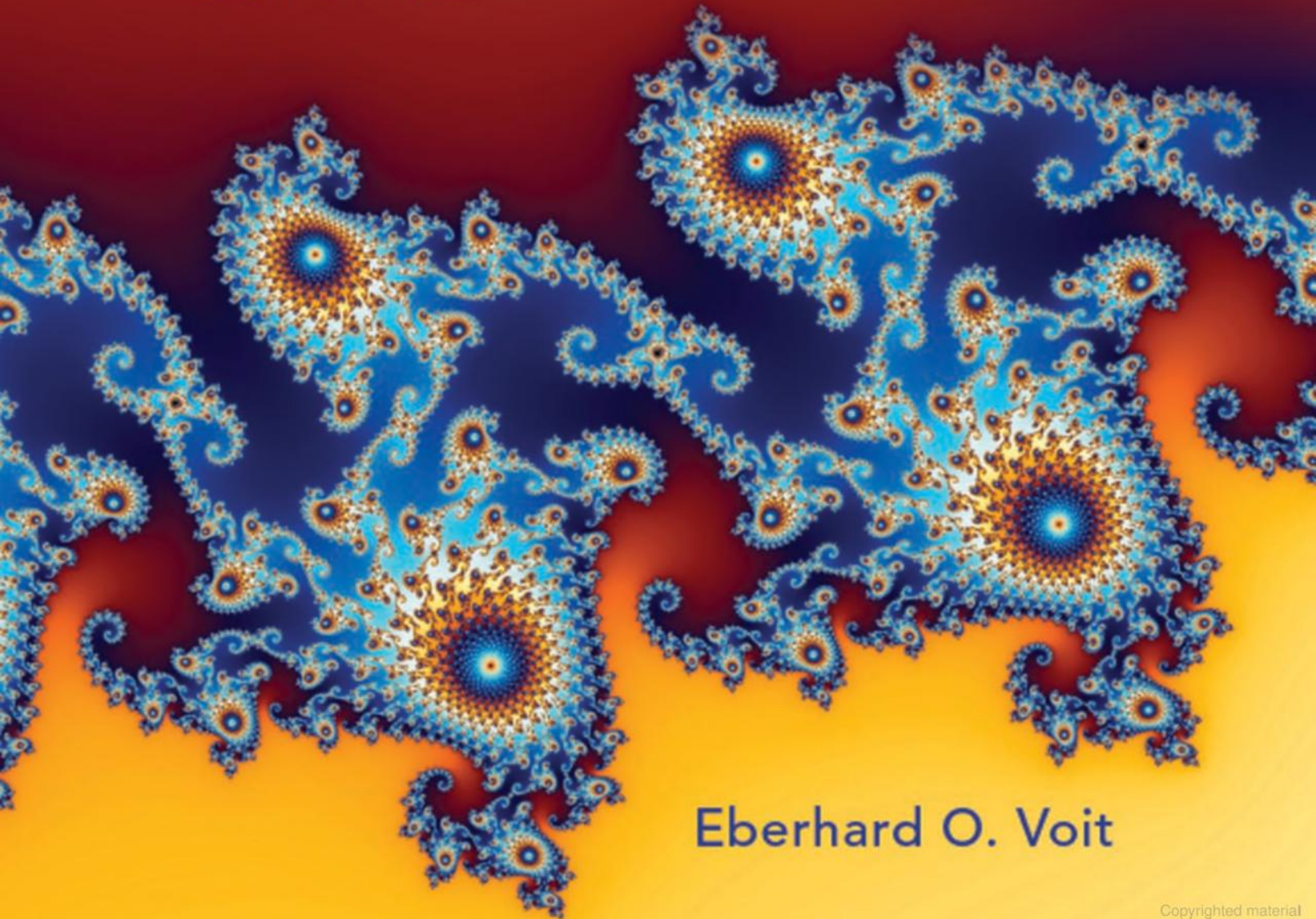


A FIRST COURSE IN
**SYSTEMS
BIOLOGY**

SECOND
EDITION



Eberhard O. Voit

A FIRST COURSE IN
SYSTEMS
BIOLOGY

SECOND
EDITION

Eberhard O. Voit

 **Garland Science**
Taylor & Francis Group
NEW YORK AND LONDON

Garland Science

Vice President: Denise Schanck
Senior Editorial Assistant: Katie Laurentiev
Assistant Editor: David Borrowdale
Production Editor: Georgina Lucas
Illustrations: Nigel Orme
Copyeditor: H.M. (Mac) Clarke
Typesetting: NovaTechset Pvt Ltd
Proofreader: Sally Huish
Indexer: Nancy Newman
Cover Design: Andrew Magee

Front cover image. The beautiful geometric shape of the fractal is called self-similar because it has the same appearance at smaller and smaller scales. It reminds us of fundamental design features like feedback loops that we encounter at many organizational levels of biological systems. Fractals are generated with nonlinear recursive models, and they are discussed with simpler examples in Chapter 4. (Courtesy of Wolfgang Beyer under Creative Commons Attribution-Share Alike 3.0 Unported license.)

© 2018 by Garland Science, Taylor & Francis Group, LLC

This book contains information obtained from authentic and highly regarded sources. Reprinted material is quoted with permission, and sources are indicated. Reasonable efforts have been made to publish reliable data and information, but the author and the publisher cannot assume responsibility for the validity of all materials or for the consequences of their use. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means – graphic, electronic, or mechanical, including photocopying, recording, taping, or information storage and retrieval systems – without permission of the copyright holder.

ISBN 978-0-8153-4568-8

Library of Congress Cataloging-in-Publication Data

Names: Voit, Eberhard O., author.
Title: A first course in systems biology / Eberhard O. Voit.
Description: Second edition. | New York : Garland Science, 2017.
Identifiers: LCCN 2017017580 | ISBN 9780815345688 (alk. paper)
Subjects: LCSH: Systems biology. | Computational biology.
Classification: LCC QH324.2 .V65 2017 | DDC 570.1/13–dc23
LC record available at <https://lcn.loc.gov/2017017580>

Published by Garland Science, Taylor & Francis Group, LLC, an informa business,
711 Third Avenue, 8th Floor, New York NY 10017, USA,
and 2 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 website at <http://www.garlandscience.com>

Contents

Chapter 1: Biological Systems	1		
Reductionism and Systems Biology	5		
Even Simple Systems Can Confuse Us	8		
Why Now?	10		
Communicating Systems Biology	13		
The Task Before Us	16		
Exercises	17		
References	17		
Further Reading	18		
Chapter 2: Introduction to Mathematical Modeling	19		
Goals, Inputs, and Initial Exploration	24		
2.1 Questions of Scale	24		
2.2 Data Availability	25		
Model Selection and Design	26		
2.3 Model Structure	27		
2.4 System Components	30		
2.5 Model Equations	35		
2.6 Parameter Estimation	36		
Model Analysis and Diagnosis	37		
2.7 Consistency and Robustness	38		
2.8 Exploration and Validation of Dynamical Features	40		
Model Use and Applications	43		
2.9 Model Extensions and Refinements	43		
2.10 Large-Scale Model Assessments	45		
2.11 Questions of Design	46		
2.12 Simplicity versus Complexity	47		
Exercises	49		
References	50		
Further Reading	50		
Chapter 3: Static Network Models	51		
Strategies of Analysis	52		
Interaction Graphs	53		
3.1 Properties of Graphs	54		
3.2 Small-World Networks	58		
Dependencies Among Network Components	62		
3.3 Causality Analysis	62		
3.4 Mutual Information	62		
Bayesian Reconstruction of Interaction Networks	63		
3.5 Application to Signaling Networks	66		
3.6 Applications to Other Biological Networks	69		
Static Metabolic Networks and Their Analysis	69		
3.7 Stoichiometric Networks	70		
3.8 Variants of Stoichiometric Analysis	73		
3.9 Metabolic Network Reconstruction	73		
3.10 Metabolic Control Analysis	74		
Exercises	78		
References	80		
Further Reading	82		
Chapter 4: The Mathematics of Biological Systems	83		
Discrete Linear Systems Models	85		
4.1 Recursive Deterministic Models	85		
4.2 Recursive Stochastic Models	88		
Discrete Nonlinear Systems	91		
Continuous Linear Systems	93		
4.3 Linear Differential Equations	94		
4.4 Linearized Models	95		
Continuous Nonlinear Systems	100		
4.5 Ad hoc Models	101		
4.6 Canonical Models	102		
4.7 More Complicated Dynamical Systems Descriptions	110		
Standard Analyses of Biological Systems Models	110		
4.8 Steady-State Analysis	110		
4.9 Stability Analysis	115		
4.10 Parameter Sensitivity	118		

