporated into membranes. It is clear from images such as those in Figures 2.24 and 2.25 that the endoplasmic reticulum can assume different geometries in different cell types and under different conditions. How much total membrane area is taken up by the endoplasmic reticulum? How strongly does the specific membrane morphology affect the total size of the organelle?

ESTIMATE

Estimate: Membrane Area of the Endoplasmic Reticulum

One of the most compelling structural features of the endoplasmic reticulum is its enormous surface area. To estimate the area associated with the endoplasmic reticulum, we take our cue from Figure 2.24, which suggests that we think of the endoplasmic reticulum as a series of concentric spheres centered about the nucleus. We follow Fawcett (1966), who characterizes the endoplasmic reticulum as forming "lamellar systems of flat cavities, rather uniformly spaced and parallel to one another" as shown in Figure 2.24.

An estimate can be made by adding up the areas from each of the concentric spheres making up our model endoplasmic reticulum. This can be done by simply noticing that the volume enclosed by the endoplasmic reticulum can be written as

$$V_{ER} = \sum_i A_i d, \tag{2.17}$$

where A_i is the area of the ith concentric sphere and d is the distance between adjacent cisternae. Since two membranes bound each cisterna, the total area of the endoplasmic reticulum membrane is $A_{ER} = 2 \times \sum_i A_i$. In our model, the total volume of the endoplasmic reticulum can be written as the difference between the volume taken up by the outermost sphere and the volume of the innermost concentric sphere (which is the same as the volume of the nucleus). This results in

$$V_{ER} = \frac{4\pi}{3} R_{out}^3 - \frac{4\pi}{3} R_{nucleus}^3. \tag{2.18}$$

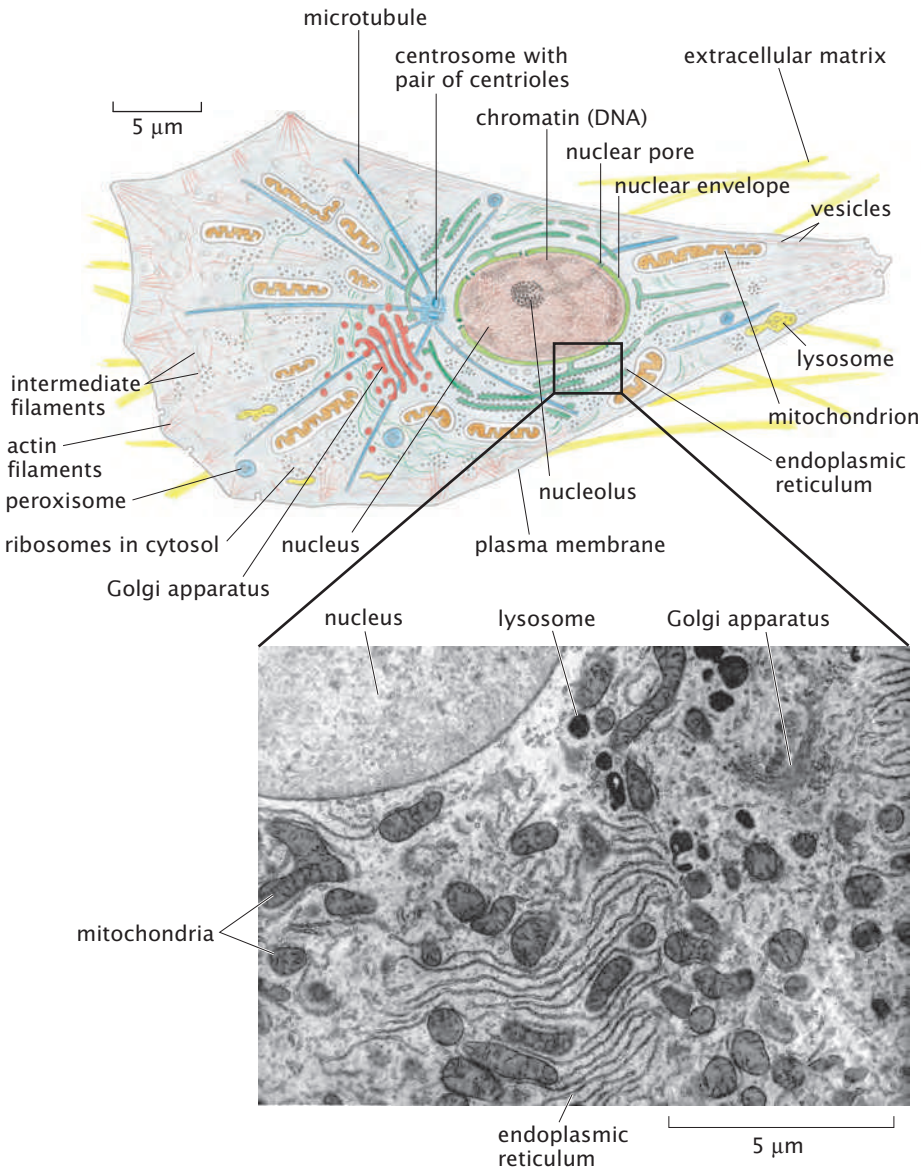


Figure 2.23: Eukaryotic cell and its organelles. The schematic shows a eukaryotic cell and a variety of membrane-bound organelles. A thin-section electron microscopy image shows a portion of a rat liver cell approximately equivalent to the boxed area on the schematic. A portion of the nucleus can be seen in the upper left-hand corner. The most prominent organelles visible in the image are mitochondria, lysosomes, the rough endoplasmic reticulum, and the Golgi apparatus. (Eukaryotic cell from Alberts et al., Molecular Biology of the Cell, 5th ed., New York, Garland Science, 2008; electron micrograph from D. W. Fawcett, The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders, 1966.)

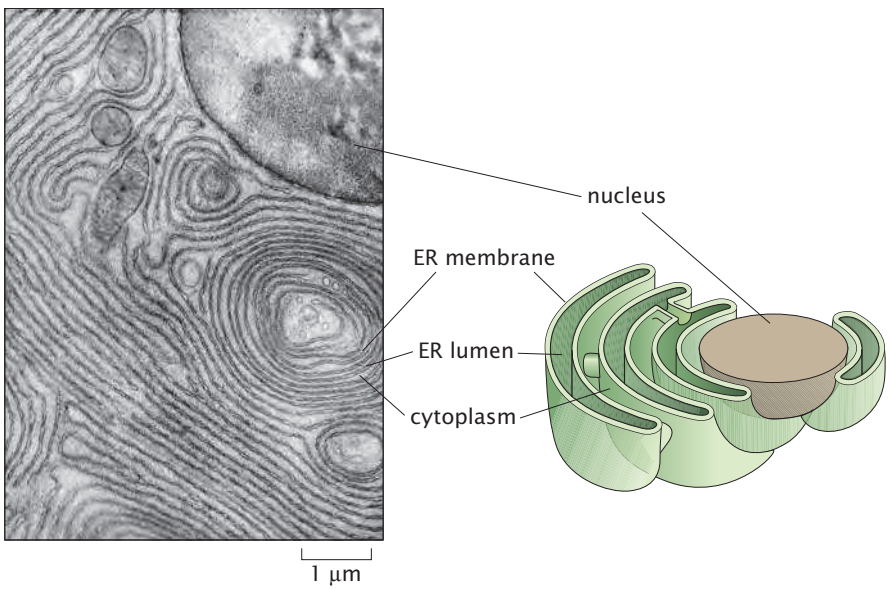


Figure 2.24: Electron micrograph and associated schematic of the endoplasmic reticulum (ER). The left-hand panel shows a thin-section electron micrograph of an acinar cell from the pancreas of a bat. The nucleus is visible at the upper right and the dense and elaborate endoplasmic reticulum structure is strikingly evident. The right-hand panel shows a schematic diagram of a model for the three-dimensional structure of the endoplasmic reticulum in this cell. Notice that the size of the lumen in the endoplasmic reticulum in the schematic is exaggerated for ease of interpretation. (Electron micrograph from D. W. Fawcett, The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders, 1966.)

Combining the two ways of computing the volume of the endoplasmic reticulum, Equations 2.17 and 2.18, we arrive at an expression for the endoplasmic reticulum area,

$$A_{ER} = \frac{8\pi}{3d}\left(R_{out}^3 - R_{nucleus}^3\right). \tag{2.19}$$

Using the values $R_{nucleus} = 5\,\mu m$, $R_{out} = 10\,\mu m$, and $d = 0.05\,\mu m$, we get, at an estimate, $A_{ER} = 15 \times 10^4\,\mu m^2$. This result should be contrasted with a crude estimate for the outer surface area of a fibroblast, which can be obtained by using the dimensions in Figure 2.22(C) and which yields an area of $10^4\,\mu m^2$ for the cell membrane itself. To estimate the area of the endoplasmic reticulum when it is in reticular form, we describe its structure as interpenetrating cylinders of diameter $d \approx 10\,nm$ separated by a distance $a \approx 60\,nm$, as shown in Figure 2.25. The completion of the estimate is left to the problems, but results in a comparable membrane area to the calculation here assuming concentric spheres.

The other major organelles found in most cells and visible in Figure 2.23 include the Golgi apparatus, mitochondria, and lysosomes. The Golgi apparatus, similar to the endoplasmic reticulum, is largely involved in processing and trafficking of membrane-bound and secreted proteins (for another view, see Figure 11.38 on p. 463). The Golgi apparatus is typically seen as a pancake-like stack of flattened compartments, each of which contains a distinct set of enzymes. As proteins are processed for secretion, for example by addition and remodeling of attached sugars, they appear to pass in an orderly fashion through each element in the Golgi stack. The mitochondria are particularly striking organelles with a smooth outer surface housing an elaborately folded system of internal membrane structures (for a more detailed view, see Figure 11.39 on p. 464). The mitochondria are the primary site of ATP synthesis for cells growing in the presence of oxygen, and their fascinating physiology and structure have been well studied. We will return to the topic of mitochondrial structure in Chapter 11 and discuss the workings of the tiny machine responsible for ATP synthesis in Chapter 16. Lysosomes serve a major role in the degradation of cellular components. In some specialized cells such as macrophages, lysosomes also serve as the compartment where bacterial invaders can be degraded. These membrane-bound organelles are filled with acids, proteases, and other degradative enzymes. Their shapes are polymorphous; resting lysosomes are simple and nearly spherical, whereas lysosomes actively involved in degradation of cellular components or of objects taken in from the outside may be much larger and complicated in shape.

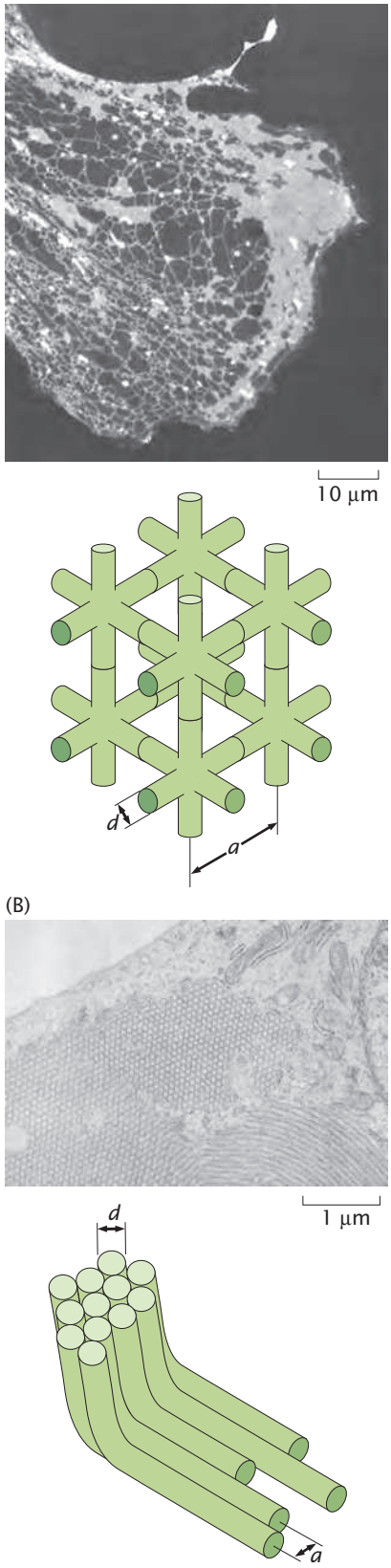


Figure 2.25: Variable morphology of the endoplasmic reticulum. (A) In most cultured cells, the endoplasmic reticulum is a combination of a web-like reticular network of tubules and larger flattened cisternae. In this image, a cultured fibroblast was stained with a fluorescent dye called DiOC6 that specifically labels endoplasmic reticulum membrane. On the bottom is a schematic of an idealized three-dimensional reticular network. (B) Some specialized cells and those treated with drugs that regulate the synthesis of lipids reorganize their endoplasmic reticulum to form tightly-packed, nearly crystalline arrays that resemble piles of pipes. (A, adapted from M. Terasaki et al., *J. Cell. Biol.* 103:1557, 1986; B, adapted from D. J. Chin et al., *Proc. Natl Acad. Sci. USA* 79:1185, 1982.)

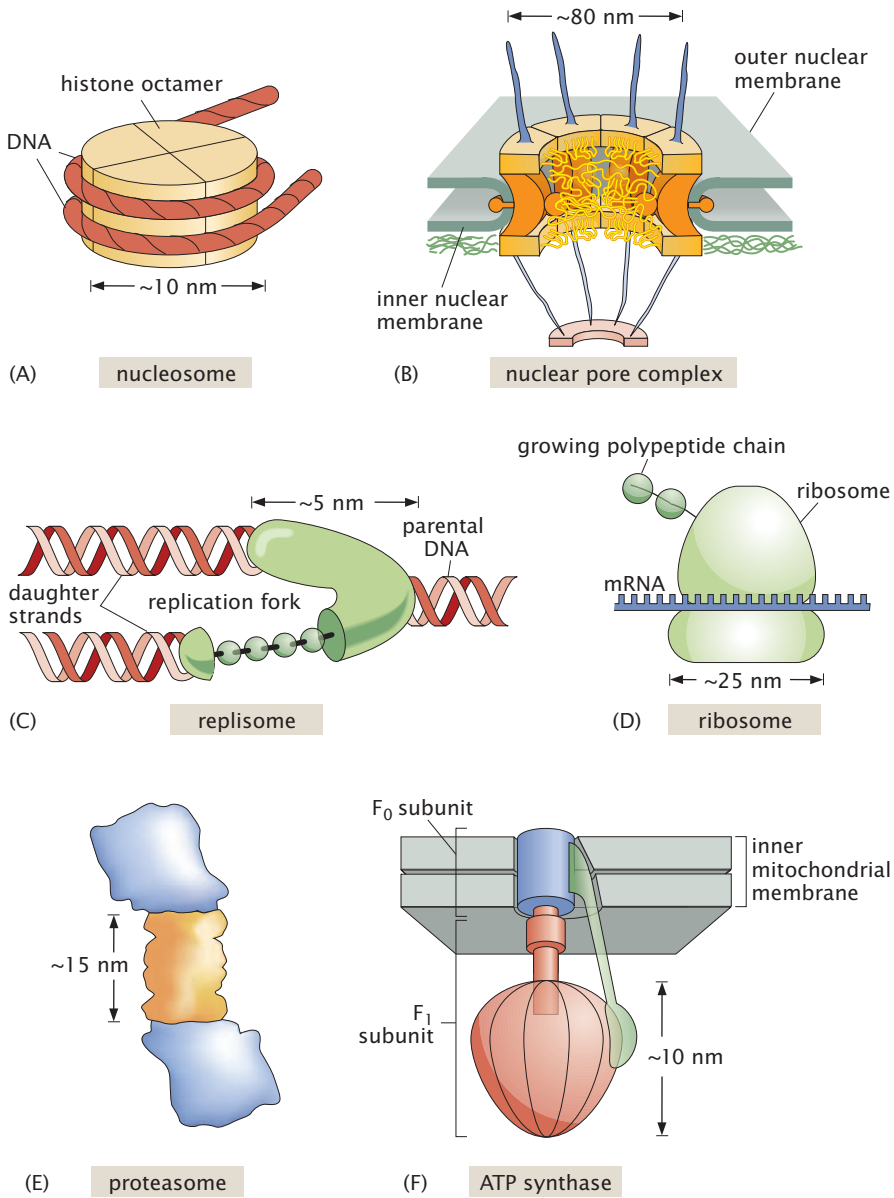


Figure 2.26: The macromolecular assemblies of the cell. (A) The nucleosome is a complex of eight protein units and DNA. (B) The nuclear pore complex mediates the transport of material in and out of the nucleus. (C) The replisome is a complex of proteins that mediate the copying of the genetic material. (D) The ribosome reads mRNA and synthesizes the corresponding polypeptide chain. (E) The proteasome degrades proteins. (F) ATP synthase is a large complex that synthesizes new ATP molecules.

These common organelles are only a few of those that can be found in eukaryotic cells. Some specialized cells have remarkable and highly specialized organelles that can be found nowhere else, such as the stacks of photoreceptive membranes found in the rod cells of the visual system as indicated schematically in Figure 2.16(G) (p. 53). The common theme is that all organelles are specialized subcompartments of the cell that perform a particular subset of cellular tasks and represent a smaller, discrete layer of organization one step down from the whole cell.

2.2.3 Macromolecular Assemblies: The Whole is Greater than the Sum of the Parts

Macromolecules Come Together to Form Assemblies

Proteins, nucleic acids, sugars, and lipids often work as a team. Indeed, as will become clear throughout the remainder of this book,

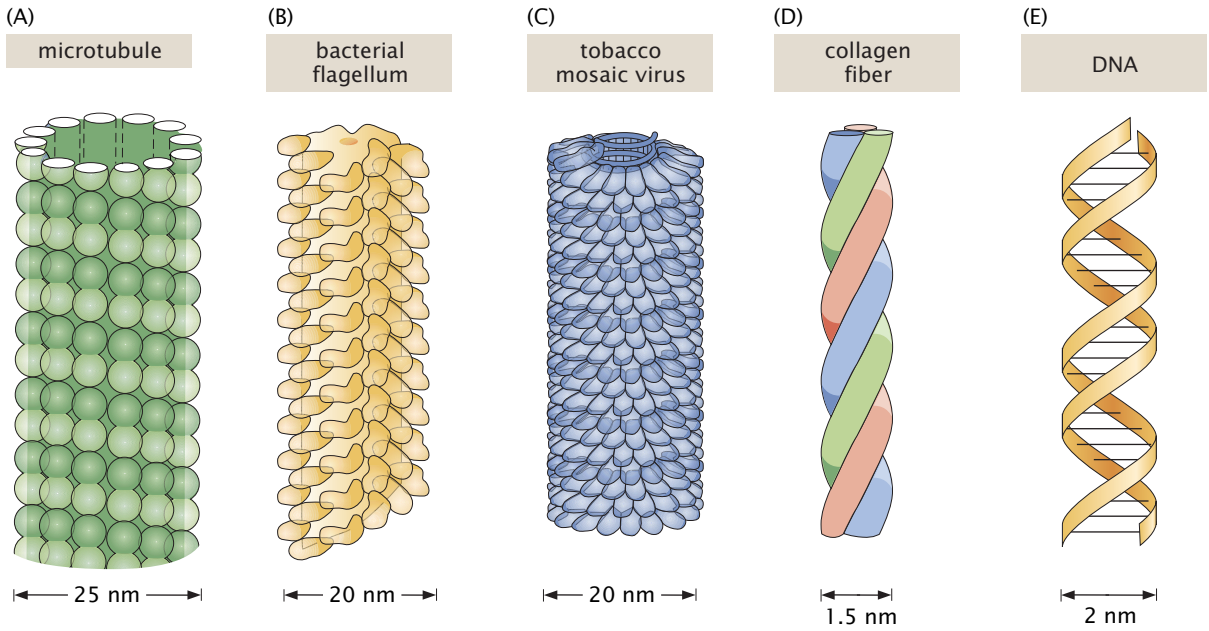


Figure 2.27: Helical motifs of molecular assemblies. Macromolecular assemblies have a variety of different helical structures, some formed from individual monomeric units (such as (A)–(C)), others resulting from coils of proteins (D), and yet others made up of paired nucleotides (E).

these macromolecules often come together to make assemblies. We think of yet another factor-of-10 magnification relative to the previous section, and with this increase of magnification we see assemblies such as those shown in cartoon form in Figure 2.26. The genetic material in the eukaryotic nucleus is organized into chromatin fibers, which themselves are built up of protein–DNA assemblies known as nucleosomes (Figure 2.26A). Traffic of molecules such as completed RNA molecules out of the nucleus is mediated by the elaborate nuclear pore complex (Figure 2.26B). The replication complex that copies DNA before cell division is similarly a collection of molecules, which has been dubbed the replisome (Figure 2.26C). When the genetic message is exported to the cytoplasm for translation into proteins, the ribosome (an assembly of proteins and nucleic acids) serves as the universal translating machine that converts the nucleic acid message from the RNA into the protein product written in the amino acid alphabet (Figure 2.26D). When proteins have been targeted for degradation, they are sent to another macromolecular assembly known as the proteasome (Figure 2.26E). The production of ATP in mitochondria is similarly mediated by a macromolecular complex known as ATP synthase (Figure 2.26F) . The key idea of this subsection is to show that there is a very important level of structure in cells that is built around complexes of individual macromolecules and with a characteristic length scale of 10 nm.

Helical Motifs Are Seen Repeatedly in Molecular Assemblies

A second class of macromolecular assemblies, characterized not by function but rather by structure, is the wide variety of helical macromolecular complexes. Several representative examples are depicted in Figure 2.27. In Figure 2.27(A), we show the geometric structure of microtubules. As will be described in more detail later, these structures are built up of individual protein units called tubulin. A second example shown in Figure 2.27(B) is the bacterial flagellum of *E. coli*. Here too, the same basic structural idea is repeated, with the helical geometry built up from individual protein units, in this case

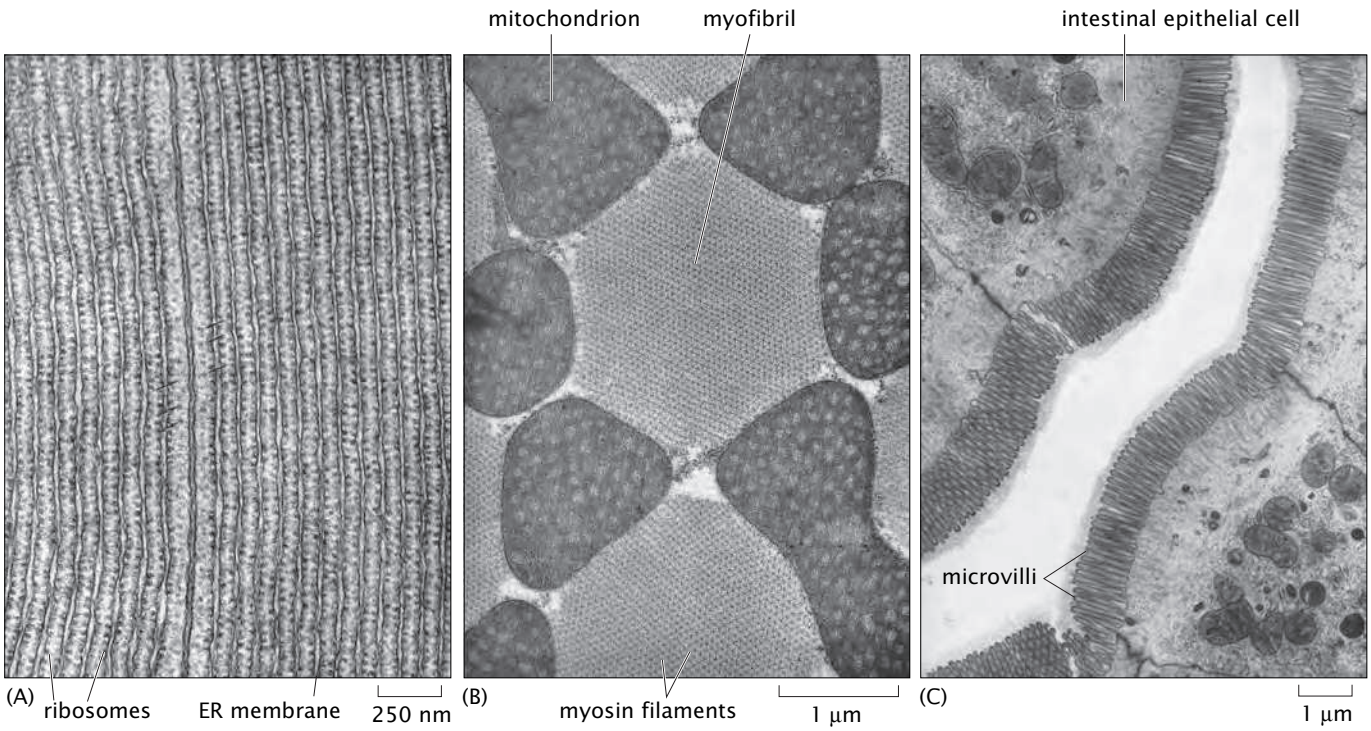


Figure 2.28: Ordered macromolecular assemblies. Collage of examples of macromolecules organized into superstructures. (A) Ribosomes on the endoplasmic reticulum ("rough ER"), (B) myofibrils in the flight muscle, and (C) microvilli at the epithelial surface. (A–C, adapted from D. W. Fawcett, The Cell, Its Organelles and inclusions: An Atlas of Fine Structure. W. B. Saunders, 1966.)

flagellin. The third example given in the figure is that of a filamentous virus, with tobacco mosaic virus (TMV) chosen as one of the most well-studied viruses.

The helical assemblies described above are characterized by individual protein units that come together to form helical filaments. An alternative and equally remarkable class of filaments are those in which α-helices (chains of amino acids forming protein subunits with a precise, helical geometry) wind around each other to form superhelices. The particular case study that will be of most interest in subsequent discussions is that of collagen, which serves as one of the key components in the extracellular matrix of connective tissues and is one of the major protein products of the fibroblast cells introduced earlier in the chapter (see Figure 2.16 on p. 53). The final example is the iconic figure of the DNA molecule itself.

Macromolecular Assemblies Are Arranged in Superstructures

Assemblies of macromolecules can interact with each other to create striking instances of cellular hardware with a size comparable to organelles themselves. Figure 2.28 depicts several examples. Figure 2.28(A) shows the way in which ribosomes are organized on the endoplasmic reticulum with a characteristic spacing that is comparable to the size of the ribosomes (≈20 nm). A second stunning example is the organization of myofibrils in muscles as shown in Figure 2.28(B). This figure shows the juxtaposition of the myofibrils and mitochondria. The myofibrils themselves are an ordered arrangement of actin filaments and myosin motors, as will be discussed in more detail in

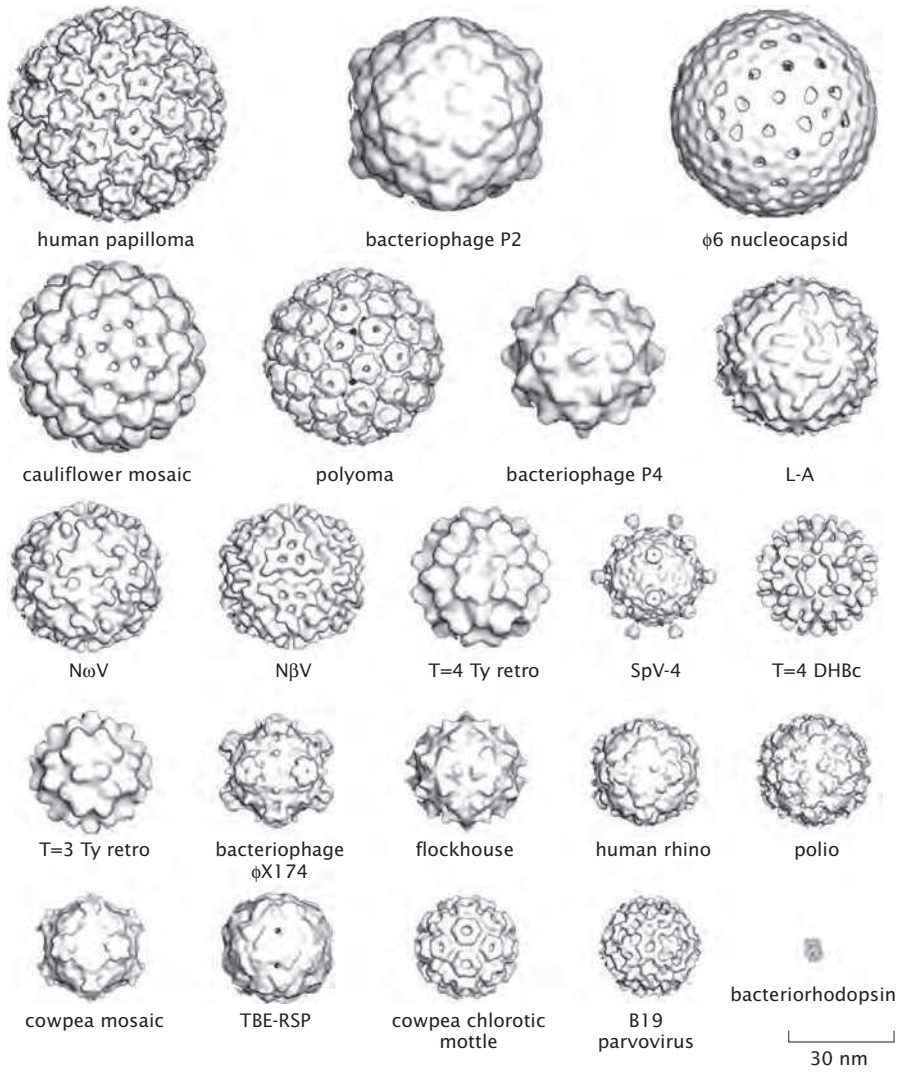


Figure 2.29: Structures of viral capsids. The regularity of the structure of viruses has enabled detailed, atomic-level analysis of their construction patterns. This gallery shows a variety of the different geometries explored by the class of nearly spherical viruses. For size comparison, a large protein, bacteriorhodopsin, is shown in the bottom right. (Adapted from T. S. Baker et al., *Microbiol. Mol. Biol. Rev.* 63:862, 1999.)

Chapter 16. The last example, shown in Figure 2.28(C), is of the protrusions of microvilli at the surface of an epithelial cell. These microvilli are the result of collections of parallel actin filaments. The list of examples of orchestration of collections of macromolecules could go on and on and should serve as a reminder of the many different levels of structural organization found in cells.

2.2.4 Viruses as Assemblies

Viruses are one of the most impressive and beautiful examples of macromolecular assembly. These assemblies are a collection of proteins and nucleic acids (though many viruses have lipid envelopes as well) that form highly ordered and symmetrical objects with characteristic sizes of tens to hundreds of nanometers. The architecture of viruses is usually a protein shell where the so-called capsid is made up of a repetitive packing of the same protein subunits over and over to form an icosahedron. Within the capsid, the virus packs its genetic material, which can be either single-stranded or double-stranded DNA or RNA depending upon the type of virus. Figure 2.29 is a gallery of the capsids of a number of different viruses. Different viruses have different elaborations on this basic structure and can include lipid coats,

surface receptors, and internal molecular machines such as polymerases and proteases. One of the most amazing features of viruses is that by hijacking the machinery of the host cell, the viral genome commands the construction of its own inventory of parts within the host cell and then, in the crowded environment of that host, assembles into infectious agents prepared to repeat the life cycle elsewhere.

Human immunodeficiency virus (HIV) is one of the viruses that has garnered the most attention in recent years. Figure 2.30 shows cryo-electron microscopy images of mature HIV virions and gives a sense of both their overall size (roughly 130 nm) and their internal structure. In particular, note the presence of an internal capsid shaped like an ice-cream cone. This internal structure houses the roughly 10 kilobase (kb) RNA viral genome. As with our analysis of the inventory of a cell considered earlier in the chapter, part of developing a "feeling for the organism" is to get a sense of the types and numbers of the different molecules that make up that organism. In the case of HIV, these numbers are interesting for many reasons, including that they say something about the "investment" that the infected cell has to make in order to construct new virions.

For our census of an HIV virion, we need to examine the assembly of the virus. In particular, one of the key products of its roughly 10 kb genome is a polyprotein, a single polypeptide chain containing what will ultimately be distinct proteins making up the capsid, matrix, and nucleocapsid, known as Gag and shown schematically in Figure 2.31. The formation of the *immature* virus occurs through the association of

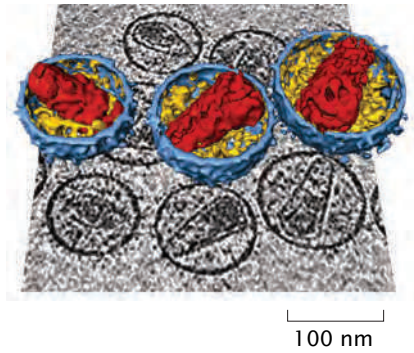


Figure 2.30: Structure of HIV viruses. The planar image shows a single frame from an electron microscopy tilt series. The three-dimensional images show reconstructions of the mature viruses featuring the ice-cream-cone-shaped capsid on the interior. (Adapted from J. A. G. Briggs et al., *Structure* 14:15, 2006.)

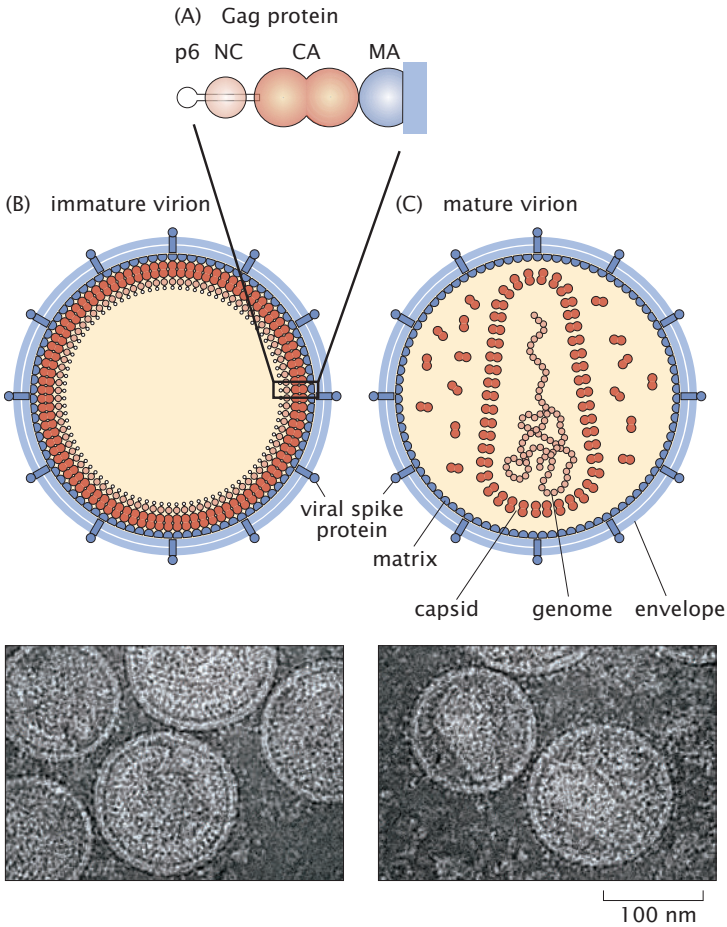


Figure 2.31: HIV architecture.
(A) Schematic of the Gag polyprotein, a 41,000 Da architectural building block.
(B) Immature virions showing the lipid bilayer coat and the uncut Gag shell on the interior.
(C) Mature virions in which the Gag protein has been cut by proteases and the separate components have assumed their architectural roles in the virus. The associated electron microscopy images show actual data for each of the cartoons. (Adapted from J. A. G. Briggs et al., *Nat. Struct. Mol. Biol.* 11:672, 2004.)

the N-terminal ends of these Gag proteins with the lipid bilayer of the host cell, with the C-termini pointing radially inward like the spokes of a three-dimensional wheel. (N-terminus and C-terminus refer to the two structurally distinct ends of the polypeptide chain. During protein synthesis, translation starts at the N-terminus and finishes at the C-terminus.) As more of these proteins associate on the cell surface, the nascent virus begins to form a bud on the cell surface, ultimately resulting in spherical structures like those shown in Figure 2.31. During the process of viral maturation, a viral protease (an enzyme that cuts proteins) clips the Gag protein into its component pieces known as matrix (MA), capsid (CA), nucleocapsid (NC), and p6. The matrix forms a shell of proteins just inside of the lipid bilayer coat. The capsid proteins form the ice-cream-cone-shaped object that houses the genetic material and the nucleocapsid protein is complexed with the viral RNA.

ESTIMATE

Estimate: Sizing Up HIV Unlike many of their more ordered viral counterparts, HIV virions have the intriguing feature that the structure from one to the next is not exactly the same. Indeed, they come in both different shapes and different sizes. As a result, our attempt to "size up" HIV will be built around some representative numbers for these viruses, but the reader is cautioned to think of a statistical distribution of sizes and shapes. As shown in the cryo-electron microscopy picture of Figure 2.30, the diameter of the virion is between 120 nm and 150 nm and we take a "canonical" diameter of 130 nm.

We begin with the immature virion. To find the number of Gag proteins within a given virion, we resort to simple geometrical reasoning. Since the radius of the overall virion is roughly 65 nm, and the outer 5 nm of that radius is associated with the lipid bilayer, we imagine a sphere of radius 60 nm that is decorated on the inside with the inward facing spokes of the Gag proteins. If we think of each such Gag protein as a cylinder of radius 2 nm, this means they take up an area $A_{\text{Gag}} \approx 4\pi \text{nm}^2$. Using this, we can find the number of such Gag proteins as

$$N_{\text{Gag}} = \frac{\text{surface area of virion}}{\text{area per Gag protein}} \approx \frac{4\pi(60\,\text{nm})^2}{4\pi\,\text{nm}^2} \approx 3500. \quad (2.20)$$

The total mass of these Gag proteins is roughly

$$M_{\text{Gag}} \approx 3500 \times 40{,}000\ \text{Da} \approx 150\ \text{MDa}, \quad (2.21)$$

where we have used the fact that the mass of each Gag polyprotein is roughly 40 kDa. This estimate for the number of Gag proteins is of precisely the same magnitude as those that have emerged from cryo-electron microscopy observations (see Briggs et al., 2004).

The number of lipids associated with the HIV envelope can be estimated similarly as

$$N_{\text{lipids}} \approx \frac{2 \times 4\pi(65\,\text{nm})^2}{1/2\,\text{nm}^2} \approx 200{,}000\ \text{lipids}, \quad (2.22)$$

where the factor of 2 accounts for the fact that the lipids form a bilayer, and we have used a typical area per lipid of $0.5\,\text{nm}^2$. The lipid census of HIV has been taken using mass spectrometry, which permits the measurement of each of the

different types of lipids forming the viral envelope (see Brügger et al., 2006). Interestingly, the diversity of lipids in the HIV envelope is enormous, with the lipid composition of the viral envelope distinct from that of the host cell membrane. The measured total number of different lipids is roughly 300,000. Further analysis of the parts list of HIV is left to the problems at the end of the chapter.

Ultimately, viruses are one of the most interesting cases of the power of macromolecular assembly. These intriguing machines occupy a fuzzy zone at the interface between the living and the nonliving.

2.2.5 The Molecular Architecture of Cells: From Protein Data Bank (PDB) Files to Ribbon Diagrams

If we continue with another factor of 10 in our powers-of-10 descent, we find the individual macromolecules of the cell. In particular, this increase in spatial resolution reveals four broad categories of macromolecules: lipids, carbohydrates, nucleic acids, and proteins. As was shown in Chapter 1 (p. 3), these four classes of molecules make up the stuff of life and have central status in building cells both architecturally and functionally. Though often these molecules are highly anisotropic (for example, a DNA molecule is usually many orders of magnitude longer than it is wide), their characteristic scale is between 1 and 10 nm. For example, as shown earlier in the chapter, a "typical" protein has a size of several nanometers. Lipids are more anisotropic, with lengths of 2–3 nm and cross-sectional areas of roughly 0.5 nm^2.

The goal of this section is to provide several different views of the molecules of life and how they fit into the structural hierarchy described throughout the chapter.

Macromolecular Structure Is Characterized Fundamentally by Atomic Coordinates

The conjunction of X-ray crystallography, nuclear magnetic resonance, and cryo-electron microscopy has revealed the atomic-level structures of a dazzling array of macromolecules of central importance to the function of cells. The list of such structures includes molecular motors, ion channels, DNA-binding proteins, viral capsid proteins, and various nucleic acid structures. The determination of new structures is literally a daily experience. Indeed, as will be asked of the reader in the problems at the end of the chapter, a visit to websites such as the Protein Data Bank or VIPER reveals just how many molecular and macromolecular structures are now known.

Though the word *structure* can mean different things to different people (indeed, that is one of the primary messages of this chapter and Chapter 8), most self-identified structural biologists would claim that the determination of structure ultimately refers to a list of atomic coordinates for the various atoms making up the structure of interest. As an example, Figure 1.3 (p. 6) introduced detailed atomic portraits of nucleic acids, proteins, lipids, and sugars. In such descriptions, the structural characterization of the system amounts to a set of coordinates

$$\mathbf{r}_i = x_i\mathbf{i} + y_i\mathbf{j} + z_i\mathbf{k}, \tag{2.23}$$

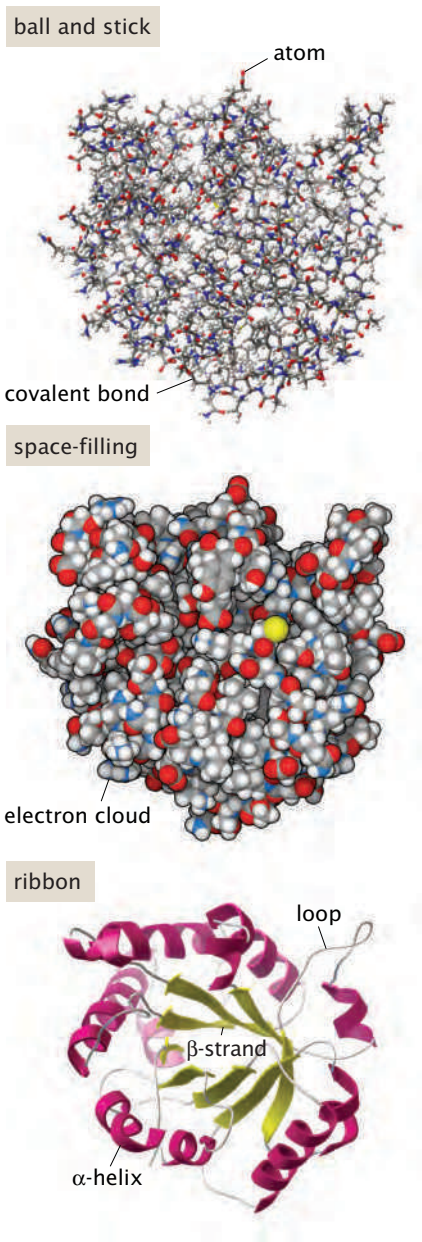


Figure 2.32: Three representations of triose phosphate isomerase. This enzyme is one of the enzymes in the glycolysis pathway (PDB 3tim). (Courtesy of D. Goodsell.)

where, having chosen some origin of coordinates, the coordinates of the ith atom in the structure are given by (x_i, y_i, z_i). That is, we have some origin of Cartesian coordinates and every atomic position is an address on this three-dimensional grid.

Because the macromolecules of the cell are subject to incessant jiggling due to collisions with each other and the surrounding water, a static picture of structure is incomplete. The structural snapshots embodied in atomic coordinates for a given structure miss the fact that each and every atom is engaged in a constant thermal dance. Hence, the coordinates of Equation 2.23 are really of the form

$$\mathbf{r}_i(t) = x_i(t)\mathbf{i} + y_i(t)\mathbf{j} + z_i(t)\mathbf{k}, \tag{2.24}$$

where the t reminds us that the coordinates depend upon time and what is measured in experiments might be best represented as $\langle \mathbf{r}_i(t) \rangle_{\text{time}}$, where the brackets $\langle\ \rangle_{\text{time}}$ signify an average over time.

An example of an atomic-level representation of one of the key proteins of the glycolysis pathway is shown in Figure 2.32. We choose this example because glycolysis will arise repeatedly throughout the book (see p. 191 for example) as a canonical metabolic pathway. The figure also shows several alternative schemes for capturing these structures such as using ribbon diagrams which highlight the ways in which the different amino acids come together to form elements of secondary structure such as α-helices and β-sheets.

Chemical Groups Allow Us to Classify Parts of the Structure of Macromolecules

When thinking about the structures of the macromolecules of the cell, one of the most important ways to give those structures functional meaning (as opposed to just a collection of coordinates) is through reference to the chemical groups that make them up. For example, the structure of the protein shown in Figures 1.3 and 2.32 is not just an arbitrary arrangement of carbons, nitrogens, oxygens, and hydrogens. Rather, this structure reflects the fact that the protein is made up of a linear sequence of amino acids that each have their own distinct identity as shown in Figure 2.33. The physical and chemical properties of these amino acids dictate the folded shape of the protein as well as how it functions.

Amino acids are but one example of a broader class of sub-nanometer-scale structural building blocks known as "chemical groups." These chemical groups occur with great frequency in different macromolecules and, like the amino acids, each has its own unique chemical identity. Figure 2.34 shows a variety of chemical groups that are of interest in biochemistry and molecular biology. These are all biologically important chemical functional groups that can be attached to a carbon atom as shown in Figure 2.34 and are all found in protein structures. The methyl and phenyl groups contain only carbon and hydrogen and are hence hydrophobic (unable to form hydrogen bonds with water). To the right of these are shown two chemically similar groups, alcohol and thiol, consisting of oxygen or sulfur plus a single hydrogen. The key feature of these two groups is that they are highly reactive and can participate in chemical reactions forming new covalent bonds. Amino acids containing these functional groups (serine, threonine, tyrosine, and cysteine) are frequently important active site residues in enzyme-catalyzed reactions. The next row starts with a nitrogen-containing amino group, which is

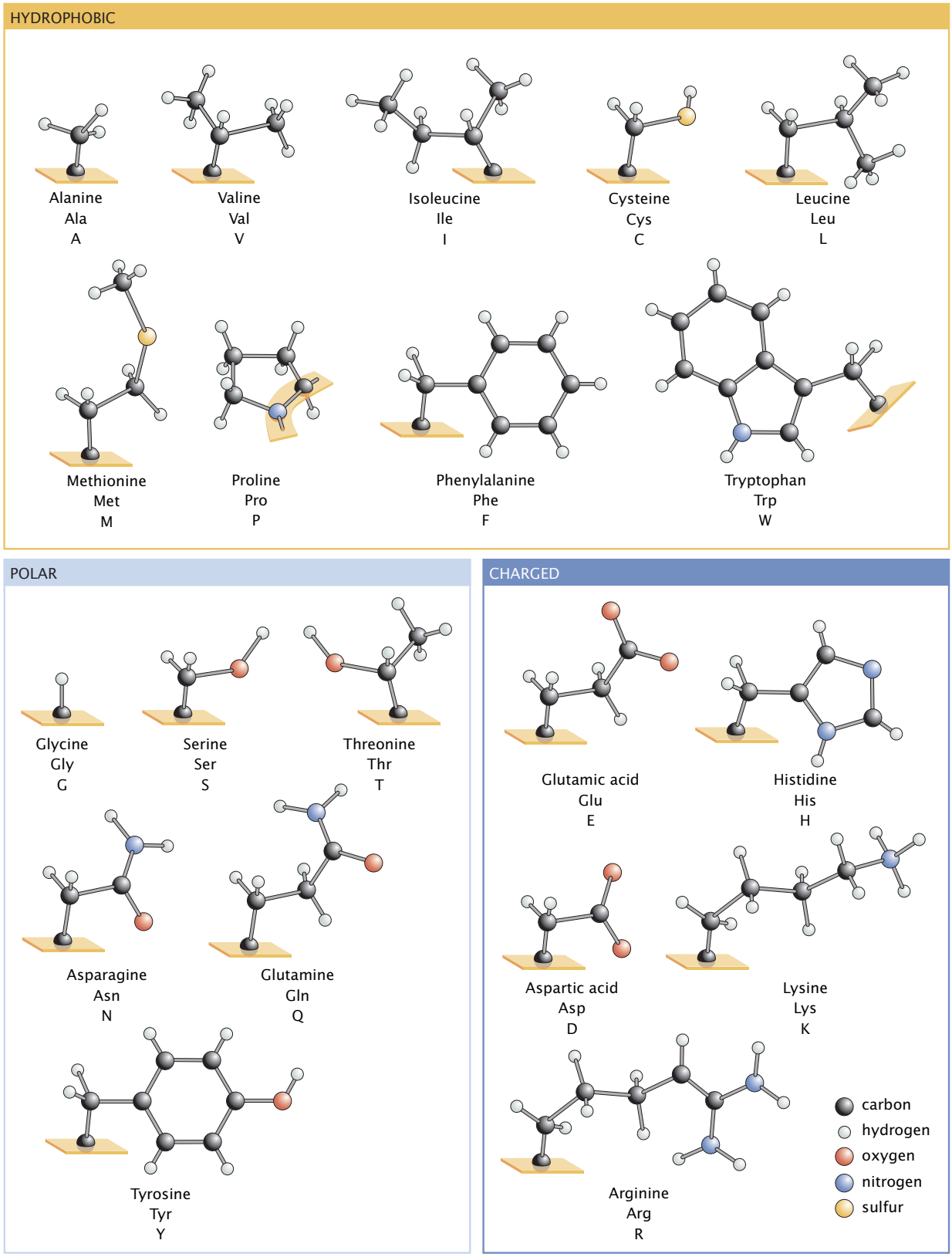


Figure 2.33: Amino acid side chains. The amino acids are represented here in ball-and-stick form, where a black ball indicates a carbon atom, a small white ball a hydrogen atom, a red ball an oxygen atom, a blue ball a nitrogen atom, and a yellow ball a sulfur atom. Only the side chains are shown. The peptide backbone of the protein to which these side chains are attached is indicated by an orange tile. The amino acids are subdivided based upon their physical properties. The group shown at the top are hydrophobic and tend to be found on the interior of proteins. Those at the bottom are able to form hydrogen bonds with water and are often found on protein surfaces.

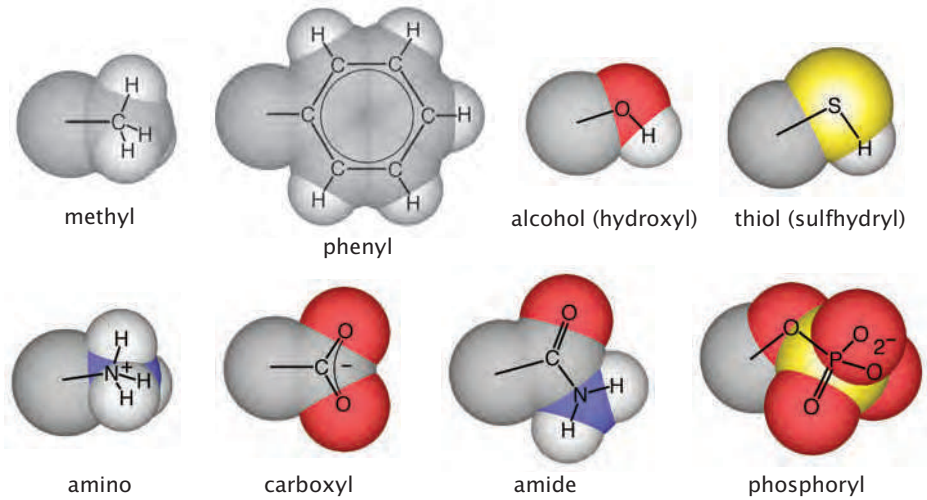


Figure 2.34: Chemical groups. These are some of the most common groups found in organic molecules such as proteins. (Courtesy of D. Goodsell.)

usually positively charged at neutral pH, and a carboxylic acid, which is usually negatively charged. All amino acids in monomeric form have both of these groups. In a protein polymer, there is a free amino group at the N-terminus of the protein and a free carboxylic acid group at the C-terminus of the protein. Several amino acids also contain these groups as part of their side chains and the charge-based interactions are frequently responsible for chemical specificity in molecular recognition as well as some kinds of catalysis. An amide group is shown next. This group is not generally charged but is able to participate in a variety of hydrogen bonds. The last group shown is a phosphoryl group, which is not part of any amino acid that is incorporated by the ribosome in a polypeptide chain during translation. However, phosphoryl groups are frequently added to proteins as a post-translational modification and perform extremely important regulatory functions.

Nucleic acids can similarly be thought of from the point of view of chemical groups. In Figure 1.3 (p. 9), we showed the way in which individual groups can be seen as the building blocks of DNA structures such as that shown in Figure 1.3(A). In particular, we note that the backbone of the double helix is built up of sugars (represented as pentagons) and phosphates. Similarly, the nitrogenous bases that mediate the pairing between the complementary strands of the backbone are represented diagrammatically via hexagons and pentagons, with hydrogen bonds depicted as shown in the figure.

This brief description of the individual molecular units of the machines of the cell brings us to the end of our powers-of-10 descent that examined the structures of the cell. Our plan now is to zoom out from the scale of individual cells to examine the structures they form together.

2.3 Telescoping Up in Scale: Cells Don't Go It Alone

Our powers-of-10 journey has thus far shown us the way in which cells are built from structural units going down from organelles to macromolecular assemblies to individual macromolecules to chemical groups, atoms, and ions. Equally interesting hierarchies of structures are revealed as we reduce the resolution of our imaginary camera and zoom out from the scale of individual cells. What we see once we

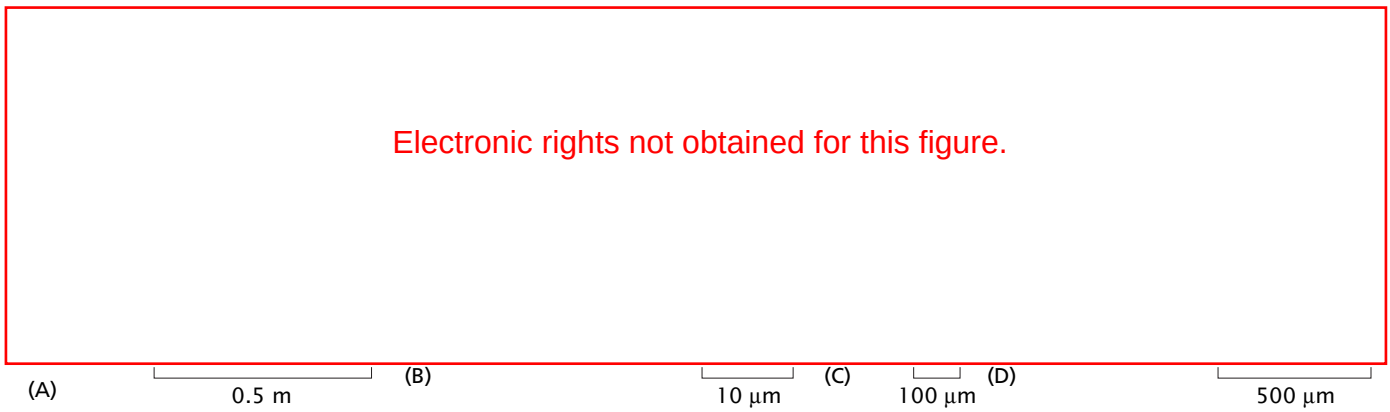


Figure 2.35: Representative examples of different communities of cells. (A) These stromatolites living in Shark Bay (Western Australia) are layered structures built by cyanobacteria. Stromatolites dominate the fossil record between 1 and 2 billion years ago, but are rare today. (B) This scanning electron micrograph showing a bacterial biofilm growing on an implanted medical device illustrates the morphological variety of bacterial species that can cooperate to form an interdependent community. (C) The social amoeba *Dictyostelium discoideum* forms fruiting bodies. The picture shows a fruiting body with spores—the tall stalk with a bulb at the top is a collection of amoebae. (D) A close-up photograph of the head of the robber fly, *Holocephala fusca*, shows the ordered structure of the compound eyes found in most insects. (A, Courtesy of Bellarmine University Department of Biology; B, Reproduced with permission of the ASM Microbe Library; C, Courtesy of Pauline Schaap; D, Courtesy of Thomas Shahan.)

begin to zoom out from the scale of single cells is the emergence of communities in which cells do not act independently.

2.3.1 Multicellularity as One of Evolution's Great Inventions

Life has been marked by several different evolutionary events that each wrought a wholesale change in the way that cells operate. One important category of such events is the acquisition of the ability of cells to communicate and cooperate with one another to form multicellular communities with common goals. This has happened many times throughout all branches of life and has culminated in extremely large organisms such as redwood trees and giraffes. In this section, we explore the ways in which cell–cell communication and cooperation have given rise to new classes of biological structures. Figure 2.35 shows a variety of different examples of cellular communities, some of which form the substance of the remainder of the chapter.

Bacteria Interact to Form Colonies such as Biofilms

The oldest known cellular communities recorded in the fossil record are essentially gigantic bacterial colonies called stromatolites. Although most stromatolites were outcompeted in their ecological niches by subsequent fancier forms of multicellular life, a few can still be found today taking essentially the same form as their 2-billion-year-old fossil brethren. Bacterial colonies of this type can still be found today in Australia, such as those shown in Figure 2.35(A). These communities have a characteristic size of about a meter and reflect collections of bacterial cells held together by an extracellular matrix secreted by these cells.