4.11 Analysis of Systems Dynamics	119	6.13 Transcription Factors	188
Other Attractors	122	6.14 Models of Gene Regulation	190
4.12 Limit Cycles	123	Measuring Gene Expression	191
4.13 Chaotic Attractors	126	Localization of Gene Expression	194
Exercises	128	Outlook	196
References	132	Exercises	196
Further Reading	133	References	198
		Further Reading	200
Chapter 5: Parameter Estimation	135	Chapter 7: Protein Systems	201
Parameter Estimation for Linear Systems	136	Chemical and Physical Features of Proteins	202
5.1 Linear Regression Involving a Single Variable	136	7.1 Experimental Protein Structure Determination and Visualization	206
5.2 Linear Regression Involving Several Variables	138	An Incomplete Survey of the Roles and Functions of Proteins	208
Parameter Estimation for Nonlinear Systems	141	7.2 Enzymes	209
5.3 Comprehensive Grid Search	143	7.3 Transporters and Carriers	211
5.4 Nonlinear Regression	145	7.4 Signaling and Messenger Proteins	214
5.5 Genetic Algorithms	146	7.5 Proteins of the Immune System	215
5.6 Other Stochastic Algorithms	148	7.6 Structure Proteins	216
5.7 Typical Challenges	149	Current Challenges in Protein Research	218
Parameter Estimation for Systems of Differential Equations	153	7.7 Proteomics	218
Structure Identification	160	7.8 Structure and Function Prediction	220
Exercises	161	7.9 Localization	222
References	166	7.10 Protein Activity and Dynamics	224
Further Reading	167	Exercises	226
		References	228
		Further Reading	230
Chapter 6: Gene Systems	169	Chapter 8: Metabolic Systems	231
The Central Dogma	169	Biochemical Reactions	232
Key Properties of DNA and RNA	171	8.1 Background	232
6.1 Chemical and Physical Features	171	8.2 Mathematical Formulation of Elementary Reactions	234
6.2 Size and Organization of DNA	174	8.3 Rate Laws	235
6.3 Genes and Noncoding DNA	175	Pathways and Pathway Systems	240
6.4 Eukaryotic DNA Packing	178	8.4 Biochemistry and Metabolomics	240
6.5 Epigenetics	178	8.5 Resources for Computational Pathway Analysis	241
RNA	178	8.6 Control of Pathway Systems	244
6.6 Messenger RNA (mRNA)	179	Methods of Metabolomic Data Generation	246
6.7 Transfer RNA (tRNA)	182	8.7 Sampling, Extraction, and Separation Methods	247
6.8 Ribosomal RNA (rRNA)	182	8.8 Detection Methods	247
6.9 Small RNAs	183	8.9 Flux Analysis	249
6.10 RNA Viruses	184		
Gene Regulation	185		
6.11 The <i>lac</i> Operon	186		
6.12 Modes of Regulation	187		

From Data to Systems Models	250	Chapter 11: Integrative Analysis of Genome, Protein, and Metabolite Data: A Case Study in Yeast	303
8.10 Case Study 1: Analyzing Metabolism in an Incompletely Characterized Organism	250	On the Origin of Models	304
8.11 Case Study 2: Metabolic Network Analysis	251	A Brief Review of the Heat Stress Response in Yeast	306
8.12 Case Study 3: Extraction of Dynamic Models from Experimental Data	251	11.1 The Trehalose Cycle	308
Exercises	252	Modeling Analysis of the Trehalose Cycle	310
References	254	11.2 Design and Diagnosis of a Metabolic Pathway Model	310
Further Reading	255	11.3 Analysis of Heat Stress	312
Chapter 9: Signaling Systems	257	11.4 Accounting for Glucose Dynamics	314
Static Models of Signal Transduction Networks	259	11.5 Gene Expression	315
9.1 Boolean Networks	259	Multiscale Analysis	318
9.2 Network Inference	261	11.6 <i>In Vivo</i> NMR Profiles	318
Signal Transduction Systems Modeled with Differential Equations	261	11.7 Multiscale Model Design	320
9.3 Bistability and Hysteresis	261	11.8 The Trehalase Puzzle	324
9.4 Two-Component Signaling Systems	266	Concluding Comments	327
9.5 Mitogen-Activated Protein Kinase Cascades	270	Exercises	328
9.6 Adaptation	273	References	329
9.7 Other Signaling Systems	274	Further reading	330
Exercises	278	Chapter 12: Physiological Modeling: The Heart as an Example	331
References	279	Hierarchy of Scales and Modeling Approaches	332
Further reading	281	12.1 Basics of Heart Anatomy	333
Chapter 10: Population Systems	283	12.2 Modeling Targets at the Organ Level	334
Population Growth	283	12.3 Modeling Targets at the Tissue Level	335
10.1 Traditional Models of Population Growth	284	12.4 Modeling Targets at the Cell Level	337
10.2 More Complex Growth Phenomena	286	Simple Models of Oscillations	339
Population Dynamics Under External Perturbations	288	12.5 Black-Box Models of Oscillations	339
Analysis of Subpopulations	289	12.6 Summary of Black-Box Oscillation Models	342
Interacting Populations	292	12.7 From a Black Box to Meaningful Models	343
10.3 General Modeling Strategy	292	Electrochemistry in Cardiomyocytes	345
10.4 Phase-Plane Analysis	292	12.8 Biophysical Description of Electrochemical Processes at the Membrane of Cardiomyocytes	347
10.5 More Complex Models of Population Dynamics	297	12.9 Resting Potentials and Action Potentials	348
Exercises	299	12.10 Models of Action Potentials	350
References	301	12.11 Repeated Heartbeats	354
Further reading	302	Issues of a Failing Heart	355

12.12 Modeling Heart Function and Failure Based on Molecular Events	356	14.3 Design Principles	404
Outlook for Physiological Multiscale Modeling	361	14.4 Operating Principles	406
Exercises	362	Goal-Oriented Manipulations and Synthetic Design of Biological Systems	407
References	365	14.5 Metabolic Engineering	407
Further Reading	366	14.6 Synthetic Biology	408
Chapter 13: Systems Biology in Medicine and Drug Development	369	Case Studies of Synthetic Biological Systems Designs	411
Are you Unique?	369	14.7 Elementary Mode Analysis in Metabolic Engineering	411
13.1 Biological Variability and Disease	369	14.8 Drug Development	414
13.2 Modeling Variability and Disease	370	14.9 Gene Circuits	415
Personalized Medicine and Predictive Health	372	The Future Has Begun	419
13.3 Data Needs and Biomarkers	373	Exercises	419
13.4 Personalizing Mathematical Models	374	References	421
The Drug Development Process	378	Further Reading	423
The Role of Systems Biology in Drug Development	380	Chapter 15: Emerging Topics in Systems Biology	425
13.5 Computational Target and Lead Identification	381	Emerging Applications	426
13.6 Receptor Dynamics	382	15.1 From Neurons to Brains	426
13.7 Pharmacokinetic Modeling	385	15.2 Complex Diseases, Inflammation, and Trauma	428
13.8 Pathway Screening with Dynamic Models	390	15.3 Organisms and their Interactions with the Environment	432
13.9 Emerging Roles of Systems Biology in Drug Development	393	Modeling Needs	435
Exercises	394	15.4 Multiscale Modeling	436
References	395	15.5 A Data-Modeling Pipeline	437
Further Reading	396	Toward a Theory of Biology . . . or Several Theories?	439
Chapter 14: Design of Biological Systems	399	References	441
Natural Design of Biological Systems	400	Further Reading	443
14.1 The Search for Structural Patterns	400	Glossary	445
14.2 Network Motifs	402	Index	459



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Biological Systems

1

When you have read this chapter, you should be able to:

- Describe the generic features of biological systems
- Explain the goals of systems biology
- Identify the complementary roles of reductionism and systems biology
- List those challenges of systems biology that cannot be solved with intuition alone
- Assemble a “to-do” list for the field of systems biology

When we think of biological systems, our minds may immediately wander to the Amazon rainforest, brimming with thousands of plants and animals that live with each other, compete with each other, and depend on each other. We might think of the incredible expanse of the world’s oceans, of colorful fish swimming through coral reefs, nibbling on algae. Two-meter-high African termite mounds may come to mind, with their huge colonies of individuals that have their specific roles and whose lives are controlled by an intricate social structure ([Figure 1.1](#)). We may think of an algae-covered pond with tadpoles and minnows that are about to restart yet another life cycle.

These examples are indeed beautiful manifestations of some of the fascinating systems nature has evolved. However, we don’t have to look that far to find biological systems. Much, much smaller systems are in our own bodies and even within our cells. Kidneys are waste-disposal systems. Mitochondria are energy-production systems. Ribosomes are intracellular machines that make proteins from amino acids. Bacteria are amazingly complicated biological systems. Viruses interact with cells in a well-controlled, systemic way. Even seemingly modest tasks often involve an amazingly large number of processes that form complicated control systems ([Figure 1.2](#)). The more we learn about the most basic processes of life, such as cell division or the production of a metabolite, the more we have to marvel the incredible complexity of the systems that facilitate these processes. In our daily lives, we usually take these systems for granted and assume that they function adequately, and it is only when, for example, disease strikes or algal blooms kill fish that we realize how complex biology really is and how damaging the failure of just a single component can be.

We and our ancestors have been aware of biological systems since the beginning of human existence. Human birth, development, health, disease, and death have long been recognized as interwoven with those of plants and animals, and with the environment. For our forebears, securing food required an understanding of seasonal changes in the ecological systems of their surroundings. Even the earliest forays into agriculture depended on detailed concepts and ideas of when and what to



Figure 1.1 Biological systems abound at all size scales. Here, a termite mound in Namibia is visible evidence of a complex social system. This system is part of a larger ecological system, and it is at once the host to many systems at smaller scales. (Courtesy of Lothar Herzog under the Creative Commons Attribution 2.0 Generic license.)

plant, how and where to plant it, how many seeds to eat or to save for sowing, and when to expect returns on the investment. Several thousand years ago, the Egyptians managed to ferment sugars to alcohol and used the mash to bake bread. Early pharmaceutical treatments of diseases certainly contained a good dose of superstition, and we are no longer convinced that rubbing on the spit of a toad during full moon will cure warts, but the beginnings of pharmaceutical science in antiquity and the Middle Ages also demonstrate a growing recognition that particular plant products can have significant and specific effects on the well-being or malfunctioning of the systems within the human body.

In spite of our long history of dealing with biological systems, our mastery of engineered systems far outstrips our capability to manipulate biological systems. We send spaceships successfully to faraway places and predict correctly when they will arrive and where they will land. We build skyscrapers exceeding by hundreds of

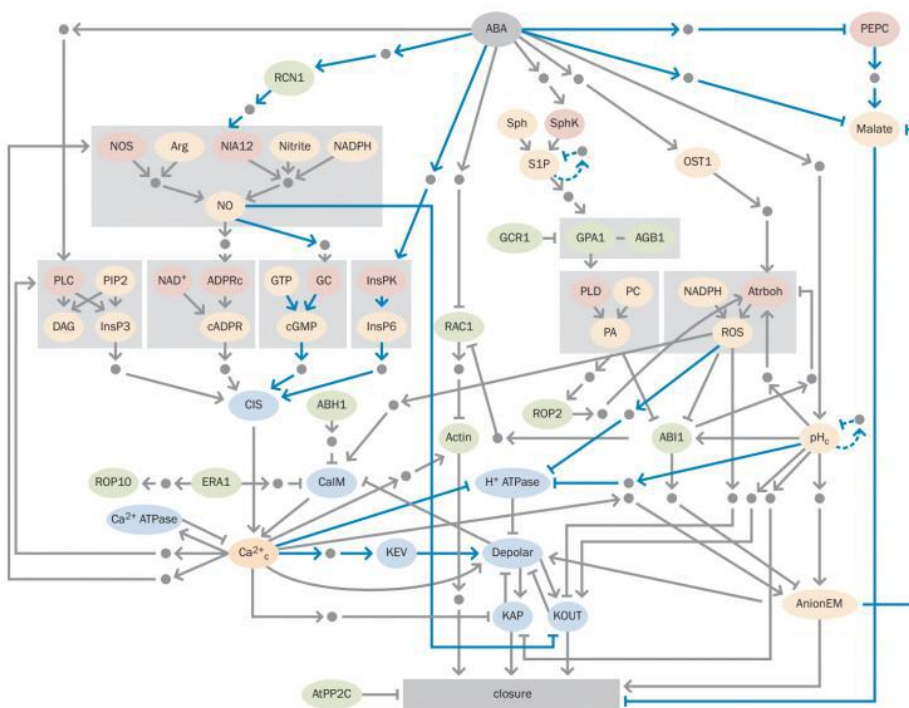


Figure 1.2 Diagram of a complicated system of molecules that coordinate the response of plants to drought. While the details are not important here, we can see that a key hormone, called abscisic acid (ABA), triggers a cascade of reactions that ultimately promote the closure of stomata and thereby reduce water evaporation [1]. Even a narrowly defined response like this closure process involves a complicated control system that contains a multitude of molecules and their interactions. In turn, this system is just one component within a much larger, physiological stress response system (cf. Figure 1.7). (From Saadatpour A, Albert I & Albert A. *J. Theor. Biol.* 266 [2010] 641–656. With permission from Elsevier.)

times the sizes of the biggest animals and plants. Our airplanes are faster, bigger, and more robust against turbulence than the most skillful birds. Yet, we cannot create new human cells or tissues from basic building blocks and we are seldom able to cure diseases except with rather primitive methods like cutting into the body or killing a lot of healthy tissue in the process, hoping that the body will heal itself afterwards. We can anticipate that our grandchildren will only shake their heads at such medieval-sounding, draconian measures. We have learned to create improved microorganisms, for instance for the bulk production of industrial alcohol or the generation of pure amino acids, but the methods for doing so rely on bacterial machinery that we do not fully understand and on artificially induced random mutations rather than targeted design strategies.

Before we discuss the roots of the many challenges associated with understanding and manipulating biological systems in a targeted fashion, and our problems predicting what biological systems will do under yet-untested conditions, we should ask whether the goal of a deeper understanding of biological systems is even worth the effort. The answer is a resounding “Yes!” In fact, it is impossible even to imagine the potential and scope of advances that might develop from biological systems analyses. Just as nobody during the eighteenth century could foresee the ramifications of the Industrial Revolution or of electricity, the Biological Revolution will usher in an entirely new world with incredible possibilities. Applications that are already emerging on the horizon are personalized medical treatments with minimal side effects, pills that will let the body regain control over a tumor that has run amok, prevention and treatment of neurodegenerative diseases, and the creation of spare organs from reprogrammed stem cells. A better understanding of ecological systems will yield pest- and drought-resistant food sources, as well as means for restoring polluted soil and water. It will help us understand why certain species are threatened and what could be done effectively to counteract their decline. Deeper insights into aquatic systems will lead to cleaner water and sustainable fisheries. Reprogrammed microbes or nonliving systems composed of biological components will dominate the production of chemical compounds from prescription drugs to large-scale industrial organics, and might create energy sources without equal. Modified viruses will become standard means for supplying cells with healthy proteins or replacement genes. The rewards of discovering and characterizing the general principles and the specifics of biological systems will truly be unlimited.

If it is possible to engineer very sophisticated machines and to predict exactly what they will do, why are biological systems so different and difficult? One crucial difference is that we have full control over engineered systems, but not over biological systems. As a society, we collectively know all details of all parts of engineered machines, because we made them. We know their properties and functions, and we can explain how and why some engineer put a machine together in a particular fashion. Furthermore, most engineered systems are modular, with each module being designed for a unique, specific task. While these modules interact with each other, they seldom have multiple roles in different parts of the system, in contrast to biology and medicine, where, for instance, the same lipids can be components of membranes and have complicated signaling functions, and where diseases are often not restricted to a single organ or tissue, but may affect the immune system and lead to changes in blood pressure and blood chemistry that secondarily cause kidney and heart problems. A chemical refinery looks overwhelmingly complicated to a layperson, but for an industrial engineer, every piece has a specific, well-defined role within the refinery, and every piece or module has properties that were optimized for this role. Moreover, should something go wrong, the machines and factories will have been equipped with sensors and warning signals pinpointing problems as soon as they arise and allowing corrective action.

In contrast to dealing with sophisticated, well-characterized engineered systems, the analysis of biological systems requires investigations in the opposite direction. This type of investigation resembles the task of looking at an unknown machine and predicting what it does (Figure 1.3). Adding to this challenge, all scientists collectively know only a fraction of the components of biological systems, and the specific roles and interactions between these components are often obscure and change over time. Even more than engineered systems, biological systems are full of sensors and signals that indicate smooth running or ensuing problems, but in most



Figure 1.3 Analyzing a biological system resembles the task of determining the function of a complicated machine that we have never seen before. Shown here as an example is the cesium fountain laser table of the United States Naval Observatory, which is used to measure time with extreme accuracy. This atomic clock is based on transitions in cesium, which have a frequency of 9,192,631,770 Hz and are used to define the second. See also [2].

cases our experiments cannot directly perceive and measure these signals and we can only indirectly deduce their existence and function. We observe organisms, cells, or intracellular structures as if from a large distance and must deduce from rather coarse observations how they might function or why they fail.

What exactly is it that makes biological systems so difficult to grasp? It is certainly not just size. **Figure 1.4** shows two networks. One shows the vast highway system of the continental United States, which covers several million miles of major

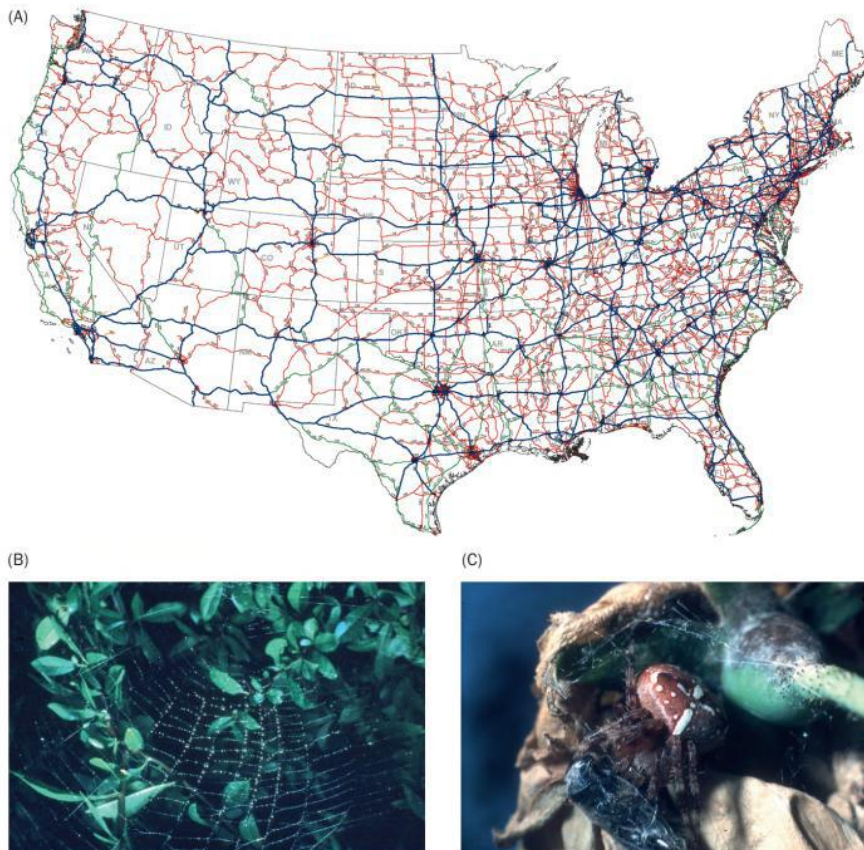


Figure 1.4 The size of a network or system is not necessarily correlated with its complexity. (A) The network of major highways in the continental United States covers over 3 million square miles. Nonetheless, its functionality is easy to grasp, and problems with a particular road are readily ameliorated with detours. (B) The web of the European diadem spider (*Araneus diadematus*) (C) is comparatively small, but the functional details of this little network are complex. Some lines are made of silk proteins that have the tensile strength of steel but can also be eaten and recycled by the spider; other lines are adhesive due to a multipurpose glue that may be sticky or rubbery depending on the situation; yet others are guide and signal lines that allow the spider to move about and sense prey. The creation of the web depends on different types of spinneret glands, whose development and function require the complex molecular machinery of the spider, and it is not yet clear how the instructions for the complicated construction, repair, and use of the web are encoded and inherited from one generation to the next. ((A) From the United States Department of Transportation.)