Many interesting kinds of bacterial communities consist of more than one species. Indeed, through a sophisticated system of signaling,

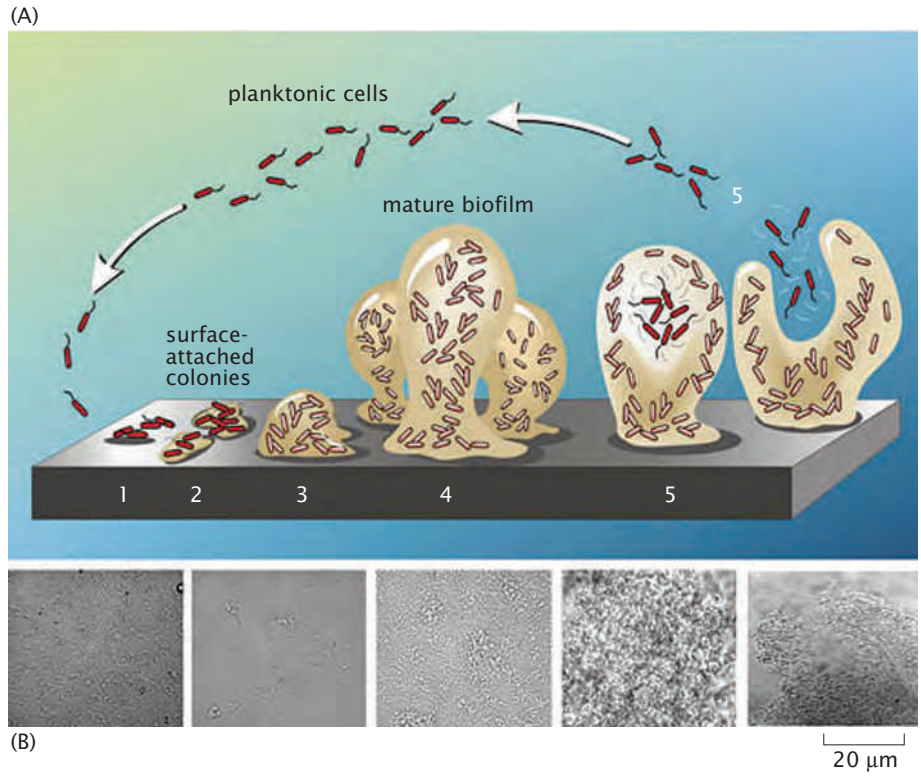


Figure 2.36: Schematic of the formation of a biofilm by bacteria. (A) The various stages in the formation of the biofilm are (1) attachment to surface, (2) secretion of extracellular polymeric substance (EPS), (3) early development, (4) maturation, and (5) shedding of cells from the biofilm. (B) The microscopy images show biofilms in various stages of the film formation process. (Adapted from P. Stoodley et al., *Annu. Rev. Microbiol.* 56:187, 2002. Original figure created by Peg Dirckx, MSU-CBE.)

detection, and organization, bacteria form colonies of all kinds ranging from biofilms to the ecosystems within animal guts. Bacterial biofilms are familiar to us all as the basis of the dentist's warning to floss our teeth every night. These communities are functionally as well as structurally interdependent. Other biofilms are noted for their destructive force when they attach to the surfaces of materials. Figure 2.35(B) shows a biofilm that grew on a silicon rubber voice prosthesis that had been implanted in a patient for about three months.

Structurally, a biofilm is formed as shown schematically in Figure 2.36. The key building blocks of such structures are a population of bacteria, a surface onto which these cells may adhere, and an aqueous environment. The formation of a biofilm results in a population of bacterial cells that are attached to a surface and enclosed in a polymeric matrix built up of molecules produced by these very same bacteria. The early stages of biofilm formation involve the adhesion of the bacteria to a surface, followed by changes in the characteristics of these bacteria such as the loss of flagella and micro-colony formation mediated by pili-twitching motility. At a larger scale, these changes at the cellular level are attended by the formation of colonies of cells and differentiation of the colonies into structures that are embedded in extracellular polysaccharides. Though there are a variety of different morphologies that are adopted by such films, roughly speaking, these biofilms are relatively porous structures (presumably to provide a conduit for import and export of nutrients and waste, respectively) that typically take on mushroom-like structures such as indicated schematically in Figure 2.36. These films have a relative proportion of something like 85% of the mass taken up by extracellular matrix, while the remaining 15% is taken up by cells themselves. A typical thickness for such films ranges from 10 to 50 μm.

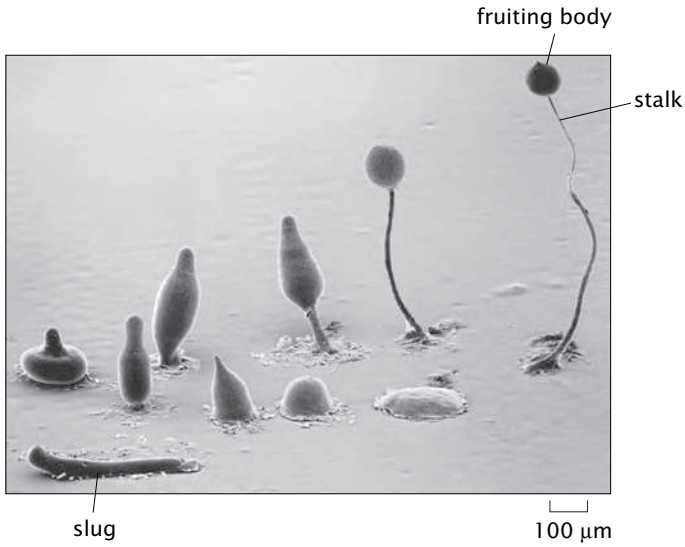


Figure 2.37: Formation of a multicellular structure during starvation. The social amoeba *Dictyostelium discoideum* responds to starvation by forming a structure made up of tens of thousands of cells in which individual cells suffer different fates. Cells near the top of the structure form spores, which are resurrected once conditions are favorable. The figure shows the developmental stages that take place on the way to making fruiting bodies, starting in the bottom right and proceeding clockwise. (Copyright, M. J. Grimson and R. L. Blanton, Biological Sciences Electron Microscopy Laboratory, Texas Tech University.)

Teaming Up in a Crisis: Lifestyle of *Dictyostelium discoideum*

Although bacteria can form communities, eukaryotes have clearly raised this to a high art. One particularly fascinating example that may give clues as to the origin of eukaryotic multicellularity is the cellular slime mold *Dictyostelium discoideum*, as shown in Figure 2.35(C). This small, soil-dwelling amoeba pursues a solitary life when times are good but seeks the comfort of its fellows during times of starvation. *Dictyostelium* is usually content to wander around as an individual with a characteristic size between 5 and 10 μm. However, when deprived of their bacterial diet, these cells undertake a radical change in lifestyle, which involves both interaction and differentiation through a series of fascinating intermediate steps as shown in Figure 2.37. A group of *Dictyostelium* amoebae in a soil sample that find themselves faced with starvation send chemical signals to one another resulting in the coalescence of thousands of separate amoebae to form a slug that looks like a small nematode worm. These cells appear to be poised on the brink between unicellular and multicellular lifestyles and can readily convert between them. Ultimately, as shown in the figure, the slug stops moving and begins to form a stalk. At the tip of the stalk is a nearly spherical bulb that contains many thousands of spores, essentially cells in a state of suspended animation. When environmental conditions are appropriate for individual amoeba to thrive, the spores undergo the process of germination, with each spore becoming a functional, individual amoeba.

Estimate: Sizing Up the Slug and the Fruiting Body The relation between the number of cells in a slug and its size is shown in Figure 2.38, where we see that these slugs can range in length between several hundred and several thousand microns with the number of cells making up the slug being between tens of thousands and several million. The next visible stage in the development is the sprouting of a stalk with a bulb at the top known as a fruiting body. The stalk is of the order of a millimeter in length, while the fruiting body itself is several hundred microns across. This fruiting body is composed of thousands of spores, that are effectively

ESTIMATE

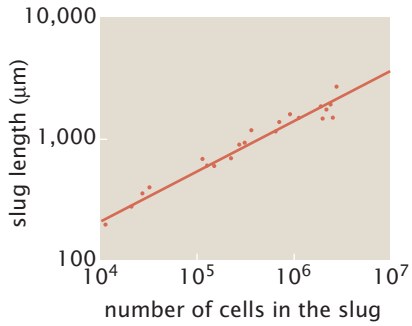


Figure 2.38: Slug size in *Dictyostelium discoideum*. The plot shows a relation between the size of the *Dictyostelium* slug and the corresponding number of cells. Points represent different individuals and the line shows the overall trend. (Adapted from Bonner, www.dictybase.org.)

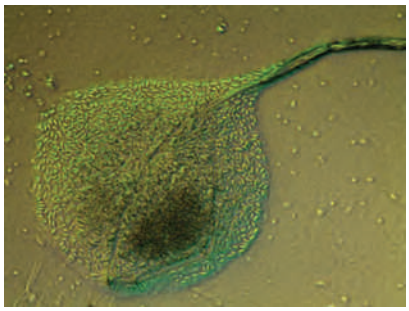


Figure 2.39: Microscopy image of a fruiting body. The fruiting body has been squished on a microscope slide, revealing both the size of the spores and their numbers.

in a state of suspended animation. An example of such a fruiting body that has been squished on a microscope slide is shown in Figure 2.39. This structure has functional consequences. In particular, those cells that become spores remain poised to respond to a better day, while the cells that formed the stalk have effectively ended their lives for the good of those that survive.

An immediate question of interest concerning the multicellular fruiting bodies shown in Figure 2.37 is how many cells conspire to make up such structures. An estimate of the number of cells in a fruiting body can be constructed by examining the nearly hemispherical collection of cells shown in Figure 2.39. Note that we treat the fruiting body as a hemisphere since, once it is squashed on the microscope slide, its shape is flattened and resembles a hemisphere more closely than a sphere. The diameter of this hemisphere is roughly $200\,\mu\text{m}$. Our rough estimate for the number of spores in the fruiting body is obtained by evaluating the ratio

$$\text{number of cells} = \frac{V_{\text{body}}}{V_{\text{cell}}}, \tag{2.25}$$

where we assume that the entirety of the fruiting body volume is made up of cells. If we assume that the cell is $4\,\mu\text{m}$ in diameter, this yields

$$\text{number of cells} = \frac{\frac{2}{3}\pi(100\,\mu\text{m})^3}{\frac{4}{3}\pi(2\,\mu\text{m})^3} \approx 6 \times 10^4 \text{ cells}. \tag{2.26}$$

Note that the size of the ball of cells in a fruiting body can vary dramatically from the $200\,\mu\text{m}$ scale shown here to several times larger, and, as a result, the estimate for the number of cells in a fruiting body can vary. Also, note that a factor-of-2 error in our estimate of the size of an individual cell will translate into a factor-of-8 error in our count of the number of cells in the slug or fully formed fruiting body.

Multicellular Organisms Have Many Distinct Communities of Cells

The three branches of life that have most notably exploited the potential of the multicellular lifestyle are animals, plants, and fungi. While bacterial and protist colonial organisms rarely form communities with characteristic dimensions of more than a few millimeters (with the exception of stromatolites), individuals in these three groups routinely grow to more than a meter in size. Their enormous size and corresponding complexity can be attributed to at least three factors: (i) production of extracellular matrix material that can provide structural support for large communities of cells, (ii) a predilection towards cellular specialization such that many copies of cells with the same genomic content can develop to perform distinct functions, and (iii) highly sophisticated mechanisms for the cells to communicate with one another within the organism. We emphasize that these traits are not unique to animals, plants, and fungi, but they are used more extensively there than elsewhere. A beautiful illustration of these principles is seen in the eye of a fly as shown in Figure 2.35(D). The eye is made up of hundreds of small units called ommatidia, each of which contains a group of eight photoreceptor cells, support cells,

and a cornea. During development of the eye, these cells signal to one another to establish their identities and relative positions to create a stereotyped structure that is repeated many times. The overarching theme of the remainder of this chapter is the exploration of how cells come together to form higher-order structures and how these structures fit into the overall hierarchy of structures formed by living organisms.

2.3.2 Cellular Structures from Tissues to Nerve Networks

Multicellular structures are as diverse as cells themselves. Often, the nature of these structures is a reflection of their underlying function. For example, the role of epithelia as barriers dictates their tightly packed, planar geometries. By way of contrast, the informational role of the network of neurons dictates an entirely different type of multicellular structure.

One Class of Multicellular Structures is the Epithelial Sheets

Epithelial sheets form part of the structural backdrop in organs ranging from the skin to the bladder. Functionally, the cells in these structures have roles such as serving as a barrier to transport of molecules, providing an interface at which molecules can be absorbed into cells, and serving as the seat of certain secretions. Several different views of these structures are shown in Figure 2.40.

The morphology of epithelial sheets is diverse in several ways. First, the morphology of the individual cells can be different (isotropic versus anisotropic). In addition, the assemblies of cells themselves have different shapes. For example, the different structures can be broadly classified into those that are a monolayer sheet (simple epithelium) and those that are a multilayer (stratified epithelium). Within these two broad classes of structures, the cells themselves have different morphologies. The cells making up a given epithelial sheet can be flat, pancake-like cells, denoted as *squamous* epithelium. If the cells making up the epithelial sheet have no preferred orientation, they are referred to as *cuboidal*, while those that are elongated perpendicular to the extracellular support matrix are known as *columnar* epithelia. Epithelial sheets have as one of their functions (as do lipid bilayers) the segregation of different media, which can have highly different ionic concentrations, pH, macromolecular concentrations, and so on.

Tissues Are Collections of Cells and Extracellular Matrix

We have seen that cells can interact to form complexes. An even more intriguing example of a multicellular structure is provided by tissues such as that shown in Figure 2.41. These connective tissues are built up from a diverse array of cells and materials they secrete. Beneath the epithelial surface, fibroblasts construct extracellular matrix. This connective tissue is built up of three main components: cells such as fibroblasts and macrophages, connective fibers, and a supporting substance made up of glycosaminoglycans (GAGs). What is especially appealing and intriguing about these tissues is the orchestration, in both space and time, leading to the positioning of cells and fibers. The fibroblasts, which were featured earlier in this chapter, serve as factories for the proteins that make up the extracellular matrix. In

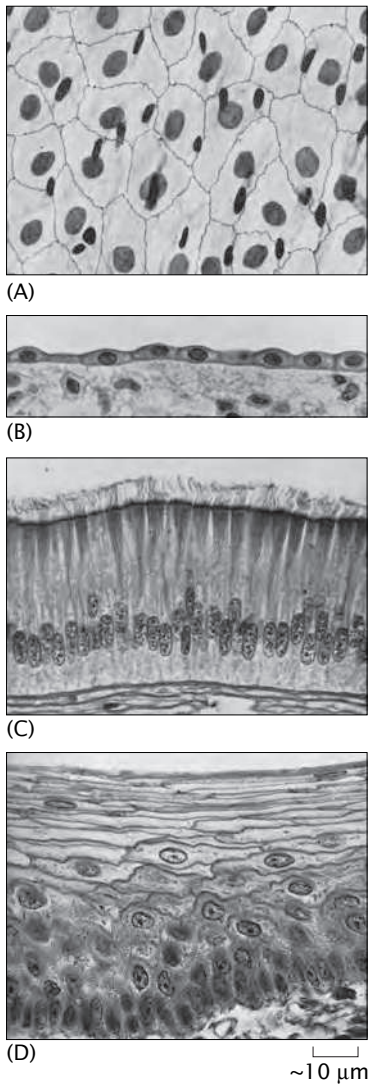


Figure 2.40: Shapes and architecture of epithelial cells. Epithelia are tissues formed by continuous sheets of cells that make tight contacts with one another. (A) Viewed from above, a simple epithelial sheet resembles a tiled mosaic. The dark ovals are the cell nuclei stained with silver. (B) Viewed from the side, simple epithelia such as this from the dog kidney may form a single layer of flattened cells above a loose, fibrous connective tissue. (C) In some specialized epithelia, the cells may extend upwards forming elongated columns and develop functional specializations at the top surface such as the beating cilia shown here. This particular ciliated columnar epithelium is from the alimentary tract of a freshwater mussel. (D) In other epithelia such as skin or the kitten's gum shown here, epithelial cells may form multiple layers. (A–D, adapted from D. W. Fawcett, The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure. W. B. Saunders, 1966.)

particular, they synthesize proteins such as collagen and elastin that, when secreted, assemble to form fibrous structures that can support mechanical loads. The medium within which these fibers (and the cells) are embedded is made up of the third key component of the connective tissue, namely, the hydrated gel of GAGs.

Figure 2.41: Tissue organization. This thin section of a villus in the human small intestine was stained with hematoxylin and eosin, a mixture of dyes that stains DNA and other negatively charged macromolecules blue or dark purple, and stains most proteins pink. In this image, a finger-shaped intestinal villus sticks up into the lumen of the intestine, covered with a single continuous monolayer of epithelial cells, whose function is to absorb nutrients. The supporting structure underneath the epithelial layer is composed of extracellular matrix secreted by fibroblasts. Numerous immune system cells also patrol this territory. (Adapted from L. G. van der Flier and H. Clevers, *Annu. Rev. Physiol.* 71:241, 2009.)

Nerve Cells Form Complex, Multicellular Complexes

A totally different example of structural organization involving multiple cells is illustrated by collections of neurons. Neurons are the specialized cells in animals that we associate with thinking and feeling. These cells allow for the transmission of information over long distances in the form of electrical signals. Neurons are constructed with many input terminals known as dendrites and a single output path known as the axon. The fascinating structural feature of these cells is that they assemble into complex networks that are densely connected in patterns where dendrites from a given cell reach out to many others, from which they take various inputs. An example of a collection of fluorescently labeled neurons is shown in Figure 2.42. Note that the branches (dendrites and axons) that reach out from the various cells have lengths far in excess of the 10 μm scale characteristic of typical eukaryotes. Indeed, axons of some neurons can have lengths of centimeters and more.

One fascinating example of neuronal contact is offered by the so-called neuromuscular junction as shown in Figure 2.43. These junctions are the point of contact between motor neurons, which convey the marching orders for a given muscle, and the muscle fiber itself. As is seen in the figure, the axon from a given motor neuron makes contacts with various muscle fibers. As will be described in more detail in Chapter 17, when an electrical signal (action potential) arrives at the contact point known as a synapse, chemical neurotransmitters are released into the space between the nerve and muscle. These neurotransmitters result in the opening of ligand-gated ion channels in the muscle, which results in a change in the electrical state of the muscle and leads to its contraction. These contrasting examples provide a small window into the diversity of cell–cell contacts.

2.3.3 Multicellular Organisms

The highest level in the structural hierarchy to be entertained here is that of individual multicellular organisms. The diversity of multicellular organisms is legendary, ranging from roses to hummingbirds, from Venus flytraps to the giant squid. What is especially remarkable about this diversity is contained in the simple statement of the cell theory: each and every one of these organisms is a collection of cells and their products. And, each and every one of these organisms is the result of a long history of evolution resulting in specialization and diversification of the various cells that make up that organism.

Cells Differentiate During Development Leading to Entire Organisms

The fruit fly *Drosophila melanogaster* has had a long and rich history as one of the key "model" organisms of biology and is a useful starting point for thinking about the size of organisms. As shown in

Figure 2.44, the mature fly has a size of roughly 3 mm that can be thought of morphologically as being built up of 14 segments: 3 segments making up the head, 3 segments making up the thorax and 8 segments making up the abdomen. We will further explore both the life cycle and the biological significance of *Drosophila* in Chapters 3 (p. 87) and 4 (p. 137), respectively, and for now content ourselves with examining the spatial scales associated with both the mature fly (Figure 2.44) and the embryos from which they are derived.

Drosophila has attained its legendary status in part because of the way it has revealed so many different ideas about genetics and embryonic development. One of the most well-studied features of the development of *Drosophila* is the way in which it lays down its anterior–posterior architecture during early development. This body plan is dictated by different cells adopting different patterns of gene expression depending upon where they are within the embryo. The pattern of expression of the *even-skipped* gene is shown in Figure 2.45. The gene *even-skipped* (*eve*) is expressed in seven stripes corresponding to 7 of the 14 *Drosophila* segments. One of the most remarkable features of the pattern of gene expression illustrated in the figure is its spatial sharpness, with the discrimination between the relevant gene being "on" or "off" taking place on a length scale comparable to the cell size itself. A fundamental challenge in the study of embryonic development has been the quest to understand the connection between patterns of gene expression in the embryo and the generation of macroscopic form.

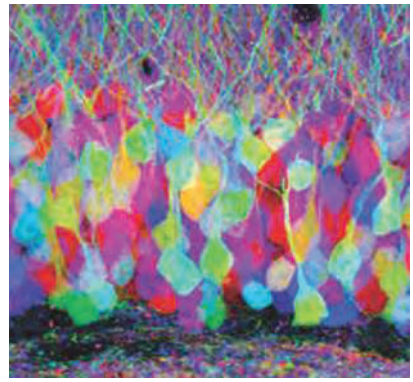


Figure 2.42: Illustration of a complex network of cells formed by neurons. Using multiple fluorophores simultaneously to create a BrainBow, the structure of the nerves is elucidated. (Adapted from J. W. Lichtman et al., *Nat. Rev. Neurosci.* 9:417, 2008.)

Estimate: Sizing Up Stripes in *Drosophila* Embryos The sharp patterns in gene expression exhibited by the pair-rule genes such as the *even-skipped* gene shown in Figure 2.45 result from spatial gradients of DNA-binding proteins known as transcription factors that either increase (activation) or decrease (repression) the level of gene expression. To get a feeling for the scales associated with the gradients in transcription factors that dictate developmental decisions and the features they engender, we idealize the *Drosophila* embryo as a spherocylinder. We characterize the geometry by two parameters, namely, the length of the cylindrical region, L, and its radius R,

ESTIMATE

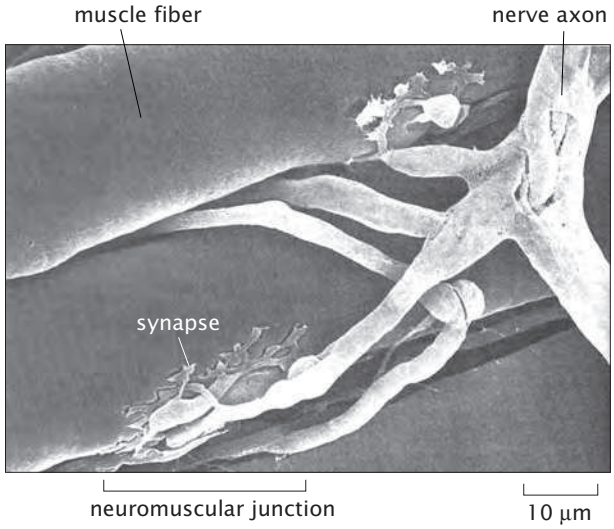


Figure 2.43: Neuromuscular junction. The axon from a nerve cell makes contact with various skeletal muscle fibers. Each muscle fiber is a giant multinucleated cell formed by the fusion of hundreds or thousands of precursor cells. Neurotransmitters secreted by the nerve cell at the synapse initiate contraction of the muscle fibers. (Adapted from D. W. Fawcett and R. P. Jensh, Bloom and Fawcett's Concise Histology. Hodder Arnold, 1997.)

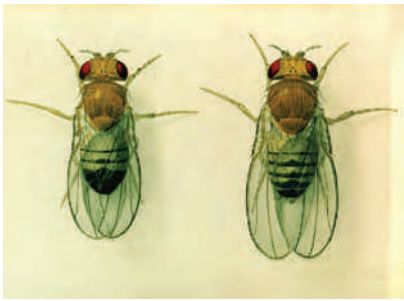


Figure 2.44: The male (left) and female (right) adult wild-type *Drosophila melanogaster*. After the 1934 drawing by Edith M. Wallace. (Courtesy of the Archives, California Institute of Technology.)

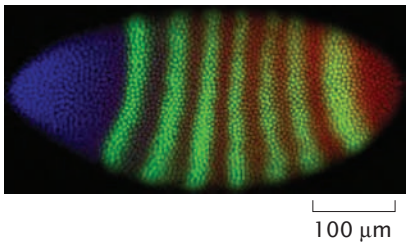


Figure 2.45: Pattern of gene expression in the *Drosophila* embryo. This early-stage fly embryo has undergone 13 rounds of cell division. The protein products of three different genes have been labeled using immunofluorescence: Bicoid in blue, Even-skipped in green, and Caudal in red. The small dots are individual nuclei. (Embryo be10 from the FlyEx database: http://urchin.spbcas.ru/flyex)

where we use approximate values of 500 μm for the length of the cylindrical region and 100 μm for the radius. In the embryo, many of the nuclei are on the surface and, as a result, our estimates will depend upon having a rough estimate of their areal density. The area of the embryo in the simple geometrical model used here is given by

$$A = 4\pi R^2 + 2\pi RL. \tag{2.27}$$

We consider the embryo after 13 divisions and before gastrulation, which brings the precursors of the muscles and internal organs to the interior of the embryo. One interesting fact about fruit fly development is that these initial divisions occur as nuclear divisions without the formation of cellular membranes which would give rise to cells. Also, note that there are not 2^{13} nuclei, since, during the preceding divisions, not all of the nuclei went to the surface of the embryo, and those that are left behind in the interior of the embryo do not keep pace with those on the surface during subsequent divisions. Instead, at this stage in development, there are on the order of $N = 6000$ nuclei on the surface. See Zalokar and Erk (1976) for details of the nuclear census during embryonic development. The areal density (that is, the number per unit area) is given by

$$\sigma = \frac{N}{4\pi R^2 + 2\pi RL}. \tag{2.28}$$

Using the numbers described above leads to an areal density of 0.01 cells/μm^2. As seen in Figure 2.45, the stripe patterns associated with the *Drosophila* embryo are very sharp and reflect cells making decisions at a very localized level.

Inspection of Figure 2.45 reveals that the stripes are roughly 20 μm wide. As a result, we can estimate the total number of nuclei participating in each stripe as

$$\begin{aligned} n = \sigma 2\pi R l_{\text{stripe}} &\approx 0.01 \text{ cells}/\mu\text{m}^2 \times 2 \times 3.14 \times 100\ \mu\text{m} \times 20\,\mu\text{m} \\ &\approx 125 \text{ nuclei}. \end{aligned} \tag{2.29}$$

Note that the average area per nucleus is given by $1/\sigma \approx 100\,\mu$m, suggesting that the radii of these nuclei is roughly 6 μm. Our main purpose in carrying out this exercise is to demonstrate the length scale of the structures that are put down during embryonic development. In this case, what we have seen is that out of the roughly 6000 nuclei that characterize the *Drosophila* embryo at the time of gastrulation, groups of roughly 120 nuclei have begun to follow distinct pathways as a result of differential patterns of gene expression that foreshadow the heterogeneous anatomical features of the mature fly.

The Cells of the Nematode Worm, *Caenorhabditis Elegans*, Have Been Charted, Yielding a Cell-by-Cell Picture of the Organism

A more recently popularized model organism for studying the genetics and development of multicellular animals is the nematode worm *Caenorhabditis elegans* (commonly abbreviated *C. elegans*) shown in Figure 2.46. Several factors that make this worm particularly attractive

Figure 2.46: *C. elegans*. This differential interference contrast image of a single adult worm was assembled from a series of high-resolution micrographs. The worm's head is at the top left-hand corner and its tail is at the right. Its gut is visible as a long tube going down the animal's body axis. Its egg cells are also visible as giant ovals towards the bottom of the body. (Courtesy of I. D. Chin-Sang.)

in its capacity as a model multicellular eukaryote are that (i) its complete genome has been determined, (ii) the identity of each and every one of its 959 cells has been determined (see Figure 2.47), and (iii) the worms are transparent, permitting a variety of structural investigations. Amazingly, *all* of the cells of this organism have had their lineages traced from the single cell that is present at the moment of fertilization. What this means precisely is that all of the roughly 1000 cells making up this organism can be assigned a lineage of the kind cell A begat cell B which begat cell C, and so on.

As shown in Figure 2.46, these worms are roughly 1 mm in length and 0.05 mm across. Like *Drosophila*, they too have been subjected to a vast array of different analyses, including, for example, how their behavior is driven by the sensation of touch. One of the most remarkable outcomes of the series of experiments leading to the lineage tree for *C. elegans* was the determination of the connectivity of the 302

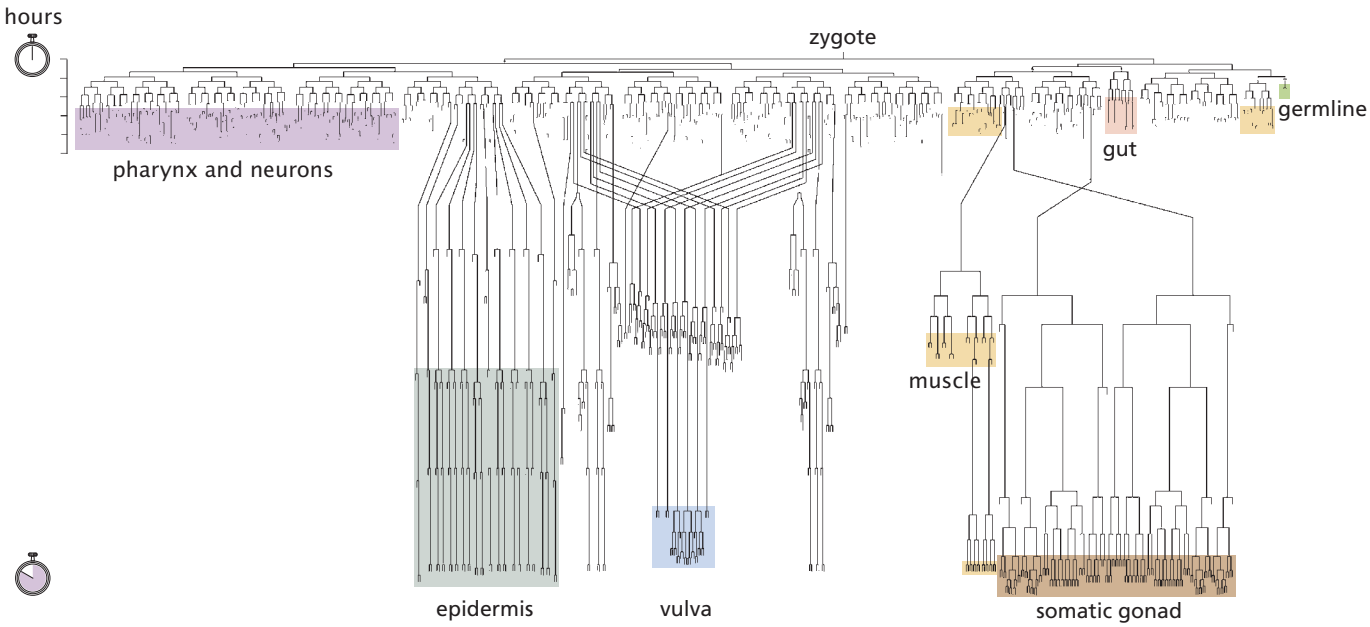


Figure 2.47: Cell lineages in *C. elegans*. The developmental pattern of every cell in the worm is identical from one animal to another. Therefore, it has been possible for developmental biologists to determine the family tree of every cell present in the entire animal by patient direct observation. In this schematic representation, the cell divisions that occur during embryogenesis are shown in the band across the top. The later cell divisions of the epidermis, vulva, and somatic gonad all take place after the animal has hatched.

neurons present in this nematode. By using serial thin sections from electron microscopy, it was possible to map out the roughly 7000 neuronal connections in the nervous system of this tiny organism. The various nerve cells are typically less than 5 μm across.

Estimate: Sizing Up *C. elegans* As an estimate of the cellular content of a "simple" animal, *C. elegans* provides a way of estimating the characteristic size of eukaryotic cells. As suggested by Figure 2.46, these small worms can be thought of as cylinders of length 1 mm and with a width of 0.1 mm. The total volume of such a worm is computed simply as roughly $7 \times 10^6\,\mu m^3$. If we use the fact that there are roughly 1000 cells in the organism, this tells us that the characteristic volume of the cells is $7000\,\mu m^3$, implying a typical radius for these cells of roughly 10 μm. A reasonable estimate for the volume of eukaryotic cells is somewhere between 2000 and $10{,}000\,\mu m^3$.

Many different species of nematodes share a common body plan involving a small number of well-defined cells; however, they may vary greatly in size. One of the largest nematodes, *Ascaris*, is a common parasite of pigs and humans. It has been estimated that up to 1 billion people on the planet carry *Ascaris* in their intestine. These worms closely resemble *C. elegans* except that they may be up to 15 cm in length and their cells are correspondingly larger. This kind of observation is not unusual throughout the animal kingdom. Many species have close relatives that differ enormously in size. The reasons and mechanisms that determine the overall size of multicellular organisms remain poorly understood.

Higher-Level Structures Exist as Colonies of Organisms

Organisms do not exist in isolation. Every organism on the planet is part of a larger ecological web that features both cooperation and competition with other individuals and species ultimately ranging up to the interconnected ecosystem of planet Earth. Coral reefs represent a vivid example of the interdependence of huge varieties of species living together in close quarters. An equally vivid and diverse community, though one frequently less appreciated for its intrinsic beauty, is found closer to home within our own intestines, which are inhabited by a teeming variety of bacteria. It has been estimated that the human body harbors nearly 10 times more microbial cells than it does human cells, and at least 100 times more bacterial genes than human genes. Thus, we should think of ourselves not really as individuals but rather as complex ecosystems of which the human cells form only a small part. Thus, our story of the hierarchy of structures that make up the living world began with the bacterium *E. coli* and will end there now. It is an irony that we as humans, it might be argued, have been colonized by our standard ruler *E. coli*.

Though we have examined biological structures over a wide range of spatial scales, our powers-of-10 journey still falls far short of being comprehensive. Ultimately, what we learn from this is that the handiwork of evolution has resulted in biological structures ranging from the nanometer scale of molecular machines all the way to the scale of the planet itself. Trying to understand what physical and biological factors drive the formation of these structures will animate much of the remainder of this book.

2.4 Summary and Conclusions

Biological structures range from the scale of nanometers (individual proteins) to hundreds of meters (redwood trees) and beyond (ecological communities). In this chapter, we have explored the sizes and numbers of biological entities, starting with the unit of life, the individual cell, and working our way down and up. We have found that sometimes biological objects show up in very large numbers of identical copies such that averages, for example of the concentration of ribosomes in the *E. coli* cytoplasm, are reasonable approximations. On the other hand, sometimes biological things show up in very few copies such that the exact number can make a big difference in the behavior of the system. We have found that cells are crowded, that there really is a world in a grain of sand (or in a biological cell). We have also explored some of the ways that biological units such as proteins or cells may interact with one another to form larger and more complex entities. Having developed an intuitive feeling for size and scale will better enable us to realistically envision the biological processes and problems described throughout the remainder of the book.

2.5 Problems

Key to the problem categories: • Model refinements and derivations • Estimates • Computational simulations

• 2.1 Revisiting the *E. coli* mass

We made a simple estimate of the mass of an *E. coli* cell by assuming that such cells have the same density of water. However, a more reasonable estimate is that the density of the macromolecules of the cell is 1.3 times that of water (BNID 101502, 104272). As a result, the estimate of the mass of an *E. coli* cell is off by a bit. Using that two-thirds of the mass is water and that the remaining one-third is macromolecular, compute the percentage error made by treating the macromolecular density as the same as that of water.

• 2.2 A feeling for the numbers: the chemical composition of a cell by pure thought

Make an estimate of the composition of carbon, hydrogen, oxygen, and nitrogen in the dry mass of a bacterium. Using knowledge of the size and mass of a bacterium, the fraction of that mass that is "dry mass" (that is, $\approx 30\%$) and the chemical constituents of a cell, figure out the approximate small integers (<10) for the composition $C_mH_nO_pN_q$, that is, find m, n, p, and q.

• 2.3 A feeling for the numbers: microbes as the unseen majority

(a) Use Figure 2.1 to justify the assumption that a typical bacterial cell (that is, *E. coli*) has a surface area of $6\,\mu m^2$ and a volume of $1\,\mu m^3$. Also, express this volume in femtoliters. Make a corresponding estimate of the mass of such a bacterium.

(b) Roughly 2–3 kg of bacteria are harbored in your large intestine. Make an estimate of the total number of bacteria inhabiting your intestine. Estimate the total number of human cells in your body and compare the two figures.

(c) The claim is made (see Whitman et al., 1998) that in the top 200 m of the world's oceans, there are roughly 10^{28} prokaryotes. Work out the total volume taken up by these cells in m^3 and km^3. Compute their mean spacing. How many such cells are there per milliliter of ocean water?

• 2.4 A feeling for the numbers: molecular volumes and masses

(a) Estimate the volumes of the various amino acids in units of nm^3.

(b) Estimate the mass of a "typical" amino acid in daltons. Justify your estimate by explaining how many of each type of atom you chose. Compare your estimate with the actual mass of several key amino acids such as glycine, proline, arginine, and tryptophan.

(c) On the basis of your result for part (b), deduce a rule of thumb for converting the mass of a protein (reported in kDa) into a corresponding number of residues. Apply this rule of thumb to myosin, G-actin, hemoglobin, and hexokinase and compare your results with the actual number of residues in each of these proteins. Relevant data for this problem are provided on the book's website.

• 2.5 Minimal media and *E. coli*

Minimal growth medium for bacteria such as *E. coli* includes various salts with characteristic concentrations in the mM range and a carbon source. The carbon source is typically glucose and it is used at 0.5% (a concentration of 0.5 g/100 mL). For nitrogen, minimal medium contains ammonium chloride (NH_4Cl) with a concentration of 0.1 g/100 mL.

(a) Make an estimate of the number of carbon atoms it takes to make up the macromolecular contents of a bacterium such as *E. coli*. Similarly, make an estimate of the number of nitrogens it takes to make up the macromolecular contents of a bacterium? What about phosphate?

(b) How many cells can be grown in a 5 mL culture using minimal medium before the medium exhausts the carbon? How many cells can be grown in a 5 mL culture using minimal medium before the medium exhausts the nitrogen? Note that this estimate will be flawed because it neglects the *energy* cost of synthesizing the macromolecules of the cell. These shortcomings will be addressed in Chapter 5.

• 2.6 Atomic-level representations of biological molecules

(a) Obtain coordinates for several of the following molecules: ATP, phosphatidylcholine, B-DNA, G-actin, the lambda repressor/DNA complex or Lac repressor/DNA complex, hemoglobin, myoglobin, HIV gp120, green fluorescent protein (GFP), and RNA polymerase. You can find the coordinates on the book's website or by searching in the Protein Data Bank and various other Internet resources.

(b) Download a structural viewing code such as VMD (University of Illinois), Rasmol (University of Massachusetts), or DeepView (Swiss Institute of Bioinformatics) and create a plot of each of the molecules you downloaded above. Experiment with the orientation of the molecule and the different representations shown in Figure 2.32.

(c) By looking at phosphatidylcholine, justify (or improve upon) the value of the area per lipid ($0.5\,\mathrm{nm}^2$) used in the chapter.

(d) Phosphoglycerate kinase is a key enzyme in the glycolysis pathway. One intriguing feature of such enzymes is their enormity in comparison with the sizes of the molecules upon which they act (their "substrate"). This statement is made clear in Figure 5.5 (p. 195). Obtain the coordinates for both phosphoglycerate kinase and glucose and examine the relative size of these molecules. The coordinates are provided on the book's website.

• 2.7 Coin flips and partitioning of fluorescent proteins

In the estimate on cell-to-cell variability in the chapter, we learned that the standard deviation in the number of molecules partitioned to one of the daughter cells upon cell division is given by

$$\langle n_1^2 \rangle - \langle n_1 \rangle^2 = Npq. \tag{2.30}$$

(a) Derive this result.

(b) Derive the simple and elegant result that the average difference in intensity between the two daughter cells is given by

$$\langle (I_1 - I_2)^2 \rangle = \alpha I_{\mathrm{tot}}, \tag{2.31}$$

where I_1 and I_2 are the intensities of daughters 1 and 2, respectively, and I_{tot} is the total fluorescence intensity of the mother cell and assuming that there is a linear relation between intensity and number of fluorophores of the form $I = \alpha N$.

• 2.8 HIV estimates

(a) Estimate the total mass of an HIV virion by comparing its volume with that of an *E. coli* cell and assuming they have the same density.

(b) The HIV maturation process involves proteolytic clipping of the Gag polyprotein so that the capsid protein CA can form the shell surrounding the RNA genome and nucleocapsid NC can complex with the RNA itself. Using Figures 2.30 and 2.31 to obtain the capsid dimensions, estimate the number of CA proteins that are used to make the capsid and compare your result with the total number of Gag proteins.

• 2.9 Areas and volumes of organelles

(a) Calculate the average volume and surface area of mitochondria in yeast based on the confocal microscopy image of Figure 2.18(C).

(b) Estimate the area of the endoplasmic reticulum when it is in reticular form using a model for its structure of interpenetrating cylinders of diameter $d \approx 10\,\mathrm{nm}$ separated by a distance $a \approx 60\,\mathrm{nm}$, as shown in Figure 2.25.

• 2.10 An open-ended "feeling for the numbers": the cell

Using the figures of cells and their organelles provided on the book website, carry out estimates of the following:

(a) The number of nuclear pores in the nucleus of a pancreatic acinar cell.

(b) The spacing between mitochondrial lamellae and the relative area of the inner and outer mitochondrial membranes.

(c) The spacing and areal density of ribosomes in the rough endoplasmic reticulum of a pancreatic acinar cell.

(d) The DNA density in the head of a sperm.

(e) The volume available in the cytoplasm of a leukocyte.

2.6 Further Reading

Goodsell, D (1998) The Machinery of Life, Springer-Verlag. This book illustrates the crowded interior of the cell and the various macromolecules of life.

Fawcett, DW (1966) The Cell, Its Organelles and Inclusions: An Atlas of Fine Structure, W. B. Saunders. We imagine our reader comfortably seated with a copy of Fawcett right at his or her side. Fawcett's electron microscopy images are stunning.

Fawcett, DW, & Jensh, RP (2002) Bloom and Fawcett: Concise Histology, 2nd ed., Hodder Arnold. This book shows some of the beautiful diversity of both cells and their organelles.

Gilbert, SF (2010) Developmental Biology, 9th ed., Sinauer Associates. Gilbert's book is a source for learning about the architecture of a host of different organisms during early development. Chapter 6 is especially relevant for the discussion of this chapter.

Wolpert, L, Tickle, C, Lawrence, P, et al. (2011) Principles of Development, 4th ed., Oxford University Press. Wolpert's book is full of useful cartoons and schematics that illustrate many of the key ideas of developmental biology.

2.7 References

Alberts, B, Johnson, A, Lewis, J, et al. (2008) Molecular Biology of the Cell, 5th ed., Garland Science.

Baker, TS, Olson, NH, & Fuller, SD (1999) Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs, *Microbiol. Mol. Biol. Rev.* **63**, 862.

Birk, DE, & Trelstad, RL (1986) Extracellular compartments in tendon morphogenesis: collagen fibril, bundle, and macroaggregate formation, *J. Cell Biol.*, **103**, 231.

Brecht, M, Fee, MS, Garaschuk, O, et al. (2004) Novel approaches to monitor and manipulate single neurons *in vivo*, *J. Neurosci.* **24**, 9223.

Briggs, JAG, Simon, MN, Gross, I, et al. (2004) The stoichiometry of Gag protein in HIV-1, *Nat. Struct. Mol. Biol.* **11**, 672.

Briggs, JAG, Grünewald, K, Glass, B, et al. (2006) The mechanism of HIV-1 core assembly: Insights from three-dimensional reconstruction of authentic virions, *Structure* **14**, 15.

Brügger, B, Glass, B, Haberkant, P, et al. (2006) The HIV lipidome: a raft with an unusual composition, *Proc. Natl Acad. Sci. USA* **103**, 2641.

Chin, DJ, Luskey, KL, Anderson, RGW, et al. (1982) Appearance of crystalloid endoplasmic reticulum in compactin-resistant Chinese hamster cells with a 500-fold increase in 3-hydroxy-3-methylglutaryl-coenzyme A reductase, *Proc. Natl Acad. Sci. USA* **79**, 1185.

Cross, PC, & Mercer, KL (1993) Cell and Tissue Ultrastructure, W. H. Freeman. A great pictorial source for anyone interested in the beauty of cells and their organelles.

Egner, A, Jakobs, S, & Hell, SW (2002) Fast 100-nm resolution three-dimensional microscope reveals structural plasticity of mitochondria in live yeast, *Proc. Natl. Acad. Sci.*, **99**, 3370.

Fawcett, DW (1986) Bloom and Fawcett—A Textbook of Histology, 11th ed., W. B. Saunders.

Finlay, BJ (2002) Global dispersal of free-living microbial eukaryote species, *Science* **296**, 1061.

Ghaemmaghami, S, Huh, W-K, Bower, K, et al. (2003) Global analysis of protein expression in yeast, *Nature* **425**, 737.

Godin, M, Delgado, FF, Son, S, et al. (2010) Using buoyant mass to measure the growth of single cells, *Nat. Methods*, **7**, 387.

Golding, I, Paulsson, J, Zawilski, SM, & Cox, EC (2005) Real-time kinetics of gene activity, *Cell* **123**, 1025.

Jorgensen, P, Nishikawa, JL, Breitkreutz, BJ, & Tyers M (2002) Systematic identification of pathways that couple cell growth and division in yeast, *Science*, **297**, 395.

Lichtman, JW, Livet, J, & Sanes, JR (2008) A technicolour approach to the connectome, *Nat. Rev. Neurosci.* **9**, 417.

Lu, P, Vogel, C, Wang, R, et al. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation, *Nat. Biotechnol.*, **25**, 117.

Neidhardt, FC, Ingraham, JL, & Schaechter, M (1990) Physiology of the Bacterial Cell, Sinauer Associates. This outstanding book is full of interesting facts and insights concerning bacteria.

Pedersen, S, Bloch, PL, Reeh, S, & Neidhardt, FC (1978) Patterns of protein synthesis in *E. coli*: A catalog of the amount of 140 individual proteins at different growth rates, *Cell* **14**, 179.

Poustelnikova, E, Pisarev, A, Blagov, M, et al. (2004) A database for management of gene expression data *in situ*, *Bioinformatics* **20**, 2212.

Radmacher, M (2007) Studying the mechanics of cellular processes by atomic force microscopy, *Methods Cell. Biol.*, **83**, 347.

Rosenfeld, N, Young, JW, Alon, U, Swain, PS, & Elowitz, MB (2005) Gene regulation at the single-cell level, *Science* **307**, 1962.

Schaechter, M, Ingraham, JL, & Neidhardt, FC (2006) Microbe, ASM Press.

Schulze, E, & Kirschner, M (1987) Dynamic and stable populations of microtubules in cells, *J. Cell Biol.* **104**, 277.

Stoodley, P, Sauer, K, Davies, DG, & Costerton, JW (2002) Biofilms as complex differentiated communities, *Annu. Rev. Microbiol.* **56**, 187.

Taniguchi, Y, Choi, PJ, Li, GW, et al. (2010) Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells, *Science*, **329**, 533.

Terasaki, M, Chen, LB, & Fujiwara, K (1986) Microtubules and the endoplasmic reticulum are highly interdependent structures, *J. Cell. Biol.* **103**, 1557.

van der Flier, LG, & Clevers, H (2009) Stem cells, self-renewal, and differentiation in the intestinal epithelium, *Annu. Rev. Physiol.*, **71**, 241.

Visser, W, Van Sponsen, EA, Nanninga, N, et al. (1995) Effects of growth conditions on mitochondrial morphology in *Saccharomyces cerevisiae*, *Antonie van Leeuwenhoek* **67**, 243.

Walker, GM (1998) Yeast, Physiology and Biotechnology, Wiley.

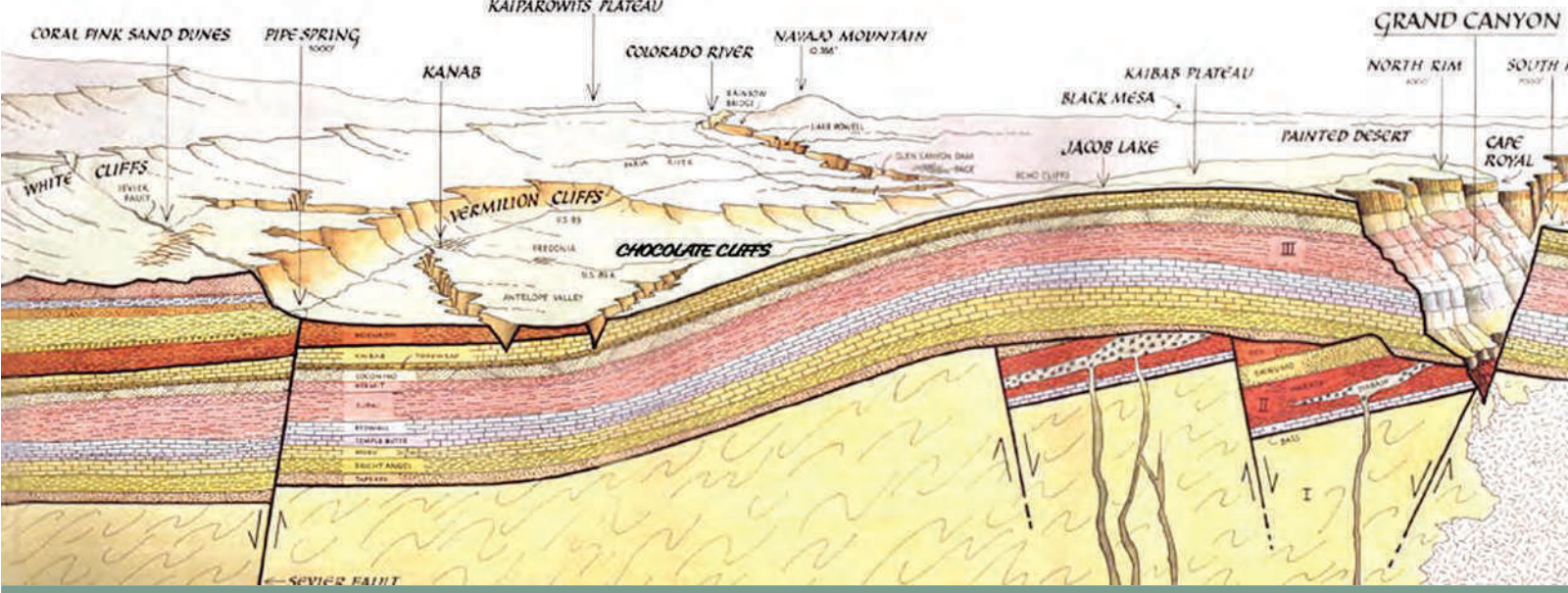
Whitman, WB, Coleman, DC, & Wiebe, WJ (1998) Prokaryotes: the unseen majority, *Proc. Natl Acad. Sci. USA* **95**, 12, 6578.

Wu, J-Q, & Pollard, TD (2005) Counting cytokinesis proteins globally and locally in fission yeast, *Science* **310**, 310. This paper illustrates the strategy of using fluorescence microscopy for taking the census of a cell.

Zalokar, M, & Erk, I (1976) Division and migration of nuclei during early embryogenesis of *Drosophila melanogaster*, *J. Microscopie Biol. Cell.* **25**, 97. This paper makes a census of the number of cells in the *Drosophila* embryo at various stages in development.

Zimmerman, SB, & Trach, SO (1991) Estimation of macromolecule concentrations and excluded volume effects for the cytoplasm of *Escherichia coli*, *J. Mol. Biol.* **222**, 599. Discussion of crowding in the bacterial cell.

This page intentionally left blank
to match pagination of print book



When: Stopwatches at Many Scales

3

Overview: In which various stopwatches are used to measure the rate of biological processes

Just as biological structures exist over a wide range of spatial scales, biological processes take place over time scales ranging from less than a microsecond to the age of the Earth itself. Using the cell cycle of *E. coli* as a standard stopwatch, this chapter develops a feeling for the rates at which different biological processes occur and illustrates the great diversity of these processes, mirroring the structural diversity seen in the previous chapter. With this "feeling for the numbers" in hand, we then explore several different views of the passage of biological time. In particular, we show how sometimes cells manage time by stringing together processes in succession while at other times they manipulate time (using enzymes to speed up reactions, for example) to alter the intrinsic rates of processes.

> "Some day people may look back on the isotope as being as important to medicine as the microscope."
>
> **Archibald Hill**

3.1 The Hierarchy of Temporal Scales

One of the defining features of living systems is that they are dynamic. The time scales associated with biological processes run from the nanosecond (and faster) scale of enzyme action to the more than 10^9 years that cover the evolutionary history of life and the Earth itself, as indicated in the map that opens this chapter. The inexorable march of biological time is revealed over many orders of magnitude in time scale, as illustrated in Figure 3.1. If we watch biological systems unfold with different stopwatches in hand, the resulting phenomena will be different: at very fast time scales, we will see the molecular dance of different biochemical species as they interact and change identity; at much slower scales, we see the unfolding of the lives of individual cells. If we slow down our stopwatch even more, what we see are the trajectories of entire species. To some extent, there is a coupling between the temporal scales described in this chapter and

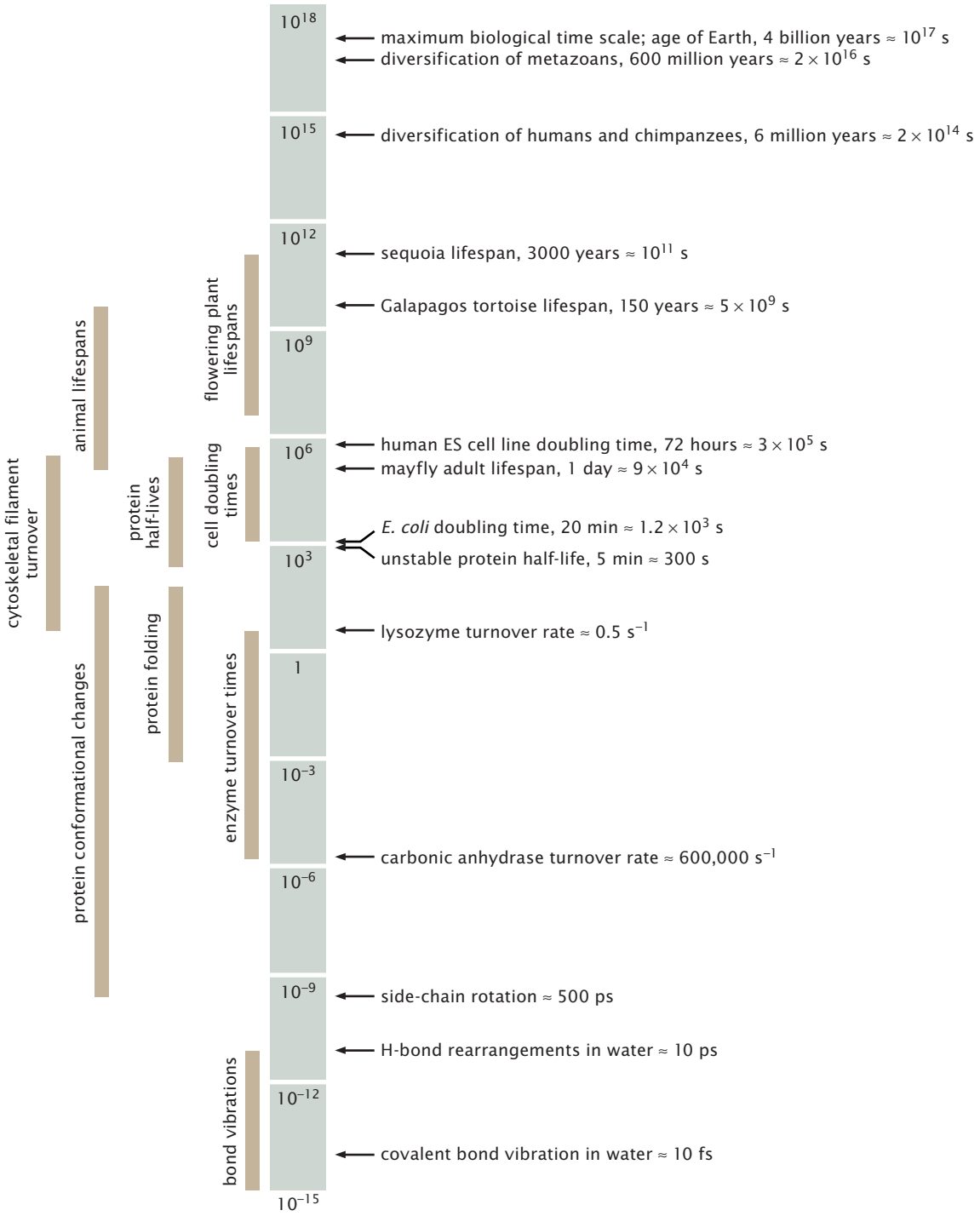


Figure 3.1: Gallery of biological time scales. Logarithmic scale showing the range of time scales associated with various biological processes. The time scale is in seconds.

the spatial scales described in the previous chapter; small things such as individual molecules tend to operate at fast rates, and large things such as elephants tend to move around at slow rates.

This chapter describes the time scales of biological phenomena from a number of different perspectives. In this section, we develop a feeling for biological time scales by examining the range of different time scales seen in molecular, cell, developmental, and evolutionary biology. This discussion is extended by describing the experimental basis for what we know about time scales in biology. As in Chapter 2, we once again invoke *E. coli* as our reference system, this time by using

the cell cycle of our "reference cell" as the standard stopwatch. The remainder of the chapter is built around viewing time in biology from three distinct perspectives. In Section 3.2, we show how the time scale of certain biological processes is dictated by how long it takes some particular procedure (such as replication) to occur. We will refer to this as *procedural time*. Section 3.3 explores time from a different angle. In this case, we consider a broad class of biological processes whose timing is of the "socks before shoes" variety. That is, processes are linked in a sequential string, and in order for one process to begin, another must have finished. We will refer to this kind of time keeping as *relative time*. Finally, Section 3.4 reveals a third way of viewing time in biological processes, as a commodity to be manipulated. In a process we will call *manipulated time*, we show how cells and organisms find ways to either speed up or slow down key processes such as replication and metabolism.