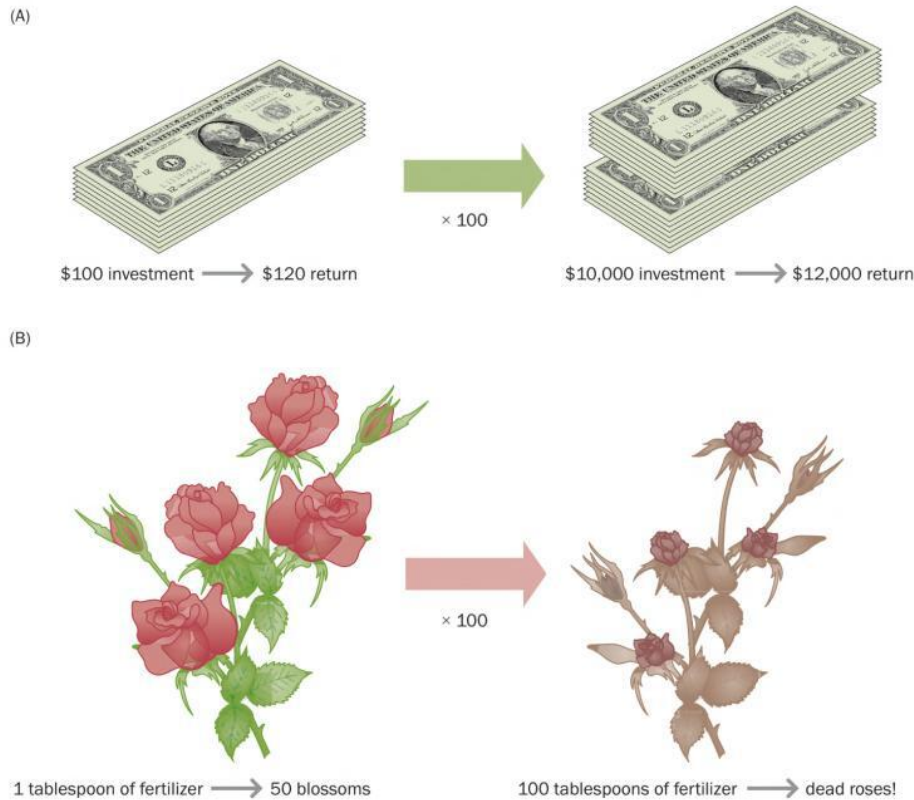


Figure 1.5 Biological phenomena are often difficult to understand, because our minds are trained to think linearly. (A) The return on an investment grows (or decreases) linearly with the amount invested. (B) In biology, more is not necessarily better. Biological responses often scale within a modest range, but lead to an entirely different response if the input is increased a lot.

highways. It is a very large system, but it is not difficult to understand its function or malfunction: if a highway is blocked, it does not take much ingenuity to figure out how to circumvent the obstacle. The other network is a comparably tiny system: the web of a diadem spider. While we can observe the process and pattern with which Ms. Spider spins her web, we do not know which neurons in her brain are responsible for different phases of the complicated web production process and how she is able to produce the right chemicals for the spider silk, which in itself is a marvel of material science, let alone how she manages to survive, multiply, and maybe even devour her husband.

Biological systems often consist of large numbers of components, but they pose an additional, formidable challenge to any analysis, because the processes that govern them are not linear. This is a problem, because we are trained to think in linear ways: if an investment of \$100 leads to a return of \$120, then an investment of \$10,000 leads to a return of \$12,000. Biology is different. If we fertilize our roses with 1 tablespoon of fertilizer and the rose bushes produce 50 blossoms, a little bit more fertilizer may increase the number of blossoms, but 100 tablespoons of fertilizer will not produce 5000 blossoms but almost certainly kill the plants (**Figure 1.5**). Just a small amount of additional sun exposure turns a tan into sunburn. Now imagine that thousands of components, many of which we do not know, respond in such a fashion, where a small input does not evoke any response, more input evokes a physiological response, and a little bit more input causes the component to fail or exhibit a totally different “stress” response. We will return to this issue later in this and other chapters with specific examples.

REDUCTIONISM AND SYSTEMS BIOLOGY

So the situation is complicated. But because we humans are a curious species, our forebears did not give up on biological analysis and instead did what was doable, namely collecting information on whatever could be measured with the best current methods (**Figure 1.6**). By now, this long-term effort has resulted in an amazing list of biological parts and their roles. Initially, this list contained new plant and animal



Figure 1.6 Collecting information is the first step in most systems analyses. The eighteenth-century British explorer Captain James Cook sailed the Pacific Ocean and catalogued many plants and animal species that had never been seen before in Europe.

species, along with descriptions of their leaves, berries, and roots, or their body shapes, legs, and color patterns. These external descriptions were valuable, but did not provide specific clues on how plants and animals function, why they live, and why they die. Thus, the next logical step was to look inside—even if this required stealing bodies from the cemetery under a full moon! Cutting bodies open revealed an entirely new research frontier. What were all those distinct body parts and what did they do? What were organs, muscles, and tendons composed of? Not surprisingly, this line of investigation eventually led to the grand-challenge quest of discovering and measuring *all* parts of a body, the parts of the parts (. . . of the parts), as well as their roles in the normal physiology or pathology of cells, organs, and organisms. The implicit assumption of this reductionist approach was that knowing the building blocks of life would lead us to a comprehensive understanding of how life works.

If we fast-forward to the twenty-first century, have we succeeded and assembled a complete parts catalog? Do we know the building blocks of life? The answer is a combination of yes's and no's. The catalog is most certainly not complete, even for relatively simple organisms. Yet, we have discovered and characterized genes, proteins, and metabolites as the major building blocks. Scientists were jubilant when the sequencing of the human genome in the early years of this millennium was declared complete: we had identified the ultimate building blocks, our entire blueprint. It turned out to consist of roughly three billion nucleotide pairs of DNA.

The sequencing of the human genome was without any doubt an incredible achievement. Alas, there is much more to a human body than genes. So, the race for building blocks extended to proteins and metabolites, toward individual gene variations and an assortment of molecules and processes affecting gene expression, which changes in response to external and internal stimuli, during the day, and throughout our lifetimes. As a direct consequence of these ongoing efforts, our parts list continues to grow at a rapid pace: A parts catalog that started with a few organs now contains over 20,000 human genes, many more genes from other organisms, and hundreds of thousands of proteins and metabolites along with their variants. In addition to merely looking at parts in isolation, we have begun to realize that most biological components are affected and regulated by a variety of other components. The expression of a gene may depend on several transcription factors, metabolites, and a variety of small RNAs, as well as molecular, epigenetic attachments to its DNA sequence. It is reasonable to expect that the list of processes within the body is much larger than the number of components on our parts list. Biologists will not have to worry about job security any time soon!

The large number of components and processes alone, however, is not the only obstacle to understanding how cells and organisms function. After all, modern computers can execute gazillions of operations within a second. Our billions of telephones worldwide are functionally connected. We can make very accurate

predictions regarding a gas in a container, even if trillions of molecules are involved. If we increase the pressure on the gas without changing the volume of the container, we know that the temperature will rise, and we can predict by how much. Not so with a cell or organism. What will happen to it if the environmental temperature goes up? Nothing much may happen, the rise in temperature may trigger a host of physiological response processes that compensate for the new conditions, or the organism may die. The outcome depends on a variety of factors that collectively constitute a complex stress response system (Figure 1.7). Of course, the comparison to a gas is not

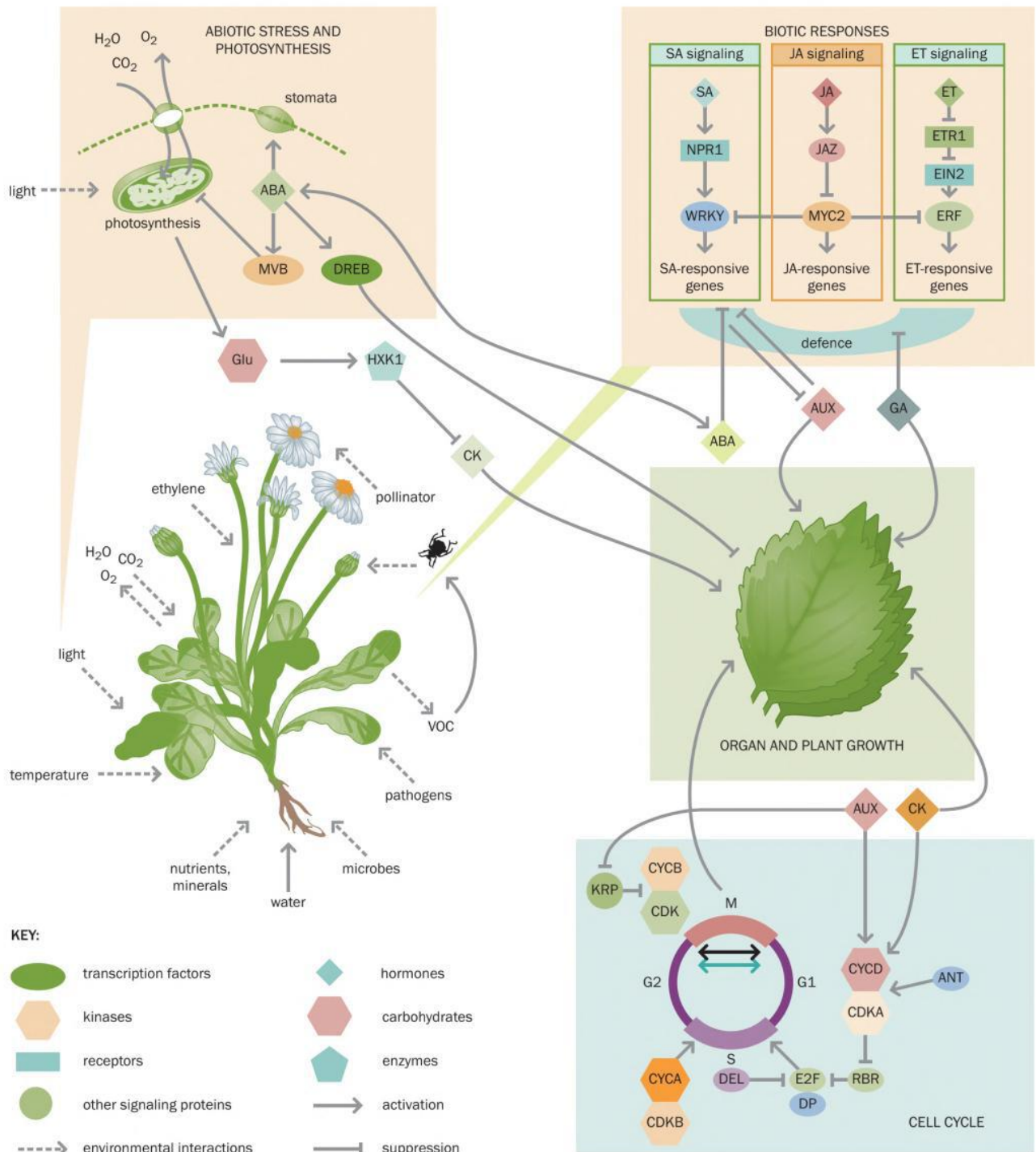


Figure 1.7 Stress responses are coordinated by systems at different levels of organization (cf. Figure 1.2). At the physiological level, the stress response system in plants includes changes at the cellular, organ, and whole-plant levels and also affects interactions of the plant with other species. (From Keurentjes JJB, Angenent GC, Dicke M, et al. *Trends Plant Sci.* 16 [2011] 183–190. With permission from Elsevier.)

quite fair, because, in addition to their large number, the components of a cell are not all the same, which drastically complicates matters. Furthermore, as mentioned earlier, the processes with which the components interact are nonlinear, and this permits an enormous repertoire of distinctly different behaviors with which an organism can respond to a perturbation.

EVEN SIMPLE SYSTEMS CAN CONFUSE US

It is easy to demonstrate how quickly our intuition can be overwhelmed by a few nonlinearities within a system. As an example, let's look at a simple chain of processes and compare it with a slightly more complicated chain that includes regulation [3]. The simple case merely consists of a chain of reactions, which is fed by an external input (Figure 1.8). It does not really matter what X , Y , and Z represent, but, for the sake of discussion, imagine a metabolic pathway such as glycolysis, where the input, glucose, is converted into glucose 6-phosphate, fructose 1,6-bisphosphate, and pyruvate, which is used for other purposes that are not of interest here. For illustrative purposes, let's explicitly account for an enzyme E that catalyzes the conversion of X into Y .

We will learn in the following chapters how one can formulate a model of such a pathway system as a set of differential equations. And while the details are not important here, it does not hurt to show such a model, which might read

$$\begin{aligned} \dot{X} &= \text{Input} - aEX^{0.5}, \\ \dot{Y} &= aEX^{0.5} - bY^{0.5}, \\ \dot{Z} &= bY^{0.5} - cZ^{0.5}. \end{aligned} \tag{1.1}$$

Here, X , Y , and Z are concentrations, E is the enzyme activity, and a , b , and c are rate constants that respectively represent how fast X is converted into Y , how fast Y is converted into Z , and how quickly material from the metabolite pool Z leaves the system. The dotted quantities on the left of the equal signs are differentials that describe the change in each variable over time, but we need not worry about them at this point. In fact, we hardly have to analyze these equations mathematically to get an idea of what will happen if we change the input, because intuition tells us that any increase in *Input* should lead to a corresponding rise in the concentrations of the intermediates X , Y , and Z , whereas a decrease in *Input* should result in smaller values of X , Y , and Z . The increases or decreases in X , Y , and Z will not necessarily be exactly of the same extent as the change in *Input*, but the direction of the change should be the same. The mathematical solution of the system in (1.1) confirms this intuition. For instance, if we reduce *Input* from 1 to 0.75, the levels of X , Y , and Z decrease, one after another, from their initial value of 1 to 0.5625 (Figure 1.9).

Now suppose that Z is a signaling molecule, such as a hormone or a phospholipid, that activates a transcription factor TF that facilitates the up-regulation of a gene G that codes for the enzyme catalyzing the conversion of X into Y (Figure 1.10). The simple linear pathway is now part of a functional loop. The organization of this loop is easy to grasp, but what is its effect? Intuition might lead us to believe that the positive-feedback loop should increase the level of enzyme E , which would result in more Y , more Z , and even more E , which would result in even more Y and Z . Would the concentrations in the system grow without end? Can we be sure about this prediction? Would an unending expansion be reasonable? What will happen if we increase or decrease the input as before?

The overall answer will be surprising: the information given so far does not allow us to predict particular responses with any degree of reliability. Instead, the answer depends on the numerical specifications of the system. This is bad news for the unaided human mind, because we are simply not able to assess the numerical consequences of slight changes in a system, even if we can easily grasp the logic of a system as in Figure 1.10.

To get a feel for the system, one can compute a few examples with an expanded model that accounts for the new variables (for details, see [3]). Here, the results are more important than the technical details. If the effect of Z on TF is weak, the

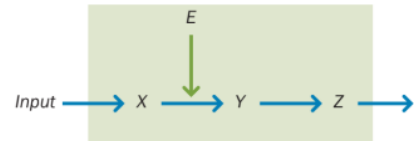


Figure 1.8 The human brain handles linear chains of causes and events very well. In this simple pathway, an external input is converted sequentially into X , Y , and Z , which leaves the system. The conversion of X into Y is catalyzed by an enzyme E . It is easy to imagine that any increase in *Input* will cause the levels of X , Y , and Z to rise.

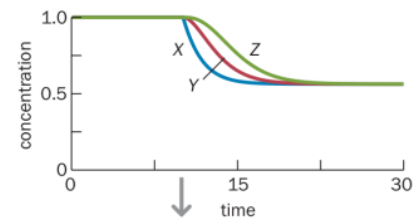


Figure 1.9 Simulations with the system in (1.1) confirm our intuition: X , Y , and Z reflect changes in *Input*. For instance, reducing *Input* in (1.1) to 75% at time 10 (arrow) leads to permanent decreases in X , Y , and Z .

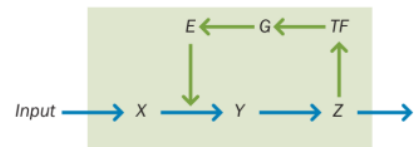


Figure 1.10 Even simple systems may not allow us to make reliable predictions regarding their responses to stimuli. Here, the linear pathway from Figure 1.8 is embedded into a functional loop consisting of a transcription factor TF and a gene G that codes for enzyme E . As described in the text, the responses to changes in *Input* are no longer obvious.

response to a decrease in *Input* is essentially the same as in Figure 1.9. This is not too surprising, because the systems in this case are very similar. However, if the effect of *Z* on *TF* is stronger, the concentrations in the system start to oscillate, and after a while these oscillations dampen away (Figure 1.11A). This behavior was not easy to predict. Interestingly, if the effect is further increased, the system enters a stable oscillation pattern that does not cease unless the system input is changed again (Figure 1.11B).

The hand-waving explanation of these results is that the increased enzyme activity leads to a depletion of *X*. A reduced level of *X* leads to lower levels of *Y* and *Z*, which in turn lead to a reduced effect on *TF*, *G*, and ultimately *E*. Depending on the numerical characteristics, the ups and downs in *X* may not be noticeable, they may be damped and disappear, or they may persist until another change is introduced. Intriguingly, even if we know that these alternative responses are possible, the unaided human mind is not equipped to integrate the numerical features of the model in such a way that we can predict which system response will ensue for a specific setting of parameters. A computational model, in contrast, reveals the answer in a fraction of a second.