3.1.1 The Pageant of Biological Processes

Biological Processes Are Characterized by a Huge Diversity of Time Scales

A range of different processes associated with individual organisms, and their associated time scales, is shown in Figure 3.2 (we leave a discussion of evolutionary processes for the next section). Broadly speaking, this figure shows a loose powers-of-10 representation of different biological processes in the same spirit as Figure 2.15 (p. 52) showed a powers-of-10 representation of spatial scales. Although this serves as a useful starting point, as we will see later in the chapter, an *absolute* measure of time in seconds or minutes is sometimes not the most useful way to think about the passage of time within cells. For example, embryonic development for humans takes drastically longer than for chickens, but the relative timing of common events can be meaningfully compared. For the moment, our discussion of Figure 3.2 is designed to give a feeling for the numbers; how long do various key biological processes actually take in absolute terms as measured in seconds, minutes and hours?

We begin (Figure 3.2A,B) with some of the processes associated with the development of the fruit fly *Drosophila melanogaster*. *Drosophila* has been one of the key workhorses of developmental biology, and much that we know about embryonic development was teased out of watching the processes that take place over the roughly 10 days between fertilization of the egg and the emergence of a fully functioning adult fly. If we increase our temporal resolution by a factor of 10, we see the processes in the development of the fly embryo. Over the first 10 hours or so after fertilization, as shown in Figure 3.2(B), a single cell is turned into an organized collection of thousands of cells with particular spatial positions and functions. One of the most dramatic parts of embryonic development is the process of gastrulation when the future gut forms as a result of a series of folding events in the embryo. This process is indicated schematically in Figure 3.2(B).

Individual cells have a natural developmental cycle as well. The *cell cycle* refers to the set of processes whereby a single cell, through the process of cell division, becomes two daughter cells. The time scales associated with the cell cycle of a bacterium such as *E. coli* are shown in Figure 3.2(C), with a characteristic scale of several thousand seconds. The lives of individual cells are fascinating and complex. If we

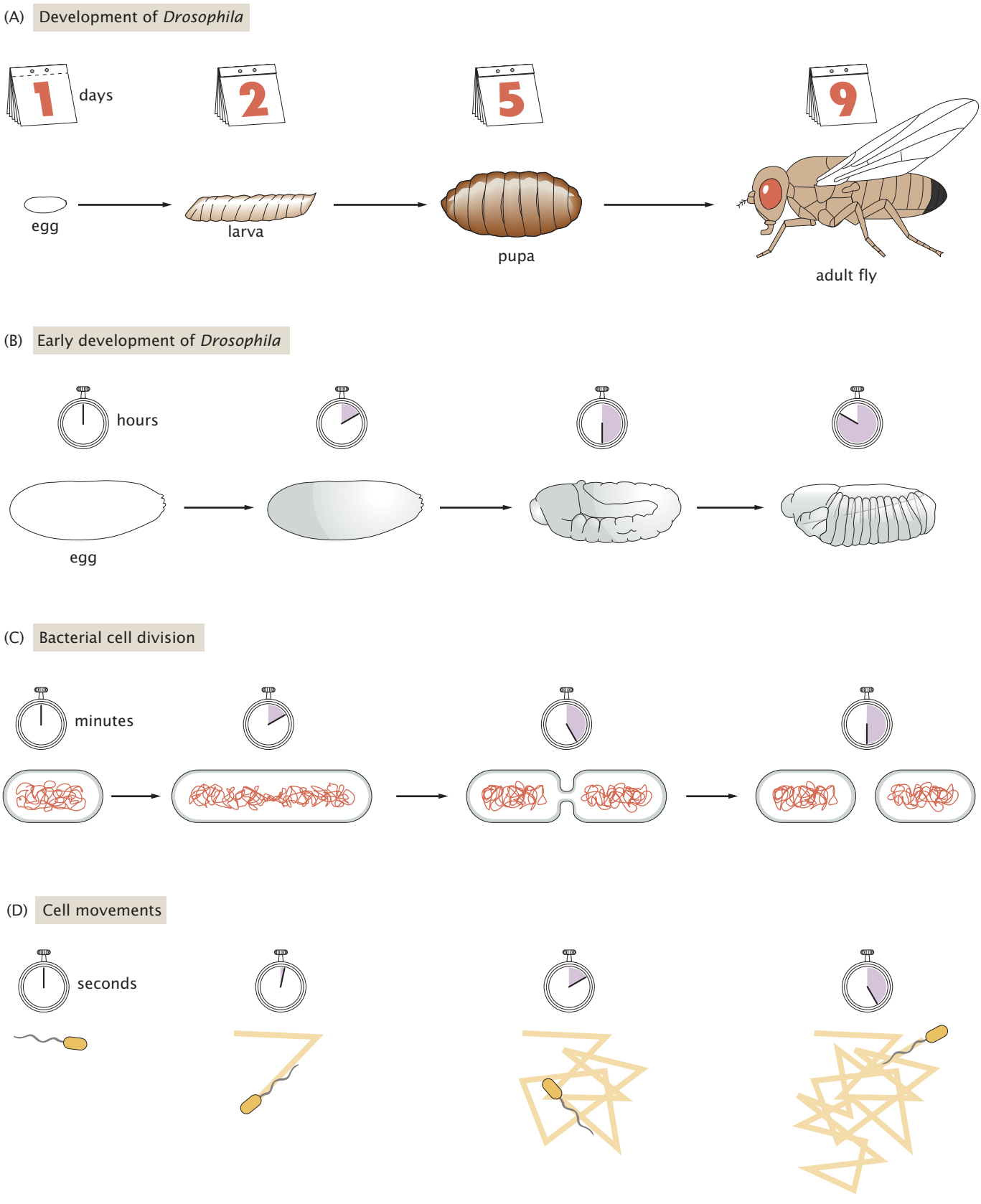


Figure 3.2: Hierarchy of biological time scales. Cartoon showing range of time scales associated with different biological processes. (A) Development of *Drosophila*. (B) Early development of *Drosophila* embryo. (C) Bacterial cell division. (D) Cell movements. (E) Protein synthesis. (F) DNA transcription. (G) Gating of ion channels. (H) Enzyme catalysis.

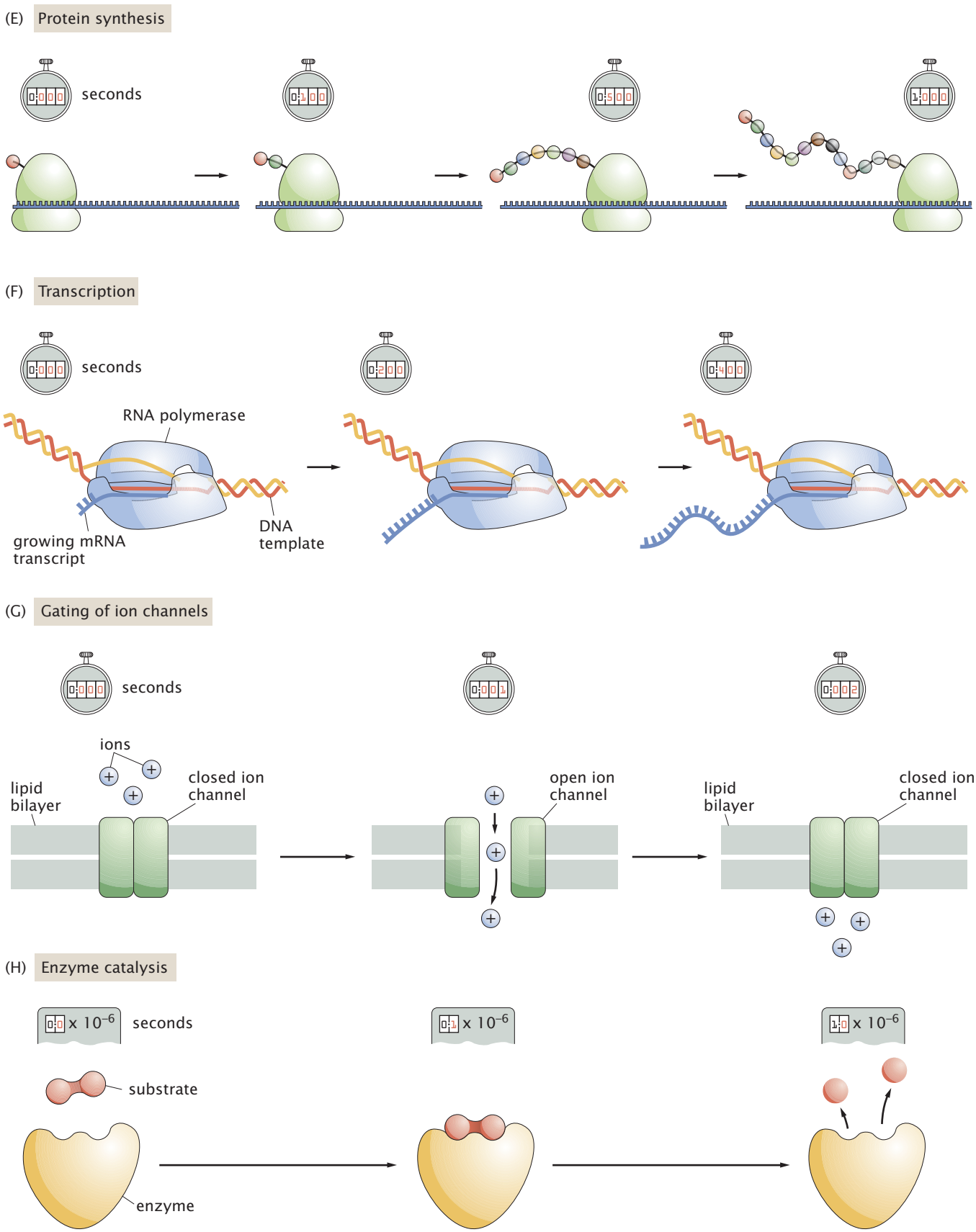


Figure 3.2: Continued.

were to dissect the activities of an individual cell as it goes about its business between cell divisions, we would find a host of processes taking place over a range of different time scales. If we stare down a microscope at a swimming bacterium for several seconds, we will notice episodes of directed motion, punctuated by rapid directional changes. Figure 3.2(D) shows the time scales over which an individual bacterium such as *E. coli* exercises its random excursion during movement. If our stopwatch runs a factor of 10 faster, we are now operating at the scale of deciseconds, a scale that characterizes the rate of amino acid incorporation during protein synthesis, a process represented in Figure 3.2(E). Macromolecular synthesis is one of the most important sets of processes that any cell must undertake to make a new cell. Another key part of the macromolecular synthesis required for cell division is the process of transcription, which is the intermediate step connecting the genetic material as contained in DNA and the readout of that message in the form of proteins. Transcription refers to the synthesis of mRNA molecules as faithful copies of the nucleotide sequence in the DNA, a polymerization process catalyzed by the enzyme RNA polymerase. The incorporation by RNA polymerase of nucleotides onto the mRNA during transcription, as depicted schematically in Figure 3.2(F), happens a few times faster than amino acid incorporation by ribosomes during protein synthesis.

In the moment-to-moment life of the cell, proteins do most of the work. Many proteins are able to operate at time scales much faster than the relatively stately machinery carrying out the central dogma operations. For example, a great number of biological processes are dictated by the passage of ions across ion channels, with a characteristic time scale of milliseconds as shown in Figure 3.2(G). A factor-of-1000 speed-up of our stopwatch brings us to the world of enzyme kinetics at the microsecond time scale (Figure 3.2H) and faster. It is important to note that these time scales merely represent a general rule of thumb. For example, turnover rates for individual enzymes may range from $0.5\,s^{-1}$ to $600{,}000\,s^{-1}$.

Before proceeding, one of the questions we wish to consider is how the time scales depicted in Figure 3.2 are actually known. As with much of our story, the stopwatches associated with each of the cartoons in that figure have been determined as the results of many kinds of complementary experiments.

Experiments Behind the Facts: Measurements of Biological Time Broadly speaking, the experiments that characterize the dynamics of cells and the molecules that populate them are ultimately based on tracking transformations. We can divide these experiments into four broad categories that can be applied across all levels of spatial scale from molecular to ecological. These methods are summarized in Figure 3.3.

EXPERIMENTS

1. *Direct observation* The first and most obvious way to characterize time in a biological process is simply to observe the process unfold and to record the absolute time at which transformation occurs. An example of this strategy is shown in Figure 3.3, which shows the motion of a neutrophil. Similarly, looking down a microscope at a mammalian cell in tissue culture, it is possible to observe many of the steps in the cell cycle (see Figure 3.21 on p. 116) unfolding over real time, including condensation of the chromosomes, alignment of the chromosomes

METHOD	direct observation	fixed time points	pulse–chase	product accumulation
TYPICAL TIME SCALES	milliseconds to hours	microseconds to years	minutes to days	minutes to days
TYPES OF PROCESSES	individual transformations	population changes	continuous (e.g., metabolism, transport)	biosynthetic or enzymatic
EXAMPLE	cell crawling	bacterial growth curve	axonal transport	GFP expression

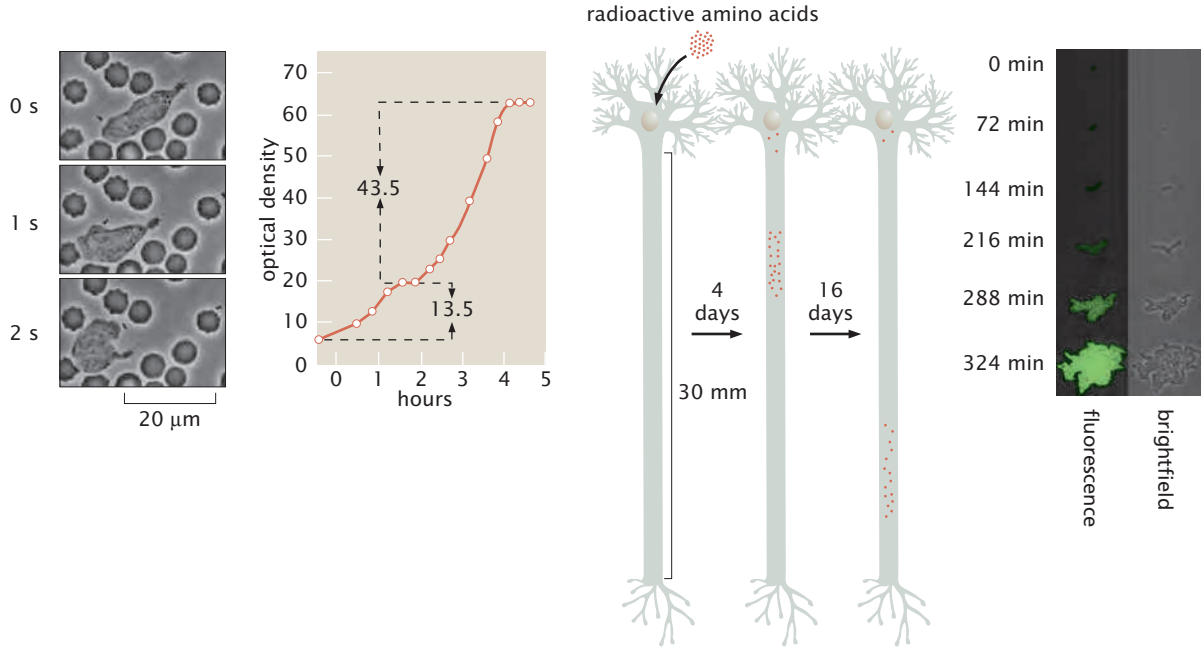


Figure 3.3: Experiments to measure the timing of biological processes. The figure summarizes four strategies for measuring biological rates. For *direct observation*, the example shows three frames from a video sequence taken by David Rogers in the 1950s of a single white blood cell (neutrophil) pursuing a bacterium through a forest of red blood cells. The movement of the cell is sufficiently fast that it can be directly observed by the human eye. For *fixed time points*, the experiment shown is a classic performed by Monod, who tracked the growth of *E. coli* in a single culture when two different nutrient sugars were mixed together. The bacteria initially consumed all of the available glucose and then their growth rate slowed as they switched over into a new metabolic mode enabling them to use lactose. For *pulse–chase*, labeling proteins at their point of synthesis in a neuron cell body with a pulse of radioactive amino acids followed by a chase of unlabeled amino acids was used to measure the rate of continuous axonal transport. *Product accumulation* is illustrated by the expression of GFP under a regulated promoter of interest in a bacterial cell. The rate of gene transcription off of such promoter can be inferred by measuring the amount of GFP present as a function of time. (Growth curve adapted from B. Müller-Hill, The *lac* Operon: A Short History of a Genetic Paradigm, Walter de Gruyter, 1996; neuronal transport adapted from B. Droz and C. P. Leblond, *Science* 137:1047, 1962; fluorescent image series adapted from N. Rosenfeld et al., *Science* 307:1962, 2005.)

through the action of the mitotic spindle, their segregation into daughter nuclei, and finally cytokinesis when the cell is pinched into two fully formed daughter cells. Although this is easy to do for processes that take minutes to hours and occur over spatial scales that can be observed with a light microscope or the unaided human eye, it is extremely difficult to measure time simply by observation for events that are very fast, very slow, very small, or very large. Over the past few decades, there have been vast experimental improvements in direct or near-direct observation of single molecules such that this naturalistic approach to "observing a lot just by

watching" can be applied all the way down to the molecular scale. We will see many examples of this approach throughout the book.

2. *Fixed time points* When events of interest cannot be directly observed, there are other ways to probe their duration. Rather than continuously observing an individual over time, one can draw individuals from a population at given time intervals and examine their properties at this series of fixed time points. For example, a bacterial population in a liquid culture started from a single cell will grow exponentially, then plateau, and eventually die off over a period of several days. Rather than staring at the tube continuously for several days, the essential kinetics of this process can be measured simply by examining cell density at some fixed interval such as every hour as shown in Figure 3.3. Similarly, the events of embryonic development for useful model organisms such as flies and frogs unfold over a period of days to weeks. However, under a given set of environmental conditions, the sequence and timing of these events are stereotyped from one individual to another. Therefore, the investigator can accurately describe the sequence of events in frog development by examining one dish of embryos an hour after fertilization, a second dish of embryos two hours after fertilization, etc. This is useful when the methods used to examine the embryos result in their death, for example, fixing them and staining for a particular protein of interest or preparing them for electron microscopy examination. At a much smaller spatial scale and faster time, the method of stopped-flow kinetics enables investigators to follow enzymatic events by mixing together an enzyme and its substrate and then squirting the mixture into a denaturing acid bath after fixed intervals of time. These methods are all more indirect than direct observation, but in many cases are technically easier, and different kinds of complementary information can be gleaned by comparing both for a single process.
3. *Pulse–chase* Many biological processes operate in a continuous fashion. For example, bacteria constantly take in sugar from their medium for energy and to generate the molecular building blocks to synthesize new constituents. The process of glycolysis converts a molecule of glucose into two molecules of pyruvate. Because glucose is continuously taken up and pyruvate is continuously generated, it is extremely difficult to measure how long the conversion process takes. The set of methods used to tackle these kinds of problems are generally called pulse–chase experiments. In this particular example, a bacterial cell may be fed glucose tagged with radioactive carbon for a very brief period of time, for example 1 minute. This is followed by feeding with nonradioactive glucose. Cells are then removed from the bacterial culture at various time intervals and their metabolites are examined to see which contain the

radioactive carbon. Over time, the amount of labeled glucose will decrease and the major radioactive species will pass through a series of intermediates until finally most of the radioactive carbon will be found in pyruvate. Thus, a pulse–chase experiment can be used to determine the order of intermediates in a metabolic pathway and also the amount of time it takes for the cell to perform each transformation. A classic example of this strategy to examine transport in neurons is shown in Figure 3.3. Essentially the same method is used by naturalists examining dispersion times of birds and other animals by tagging individuals with a band or radio-transmitter and releasing them back into their natural population to see where they end up and when.

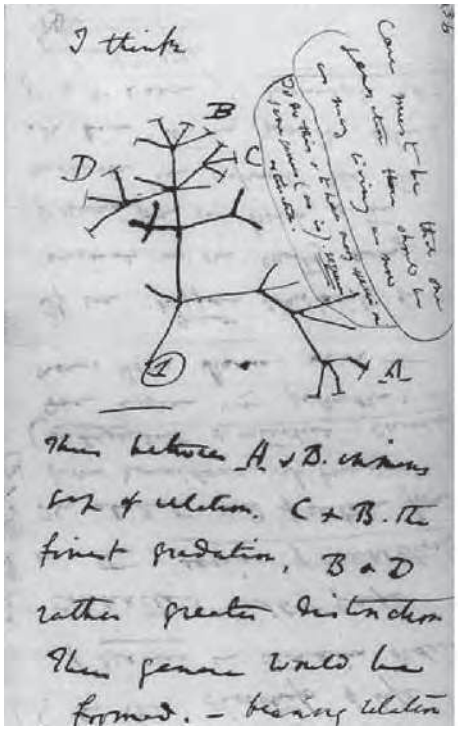
4. *Product accumulation* The final type of experiment used to determine biological rates is exemplified by an assay with a purified enzyme where a colorless substrate is converted into a colored product over time. By measuring the concentration of the colored product as a function of time, the investigator can extrapolate the average turnover rate given the known concentration of enzymes in the test tube. Similar experiments where observation of the accumulation of a product can be used as a surrogate for rate measurements can also be performed in living cells. A particularly useful example is expressing GFP (green fluorescent protein) downstream of a promoter of interest as shown in Figure 3.3. When the promoter is induced (that is, for example, by exposing the cells to some molecule that turns the gene of interest on) GFP begins to accumulate and the amount of fluorescence can be directly measured and converted into numbers of GFP molecules. Because GFP is remarkably stable, its accumulation can often represent a more accurate reporter for promoter activity than the promoter's natural product, which may be subject to other layers of regulation including rapid degradation.

3.1.2 The Evolutionary Stopwatch

The general rule that all biological processes are dynamic and undergo change over time applies to molecules, cells, organisms, and species. The evolutionary clock started more than 3 billion years ago with the appearance of the first cellular life forms on Earth. It is generally accepted that there were complex life-like processes occurring prior to the emergence of the first recognizable cells, though we cannot learn very much about what they were like either from the fossil record or from comparative studies among organisms living today.

All of the astonishing diversity of cellular life currently existing on the planet, ranging from archaea living in geothermal vents deep in the ocean to giant squid to redwood trees to the yeast that make beer, was descended from a universal common ancestor, probably a population of cells rather than an individual. This last universal common ancestor (LUCA) would have been clearly recognizable as cellular: it contained DNA as a genetic material, it transcribed its DNA into mRNA, and it translated mRNA into proteins using ribosomes. It

(A)



(B)

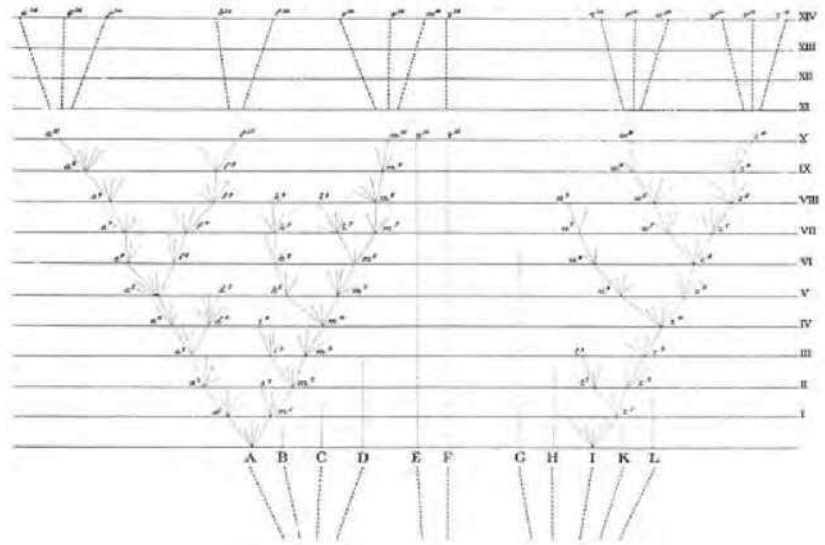


Figure 3.4: Two versions of Darwin's phylogenetic tree. (A) In his notebooks, Darwin drew the first version of what we now recognize as a common schematic demonstrating the relatedness of organisms. He introduced this speculative sketch with the words "I think," as his theory was beginning to take form. (B) In the final published version of *On the Origin of Species*, the tree had assumed more detail, showing the passage of time and explicitly indicating that most species have gone extinct. (Adapted from C. Darwin, *On the Origin of Species*, John Murray, 1859. Courtesy of The American Museum of Natural History.)

also processed sugar to make energy through the process of glycolysis and contained a rudimentary cytoskeleton consisting of an actin-like molecule and a tubulin-like molecule. We can attribute all of these features to LUCA because they are universally shared among all existing branches of cellular life. However, the demonstrable differences between redwood trees and giant squids accumulated slowly over evolutionary time as individual cellular populations became genetically isolated from one another and underwent change and divergence to fill different ecological niches. As the planet Earth is constantly being reshaped and remodeled by the uncounted legion of organisms that inhabit it, environmental niches are always unstable and can be changed by geological processes, global and local climate alterations, or the actions of competing organisms.

We can fruitfully think of evolution as the process of change in the genetic information carried by a population of related organisms. Sometimes, a single lineage can be seen as altering over time as its environment changes. More commonly, a single population will subdivide into populations that will become isolated and suffer different fates. Some will die off, some will remain similar to the ancestral population, and some will undergo significant biochemical, morphological, or behavioral alterations over time until they are ultimately recognized as new species. These basic ideas were beautifully articulated by Charles Darwin in *On the Origin of Species* and illustrated by the single figure in that book reproduced here as Figure 3.4.

How long does this evolutionary process take and how can we measure the passage of evolutionary time? It is unsatisfying to rely on

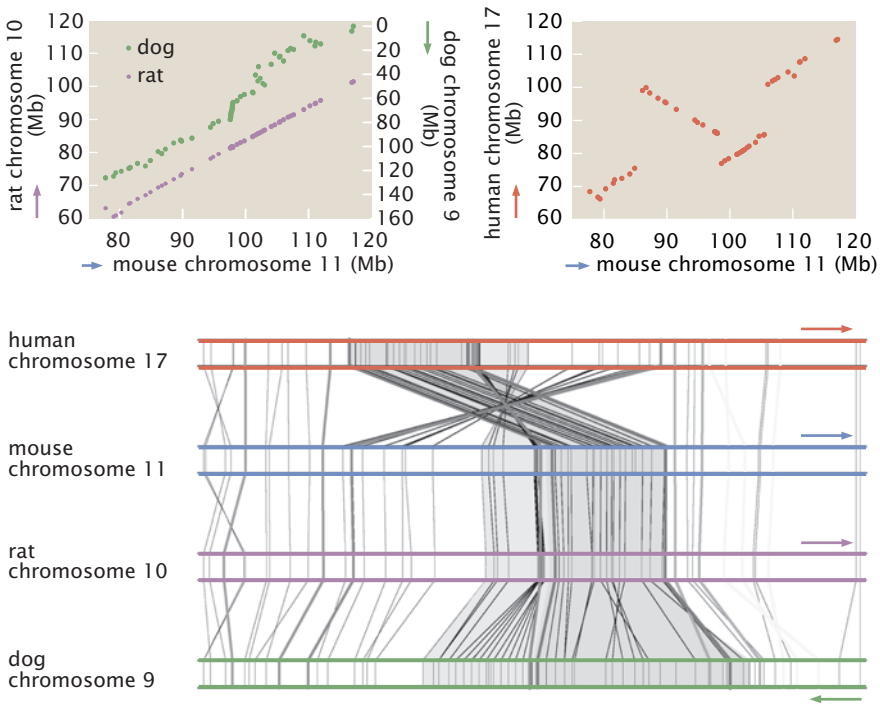


Figure 3.5: Inferring evolutionary relatedness by chromosome alignment. Equivalent regions of four chromosomes from human, mouse, rat, and dog were compared to find the location of homologous genes. The graphs at the top show the position of each gene in the rat, dog, and human sequences as a function of their positions on the mouse sequence. Because little change has occurred in chromosomal structure between the mouse and the rat, the points representing the locations of homologous genes form a nearly perfectly straight line. On the equivalent chromosomal segment from the dog, the genes are again mostly in the same order, but the spacing between them has changed substantially. The numbering convention for positions along the dog chromosome runs in the opposite direction compared with the mouse and rat chromosomes, as indicated by the arrows. Comparing the human with the mouse, a large inversion can be detected. The same data are shown in a different form in the chart at the bottom. Each vertical line on the chromosome represents a particular gene and the diagonal lines between the chromosomes link up homologs between human and mouse, mouse and rat, and rat and dog. The gray region shows a portion of the chromosome that can be followed through all the species; it is inverted between human and mouse, very similar in position and size between mouse and rat, and similar in orientation but increased in size between rat and dog because of changes in the spacing between the genes. (Adapted from G. Andelfinger et al., *Genomics* 83:1053, 2004.)

the extrapolation of mutation rates measured in artificial laboratory experiments to the evolution of species over time in the real world. Real-world conditions are much less stable or controlled than laboratory conditions, and furthermore the time scales of greatest interest for studying the evolution of species are much longer than can be achieved in the laboratory by even the most patient experimentalist. Traditionally, our understanding of evolutionary alterations depended upon two kinds of observations, namely, comparisons among currently living species and examination of the fossil record. Information about the age of particular fossils can be inferred from identification of the geological strata in which they are found, and also by examining the proportions of different radioisotopes that decay at a regular rate and thereby provide information about when the rock was formed.

Comparison of living species to ascertain the degree of relatedness was carried out for many hundreds of years before the modern theory of evolution was first described. It is immediately obvious that some organisms are more closely related than others. For example, horses and donkeys are clearly more similar to each other than either is to a dog, but horses and dogs are more similar to each other than either is to a squid. These obvious morphological differences have been the basis of the science of systematics going back to Linnaeus in the eighteenth century. In the modern era of molecular genetics, we can more easily ascribe a universal metric for genetic similarity among organisms based on similarities and differences in DNA sequence. As a population evolves over time, its DNA complement will change by several mechanisms. First, small-scale mutations or large-scale rearrangements of its genome may occur (an illustration of the consequences of this kind of rearrangement is shown in Figure 3.5). Second, the population may acquire new genes or even entire groups of genes by horizontal transfer from other organisms (horizontal transfer refers to cells passing genetic material to cells other than their own descendants). And third, the population may simply lose large chunks

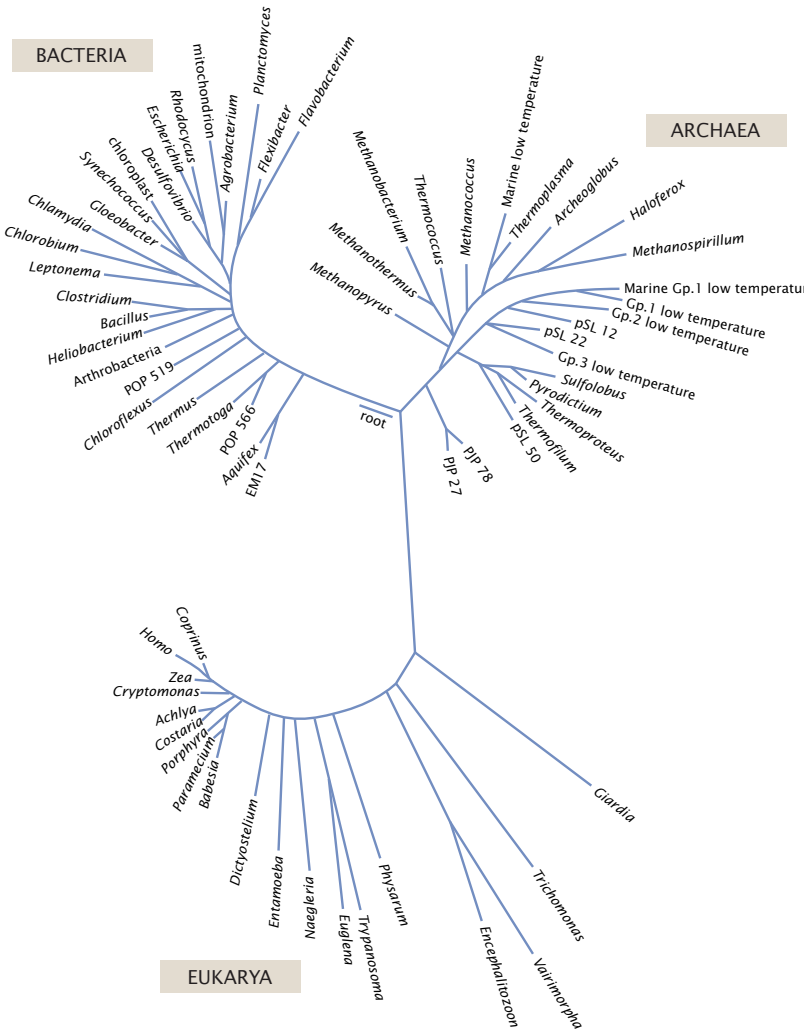


Figure 3.6: Universal phylogenetic tree. This diagram shows the similarity among ribosomal RNA sequences for representative organisms from all major branches of life on Earth. The total length of the lines connecting any two organisms indicates the magnitude of the difference in their ribosomal RNA sequences.

of DNA. Thus different organisms contain different complements of genes as well as sequence differences between homologous copies of the same gene. The term "homologous" refers to descent from a common ancestor. For example, ribosomal RNAs are homologous in all cells. In Chapter 21, we will give some examples of ribosomal RNA sequences and show how they can be used to build a universal phylogenetic tree (see Figure 21.32 on p. 1009). One example of a tree based on ribosomal RNA sequences that attempts to demonstrate the relatedness among all branches of existing life is shown in Figure 3.6.

Phylogenetic trees established by molecular methods tend to be in excellent agreement with analogous trees of similarity based on morphological or biochemical criteria as have been established by botanists, zoologists, and microbiologists over the past several hundred years. We will examine statistical methods for constructing such trees in Chapter 21.

What does any of this have to do with the determination of evolutionary time? In the laboratory, we can observe that certain types of changes in DNA sequences within a population happen frequently (for example, single-point mutations changing a C to a T), while others happen more rarely (cross-over events reversing the order of all the genes within a segment of a chromosome). We can even measure the time constants that characterize such events. If we assume that these kinds of mutational events happen with the same frequency in

wild populations as they do in the laboratory, then we can estimate divergence times for organismal populations based on calculating how long on average it would take to achieve the observed number of sequence alterations given known rates of sequence alteration events. In a few cases, these time estimates can be anchored by reference to the fossil record. In reality, inferring evolutionary time from sequence similarity is fraught with peril because not all sequence alterations are equally likely to be randomly incorporated into the genetic heritage of a population of organisms. Some mutations will prove to be unfavorable for a given organism's lifestyle and individuals carrying those mutations will be eliminated from the population by natural selection. Other mutations will prove to be advantageous and organisms carrying those mutations will quickly outcompete other members of their species. These selection effects can make the sequence-determined evolutionary clock appear to run too slow or too fast. Biologists face challenges similar to those faced by astronomers. In the astronomical setting, continual refinements in cosmological distance scales based on various types of standard candle (light sources of known absolute intensity) have led to increasingly refined measurements of astronomical distance. Similarly, biologists have a number of different standard stopwatches that can be used to calibrate the flow of evolutionary time.

3.1.3 The Cell Cycle and the Standard Clock

The *E. coli* Cell Cycle Will Serve as Our Standard Stopwatch

In Figure 2.1 (p. 37) we used the size of an *E. coli* cell as our standard measuring stick. Similarly, we now invoke the time scale of the *E. coli* cell cycle as our standard stopwatch. The goal of Figure 3.2 was to illustrate the variety of different processes that occur in cell biology and the time scales over which they are operative. As with our discussion of structural hierarchies, we once again use the trick of invoking *E. coli* as our reference, this time with the several thousand seconds of its cell cycle as our reference time scale.

As shown in Figure 3.2(C), the bacterial cell cycle will be defined as the time between the "birth" of a given cell resulting from division of a parental cell to the time of its own subsequent division. This cell cycle is characterized structurally by the segregation of the duplicated bacterial chromosome into two separate clumps and the construction of a new portion of the cell wall, or septum, that separates the original cell into two daughters. These processes are illustrated in more detail in Figure 3.7. Because *E. coli* is a roughly cylindrical cell that maintains a nearly constant cross-sectional area as it grows longer, the total cell volume can be easily estimated simply from measuring the length, and this also provides a guide as to the point in the cell cycle. As cell division proceeds, *E. coli* doubles in length and hence also doubles in volume. The time scale associated with the binary fission process of interest here is of the order of an hour (to within a factor of 2), though division can take place in under 20 minutes under optimal growth conditions.

In the previous chapter, we argued that having a proper molecular inventory of a cell is a prerequisite to building models of many problems of biological interest. Here we argue that a similar "feeling for the numbers" is needed concerning biological time scales. How long does

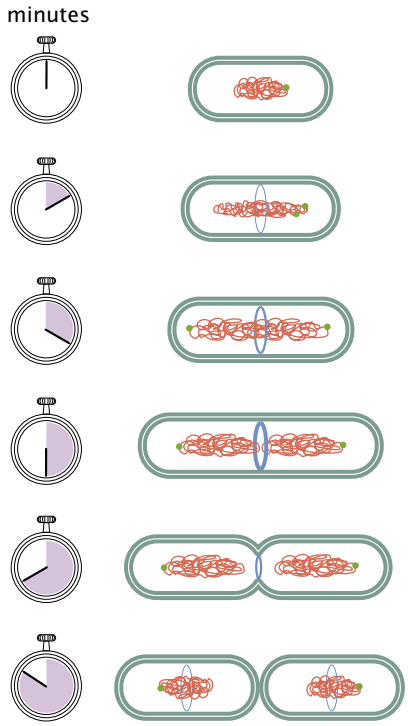


Figure 3.7: Schematic of an idealized bacterial cell cycle. A newborn cell shown at the top has a single chromosome with a single origin of replication marked by the green dot. The cell cycle initiates with the duplication of the origin, and DNA replication then proceeds in an orderly fashion around the circular chromosome. At the same time, a group of cell division proteins beginning with the tubulin analog FtsZ form a ring at the center of the cell that will dictate the future site of septum formation. As DNA replication proceeds and the cell elongates, the two origins become separated from each other, with one traveling the entire length of the cell to take up residence at the opposite pole. As the septum begins to close down, the two chromosomal masses are physically separated into the two daughter cells, where the cycle can begin anew.

it take for an *E. coli* cell to copy its genome and is this rate consistent with the speed of the molecular machine (DNA polymerase) that does this copying? On what time scale do newly formed proteins in neurons reach the ends of their axons and can this be explained by diffusion? Often, the time scale associated with a given process will provide a clue about what physical mechanisms are in play. In addition, one of our biggest concerns in coming chapters will be to figure out under what conditions we are justified in using ideas from equilibrium physics (as opposed to nonequilibrium physics). The answer to this question will be determined by whether or not there is a separation of time scales, and the only way we can know that is by having a feeling for what time scales are operative in a given problem. To that end, we begin by taking stock of the processes that an *E. coli* cell must undergo to copy itself.

For estimates in this book, we will choose a standard for bacterial growth in a minimal defined medium with glucose as the sole carbon source. As mentioned previously, the rate of cell division can vary more than 10-fold depending upon nutrient availability and temperature, so we must define the terms under which we will proceed with our estimates. The choice of minimal medium with glucose at 37 °C is a practical one since many quantitative experiments have been performed under these conditions. With sufficient aeration, *E. coli* in this medium typically double in 40–50 minutes, and we will use 3000 s as our canonical cell cycle time. In general, time scales for biological processes are much more variable than spatial scales, although it is true that rapidly growing *E. coli* are slightly larger than slowly growing *E. coli*. The difference in size for cells growing under different conditions may be an order of magnitude less than the difference in cycle time.

Computational Exploration: Timing *E. coli* We noted above that the division time for *E. coli* has a characteristic time scale of a few thousand seconds. In this Computational Exploration, we make a crude estimate of the mean division time by using a time lapse video of dividing bacteria like that shown in Figure 3.8. In such a movie, a microscope with a 100× objective is set up to take a snapshot roughly every 10 minutes, though here we only show images from every 30 minutes. When viewed one after the other, these movies reveal that the cells undergo a succession of growth and division processes.

COMPUTATIONAL EXPLORATION

The idea in this exercise is to write a Matlab script that examines the images and determines the total area of the cells as a function of time. The phase contrast microscopy used to make this video is an imaging technique that makes it possible to clearly visualize objects like *E. coli* that would otherwise be nearly transparent when illuminated directly by white light. The first step is to find the cells in some automated fashion, a process known as segmentation, though in the simple algorithm described here we fall short of the identification of individual cells. There are many different techniques for carrying out the segmentation process. One of the simplest ideas is to use thresholding, in which all intensities below (or above) a certain threshold are assigned a value of zero (or one). One of the problems with this approach is that it is indiscriminate in the sense that objects other than cells that have different intensities than the background will be detected by the thresholding. To eliminate these false positives, we could

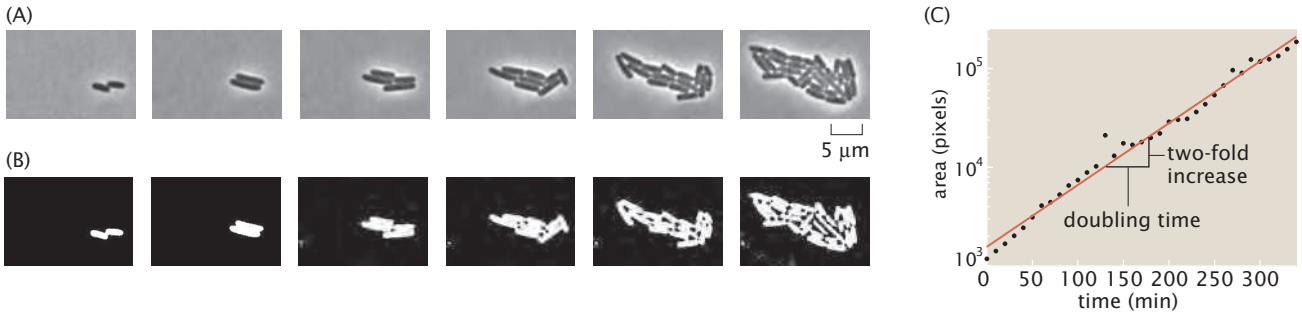


Figure 3.8: Time lapse movie of dividing *E. coli* cells. (A) Phase contrast snapshots taken 30 minutes apart showing dividing cells. (B) Thresholded versions of the same images shown in (A). (C) Plot of the area of the cells as a function of time in units of pixels. A division time of roughly 45 minutes can be read off the graph directly, which jibes with the fact that these cells were grown on agar pads with minimal media. The full set of images for this movie needed for the Computational Exploration can be found on the book's website.

go further, for example by accepting only those objects whose area falls between some upper and lower bound and with an aspect ratio that is appropriate.

For the reader who has not had much experience with image analysis, the unsophisticated treatment recommended here is intended to develop intuition in segmentation techniques such as thresholding. To be concrete, as shown in Figure 3.8(A), we have a series of images of dividing cells. We suggest that the reader start with the first image and read it into Matlab. Then, rescale the values in the image from 0 to 1. Make a plot of this new rescaled image. The next step is to set all pixels lower than some threshold value (for example, 0.4) to the value of 1 and all pixels with intensity higher than this value to 0. This algorithm is shown in schematic form in Figure 3.9. The outcome of this procedure is a series of new images like those shown in Figure 3.8(B). The reader can then compare different choices of the threshold used to see how this choice works both for this image and others from the image sequence. With a satisfactory choice of threshold, compute the area of the "cells" from each image by summing over the values on all of the pixels in the image and make a plot of this area as shown in Figure 3.8(C).

If we assume that the cells are doubling steadily, this implies that the number of cells as a function of time can be written as $N(t) = N_0 e^{kt}$. On the assumption that all cells have the same area, by multiplying both sides by the area per cell A_0, we can rewrite this equation as $A(t) = A_0 e^{kt}$. The time constant for doubling can be determined by noting that $2N_0 = N_0 e^{kt_{\text{double}}}$, which results in $t_{\text{double}} = (1/k) \ln 2$. The rate constant, in turn, can be read off of a graph like that shown in Figure 3.8(C).

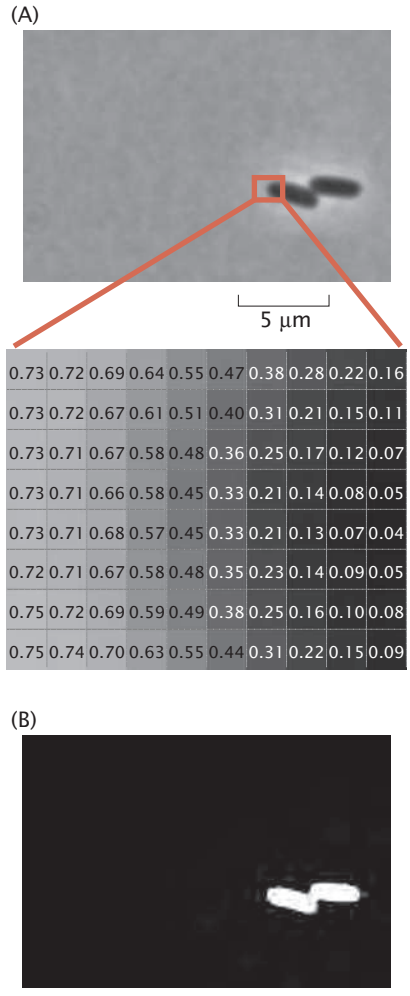


Figure 3.9: Segmentation of cells using thresholding. (A) A phase contrast image of part of a field of dividing cells. The figure below zooms in on a small region of the image and shows the intensity values of the image for all pixels within that region. (B) The result of segmenting the image with a threshold value of 0.4.

Estimate: Timing *E. coli* In Section 2.1.2 (p. 38), we sized up *E. coli* by giving a series of rough estimates for its parts list. We now borrow those estimates to gain an impression of the rates of various processes in the *E. coli* cell cycle. The simple idea behind these estimates is to take the total quantity of material that must be used to make a new cell and to divide by the time ($\approx$3000 s) of the cell cycle. When *E. coli* is grown on minimal medium with glucose as the sole carbon source, six atoms of carbon are added to the cellular inventory for each molecule of

ESTIMATE

glucose taken up. In the previous chapter, we estimated that the number of carbon atoms it takes to double the material in a cell so that it can divide in two (just the construction material) is of the order of 10^{10}. For this estimate, we ignored the material released as waste products and the reader will have the opportunity to estimate this contribution in the problems at the end of the chapter. We are also deliberately ignoring the glucose molecules that must be consumed to generate energy for the synthesis reactions—this topic will be taken up in Chapter 5. At this point, we can estimate the rate of sugar uptake required simply to deliver the 10^{10} carbon molecules necessary for building the material of the new cell: 10^{10} carbons must be captured over 3000 s, with 6 carbons per glucose molecule, giving an average rate of roughly 5×10^5 glucose molecules every second.

Of course, having the carbon present is merely a prerequisite for the macromolecular synthesis required to make a new cell. One of the most important processes in the cell cycle is replication. Given that the complete *E. coli* genome is about 5×10^6 bp in size, we can estimate the required rate of replication as

$$\frac{\mathrm{d}N_{\mathrm{bp}}}{\mathrm{d}t} \approx \frac{N_{\mathrm{bp}}}{\tau_{\mathrm{cell}}} \approx \frac{5 \times 10^6\,\mathrm{bp}}{3000\,\mathrm{s}} \approx 2000\,\mathrm{bp/s}. \tag{3.1}$$

Note that we have rounded to the nearest thousand.

Similarly, the rate of protein synthesis can be estimated by recalling from the previous chapter that the total number of proteins in *E. coli* is roughly 3×10^6, implying a protein synthesis rate of

$$\frac{\mathrm{d}N_{\mathrm{protein}}}{\mathrm{d}t} \approx \frac{N_{\mathrm{protein}}}{\tau_{\mathrm{cell}}} \approx \frac{3 \times 10^6\ \mathrm{proteins}}{3000\,\mathrm{s}} \approx 1000\,\mathrm{proteins/s}. \tag{3.2}$$

A similar estimate can be performed for the rate of lipid synthesis, resulting in

$$\frac{\mathrm{d}N_{\mathrm{lipid}}}{\mathrm{d}t} \approx \frac{N_{\mathrm{lipid}}}{\tau_{\mathrm{cell}}} \approx \frac{5 \times 10^7\ \mathrm{lipids}}{3000\,\mathrm{s}} \approx 20{,}000\,\mathrm{lipids/s}. \tag{3.3}$$

Yet another intriguing aspect of the mass budget associated with the cell cycle is the control of the water content within the cell. Recalling our estimate from the previous chapter that an *E. coli* cell has roughly 2×10^{10} water molecules results in the estimate that the rate of water uptake during the cell cycle is

$$\frac{\mathrm{d}N_{\mathrm{H_2O}}}{\mathrm{d}t} \approx \frac{N_{\mathrm{H_2O}}}{\tau_{\mathrm{cell}}} \approx \frac{2 \times 10^{10}\ \mathrm{waters}}{3000\,\mathrm{s}} \approx 7 \times 10^6\,\mathrm{waters/s}. \tag{3.4}$$

This rate of water uptake can be considered slightly differently by working out the average mass flux across the cell membrane. The flux is defined as the amount of mass crossing the membrane per unit area and per unit time, and in this instance is given by

$$j_{\mathrm{water}} \approx \frac{\mathrm{d}N_{\mathrm{H_2O}}/\mathrm{d}t}{A_{E.coli}} \approx \frac{7 \times 10^6\ \mathrm{waters/s}}{6 \times 10^6\,\mathrm{nm}^2} \approx 1\ \mathrm{water/(nm^2 s)}, \tag{3.5}$$

though we also note that this mass transport is mediated primarily by proteins that are distributed throughout the membrane.

Each of these estimates tells us something about the nature of the machinery that mediates the processes of the cell. In the remaining sections, these estimates will serve as our jumping-off point for estimating the rate at which individual molecular machines carry out the processes of synthesis and transport needed to support metabolism and the cell cycle.

COMPUTATIONAL EXPLORATION

Computational Exploration: Growth Curves and the Logistic Equation One of the consequences of all of this busy molecular synthesis and cell division is that cells grow and divide, grow and divide, over and over again. After less than 24 hours, a 5 mL culture that starts out with just one cell will be populated with roughly 5×10^9 cells. Of course, such growth cannot go on unchecked indefinitely. In the remainder of the book, we will repeatedly have occasion to ask for a detailed accounting of the trajectories of some quantity that is changing over time, such as the number of cells in our culture, $N(t)$ described above. In many cases, we will be able to resort to an analytic treatment of the differential equations that describe these trajectories. However, there will also be times where the differential equation of interest is beyond our means to attack analytically and we will have to resort to numerical methods. To see how the numerical approach to these problems unfolds, we consider bacterial growth.

The study of growth curves was once referred to by Jacques Monod as "the basic method of microbiology." We have already seen an example of such a growth curve in Figure 3.3, which shows data from one of the original Monod experiments that resulted in the first understanding of transcriptional regulation. This fervor for the study of bacterial growth was articulated even more forcefully and amusingly by Frederick Neidhardt in his article aptly entitled "Bacterial growth: constant obsession with $\mathrm{d}N/\mathrm{d}t$" where he reports that "One of life's inevitable disappointments—one felt often by scientists and artists, but not only by them—comes from expecting others to share the particularities of one's own sense of awe and wonder. This truth came home to me recently when I picked up Michael Guillens fine book 'Five Equations That Changed the World' and discovered that my equation—the one that shaped my scientific career—was not considered one of the five."

The simplest version of Neidhardt's life-changing equation is

$$\frac{\mathrm{d}N}{\mathrm{d}t} = rN. \tag{3.6}$$

This equation has far-reaching consequences since it shows how some quantity grows over time and has the familiar solution

$$N(t) = N_0 \mathrm{e}^{rt} \tag{3.7}$$

that signals exponential growth. With the inclusion of a minus sign in front of the right-hand side, this same equation

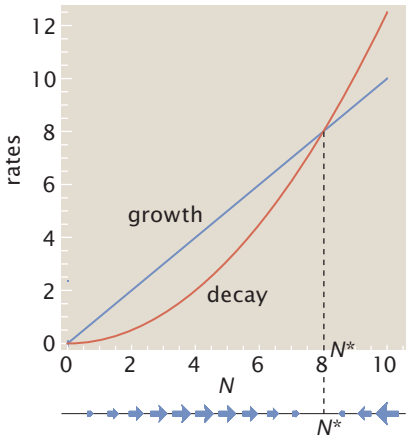


Figure 3.10: Rates as a function of number of cells. The blue curve shows how the growth rate depends upon the number of cells. The red curve shows the negative value of the decay rate. The vectors under the figure show the net growth rate. The length and size of these vectors is proportional to the rate and their direction signifies whether there is net growth or decay in the population. N^* is the fixed point of this simple dynamical system.

captures problems ranging from radioactive decay to the degradation of messenger RNAs in the cellular interior.

Of course, as noted above, despite the best efforts of humans with both population and spending, such growth cannot go unchecked. One of the simplest models that respects the fact that such growth cannot be sustained is the so-called logistic equation given by

$$\frac{\mathrm{d}N}{\mathrm{d}t} = rN\left(1 - \frac{N}{K}\right). \tag{3.8}$$

In this equation, K represents the maximum population that can be sustained by the environment. When N is very small compared with K, the equation shows nearly the same exponential behavior as the simple growth described by Equation 3.6. However, as N gets larger, the growth rate slows down, until when $N = K$ (that is, the population has reached the maximum carrying capacity of the environment), $\mathrm{d}N/\mathrm{d}t$ falls to 0. An alternative way of looking at Equation 3.8 is to think of the cellular birth and death processes independently. This can be depicted graphically as shown in Figure 3.10. In particular, note that there is a growth rate characterized by the term rN and a death rate characterized by $-rN^2/K$. Each of these rates depends on N with different functional forms as shown in Figure 3.10. Further, the fixed point of the system is revealed by the vectors plotted beneath the graph, which show a simple one-dimensional phase portrait that tells us whether the *net* growth will be positive or negative as a function of the population size. These phase portraits allow us to run our dictum of the necessity of mathematicizing the cartoons in reverse by making it possible to cartoonize the mathematics in a clear and simple way that reveals the key features of the dynamical system. We will use such phase portraits to cartoonize the mathematics throughout this book.

Though the logistic equation given in Equation 3.8 is amenable to analytic resolution with the solution given by

$$N(t) = \frac{KN_0\mathrm{e}^{rt}}{K + N_0\left(\mathrm{e}^{rt} - 1\right)}, \tag{3.9}$$

we will use it as the basis for our illustration of the simplest numerical method for integrating such equations.

Both of the differential equations written above for the time evolution of our growing bacterial culture are special cases of the more general form

$$\frac{\mathrm{d}N}{\mathrm{d}t} = f(N, t). \tag{3.10}$$

In this case, the function $f(N, t)$ is the growth rate and tells us how the rate depends upon the current population size. To integrate this equation, we note that we can rewrite it in terms of the discrete approximation to the derivative as

$$\frac{N(t + \Delta t) - N(t)}{\Delta t} \approx f(N, t). \tag{3.11}$$

Using this definition, we can rearrange the terms to find

$$N(t + \Delta t) = N(t) + f(N, t)\Delta t. \tag{3.12}$$

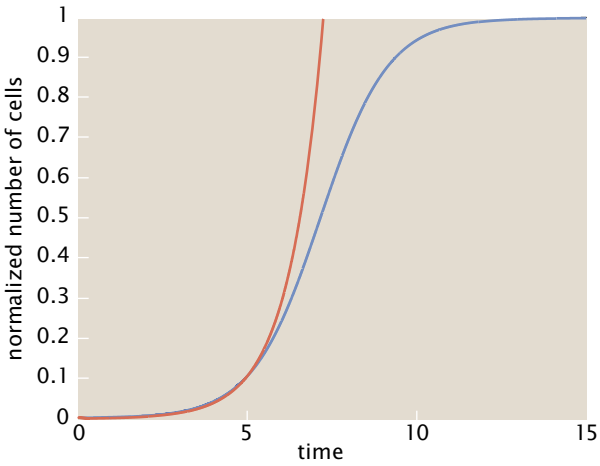


Figure 3.11: Integration of differential equations for bacterial growth. Using the simple algorithm described in the Computational Exploration, we integrate both the equation for exponential growth and the logistic equation "by hand." The red line shows the result of integration of Equation 3.6 for the case in which $r = 1$. The blue line shows the result of integration of Equation 3.8 for the case in which $r = 1$ and $K = 1$.

This formula can be easily implemented in a number of different ways within Matlab, but perhaps the easiest scheme is simply to construct a loop that repeatedly takes the current population and increments it to produce the population size a time step Δt later. The result of integrating both of the equations given above using this integration scheme is shown in Figure 3.11.

The Matlab code included on the book's website provides two versions of this exercise. One of these versions illustrates the numerical integration "by hand" described above and the other shows how to use the built-in function within Matlab for integrating differential equations. Generally, rather than writing the integration routine itself, it is often easier to exploit such built-in functions.

3.1.4 Three Views of Time in Biology

Modern humans have built much of the activity of our societies around an obsession with absolute time. This obsession is revealed by the propensity for events to occur at a certain time of day, for example, class starts at 9 a.m., or scheduling our activity by measured blocks of time, for example, you must practice the piano for half an hour. It is not clear, however, that other organisms relate to time in this manner. In the remainder of this chapter, we will discuss three different views of time that seem to be important to life, and we will term them *procedural time*, *relative time*, and *manipulated time*.

In the previous chapter, we explored the question of why biological things are a certain size, and the ultimate reason is the finite extent of the atoms that make up biological molecules. Here we are trying to understand why biological processes take a certain amount of time, a difficult task. For the most part, the size of things does not strongly depend on environment and external conditions, but the time scale of processes often does. For example, bacteria growing in leftover potato salad will replicate rapidly when the salad is left on a picnic table in full sun but much more slowly in a refrigerator. The fundamental reason for the difference in replication rates as a function of temperature can be attributed to the slowing or the acceleration of the many individual enzymatic steps that must take place for the cell to double in size and divide. In this sort of context, it appears that organisms pay attention to *procedural time* (an idea to be fleshed out

in Section 3.2) rather than absolute time: they do something for as long as it takes to get it done since there is some procedure such as DNA replication dictated by an enzymatic rate. A particularly interesting class of procedural time mechanisms are those that organisms use to build clocks that are extremely good at keeping track of absolute time without regard to perturbation by external conditions. One fascinating example of this that we will explore in more detail later in the chapter is the diurnal clock that enables an organism to perform different acts at different times of the day, even in the absence of external signals such as the rising and setting of the sun. For these clocks to work, organisms must have a way to convert procedural time into absolute time so as to ignore external conditions, including temperature.

Although calculating procedural time for a process of interest can often put a lower limit on how fast that process can occur, cells often seem to put as much effort into making sure that processes occur in the correct order as in making sure that they occur quickly. In the context of cell division, for example, it would be problematic for a cell to try to segregate its chromosomes into the two daughters until the process of DNA replication is complete. The result would be that at least one daughter would lack the full genetic complement of the mother cell. We will refer to processes where one must be complete before another can start under the category of *relative time* (that is, before or after rather than how long). This topic will be explored in Section 3.3.

Third, and perhaps most interestingly, it appears that living organisms are rarely content to accept time as it is. In some cases, they seem to be impatient, demanding that their life processes occur more quickly than permitted by the underlying chemical and physical mechanisms. Rate acceleration by enzyme catalysis is a prime example. In other cases, they seem to delay the intrinsic proceeding of events, freezing time in "suspended animation" as in formation of bacterial spores that can survive for hundreds or thousands of years, only to be reanimated when conditions become favorable. In Section 3.4, we will argue that these processes are examples of what we will refer to as *manipulated time*.

3.2 Procedural Time

The underlying idea of measurements of procedural time is simply that the chemical and physical transformations characteristic of life do not happen instantaneously. Complex processes can be thought of as being built up from many small steps, each of which takes a finite amount of time. For many biological processes that are intrinsically repetitive, such as the replication of DNA or the synthesis of proteins, the same step is used over and over again, namely, addition of single nucleotides to a growing daughter strand or addition of single amino acids to a growing polypeptide chain. In this section on procedural time, we will begin by making some estimates about these processes of the central dogma as an example of the general issues of computing procedural time for multistep biological processes. Then we will move on to the interesting special examples of clocks and oscillators where procedural times are calibrated so that cell cycles and diurnal cycles can follow the constant ticking of a reliable clock.

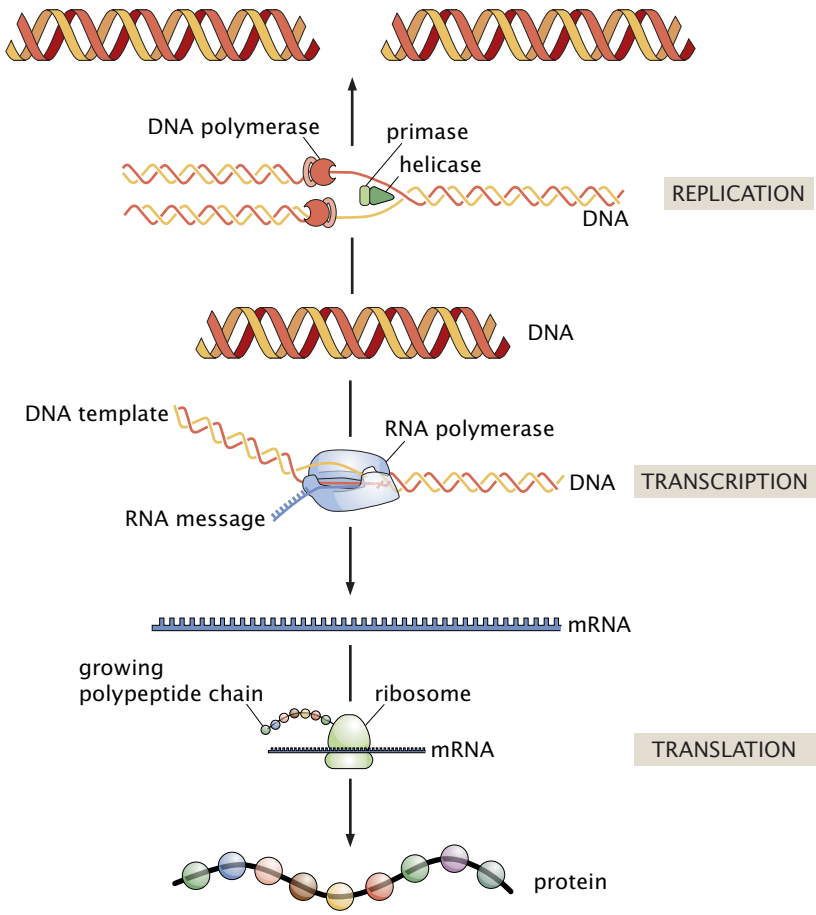


Figure 3.12: The processes of the central dogma. DNA is replicated to make a second copy of the genome. Transcription refers to the process in which RNA polymerase makes an mRNA molecule. Translation refers to the synthesis of a polypeptide chain whose sequence is dictated by the arrangement of nucleotides on mRNA.

3.2.1 The Machines (or Processes) of the Central Dogma

The Central Dogma Describes the Processes Whereby the Genetic Information Is Expressed Chemically

One of the most important classes of processes in cellular life is that associated with the so-called central dogma of molecular biology. The suite of processes associated with the central dogma are those related to the polymerization of the polymer chains that make up the nucleic acids and proteins that are at the heart of cellular life. The fundamental processes of replication, transcription, and translation and their linkages are shown in Figure 3.12. The basic message of this "dogma" in its least sophisticated form is that DNA leads to RNA, which leads to proteins. From the standpoint of cellular timing, the processes of the central dogma will serve as a prime example of procedural time. A typical circular bacterial genome, for example, is replicated by just two DNA polymerase complexes that take off in opposite directions from the origin of replication and each travels roughly half-way around the bacterial genome to meet on the opposite side. The time to replicate a bacterial genome is governed by the rate at which these polymerase motors travel to copy the roughly 5×10^6 bp of the bacterial genome. Similarly, the time to synthesize a new protein is governed by the rate of incorporation of amino acids by the ribosome.

The Processes of the Central Dogma Are Carried Out by Sophisticated Molecular Machines

One of the primary processes shown in Figure 3.12 is the copying of the genome, also known as replication. DNA replication must take place before a cell divides. As shown in Figure 3.12, the process of DNA replication is mediated by a macromolecular complex (the replisome), which has a variety of intricate parts such as the enzyme DNA polymerase, which incorporates new nucleotides onto the nascent DNA molecule, and helicases and primases that pry apart the DNA at the replication fork and prime the polymerization reaction, respectively.

The DNA molecule serves as a template in two different capacities. As described above, a given DNA molecule serves as a template for its own replication (DNA → DNA). However, in its second capacity as the carrier of the genetic material, a DNA molecule must also dictate the synthesis of proteins (the expression of its genes). The first stage in this process of gene expression is the synthesis of an mRNA molecule (DNA → RNA) with a nucleotide sequence complementary to the DNA strand from which it was copied, which will serve as the template for protein synthesis. This transcription process is carried out by a molecular machine called RNA polymerase that is shown schematically in Figure 3.12. In eukaryotes, transcription takes place in the nucleus while subsequent protein synthesis takes place in the cytoplasm, so there must be an intermediate step of mRNA export.

Once the messenger molecule (mRNA) has been synthesized, the translation process can begin in earnest (RNA → protein). As already described in Section 2.2.3 (p. 63), translation is mediated by one of the most fascinating macromolecular assemblies, namely, the ribosome. The ribosome is the apparatus that speaks both of the two great polymer languages and, in particular, forms a string of amino acids (a polypeptide chain) that are dictated by the codons (represented by three letters) on the mRNA molecule. The structure of the ribosome is indicated in cartoon form in Figure 3.12. As might be expected for a bilingual machine, the ribosome contains structural components of both RNA and protein. The two halves of the ribosome clamp an mRNA and then the ribosome moves processively down the length of the mRNA. As the ribosome moves along, successive triplets of nucleotides are brought into registry with active sites in the ribosomal machinery that align special RNA molecules (tRNA), charged with various amino acids, to recognize the complementary triplet codon. Subsequently, the ribosome catalyzes transfer of the correct amino acid from the tRNA onto a growing polypeptide chain and releases the now empty tRNA. This set of processes was indicated schematically in Figure 3.2(E). As shown in Figure 3.13, the nascent mRNA molecules in bacteria are immediately engaged by ribosomes so that protein translation can occur before transcription is even finished.

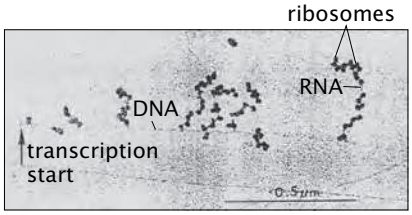


Figure 3.13: Electron microscopy image of simultaneous transcription and translation. The image shows bacterial DNA and its associated mRNA transcripts, each of which is occupied by ribosomes. (Adapted from O. L. Miller et al., *Science* 169:392, 1970.)

The timing of all three of these processes is dictated by the intrinsic rate at which these machines carry out their polymerization reactions. All of them can be thought of in the same framework as repetitions of N essentially identical reactions, each of which takes an average time Δt to perform (though clearly there are substantial fluctuations). We will now estimate total times for each of the three central processes of the central dogma.

Estimate: Timing the Machines of the Central Dogma

ESTIMATE

The estimates concerning the mass budget of dividing cells from Chapter 2 (p. 35) can be used as a springboard for contemplating the rates of the machines that mediate the processes of the central dogma. In our first estimate, we expand upon the estimate obtained earlier in the chapter of the rate at which the genome of an *E. coli* cell is copied, with the aim of learning more about the speed of the DNA replication complex. DNA replication in bacteria such as *E. coli* is undertaken by two replication complexes, which travel in opposite directions away from the origin of replication on the circular chromosome.

Given that the complete *E. coli* genome is about 5×10^6 bp in size and it is copied in the 3000 s of the cell cycle, we have already found that the rate of DNA synthesis is roughly 2000 bp/s, or 1000 bp/s per DNA replication complex (replisome) since there are two of these complexes moving along the DNA simultaneously. Biochemical studies have found rates for the DNA polymerase complex in the 250–1000 bp/s range. As we have mentioned, *E. coli* are actually capable of dividing in much less than 3000 s, in fact, in as little as 1000 s, although their DNA replication machinery cannot proceed any faster than this absolute speed limit. How do they achieve this? For now, we will leave this as an open mystery and will return to the question in the final section of the chapter on manipulating time.

For a bacterial cell, transcription involves the synthesis of mRNA molecules with a typical length of roughly 1000 bases. Our reasoning is that the typical protein has a length of 300 amino acids, with 3 bases needed to specify each such amino acid. Both bulk and single-molecule studies have revealed that a characteristic transcription rate is tens of nucleotides per second. Using 40 nucleotides/s, we estimate that the time to make a typical transcript is roughly 25 s.

Yet another process of great importance in the central dogma is protein synthesis by ribosomes. Recall from our estimates in the previous chapter that the number of proteins in a "typical" bacterial cell like *E. coli* is of the order of 3×10^6. This suggests, in turn, that there are of the order of 9×10^8 amino acids per *E. coli* cell, which are incorporated into new proteins over the roughly 3000 s of the cell cycle. We have made the assumption that each protein has 300 amino acids. This implies that the mean rate of amino acid incorporation per second is given by

$$\frac{dN_{aa}}{dt} \approx \frac{9 \times 10^8 \text{ amino acids}}{3000 \text{ s}} \approx 3 \times 10^5 \text{ amino acids/s}. \quad (3.13)$$

The number of ribosomes at work on synthesizing these new proteins is roughly 20,000, which implies that the rate per ribosome is 15 amino acids/s, while the measured value is 25 amino acids incorporated per second. These numbers also imply that the mean time to synthesize a typical protein is roughly 20 s.

One of our conclusions is that the rate of protein synthesis by the ribosome is slower than the rate of mRNA synthesis by RNA polymerase. However, as shown in Figure 3.13, multiple ribosomes can simultaneously translate a single mRNA by

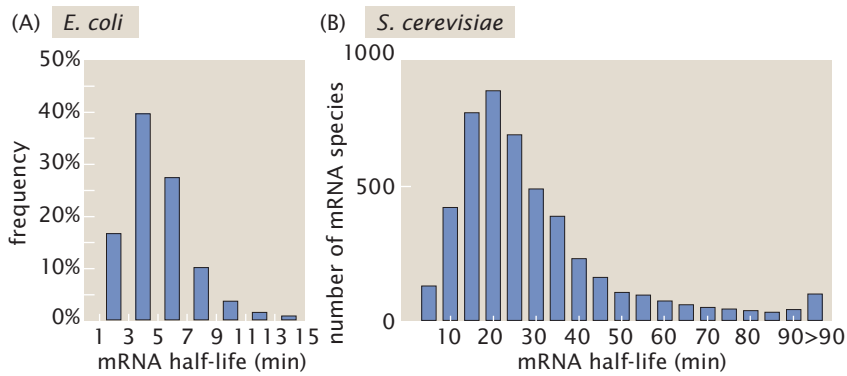


Figure 3.14: mRNA lifetimes. (A) Measurements of the lifetimes of thousands of different mRNA transcripts in *E. coli* using microarrays. The mean lifetime is slightly in excess of 5 minutes. (B) Measurements of the lifetimes of thousands of different mRNA transcripts in the yeast *S. cerevisiae* using microarrays. (A, adapted from J. A. Bernstein et al., *Proc. Natl Acad. Sci. USA* 99:9697, 2002; B, adapted from Y. Wang et al., *Proc. Natl Acad. Sci. USA* 99:5860, 2002.)

proceeding linearly in an orderly fashion and indeed, multiple RNA transcripts may exist in different degrees of completion, being transcribed from the same genetic locus. Thus, when considering the net rates of processes in cells, the number of molecular players is clearly as important as the intrinsic rate.

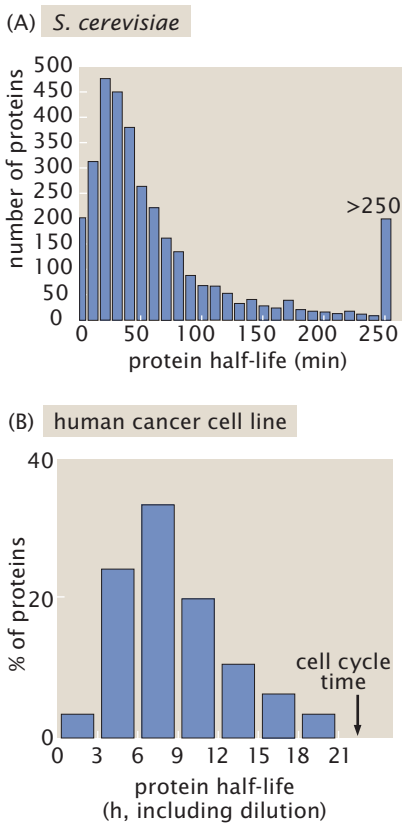


Figure 3.15: Protein lifetimes. Measurements of the lifetimes of thousands of different proteins. (A) Distribution of lifetimes of proteins in the yeast *S. cerevisiae*. (B) Distribution of lifetimes of proteins in human cells. (A, adapted from A. Belle et al., *Proc. Natl Acad. Sci. USA* 103:13004, 2006; B, adapted from E. Eden et al., *Science* 331:764, 2011.)

The estimates we have constructed for the rates at which the processes of the central dogma play are not the entire story of the overall census of mRNAs and their protein partners. Both mRNA transcripts and their downstream protein products are subjected to decay processes that take them out of circulation. To see this more explicitly, consider the results presented in Figure 3.14, which reflect measurements on literally thousands (that is, >2000) of different mRNA transcripts in *E. coli* and yeast. In both of the studies presented here, the factors determining mRNA lifetime were more subtle than some of the favored hypotheses such as a relation to secondary structure in the mRNA or the abundance of the gene products. One of the fascinating outcomes of the yeast study was the finding that mRNAs whose protein products are part of the same macromolecular assembly tend to have similar lifetimes.

Generally, the decay times for proteins are longer than those for mRNAs, as shown in Figure 3.15. Indeed, the half-lives are generally sufficiently long that the dominant effect in bacteria is protein elimination through cell division rather than by active degradation. That is, when cells divide, the average number of proteins per cell is reduced by a factor of two. In the case of eukaryotes like those shown in Figure 3.15, the lifetimes are between tens and hundreds of minutes. Degradation dynamics is explored further in the problems at the end of the chapter.

3.2.2 Clocks and Oscillators

In the context of the central dogma, we have described measurements of procedural time for processes that essentially happen once and run to completion, such as the synthesis of a protein molecule. However, many cellular processes run in repeated regular cycles. These cyclic or oscillatory processes frequently represent control systems where procedural times of some subprocess can be used to set the oscillation period. Two widely studied examples are the oscillators used to drive the cell division cycle and the mechanisms governing behavioral switches between daytime and nighttime, which will be explored in detail below. These daily clocks are called circadian or diurnal

oscillators. Other everyday oscillators run the beating of our hearts and the pattern of our breathing.

Developing Embryos Divide on a Regular Schedule Dictated by an Internal Clock

One of the best understood examples of an oscillatory clock used by cells is seen in the early embryonic cell cycle of many animals. The best-studied example is the South African clawed frog, *Xenopus laevis* (abbreviated *X. laevis*). After the giant egg (1 mm) is fertilized, a cell division cycle proceeds roughly every 20 minutes until the egg has been cleaved into approximately 4000 similarly sized cells. The regularity and synchrony of these cell divisions reflects an underlying oscillatory clock based on a clever manipulation of procedural time. The clock starts each cell division with the synthesis of a protein called cyclin. Cyclin is made from a relatively rare mRNA. As a result, the protein accumulates slowly. The biological function of cyclin is to activate a protein kinase, an enzyme that covalently attaches phosphate groups to amino acid side chains on other proteins. This process, known as phosphorylation, is one of the key ways that protein activity can be controlled after translation. Essentially, the protein is inactive in the absence of its phosphate group. Kinase activation cannot begin until the cyclin protein has accumulated to a certain threshold level. After the kinase is activated, one of its targets is an enzyme, which in turn catalyzes the destruction of the cyclin protein. All cyclin in the cell is quickly destroyed with a half-life of 90 s, resetting the clock to its zero position.

The regularity of this oscillatory clock depends upon several measurements of procedural time. First, accumulation of the cyclin protein to its threshold level depends upon the rate of ribosomal synthesis of that protein. Second, activation of the cyclin-dependent protein kinase kicks off a second procedural time measurement, which reflects the length of time required by the kinase to encounter and phosphorylate its enzymatic substrates. Third, the degradation of the cyclin protein also requires a fixed, but brief, amount of time. The sum of these three procedural times gives the total period of the clock. The outcome of these molecular events in terms of molecular concentrations is illustrated in Figure 3.16. Just as for all the examples of procedural time described above, the amount of absolute time in seconds, minutes, or hours may change depending upon external conditions such as temperature.

This cyclin-driven cell-cycle oscillator is one example of a very general category of two-component oscillatory systems found throughout biology. A simplified idealized representation of such an oscillator is shown in Figure 3.17(A), while a more accurate representation of the real cell-cycle control system from the yeast *S. cerevisae* is shown in Figure 3.17(B).

Diurnal Clocks Allow Cells and Organisms to Be on Time Everyday

A second example of the use of procedural time to build a clock is when cells arrange a series of molecular processes in such a way that they can measure an absolute time. Unlike the cell-cycle clock, it is critical that the diurnal clock does not change its period when the temperature changes, such as during the change of seasons. Many

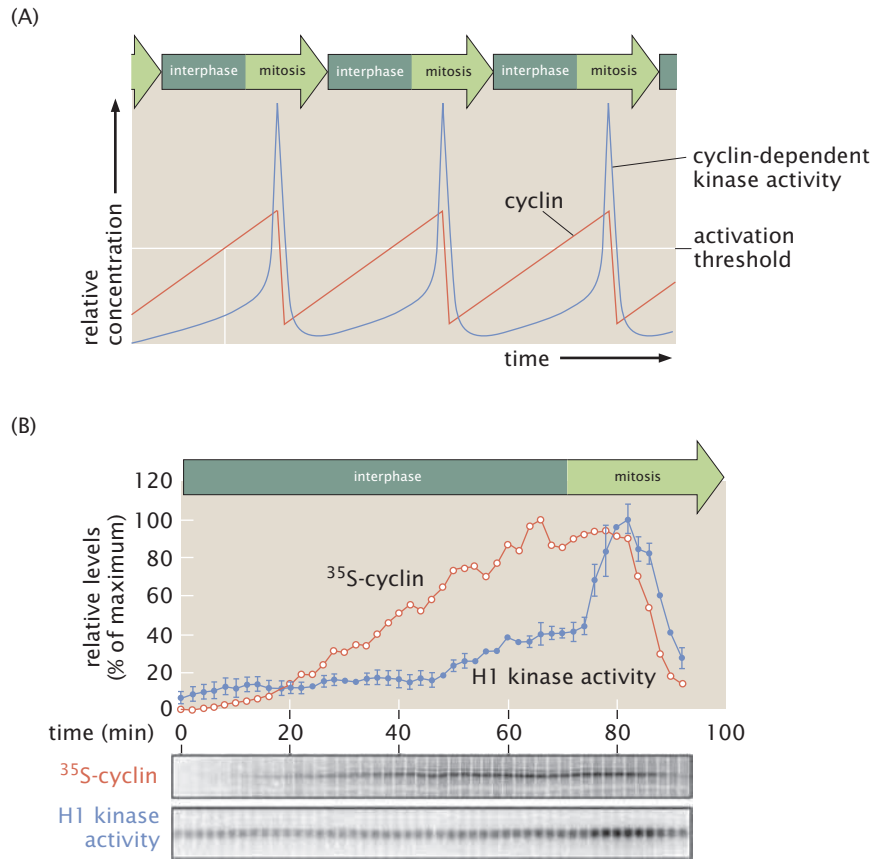
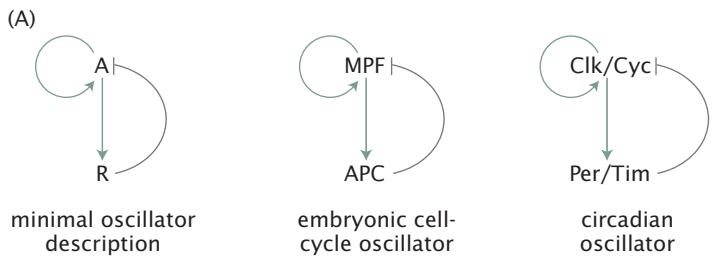


Figure 3.16: The oscillatory cell-cycle clock. This diagram shows the procedural events that underlie the regular oscillations of the cell-cycle clock in the *X. laevis* embryo. (A) Cyclin protein concentration rises slowly over time until it reaches a threshold at which point it activates cyclin-dependent kinase. Cyclin-dependent kinase activity increases sharply at this threshold and in turn activates enzymes involved in cyclin protein degradation. Once the degradation machinery is turned on, cyclin protein levels quickly fall back to zero. Cyclin-dependent kinase activity also falls and the degradation machinery inactivates. This oscillatory cycle is repeated many times. (B) Actual data for a single cycle of the oscillation. (Adapted from J. R. Pomerening et al., *Cell* 122:565, 2005.)



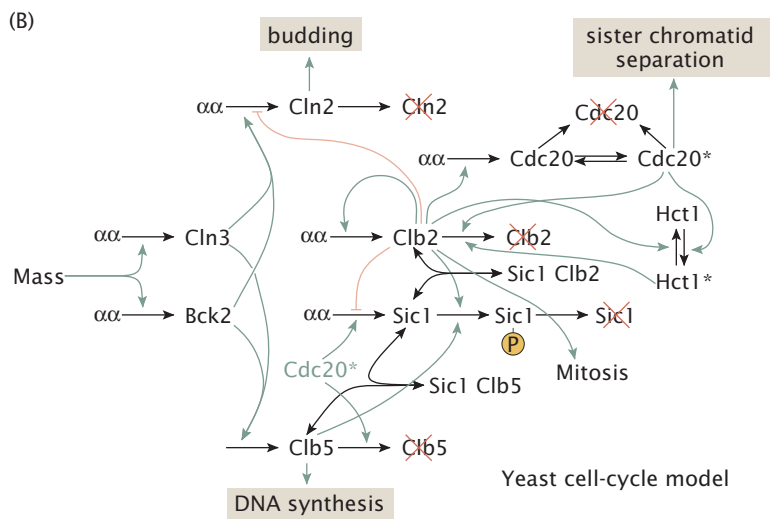


Figure 3.17: Logic diagrams for the construction of cell-cycle oscillators. (A) The minimal oscillator requires only two components. The first component activates the second component, for example by catalyzing its synthesis. The second component inhibits the first, for example by catalyzing its degradation. (B) A biochemically realistic representation of the cell-cycle oscillator in yeast is outrageously more complicated. This is because the real oscillator must work under a wide variety of conditions, be insensitive to fluctuations in the concentrations or activities of its components, and be subject to multiple kinds of regulatory inputs. (Adapted from N. W. Ingolia and A. W. Murray, *Curr. Biol.* 14:R771, 2004.)

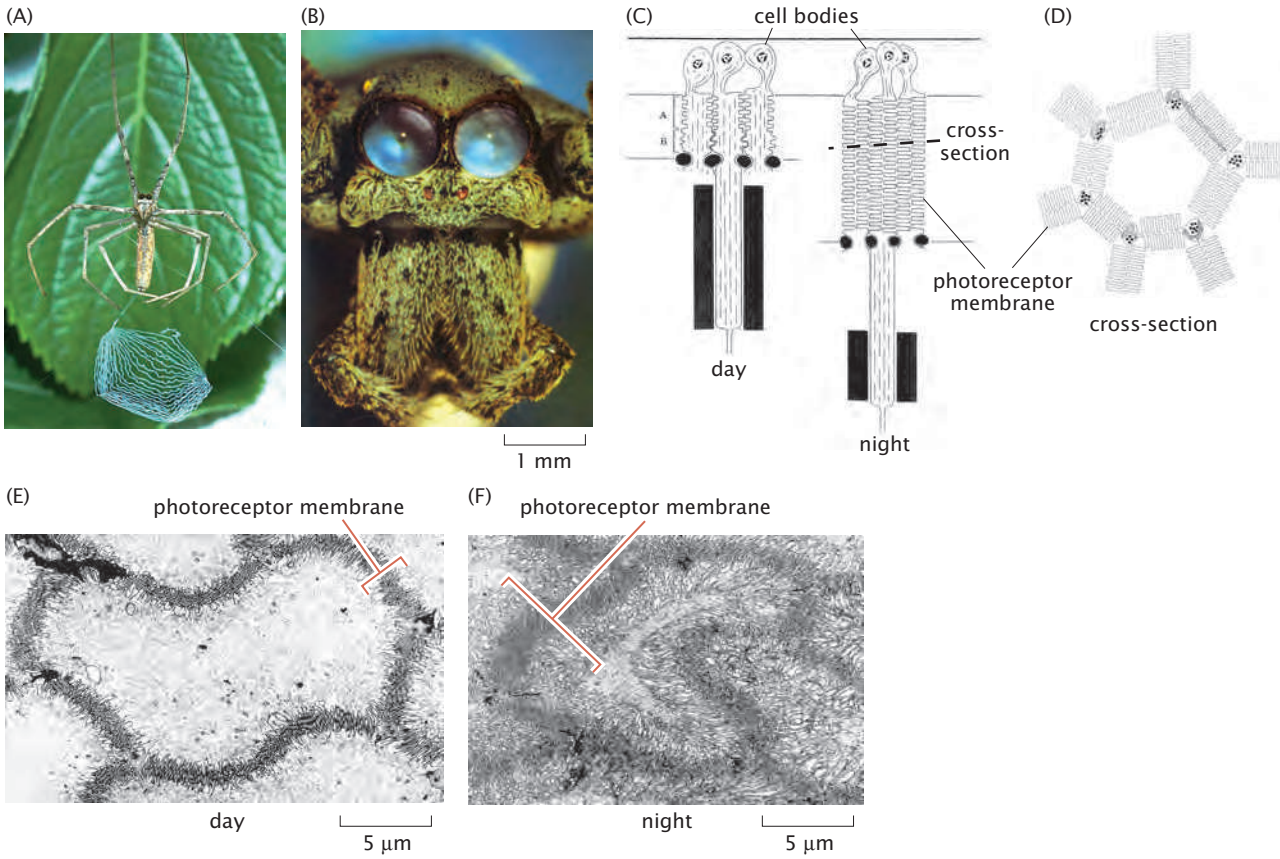


Figure 3.18: An extreme example of a structural change driven by the diurnal clock. (A) The net-casting spider, *Deinopis subrufa*, is a nocturnal hunter with an unusual strategy. It spins a small net, which it holds with its legs and tosses to entangle unwary passing prey. (B) In order to see the prey and know when to toss its net, the spider must have excellent night vision. Two of its eight eyes are extremely enlarged and exquisitely light-sensitive. (C) The light sensitivity of the spider's eyes changes by a factor of approximately 1000 between daytime and nighttime. During the day, the photoreceptor cell processes are short and fairly disorganized. At night, the total amount of membrane containing light sensors increases both by lengthening of the cells and by the construction of convoluted membrane folds, all packed with photoreceptor molecules. (D) In cross-section, the photosensitive membranes of neighboring cells abut each other, forming a regular tile-like pattern. (E) A cross-section through the photoreceptor cells of a spider sacrificed during the day shows relatively modest thickening of the boundary membranes. (F) An equivalent section taken from a spider sacrificed at night shows a vast increase in the number and size of the membrane folds. At dawn, these membranes will all be degraded, only to be resynthesized the following dusk. (A, courtesy of Mike Gray, Australian Museum; B, courtesy of The Smith-Kettlewell Eye Research Institute; C–F, adapted from A. D. Blest, *Proc. R. Soc. Lond.* B200:463, 1978.)

organisms perform some specific task at the same time everyday. A spectacular example is shown in Figure 3.18, where an animal alters the light sensitivity of its eyes in anticipation of sundown. While we might imagine that these kind of daily changes are triggered by, for example, the intensity of sunlight, it has been demonstrated for many organisms that they continue to perform their diurnal cycle even when kept in the dark. Direct observation of these cycles over long periods of time in cyanobacteria has demonstrated that they can operate with tight precision over weeks without any external cues about absolute time.

Different organisms use information about the time of day for vastly different purposes. Nevertheless, as illustrated in Figure 3.19, the molecular circuitry governing their circadian rhythms conserves certain common features. Generally, these systems include positive elements that activate transcription of so-called clock genes that drive rhythmic biological outputs as well as promoting the expression of

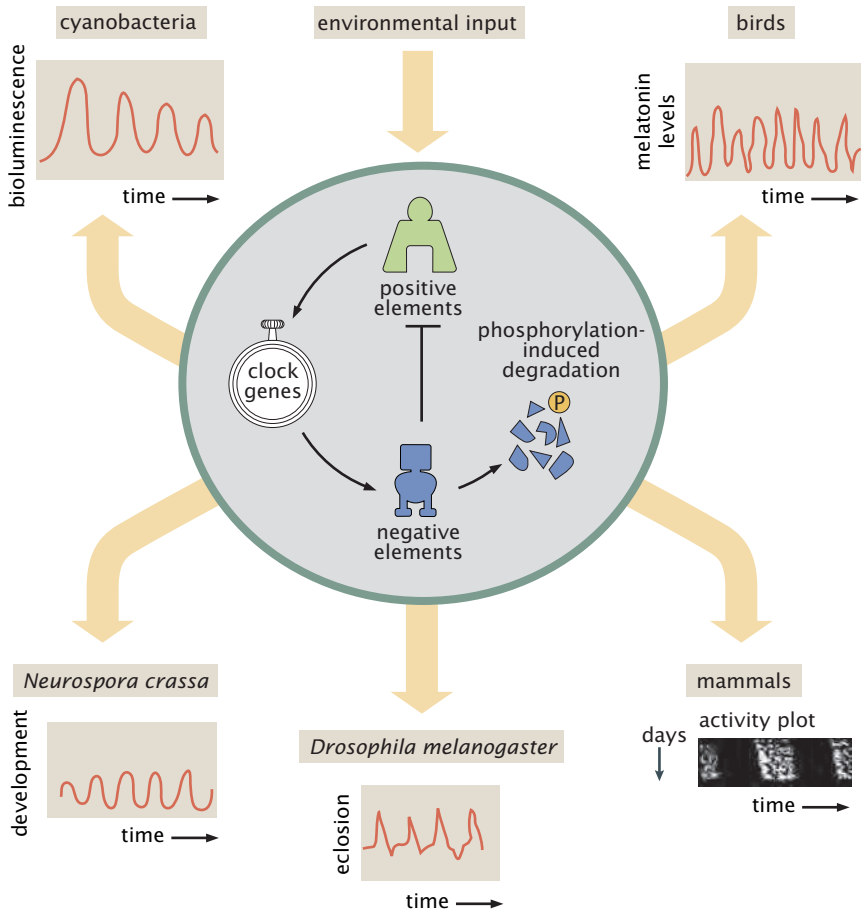


Figure 3.19: Schematic showing generic features of circadian clocks. Circadian clock mechanisms are autonomously driven oscillators that can be modulated by external inputs. Different organisms ranging from cyanobacteria to fungi to insects, birds, and mammals use their circadian timers to regulate different kinds of biological outputs and also use very different kinds of protein components in the internal circuitry. (Adapted from D. Bell-Pedersen et al., *Nat. Rev. Genet.* 6:544, 2005.)

negative elements that inhibit the activities of the positive elements. Phosphorylation of the negative elements leads to their degradation, allowing them to restart the cycle. Although the circadian oscillators are capable of continuing to measure time in constant light or constant darkness, they can nevertheless accept inputs from environmental signals such as the sun to reset their phase. Humans commonly experience the inefficiency of the phase-resetting mechanisms as the phenomenon of jetlag.

The circadian oscillator known to function with the fewest components is the one from the photosynthetic cyanobacterium *Synechococcus elongatus*. This organism's clock requires just three proteins, called KaiA, KaiB, and KaiC, and remarkably it appears that neither gene transcription nor protein degradation is necessary for this clock to function. A purified mixture of just these three proteins together with ATP is capable of sustaining an oscillatory cycle of KaiC protein phosphorylation over periods of at least several days. KaiC is able to catalyze both its own phosphorylation and its own dephosphorylation. KaiA enhances KaiC autophosphorylation and KaiB inhibits the effect of KaiA. Figure 3.20 shows the data supporting this remarkable finding.

3.3 Relative Time

The examples in the previous section on procedural time have emphasized the ways that cells set and measure the time that it takes to accomplish specific tasks. Some processes are rapid and others are

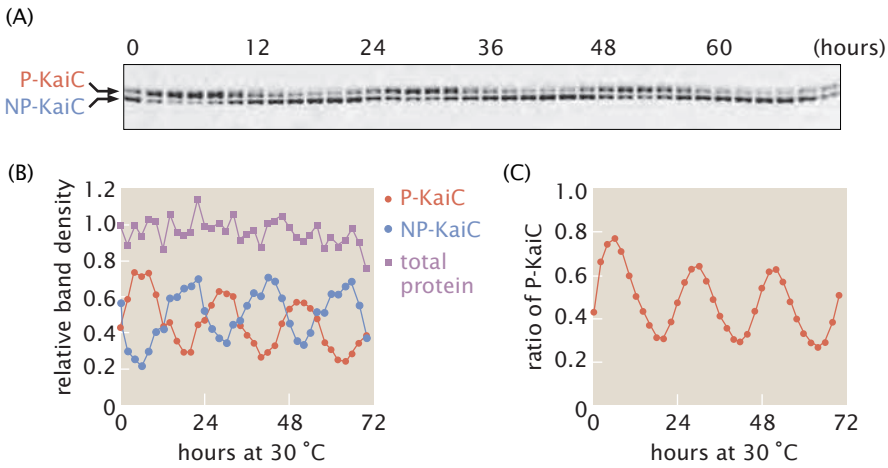


Figure 3.20: Reconstitution of the circadian oscillator. (A) In a mixture of purified KaiC protein together with KaiA, KaiB, and ATP, the molecular weight of KaiC can be seen to shift up and down slightly over a 24 h period. The upper band on this gel shows the phosphorylated form of KaiC (P-KaiC) protein and the lower band is the nonphosphorylated form (NP-KaiC). (B) Quantitation of the density of these two bands over time reveals that their concentration oscillates in a reciprocal manner such that the total amount of KaiC protein remains roughly constant. (C) The ratio of the amount of phosphorylated KaiC to total KaiC oscillates with a regular period of slightly under 24 h. (Adapted from M. Nakajima et al., *Science* 308:414, 2005.)

slow because of intrinsic features or environmental circumstances. In the well-regulated life of the cell, it is frequently important that fast and slow processes not be permitted to run independently of one another, but instead be linked in a logical sequence that depends upon the cell's needs. In this context, we now turn to what we will call *relative time*, which includes the governing mechanisms that ensure that related processes can be strung together in a "socks before shoes" fashion in which event A must be completed before event B can begin. Event C dutifully awaits the completion of event B before it begins, and so on.

3.3.1 Checkpoints and the Cell Cycle

In our initial discussion of the eukaryotic cell division cycle in the context of clocks and oscillators, we used the example of early embryonic divisions in the frog *X. laevis* and asserted that the underlying driver was a simple two-component oscillator. Once past the earliest stages of embryonic development, the cell cycle becomes much more complex and, in particular, becomes sensitive to feedback control from the cell's environment. The points in the cell cycle that are subject to interruption by external signals are referred to as checkpoints. These checkpoints ensure, for example, that chromosomes are not segregated until the DNA replication process is complete.

The Eukaryotic Cell Cycle Consists of Four Phases Involving Molecular Synthesis and Organization

Figure 3.21 shows the key features of the eukaryotic cell cycle, with an emphasis on the regulatory checkpoints that ensure that all processes will occur in the correct order. There is no single universal time scale for the eukaryotic cell cycle, which can vary greatly from one cell type to the next. In the human body, some cells in the intestinal lining can divide in as little as 10–12 hours, while others such as some tissue stem cells have cell cycles measured in days or weeks. The eukaryotic cell cycle is usually described in terms of four stages denoted as G_1, S, G_2, and M, where the M phase includes the most recognized features, namely, nuclear division (mitosis) and cell division (cytokinesis), the two G (gap) phases are periods of growth, and the S (synthesis) phase is the period during which the nuclear DNA is replicated. Together, the phases other than the M phase constitute

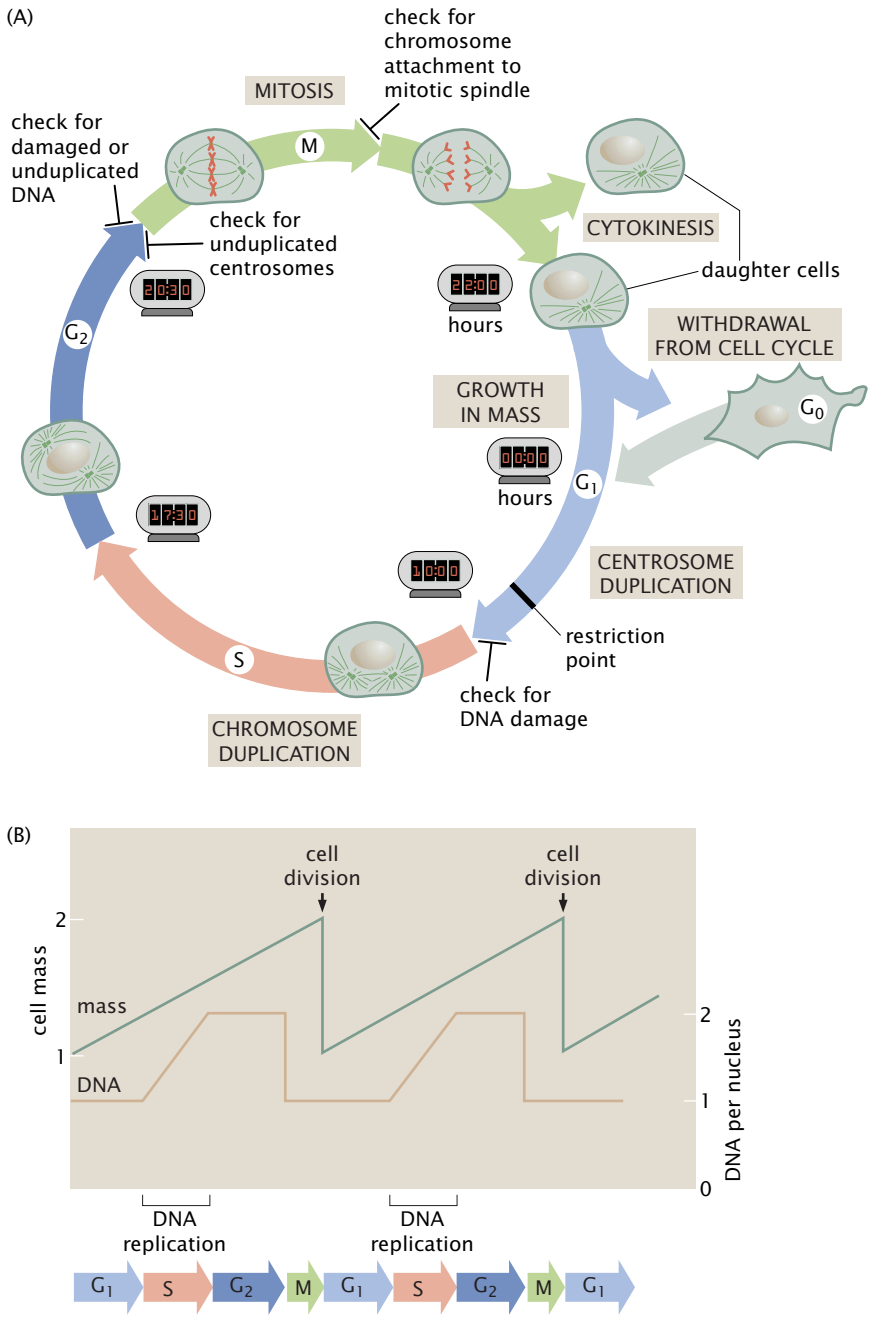


Figure 3.21: The eukaryotic cell cycle. (A) This cartoon shows some of the key elements of the process of cell division, including the four phases G_1, S, G_2, and M, as well as some of the most important checkpoints. Cells that stop proliferating can exit the cell cycle at G_1 and enter a resting phase, called G_0. Most fully differentiated cells in the adult human body are in G_0. Under particular circumstances, non-dividing cells in G_0 can reenter the cell cycle at G_1. (B) Time course of cell mass and DNA content during the cell cycle. Cell mass can increase continuously, while DNA content increases only during S phase. (A, adapted from T. D. Pollard and W. C. Earnshaw, Cell Biology, W. B. Saunders, 2007; B, adapted from A. Murray and T. Hunt, The Cell Cycle, Oxford University Press, 1993.)

interphase. During interphase, the mass content of the cell increases, as does its size.

If we use a cultured animal cell such as a fibroblast as our standard, G_1 is roughly 10 hours long, is characterized by a significant increase in cell mass, and culminates in a checkpoint to insure sufficient cell size and appropriate environmental conditions to pass to the next stage. Throughout most of G_1, the cell has only a single centrosome, forming the core of the microtubule organizing center. Near the end of G_1, the centrosome is duplicated so that the two daughter centrosomes can eventually form the two poles of the microtubule-rich mitotic spindle. At this point, the cell examines itself for DNA damage. Any signs of damage such as double-strand breaks will trigger a checkpoint control that prevents the cell from initiating DNA replication until the damage is completely repaired. This checkpoint also insures

a critical aspect of the regulation of relative time for the cell's replication, specifically, that it must have grown to approximately twice its prior size before it is allowed to begin to divide. If this checkpoint is successfully passed, then DNA replication can begin. In the S phase, the eukaryotic DNA is replicated over a period of about 6 hours.

Following the S phase and a shorter gap phase called G_2, another checkpoint verifies that every chromosome has been completely replicated and the centrosome has been properly duplicated before the initiation of the assembly of the mitotic spindle, the microtubule-based apparatus that physically separates the chromosomes into the two daughter cells. This enforcement of relative time is particularly critical because if a cell were to try to segregate the chromosomes before replication was complete, then at least one of the daughters would inherit an incomplete copy of the genome. After passing this checkpoint, the M phase begins. This relatively brief period of the cell cycle (of the order of 1 hour) involves most of the spectacular events of cell division that can be directly observed in a light microscope. Within the M phase, again it is critical that events occur in the proper order. The bipolar mitotic spindle built from microtubules forms symmetric attachments to each pair of replicated sister chromosomes. When they have all been attached to the spindle, the chromosomes all suddenly and simultaneously release their sisters and are pulled to opposite poles. A spindle-assembly checkpoint insures that every chromosome is properly attached before segregation begins. The molecular mechanisms governing the enforcement of relative time in the cell cycle involve protein phosphorylation and degradation events as well as gene transcription. In order to delve deeper into the principles governing the measurement and enforcement of relative time, we will now turn to a different example where gene transcription is the principal site of regulation.

3.3.2 Measuring Relative Time

There is great regulatory complexity involved in orchestrating the cell cycle. That is, the time ordering of the expression of different genes follows a complex program, with certain parts clearly following a progression in which some processes must await others before beginning. By measuring the pattern of gene expression, it becomes possible to explore the relative timing of events in the cell cycle. To get a better idea of how this might work, we need to examine how networks of genes are coupled together.

Genetic Networks Are Collections of Genes Whose Expression Is Interrelated

Sets of coupled genes are shown schematically in Figure 3.22. For simplicity, this diagram illustrates how the product of one gene can alter the expression of some other gene. Perhaps the simplest regulatory motif is direct negative control in which a specific protein binds to the promoter region on DNA of a particular gene and physically blocks binding of RNA polymerase and subsequent transcription. This protein is itself the result of some other gene, which can in turn be subject to control by yet other proteins (or perhaps the output of the gene that it controls). The second broad class of regulatory motifs is referred to as activation and results when a regulatory protein (a transcription

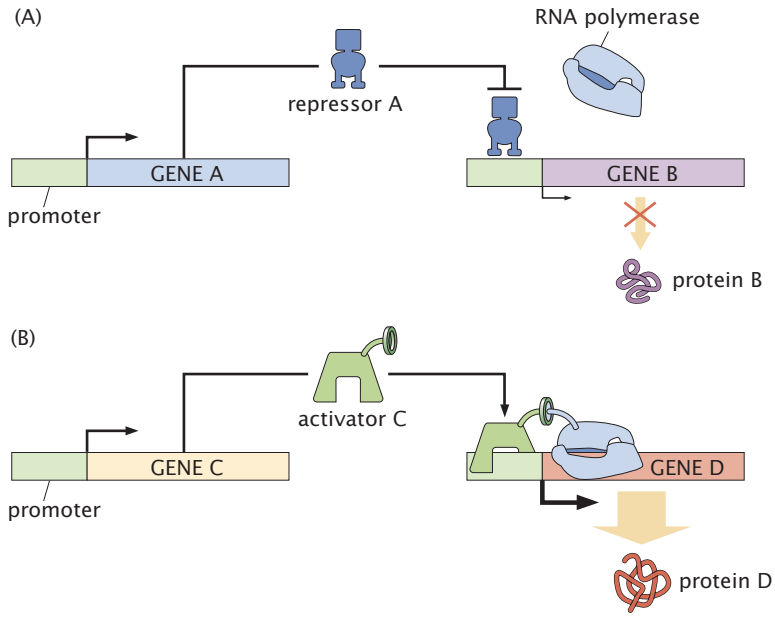


Figure 3.22: Simple networks of interacting genes. (A) Gene A encodes a repressor protein, which binds to the promoter region of gene B and prevents it from being transcribed by RNA polymerase. Only very small amounts of protein B are expressed. (B) Gene C encodes an activator protein, which helps to recruit RNA polymerase to the promoter of gene D. Large amounts of protein D are expressed.

factor called an activator) binds in the vicinity of the promoter and "recruits" RNA polymerase to its promoter.

One way to measure the extent of gene expression is to use a DNA microarray. The idea behind this technique is that a surface is decorated in an ordered *x*–*y* array with fragments of DNA, and the sequence of each spot on this array is different. To take a census of the current mRNA contents of a cell (which gives a snapshot of the current expression level of all genes), the cell is broken open and the mRNA contents are hybridized (that is, they bind to their complementary strand via base pairing) to the DNA on the microarray. DNA molecules on the surface that are complementary to RNA molecules in the cell lysate will hybridize with their complementary fragments, and those molecules that bind are labeled with fluorescent tags so that the fluorescence intensity can be used as a readout of the number of bound molecules. There is a bit more subtlety in the procedure than we have described, since really the mRNA is turned into DNA first, but we focus on the concept of the measurement rather than its practical implementation. The intensity of the spots on the microarray report the extent to which each gene of interest was expressed. By repeating this measurement again and again at different time points, it is possible to profile the state of gene expression for a host of interesting genes at different times in the cell cycle. These measurements yield a map of the *relative* timing of different genes.

One of the key model systems for examining the bacterial cell cycle is *Caulobacter crescentus*. In this system, DNA microarray analysis was used to examine the timing of various processes during the course of the cell cycle. In a beautiful set of experiments, roughly 20% of the *Caulobacter* genome was implicated in cell-cycle control as a result of time-varying mRNA concentrations that were slaved to the cell cycle itself. The idea of the experiment was to break open synchronized cells every 15 minutes and to harvest their mRNA. Then, by using a DNA microarray to find out which genes were being expressed at that moment, it was possible to put together a profile of which genes were expressed when. The outcome of this experiment is shown in

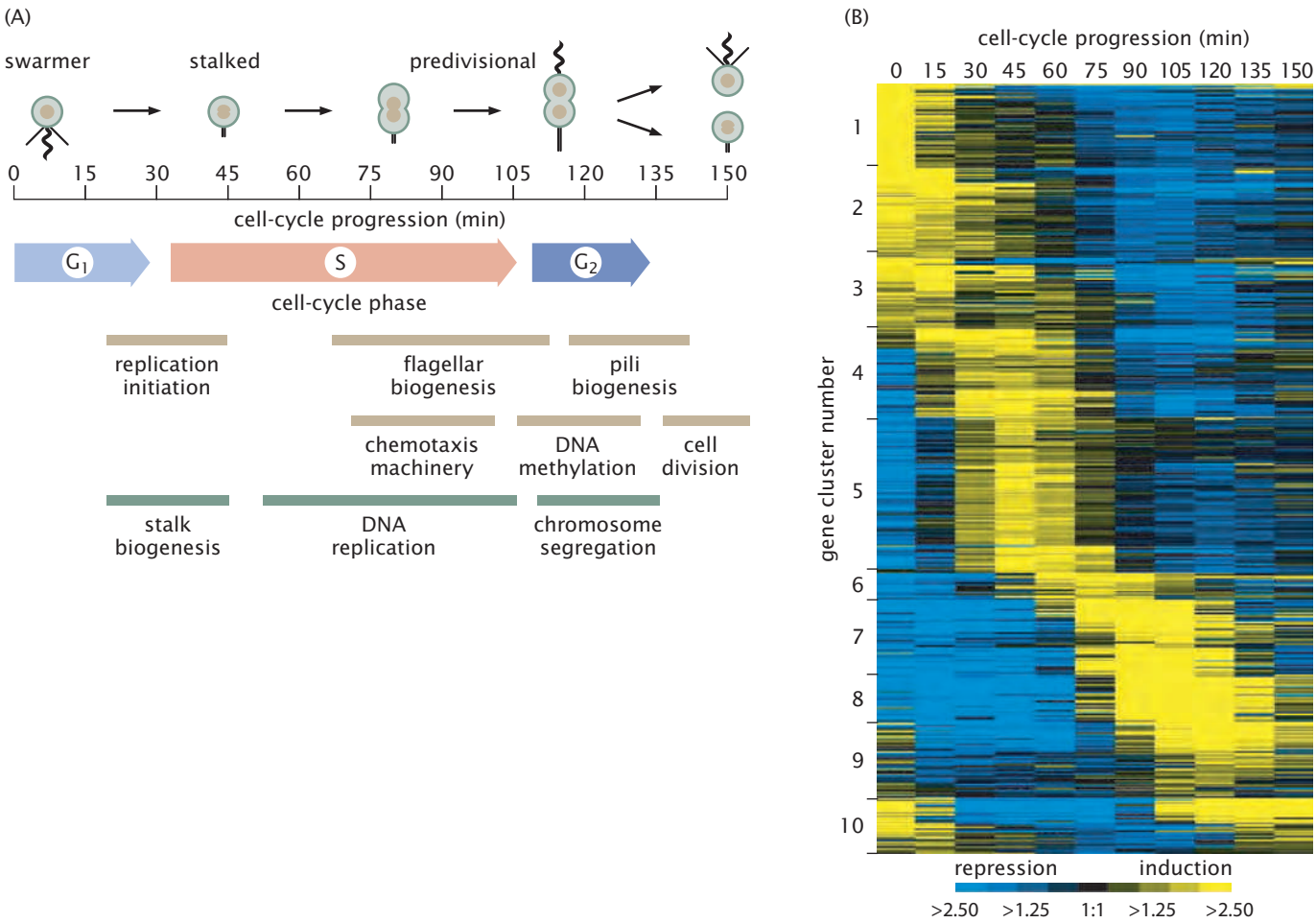


Figure 3.23: Gene expression during the cell cycle for *Caulobacter crescentus*. (A) The 150 minute cell cycle of *Caulobacter* is shown, highlighting some of the key morphological and metabolic events that take place during cell division. (B) The microarray data show how different batteries of related genes are expressed in a precise order. The genes are organized by time of peak expression. Each row corresponds to the time history of expression for a given gene, with time running from 0 to 150 minutes from left to right. For each gene, yellow indicates the time with the highest level of gene expression and blue indicates the time with the lowest level of gene expression. From top to bottom, the genes are organized into different clusters associated with different processes such as DNA replication and chromosome segregation. (Adapted from M. T. Laub et al., *Science* 290:2144, 2000.)

Figure 3.23. What these experiments revealed is the *relative* timing of different events over the roughly 150 minutes of the *Caulobacter* cell cycle.

The Formation of the Bacterial Flagellum Is Intricately Organized in Space and Time

A higher-resolution look at the relative timing of cellular events is offered by the macromolecular synthesis of one of the key organelles for cell motility, the bacterial flagellum. Figure 3.24 shows the various gene products (FlgK, MotB, etc.) that are involved in the formation of the bacterial flagellum. Essentially, each of these products corresponds to one of the protein building blocks associated with flagellar construction. Once the flagella are assembled, the cell propels itself around by spinning them. The dynamical question posed in the experiment is the extent to which the expression of the genes associated with these different building blocks is orchestrated in time.