The specific details of the example are not as important as the take-home message: If a system contains regulatory signals that form functional loops, we can no longer rely on our intuition for making reliable predictions. Alas, essentially all realistic systems in biology are regulated—and not just with one, but with many control loops. This leads to the direct and sobering deduction that intuition is not sufficient and that we instead need to utilize computational models to figure out how even small systems work and why they might show distinctly different responses or even fail, depending on the conditions under which they operate.

The previous sections have taught us that biological systems contain large numbers of different types of components that interact in potentially complicated ways and are controlled by regulatory signals. What else is special about biological systems? Many answers could be given, some of which are discussed throughout this book. For instance, two biological components are seldom 100% the same. They vary from one organism to the next and change over time. Sometimes these variations are inconsequential, at other times they lead to early aging and disease. In fact, most

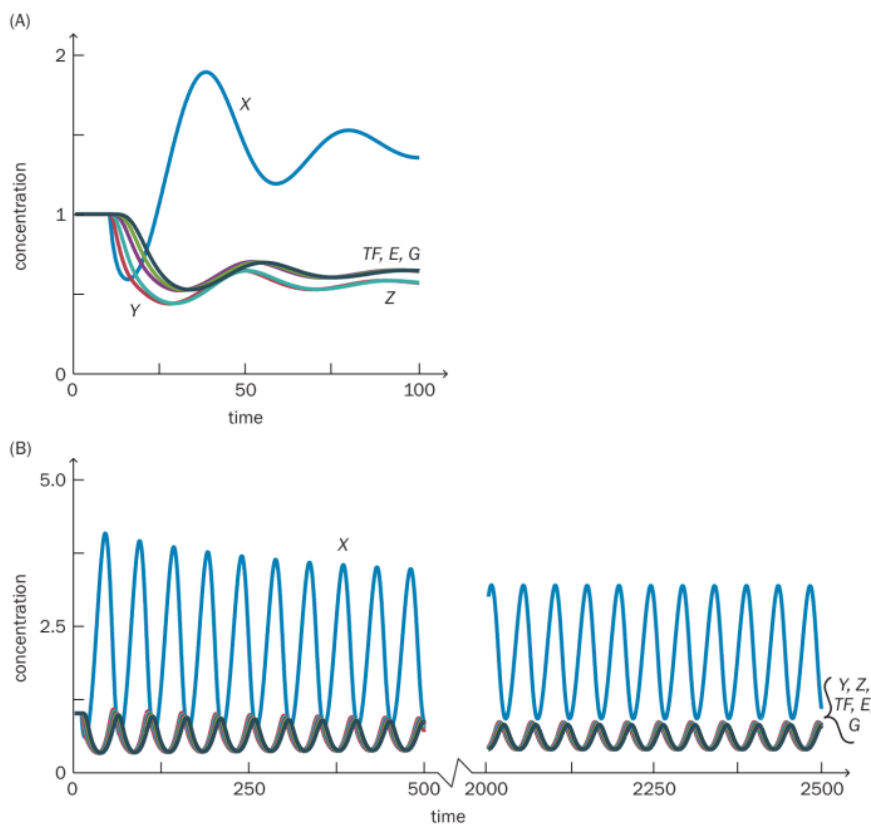


Figure 1.11 Simulation results demonstrate that the looped system in Figure 1.10 may exhibit drastically different responses. If the effect of *Z* on *TF* is very small, the response is essentially like that in Figure 1.9 (results not shown). (A) If the effect of *Z* on *TF* is relatively small, the functional feedback loop causes the system to go through damped oscillations before assuming a new stable state. (B) For stronger effects of *Z* on *TF*, the system response is a persistent oscillation.

diseases do not have a single cause, but are the consequence of an unfortunate combination of slight alterations in many components. Another feature that complicates intuition is the delay in many responses to stimuli. Such delays may be of the order of seconds, hours, or years, but they require the analyst to study not merely the present state of a biological system but also its history. For instance, recovery from a severe infection depends greatly on the preconditioning of the organism, which is the collective result of earlier infections and the body's responses [4].

Finally, it should be mentioned that different parts of biological systems may simultaneously operate at different scales, with respect to both time and space. These scales make some aspects of their analysis easier and some harder. Let's begin with the temporal scale. We know that biology at the most basic level is governed by physical and chemical processes. These occur on timescales of the order of milliseconds, if not faster. Biochemical processes usually run on a scale of seconds to minutes. Under favorable conditions, bacteria divide every 20–30 minutes. Our human lifespan extends to maybe 120 years, evolution can happen at the genetic level with lightning speed, for instance, when radiation causes a mutation, while the emergence of an entirely new species may take thousands or even millions of years. On one hand, the drastically different timescales make analyses complicated, because we simply cannot account for rapid changes in all molecules of an organism over an extended period of time. As an example, it is impossible to study aging by monitoring an organism's molecular state every second or minute. On the other hand, the differences in timescales justify a very valuable modeling "trick" [5, Chapter 5]. If we are interested in understanding some biochemical process, such as the generation of energy in the form of adenosine triphosphate (ATP) by means of the conversion of glucose into pyruvate, we can assume that developmental and evolutionary changes are so slow in comparison that they do not change during ATP production. Similarly, if we study the phylogenetic family tree of species, the biochemical processes in an individual organism are comparatively so fast that their details become irrelevant. Thus, by focusing on just the most relevant timescale and ignoring much faster and much slower processes, any modeling effort is dramatically simplified.

Biology also happens on many spatial scales. All processes have a molecular component, and their size scale is therefore of the order of ångströms and nanometers. If we consider a cell as the basic unit of life, we are dealing with a spatial scale of micrometers to millimeters, with some exceptions such as cotton "fiber" cells reaching the length of a few centimeters [6] and the afferent axons of nerve cells in giraffes, reaching from toe to neck, extending to 5 meters [7, p. 14]. The sizes of typical cells are dwarfed by higher plants and animals and by ecosystems such as our oceans, which may cover thousands of square kilometers. As with the different temporal scales, and using analogous arguments, models of biological systems often focus on one or two spatial scales at a time [5]. Nonetheless, such simplifications are not always applicable, and some processes, such as aging and algal blooms, may require the simultaneous consideration of several temporal and spatial scales. Such multiscale assessments are often very complicated and constitute a challenging frontier of current research (see Chapter 15).

WHY NOW?

Many of the features of biological systems have been known for quite a while, and, similarly, many concepts and methods of systems biology have their roots in its well-established parent disciplines, including physiology, molecular biology, biochemistry, mathematics, engineering, and computer science [8–11]. In fact, it has been suggested that the nineteenth-century scientist Claude Bernard might be considered the first systems biologist, since he proclaimed that the "application of mathematics to natural phenomena is the aim of all science, because the expression of the laws of phenomena should always be mathematical" [12, 13]. A century later, Ludwig von Bertalanffy reviewed in a book his three decades of attempting to convince biologists of the systemic nature of living organisms [14, 15]. At the same time, Mihajlo Mesarović used the term "Systems Biology" and declared that "real advance . . . will come about only when biologists start asking questions which are based on systems-theoretic concepts" [16]. The same year, a book review in *Science*

envisioned "... a field of systems biology with its own identity and in its own right" [17]. A few years later, Michael Savageau proposed an agenda for studying biological systems with mathematical and computational means [5].

In spite of these efforts, systems biology did not enter the mainstream for several more decades. Biology kept its distance from mathematics, computer science, and engineering, primarily because biological phenomena were seen as too complicated for rigorous mathematical analysis and mathematics was considered applicable only to very small systems of little biological relevance. The engineering of biological systems from scratch was impossible, and the budding field of computer science contributed to biology not much more than rudimentary data management.

So, why has systems biology all of the sudden moved to the fore? Any good detective will know the answer: motive and opportunity. The motive lies in the realization that reductionist thinking and experimentation alone are not sufficient if complex systems are involved. Reductionist experiments are very good in generating detailed information regarding specific components or processes of a system, but they often lack the ability to characterize, explain, or predict **emergent properties** that cannot be found in the parts of the system but only in their web of interactions. For instance, the emergence of oscillations in the example system represented by the equations in (1.1) cannot be credited to a single component of the system but is a function of its overall organization. Although we had complete knowledge of all details of the model pathway, it was very difficult to foresee its capacity either to saturate or oscillate in a damped or stable fashion. Biology is full of such examples.

A few years ago, Hirotada Mori's laboratory completed the assembly of a complete catalogue of single mutants in the bacterium *Escherichia coli* [18]. Yet, the scientific community is still not able to foresee which genes the bacterium will up- or down-regulate in response to new environmental conditions. Another very challenging example of emergent system properties is the central nervous system. Even though we understand quite well how action potentials are generated and propagated in individual neurons, we do not know how information flows, how memory works, and how diseases affect the normal functioning of the brain. It is not even clear how information in the brain is represented (see also [Chapter 15](#)). Thus, while reductionist biology has been extremely successful and will without any doubt continue to be the major driving force for future discovery, many biologists have come to recognize that the detailed pieces of information resulting from this approach need to be complemented with new methods of system integration and reconstruction [19].

The opportunity for systems biology is the result of the recent confluence and **synergism** of three scientific frontiers. The first is of course the rapid and vast accumulation of detailed biological information at the physiological, cellular, molecular, and submolecular levels. These targeted investigations of specific phenomena are accompanied by large-scale, high-throughput studies that were entirely infeasible just a couple of decades ago. They include quantification of genome-wide expression patterns, simultaneous identification of large arrays of expressed proteins, comprehensive profiling of cellular metabolites, characterization of networks of molecular interactions, global assessments of immune systems, and functional scans of nervous systems and the human brain. These exciting techniques are generating unprecedented amounts of high-quality data that are awaiting systemic interpretation and integration ([Figure 1.12](#)).