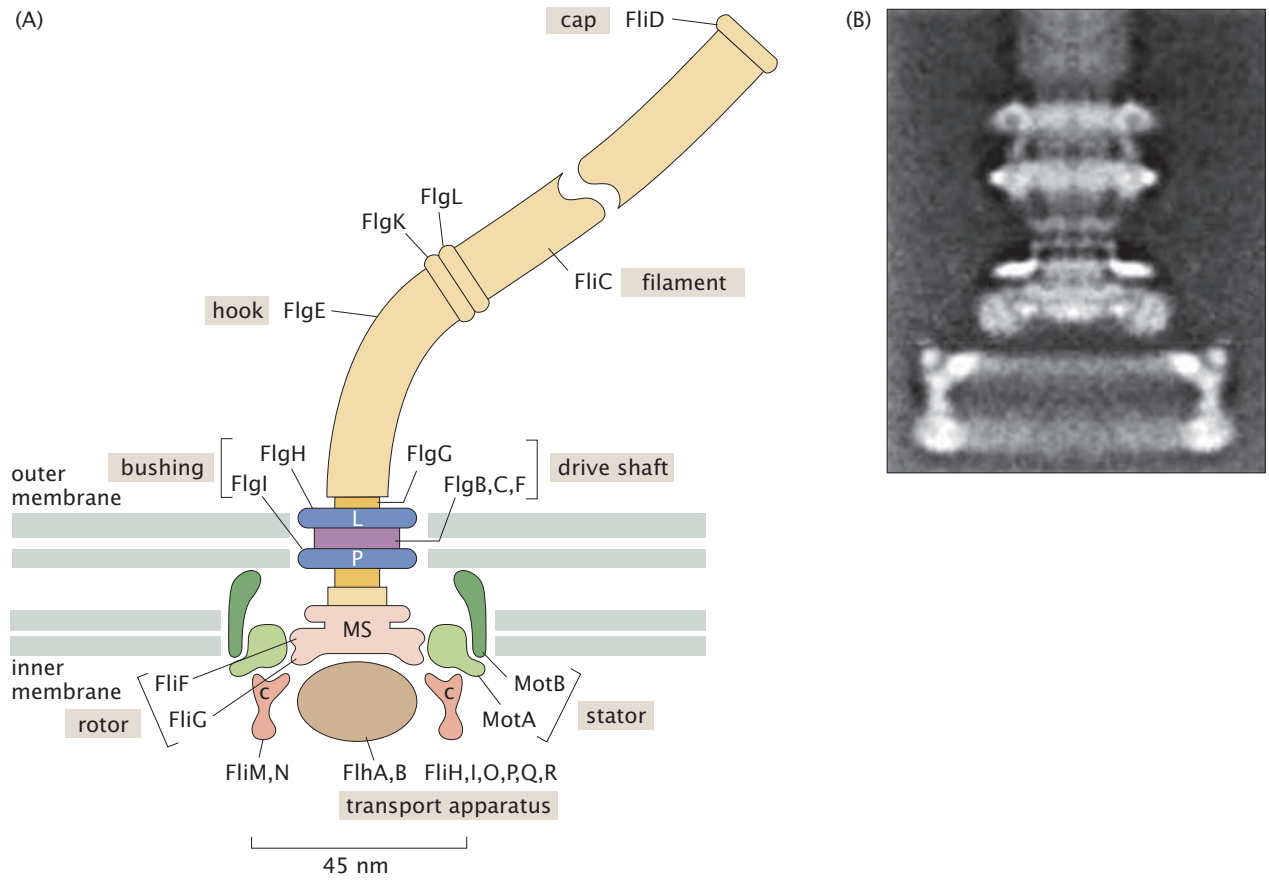


Figure 3.24: Molecular architecture of the bacterial flagellum. (A) The diagram shows the membrane-bound parts of the flagellar apparatus as well as the flagellum itself. The labels refer to the various gene products involved in the assembly of the flagellum. (B) The large ring at the bottom is the C ring, found on the cytoplasmic face. The smaller ring immediately above it is the MS ring, which spans the inner membrane, and the L and P rings (which span the outer membrane) are seen near the top. (Adapted from H. C. Berg, *Phys. Today* 53:24, 2000.)

The idea of the experiment is to induce the growth of flagella in starved *E. coli* cells (which lack flagella) and to use a reporter gene, namely, a gene leading to the expression of GFP, to report on when each of the different genes associated with the flagellar pathway is being expressed. This experiment permits us to peer directly into the dynamics of assembly of the bacterial flagellum, which reveals a sequence of events that are locked into succession in exactly the sort of way that we argued is characteristic of relative time. Figure 3.25 shows the results of this experiment. To deduce a time scale from this figure, we consider the band of expression and note that the roughly 15 genes are turned on sequentially over a period of roughly 180 minutes starting about 5 hours after the beginning of the experiment. This implies an approximate delay time between each product of roughly 12 minutes. Note that the entire experiment was run over a period of roughly 10 hours, though induction of new flagella (the part of the experiment of interest here) begins after three or four rounds of cell division have occurred.

3.3.3 Killing the Cell: The Life Cycles of Viruses

Cells are not the only biological entities that care about relative timing. Once viruses have infected a host cell, they are like a ticking time bomb with an ever-shortening fuse of early, middle, and late genes. Once these genes have been expressed and their products assembled, hundreds of new viruses emerge from the infected (and now defunct) cell to repeat the process elsewhere.

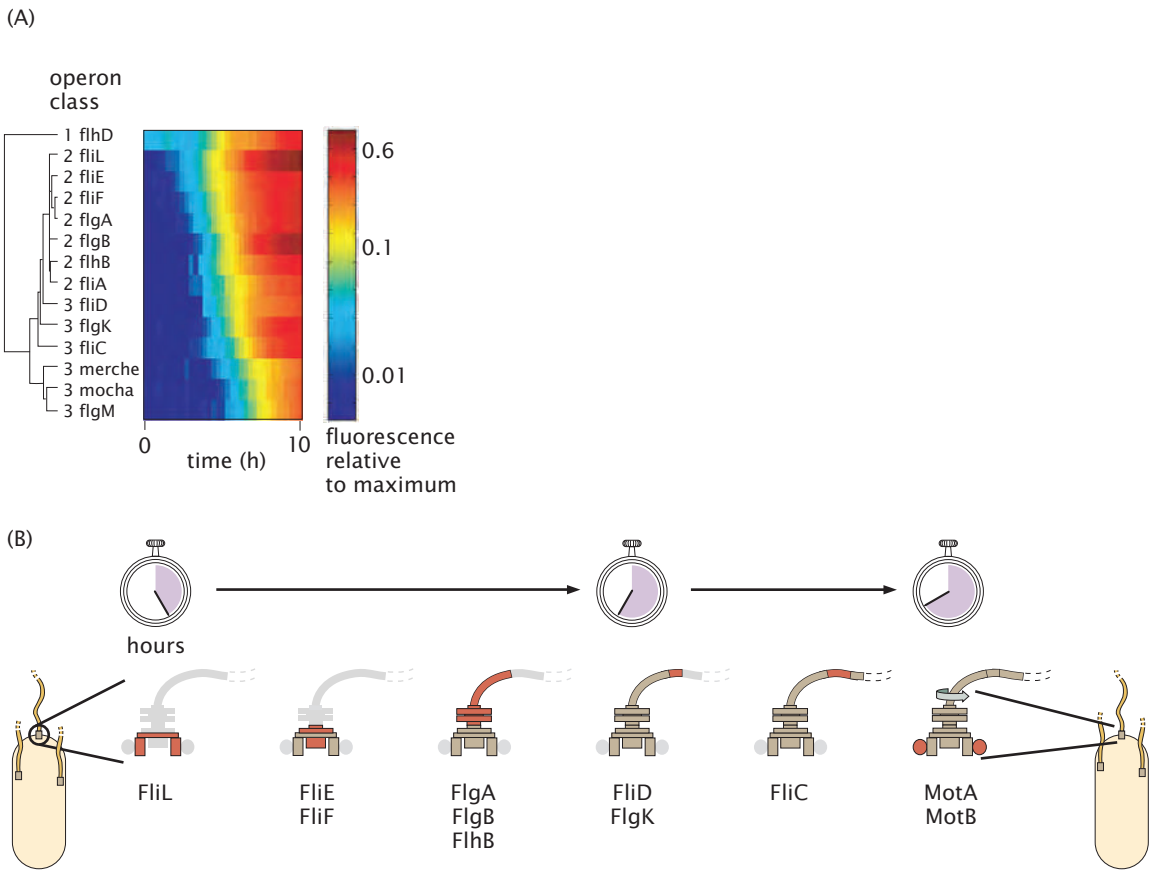


Figure 3.25: Timing of gene expression during the process of formation of the bacterial flagellum. (A) The extent of gene expression (as measured by fluorescence intensity) of each of the gene products as a function of time. (B) The cartoons show the timing of synthesis of different parts of the flagellum. The stopwatch times are chosen by assuming that in total the process takes roughly 3 hours and that flagellar synthesis begins roughly 5 hours after the beginning of the experiment. (Adapted from S. Kalir et al., *Science* 292:2080, 2001.)

Viral Life Cycles Include a Series of Self-Assembly Processes

We have already described the cell cycle as a master process characterized by an enormous number of subprocesses. A more manageable example of an entire "life cycle" is that of viruses, which illustrates the intricate relative timing of biological processes. An example of the viral life cycle of a bacteriophage (introduced in the last chapter as a class of viruses that attack bacterial cells) is shown in Figure 3.26, which depicts the key components in the life history of the virus. The key stages in this life cycle are captured in kinetic verbs such as infect, transcribe, translate, assemble, package, and lyse. Infection is the process of entry of the viral DNA into the host cell. Transcription and translation refer to the hijacking of the cellular machinery so as to produce viral building blocks (both nucleic acid and proteins). Assembly is the coming together of these building blocks to form the viral capsid. Packaging, in turn, is the part of the life cycle when the viral genome is enclosed within the capsid. Finally, lysis refers to the dissolution of the host cell and the emergence of a new generation of phage that go out and repeat their life cycle elsewhere.

As illustrated by the cartoon in Figure 3.26 and, in particular, by our use of the stopwatch motifs, the time between the arrival of the virus at the bacterial surface and the destruction of that very same membrane during the lysis phase when the newly formed viruses are released seems very short at 30 minutes. Indeed, one of our goals in the chapters that follow will be to come to terms with the 30 minute characteristic time scale of the viral life cycle and the various processes that make it up. On the other hand, though the absolute units (30 minutes) are interesting, it is important to emphasize that this

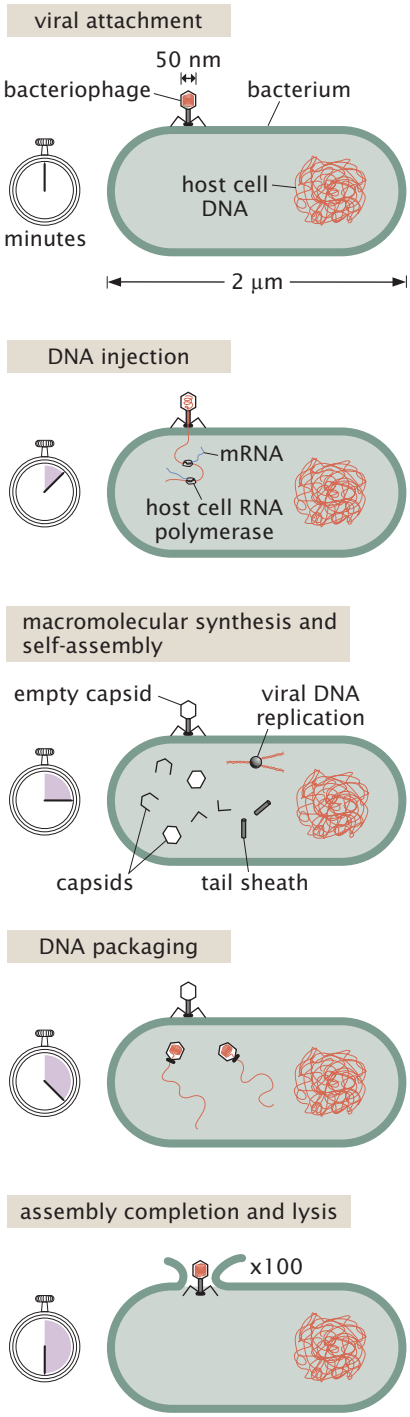


Figure 3.26: Timing the life cycle of a bacteriophage. The cartoon shows stages in the life cycle of a bacteriophage and roughly how long after infection these processes occur. Note that the head and the tail follow distinct assembly pathways and join only after they have separately assembled.

set of processes is locked together sequentially in a progression of relative timing events (as with the synthesis of the bacterial flagellar apparatus).

Because of their stunning structures and rich lifestyles, we now examine a second class of viruses, namely, RNA animal viruses such as HIV already introduced from a structural perspective in Section 2.2.4 (p. 66). As shown in Figure 3.27, the infection process for these viruses is quite distinct from that in bacterial viruses. In particular, we note the presence of a membrane coat on the virus, which allows the entire virus to attach to membrane-bound receptors on the victim cell. As a result of this interaction between the virus and the host cell, the virus is swallowed up by the cell that is under attack in a process of membrane fusion. Once the virus has entered the embattled cell, the genetic material (RNA) is released and reverse transcriptase creates a DNA molecule encoding the viral genome using the viral RNA as a template, which is then delivered to the host nucleus and incorporated into its genome. After the genetic material has been delivered to the nucleus, a variety of synthesis processes are undertaken that result in copies of the viral RNA as well as fascinating polyproteins (the Gag proteins described in Chapter 2), which are exported to the plasma membrane, where they undergo an intricate process of self-assembly at the membrane of the infected cell. Once the newly formed virus is exported, it undergoes a maturation process resulting in new, fully infectious, viral particles. Each of these processes is locked in succession in a pageant of relatively timed events.

3.3.4 The Process of Development

As already demonstrated in Section 2.3.3 (p. 78), one of the most compelling, mysterious, and visually pleasing processes in biology is the development of multicellular organisms. Like the cell cycle of individual cells, development depends upon a fixed relative ordering of events. Perhaps the most studied organism from the developmental perspective is the fruit fly *Drosophila melanogaster*. The process of *Drosophila* development has already been schematized in Figures 3.2(A) and (B). Embryonic development represents the disciplined outcome of an encounter between an egg and its partner sperm. In the hours that follow this encounter for the fruit fly, the nascent larva undergoes a series of nuclear divisions and migrations as shown in Figure 3.28. In particular, as the nuclei divide to the 10th generation (512 nuclei), they also start to collect near the surface of the developing larva, forming the syncytial blastoderm, an ovoid object with nearly all of the cells localized to the surface. At the 13th generation, the individual nuclei are enclosed by their own membranes to form the cellular blastoderm. The structural picture by the end of this process is a collection of roughly 5000 cells that occupy the surface of an ovoid object (roughly), which is 500 μm in length and roughly 200 μm in cross-sectional diameter.

Accompanying these latter stages of the developmental pathway is the beginning of a cellular dance in which orchestrated cell movements known as gastrulation lead to the visible emergence of the macroscopic structures associated with the nascent embryo. Snapshots from this process are shown in Figure 3.29 with a time scale associated with the process of gastrulation of the order of hours. We have already proclaimed the importance and beauty of the temporal organization of gene expression associated with the cell cycle. We now add to that compliment by noting that during the development of the

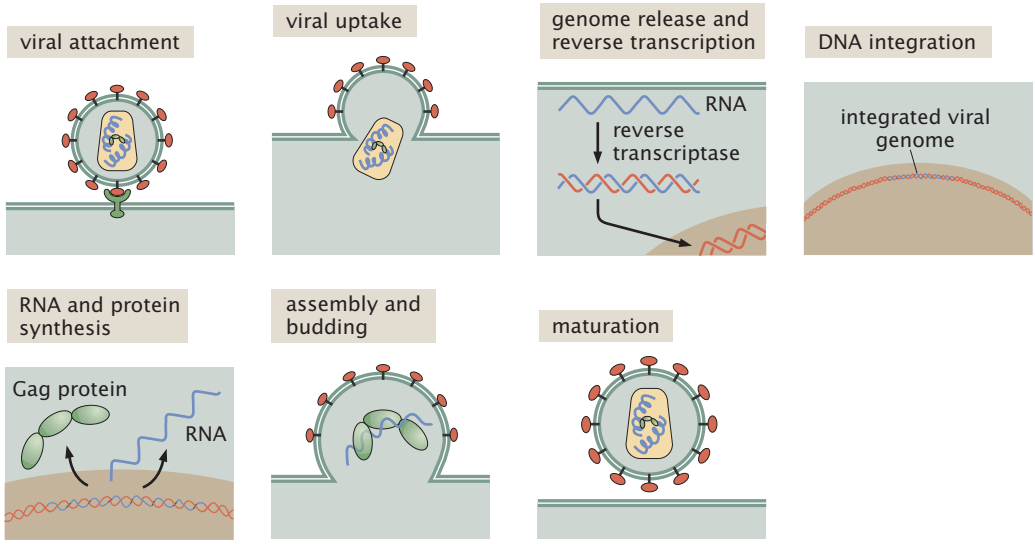


Figure 3.27: Stages in the life cycle of HIV. The timing of the events in the life cycle of HIV can be explored in Reddy and Yin (1999).

Drosophila body plan, there is an ordered spatial pattern of expression of genes with colorful names such as *hunchback* and *giant*, which determine the spatial arrangement of different cells.

These developmental processes make their appearance here because they too serve as an example of relative time. In particular, an example of the time ordering is the cascade of genes associated with the segmentation of the fly body plan into its anterior and posterior parts as already introduced in Section 2.3.3 (p. 78). The long axis of the *Drosophila* embryo is subject to increasing structural refinement as a result of a cascade of genes known collectively as segmentation genes. This collection of genes acts in a cascade, which is a code word for precisely the kind of sequential processes that are behind *relative time* as introduced in this section. The first set of genes in the cascade are known as the *gap* genes. These genes divide the embryo into three regions: the anterior, middle, and posterior. The *gap* genes have as protein products transcription factors that control the next set of genes in the cascade, which are known as *pair-rule* genes. The *pair-rule* genes begin to form the identifiable set of seven stripes of cells. Finally, the *segment polarity* genes are expressed in 14 stripes.

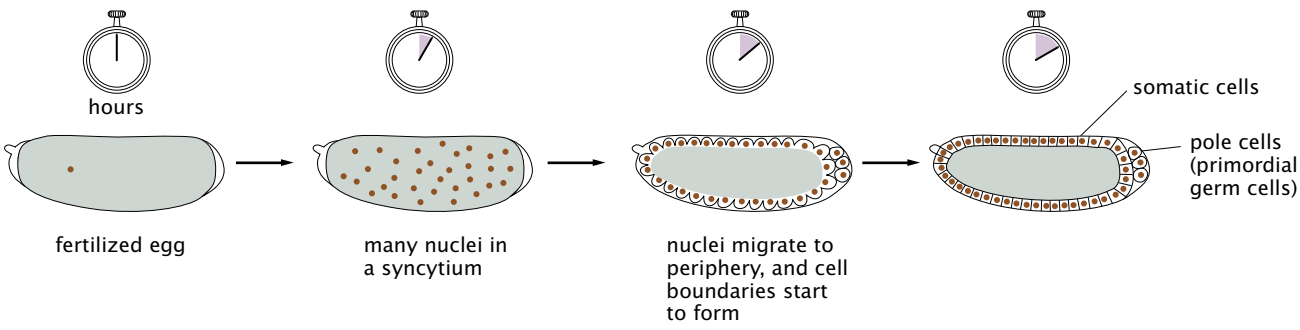


Figure 3.28: Early development of the *Drosophila* embryo. After fertilization, the single nucleus undergoes a series of eight rapid divisions producing 256 nuclei that all reside in a common cytoplasm. At this point, the nuclei begin to migrate toward the surface of the embryo while continuing to divide. After reaching the surface and undergoing a few more rounds of division, cell boundaries form by invaginations of the plasma membrane. At this early stage, the cells that are destined to give rise to sperm or eggs segregate themselves and cluster at the pole of the embryo. All of these events happen within roughly 2 hours.

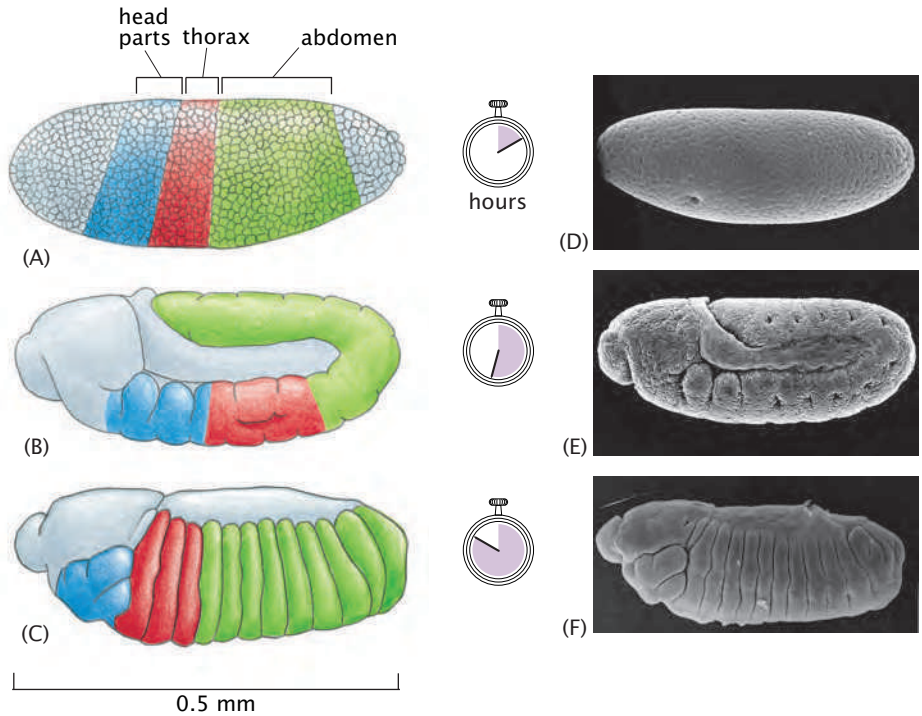


Figure 3.29: Early pattern formation in the *Drosophila* embryo. (A)–(C) show schematics of the shape of the *Drosophila* embryo at 2, 6, and 10 hours after fertilization, respectively. At 2 hours, no obvious structures have yet begun to form, but the eventual fates of the cells that will form different parts of the animal's body have already been determined. By 6 hours, the embryo has undergone gastrulation and the body axis of the embryo has lengthened and curled back on itself to fit in the egg shell. By 10 hours, the body axis has contracted and the separate segments of the animal's body plan have become clearly visible. (D), (E) and (F) show scanning electron micrographs of embryos at each of these stages. Remarkably, all of this pattern formation takes place without any growth. (A, adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed., Garland Science, 2008. D and E, courtesy of F. R. Turner and A. P. Mahowald, *Dev. Biol.* 50:95–108, 1976; F, from J. P. Petschek, N. Perrimon and A. P. Mahowald, *Dev. Biol.* 119:175–189, 1987. All with permission from Academic Press.)

Estimate: Timing Development A simple estimate of the number of cells associated with a developing organism can be obtained by assuming perfect synchrony from one generation to the next,

$$\text{number of cells} \approx 2^N, \tag{3.14}$$

where N is the number of generations. Further, if we assume that the cell cycle is characterized by a time τ_{cc}, then the number of cells as a function of time can be written simply as

$$\text{number of cells} \approx 2^{t/\tau_{cc}}. \tag{3.15}$$

Interestingly, in the early stages of *Drosophila* development, since it is only nuclear division (and hence largely DNA replication) that is taking place, the mean doubling time is 8 minutes. Thus after 100 minutes, roughly 10 generations worth of nuclear division will have occurred, with the formation of the approximately 1000 cells that form the syncytial blastoderm.

ESTIMATE

3.4 Manipulated Time

Sometimes, the cell is not satisfied with the time scales offered by the intrinsic physical rates of processes and has to find a way to beat these speed limits. For example, in some cases, the bare rates of biochemical reactions are prohibitively slow relative to characteristic cellular time scales and, as a result, cells have tied their fate to enzymatic manipulation of the intrinsic rates. In a similar vein, diffusion as a means of intracellular transport is ineffective over large distances. In this case, cells have active transport mechanisms involving molecular motors and cytoskeletal filaments that can overcome the diffusive speed limit, or alternatively, the cell might resort to reduced-dimensionality diffusion to beat the diffusive speed limit. There are even more tricky ways in which cells manipulate time, such as in the case of beating the bacterial replication limit. These examples and others will serve as the basis of our discussion of the way cells manipulate time.

3.4.1 Chemical Kinetics and Enzyme Turnover

Some chemical reactions proceed much more slowly than would be necessary for them to be biologically useful. For example, the hydrolysis of the peptide bonds that make up proteins would take times measured in years in the absence of proteases, which are the enzymes that cleave these bonds. Triose phosphate isomerase, the enzyme in the glycolysis pathway featured in Figure 2.32 (p. 70), is responsible for a factor-of-10^9 speed-up in the glycolytic reaction it catalyzes. What these numbers show is that even if a given reaction is favorable in terms of free energy, the energy barrier to that reaction can make it prohibitively slow. As a result, cells have found ways to manipulate the timing of reactions using enzymes as catalysts. Indeed, the whole of biochemistry is in some ways a long tale of catalyzed reactions, many of which take place on time scales much shorter than milliseconds, whereas, in the absence of these enzymes, they might not take place in a year! The individual players in the drama of glycolysis such as hexokinase, phosphofructokinase, triose phosphate isomerase, and pyruvate kinase reveal their identity as enzymes with the ending *ase* in their names. Enzymes are usually denoted by the ending *ase* and are classified according to the reactions they catalyze.

The basis of enzyme action is depicted in Figure 3.30. For concreteness, we consider an isomerization reaction where a molecule starts out in some high-energy state A and we interest ourselves in the transitions to the lower-energy state B. For the molecule schematized in the figure, the energy associated with the conformation has an electrostatic contribution and an "internal" contribution. As seen in Figure 3.30(A), the internal contribution has two minima, while the electrostatic contribution to the free energy is a monotonically decreasing function. In the presence of the enzyme, the height of the barrier is suppressed because the charges on the molecule of interest have an especially favorable interaction with the enzyme at the transition state.

The key point about the reaction rate is that, as shown in the figure, it depends upon the energy barrier separating the two states according to

$$\Gamma_{A\to B} \propto \mathrm{e}^{-G_{\mathrm{barrier}}/k_{\mathrm{B}}T}, \tag{3.16}$$

where $\Gamma_{A\to B}$ is the transition rate with units of s^{-1}, G_{barrier} is the size of the energy barrier and $k_{\mathrm{B}}T$ is the thermal energy scale with k_{B} known

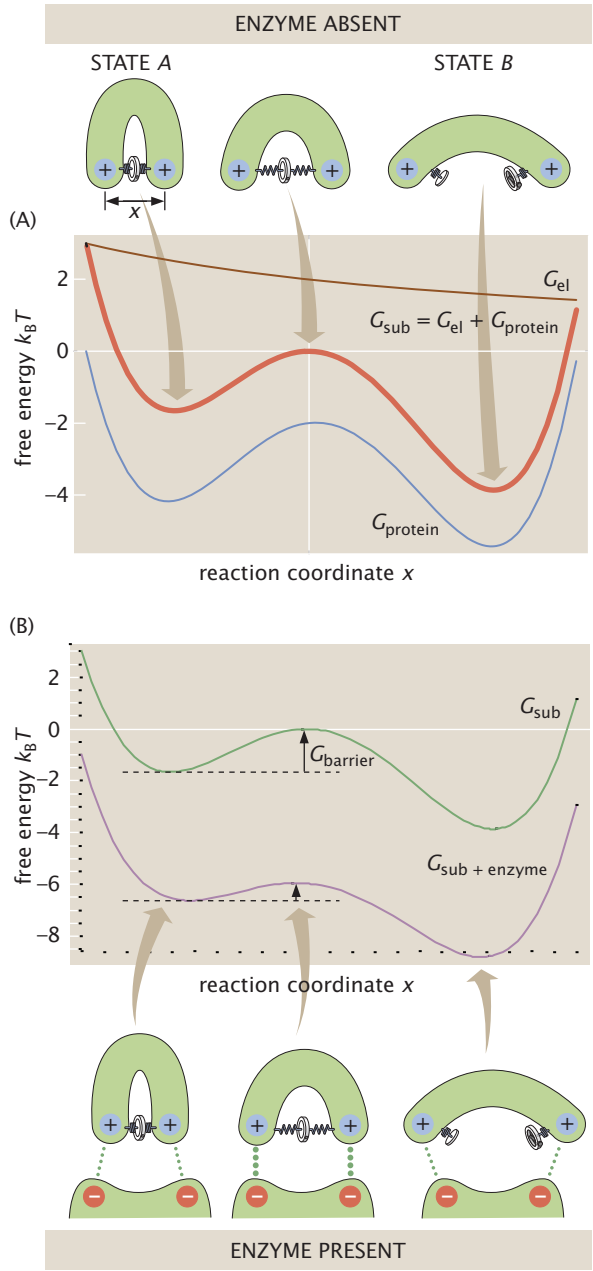


Figure 3.30: Enzymes and biochemical rates. (A) The electrostatic (G_{el}) and internal ($G_{protein}$) contributions to the overall free energy. The sum of these two contributions gives the total conformational free energy G_{sub} (in the absence of the enzyme). (B) A simple one-dimensional representation of the energy landscape for a biochemical reaction in the absence and presence of an enzyme to catalyze the reaction. The presence of the enzyme lowers the energy barrier through favorable charge interactions at the transition state and increases the rate of the reaction.

as Boltzmann's constant. As will be shown in Chapters 5 and 6, k_BT sets the scale of thermally accessible fluctuations in a system. Even though the energy of state *B* might be substantially lower than that of state *A*, the transitions can be exceedingly slow because of large barrier heights (that is, $G_{barrier} \gg k_BT$). The presence of an enzyme does not alter the end states or their relative energies, but it suppresses the barrier between the two states.

3.4.2 Beating the Diffusive Speed Limit

A second example of the way cells manipulate time is the way in which they perform transport and trafficking. Organelles, proteins, nucleic acids, etc. are often produced in one part of the cell only to be transported to another part where they are needed. For example, the mRNA molecules produced in the nucleus need to make their way to

the ribosomes, which are found in the cytoplasm on the endoplasmic reticulum. One physical process that can move material around is passive diffusion.

Diffusion Is the Random Motion of Microscopic Particles in Solution

Ions, molecules, macromolecular assemblies, and even organelles wander around aimlessly as a result of diffusion. Diffusion refers to the random motions suffered by microscopic particles in solution, and is sometimes referred to as Brownian motion in honor of the systematic investigations made by Robert Brown in the 1820s. Brown noticed the random jiggling of pollen particles suspended in solution, even for systems that are ostensibly in equilibrium and have no energy source. Indeed, so determined was he to find out whether or not this was some effect intrinsic to living organisms, he even examined exotic suspensions using materials such as the dust from the Sphinx and found the jiggling there too. The effects of Brownian motion are palpable for particles in solution that are micron size and smaller, exactly the length scales that matter to cells. Diffusion results from the fact that in the cell (and for microscopic particles in solution), deterministic forces are on a nearly equal footing with thermal forces, an idea to be fleshed out in Section 5.1.1 (p. 189) and illustrated graphically in Figure 5.1 (p. 190). Thermal forces result from the random collisions between particles that can be attributed to the underlying jiggling of atoms and molecules. This fascinating topic will dominate the discussion of Chapter 13.

ESTIMATE

Estimate: The Thermal Energy Scale One way to quantify the relative importance of the energy scale of a given process and thermal energies is by measuring the energy of interest in $k_B T$ units. At room temperature, typical thermal energies are of the order of $k_B T$, with a value of

$$k_B T = 1.38 \times 10^{-23}\,\text{J/K} \times 300\,\text{K} \approx 4.1 \times 10^{-21}\,\text{J} = 4.1\,\text{pN nm}. \tag{3.17}$$

One way to see the importance of this energy scale is revealed by Equation 3.16 (and will also be revealed by the Boltzmann distribution (to be described in Chapter 6), which says that the probability of a state with energy E_i is proportional to $e^{-E_i/k_B T}$). These expressions show that when the energy is comparable to $k_B T$, barriers will be small (and probabilities of microstates high). The numerical value ($k_B T \approx 4\,\text{pN nm}$) is especially telling since many of the key molecular motors relevant to biology act with piconewton forces over nanometer distances, implying a competition between deterministic and thermal forces. This discussion tells us that for many problems of biological interest, thermal forces are on a nearly equal footing with deterministic forces arising from specific force generation.

Diffusion Times Depend upon the Length Scale

One simple and important biological example of diffusion is the motion of proteins bound to DNA, which can be described as one-dimensional diffusion along the DNA molecule. For example, it is

thought that regulatory proteins (transcription factors) that control gene expression find their specific binding sites on DNA partly through this kind of one-dimensional diffusive motion. Another example is provided by the arrival of ligands at their specific receptors. The basic picture is that of molecules being battered about and every now and then ending up by chance in the same place at the same time. To get a feeling for the numbers, it is convenient to consider one of the key equations that presides over the subject of diffusion, namely,

$$t_{\text{diffusion}} \approx x^2/D, \tag{3.18}$$

where D is the diffusion constant. This equation tells us that the time scale for a diffusing particle to travel a distance x scales as the square of that distance.

Estimate: Moving Proteins from Here to There For molecules and assemblies that move passively within the cell, the time scale can be estimated using Equation 3.18. For a protein with a 5 nm diameter, the diffusion constant in water is roughly $100\,\mu\text{m}^2/\text{s}$; this estimate can be obtained from the Stokes–Einstein equation (to be discussed in more detail in Chapter 13 in Equation 13.62 on p. 531), which gives the diffusion constant of a sphere of radius R moving through a fluid of viscosity η at temperature T as $D = k_B T/6\pi\eta R$. The time scale for such a typical protein to diffuse a distance of our standard ruler (that is, across an *E. coli*) is

$$t_{E.coli} \approx \frac{L^2_{E.coli}}{D} \approx \frac{1\,\mu\text{m}^2}{100\,\mu\text{m}^2/\text{s}} \approx 0.01\,\text{s}. \tag{3.19}$$

This should be contrasted with the time scale required for diffusion to transport molecules from one extremity of a neuron to the other as shown in Figure 3.3. In particular, the diffusion time for the squid giant axon, which has a length of the order of 10 cm, is $t_{\text{diffusion}} \approx 10^8$ s! The key conclusion to take away from such an estimate is the impossibly long time scales associated with diffusion over large distances. Nature's solution to this conundrum is to exploit *active* transport mechanisms in which ESTIMATE ATP is consumed in order for motor molecules to carry out directed motion.

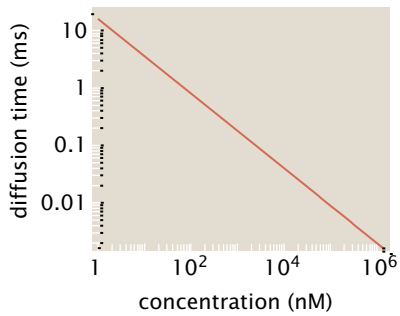


Figure 3.31: Concentration measured in diffusion times. The characteristic time scale for diffusion-mediated molecular encounters depends upon the concentration of the diffusing species.

Another way of coming to terms with diffusion is suggested by thinking about concentration. In Figure 2.12 (p. 48), we described physical interpretations of concentration in terms of number of molecules per *E. coli* cell and in terms of mean molecular spacings. We can also think of concentrations in terms of diffusion times by using Equation 3.18. In particular, this equation can be used to estimate a characteristic time scale for reactants to find each other as a function of the concentration as shown in Figure 3.31.

Diffusive Transport at the Synaptic Junction Is the Dynamical Mechanism for Neuronal Communication

One of the classic episodes in the history of biology centered around the battle between Golgi, Ramon y Cajal, and their followers as to the nature of nerve transmission. At the heart of the controversy over

soup versus sparks was the question of whether nerves are reticula (the view of Golgi) or rather built up of individual nerve cells, with communication between adjacent cells carried out by molecular transport. This debate was largely settled in favor of the "soup" view.

Estimate: Diffusion at the Synaptic Cleft The ideas introduced above can help us understand the dynamics of neurotransmitters at synapses. An example of the geometry of such a synapse is shown in Figure 3.32. Using exactly the same ideas as developed in the previous section, we can work out the time scale for neurotransmitters released at one side of the synapse to reach the receptors on the neighboring cell.

ESTIMATE

To be concrete, consider the diffusion of acetylcholine across a synaptic cleft with a size of roughly 20 nm. Given a diffusion constant for acetylcholine of $\approx 100\ \mu m^2/s$, the time for these molecules to diffuse across the cleft is

$$t = L^2/D \approx \frac{400\ \text{nm}^2}{10^8\ \text{nm}^2/\text{s}} \approx 4\,\mu\text{s}. \qquad (3.20)$$

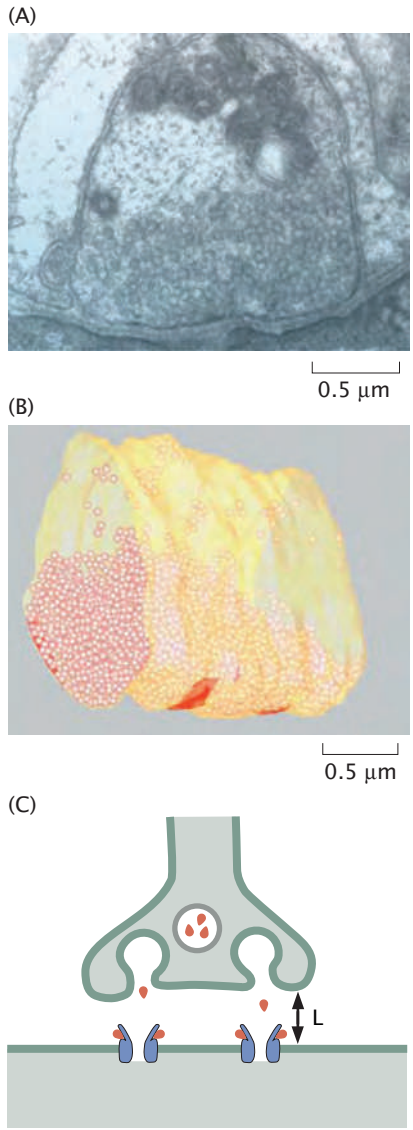


Figure 3.32: Geometry of the synapse. (A) Electron micrograph image of a nerve terminal. (B) Reconstruction of the synaptic geometry for a nerve terminal segment with a length of roughly 2 μm. The many vesicles that are responsible for the active response of the synapse are shown in red. (C) Schematic of the diffusive transport across a synaptic cleft. (A and B, adapted from S. O. Rizzoli and W. J. Betz, *Nat. Rev. Neurosci.*, 6:57, 2005.)

Molecular Motors Move Cargo over Large Distances in a Directed Way

In many instances, particularly for movement over long distances, diffusion is too slow to be of any use for intracellular transport. To beat the diffusive speed limit, cells manipulate time with a sophisticated array of molecular machines (usually proteins) that result in directed transport. These processes are collectively powered by the consumption of some energy source (usually ATP). Broadly construed, the subject of active transport allows us to classify a wide variety of molecules as *molecular motors*, the main subject of Chapter 16 (p. 623). We have already seen the existence of such motors in a number of different contexts, with both DNA polymerase and RNA polymerase, which were introduced earlier in the chapter, satisfying the definition of active transport because they use chemical energy to move along their DNA substrate in a directed way.

Concretely, the class of motors of interest here are those that mediate transport of molecules from one place in the cell to another in eukaryotes. Often, such transport takes place as vesicular traffic, with the cargo enclosed in a vesicle (a flexible spherical shell made up of lipid molecules in the form of a lipid bilayer) that is in turn attached to some molecular motor. These molecular motors travel in a directed fashion on the cytoskeletal network that traverses the cell. For example, traffic on microtubules runs in both directions as a result of two translational motors, kinesin and dynein. Molecular-motor-mediated transport on actin filaments is shown in cartoon form in Figure 3.33. Note that this cartoon gives a rough idea of the relative proportions of the motors and the actin filaments on which they move and conveys the overall structural features, such as two heads, of the motors themselves. In addition, Figure 3.33(B) shows a time trace of the position of a fluorescently labeled myosin motor, which illustrates the discrete steps of the motor, also permitting a measurement of the mean velocity.

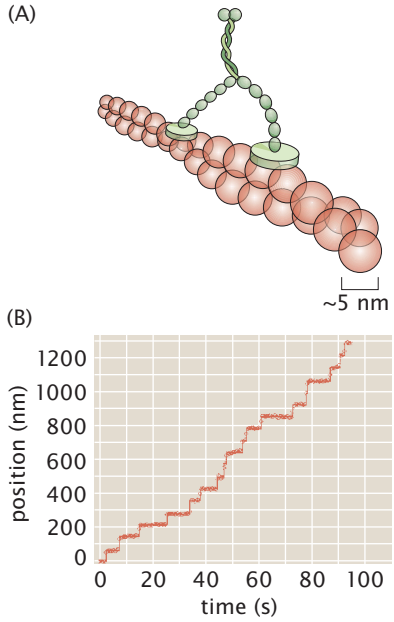


Figure 3.33: Motion of myosin on an actin filament. (A) Schematic of the motor on an actin filament. Note that the step size is determined by the periodicity of the filament. (B) Position as a function of time for the motor myosin V as measured using single-molecule techniques. (Adapted from A. Yildiz et al., *Science* 300:2061, 2003.)

Estimate: Moving Proteins from Here to There, Part 2
We have already noted that biological motility is based in large measure on diffusion. On the other hand, there are a host of processes that cannot wait the time required for diffusion. In particular, recall that our estimate for the diffusion time for a typical protein to traverse an axon was a whopping 10^8 s, or roughly 3 years. For comparison, we can estimate the transport time for kinesin moving on a microtubule over the same distance. As the typical speed of kinesin in a living cell is 1 μm/s, the time for it to transport a protein over a distance of 10 cm is 10^5 s, or just over a day.

To see these ideas play out more concretely, we can return to Figure 3.3 (p. 93). The classic experiment highlighted there traces the time evolution of radioactively labeled proteins in a neuron. The figure shows that the radio-labeled proteins travel roughly 18 mm in 12 days, which translates into a mean speed of roughly 20 nm/s. Observed axonal transport speeds for single motors are a factor of 10 or more larger, but we can learn something from this as well. In particular, motors are not perfectly "processive"—that is, they fall off of their cytoskeletal tracks occasionally and this has the effect of reducing their mean speed. Observed motor velocities are reported, on the other hand, on the basis of tracking individual motors during one of their processive trajectories.

ESTIMATE

Membrane-Bound Proteins Transport Molecules from One Side of a Membrane to the Other

Another way in which cells manipulate transport rates is by selectively and transiently altering the permeability of cell membranes through protein channels and pumps. Many ionic species are effectively unable to permeate (at least on short time scales) biological membranes. What this means is that concentration gradients can be maintained across these membranes until these protein channels open, which then permits a flow of ions down their concentration gradient. In fancier cases, such as indicated schematically in Figure 3.34, ions can be pumped up a concentration gradient through mechanisms involving ATP hydrolysis. In Figure 3.2(G) we showed the process of ion transport across a membrane with a characteristic time of microseconds.

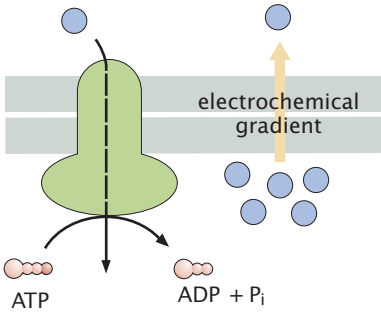


Figure 3.34: Active transport across membranes. Molecular pumps consume energy in the form of ATP hydrolysis and use the liberated free energy to pump molecules against their concentration gradient.

Estimate: Ion Transport Rates in Ion Channels An ion channel embedded in the cell membrane can be thought of as a tube with a diameter of approximately $d = 0.5$ nm (size of hydrated ion) and a length $l = 5$ nm (width of the lipid bilayer). With these numbers in hand, and a typical value of the diffusion constant for small ions (e.g. sodium), $D \approx 2000\,\mu\text{m}^2/\text{s}$, we can estimate the flux of ions through the channel, assuming that their motion is purely diffusive.

To make this estimate, we invoke an approximate version of Fick's law (to be described in detail in Chapter 8) that says that the flux (number of molecules crossing unit area per unit time) is proportional to the difference in concentration and inversely proportional to the distance between the two

ESTIMATE

"reservoirs." Mathematically, this can be written as

$$J_{\text{ion}} \approx D\frac{\Delta c}{l}, \tag{3.21}$$

where Δc is the difference in ion concentration across the cell membrane. For typical mammalian cells the concentration difference for sodium or potassium is $\Delta c \approx 100\,\text{mM}$ (or $\approx 6 \times 10^{-2}$ molecules/nm^3) and the distance across the membrane is $l \approx 5\,\text{nm}$, resulting in

$$\begin{aligned} J_{\text{ion}} &\approx 2 \times 10^9 \frac{\text{nm}^2}{\text{s}} \times \frac{6 \times 10^{-2}\ \text{molecules/nm}^3}{5\,\text{nm}} \\ &\approx 2 \times 10^7\,\text{nm}^{-2}\,\text{s}^{-1}, \end{aligned} \tag{3.22}$$

where we have used a diffusion constant of $D = 2 \times 10^9\,\text{nm}^2/\text{s}$ appropriate for an ion. Given the cross-sectional area of a typical channel $A_{\text{channel}} = d^2\pi/4 \approx 0.2\,\text{nm}^2$, the number of ions traversing the membrane per second is estimated to be

$$\frac{\text{d}N_{\text{ion}}}{\text{d}t} = J_{\text{ion}}A_{\text{channel}} \approx 2 \times 10^7\,\text{nm}^{-2}\,\text{s}^{-1} \times 0.2\,\text{nm}^2 = 4 \times 10^6\,\text{s}^{-1}. \tag{3.23}$$

This estimate does remarkably well at giving a sense of the time scales associated with ion transport across ion channels.

Enzymes, molecular motors, and ion channels (and pumps) are all ways in which the cell uses proteins to circumvent the intrinsic rates of different physical or chemical processes.

3.4.3 Beating the Replication Limit

The most fundamental process of cellular life is to form two new cells. A minimal requirement for this to take place is that an individual cell must duplicate its genetic information. Replication of the genetic material proceeds through the action of DNA polymerase, an enyzme that copies the DNA sequence information from one DNA strand into a complementary strand. Like all biochemical reactions, this requires a certain amount of time, which we estimated earlier in the chapter. Why should any cell accept this speed limit on its primary directive of replication? When we calculated the replication time for the *E. coli* chromosome, we concluded that two replication forks operating at top speed would be sufficient to replicate the chromosome in approximately the 3000 s division time that we stipulated for a bacterium growing in a minimally defined medium with glucose as the sole carbon source under a continuous supply of oxygen. While this set of conditions is in many ways convenient for the human experimentalist, it is by no means ideal for the bacterium. If instead of only supplying glucose we add a rich soup of amino acids, *E. coli* will grow much faster, with a doubling time of order 20 minutes (1200 s).

How can the bacterium double more quickly than its chromosome can replicate? For *E. coli* and other fast-growing bacteria, the answer is a simple and elegant manipulation of the procedural time limit imposed by the DNA replication apparatus. The bacterium begins replicating its chromosome a second time before the first replication is complete. In a rapidly growing *E. coli*, there may be between four

and eight copies of the chromosomal DNA close to the replication origin, even though there may be only one copy of the chromosome close to the replication terminus. In other words, the bacterium has started replicating its daughter's, grandaughter's or even great grandaughter's chromosome before its own replication is complete. The newborn *E. coli* cell is thus essentially already pregnant with partially replicated chromosomes preparing for the next one or two generations.

As we have noted above, the genome size for bacteria tends to be substantially smaller than the genome size for eukaryotic cells. Nonetheless, eukaryotic cells are still capable of replicating at a remarkably fast rate. For example, early embryonic cells of the South African clawed frog *X. laevis* can divide every 30 minutes. Despite the fact that its genome (3100 Mbp) is more than 600-fold larger than the genome of *E. coli*, two mechanical changes enable rapid replication of the *Xenopus* genome (and the genome of other eukaryotes). First, the genome is subdivided into multiple linear chromosomes rather than a single circular chromosome. Second, and more importantly, replication is initiated simultaneously from many different origins sprinkled throughout the chromosome as opposed to the single origin of bacterial chromosomes. This parallel processing for the copying of genomic information enables the task to be completed more rapidly than would be dictated by the procedural time limit imposed by a single molecule of DNA polymerase.

3.4.4 Eggs and Spores: Planning for the Next Generation

We have been considering the processes of cell growth and division as though they are tightly coupled with one another. In some cases, however, organisms may separate the processes of growth and division so that they occur over different spans of time. The most dramatic example of this is in the growth of giant egg cells followed by extremely rapid division of early embryos after fertilization. For example, in the frog *X. laevis*, each individual egg cell is enormous—up to 1 mm across—and grows gradually within the body of the female over a period of 3 months. Following fertilization, cell division occurs without growth so that a tadpole hatches after 36 hours that has the same mass as the egg from which it is derived.

Even for organisms where cell growth normally is coupled to cell division, there are several mechanisms whereby cells may choose to postpone either growth or division if conditions are unfavorable. For example, many cells ranging from bacteria through fungi to protozoans such as *Dictyostelium* are capable of creating spores. Spores are nearly metabolically inert and serve as a storage form for the genomic information of the species that can survive periods of drought or low nutrient availability. This is a mechanism by which an organism can effectively exist in suspended animation, waiting for however much time is needed to pass until conditions become favorable again. When fortune finally favors the spore, it can germinate, releasing a rapidly growing cell. The maximum survival time of spores is unknown. However, there are *Bacillus* spores that were put into storage by Louis Pasteur in the late nineteenth century that appear to be fully viable today and it is generally accepted that some spores may be able to survive for thousands of years. Some controversial reports even suggest that viable bacterial spores can be recovered from insect bodies trapped in amber for a few million years. The seeds of flowering plants perform a similar role, though they are typically not as robust as spores.

Although animals do not form true spores, several do have forms that permit long-term survival under starvation conditions. The most familiar examples are hibernation of large animals such as bears, which can survive an entire winter season without eating. Smaller animals perform similar tricks. These include dauer (German for "enduring") form larvae of several worms. The most impressive example of "suspended animation" among animals is presented by the tardigrade or water bear, a particularly adorable segmented metazoan that rarely grows larger than 1 mm. The tardigrade normally lives in water, but when it is dried out, it slows its metabolism and alters its body shape, extruding almost all of its water, to form a dried out form called a tun. Tardigrade tuns can be scattered by wind and can survive extreme highs and lows of temperature and pressure. When the tun falls into a favorable environment like a pond, the animal will reanimate. Each of these examples shows how organisms have evolved mechanisms that are completely indifferent to the absolute passage of time.

3.5 Summary and Conclusions

Because life processes are associated with constant change, it is important to understand how long these processes take. In the case of the diurnal clock, organisms are able to measure the passage of time with great regularity to determine their daily behaviors. For most other kinds of biological processes, times are not absolute. In this chapter, we explored several different views of time in biological systems starting with the straightforward assignment of procedural time as measured by the amount of time it takes to complete some process. In complex biological systems, processes that occur at intrinsically different rates may be linked together such that one must be completed before another can begin. Examples of this kind of measurement of relative time are found in the regulation of the cell cycle, the assembly of complex structures such as the bacterial flagellum, etc. Finally, we briefly explored some of the ways that organisms manipulate biological processes to proceed faster or slower than the normal intrinsic rates. Armed with these varying views of time, we will use time as a dimension in our estimates and modeling throughout the remainder of the book. Time will particularly take center stage in Part 3, Life in Motion, where dynamic processes will be revealed in all their glory.

3.6 Problems

Key to the problem categories: ● Model refinements and derivations ● Estimates

● 3.1 Growth and the logistic equation

In the chapter, we described the logistic equation as a simple toy model for constrained growth of populations. In this problem, the goal is to work out the dynamics in more detail.

(a) Rewrite the equation in dimensionless form and explain what units this means time is measured in.

(b) Find the value of N at which the growth rate is maximized.

(c) Find the maximum growth rate.

(d) Use these results to make a one-dimensional phase portrait like that shown in Figure 3.10.

● 3.2 Protein synthesis and degradation in the *E. coli* cell cycle

Improve the estimates for the protein synthesis rates of *E. coli* during a cell cycle from Sections 3.1.3 and 3.2.1 by including the effect of protein degradation. For simplicity, assume that all proteins are degraded at the same rate with a half-life of 60 minutes and work out the number of ribosomes needed to produce the protein content of a new bacterium given that part of the synthesis is required for the replacement of degraded proteins. Compare with the results from the estimate in Section 2.1.2 (p. 38).

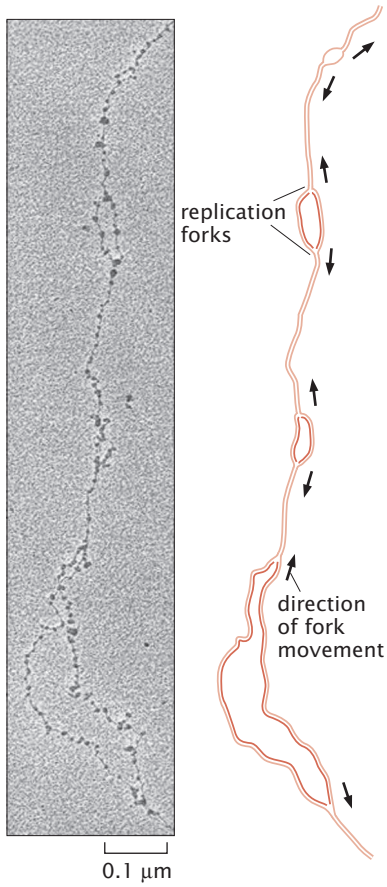


Figure 3.35: Replication forks in *D. melanogaster*. Replication forks move away in both directions from replication origins. (Electron micrograph courtesy of Victoria Foe. Adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed. Garland Science, 2008.)

• 3.3 DNA replication rates

Assuming that Figure 3.35 is a representative sample of the replication process:

(a) Estimate the fraction of the total fly genome shown in the micrograph. Note that the fly genome is about 1.8×10^8 nucleotide pairs in size.

(b) Estimate the number of DNA polymerase molecules in a eukaryotic cell like this one from the fly *D. melanogaster*.

(c) There are eight forks in the micrograph. Estimate the lengths of the DNA strands between replication forks 4 and 5, counting up from the bottom of the figure. If a replication fork moves at a speed of roughly 40 bp/s, how long will it take for forks 4 and 5 to collide?

(d) Given the mean spacing of the bubbles, estimate how long it will take to replicate the entire fly genome.

• 3.4 RNA polymerase and ribosomes

(a) If RNA polymerase subunits β and β′ together constitute approximately 0.5% of the total mass of protein in an *E. coli* cell, how many RNA polymerase molecules are there per cell, assuming each β and β′ subunit within the cell is found in a complete RNA polymerase molecule? The subunits have a mass of 150 kDa each. (Adapted from Problem 4.1 of Schleif, 1993.)

(b) Rifampin is an antibiotic used to treat *Mycobacterium* infections such as tuberculosis. It inhibits the initiation of transcription, but not the elongation of RNA transcripts. The time evolution of an *E. coli* ribosomal RNA (rRNA) operon after addition of rifampin is shown in Figures 3.36(A)–(C). An operon is a collection of genes transcribed as a single unit. Use the figure to estimate the rate of transcript elongation. Use the beginning of the "Christmas-tree" morphology on the left of Figure 3.36(A) as the starting point for transcription.

(c) Using the calculated elongation rate, estimate the frequency of initiation off of the rRNA operon. These genes are among the most transcribed in *E. coli*.

(d) As we saw in the chapter, a typical *E. coli* cell with a division time of 3000 s contains roughly 20,000 ribosomes. Assuming there is no ribosome degradation, how many RNA polymerase molecules must be synthesizing rRNA at any instant? What percentage of the RNA polymerase molecules in *E. coli* are involved in transcribing rRNA genes?

• 3.5 Cell cycle and number of ribosomes

Read the paper "Ribosome content and the rate of growth of *Salmonella typhimurium*" by R.E. Ecker and M. Schaechter (*Biochim. Biophys. Acta* **76**, 275, 1963).

(a) The authors show that their data can be fit by a simple assumption that the rate of production of soluble protein, P, is proportional to the number of ribosomes. But this leaves out the need for the ribosomes to also make ribosomal protein, in an amount, R. The simplest assumptions are that the total protein production rate, dA/dt, is proportional to the number of ribosomes, and the total amount of non-ribosomal protein (also knows as soluble protein) needed per cell is independent of growth rate. Given these assumptions, how would the ratio R/A, with A the total protein content, depend on the growth rate? Does this give a maximum growth rate? If so, what is it?

(b) With a 3000 s division time for *E. coli*, about 25% of its protein is ribosomal. Note that the microbe *Salmonella typhimurium* considered here is very similar to *E. coli*. Using these numbers and your results from above, what fraction of the protein would be ribosomal for the highest growth rate studied in the paper? How does this compare to their measured ratio of ribosomes to soluble protein, R/P at these growth rates? How does the predicted R/P at high growth rates change if you now assume that dP/dt is proportional to R as they did in the paper?

(c) Another factor that needs to be taken into account is the decay of proteins. If all proteins decayed at the same rate, γ, how would this modify your results from (a)? How does the predicted functional from of R/A versus growth rate change? Explain why the data rule out γ being too large and hence infer a rough lower bound for the lifetime, $1/\gamma$, of "average" proteins.

(Problem courtesy of Daniel Fisher.)

• 3.6 The bleach–chase method and protein degradation

Protein production is tempered both by active degradation and by dilution due to cell division. A clever recent method (see Eden et al. 2011) makes it possible to measure net degradation rates by taking two populations of the same cells that have a fluorescent protein fused to the protein of interest. In one of those populations, the fluorescent proteins are photobleached and in the other they are not. Then, the fluorescence in the two populations is monitored over time as the photobleached population has its fluorescence replenished by protein production. By monitoring the time dependence in the difference between these two populations, the degradation rate constant can be determined explicitly.

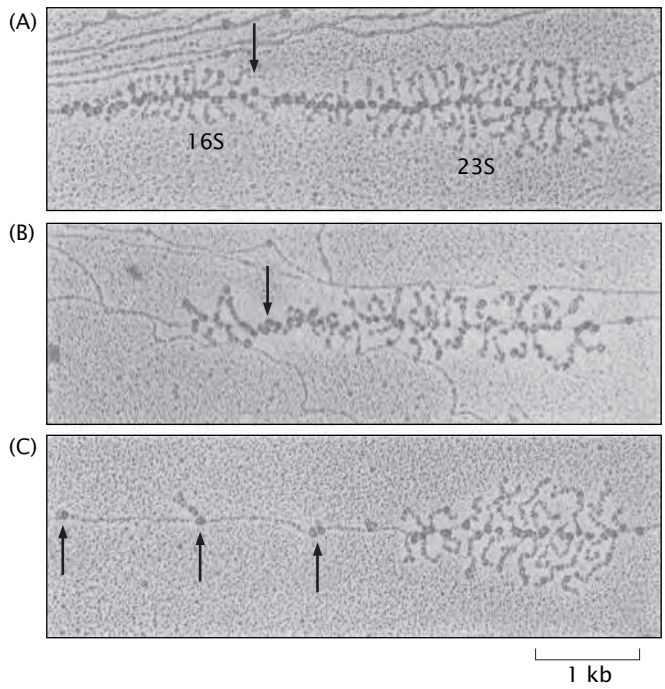


Figure 3.36: Effect of rifampin on transcription initiation. Electron micrographs of *E. coli* rRNA operons: (A) before adding rifampin, (B) 40 s after addition of rifampin, and (C) 70 s after exposure. No new transcripts have been initiated, but those already initiated are carrying on elongation. In (A) and (B) the arrow is used as a fiducial marker and signifies the site where RNaseIII cleaves the nascent RNA molecule producing 16S and 23S ribosomal subunits. RNA polymerase molecules that have not been affected by the antibiotic are marked by the arrows in (C). (Adapted from L. S. Gotta et al., *J. Bacteriol.* 20:6647, 1991.)

(a) At a certain instant in time, we photobleach one of the two populations so that their fluorescent intensity is now reduced relative to its initial value and relative to the value in the unphotobleached population. The number of fluorescent proteins N_f in the unphotobleached population varies in time according to the simple dynamical equation

$$\frac{dN_f}{dt} = \beta - \alpha N_f, \tag{3.24}$$

which acknowledges a rate of protein production β and a degradation rate α. Explain what this equation means and what it implies about the steady-state value of the number of fluorescent proteins per cell.

(b) A similar equation describes the dynamics of the unphotobleached molecules in the cells that have been subjected to photobleaching, with the number that that are unphotobleached given by N_u and described by the dynamical equation

$$\frac{dN_u}{dt} = \beta - \alpha N_u. \tag{3.25}$$

On the other hand, the number of photobleached proteins is subject to a different dynamical evolution described by the equation

$$\frac{dN_p}{dt} = -\alpha N_p, \tag{3.26}$$

since all that happens to them over time is that they degrade. Explain why there are two populations of proteins within the photobleached cells and why these are the right equations.

(c) In the paper, the authors then tell us to evaluate the difference in the number of fluorescent proteins in the two populations. A critical assumption then is that

$$\frac{dN_f}{dt} = \frac{dN_u}{dt} + \frac{dN_p}{dt} \tag{3.27}$$

The point is that, over time after photobleaching, the photobleached cells will become more fluorescent again as new fluorescent proteins are synthesized.

Plot the difference between the intensity of the cells that were not disturbed by photobleaching and those that were. In particular, we have

$$\frac{d(N_f - N_u)}{dt} = -\alpha(N_f - N_u). \tag{3.28}$$

Note that this quantity is experimentally accessible since it calls on us to measure the level of fluorescence in the two populations and to examine the difference between them. Integrate this equation and show how the result can be used to determine the constant α that characterizes the dynamics of protein decay.

• 3.7 The sugar budget in minimal medium

In rapidly dividing bacteria, the cell can divide in times as short as 1200 s. Make a careful estimate of the number of sugars (glucose) needed to provide the carbon for constructing the macromolecules of the cell during one cell cycle of a bacterium. Use this result to work out the number of carbon atoms that need to be taken into the cell each second to sustain this growth rate.

• 3.8 Metabolic rates

Assume that 1 kg of bacteria burn oxygen at a rate of 0.006 mol/s. This oxygen enters the bacterium by diffusion through its surface at a rate given by $\Phi = 4\pi DRc_0$, where $D = 2\,\mu m^2/ms$ is the diffusion constant for oxygen in water, $c_0 = 0.2\,mol/m^3$ is the oxygen concentration, and R is the radius of the typical bacterium, which we assume to be spherical.

(a) Show that the amount of oxygen that diffuses into the bacterium is greater than the amount used by the bacterium in metabolism. For simplicity, assume that the bacterium is a sphere.

(b) What conditions does (a) impose on the radius R for the bacterial cell? Compare it with the size of *E. coli*.

• 3.9 Evolutionary time scales

To get a sense of evolutionary time scales, it is useful to draw an analogy with human time scales. If the age of life on Earth is analogous to one year, what would, proportionally, various other time scales be analogous to? Specifically, the time since the origin of animal life? Of primates? Of the genus *Homo*? Of civilization? Since the extinction of dinosaurs? A human life span? (Problem courtesy of Daniel Fisher.)

3.7 Further Reading

Murray, A, & Hunt, T (1993) The Cell Cycle, Oxford University Press. This book has a useful discussion of the cell cycle.

Morgan, DO (2007) The Cell Cycle: Principles of Control, New Science Press. Morgan's book is full of interesting insights into the cell cycle.

Bier, E (2000) The Coiled Spring: How Life Begins, Cold Spring Harbor Laboratory Press. This is the best introduction to developmental biology of which we are aware. This book is a wonderful example of the seductive powers of development.

Carroll, SB (2005) Endless Forms Most Beautiful, W. W. Norton. One of us (RP) read this book twice in the first few weeks after it hit the shelves. From the perspective of the present chapter, this book illustrates the connection between developmental and evolutionary time scales.

Neidhardt, F (1999) Bacterial growth: constant obsession with dN/dt, *J. Bacteriol.* **181**, 7405. Bacterial growth curves are one of the simplest and most enlightening tools for peering into the inner workings of cells. Neidhardt's ode to growth curves is both entertaining and educational.

Dressler, D, & Potter, H (1991) Discovering Enzymes, W. H. Freeman. This book is full of fascinating insights into enzymes.

3.8 References

Alberts, B, Johnson, A, Lewis, J, et al. (2008) Molecular Biology of the Cell, 5th ed., Garland Science.

Andelfinger, G, Hitte, C, Etter, L, et al. (2004) Detailed four-way comparative mapping and gene order analysis of the canine *ctvm* locus reveals evolutionary chromosome rearrangements, *Genomics* **83**, 1053.

Bell-Pedersen, D, Cassone, VM, Earnest, DJ, et al. (2005) Circadian rhythms from multiple oscillators: lessons from diverse organisms, *Nat. Rev. Genet.* **6**, 544.

Belle, A, Tanay, A, Bitincka, L, et al. (2006), Quantification of protein half-lives in the budding yeast proteome, *Proc. Natl Acad. Sci. USA* **103**, 13004.

Berg, HC (2000) Motile behavior of bacteria, *Phys. Today* **53**(1), 24.

Blest, AD (1978) The rapid synthesis and destruction of photoreceptor membrane by a dinopid spider: a daily cycle, *Proc. R. Soc. Lond.* **B200**, 463.

Darwin, C (1859) On the Origin of Species, John Murray.

Droz, B, & Leblond, CP (1962) Migration of proteins along the axons of the sciatic nerve, *Science* **137**, 1047.

Eden, E, Geva-Zatorsky, N, Issaeva I, et al. (2011) Proteome half-life dynamics in living human cells, *Science* **331**, 764.

Gotta, LS, Miller, OL, Jr, & French, SL (1991) rRNA transcription rate in *Escherichia coli*, *J. Bacteriol.* **20**, 6647.

Ingolia, NW, & Murray, AW (2004) The ups and downs of modeling the cell cycle, *Curr. Biol.* **14**, R771.

Kalir, S, McClure, J, Pabbaraju, K, et al. (2001) Ordering genes in a flagella pathway by analysis of expression kinetics from living bacteria, *Science* **292**, 2080.

Laub, MT, McAdams, HH, Feldblyum, T, et al. (2000) Global analysis of the genetic network controlling a bacterial cell cycle, *Science* **290**, 2144.

Miller, OL, Jr, Hamkalo, BA, & Thomas, CA, Jr (1970) Visualization of bacterial genes in action, *Science* **169**, 392.

Müller-Hill, B (1996) The *lac* Operon: A Short History of a Genetic Paradigm, Walter de Gruyter. A very stimulating and interesting book to read.

Nakajima, M, Imai, K, Ito, H, et al. (2005) Reconstitution of circadian oscillation of cyanobacterial KaiC phosphorylation *in vitro*, *Science* **308**, 414.

Pollard, TD, & Earnshaw, WC (2007) Cell Biology, 2nd ed., W. B. Saunders.

Pomerening, JR, Kim, SY, & Ferrell, JE (2005) Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations, *Cell* **122**, 565.

Reddy, B, & Yin, J (1999) Quantitative intracellular kinetics of HIV type 1, *AIDS Res. Hum. Retroviruses* **15**, 273.

Rizzoli, SO, & Betz, WJ (2005), Synaptic vesicle pools, *Nat. Rev. Neurosci.* **6**, 57.

Rosenfeld, N, Young, JW, Alon, U, et al. (2005) Gene regulation at the single-cell level, *Science* **307**, 1962.

Schleif, R (1993) Genetics and Molecular Biology, The Johns Hopkins University Press.

Yildiz, A, Forkey, JN, McKinney, SA, et al. (2003) Myosin V walks hand-over-hand: single fluorophore imaging with 1.5-nm localization, *Science* **300**, 2061.

Who: "Bless the Little Beasties" 4

Overview: In which key model systems are introduced

Most organisms on the planet share some fundamental similarities, as epitomized by the near universality of the genetic code. For exploring these kinds of processes, the choice of any cell as a subject should be as good as any other. Yet, for various reasons, often certain organisms become the beloved model systems that propel biological investigation forward. At the same time, many organisms exhibit fantastic specializations such that they may be considered experts in particular types of biological processes. For studying these specialized processes, such as rapid electrical conductance down the squid giant axon, it is almost always best to go to the expert. Many kinds of biological experimentation make other demands on their subjects. The subjects must be willing to grow in laboratory conditions, they must be amenable to genetic or biochemical analysis, and they must not be terribly hazardous to human researchers. These competing demands have resulted in the choice among biological researchers of a certain subset of living organisms that are widely used as model systems. In this chapter, we will explore a brief history of the contributions of several of the most noted model organisms who will reappear time and again throughout the remainder of the book.

> "To see a World in a Grain of Sand, And a Heaven in a Wild Flower, Hold Infinity in the Palm of your hand, And Eternity in an hour."
>
> **William Blake,**
>
> *Auguries of Innocence*

4.1 Choosing a Grain of Sand

The preceding two chapters have celebrated the extraordinary diversity of living organisms on Earth. We have emphasized the fascinating and useful fact that all organisms alive today are part of a common lineage and therefore share many features, underlying their extraordinary diversity. From this perspective, it should be possible to undertake the exploration of biological science using any living organism as a "model system" for study. In practice though, a handful of organisms have been most widely used and proved to be most

informative. In Chapter 2, we introduced *E. coli* as the ultimate experimental workhorse, and certainly for many processes represented in the bacterial domain of life *E. coli* has provided most of our understanding. There are many things that *E. coli* cannot do, such as the kind of sexual reproduction characteristic of multicellular eukaryotes or differentiation into multiple cell types to form a multicellular organism. From each of the different branches of the great tree of life, one or a few organisms have been selected as models that are experimentally tractable and that provide a means of introduction for the curious biologist to the less accessible organisms within their branch. The examples that we have chosen to discuss in this chapter will reappear throughout the book. However, a quick perusal of the biological literature reveals many other model organisms that are used with comparable frequency to those discussed here, and this chapter should be considered introductory rather than comprehensive.

Modern Genetics Began with the Use of Peas as a Model System

In the late nineteenth century at about the same time that Darwin and Wallace were first elucidating the interconnectedness of all living organisms and Louis Pasteur was articulating the biochemical unity of life, a humble Austrian monk, Gregor Mendel, was establishing a new science of genetics. He did this not by focusing on the broad sweep of biology as had these other contemporaneous giants, but rather by focusing in great detail on a single organism, the pea plant *Pisum sativum*. Mendel noticed that variations in biological form and function occurred not just among different species, but also among individuals of the same species such as the pea. Furthermore, plants with a particular trait such as unusually wrinkly peas seemed particularly likely to give rise to offspring that shared that same trait. Through a series of carefully controlled pea breeding experiments, Mendel established the statistical rules governing inheritance and provided compelling and concrete evidence for the idea that large-scale traits of organisms were passed from parent to offspring in the form of "irreducible units" of inheritance that passed unaltered.

Peas were Mendel's model organism of choice not merely because of their central role in the monastery kitchen garden, but also because they have a short life cycle, a wide variety of easily measured heritable traits, and ease of large-scale stable cultivation. These same traits have been sought by subsequent generations of biologists in the choice of many other model systems that we will explore, ranging from bacteriophage to bacteria to yeast to fruit flies. The choice of a model system is, of course, only one step towards exploring the mechanistic basis of complex biological phenomena. The organism must lend itself to certain kinds of experimentation as detailed below.

4.1.1 Biochemistry and Genetics

What makes an organism a useful model organism? Throughout the twentieth century, a major focus of biological research was on understanding the underlying molecular basis for complex life processes. This meant that a useful model organism must in one way or another facilitate the process of making connections between events at the molecular scale and events at the level of whole organisms. Two very general complementary approaches encompass most of this work: genetics and biochemistry. Genetics is the study of heritable

variation and takes advantage of the fact that living organisms give rise to offspring that are typically similar but not identical to themselves. Biochemistry demands the isolation of a component of a living organism that retains some interesting function. This takes advantage of the fact that the structure of organisms is divisible.

The goal of both biochemistry and genetics is to identify the molecules that are key players in important biological processes. But the approach of these two disciplines to molecular identification is nearly opposite. In the field of genetics, important molecules are identified within the context of an enormously complex living organism. By removing one component at a time and then asking whether the organism can still perform the process of interest, this procedure will determine whether a particular gene product is *necessary* for the process under investigation. In the complementary approach of biochemistry, the complex organism is subdivided into simpler and simpler subsets or fractions and each fraction is then tested to determine whether it can perform the process of interest. In many classic cases, a single polypeptide chain, laboriously purified from an organism, can perform an important biological process completely on its own in a test tube. This kind of analysis can be used to determine whether a particular molecule or set of molecules is *sufficient* to perform the process of interest. In the modern era of molecular analysis, we consider a process to be "understood" in terms of its molecular players if we can identify all molecules that are necessary and sufficient.

EXPERIMENTS

Experiments Behind the Facts: Genetics The practice of genetics involves identification and isolation of genes responsible for heritable variations within a species. In effect, this is applied Darwinism in action since it takes advantage of the two pillars of the theory of evolution, namely, (i) heritable variation and (ii) selection. One of the most familiar long-term genetics projects undertaken by humans has been the creation of hundreds of distinct dog breeds from a common wolf-like ancestor. In the laboratory, the first step in a genetics experiment is usually the isolation of a mutant that has either lost or gained some property of interest to the experimenter. For example, isolation of a fruit fly mutant with white instead of the normal red eyes can be the first step towards characterizing the pathway responsible for the generation of red eye color.

The earliest geneticists such as Mendel took advantage of naturally occurring mutants and chose to cross only those individuals carrying the trait of interest. Impatient modern biologists will usually start by deliberately generating a large collection of random mutants and then sorting through them to find those with interesting properties.

Kinds of mutagenesis Generation of mutants implies deliberate damage to DNA. In the mid-twentieth century, this was most commonly done with nonspecific chemical or radiological insults. X-rays or ultraviolet radiation or nonspecific chemical agents such as nitrosoguanidine that react with DNA bases induce random DNA damage that often the cell cannot accurately repair. These agents can cause alterations in single base pairs or small insertions or deletions that may alter the reading frame of a gene. An alternative approach that renders the subsequent steps of locating the site of the mutation much easier is to induce mutagenesis by introducing a transposon, a

small, self-replicating fragment of DNA that will insert itself into the genome of the organism and typically disrupt the expression of genes at the site of insertion. In some cases, a biologist may wish to target the mutations more narrowly in the pool and perhaps explore only sequence variations within a particular gene. For these narrow purposes, many techniques of site-directed mutagenesis have been developed that take advantage of recombinant DNA and polymerase chain reaction (PCR) technology specifically to alter DNA in a designed, nonrandom way.

Selections and screens With the collection of mutants in hand, the next step is to identify the individuals with the desired phenotype. An organism's phenotype consists of its observable properties; in contrast, its genotype is the underlying heritable information that encodes for the observable properties. The easiest way to identify mutants with the desired phenotype is to design an experiment that will kill all of the organisms that lack the desired trait. For example, bacterial mutants resistant to a certain antibiotic can be readily isolated by simply pouring the antibiotic into the bacterial culture and isolating the survivors. This kind of mutant hunt is called a *selection*.

Because it is rare that such a convenient life or death situation can be established, it is more common that geneticists will design *screens*, methods of sorting mutant populations to enrich or identify the mutants of interest. For *Drosophila*, a screen may be as simple as a visual examination of the mutant pool. This approach is sufficient for isolating white-eyed flies or wrinkled peas. More typically, the organisms must be challenged or assayed in some way. For example, identification of *E. coli* mutants that are unable to grow on lactose as a carbon source requires replica plating as shown in Figure 4.1. Clever biochemists, wary of replica plating, have worked out complementary methods by designing chemicals that serve as reporters for functional biological pathways. For example, normal *E. coli* are able to convert the colorless chemical nicknamed X-gal (whose real name is 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) into a blue product if and only if their lactose utilization pathway is intact. When grown on plates containing X-gal, mutant colonies stand out by virtue of their white (clear) color.

Conditional mutants In studying processes that are essential to life (for example, the ability to replicate DNA, the ability to generate ATP, etc.) it is obvious that there must be a catch. Genetics can only study heritable variation in individuals that are able to survive and reproduce. However, these essential life processes are certainly of interest to biological scientists. The indispensable trick that has been developed to allow geneticists to study these essential processes is the generation of "conditional mutants," that is, genes whose mutant phenotype is expressed under only some conditions and not others. A classic example is the isolation of temperature-sensitive mutants in microbial species. Normal *E. coli* can replicate over a wide range of temperatures from below 10°C to over 45°C. Bacteria carrying a mutation in DNA polymerase that alters protein stability such that the polymerase is functional at a low temperature

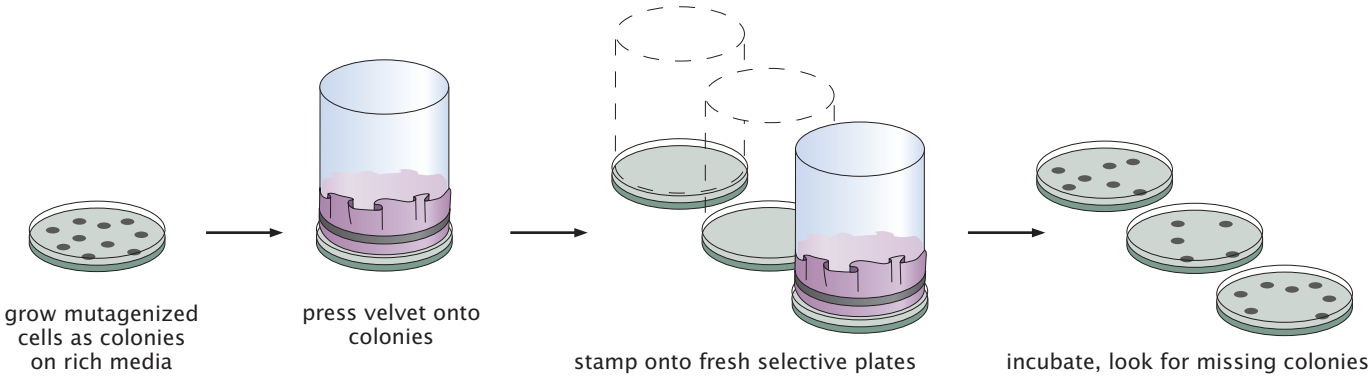


Figure 4.1: Replica plating method for isolating bacterial mutants. In these experiments, an entire distribution of colonies is lifted off one plate and put onto other plates that have some different nutrients. By comparing the original and secondary plates, it is possible to find colonies that did *not* grow on the secondary plates and identify them as mutants in some characteristic of interest.

(30 °C) but unstable at a high temperature (40 °C) can be propagated at low temperature even though they carry a lethal mutation in an essential product. Modern implementations of this approach involve construction of altered genes whose products are only expressed when the organisms are given an artificial chemical signal.

Experiments Behind the Facts: Biochemistry The practice of biochemistry is complementary to that of genetics and involves an application of reductionism, the scientific philosophy that complex processes can be broken down into simpler constituents. One of the earliest applications of biochemistry followed up the brilliant insight that the desirable ability of yeast to convert sugar into alcohol does not require the presence of living yeast cells, but rather only the presence of a few of its critical proteins that catalyze this chemical transformation. The central philosophy of biochemistry can be summarized as "grind and find" as exemplified by putting an organism into a blender and sorting out the molecules of the resulting organismal smoothie to find the one responsible for the process of interest.

Assay development The most important feature of a biochemical project is the development of an appropriate assay, a testing mechanism to determine whether a given test tube contains the molecule or molecules responsible for a process of interest. For example, we mentioned above the chromogenic substrate X-gal, which turns blue when acted upon by an enzyme in the *E. coli* pathway for lactose utilization. Besides being used on plates to assay the phenotypes of intact cells, X-gal can also be added to a test tube containing a purified protein and the development of blue color can be measured quantitatively. This is a simple example. Most biochemical assay development requires substantially more effort and ingenuity.

Fractionation With an assay in hand, the next necessary step is to subdivide the components of the organism. Numerous techniques have been developed for separating molecular constituents on the basis of their chemical or physical properties. Among the most powerful of these are the many varieties of chromatography. The basic idea is to pass a complex mixture of

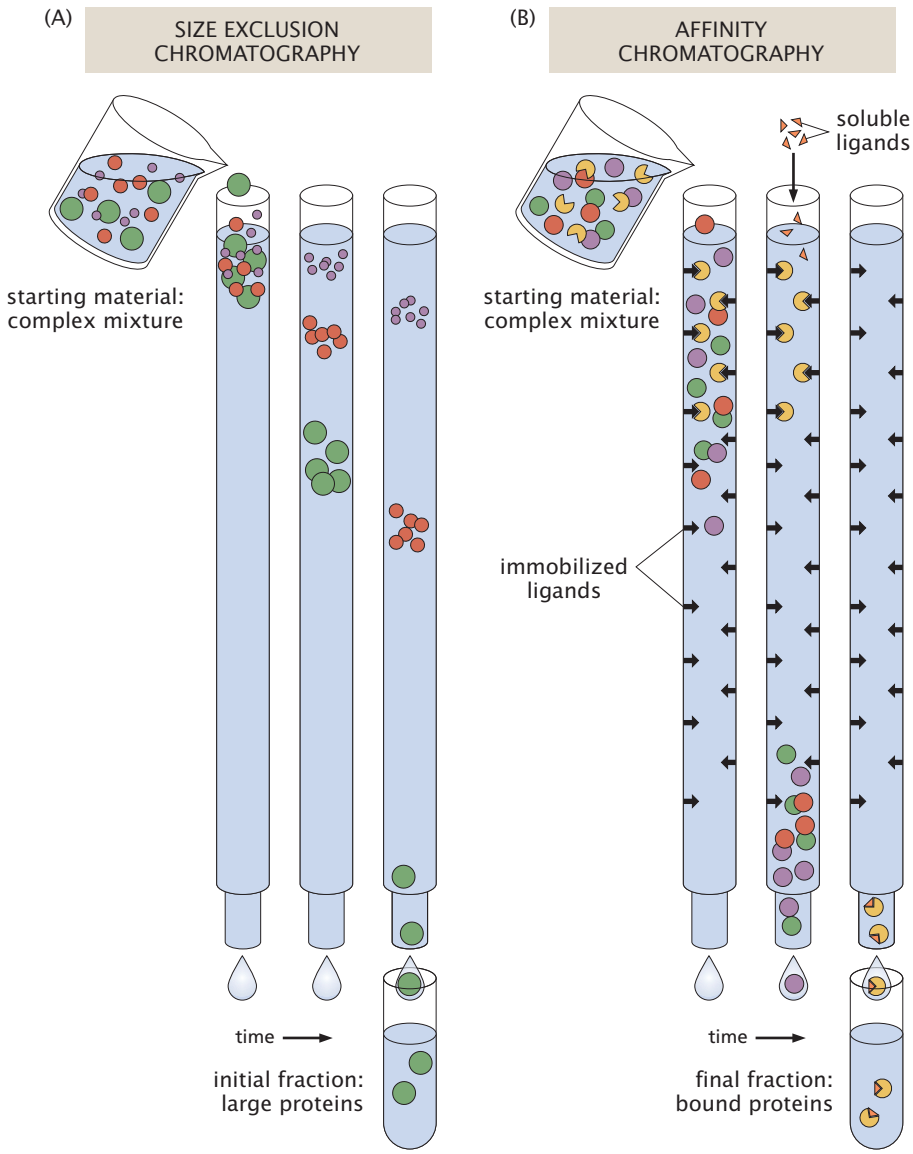


Figure 4.2: Biochemical columns. The column filters different molecular species by virtue of the different speeds that these species have as they traverse the column. (A) In size-exclusion chromatography, the column is filled with a matrix of small beads that have many pores of varying sizes. When a mixture of molecules of different sizes is poured in at the top of the column, the largest molecules will flow through fastest because they cannot enter the small pits and holes in the beads. The smallest molecules will move slowest. (B) In affinity chromatography, the matrix is chemically conjugated to a molecule that will bind to the protein of interest. When a complex mixture is poured into this kind of column, most of the proteins will flow through immediately and only those that can bind to the ligand in the column will be retained. To elute the protein of interest from the column, a soluble form of the same ligand can be added. This will displace the protein from the column matrix and allow it to be collected as a fraction at the bottom.

components through a matrix of material where some components are more free to move than others. A useful geometry is the biochemical column shown in Figure 4.2, in which a small volume of material is added to the top and fractions are collected from the bottom. Molecules that interact only weakly with the material of the column pass through quickly and can be collected in early fractions. Molecules whose transit time is delayed appear at the bottom, only later. Different kinds of columns can separate different molecules based on their physical size, net charge in solution, ability to bind to particular molecular targets, etc. By using a series of fractionation steps operating on different principles, it is frequently possible to isolate a single molecular species or complex that can perform the desired function.

Armed with a sense of two of the central experimental pillars of biology, we now explore some of the hall-of-fame model systems that have led to the emergence of modern biology and that can serve as testbeds for physical biology as well.

4.2 Hemoglobin as a Model Protein

This section explores the question of how the study of the structure and function of the hemoglobin molecule has served as one of the dominant themes in biochemistry for more than a century. As noted in the previous section, genetics and biochemistry have served as two of the driving forces behind the emergence of modern biology. One of the most significant goals of biochemistry has been the dissection of the structure and function of all of the molecules of life. This quest has been spectacularly successful, and can be illustrated through the case study of hemoglobin, a molecule we will appeal to repeatedly to illustrate different ideas in physical biology.

4.2.1 Hemoglobin, Receptor–Ligand Binding, and the Other Bohr

The Binding of Oxygen to Hemoglobin Has Served as a Model System for Ligand–Receptor Interactions More Generally

The function of macromolecules in the living world emerges from interaction. That is, it is only in their relations with one another (Lac repressor and DNA, hemoglobin and oxygen, lipids and water, etc.) that the molecules of life begin to be animated with the first shadowy resemblance of their role in the living realm. Figure 4.3 shows some examples of receptor–ligand linkages that will be revisited throughout the book. Klotz (1997), himself paraphrasing Paul Ehrlich, noted "a substance is not effective unless it is linked to another." Probably the most studied example of such linkage is that of the hemoglobin (Hb) molecule in its partnership with oxygen. Hemoglobin is present in red blood cells and is actively responsible for transporting oxygen to cells. Target tissues such as muscle also contain oxygen-binding proteins. The most abundant of these is myoglobin. Whereas hemoglobin has four binding sites for oxygen molecules, myoglobin has only one. As we will see below, this structural difference between the two proteins causes very interesting functional differences. This cooperative relationship (ligand (O_2) and receptor (Hb)) is but one example of the nearly endless variety of such relationships that occur in biological chemistry. From a physical model-building viewpoint, the key questions that will occupy our attention center on estimating the fraction of receptors that will be bound by ligands as a function of ligand concentration and external conditions such as pH and temperature.

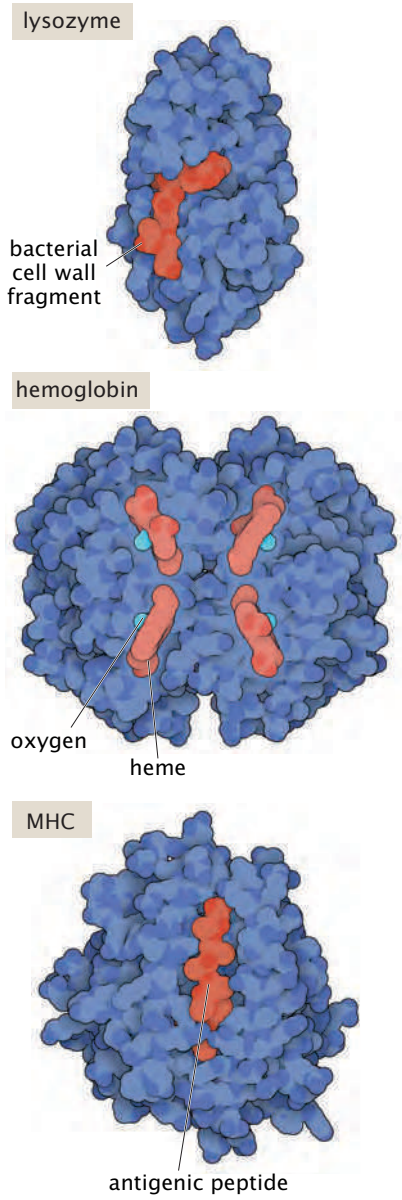


Figure 4.3: Examples of ligand–receptor interactions. From top to bottom the proteins are lysozyme (PDB 148l), hemoglobin (PDB 1hho), and major histocompatability complex protein (MHC, PDB 1hsa). Lysozyme binds molecules making up the bacterial cell wall resulting in its degradation, hemoglobin binds oxygen, and, in its role in the immune system, MHC presents antigens at the cell surface. (Courtesy of D. Goodsell.)

ESTIMATE

Estimate: Hemoglobin by the Numbers There are hundreds of millions of hemoglobin molecules in each of the red blood cells that populate our bodies. To get a feeling for the numbers, we examine the results of a complete blood count (CBC) test at the doctor. Such tests report a number of interesting features about our blood. For an example of the outcome of such a test, see Table 4.1 in the problems at the end of the chapter. One of the outcomes of such a test is the observation that typically there are roughly 5×10^6 red blood cells per microliter of blood and that there is roughly 15 g of hemoglobin for every deciliter of blood (the so-called mean corpuscular hemoglobin (MCH)). If we estimate that there are roughly 5 L of blood in a typical adult human, this amounts to roughly 25×10^{12} red blood cells, containing in total some 750 g of hemoglobin. Given that the mass of a hemoglobin molecule

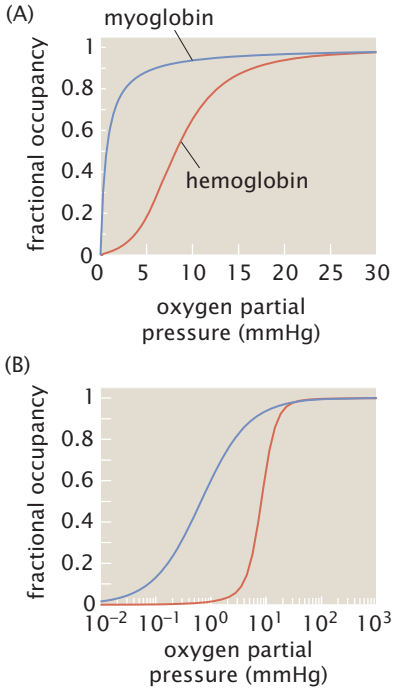


Figure 4.4: Binding curves for oxygen uptake by myoglobin and hemoglobin. The fractional occupancy refers to the fraction of available oxygen-binding sites that are occupied in equilibrium as a function of the oxygen concentration as dictated by the pressure. (A) Binding curves on a linear scale. (B) Binding curves on a log–linear plot. The curves were generated based on experimental data. (Data for hemoglobin from K. Imai, *Biophys. Chem.* 37:197, 1990; data for myoglobin from A. Rossi-Fanelli and E. Antonini, *Arch. Biochem. Biophys.* 77:478, 1958.)

is roughly 64,000 Da, we arrive at the conclusion that there are of the order of 7×10^{21} hemoglobin molecules in the blood of the average adult. Finally, these observations from an everyday blood test thus lead to the conclusion that in each red blood cell there are of the order of 3×10^{8} hemoglobin molecules. This large number of molecules trapped inside a membrane results in an osmotic pressure, as we will calculate in Chapter 14. Estimates based on the quantities measured in a blood test are explored further in the problems at the end of the chapter.

Quantitative Analysis of Hemoglobin Is Based upon Measuring the Fractional Occupancy of the Oxygen-Binding Sites as a Function of Oxygen Pressure

One of the key questions that one might ask about the binding relationship between a ligand and its receptor is what fraction of the receptor-binding sites is occupied as a function of the concentration of ligand molecules. This question is of physiological interest in the setting of hemoglobin because it tells us something about the uptake and delivery of oxygen as a function of the environment within which the hemoglobin finds itself. If a hemoglobin molecule is in an oxygen-starved environment, it has a higher propensity to drop bound oxygen molecules. These insights are captured quantitatively in binding curves like those in Figure 4.4, which show the fractional occupancy of sites in both hemoglobin and myoglobin as a function of the oxygen pressure. What we note in such curves is that when the concentration of oxygen is very high, essentially all of the available sites on hemoglobin (or myoglobin) are occupied by O_2 molecules. By way of contrast, in those environments such as active muscle where the concentration of oxygen is "low" (as defined precisely by the binding curve itself), the hemoglobin molecule has largely delivered its cargo.

Interestingly, much of the work that ushered in the quantitative approach to hemoglobin structure and function was carried out by Christian Bohr, a Danish physiologist, perhaps more familiar to some of our readers as the father of Niels Bohr. Indeed, there is an effect associated with hemoglobin known as the Bohr effect. This effect refers to data like those shown in Figure 4.5, where there are shifts in the oxygen-binding curve as a function of the pH. The Bohr effect is of interest to athletes since it reveals the fact that in a more acidic environment (like starved muscles), the competition of H^+ to bind to hemoglobin makes the release of O_2 even more likely than at normal pH. From the point of view of the present book, the beauty of measurements like those shown in the figure is that they were instrumental in driving research aimed at quantitative modeling of protein structure and function and serve as a compelling example of our battle cry that quantitative data demand quantitative models.

4.2.2 Hemoglobin and the Origins of Structural Biology

In parallel with efforts to dissect the functional characteristics of hemoglobin, there was a strong push to understand the structure of this important molecule. The idea of connecting structure and function is such a central tenet of current thinking that it is perhaps difficult to realize how much needs to be known before one can embark upon the attempt to effect this linkage.

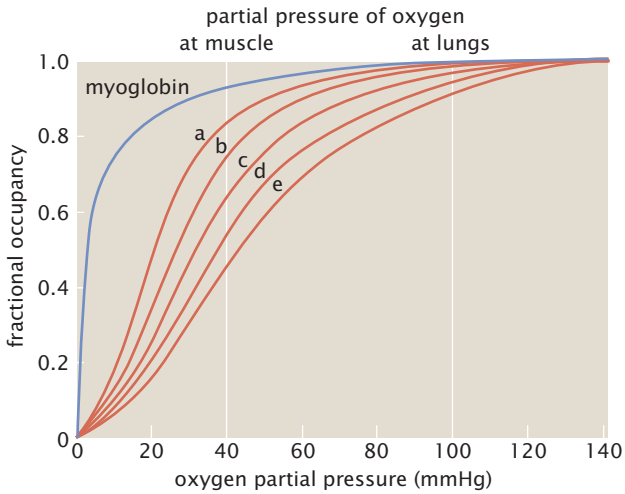


Figure 4.5: Binding curves for oxygen uptake by hemoglobin as a function of pH revealing the Bohr effect. The hemoglobin binding curves are shown for five values of the pH: (a) 7.5, (b) 7.4, (c) 7.2, (d) 7.0, and (e) 6.8. The vertical lines indicate the partial pressures experienced in muscle and in the lungs. (Adapted from R. E. Dickerson and I. Geis, Hemoglobin: Structure, Function, Evolution and Pathology. Benjamin/Cummings, 1983.)

The Study of the Mass of Hemoglobin Was Central in the Development of Centrifugation

One of the first "structural" questions that was tackled (with great difficulty) is what is the mass of the hemoglobin molecule? It is easy to take the answer to such questions for granted. Historically, the proper response to this question was tied to the development of one of the most important tools in the biochemist's and biologist's modern arsenal, the centrifuge. Svedberg's development of the centrifuge represents one of the key technological breakthroughs that permitted the advance of biological science, with particular implications for the debate over the nature of proteins and the constitution of hemoglobin (see Tanford and Reynolds, 2001—referenced in Further Reading at the end of the chapter). In particular, as a result of centrifugation studies, it was concluded that hemoglobin is a tetramer made up of four units each with a mass of roughly 16,000 Da.

Structural Biology Has Its Roots in the Determination of the Structure of Hemoglobin

The Protein Data Bank has obscured the extreme difficulty underlying the determination of protein structure. With a simple mouse click, the structures of thousands of different proteins are available for anyone to see. However, we are the beneficiaries of more than 30 years of dogged effort on the part of the founders of structural biology. With reference to the structure of hemoglobin, Max Perutz is the prime hero. One of the most inspiring features of Perutz's battle to obtain the structure of hemoglobin is the length of that battle. From the time (1937) at which he set himself the task of determining this structure to the time of publication of the first structural reports was a period of nearly 25 years during which a great number of technical obstacles were overcome.

The outcome of the structural studies of Perutz (and others) was the recognition that hemoglobin is a tetramer composed of four subunits, two α-chains and two β-chains. The focal point of each subunit is a heme group, which is made up of an iron atom at the center of a porphyrin ring. As noted above, each of these subunits has a molecular mass of roughly 16,000 Da and an amino acid sequence of roughly 145 amino acids (the sequence can be seen in Figure 21.23 on p. 995).

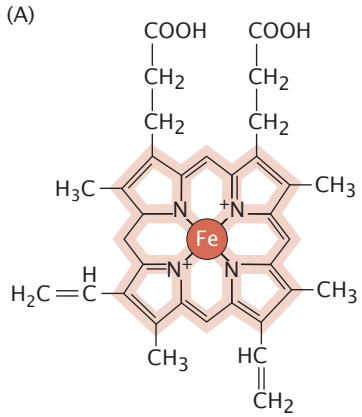


Figure 4.6: Structure of hemoglobin. (A) The heme group includes a porphyrin ring (light red line) that can bind an iron ion. The iron can bind to oxygen or carbon monoxide molecules and is responsible for the red color of blood. (B) The protein hemoglobin consists of four separate polypeptide chains shown in different colors, two α-chains and two β-chains (PDB 1hho). Each one carries a heme group, so the overall protein can bind up to four molecules of oxygen.

The heme groups are the focal point of a functional discussion of hemoglobin because they are the site of O_2 binding. One useful structural description is that each subunit comprises eight α-helical regions that form a box around the heme group as shown in Figure 4.6.

Some of the most important lessons to have emerged from the analysis of protein structures in the hands of Perutz and others can be summarized as follows. (i) The structure of hemoglobin with bound oxygen (oxyhemoglobin) and that of hemoglobin with no bound oxygen (deoxyhemoglobin) are not the same, paving the way for ideas about the connection of structure and function. (ii) The similarities between horse and human hemoglobin are the primitive roots of some of the fundamental questions raised in bioinformatics concerning both sequence and structure homology (a topic we take up in Chapter 21), and strike right to the heart of how structural biologists have contributed to our further understanding of evolution.

4.2.3 Hemoglobin and Molecular Models of Disease

Another fascinating episode in the history of hemoglobin as a model system is the quest to understand sickle-cell disease (sometimes called sickle-cell anemia). This disease is characterized morphologically by severe alterations in the structure of the red blood cell and is manifested physiologically as trapping of the red blood cells in capillaries, cutting off oxygen to some tissues and causing severe pain. Figure 2.21 (p. 59) shows the passage of healthy red blood cells through narrow man-made capillaries. The proposal made by Pauling was that sickle-cell disease is a "molecular disease" resulting from a defect in hemoglobin. The histological manifestation of this defect is the banana-shaped cells characteristic of sickle-cell disease rather than the more familiar biconcave disc shape characteristic of healthy red blood cells (shown in Figure 5.10 on p. 204).

The particular fault with the hemoglobin molecule in the sickle-cell case is the replacement of a charged glutamic acid residue with a hydrophobic valine at one specific location on the β-chains in the hemoglobin tetramer. The formation of the peculiar shapes of the red blood cells in those with the disease is a result of the fact that the hemoglobin molecules aggregate to form fibers. These fibers push on the cell membrane and deform it. Our estimate at the beginning of this section led us to the conclusion that there are 3×10^8 hemoglobin molecules per red blood cell. To get an idea of the mean spacing of these molecules, we can use the formula $d = (V/N)^{1/3}$ (this formula appeared as Equation 2.12 on p. 49), which, taking a volume of 30 μm^3 per red blood cell, leads us to a center-to-center distance between hemoglobin molecules of less than 5 nm, to be compared with a typical protein radius of the order of 2–3 nm. What we learn from these numbers will be revealed as a recurring theme throughout the book (indeed, all of Chapter 14 will discuss these effects), namely, that cells are very crowded. With regard to hemoglobin in red blood cells, the crowding is so high that apparently replacing a charged residue with a hydrophobic residue suffices to drive crystallization of the hemoglobin molecules.

4.2.4 The Rise of Allostery and Cooperativity

One of the related roles in which hemoglobin has served as a model system of paramount importance is in the context of the physical

origins of the curves shown in Figure 4.4. In particular, as will be shown in Chapter 7, the most naive model that one would be tempted to apply to the problem of oxygen binding to a protein does a remarkable job of reproducing the features shown in the *myoglobin* binding curve. However, the sigmoidal shape of the hemoglobin binding curve is a different matter altogether and helped give rise to two key notions, namely, cooperativity and allostery. As noted by Brunori (1999), "I suspect that no other protein will ever be studied in comparable detail: many facets of the biology, chemistry and physics of proteins were addressed initially by researchers working on hemoglobin." As we will show in Chapter 7, cooperativity refers to the fact that binding of ligands on different sites on the same molecule is not independent. For hemoglobin, binding of an oxygen molecule at one of the four binding sites provided by the heme groups effectively increases the likelihood that a second oxygen molecule will bind at a second site, relative to completely deoxygenated hemoglobin. A new molecule of oxygen is even more likely to bind to hemoglobin where two of the oxygen-binding sites are occupied, and still more likely to bind to hemoglobin where three sites are already occupied. Consequently, it is usually the case that a solution of hemoglobin with a limiting amount of oxygen present will contain many individual hemoglobin molecules that have no bound oxygen at all, and some hemoglobin molecules that have four bound oxygens. The idea that the binding of a ligand at one site on a protein can affect the properties of the protein at a distant site is generally referred to as allostery. For hemoglobin, the binding of an oxygen at one site allosterically increases its affinity for oxygen at the other three sites. For many enzymes, the binding of an allosteric activator or inhibitor at one site will affect the enzymatic activity of the enzyme at a distant site. This idea, which has been served immensely by studies on hemoglobin, has turned out to be one of the central ideas concerning the biological uses of binding.

In addition to the notion of cooperativity, studies of hemoglobin have provided the basis for the development of the idea that proteins can exist in multiple conformational states. This is a theme we will develop in detail in Chapter 7 and explains some of the added complexity that proteins adopt as a result of post-translational modifications such as phosphorylation. For enzymes, it is frequently useful to imagine that the protein can exist in several distinct structural states and that it is constantly switching back and forth among these states. One of these structural states can be thought of as the active or functional state, and environmental changes such as tuning the pH or binding of ligands to allosteric sites on these molecules can alter the fraction of time that the molecule spends in each state. The idea that ligand binding can induce conformational change is one that saw concrete experimental realization in the context of hemoglobin and inspired a classic model known as the Monod–Wyman–Changeux (MWC) model, a two-state model of allostery that had as one of its proving grounds the analysis of oxygen-binding curves in hemoglobin. This model will be described in Chapter 7.

4.3 Bacteriophages and Molecular Biology

One of the most enticing classes of biological organisms that has had flagship status as a model system consists of the bacterial viruses known as bacteriophages. The present section highlights a

Figure 4.7: Electron micrograph of the T2 (large) and ϕ29 (small) bacteriophages. (Adapted from S. Grimes et al., *Adv. Virus Res.* 58:255, 2002.)

few representative examples of the way in which bacteriophages have figured centrally (and still do) in the explication of biological phenomena.

4.3.1 Bacteriophages and the Origins of Molecular Biology

Bacteriophages Have Sometimes Been Called the "Hydrogen Atoms of Biology"

The choice of model organisms often reflects experimental convenience. Bacteriophages and their bacterial hosts were so experimentally convenient that together they served to launch modern molecular biology. The use of bacterial viruses as model organisms resulted from a broad variety of such conveniences, including a 20 minute life cycle and the fact that huge quantities (in excess of 10^{10}) of such viruses could be assayed simultaneously. Further, bacteriophages have relatively compact genomes (usually much less than 1 Mbp) and are constructed from a limited parts list. Because of this simplicity, Max Delbrück referred to bacteriophages as the "hydrogen atom of biology," a reference to the fact that in physics the hydrogen atom is the simplest of atoms and that lessons from this simple case are instructive in building intuition about more complex atoms as well. The basic architectural element of the bacteriophage is the presence of a protein shell (the capsid) that is filled with genetic material. Several examples of bacteriophages are shown in Figure 4.7, which has been dubbed the phage family portrait, though the diversity of bacterial viruses goes well beyond the two examples shown here. This particular electron micrograph shows the morphology and relative size of the T2 and ϕ29 phages.

Experiments on Phages and Their Bacterial Hosts Demonstrated That Natural Selection Is Operative in Microscopic Organisms

The unity of life is an idea that we now take for granted. Nevertheless, less than 150 years ago, it was not clear that microbes were even "cells." As will be discussed in more detail in Chapter 21, only in the latter half of the twentieth century were the Archaea discovered. It has always been a challenge to try to figure out the relations between the world of microscopic living organisms on the one hand and the macroscopic living organisms of everyday experience such as giraffes and redwood trees on the other.

One of the hard-won insights into our understanding of microbes and the viruses that infect them had to do with the applicability of the ideas of evolution to microorganisms. The question was: are microbes subject to variation and selection in the same way as larger organisms such as peas and fruit flies? The answer to this question came from a famous experiment by Delbrück and Luria known as the "fluctuation test" that relied on bacteriophages.

Their reasoning was based on the observation that infection of a culture of bacteria with bacteriophages led to some subset of cells that were resistant. Their aim was to determine whether this resistance arose as a result of spontaneous mutations or was somehow induced by the phages themselves. If the resistance was induced, they reasoned that if they diluted some original culture into several different cultures and infected them with phages, roughly the same number of resistant colonies should be found per culture. On the other hand, if

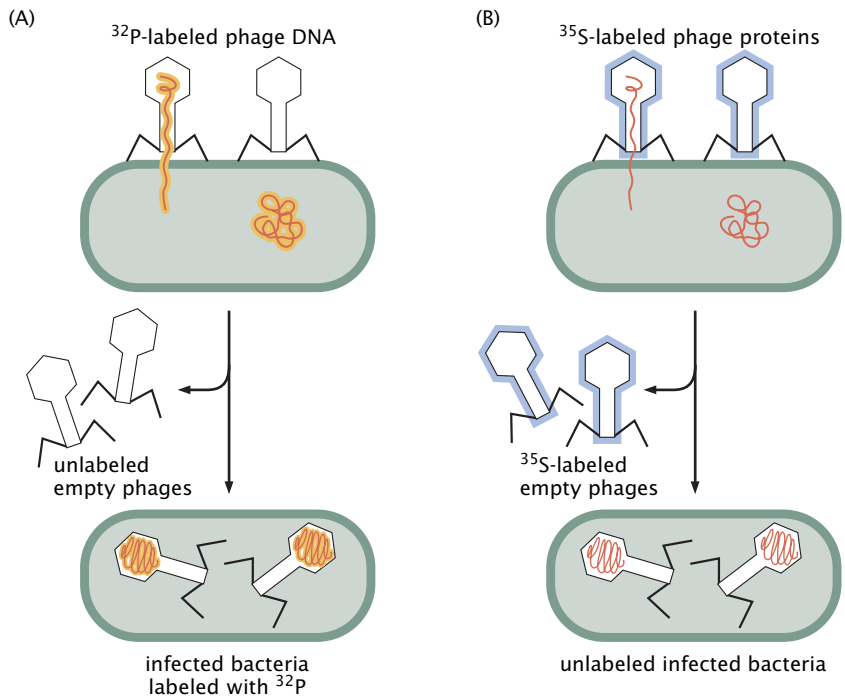


Figure 4.8: The Hershey–Chase experiment. (A) For the case in which the nucleic acid of the phage is labeled, radioactive phosphorus is found in the infected bacteria. (B) For the case in which the protein of the phage is labeled, no radioactivity is found in the infected bacteria.

the resistance arose spontaneously, there should be a large variation in the number of resistant colonies from one culture to the next. The latter eventuality is what was observed, demonstrating the important role of spontaneous mutations as the raw material for evolution, even in the case of microscopic organisms.

The Hershey–Chase Experiment Both Confirmed the Nature of Genetic Material and Elucidated One of the Mechanisms of Viral DNA Entry into Cells

One of the classic experiments in the history of molecular biology that exploited bacteriophages is the so-called Hershey–Chase experiment shown in Figure 4.8. The idea of the experiment is to build phage particles with packaged DNA that is radioactively labeled with the phosphorus isotope ^{32}P and proteins associated with the capsid containing the sulfur isotope ^{35}S. These labeled phages are used to infect bacteria and then the phages and the bacteria are separated by vigorous stirring in a blender. By querying the bacteria to find out which radioactive species they contain, it is possible to determine whether protein or DNA is the "genetic material" of the viruses themselves.

The Hershey–Chase experiment was a direct and convincing experimental confirmation that DNA (and not proteins) carries the genetic material. However, in addition to this deep-rooted insight with implications for all of biology, several more specialized insights emerged from this experiment as well. First, this experiment provides insights into the nature of viral infection, at least in the case of bacteriophages with the viral DNA delivered to the bacterial interior while the protein shell remains outside. In Chapters 10 and 16, we will explore how physical reasoning can be used to understand viral packing and ejection, respectively. A second powerful set of insights embodied in this experiment concern the nature of viral self-assembly, in particular, and biological self-assembly, more generally. Specifically,

this experiment showed that the synthetic machinery (for both DNA and proteins) of the host cell had to be hijacked to do the bidding of the virus; namely, it had to make copies of the viral DNA and express the proteins required for the assembly of virulent new phage particles.

Experiments on Phage T4 Demonstrated the Sequence Hypothesis of Collinearity of DNA and Proteins

Much of what we now "know" and take for granted was hard won in the knowing. Despite the ease with which every high-school science student can rattle off the connection between the two great polymer languages, the collinearity of DNA and protein, itself an article of belief of the central dogma, was at the start nothing but hypothesis. It was the work of Seymour Benzer on phage T4 that gave this hypothesis its first steps towards being a theory.

Benzer's phage experiment was in many ways the bacteriophage partner of the types of experiments that had already been perfected by Thomas Hunt Morgan and his chromosomal geographers whose work on fruit flies will be discussed later in this chapter. Benzer used mutants of a particular gene known as *rII* of bacteriophage T4. At this point in time, the idea that genes were physical entities arranged in a linear fashion along chromosomal DNA had already been established by the work of Morgan, but most people tended to think of individual genes as being point objects that could be either normal or mutant. In Benzer's experiment, he examined what could happen when two different mutant copies of a gene were allowed to recombine with one another. He found that recombination events could glue the healthy parts of the sequences back together, resulting in a wild-type phage. The conclusion to be drawn from the recovery of wild-type phages from mutants was that in some cases mutations could exist at different locations within a single gene. Benzer could map out the geography of the *rII* gene by measuring the frequency which with wild-type T4 phages were restored between multiple pairwise combinations of distinct mutations. The more likely the restoration, the more distant the mutations within the *rII* gene.

Most fundamentally, Benzer showed experimentally that individual genes are endowed with physical extent (that is, some region of base pairs along the DNA double helix) exactly as was demanded by the model of DNA structure proposed by Watson and Crick. There was a long gap of roughly 20 years between Benzer's gene mapping experiments and the explicit sequencing of DNA, again involving a bacteriophage, though this time the phage ϕX174.

The Triplet Nature of the Genetic Code and DNA Sequencing Were Carried Out on Phage Systems

Brenner and Crick, inspired by Benzer's mapping experiments, saw that these same methods could provide an answer to the fundamental coding problem: is the genetic code built up of triplet codons? The triplet codon hypothesis was predicated on the idea that a given gene is read in groups of three letters starting from some unique point. Though seemingly arcane, it was also suspected that the chemical acridine yellow caused mutations in DNA by either adding or deleting a single base. The realization of Crick and Brenner was that if

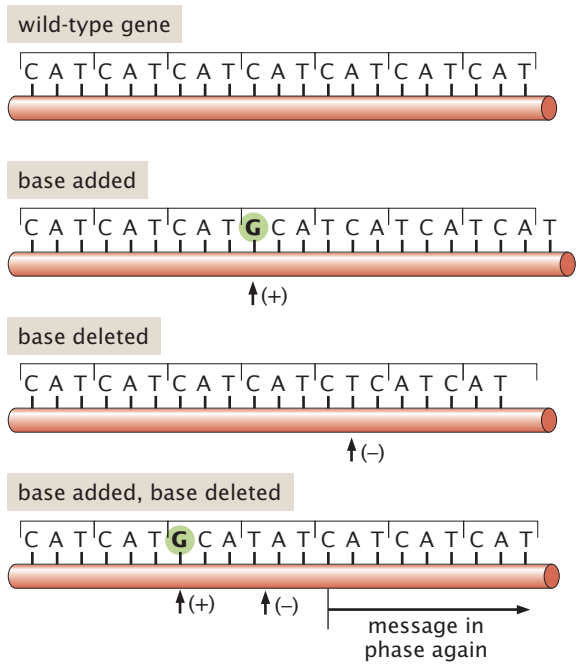


Figure 4.9: Cartoon showing the restoration of phase in DNA sequences through collections of insertion and deletions. (Adapted from H. F. Judson, The Eighth Day of Creation. Cold Spring Harbor Laboratory Press, 1996.)

the triplet hypothesis was correct, then a collection of three insertions or deletions (or alternatively an insertion and a deletion) should restore nonfunctional genes to functional status since they would restore the reading of the gene to the correct, original phase as shown in Figure 4.9. As a result, by mixing and matching insertion and deletion mutants, it was possible to show that the genetic code was functionally based on collections of triplets.

Phages Were Instrumental in Elucidating the Existence of mRNA

Though it is easy for any reader of a textbook to quote with authority the various biological roles of RNA, it is quite another matter to demonstrate such assertions experimentally, especially for the first time. Another of the classic experiments that ushered in molecular biology on the heels of phage research and that served to strengthen the experimental backdrop to the central dogma concerned the functional role of what is now known as mRNA (messenger RNA).

One of the early pieces of evidence in favor of the idea that mRNA served as an intermediate in protein synthesis came from the observation that in cells infected by a phage, there was an unstable RNA fraction (as opposed to the relatively stable RNA that makes up the ribosome) with a similar base composition as the phage DNA itself. In an experiment carried out by Brenner, Jacob, and Meselson, T4 phage infection was used to determine whether this second class of RNA was or was not related to the ribosome structure. The idea was that if the RNA synthesized after viral infection were exclusively related to ribosome structure, then this RNA should only be found in ribosomes synthesized after infection. Alternatively, if the RNA was a messenger, then it should be found in association with old and new ribosomes alike. By employing growth media in which different isotopes of nitrogen and carbon were used, heavy and light ribosomes would be synthesized and could be separated by density gradient centrifugation. The result of the experiment was the observation that the viral-derived RNA was associated with the heavy ribosomes (old) and

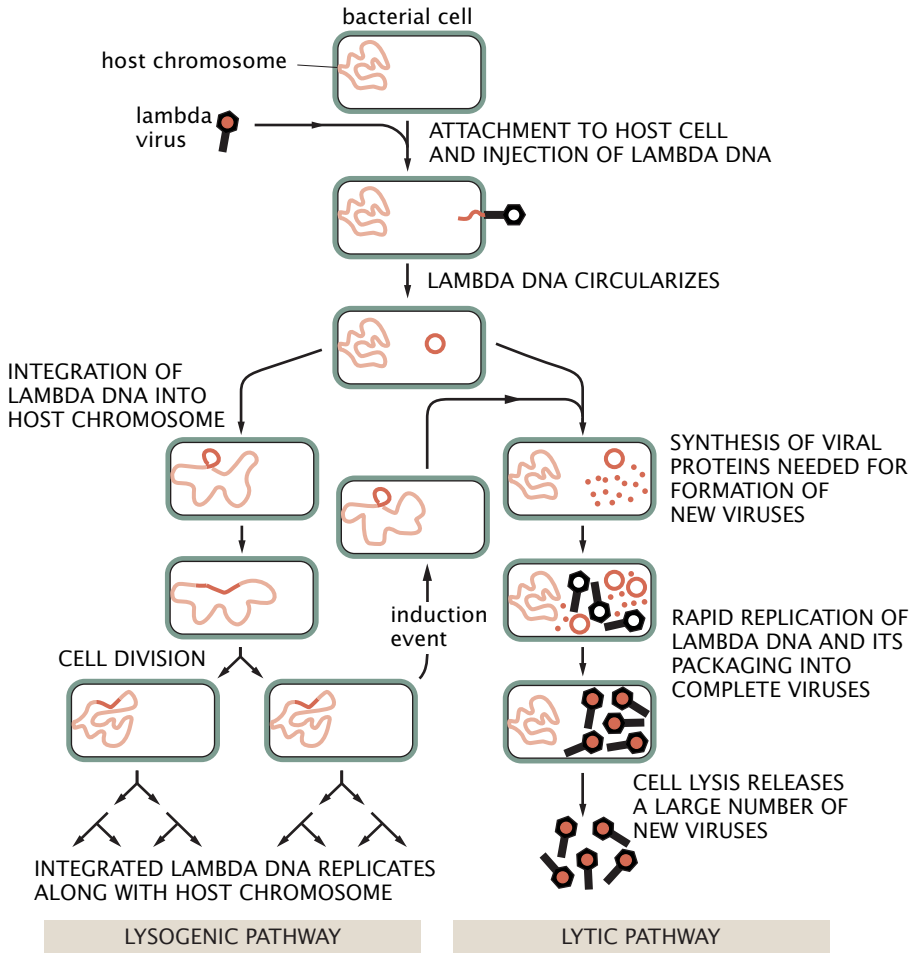


Figure 4.10: Lifestyles of bacterial viruses. Lysogenic and lytic pathways for bacteriophage lambda. In the lysogenic pathway, the viral genome is incorporated into the bacterial genome to form a "prophage." In the lytic pathway, new viruses are assembled to completion and are able to repeat their deeds elsewhere after being released from the lysed cell. (Adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed. Garland Science, 2008.)

hence served in an informational rather than structural role. For the details of this fascinating episode, see Echols (2001—referenced in Further Reading at the end of the chapter).