The second frontier is the result of ingenuity and innovation in engineering, chemistry, and material sciences, which have begun to provide us with a growing array of technologies for probing, sensing, imaging, and measuring biological systems that are at once very detailed, extremely specific, and usable *in vivo*. Many tools supporting these methods are in the process of being miniaturized, in some cases down to the nanoscale of molecules, which allows diagnoses with minute amounts of biological materials and one day maybe biopsies of individual, living cells. Devices at this scale will allow the insertion of sensing and disease treatment devices into the human body in an essentially noninvasive and harmless fashion [20–22]. Bioengineering and robotics are beginning to render it possible to measure hundreds or thousands of biomarkers from a single drop of blood. It is even becoming feasible to use molecular structures, prefabricated by nature, for new purposes in medicine, drug delivery, and biotechnology ([Figure 1.13](#)).

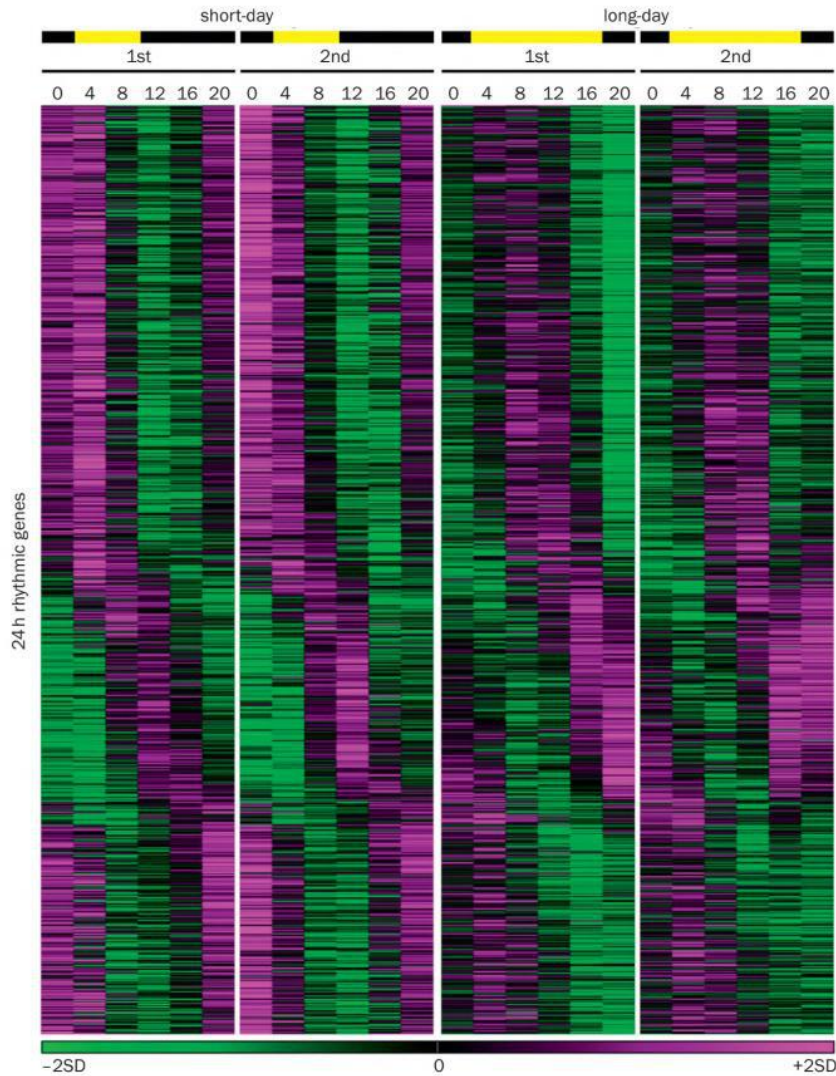


Figure 1.12 Modern high-throughput methods of molecular biology offer data in unprecedented quantity and quality.

As an example, the heat map shown here represents a genome-wide expression profile of 24-hour-rhythmic genes in the mouse under chronic short-day (left two panels) and long-day (right two panels) conditions. (From Masumoto KM, Ukai-Tadenuma M, Kasukawa T, et al. *Curr. Biol.* 20 [2010] 2199–2206. With permission from Elsevier.)

The third frontier is the co-evolution of mathematical, physical, and computational techniques that are more powerful and accessible to a much wider audience than ever before. Imagine that only a few decades ago computer scientists used punch cards that were read by optical card readers (**Figure 1.14**)! Now, there are even specific computing environments, including Mathematica® and MATLAB®, as well as different types of customized mark-up languages (XML), such as the systems biology mark-up language SBML [23] and the mark-up language AGML, which was developed specifically for analyzing two-dimensional gels in proteomics [24].

Before today's much more effective computer science techniques were available, it was not even possible to keep track of the many components of biological systems, let alone analyze them. But over the past few decades, a solid theoretical and numerical foundation has been established for computational methods specifically tailored for the investigation of dynamic and adaptive systems in biology and medicine. These techniques are now at the verge of making it possible to represent and analyze large, organizationally complex systems and to study their emergent properties in a rigorous fashion. Methods of machine learning, numerical mathematics, and bioinformatics permit the efficient mining and analysis of the most useful data from within an overwhelming amount of data that are not pertinent for a given task. Algorithmic advances permit the simulation and optimization of very large biological flux distribution networks. Computer-aided approximation approaches yield ever-finer insights into the dynamics of complex nonlinear systems, such as the control of blood flow in healthy and diseased hearts. New mathematical, physical, and computational methods are beginning to make it possible to

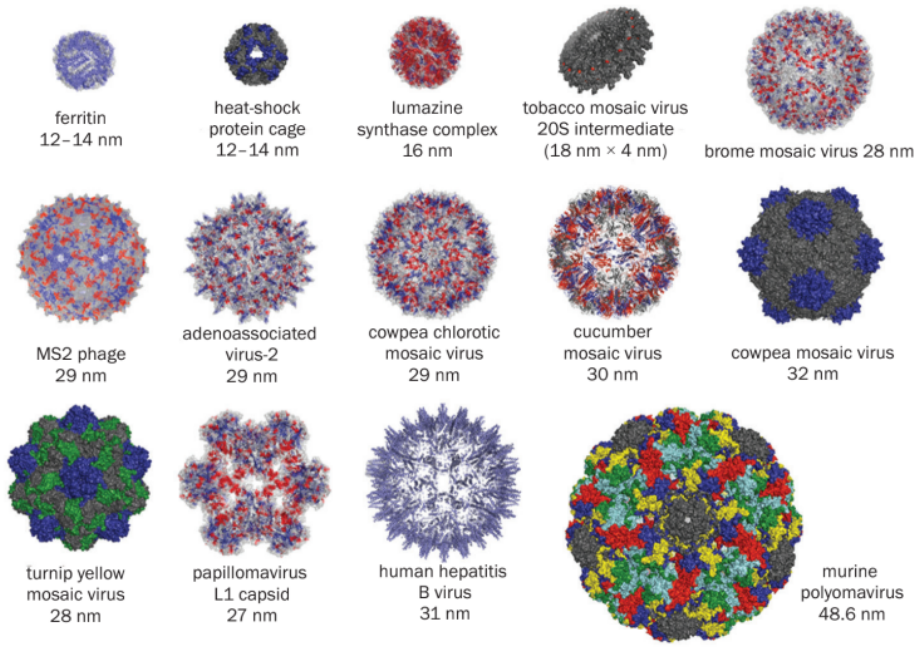


Figure 1.13 “Protein cages” are particles that have applications in bionanotechnology and nanomedicine. These particles are very interesting biological building blocks because they self-assemble into a variety of different shapes. The features of these bionanoparticles can be genetically manipulated and fine-tuned for biomedical purposes, such as drug delivery, gene therapy, tumor imaging, and vaccine development. (From Lee LA & Wang Q. *Nanomedicine 2* [2006] 137–149. With permission from Elsevier.)



Figure 1.14 Advances in computer power, accessibility, and user-friendliness over the past 40 years have been tremendous. Not too long ago, computer code had to be fed manually into the computer with punch cards. (Courtesy of Mutatis mutandis under the Creative Commons Attribution-Share Alike 3.0 Unported license.)

predict the folding of proteins and the binding between target sites and ligands. These predictions, in turn, suggest insights into specific molecular interactions and promise the potential of targeted drug interventions that minimize toxic side effects.

Motive and opportunity have met to make systems biology attractive and feasible. It has become evident that the relevant disciplines complement each other in unique ways and that the synergism among them will revolutionize biology, medicine, and a host of other fields, including biotechnology, environmental science, food production, and drug development.

COMMUNICATING SYSTEMS BIOLOGY

It is not a trivial task to talk succinctly about 25,000 genes and their expression state or about the many processes occurring simultaneously in response to a signal that a cell receives at its outer surface. Our minds are ill equipped to characterize numerical relationships, let alone discuss complicated mathematical functions, especially if these depend on many variables. If we did not have numbers, we would even have problems describing everyday features such as temperature. Of course, we can say that it is cold or hot, and we have a dozen or so adjectives in between. But if we need more accuracy, common language ceases to be sufficient. Discerning 37.0°C from 38.4°C is not easy without a numerical scale, but it is necessary to have the tools to

describe the difference, for instance, because the former reflects normal body temperature, while the latter is a sign of fever.