General Ideas about Gene Regulation Were Learned from the Study of Viruses as a Model System

We have already described the viral life cycle of certain bacteriophages in rough terms in Figure 3.26 (p. 122). However, one of the subtleties left out of that description is the observation that in certain instances, after infection, the viral genome can be peacefully incorporated into the host cell genome and can lie dormant for generations in a process known as lysogeny. These two post-infection pathways are shown in Figure 4.10. As a result of observations on the phage switch, it was wondered how the "decision" is made either to follow the lytic pathway, culminating in destruction of the host cell or, alternatively, to follow the lysogenic pathway with the result that the viral genome lies dormant. The pursuit of the answer to this question, in conjunction with studies of the *lac* operon (to be discussed in Section 4.4.2), resulted in profound insights into the origins of biological feedback and control and the emergence of the modern picture of gene regulation.

The key idea to emerge from experiments on the lifestyle choices of bacteriophages (and the metabolic choices of their bacterial hosts) is the idea that there are genes whose primary purpose is to control

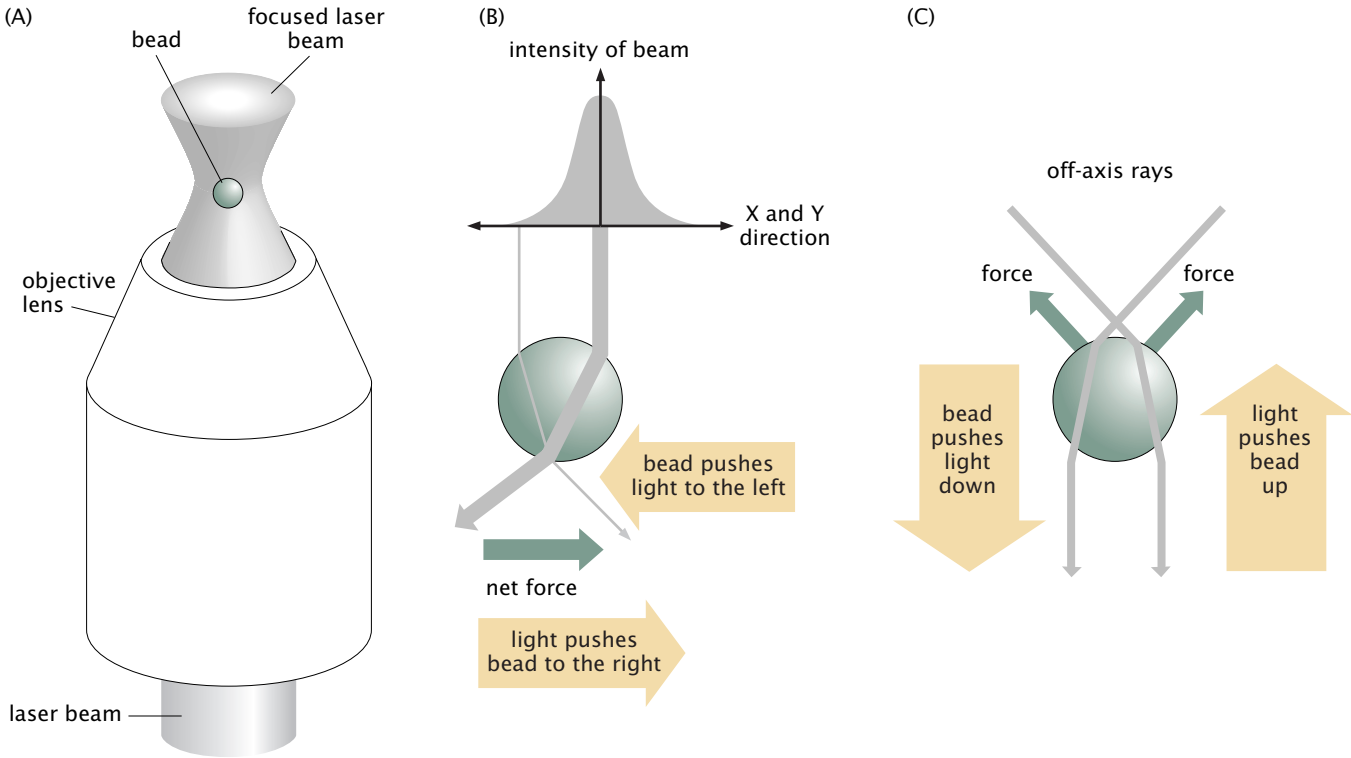


Figure 4.11: Schematic of optical tweezers. (A) A simple optical trap can be made just by focusing a laser beam through a high-numerical-aperture microscope objective. A glass or plastic bead of the order of one micron in size will be held at the focal point of the laser light. (B) In cross-section, the focused light has an approximately Gaussian intensity profile. If the bead drifts slightly to the left, the imbalanced force of the refracted light will push it back toward the center of the beam. (C) Similarly, when the bead drifts down with respect to the focal point of the laser light, the unbalanced force from the refracted light will push it back up.

other genes. Using physical biology to dissect these processes will serve as one of the threads of the entire book and will be the centerpiece of Chapter 19. In the bacteriophage setting, the examination of the decision of lysis versus lysogeny led to the emergence of the idea of genetic switches, controlled at the molecular level by a host of molecular assistants (known as transcription factors) that turn genes either off (repressors) or on (activators). Regulatory decisions are clearly one of the great themes of modern biology and the emergence of this field owes much to work on bacteriophages and their bacterial hosts.

4.3.2 Bacteriophages and Modern Biophysics

As with many of the systems described in this chapter, just when it seems that a particular model system has exhausted its usefulness, like a phoenix from the ashes, these model systems reemerge in some new context providing a model basis for some new class of phenomena. Use of bacterial viruses has seen a renaissance as a result of progress on two disparate fronts, namely, (i) the successes of structural biologists in crystallizing and determining the structures of both individual parts of viruses and entire viruses (illustrated in Figure 2.29 on p. 66) and (ii) the emergence of the discipline of single-molecule biophysics, which has permitted the investigation of processes attending the viral life cycle.

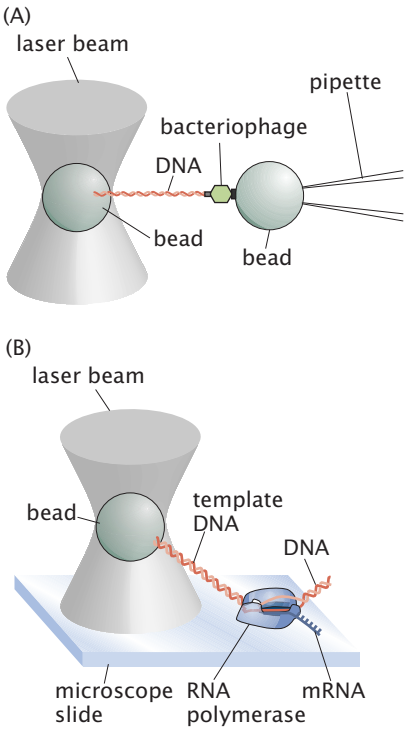


Figure 4.12: Cartoon showing two single-molecule experiments based on bacteriophages. (A) Single-molecule analysis of DNA packaging in bacteriophage $\phi 29$. The molecular motor in the phage pulls the viral genome into the capsid against the force produced by the optical trap. (B) Single-molecule studies of transcription using RNA polymerase from bacteriophage T7. During the process of *in vitro* transcription, the DNA molecule is pulled into the immobilized RNA polymerase. Since the bead is held in the optical trap, the forces produced by RNA polymerase can be measured.

Many Single-Molecule Studies of Molecular Motors Have Been Performed on Motors from Bacteriophages

One of the most exciting experimental advances of recent decades has been the advent of a host of single-molecule techniques for querying the properties of individual molecular actors in the drama of the cell. Some of the most beautiful applications of these ideas have involved individual bacteriophages and their particular gene products. One class of single-molecule experiments centers on the idea of holding a particular molecule (or set of molecules) during some process such as the packing of viral DNA into the capsid and examining how the applied force alters the ensuing dynamics.

The use of optical tweezers such as shown in Figure 4.11 permits the application of piconewton forces over nanometer length scales to a variety of molecular machines. Though there are many different single-molecule techniques available, for the moment we consider just the use of light to apply forces. The basic idea of the optical tweezers method is relatively simple: radiation pressure from laser light is used to apply forces to a micron-scale glass bead. The trick is to attach this bead to the molecule of interest such that the motion of the bead provides a surrogate for the dynamics of the macromolecule of interest. These linkages between bead and molecule are implemented using a molecular analog of Velcro. One very popular and strong Velcro-like linkage exploits the ability of the protein streptavidin to bind very specifically to the small vitamin molecule biotin. This interaction is widely used in the laboratory because biotin can be easily chemically coupled to nearly any macromolecule of interest. By applying known forces to the bead, the mechanochemistry of the attached macromolecule can be studied as a function of the applied load.

Using optical tweezers, a number of brilliant *in vitro* experiments have been performed on biological processes relevant to the bacteriophage life cycle. One such experiment involves the measurement of the forces that are operative when the portal motor is in the process of packing the viral DNA into the capsid. A schematic of the experiment is shown in Figure 4.12(A). This class of experiments reveals the build up of a resistive force as more and more DNA is packaged into the capsid. We will use simple ideas from electrostatics and elasticity to compute these forces in Chapter 10. A second class of experiments of the single-molecule variety that has been similarly revealing and founded upon the use of bacteriophage constructs involves the measurement of the rate of transcription by T7 RNA polymerase as shown in Figure 4.12(B). As will be seen in the remainder of this book (especially Chapter 16), these experiments provide a direct window onto the dynamics of the molecules that make cells work.

4.4 A Tale of Two Cells: *E. coli* As a Model System

4.4.1 Bacteria and Molecular Biology

We have already seen the important role played by bacteriophages as model systems in the creation of the field of molecular biology. There is an intimate partnership between these phages and the bacteria they infect and perhaps most notably the bacterium *E. coli*. As stated eloquently in the 1980 Nobel Lecture of Paul Berg (Berg, 1993), "Until a few years ago, much of what was known about the molecular details of

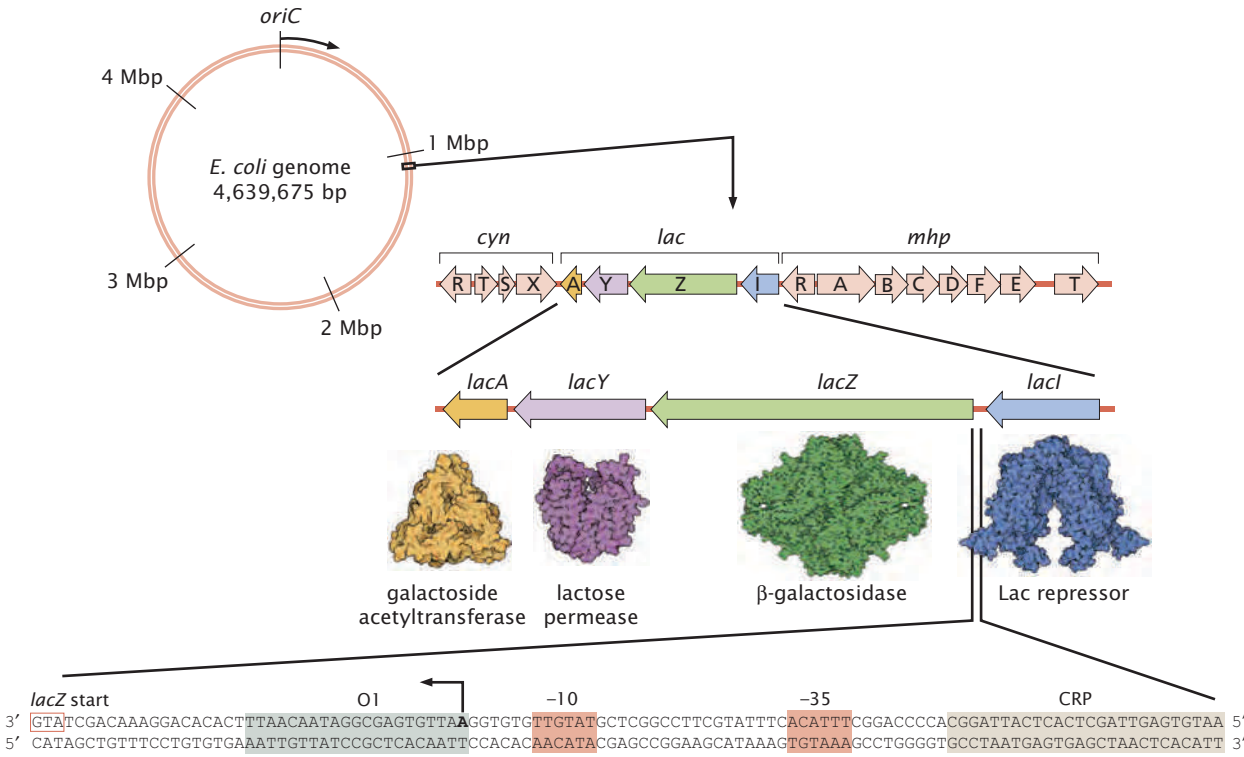


Figure 4.13: Circular map of the *E. coli* genome with a higher-resolution view of the region near the *lac* operon. At the top, the entire genome is represented as a circle. The origin of replication is referred to as *oriC* and base pairs are numbered starting from 1 in the direction of the arrow. The *lac* operon is located about one-quarter of the way around the circle, nestled between the *cyn* and *mhp* operons, whose functions are cyanate utilization and degradation of 3-(3-hydroxyphenyl)propionic acid, respectively. The protein-encoding genes in the *lac* operon are oriented such that they are all transcribed toward the origin. Each of the four genes *lacI* (pdb 1LBH, 1EFA), *lacZ* (pdb 1BGL), *lacY* (pdb 1PV7), and *lacA* (pdb 1KRV) encodes a distinct protein involved in lactose utilization. The proteins are not all drawn to the same scale since β-galactosidase, for example, is so large. The DNA sequence around the beginning of the *lacZ* gene is shown at higher magnification. The start of transcription is indicated by the arrow. The codon for the first amino acid in the sequence (methionine) is boxed. The two binding sites for RNA polymerase are indicated with red shaded boxes at the −10 and −35 positions counting from the transcriptional start site. Two important regulatory sites are shown as shaded boxes; O1 is a binding site for the Lac repressor protein (encoded by the *lacI* gene) and CRP is a binding site for another regulatory protein. (Structural illustrations from D. Goodsell.)

gene structure, organization and function had been learned in studies with prokaryote microorganisms and the viruses that inhabit them, particularly, the bacterium *Escherichia coli* and the T and lambdoid bacteriophages." The goal of this section is to explore the various ways in which *E. coli* has earned this status and to show how many of these discoveries now provide the impetus for physical biology.

The structure of *E. coli* was described in Chapter 2 and some of the key processes that attend its life cycle were described in Chapter 3. These rich structures and processes are dictated by the 4.6 Mbp genome illustrated in Figure 4.13. If we invoke our rule of thumb that each gene encodes for a protein with a length of 300 amino acids, then this amounts to roughly 1000 bp/gene. Using this rule of thumb, we estimate in turn that the *E. coli* genome codes for roughly 4600 distinct proteins. Genome sequencing has revealed that the actual number of genes is 4435, of which 4131 encode proteins, 168 encode structural RNA, and the remainder appear to be nonfunctional. While this is a relatively long "parts list," it is much simpler than the parts list for a jumbo jet and gives us hope that the functioning of the

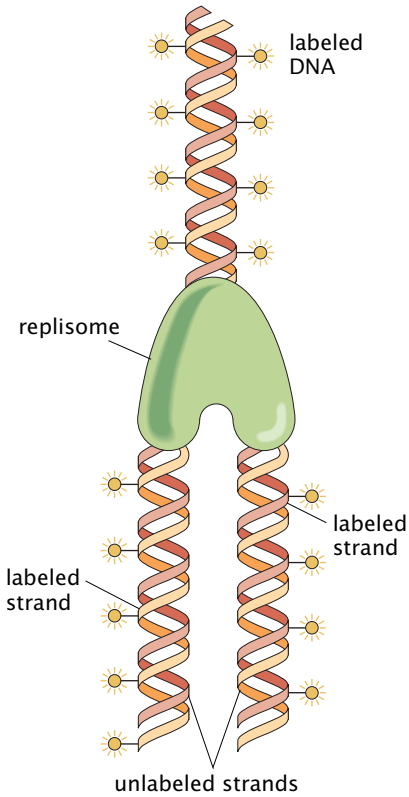


Figure 4.14: The Meselson–Stahl experiment and the hypothesis of semi-conservative replication. When DNA labeled on both strands with a heavy nitrogen isotope is copied by the replisome, each of the two daughter molecules ends up with one heavy strand and one light strand. Note that the schematic intentionally ignores subtleties about the distinction between leading and lagging strands during replication.

E. coli cell may be understandable in terms of the functions of its constituent parts.

4.4.2 *E. coli* and the Central Dogma

The *central dogma* of molecular biology (introduced in Chapter 3 and illustrated in Figure 3.12 on p. 107) specified the mechanism whereby the information content of genomes is turned into the proteins that implement cellular functions. At the outset, many of the ideas associated with the central dogma were speculative and required the first generation of molecular biologists to design and implement a series of clever experiments to work out the processes that connect the structure and information content of DNA to its downstream consequences. Much of that work was carried out using the bacterium *E. coli* and its associated viruses.

The Hypothesis of Conservative Replication Has Falsifiable Consequences

From the moment of its inception, thinking on the structure of DNA was intimately tied to ideas about how the genetic material was faithfully passed on from one generation to the next. One of the hypotheses associated with the development of the central dogma was that of semiconservative replication. That is, when a particular DNA molecule is copied, the two daughter molecules each have a single strand from the parent molecule. This model is clearly experimentally falsifiable and led to one of the most famed experiments in the history of molecular biology, the Meselson–Stahl experiment. Once again (like with the Hershey–Chase experiment), the concept of the experiment was built around the idea of using different isotopes to track the evolution of the parent and daughter molecules. From the perspective of mass conservation, the key idea is that replication involves doubling the total number of nucleotides relative to the parent molecule. One scenario to imagine is that somehow the parent molecule is copied but that the daughter molecule is built up entirely of new nucleotides. Alternatively, in the hypothesis of semiconservative replication, each of the daughter molecules has one strand from the parent and one comprising new nucleotides.

The idea of the Meselson–Stahl experiment was to use nucleotides with the nitrogen isotope ^{15}N, rather than the ordinary ^{14}N isotope. By growing *E. coli* cells in a medium that made available only heavy nitrogen, a culture of cells was grown whose DNA consisted strictly of ^{15}N. By then growing the cells for one generation in a medium in which only ^{14}N is provided, the daughter strands would bear the mark of this change in nitrogen isotopes, which would show up as a difference in mass. The semiconservative and conservative replication hypotheses give rise to different predictions for the observed masses. An example of the distinction between parent and daughter strands is shown in Figure 4.14. To separate the different molecules, the samples were centrifuged in a gradient of CsCl (cesium chloride), with the effect that each DNA molecule found the appropriate equilibrium depth where the buoyant force just matched the downward-acting effective force of gravitation (effective because the force is really due to the centrifugation). What Meselson and Stahl found is that the newly synthesized DNA had a mass intermediate between the ^{15}N and ^{14}N versions of the molecule, supporting the hypothesis of semiconservative replication.

One of the key insights to emerge from the experiment of Meselson and Stahl was the demonstration that DNA replication is a templated polymerization process. That is, the parent strand serves as a template for the daughter and one-half of the parent molecule is donated to each daughter cell. Indeed, from the standpoint of the physics or chemistry of polymerization, this is one of the resounding differences between biological polymers and ordinary synthetic polymers like polyethylene—biological polymers are built from a template and hence, in principle, every copy (as long as no mistakes are made) should have precisely the same sequence as the parent molecule.

Extracts from *E. coli* Were Used to Perform *In Vitro* Synthesis of DNA, mRNA, and Proteins

The understanding of DNA replication on the basis of experiments involving *E. coli* did not stop with Meselson and Stahl. One line of research of particular importance used extracts from *E. coli* cells in order to carry out *in vitro* synthesis of new DNA molecules. In a similar vein, another episode in the history of molecular biology using *E. coli* as the relevant experimental organism centered on the isolation of the enzyme responsible for creating mRNA. The genetic tradition launched by Mendel was consummated in one of the greatest discoveries of modern biology, namely, the cracking of the genetic code, which was also carried out using extracts of *E. coli*, resulting in the understanding embodied in Figure 1.4 (p. 10).

4.4.3 The *lac* Operon as the "Hydrogen Atom" of Genetic Circuits

Gene Regulation in *E. coli* Serves as a Model for Genetic Circuits in General

We have already described the study of lytic control in bacteriophages as one of the two key episodes in the development of the modern view of genetic control. The other such episode, developed concurrently in the Paris laboratories of Jacob and Monod, concerned the genetic decisions made in *E. coli* associated with metabolism of the sugar lactose. The remarkable idea that emerged from these studies is that there are certain genes whose task it is to control other genes. Prior to the advent of the operon concept to be described below, the notion of a gene was generally tied to what might be called a "structural gene," responsible for the construction of molecules such as hemoglobin and DNA polymerase and many others. However, the operon concept called for a new type of gene whose role was to control the expression of some other gene.

The ways in which such control can be effected are wide and varied. One important type of genetic control is known as transcriptional regulation and refers to molecular interventions that control whether or not an mRNA molecule for a gene of interest is synthesized. Negative control (or repression) is mediated by molecules known as repressors that can bind the DNA and inhibit the ability of RNA polymerase to bind and make mRNA. Positive control (or activation) refers to the case in which a gene is generically dormant unless some molecule induces transcription by "recruiting" RNA polymerase to the relevant promoter. Activation is mediated by molecules known as activators that increase the probability of RNA polymerase binding. Note that

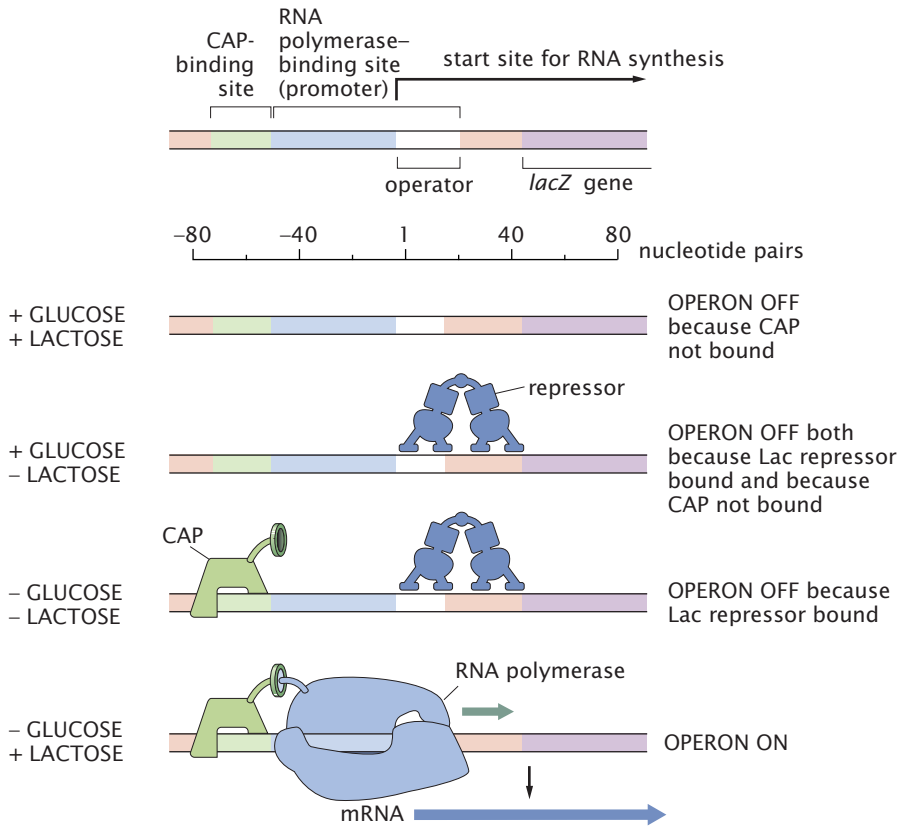


Figure 4.15: Schematic of the regulatory region of the *lac* operon. This figure shows the genography of the *lac* operon in the form of the function and products of different parts of this part of the *E. coli* genome. CAP is the catabolite activator protein, which is an activator of the *lac* operon in the absence of glucose. Combinatorial control by CAP and Lac repressor ensures that the genes of the operon are expressed only when lactose is present and glucose is absent.

there is a wide variety of other mechanisms that regulate gene expression after the polymerase is already bound, though the recruitment mechanism described here will garner the most attention throughout the book.

The *lac* Operon Is a Genetic Network That Controls the Production of the Enzymes Responsible for Digesting the Sugar Lactose

The classic example of negative control is that provided by the *lac* operon shown in Figure 4.15, one of the many profound and lasting insights to have emerged from the study of bacteria in general, and *E. coli* in particular. The work on this operon was so convincing and paradigmatic that it took over a decade for the first example of positive genetic control to be widely accepted. The chemical reactions that are tied to this genetic circuit involve the digestion of the sugar lactose in *E. coli* cells that have been deprived of other sugars and that find themselves in a medium rich in lactose. An example of a growth curve in which cells make a decision to turn on the genes responsible for metabolism of a sugar source other than glucose is shown in Figure 3.3 (p. 93), where we see that once the cells have exhausted the glucose, there is a waiting period during which the genes necessary to use lactose are turned on. The enzyme responsible for this digestion is called β-galactosidase (shown in Figure 4.13) and is substantively present only when lactose is the only choice of sugar available.

The *lac* operon refers to the region in the *E. coli* genome shown in Figure 4.13 that includes the promoter site for the binding of RNA polymerase upstream from the structural genes of the *lac* operon, the operator site where Lac repressor binds and inhibits transcription by

RNA polymerase and then the genes encoding β-galactosidase (*lacZ*), the lactose permease (*lacY*), which builds membrane pores through which lactose can be brought into the cell, and transacetylase (*lacA*), an enzyme that transfers an acetyl group from acetyl CoA to the sugar. (CoA, or coenzyme A, is a small carrier molecule that can be attached to acetyl or acyl groups and is often used as an intermediate in metabolic reactions.) The *lac* operon is shown in Figures 4.13 and 4.15, which give a feel for the number of base pairs involved, the appearance of the relevant gene products, and the relative positions of the various regions in the operon. The region of control is associated with the promoter (with the arrow signifying the start of RNA synthesis) and the operator (the repressor binding site). The promoter is the region where RNA polymerase binds prior to transcription. This region partially overlaps the operator region, which is the site where Lac repressor binds to prevent transcription. The product of the *lacI* gene is the Lac repressor molecule itself. In addition, there is an activator known as CAP (catabolite activator protein) that binds in the vicinity of the promoter and effects positive control. For the present discussion, the most important reason for bringing all of this up is the observation that studies on this seemingly obscure feature of *E. coli* served as a model system for the construction of the modern theory of gene regulation, which has implications for topics as seemingly distant as the emergence of the molecular understanding of development and the development of animal body plans.

4.4.4 Signaling and Motility: The Case of Bacterial Chemotaxis

E. coli Has Served as a Model System for the Analysis of Cell Motility

A typical bacterium such as *E. coli* gets around. Not only that, but just as these cells express metabolic preferences, their motility reveals that they can express preferences about their environments as well. In particular, if some favorable chemical species (sugars, dipeptides, and other "chemoattractants") is pipetted locally into a solution, the bacteria will actually bias their motion so as to swim towards these chemoattractants. Similarly, if repellants are pipetted into the medium, the bacteria will swim away from them. This process of chemotaxis is one of the great stories of biology and is not just some obscure practice of bacteria. Indeed, chemotaxis is part of the social decision making of *Dictyostelium* (see p. 75) and is a key ingredient of the workings of our immune systems as neutrophils hunt down foreign invaders as shown in Figure 3.3 (p. 93).

The study of bacterial chemotaxis in *E. coli* has resulted in the emergence of a beautiful story that relates molecular signaling and cell motility as shown in Figure 4.16. Using the tools of genetics and biochemistry, it has been possible to identify the molecular machinery responsible for chemotactic decision making, with many of the lessons emerging from this study relevant to other organisms as well. The study of chemotaxis and the resulting bacterial motility represents a profound meeting point of subjects, including low-Reynolds-number hydrodynamics, cell signaling, and the study of polymorphism of protein complexes such as the bacterial flagellum (also relevant in the setting of viral self-assembly).

As can be seen in an optical microscope, a motile bacterium finds its way across the field of view very quickly. The mechanistic underpinnings of this motion are tied to the rotation of the bacterial

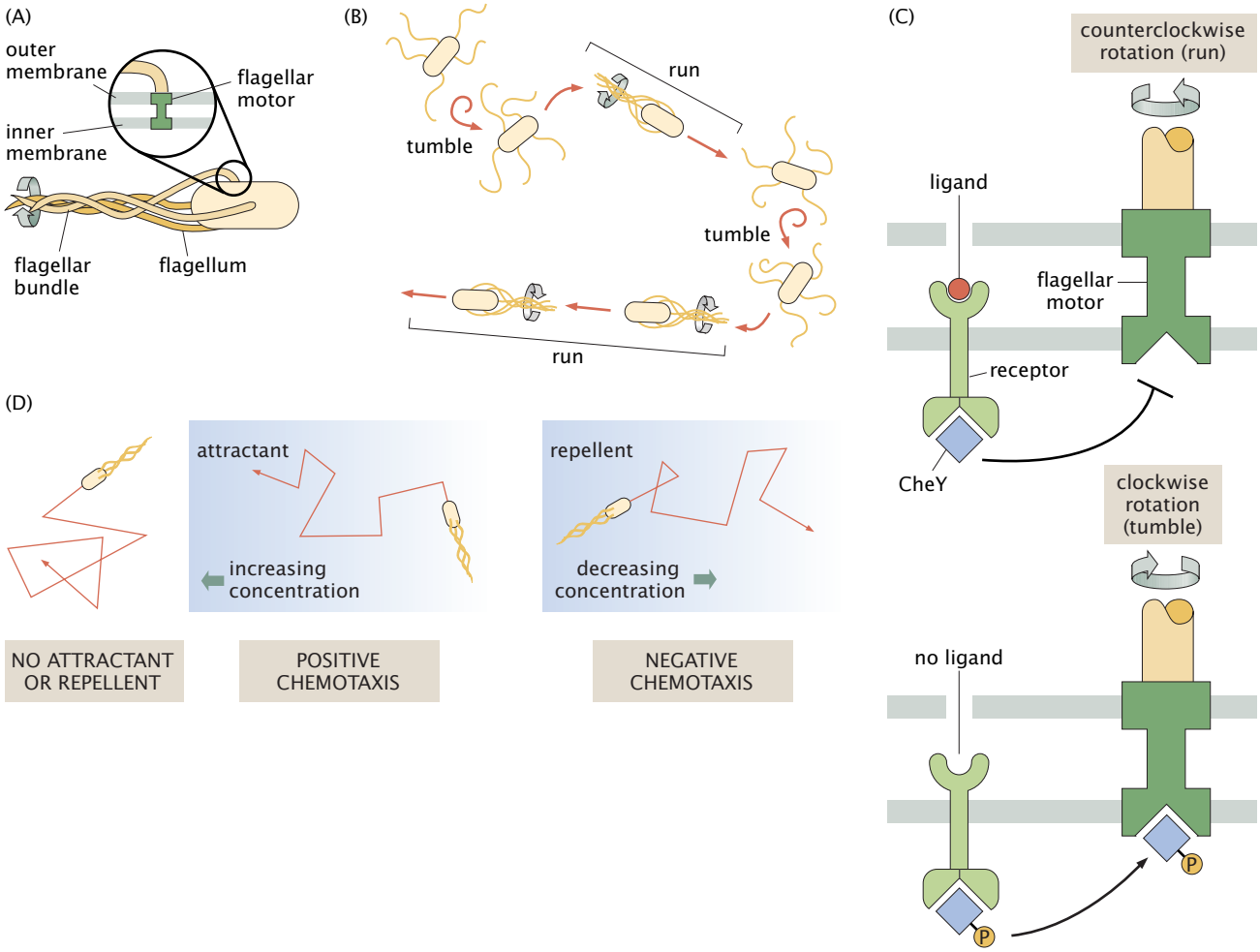


Figure 4.16: Mechanism of bacterial chemotaxis. (A) Many bacteria, including *E. coli*, are equipped with flagella. The helical flagellar filaments are turned by a motor embedded in the cell's wall and membranes. (B) When all of the flagellar motors on an individual cell are rotating counterclockwise, the filaments bundle together and the bacterium moves efficiently in a straight line called a run. When the motors reverse direction and spin clockwise, the flagellar bundle flies apart and the bacterium rotates in an apparently random manner called a tumble. Over long distances, bacterial trajectories appear as run segments arranged at random angles to one another at junctures where a tumble event occurred. (C) The switching of the motor from counterclockwise to clockwise rotation, and hence the switching of the behavior from running to tumbling, can be controlled by the presence of small molecules in the bacterium's environment. These molecules bind to receptors that then induce phosphorylation of a signaling protein, which in turn binds to and affects the mechanics of the motor. (D) Tuning of swimming behavior by small molecules can cause the bacterium to swim either toward desirable food sources, such as sugars or amino acids, or away from noxious toxins.

flagellum. A typical *E. coli* has on the order of 5 such flagella with a characteristic length of the order of 10 μm. One of the most remarkable classes of molecular motors is those that exercise rotary motion. The first such motor to be characterized in detail is that associated with the rotary motion of the *E. coli* flagellum and shown in Figure 3.24 (p. 120). This tiny rotary motor is capable of spinning the flagellum at roughly 6000 rpm using energy stored in an ion gradient across the bacterial cell membrane. The remarkable mechanical feature of this motor is that it can reverse its direction, rotating either clockwise or counterclockwise at comparable speeds. It is the likelihood of switching between counterclockwise and clockwise motion that regulates bacterial trajectories in response to chemoattractants. When all of the motors on a particular bacterium are rotating in a counterclockwise

direction, the five or so helical flagella line up into a well-ordered bundle and their spinning cooperates to propel the bacterium forward in a straight line at a rapid speed ($\approx$10–30 μm/s). When the motors reverse direction, the bundle is disrupted and the individual flagella fly apart. This disruption in the straight swimming of the bacterium is referred to as a tumble, because when the motor switches back to counterclockwise rotation and the bacterium begins swimming in a straight line again, the bacterium will be pointing in a new direction in space. During chemotaxis, bacteria swim towards attractive chemicals and away from repellant molecules simply by changing the frequency of tumbling. In Chapter 19, we will describe the molecular underpinnings of this behavior as well as physical models that have been put forth to explain the phenomenon.

All told, *E. coli* has been a remarkable model organism that has yielded a long list of biological discoveries.

4.5 Yeast: From Biochemistry to the Cell Cycle

Yeast Has Served as a Model System Leading to Insights in Contexts Ranging from Vitalism to the Functioning of Enzymes to Eukaryotic Gene Regulation

Several technical features of *E. coli* made it a nearly ideal organism for early geneticists and biochemists. First, it grows quickly. Second, it can grow on plates in colonial form such that all the descendants of a single individual cell can easily be gathered together. Third, it exhibits a variety of alternative metabolic lifestyles: for example, the ability to grow with and without oxygen, the ability to grow on different carbon sources, and the ability to either synthesize its own amino acids or take them up from the surrounding medium as environmental conditions permit. This flexibility facilitated isolation of bacterial mutants that were able to grow in one condition but not another, resulting in the identification of the enzymes involved in most major metabolic pathways. However, *E. coli* is a bacterium and is in many ways profoundly different from eukaryotic organisms such as ourselves.

To elucidate processes specific to eukaryotic organisms, another model system must be adopted that ideally would share many of the desirable features of *E. coli*. Such an organism is *Saccharomyces cerevisiae*, the brewer's or baker's yeast. Humans domesticated this wild fungus that normally grows on the skin of fruit some thousands of years ago. It has long played a critical role in the economy of human societies as it is necessary for making leavened bread and creating a variety of delicious intoxicating beverages, including beer and wine. Because of the economic importance of having reliable and reproducible brewing and baking processes, the inner secrets of *S. cerevisiae* were of intense interest to early microbiologists (see Barnett (2003) in Further Reading at the end of the chapter).

Although wild *S. cerevisiae* typically grows in unruly, stringy cellular groups, strains of domesticated *S. cerevisiae* were chosen by microbiologists to form neat, round colonies when grown on agar plates, looking much like colonies of *E. coli* but smelling substantially better. *S. cerevisiae* also grows remarkably fast for a eukaryotic cell, doubling in less than 90 minutes under favorable conditions. Also like *E. coli*, yeast can pursue a variety of metabolic lifestyle choices, and so it is similarly easy to isolate auxotrophic mutants (that is, mutants unable to synthesize an essential amino acid or other

metabolite). However, *S. cerevisiae* is a full-fledged card-carrying eukaryote. It has a membrane-enclosed nucleus and a full complement of eukaryotic membrane-bound organelles, including an endoplasmic reticulum, Golgi apparatus, and mitochondria, all of which are missing in *E. coli*. Even better, *S. cerevisiae* has sex the way that most other eukaryotes do, blending together the complete genomes of two parent cells that are then recombined in the offspring. Intriguingly, yeast can grow in either haploid or diploid forms, that is, either with one copy of every chromosome or with two copies of every chromosome where each of the two copies is derived from different parents. Thus, genetic and molecular dissection of eukaryotic-specific behaviors has been largely carried out using yeast. Furthermore, yeast (being a fungus) is much more closely related to animals than any bacterium, or indeed most other eukaryotic microorganisms. It therefore shares with us many specific pathways, including certain kinds of signaling pathways that are not found in *E. coli*. Below, we summarize some particular case studies where yeast has provided initial critical insights into processes that are important to many or all eukaryotic cells.

4.5.1 Yeast and the Rise of Biochemistry

Yeast served as a central player in the debate over vitalism and in launching the field of biochemistry. Many of the greatest chemists of the eighteenth century participated in the heated and vitriolic debate over the nature of fermentation. Fermentation was particularly important in France, at that time the premier wine-making country in the world. Prior to his decapitation by guillotine during the French Revolution, the great French chemist Antoine Lavoisier was the first to measure accurately the conversion of sugar into alcohol and carbon dioxide. Over the next several decades as the resolution of optical microscopes improved, it became possible actually to see the yeast cells responsible for fermentation and to observe their process of growth and division, establishing that fermentation was a function of a plant-like living organism and not a purely chemical process. Although it eventually did become accepted that living yeast cells were normally responsible for fermentation, many chemists held to the complementary view that life itself was not essential for fermentation to occur.

In an interesting irony of history, both sides in this heated debate were correct: as part of their normal life cycle, yeast cells convert sugar into alcohol. At the same time, the proteins (enzymes) that carry out this process within yeast can be extracted and purified to perform precisely the same biochemical conversion. Careful experiments by Eduard Buchner and others revealed that cell-free extracts from yeast could catalyze the fermentation of sugar, resulting in the realization that there are specific chemical substances that catalyze the key chemical reactions of the biological realm. Insights on the biochemistry of yeast were foundational in unearthing the features of the glycolysis pathway shown in Figure 5.2 (p. 191) and discussed in Chapter 5.

4.5.2 Dissecting the Cell Cycle

As yeast proceeds through its cell division cycle, it undergoes a series of morphological changes such that it is immediately apparent to the eye of a microscopist where a particular individual yeast cell is in its

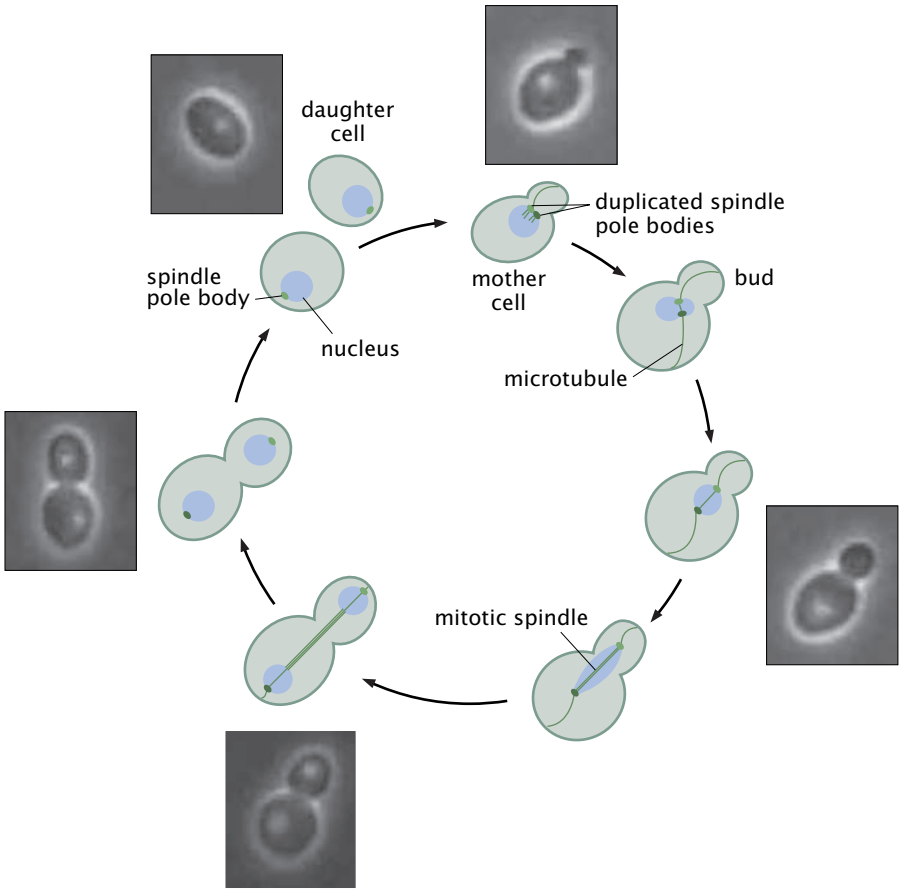


Figure 4.17: Cell cycle of yeast. Diagram of the process of yeast budding and division. As a single yeast cell progresses through the cell cycle, the future daughter cell, or bud, grows in size until it is large enough to receive one copy of the replicated genome. Then the bridge between mother and daughter cells is severed and both mother and daughter can go on to bud. The microscopy images show the yeast *S. cerevisiae* at different stages in the cell cycle. (Images from R. Menssen et al., *Curr. Biol.* 11:345, 2001.)

cell-cycle progression, as shown in Figure 4.17. It is easier to detect progression through the yeast cell cycle by visual means in *S. cerevisiae* than most other cell types because of its unusual propensity to grow by budding rather than by binary fission. The yeast cell cycle starts with an ovoid mother cell that then begins to grow a small bud from a single point on the surface. As the cell progresses through the cell cycle, the bud grows while the mother cell remains nearly the same size. After the chromosomes have been duplicated, the mitotic spindle delivers one set of chromosomes into the bud which eventually pinches off as a daughter cell. The mother cell can then go on to produce more daughters and the daughter can go on to have daughters of her own.

The correspondence between cell morphology and position in the cell cycle made it straightforward for investigators to isolate a large series of mutants involved in cell-cycle progression. Because cell division is a process essential to life, it was necessary to isolate conditional mutants (conditional mutants were defined in the Experiments Behind the Facts section on p. 140 at the beginning of the chapter). The initial strategy was to identify a large number of mutant yeast strains that were able to grow at a normal temperature (30 °C) but unable to grow at an elevated temperature (37 °C) where wild-type yeast can still flourish. Next, these temperature-sensitive growth mutants were visually evaluated to determine whether all the cells in a dying population at the high temperature appeared to be stuck at the same point in the cell cycle. Each of these mutant strains (called Cdc mutants for "cell division cycle") lacked a critical protein involved in cell-cycle progression such that a cell at any stage in the division cycle

when moved to an elevated temperature would continue to progress until it reached the point where the missing protein was essential.

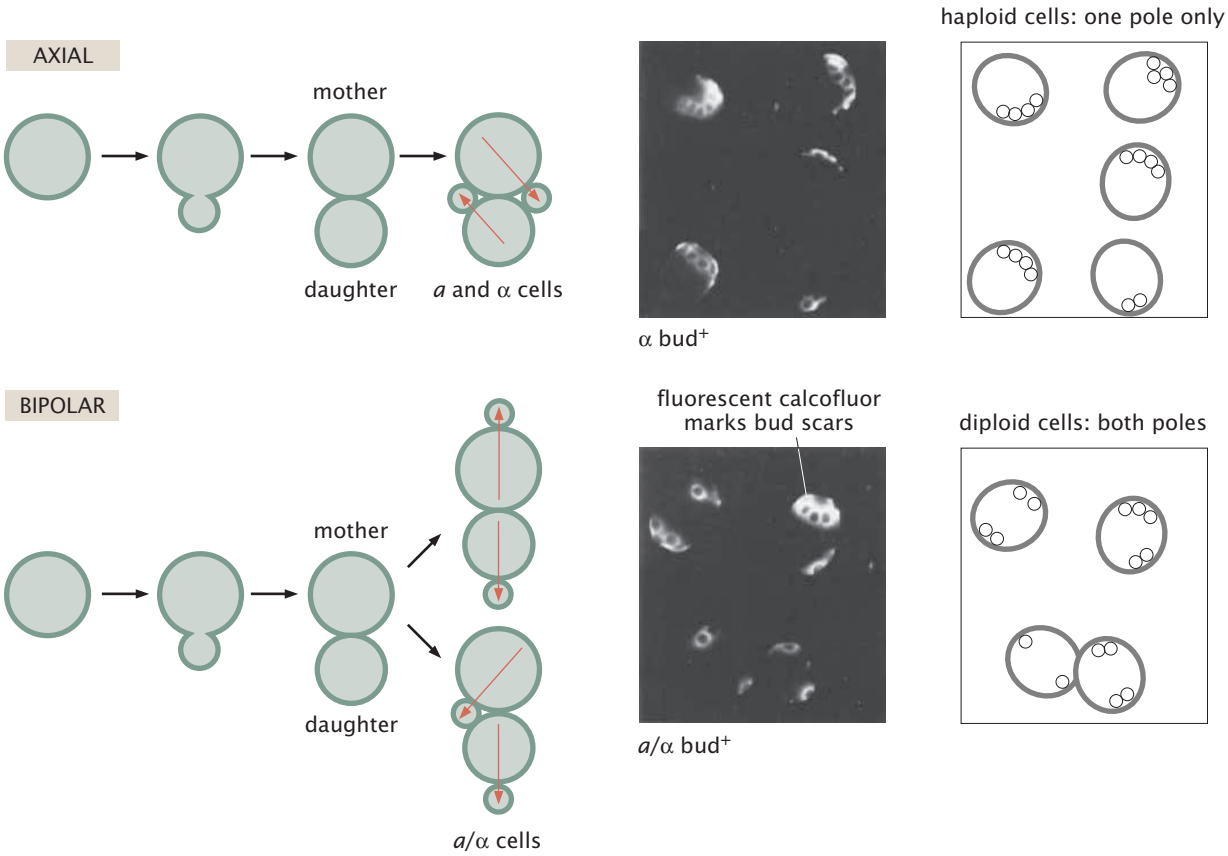
Many of the genes identified in this manner as critical to cell-cycle progression in *S. cerevisiae* have since proved to be critical for cell-cycle progression in most other eukaryotes as well, including in humans and frogs. As described in Chapter 3, biochemical experiments focused on understanding the cell-cycle oscillator in frog embryos were also used to identify proteins critical in cell-cycle progression. It is a satisfying demonstration of the complementarity of biochemistry and genetics as well as of the unity of life that both approaches pointed towards many of the same candidate molecules.

4.5.3 Deciding Which Way Is Up: Yeast and Polarity

The major benefits of yeast as an experimental organism, namely, its rapid and stereotypical growth pattern and the ease with which conditional mutants can be isolated, have been repeatedly exploited in many different kinds of studies to elucidate the basic processes of eukaryotic cell biology, analogous to those described above for the cell cycle. Another interesting and general biological process, critically informed by experiments on yeast, is the establishment of cell polarity. Polarity simply refers to the ability of cells to localize subcellular components in a nonuniform way in response to either external or internal spatial signals. Establishment of polarity is a feature of essentially all eukaryotic cells and many bacterial cells. It is particularly evident in cell types where spatial differentiation between one end of the cell and the other is critical for biological function. Some examples include motile cells that must distinguish front from back, epithelial cells forming barriers in animal tissues that must distinguish inside from outside, and neurons, which typically receive electrical signals from neighboring cells at one end and transmit them at the other end. As is clear from the images of budding yeast progressing through its cell cycle shown in Figure 4.17, the yeast cell targets growth in a polarized fashion such that the bud grows while the mother cell stays about the same size.

A series of clever genetic screens has enabled the identification of the protein machinery in yeast responsible for the establishment of large-scale cell polarity and for targeted growth. The master regulator determining cell polarity was actually first identified in the Cdc screen described above. Cells with a conditional mutation in the protein Cdc42 did not arrest at a particular point in the cell cycle, unlike the others described, but instead continued to grow uniformly and isotropically, becoming larger and larger, without ever choosing a bud site. Subsequent studies have shown that the Cdc42 protein is the primary signal that eventually directs construction of large-scale internal polarized structures, including actin cables, which run from one end of the cell to the other, that provide the underlying scaffolding for all kinds of directed transport into the growing bud. In many kinds of animal cells, the homologs of Cdc42 have also been found to play important instructional roles in establishing cytoskeletal polarity.

The same polarization machinery is used for a different purpose when yeast cells feel the urge to mate with one another. Haploid yeast cells of either of two mating types called *a* and α secrete chemical pheromones that can be detected by a nearby member of the opposite sex. Because yeast cells are unable to crawl towards one another for consummation, they instead simply grow in the direction of the



pheromone signal until they make contact. Under this circumstance, the localization of Cdc42 to a single patch on the surface is directed by signals from the pheromone receptor in the yeast membrane. This is an example of the coupling between external spatial information, in this case provided by the pheromone secreted by a nearby amorous mating partner, and the internal polarization machinery of the cell. Similar hierarchical signaling events using many of the same proteins have since been shown to be involved in the polarization of motile animal cells such as neutrophils during chemotaxis and in the polarized growth of nerve cell extensions, all in the presence of external cues.

In the absence of an external signal, how does the growing yeast cell choose where to put its Cdc42 protein and, therefore, where to put its single bud? It has long been observed that yeast cells bud in regular patterns that are determined by their haploid or diploid state as shown in Figure 4.18. Haploid cells typically start a new bud site immediately adjacent to the site from which they last budded, such that the cells in a microcolony tend to all pile up in a small clump. In contrast, diploid cells alternate their bud site selection between their two poles such that microcolonies grow in a more extended form. A visual screen was used to find yeast mutants that formed an apparently random pattern of buds on their surface. These mutants were fully viable and exhibited no significant growth defect in the laboratory. Their only deficiency was in their ability to tell which end is which.

In all the cases described above, the yeast cell grows in one and only one direction, although in the case of diploid cells growing in a bipolar fashion, this cell clearly has the capability of growing from either end. How does the cell know that it has formed a localized patch of

Figure 4.18: Bud site selection in growing yeast. Haploid cells (*a* cells or α cells) typically form a new bud immediately adjacent to their most recent site of cell division. This can be seen in two ways. First, looking at the pattern of small microcolonies (at the four-cell stage) shows every cell in contact with every other cell. Second, examining the position of bud scars on the cell surface (irreversible cell wall deformations that report on sites of previous cell division) shows that all bud scars are found in a single clump on one end of the cell. In contrast, diploid cells grow in a bipolar pattern where bud site choice alternates between the two ends of the cell. This results in microcolonies that are more extended with less cell–cell contact. Bud scars are found in two distinct clumps on opposite ends of the cell. (Adapted from J. Chant and I. Herskowitz, *Cell* 65:1203, 1991.)

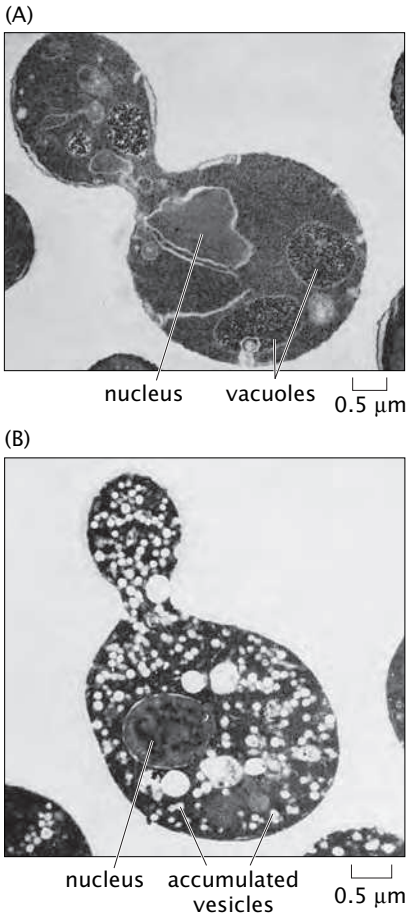


Figure 4.19: Accumulation of membrane transport vesicles in secretion-deficient yeast cells. (A) A conditional mutant in the secretory pathway gene grown at the permissive temperature showing normal internal membrane morphology. (B) Mutant grown at the restrictive temperature that has accumulated large numbers of internal transport vesicles. (Adapted from P. Novick and R. Schekman, *Proc. Natl Acad. Sci. USA* 76:1858, 1979.)

Cdc42 protein at exactly one site? Why do yeast almost never make the mistake of initiating two different buds simultaneously, even in bud site selection mutants that cannot tell one location from another? This is a specific instance of a more general problem that cells frequently have of needing to be able to count to 1. In the case of Cdc42 accumulation, the bulk of experimental evidence suggests that the unity of site selection depends on both positive feedback loops and negative feedback loops. Initial slight accumulation of Cdc42 at one location is positively reinforced as activated Cdc42 tends to self-associate and then growth of polarized actin structures directed by Cdc42 directs yet more Cdc42 to the same site. At the same time, negative feedback loops, which are not so well understood, seem to actively suppress Cdc42 accumulation everywhere else on the cell. Again, although molecular details vary from one species to another, it seems that the combination of local positive feedback and long-range negative feedback is used in many biological systems to enforce selection of one, and only one, location for subsequent polarization.

4.5.4 Dissecting Membrane Traffic

In contrast to bacteria and archaea, all eukaryotes have extensive and elaborate internal, membrane-bound organelles which perform distinct functions in cellular metabolism and behavior. These membrane-bound organelles, such as the endoplasmic reticulum and Golgi apparatus, are carefully positioned within the cell and specific subsets of their contents are trafficked from one organelle to another and also between organelles and the cell surface. In *S. cerevisiae*, directed membrane traffic is responsible for delivering newly synthesized cell wall material to sites of polarized growth as described above. Membrane traffic is also necessary for proper delivery of enzymes to their appropriate subcellular compartments, such as delivery of degradative proteases to the yeast vacuole and delivery of enzymes involved in nutrient digestion to the cell surface.

As the simplest and most genetically tractable eukaryote to exhibit all of these kinds of membrane trafficking, yeast was the target of mutational abuse designed to uncover critical genes involved in both the mechanics and signaling of membrane transport. Examples of such abused cells are shown in Figure 4.19. A critical observation was that yeast unable to deliver secretory organelles properly to the cell surface would become denser over time as protein synthesis continued without concomitant cell growth because the material responsible for cell-wall extension was no longer transported to the site of cell-wall synthesis. Furthermore, such mutants have defects in expressing secreted enzymes used for yeast metabolism, such as the enzyme invertase that clips sucrose into glucose and fructose during fermentation. To isolate mutants involved in this process, researchers first performed an enrichment by spinning yeast cells through a density gradient made of commercial floor polish. Extremely dense cells were isolated from the gradient, rinsed clean of floor polish, and then plated on a medium that would change color in the presence of invertase or other cell-surface enzymes. Several dozen different mutant strains were isolated that had various defects in processes associated with cell wall and enzyme secretion. These ranged from some cells unable to translocate proteins from the cytoplasm into the endoplasmic reticulum to others deficient in vesicle transport from the endoplasmic reticulum to the Golgi apparatus, to others deficient only

in the final fusion event delivering secretory vesicle contents to the cell surface. Members of this last class tended to accumulate very large numbers of frustrated vesicles immediately subjacent to what should have been the site of cell growth during bud formation. In parallel, biochemical experiments performed on animal cells specialized for secretion, including pancreatic cells and brain cells, identified proteins necessary for different mechanical steps of membrane fusion and vesicle budding. Nearly all of the proteins found by these biochemical methods in animal cells turned out to be homologs of genes identified in the yeast screen.

4.5.5 Genomics and Proteomics

All of the specific cell biological case studies described above have illustrated how the excellence of yeast as an organism for both biochemical and genetic experimentation has led to fundamental insights relevant to all eukaryotic cells. After the invention of recombinant DNA technology in the 1970s, classical genetics was complemented by reverse genetics. Instead of isolating mutants based on phenotype and then trying to identify the molecular lesions involved, researchers were now able to start with a particular molecule of interest and ask what phenotype would result when that molecule was disrupted. For example, for a particular protein, isolated through some biochemical means, site-specific disruption of the gene encoding that protein would generate an engineered mutant with that specific deficiency and no other, which could then be examined for phenotypic alteration. This process is called reverse genetics because it starts with known molecules and asks for phenotypes, in contrast to classical genetics, which starts with phenotypes and looks for molecules. It will come as no surprise that yeast was also one of the first eukaryotic organisms in which reverse genetics was fruitfully applied. Replication of engineered plasmids in yeast was first reported in 1978 and gene replacement by homologous recombination just a few years later in 1983. Thus, the full toolkit for site-specific DNA manipulation, originally developed in *E. coli*, was also realized in yeast. Along the same lines, yeast was the first eukaryotic cell whose genome was fully sequenced and assembled in 1996.

The sequencing of the yeast genome opened the door for a new class of genome-wide investigations that celebrated the now known and finite boundaries of yeast genetic material. With this knowledge in hand, it was possible to ask truly comprehensive questions. Rather than examining changes in transcription levels for a single gene at a time, it became conceivable to measure transcription levels for every gene in the genome. Similarly, rather than performing stepwise classical or reverse genetic analysis, one mutant at a time, it became conceivable to ask how a disruption in any of the yeast genes might affect a particular process. These large-scale, grandiose imaginings were made realizable by the relatively modest size of the yeast genome (only about 6000 genes) and its fairly simple genetic structure (very few introns interrupting gene sequences).

Over the years since the completion of the yeast genome sequence, a breathtaking set of experimental resources have been developed and made available by the community of yeast researchers. One early example was the production of complementary DNA (cDNA) microarrays that enabled the measurement of the level of every mRNA in the cell simultaneously under a variety of experimental conditions.

(cDNA is the DNA synthesized from an RNA template using the enzyme reverse transcriptase.) Massive databases collate and correlate experimental information about mRNA levels measured in this way for thousands of different individual experiments. An even more ambitious resource has been the construction of a comprehensive library of targeted yeast gene disruption mutants where, one by one, each gene in the genome has been deleted and replaced with a molecular bar code, a specific DNA sequence that identifies that mutant strain from amongst all of the others. These tagged deletion strains either can be ordered and used one by one for standard reverse genetic analysis or else can be pooled and assayed simultaneously by following their barcodes to find all of the targeted disruptions that affect a particular process of interest. This family of techniques collectively called *genomics* represents the evolution of Mendel's field of genetics into the modern era where our access to complete genome sequence information enables the possibility of performing studies that are comprehensive and complete in a way not previously imaginable. These large-scale projects in yeast paved the way for similarly ambitious genome-scale projects in many other model organisms, including humans. In many of these projects involving technique development, research in the yeast community led the way.

As we stated at the beginning of this chapter, molecular understanding of biological processes requires complementary information from experiments based in genetics and from experiments based in biochemistry. The biochemical equivalent of genomic-scale experimentation is a growing field called *proteomics*. Again, because of the relatively small number and limited complexity of protein products encoded in the yeast genome, it has been possible to perform several different kinds of comprehensive analyses of protein behavior, localization, and functional interaction. For example, in a project of similar technical scope to the generation of the deletion collection described above, every identified yeast gene has been replaced one at a time with a modified gene fused to the gene encoding GFP as indicated in Figure 2.20 (p. 58). Of the 6000 or so strains constructed in this manner, about 4000 actually expressed green fluorescence under standard laboratory growth conditions. This made it possible to catalog in a systematic way the subcellular localization of every single one of these thousands of proteins in living cells. Besides the subcellular location of a protein, much important information about its function and biological role can often be gleaned by examining its binding partners.

In a similar proteomic-scale approach, every protein-encoding gene in the yeast genome was replaced with an engineered version fused to an affinity purification tag, a specific protein sequence that is easily fished out of a complex mixture using biochemical techniques. Each tagged protein isolated in this manner could then be examined for binding partners that tended to purify along with it. The study of binding in biology will be one of the threads that runs through the entire book, with our first serious encounter with this subject to begin in Chapter 6. This is one of several kinds of large data sets that have been used to build comprehensive maps of the network of protein interactions in yeast. A small portion of this network is illustrated in Figure 4.20. The affinity purification approach will find only a subset of relevant interactions, specifically, those that are stable enough to survive disruption of the yeast cell and passage over a purification column.

Other kinds of experiments, which may now also be implemented on a genome-wide or proteome-wide scale, can reveal transient

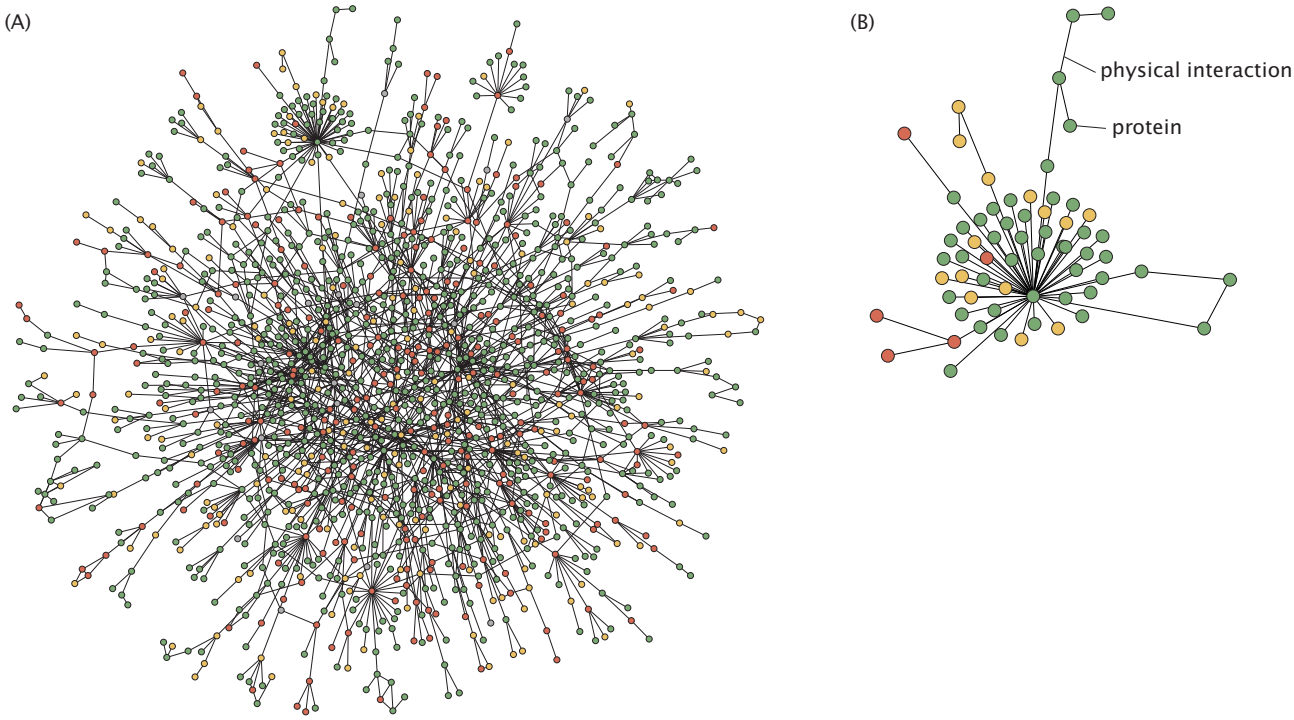


Figure 4.20: The yeast protein–protein interaction network. (A) In this diagram, each dot (or node) represents a single protein. Lines connecting the nodes indicate that those two proteins have been shown to interact with one another, biochemically or genetically. This kind of analysis reveals that all protein products in yeast participate in a web of interactions such that no protein is an island unto itself. (B) A magnified view of part of the network from (A). (Adapted from H. Jeong et al., *Nature* 411:41, 2001)

interactions or functional interactions that may be missed when searching simply for stable complexes. One commonly used approach to find interacting proteins is referred to as a two-hybrid screen. In the original implementation, a protein of choice called "bait" was fused to a partially functional protein fragment that would be able to regulate the expression of a reporter gene if and only if the bait protein was bound to a target protein carrying the other half of the functional tag. Taking a single bait protein and challenging it with a library of tagged potential partners can rapidly lead to identification of proteins that bind to the bait in living cells without the need for any biochemical purification. The rapidly growing databases that connect this kind of biochemical information with information from other large-scale screens showing genetic or functional interactions can instantly provide a researcher with a molecular-level context to understand the likely consequences of an alteration in any gene of interest.

Why has the rapid development of all these new genomic and proteomic techniques happened so quickly in yeast as compared with other model organisms? As we have described, part of the answer has to do with the properties of yeast itself. However, a perhaps equally important part of the answer comes from the nature of the yeast research community. Thousands of different individual laboratory groups use yeast as a primary or secondary model organism and this community of researchers has recognized the extraordinary power of creating and sharing large-scale resources. At present, both the concrete resources such as yeast strains and plasmids and the intangible resources such as the large-scale quantitative data sets generated by these techniques are freely and efficiently shared among all

the members of the research community. This amplifies the power of any individual group, rendering the yeast research community a large-scale analog of the genetic and proteomic networks that we have just described.

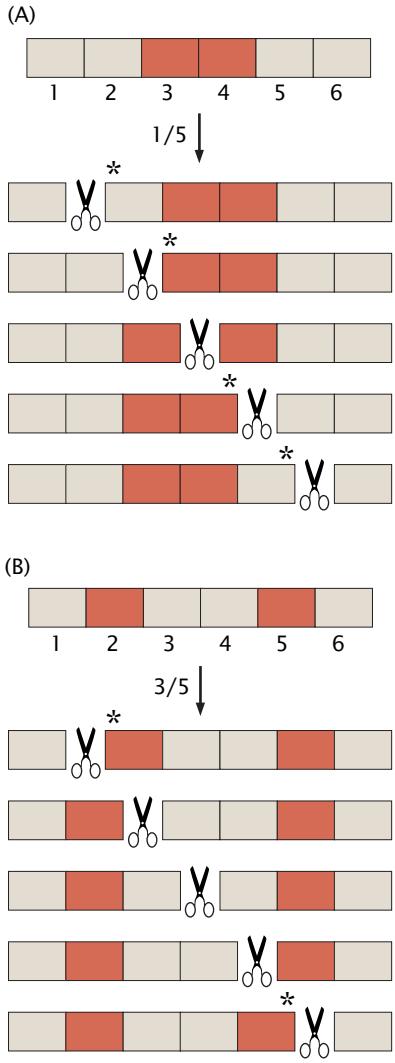


Figure 4.21: Concept of mutation correlation and physical proximity on the gene illustrated by labeling a string in two distinct points and then making random cuts of the string. The probability that the two labels will remain on the same part of the string *after* the cut depends upon their physical proximity *before* the cut. (A) Two mutations (red boxes) that are close to one another are likely to remain on the same part of the string (four times out of five). (B) Two mutations that are further apart are more likely to be separated (remaining together only two times out of five).

4.6 Flies and Modern Biology

4.6.1 Flies and the Rise of Modern Genetics

Drosophila melanogaster Has Served as a Model System for Studies Ranging from Genetics to Development to the Functioning of the Brain and Even Behavior

One of the recurring themes in the history of molecular biology is the effort that surrounds the discovery of a concrete realization of a concept that is already known abstractly. One such example is the realization that the abstract concept of a gene (a unit of inheritance) is physically manifested in chromosomes, a discovery resulting from the work of Thomas Hunt Morgan and his group on the fruit fly *D. melanogaster*. The experiments of Morgan and his students concerned the study of correlations in mutations. Morgan and his gene hunter students were, like Mendel, concerned with the nature of inheritance and mutation. The significant point of departure for their experiments was a single fruit fly with white eyes rather than the red eyes of the wild type.

The beautiful and powerful idea that ties together the various experiments from Morgan, Sturtevant, and co-workers is that correlations in inheritance give a measure of physical proximity of genes on the chromosome, as illustrated schematically in Figure 4.21. The simple idea that permitted them to map the location of particular genes on a chromosome of *Drosophila* was that during the cross-over process those genes that are physically closer on the chromosome have a greater chance of being inherited together in the offspring. Crossover is the process whereby genes are scrambled as the fruit fly is making eggs, through cutting and ligation of genomic DNA. For example, starting from a mother who carries mutations for both white eyes and miniature wings (a double mutation), the offspring can carry both mutations, one or the other, or neither. Morgan and co-workers observed that offspring of mothers carrying both white eyes and miniature wings were less likely to coinherit both mutations than offspring of mothers carrying mutations for white eyes and yellow body color. They interpreted this to mean that the gene encoding yellow bodies is physically closer on the chromosome to the gene encoding white eyes than is the gene encoding for miniature wings and therefore less likely to be separated during the process of recombination.

Another key observation enabling the evolution of the idea of a gene from an abstract unit of inheritance to a physical segment of DNA on a chromosome was facilitated by a peculiar habit of cells in the *Drosophila* salivary gland. These giant cells replicate multiple copies of their chromosomes without separating them. The multiple copies of replicated DNA remain aligned next to one another, forming a giant polytene chromosome that can be easily seen with a light microscope as shown in Figure 4.22. Characteristic dark and light banding patterns, visible on the polytene chromosomes, represent a barcode identifying certain stretches of the chromosome. The

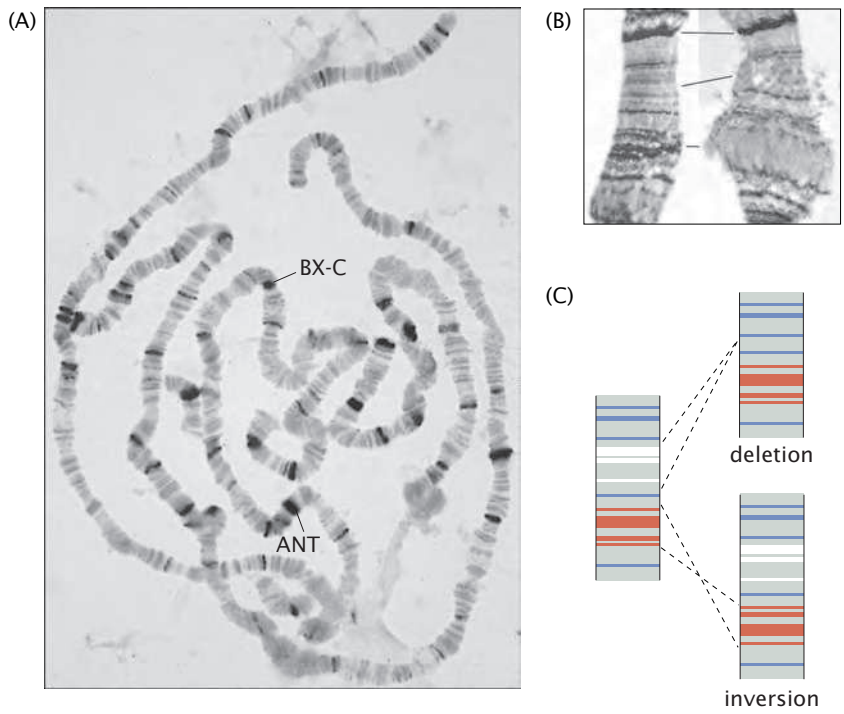


Figure 4.22: Polytene chromosomes and genetic cartography. (A) A light microscope image of a giant polytene chromosome from a *Drosophila* salivary gland cell. The positions of two genetic loci, *Antennapedia* (ANT) and *bithorax* (BX-C), are indicated. (B) A magnified view of a small portion of one of the chromosomes shows the effects of transcription on polytene chromosome structure. The images on the left and right show the same chromosomal region, before and after a specific phase of developmental transcription is initiated by the hormone ecdysone. The portion of the polytene chromosome bundle encoding a gene that is being actively transcribed decondenses to form a characteristic "puff." The identities of these developmentally regulated genes can be correlated both with their positions in the genetic map and with their physical positions on the polytene chromosome. (C) Schematic showing how a physical deletion or inversion of a portion of the chromosome is reflected in both the appearance of the polytene chromosome and the apparent order of genes on the chromosome as determined by genetic mapping. (A, adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed., Garland Science, 2008; B, adapted from I. F. Zhimulev et al., *Int. Rev. Cytol.* 241:203, 2004.)

number of distinct kinds of chromosomes visible in the microscope was intriguingly the same as the number of linkage groups (sets of genes that are generally transmitted together) identified by genetic analysis, suggesting a direct correspondence between those visible chromosomes and the linked units of inheritance. It was found that large physical rearrangements such as inversions or deletions that visibly change the banding pattern on the polytene chromosomes were correlated with genetic inversions or deletions describing changes in the linkage of particular groups of genes. This firmly established the concept that genes were physical entities aligned in order on chromosomal DNA. It is not an exaggeration to say that Morgan and his gene hunters, by studying the geography of genes on chromosomes, launched the scientific program that culminated in the sequencing of entire genomes and the biological mapmaking that is the hallmark of modern genetics.

4.6.2 How the Fly Got His Stripes

A second key role for the fruit fly has been in its role as a model system for the study of development. Unlike bacteriophages, *E. coli*, or yeast, *Drosophila* are multicellular animals, much like ourselves, with heads, eyes, and legs. Several famous *Drosophila* mutations illustrated the idea that single genetic changes could result in complete respecification of complex body parts made up of thousands of cells. For example, the mutations *eyeless* and *wingless*, as their names imply, result in the development of flies that completely fail to form a critical body part. Even more amazing are mutations such as *Antennapedia* and *bithorax* that result in transformation of one body part into another as shown in Figure 4.23. The heads of flies carrying *Antennapedia* mutations grow legs where their antennae ought to be. Flies with the *bithorax* mutation develop a second set of full-size wings directly behind their normal single pair. These kinds of observations

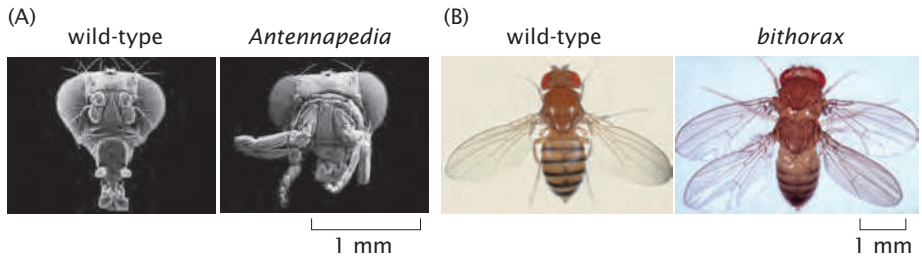


Figure 4.23: *Drosophila* mutations that transform one body part into another. (A) Scanning electron microscopy of fly heads showing the normal head with normal antennae and an *Antennapedia* mutant. (B) Fly carrying *bithorax* alongside a normal fly. (A, Courtesy of F. R. Turner.; B, courtesy of E. B. Lewis.)

established the idea that large-scale developmental patterns can be controlled by master regulatory genes. These are single genes that initiate a cascade of events leading to specification of many different cell types in carefully regulated spatial order to create body parts such as eyes, legs, and wings. We will discuss these genes further in Chapter 21.

More generally, *Drosophila* genetics and analysis of physically deficient mutant flies have provided a great deal of insight into our understanding of how cells communicate with one another during development to form large-scale complex patterns. One of the most visually striking examples is the specification of a series of segments in the fly embryo along the head-to-tail axis that will eventually become distinct segments of the adult fly. These patterns were shown in Figure 2.45 (p. 80) and discussed in that section. Specification of the segments requires several different kinds of cell–cell communication that take advantage of physical processes such as diffusion to set up steady-state distributions of signaling molecules and also uses interesting kinds of positive and negative feedback to refine positional information. In Chapter 19, we will return to this example and explore the elegant molecular and cellular interactions involved in this kind of pattern formation.

Another kind of developmental event that has been well studied in *Drosophila* and is also relevant to other kinds of animal development involves the situation where a field of identical cells becomes specified to establish distinct cellular identities. A dramatic and well-studied example is the specification of photoreceptor cells in the developing *Drosophila* eye (see Figure 2.35(D) on p. 73 for eyes in another species of fly). The *Drosophila* eye, like our eye, contains a series of light-sensitive cells that respond to different segments of the visible light spectrum enabling the animal to see in color. At early stages of development, each of the *Drosophila* photoreceptor cell precursors has the potential to become any photoreceptor cell type. As development proceeds, the cells communicate with their neighbors in such a way as to enforce the final outcome that each cluster contains one and only one copy of each cell type. This local cell–cell communication is mediated by proteins expressed on the surface of the cells that allow each cell to keep track of what its neighbor is doing and to direct the neighbors not to follow the same developmental program as the cell is undergoing itself.

Drosophila has been a model system that, because of its facility as a subject for genetic analysis, has provided some of the first molecular clues for understanding each of these kinds of developmental processes. Information gleaned from experiments on *Drosophila*, both the identities of specific key molecules and the paradigms for pattern formation events, has been key to understanding developmental processes in animals such as mice and humans which are not so genetically tractable.

4.7 Of Mice and Men

Just as yeast proved more useful for studies of eukaryotic biology than *E. coli* and *Drosophila* more useful than yeast for issues having to do with multicellularity and animal development, likewise, some biological issues of paramount interest to humans, such as development of cancer and mammalian-specific elaborations of the central nervous system (such as a highly specific sense of smell), can best be studied only in a mammalian model system. Many different kinds of mammals have been used as laboratory research subjects, ranging from horses to dogs to rodents. In general, rodents have been most widely used because of their small size and fairly rapid generation time. Rats have been particularly widely used for psychological and physiological experimentation, giving rise to the vernacular term "lab rat." Guinea pigs have also put in heroic work as laboratory subjects, giving rise to another vernacular term, being treated like a "guinea pig." But of all the rodents, and indeed, all the mammals, that have been used in biomedical research, mice (*Mus musculus*) are by far the most important.

Mice have probably been kept as pets by humans since we began storing grain and it has long been noted that occasional mutants would arise with coloring or behavior different from normal. By the eighteenth century, mouse fanciers in Japan were deliberately breeding strangely colored mice and in the nineteenth century, mouse breeding as a hobby began to catch on in Europe. The life span of mice is short (typically about 2 years) and they become sexually mature at about 6–8 weeks, facilitating genetic breeding experiments. Although it is relatively easy to isolate mutant strains of mice, for a long time it was extremely difficult to identify the molecular lesions responsible for the change in color or behavior. Advances in cloning and genotyping techniques over the past 50 years have now made it routine to identify the genetic location of a particular mutation. More importantly, it is now actually possible to deliberately genetically engineer mice in much the same way we can genetically engineer *E. coli*, yeast, or *Drosophila*. Specific genes can be deleted by site-specific recombination and altered genes can be introduced and expressed. For example, GFP can be expressed in mice either specifically in a tissue where gene expression is controlled by a known promoter or more generally, in all cells. Three such mice are shown in Figure 4.24.

One extremely interesting property of mammalian development that is not shared by flies is that every cell at a particular early stage is totipotent, that is, it can turn into any type of cell present in the adult

Figure 4.24: Genetic manipulation of the mouse. Mouse embryonic stem cells were engineered in culture so that they could express the green fluorescent protein at all stages of development. These engineered cells were then introduced into a normal mouse embryo and the resulting chimeric mouse bred to generate animals who expressed GFP in every cell in the body. Here, three GFP-expressing mouse pups are shown under fluorescent illumination along with three of their normal littermates. (Courtesy of Ralph Brinster, University of Pennsylvania.)

organism. A common manifestation of this totipotency is seen in identical twins; when the cells of the early embryo accidentally split into two groups, each group is fully able to form a complete human being. Embryonic cells with this property can be isolated and genetically engineered. Manipulation of these so-called embryonic stem cells is a critical tool in the study of mammalian developmental biology. Embryonic stem cells were first isolated from human embryos in 1998 and research on their properties and potential medical use is proceeding rapidly in many parts of the world.

4.8 The Case for Exotica

Most of the model systems discussed in this chapter are laboratory workhorses that have been used by very large numbers of laboratories to make a wide range of distinct discoveries. However, there are many cases where a single specialization in an organism has proved to be of extraordinary utility in illuminating general principles in biological function. Many of these organisms cannot be cultured in the laboratory and their scientific claim to fame may be based on a single finding. However, biology would be profoundly different if any of these beasties had not been discovered.

4.8.1 Specialists and Experts

Already we have frequently fallen into the trap of describing conditions within "the cell" as if all cells are identical or similar. Obviously, this is a drastic oversimplification. Even within a single organism such as the human body, different cells that contain the same genome are highly specialized for different purposes. Experimental biologists have long taken advantage of the diversity and specialization of different cells within organisms and different organisms within ecosystems to examine the mechanism of a particular process of interest that may be employed by very many cells but is brought to a fine art in only a few.

For example, the cytoskeletal proteins actin and myosin are involved in the production of contractile force in all eukaryotic cells. However, the skeletal muscle cells of vertebrates contain extraordinarily high concentrations of both of these proteins arranged into regular arrays that are able to produce a large amount of force and contract over long distances to move animal limbs. Thus, biochemists interested in the mechanism of force generation by actin and myosin have long focused on skeletal muscle cells as cellular specialists where these proteins are easily purified and also easily studied in their native habitats. Some animals go further and modify their already specialized cells in particularly interesting ways that are also useful for identification of important proteins involved in general processes. The electric fish *Torpedo* uses a highly specialized organ derived from muscle to generate an electric charge that can shock and stun its prey. In normal muscle cells, the electrical signaling directing their contraction occurs over a very small part of their surface area where a motor neuron makes contact with a small field of receptors called acetylcholine receptors (we will describe the mechanisms of this electrical signaling in Chapter 17). In a typical skeletal muscle cell, this field of receptors covers less than 0.1% of the cell surface, so purification of the receptors and study of their function is extremely difficult. However, in the *Torpedo* electric organ, the electric receptor domains have been

greatly expanded, covering more than 10% of the cell surface. Therefore, researchers wishing to purify and examine the acetylcholine receptors generally do so from wild-caught *Torpedo* fish.

Another example where the very large size of individual cells facilitated investigation of general cell processes is in the salivary glands of *Drosophila*. As discussed in the previous section, in these large cells, the chromosomes are duplicated many times without cell division and the four chromosomes form aligned bundles. Striped patterns on these chromosome bundles reflecting underlying differences in chromatin organization are readily visible at the level of light microscopy. This special property of these cells was of great use to early *Drosophila* geneticists in establishing the physical realization of the abstract idea of the gene.

Important specialists have also been identified outside of the animal kingdom. One that has proved particularly fruitful in the study of molecular biology has been the small teardrop-shaped ciliate *Tetrahymena*. This organism first gained celebrity status with the identification of the biosynthetic machinery that makes telomeres, specialized segments at the end of linear chromosomes that prevent information being lost due to structural difficulties of DNA replication at the ends. In human cells, telomeres represent a very small fraction of the total DNA, with only one telomere at each of the ends of our 46 chromosomes. *Tetrahymena*, however, undergoes an unusual developmental process where important segments of its chromosomes are copied many times and stored in a separate specialized nucleus, resulting in many thousands of telomeres per cell. As with the examples discussed above, this specialization facilitated biochemical identification of the enzyme complex responsible for telomere formation, called telomerase, subsequently found to be well conserved among all eukaryotes. *Tetrahymena* is also famous for being the first cell type where an RNA molecule was proved to have catalytic function. Catalytic RNAs, called ribozymes, have also since been found throughout all domains of life. This ribozyme is able to catalyze a biochemical cleavage reaction in a manner that would normally be part of the bailiwick of proteins. Structural and chemical studies of the nature of this catalysis have helped reveal principles of how biochemical catalysis works in a more general sense.

4.8.2 The Squid Giant Axon and Biological Electricity

Repeatedly in the history of biology, we will find other examples where an important molecule or biological process was easy to identify or study from an exotic cell from an exotic creature. A beautiful example that we will revisit in Chapter 17 is the giant axon from the squid *Loligo*, which was used to great effect to understand the mechanisms of electrical signaling in neurons and also the mechanisms of long-distance intracellular transport. The squid giant axon operates under the same principle as all other animal nerve cells, but, as its name suggests, it is enormous, with a size of up to 1 mm in diameter. One key challenge in studying the electrical properties of cells is how to interface our macroscopic measurement apparatus and the cell membrane. This question translates into the technical challenge of connecting electrodes to nerve cells. Because of its huge size, the squid giant axon is ideal to manipulate in the laboratory, to impale with electrodes and to observe under the microscope. As a result, the squid giant axon allowed scientists to meet this challenge such that the nerve cells could be subjected to direct electrical query through

the use of electrodes. A schematic of the squid and its axon is shown in Figure 4.25. Subsequent development of much smaller and finer electrodes enabled the study of electrical signaling in much smaller neurons and revealed that the principles elucidated in the squid are general.

Much of the business of cells is mediated by the presence of charged species both within cells and in the extracellular medium. The study of the control of spatial position and temporal evolution of charge has relevance to a huge variety of processes and most notably to the functioning of nerve cells. The present discussion will show the way in which studies on the giant axons of squids served as the basis of our current understanding of action potentials.

There Is a Steady-State Potential Difference Across the Membrane of Nerve Cells

Cell membranes mediate the compartmentalization of different chemical species (ranging from simple ions to proteins) within cells. One consequence of such compartmentalization is the possibility of a mismatch in the charge state from one side of the membrane to the other, with the consequence that an electrical potential is set up across that membrane. Such potentials are known as membrane potentials. To get a sense for the types of ionic mismatches encountered in real cells, we note that the concentration of K^+ ions within the cell is roughly 20 times greater than outside the cell. A crude estimate of the corresponding membrane potential for such concentrations can be obtained by using the Nernst equation (to be discussed in detail in Chapter 17), which tells us that

$$V = \frac{k_B T}{q} \ln \frac{C_{out}}{C_{in}}, \tag{4.1}$$

where q is the charge on the species of interest, C_{out} is the extracellular concentration of these charges, and C_{in} is their concentration within the cell. Using the concentration ratio suggested above, the Nernst equation suggests a membrane potential of the order of -75 mV, indicating that the potential inside the cell is 75 mV lower than the potential outside.

Nerve Cells Propagate Electrical Signals and Use Them to Communicate with Each Other

There are a number of different mechanisms cells use to convey information. One key mechanism involves the presence of chemical messengers that are released as the basis of cell signaling. A second key mechanism of information transmission is the use of electrical signals and, in particular, the so-called action potential. The physical picture associated with the action potential is that of a traveling pulse of membrane potential difference that propagates along nerve cells at speeds of up to 100 m/s. The successive arrival of an electrical pulse along the axon of a nerve cell is shown schematically in Figure 4.25. The idea is that the local value of the membrane potential at some point along the axon is disturbed from its equilibrium value as a result of a transient change in the permeability of the membrane due to the opening of ion channels and a concomitant flow of ions across that membrane. This spatially localized change in potential has the effect of disturbing the ion concentration and potential

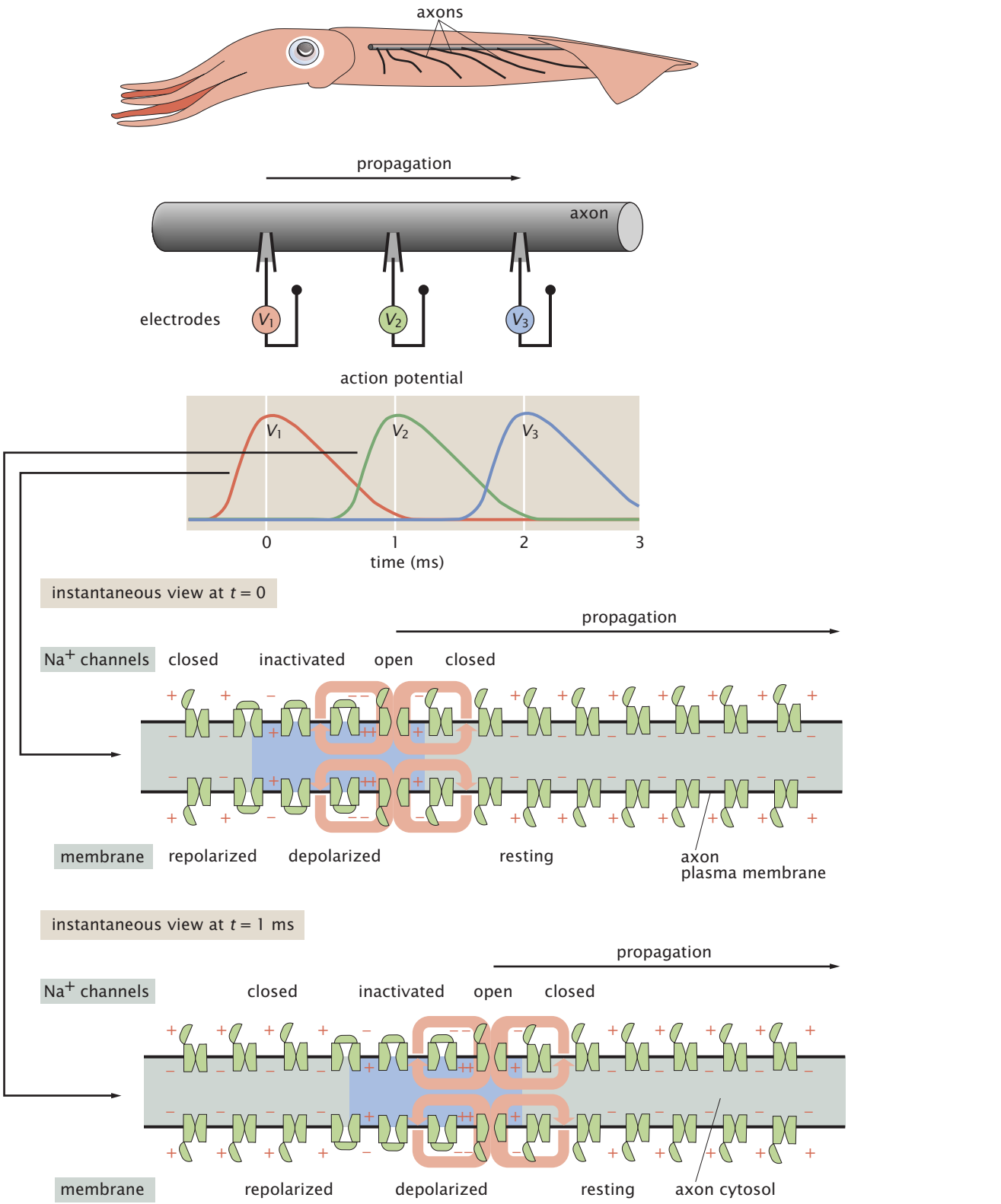


Figure 4.25: The squid giant axon. Schematic of the squid axon which shows the axon as part of the squid's anatomy, illustrates the propagation of an action potential in abstract electrical terms, and shows how that action potential is mediated by the presence of ion channels. (Adapted from B. Alberts et al., Molecular Biology of the Cell, 5th ed. Garland Science, 2008.)

in its vicinity, which unleashes transient changes in permeability in those regions with a corresponding change in the membrane potential. This cascade moves sequentially down the axon and, as noted above, can serve as the basis of signaling such as is evidenced by the contraction of muscles in response to external stimuli.

4.8.3 Exotica Toolkit

In the preceding section, we focused on organisms that specialize in producing a molecule or performing a function because the molecule or function is widespread. However, in many cases, the real strength of exotica lies in their ability to do something unique. Clever scientists have repeatedly exploited the very unusual properties of some organisms to generate reagents that can serve as a critical piece of the biologist's toolkit. The most familiar example is the green fluorescent protein (GFP). This protein is produced by the jellyfish *Aequoria victoria* as part of its system for generating light signals deep in the ocean. The ability to produce light or to fluoresce under illumination is shared by many organisms, including fireflies, many bacteria, fungi, and corals, as well as jellyfish. However, most of these light production systems rely on highly complex chemical groups that are created through elaborate biosynthetic pathways involving many enzymes. The very special feature of GFP that has made it so useful is that this protein is able to fluoresce on its own without any additional chemical groups. Thus, when a gene encoding GFP is expressed in a foreign organism from a bacterium to a tobacco plant to a rabbit, the protein is still able to fluoresce. Repeatedly throughout this book, we will see examples of the uses of GFP, sometimes as a reporter for transcription, sometimes as a probe for measuring physical properties of cytoplasm, and sometimes as a tag for showing the distribution of a particular protein of interest in a living cell.

Arguably the most important reagent ever to be identified from an exotic organism is the heat-stable DNA polymerase from *Thermus aquaticus* that enabled the development of the polymerase chain reaction (PCR). This revolutionary technique is used in all fields of biology ranging from field identification of organisms to cloning of useful products for biotechnology to identification of forensic evidence at crime scenes. PCR amplification of rare DNA sequences relies on thermal cycling where the temperature of the sample is repeatedly raised and lowered. DNA polymerase from any normal organism would be denatured by the rise in temperature. However, *T. aquaticus* prefers to live in geothermal hot springs and would die of cold at normal mammalian body temperatures. Its DNA polymerase can therefore survive many rounds of thermal cycling. Other organisms that live at extremely high temperatures have also provided heat-stable proteins useful in biotechnology and protein structural studies.

One of the major features of organisms on Earth is the large amount of energy that they devote to interacting with other organisms, largely to kill and eat them. Some predatory organisms have developed great speed or strength to catch their prey. But consider the plight of the cone snail. This family of slow-moving mollusks aspires to a predatory lifestyle, but speed and strength are not part of their repertoire. Instead, they have developed a remarkable witch's brew of potent neurotoxins that can paralyze a wide range of prey organisms. A similar strategy of hunting by chemical warfare has been developed by many snakes, spiders, and scorpions. Still other organisms create neurotoxins for defensive reasons, including sponges, frogs, and

pufferfish. Why are all these toxins useful rather than just exotic? Neurotoxins work by binding with extraordinarily high affinity to receptors in their target animals. Inhibition of these receptors at low concentrations of neurotoxin must be sufficient to paralyze or kill the target animal. Therefore, their receptors are almost always proteins critically important in neuronal signaling or muscle contraction. Because of the strong binding affinity of neurotoxins for their receptors, they can be used biochemically to purify and identify their receptors. Specific toxins are also extremely useful for studying the sequence of events involved in a particular form of signaling, since they will typically permit upstream events to occur but will block those that are downstream.

This is by no means an exhaustive list of the important contributions that have been made to biology by nonstandard organisms. The point we wish to emphasize is that the study of life processes is greatly enriched by examining the ways in which organisms are different from each other as well as by examining the ways in which they are similar.

4.9 Summary and Conclusions

This chapter has introduced many of the key model molecules and organisms that will be returned to as the basis for the development of physical biology in the remainder of the book. Our argument is that once the biology of a particular organism is well in hand, it becomes possible to shift the emphasis to the kind of rich interplay between quantitative theory and experiment that is the hallmark of physical biology to be explored throughout the book. Some of the key model systems that we will return to are hemoglobin, bacteriophages, *E. coli*, the yeast *S. cerevisiae*, and the fruit fly *D. melanogaster*. In addition to the celebrity status of these key organisms, we have also argued that there is an important role for the study of exotica. In many ways, the pleasure of biology is to explore the differences among organisms and to understand the ways in which particular specializations have arisen from the evolutionary canvas. In the remainder of the book, we will explore these themes again and again.

4.10 Problems

Key to the problem categories: • Estimates • Data interpretation

• 4.1 Structure of hemoglobin and myoglobin

(a) As in Problem 2.6, obtain the atomic coordinates for hemoglobin and myoglobin. Measure their dimensions, identify the different subunits and the heme groups.

(b) Expand the analysis of hemoglobin on p. 143 by calculating the mean spacing between hemoglobin molecules inside a red blood cell. How does this spacing compare with the size of a hemoglobin molecule?

(c) Typical results for a complete blood count (CBC) are shown in Table 4.1. Assume that an adult has roughly 5 L of blood in his or her body. Based on these values, estimate:
(i) the number of red blood cells;
(ii) the percentage in volume they represent in the blood;
(iii) their mean spacing;
(iv) the total amount of hemoglobin in the blood;
(v) the number of hemoglobin molecules per cell;
(vi) the number of white blood cells in the blood.

• 4.2 The number of phages on Earth

(a) An estimate for the number of phages on Earth can be obtained using data from Bergh et al. (1989). By taking water samples from lakes and oceans and counting the various phages, one arrives at counts of the order of 10^6–10^7 phages/mL. Using reasonable assumptions about the amount of water on Earth, make an estimate of the number of phages.

(b) Work out the mass of all of the phage particles on the Earth using your result from (a).

(Problem suggested by Sherwood Casjens.)

Table 4.1: Typical values from a CBC. (Adapted from R. W. Maxwell, Maxwell Quick Medical Reference, Maxwell Publishing, 2002.)

Test	Value
Red blood cell count (RBC)	Men: $\approx(4.3–5.7) \times 10^6$ cells/μL
	Women: $\approx(3.8–5.1) \times 10^6$ cells/μL
Hematocrit (HCT)	Men: ≈39–49%
	Women: ≈35–45%
Hemoglobin (HGB)	Men: ≈13.5–17.5 g/dL
	Women: ≈12.0–16.0 g/dL
Mean corpuscular hemoglobin (MCH)	≈26–34 pg/cell
MCH concentration (MCHC)	≈31–37%
Mean corpuscular volume (MCV)	≈80–100 fL
White blood cell count (WBC)	$\approx(4.5–11) \times 10^3$ cells/μL
Differential (% of WBC):	
Neutrophils	≈57–67
Lymphocytes	≈23–33
Monocytes	≈3–7
Eosinophils	≈1–3
Basophils	≈0–1
Platelets	$\approx(150–450) \times 10^3$ cell/μL

Table 4.2: Fraction of crossovers of six sex-linked factors in *Drosophila.* (Adapted from A. H. Sturtevant, *J. Exp. Zool.* 14:43, 1913.)

Factors	Fraction of crossovers
BR	115/324
B(C,O)	214/21736
(C,O)P	471/1584
(C,O)R	2062/6116
(C,O)M	406/898
PR	17/573
PM	109/458
BP	1464/4551
BM	260/693

• 4.3 Genetics by the numbers

Three successive nucleotides along a DNA molecule, called a *codon*, encode one amino acid. The weight of an *E. coli* DNA molecule is about 3.1×10^9 daltons. The average weight of a nucleotide pair is 660 daltons, and each nucleotide pair contributes 0.34 nm to the overall DNA length.

(a) What fraction of the mass of an *E. coli* cell corresponds to its genomic DNA?

(b) Calculate the length of the *E. coli* DNA. Comment on how it compares with the size of a single *E. coli* cell.

(c) Assume that the average protein in *E. coli* is a chain of 300 amino acids. What is the maximum number of proteins that can be coded by the *E. coli* DNA?

(d) On an alien world, the genetic code consists of two base pairs per codon. There are still four different bases. How many different amino acids can be encoded?

• 4.4 Mutation correlation and physical proximity on the gene

In Section 4.6.1, we briefly described Sturtevant's analysis of mutant flies that culminated in the generation of the first chromosome map. In Table 4.2, we show the crossover data associated with the different mutations that he used to draw the map. A crossover refers to a chromosomal rearrangement in which parts of two chromosomes exchange DNA. An illustration of the process is shown in Figure 4.26. The six factors looked at by Sturtevant are B, C, O, P, R, and M. Flies recessive in B, the black factor, have a yellow body color. Factors C and O are completely linked, they always go together and flies recessive in both of these factors have white eyes. A fly recessive in factor P has vermilion eyes instead of the ordinary red eyes. Finally, flies recessive in R have rudimentary wings and those recessive in M have miniature wings. For example, the fraction of flies that presented a crossover of the B and P factors is denoted

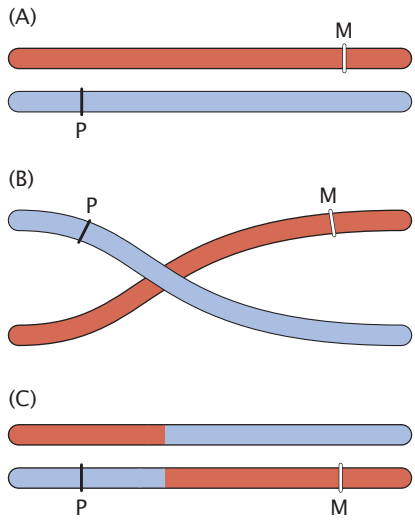


Figure 4.26: Crossing over of chromosomes. (A) Chromosomes before crossing over showing two loci labeled P and M. (B) Illustration of the crossing-over event. (C) Chromosomes after crossover.

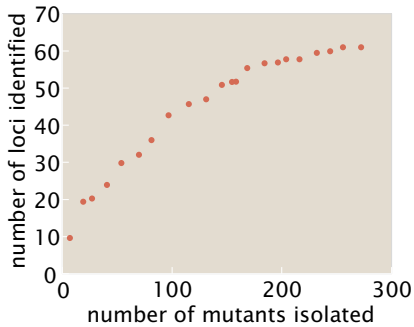


Figure 4.27: Saturation of a mutant library. Number of different identified loci as a function of the number of mutants isolated. (Adapted from C. Nusslein–Volhard et al., *Roux's Arch. Dev. Biol.* 193:267, 1984.)

as BP. Assume that the frequency of recombination is proportional to the distance between loci on the chromosome.

Reproduce Sturtevant's conclusions by drawing your own map using the first seven data points from Table 4.2.

Keep in mind that shorter "distances" are more reliable than longer ones because the latter are more prone to double crossings. Are distances additive? For example, can you predict the distance between B and P from looking at the distances B(C,O) and (C,O)P? What is the interpretation of the two last data points from Table 4.2?

• 4.5 Saturation of mutants in libraries

In a set of classic experiments, the second chromosome of *D. melanogaster* was mutagenized and the effects of these mutations characterized based on their phenotype in embryonic development. The experimenters found 272 mutants with phenotypes visibly different from wild-type embryos. However, when they determined the location of the mutations using the method outlined in Figure 4.21 and worked out in Problem 4.4, they discovered that these mutations only mapped to 61 different positions or loci on that chromosome. Figure 4.27 shows how, as more mutants were scored, ever more mutants corresponded to previously identified loci. Using a model that assumes a uniform probability of mutation in any locus, calculate the number of new loci found as a function of the number of mutants isolated. Explain the saturation effect and plot your results against the data. Provide a judgment on whether it is useful to continue searching for loci. (*Hint*: Start by writing down the probability that a specific locus has not been mapped after scoring the first M mutants). Relevant data for this problem are provided on the book's website.

• 4.6 Split genes and the discovery of introns

In a classic experiment, it was discovered that eukaryotic genomic DNA codes for proteins in a noncontinuous fashion, culminating in the idea of introns, regions in a given gene that are actually spliced out of the mRNA molecule before translation. The concept of the experiment was to hybridize the unspliced DNA with a DNA mimic of the spliced mRNA. Using Figure 4.28, estimate the sizes of the spliced regions.

• 4.7 Open reading frames in random DNA

In this problem, we will compute the probabilities of finding specific DNA sequences in a perfectly random genome, for which we assume that the four different nucleotides appear randomly and with equal probability.

(a) From the genetic code shown in Figure 1.4, compute the probability that a randomly chosen sequence of three nucleotides will correspond to a stop codon. Similarly, what is the probability of a randomly chosen sequence corresponding to a start codon?

(b) A reading frame refers to one of three possible ways that a sequence of DNA can be divided into consecutive triplets of nucleotides. An open reading frame (ORF) is a reading frame that contains a start codon and does not contain a stop codon for at least some minimal length of codons. Derive a formula for the probability of an ORF having a length of N codons (not including the stop codon).

(c) The genome of *E. coli* is approximately 5×10^6 bp long and is circular. Again assuming a that the genome is a random configuration of base pairs, how many ORFs of length 1000 bp (a typical protein size) would be expected by chance? Note that there are six possible reading frames.

(Problem courtesy of Sharad Ramanathan)

4.11 Further reading

Davis, RH (2003) The Microbial Models of Molecular Biology: From Genes to Genomes, Oxford University Press. This book examines the role of microbial model systems.

Johnson, GB (1996) How Scientists Think – Twenty-one Experiments That Have Shaped Our Understanding of Genetics and Molecular Biology, Wm C. Brown Publishers. This book makes for fascinating reading and describes many of the classic experiments discussed in this chapter in more detail.

Tanford, C, & Reynolds, J (2001) Nature's Robots: A History of Proteins, Oxford University Press. This book gives an

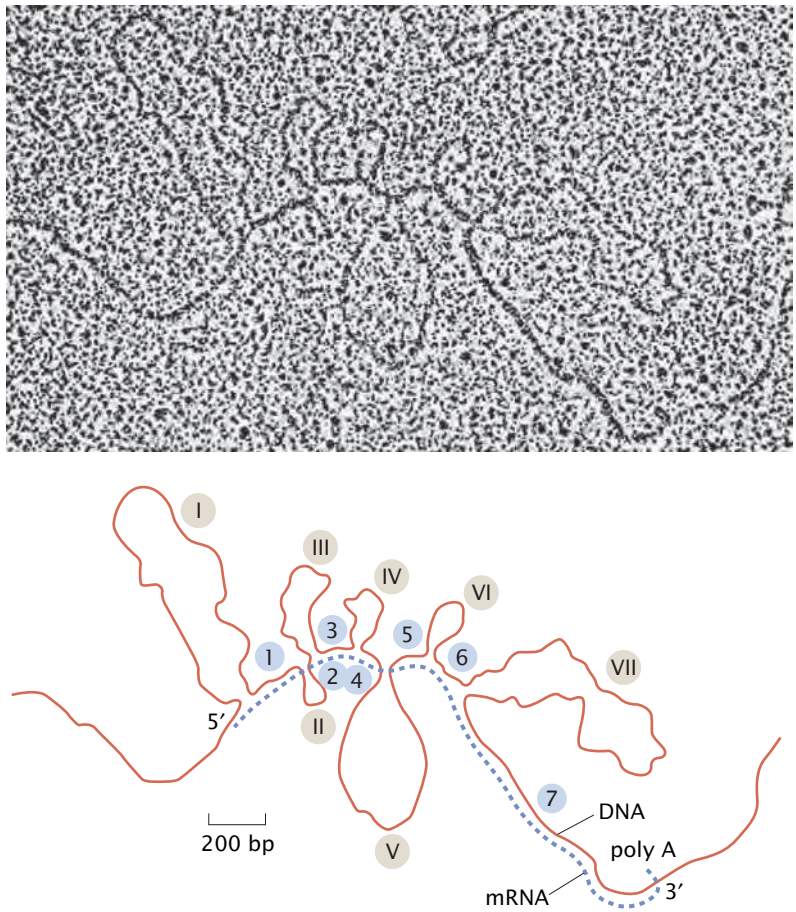


Figure 4.28: Experiment by Chambon to demonstrate the existence of introns. Unspliced DNA is hybridized to spliced mRNA. The loops correspond to regions of the DNA that have been removed in the mRNA. (Adapted from P. Chambon, *Sci. Am.* 244(5):60, 1981.)

informative account of the history of ideas associated with the determination of the nature of proteins. Within the context of the present chapter, the reader will find that hemoglobin enters this historical stage with great regularity.

Dickerson, RE (2005) Present at the Flood: How Structural Molecular Biology Came About, Sinauer Associates. This wonderful book is full of commentaries and reprints from the revolution in structural biology. For the purposes of the present chapter, it is Dickerson's story of "How to Solve a Protein Structure" with particular emphasis on the globins that touches our story.

Berg, H (2000) Motile Behavior of Bacteria, *Phys. Today*, January. Berg provides a masterful, though brief, account of the nature of motility in *E. coli*, describing many of the topics developed in the present chapter. In addition, Berg has written two wonderful books, *E. coli* in Motion (2003, Springer) and Random Walks in Biology (1993, Princeton University Press), both of which explain many fascinating features of bacteria, including chemotaxis.

Echols, H (2001) Operators and Promoters: The Story of Molecular Biology and Its Creators, University of California Press. Echols' book is a profoundly interesting and unique description of the history of much of molecular biology. In the context of the present chapter, it has a number of useful insights into the use of phages and *E. coli* as model systems, and describes topics such as gene regulation masterfully.

Judson, HF (1996) The Eighth Day of Creation: The Makers of the Revolution in Biology, Cold Spring Harbor Laboratory Press. Judson's book, like that of Echols, recounts the history of the development of molecular biology and describes many of the experiments and advances discussed in the present chapter. This book is fascinating.

Barnett, JA (2003) Beginnings of microbiology and biochemistry: the contribution of yeast research, *Microbiology* **149**, 557. Barnett has written a series of articles on the history of yeast in the emergence of biochemistry and microbiology. This article gives a flavor of the more complete set of articles, which we highly recommend.

Ghaemmaghami, S, Huh, W-K, Bower, K, et al. (2003) Global analysis of protein expression in yeast, *Nature* **425**, 737. This paper gives a flavor of the genome-wide yeast experiments that permit the study of nearly the full complement of yeast proteins.

Kornberg, A (1991) For the Love of Enzymes: The Odyssey of a Biochemist, Harvard University Press. This book describes Kornberg's work on DNA replication and much more and includes the charming story (see p. 106) which gives our chapter its title.

Weiner, J (1999) Time, Love, Memory: A Great Biologist and His Quest for the Origins of Behavior, Alfred A. Knopf. In the present chapter, we described the way in which *D. melanogaster* has repeatedly served as a model system for characterizing both genetics and development in eukaryotic systems. Weiner's book provides a thoughtful and interesting side-by-side biography of *Drosophila* and Seymour Benzer, who used this organism as the basis of his studies of the biology of behavior.

4.12 References

Alberts, B, Johnson, A, Lewis, J, et al. (2008) Molecular Biology of the Cell, 5th ed., Garland Science.

Berg, P (1993) Dissections and reconstructions of genes and chromosomes. In Nobel Lectures, Chemistry 1971–1980, World Scientific.

Bergh, O, Borsheim, KY, Bratbak, G, & Heldal, M (1989) High abundance of viruses found in aquatic environments, *Nature* **340**, 467.

Brunori, M (1999) Hemoglobin is an honorary enzyme, *Trends Biochem. Sci.* **24**, 158.

Chambon, P (1981) Split genes, *Scientific American* **244**(5), 60.

Chant, J, & Herskowitz, I (1991) Genetic control of bud site selection in yeast by a set of gene products that constitute a morphogenetic pathway, *Cell* **65**, 1203.

Dickerson, RE, & Geis, I (1983) Hemoglobin: Structure, Function, Evolution and Pathology, The Benjamin/Cummings Publishing Company, Inc. This lovely book is a testament to the status of hemoglobin as a model system. Chapter 2 gives an excellent discussion of the key role of hemoglobin as a model system with relevance to physiology, structural biology and the biophysics of allostery and cooperativity.

Grimes, S, Jardine, PJ, & Anderson, D (2002) Bacteriophage phi29 DNA packagings, *Adv. Virus Res.* **58**, 255.

Imai, K (1990) Precision determination and Adair scheme analysis of oxygen equilibrium curves of concentrated hemoglobin solution. A strict examination of Adair constant evaluation methods, *Biophys. Chem.* **37**, 197.

Jeong, H, Mason, SP, Barabasi, AL, & Oltvai, ZN (2001) Lethality and centrality in protein networks, *Nature* **411**, 41.

Klotz, IM (1997) Ligand–Receptor Energetics – A Guide for the Perplexed, John Wiley and Sons.

Maxwell, RW (2002) Maxwell Quick Medical Reference, 5th ed., Maxwell Publishing Company.

Menssen R, Neutzner A, & Seufert, W (2001) Asymmetric Spindle pole localization of yeast Cdc15 Kinase links mitotic exit and Cytokinesis, *Curr. Biol.* **11**(5), 345.

Novick, P, & Schekman, R (1979) Secretion and cell-surface growth are blocked in a temperature sensitive mutant of *Saccharomyces cerevisiae*, *Proc. Natl Acad. Sci. USA* **76**, 1858.

Nusslein-Volhard, C, Wieschaus, E, & Kluding, H (1984) Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. 1. Zygotic loci on the 2nd chromosome, *Roux's Arch. Dev. Biol.* **193**, 267.

Rossi-Fanelli, A, & Antonini, E (1958) Studies on the oxygen and carbon monoxide equilibria of human myoglobin, *Arch. Biochem. Biophys.* **77**, 478.

Sturtevant, AH (1913) The linear arrangement of six sex-linked factors in *Drosophila* as shown by their mode of association, *J. Exp. Zool.* **14**, 43.

Zhimulev, IF, Belyaeva, ES, Semeshin, VF, et al. (2004) Polytene chromosomes: 70 years of genetic research, *Int. Rev. Cytol.* **241**, 203.