We may willingly or grudgingly accept the fact that we need mathematics, which comes with its own terminology, but communication is a two-way process. If we start talking about eigenvalues and Hopf bifurcations, we are almost guaranteed to lose mainstream biologists, let alone laypeople. This is a real problem, because our results must be conveyed to biologists, who are providing us data, and to the public that pays for our research and has a right to reap the fruit of the enormous investment of resources going into science [25]. The only true solution to this challenge is the bilingual education and nurturing of systems biologists who can translate biological phenomena into math and computer code and who can explain what it really means for the biological system if the real part of an eigenvalue is positive [19].

Communication is not trivial even within biology itself, because specialization has progressed so far that different fields such as molecular biology, immunology, and nanomedicine have developed their own terminology and jargon. Let's look at this issue in the form of a parable from Indian folklore that describes six blind men exploring an elephant (Figure 1.15). This story is quite old and usually ends in utter confusion, but it is useful to analyze it further than is usually done. The story has it that each of the blind men touched a different part of an elephant and came to a different conclusion concerning the object of his research. The man touching the side thought he was touching a wall, the one feeling the leg concluded he was touching a tree. The elephant's trunk gave the impression of a snake, the tusk that of a pointed scimitar, the tail felt like a rope, and the ear appeared to be like a large leaf or fan. It is not difficult to see the analogy to a complex biological system like the onset of Alzheimer's disease. The first scientist found "the Alzheimer gene," the second discovered "a strong association between the disease and former head injuries," another scientist detected "problems with fatty acid metabolism in the brain," and yet another suggested that "aluminum in cookware might be the culprit." As in the case of the elephant, the scientists were right, to some degree.

Let's analyze the elephant story a little further. The first problem among the six blind men might have been the homogeneous pool of researchers. Including a female or a child might have provided additional clues. Also, we have to feel sorry for the Indian men for being blind. However, they were apparently not mute or deaf, so that a little discussion among them might have gone a long way. While all six were blind, it is furthermore fair to assume that they had friends with working vision, who could have set them straight. They could have used not just their hands but also their other senses,

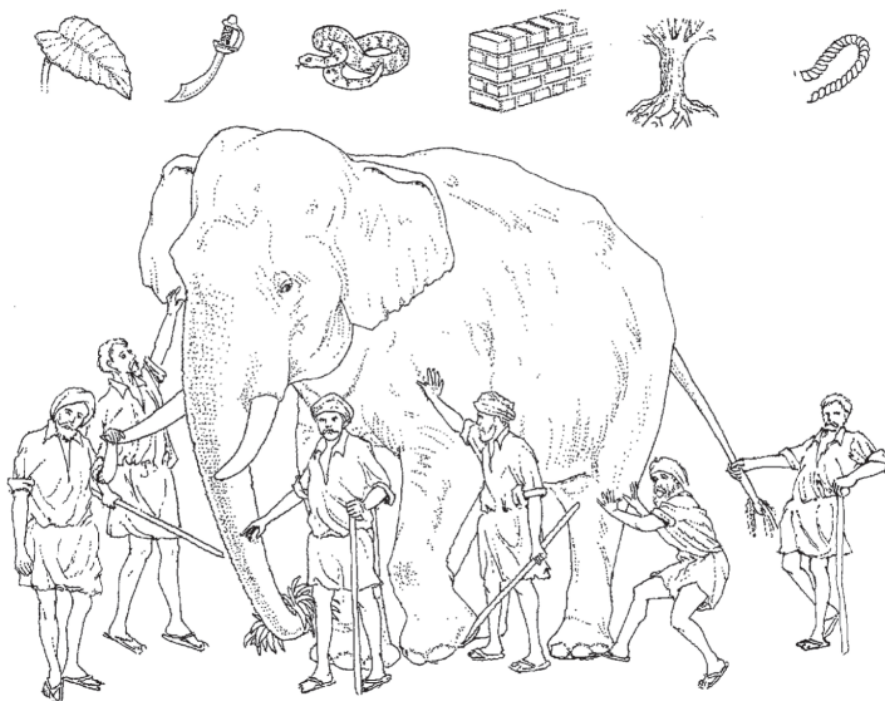


Figure 1.15 Information about isolated parts of a system alone does not always reveal the true nature of the system. An old story of six blind Indian men trying to determine what they touch is a parable for the dangers of scientific silos and the lack of good communication.

such as smell. Do tree trunks really smell like elephant feet? Finally, they apparently stayed in their one spot, thereby greatly limiting their experience base.

It is again easy to translate these issues into biology, especially when we think of purely reductionist strategies. Instead of a homogeneous pool of biologists analyzing biological systems, it is without doubt more effective to have a multidisciplinary team including different varieties of biologists, but also physicists, engineers, mathematicians, chemists, and smart people trained in the liberal arts or economics. Instead of only focusing on the one aspect right in front of our nose, communication with others provides context for singular findings. We don't know whether the Indian men spoke the same language, but we know that even if biologists, computer scientists, and physicists all use English to communicate, their technical languages and their views of the scientific world are often very different, so that communication may initially be superficial and ineffective. That is where multidisciplinary groups must engage in learning new terminologies and languages and include interdisciplinary translators. Just as the Indian men should have called upon their seeing friends, investigators need to call in experts who master techniques that have not been applied to the biological problem at hand. Finally, it behooves the trunk analyzer to take a few steps and touch the tusk and the side. Established scientific disciplines have in the past often become silos. Sometimes without even knowing it, researchers have kept themselves inside these silos, unable or unwilling to break out and to see the many other silos around, as well as a whole lot of space between them.

Systems biology does not ask the six blind men to abandon their methods and instead to run circles around the elephant. By focusing on one aspect, the reductionist "elephantologists" are poised to become true experts on their one chosen body part and to know everything there is to know about it. Without these experts, systems biology would have no data to work on. Instead, what systems biology suggests is a widening of the mindset and at least rudimentary knowledge of a second language, such as math. It also suggests the addition of other researchers, assisting the "trunkologist" and the "tuskologist" by developing new tools of analysis, by telling them in their language what others have found, by closing the gap between trunks and tusks and tails.

One strategy for accomplishing this synergism is to collect the diverse pieces of data and contextual information obtained by the six blind men and to merge them into a conceptual model. What kind of "thing" could consist of parts that feel like a snake, tree trunks, large walls, two scimitars, two fans, and a rope? How could an aberrant gene, former head injuries, and altered brain metabolism functionally interact to result in Alzheimer's disease? Well-trained systems biologists should be able to develop strategies for merging heterogeneous information into formal models that permit the generation of testable hypotheses, such as "tree-trunk-like things are connected by wall-like things." These hypotheses may be wrong, but they can nevertheless be very valuable, because they focus the scientific process on new, specific experiments that either confirm or refute the hypothesis. An experiment could be to walk along the "wall" as far as possible. Is there a "tree trunk" on the end? Are there "tree trunks" to the left and to the right? Is there a "pointed scimitar" at one or both ends? Is the "snake" connected to a "wall" or to a "tree trunk"? Does the "wall" reach the ground? Each answer to one of these questions constrains further and further what the unknown "thing" could possibly look like, and this is the reason that refuted hypotheses are often as valuable or even more valuable than confirmed hypotheses. "The wall does indeed not reach the ground!" Then, how is it supported? By "tree trunks"?

The story tells us that effective communication can solve a lot of complex questions. In systems biology, such communication is not always easy, and it requires not only mastering the terminology of several parent disciplines but also internalizing the mindset of biologists and clinicians on the one hand and of mathematicians, computer scientists, and engineers on the other. So, let's learn about biology. Let's study laboratory data and information and explore the mindset of biologists. Let's study graphs and networks with methods from computer science. Let's see how mathematicians approach a biological system, struggle with assumptions, make simplifications, and obtain solutions that are at first incomprehensible to the non-mathematician but do have real meaning once they are translated into the language of biology.

THE TASK BEFORE US

We have discussed the need to understand biological systems. But what does that really mean? Generically, it means that we should be able to explain how biological systems work and why they are constructed in the fashion as we observe them and not in a different one. Second, we should be able to make reliable predictions of responses of biological systems under yet-untested conditions. And third, we should be able to introduce targeted manipulations into biological systems that change their responses to our specifications.

This level of understanding is a tall order, and we will need many years to achieve it even for a narrowly focused domain within the huge realm of biological systems. An important component of the task is the conversion of actual biological systems into computational models, because this conversion, if it is valid, allows us almost unlimited and comparatively cheap analyses. The resulting models of biological systems come in two types. The first focuses on specific systems and includes all pertinent functional and numerical details—one might think of the analogy to a flight simulator. The second type of model is intended to help us understand the fundamental, generic features of the organization of biological systems; here one might think of elementary geometry, which offers us valuable insights into spatial features of the world by dealing with ideal triangles and circles that do not really exist in nature. The two model types point to two opposite ends of a large spectrum. The former models will be large and complex, while the latter will be as reduced and simple as feasible. In practice, many models will fall between these two extremes.

To pave the way toward these goals, this book is organized in three sections. The first of these introduces in four chapters a set of modeling tools for converting biological phenomena into mathematical and computational analog and for diagnosing, refining, and analyzing them. The second section describes in five chapters the molecular inventories that populate biological systems, and the five chapters in the third section are devoted to representative case studies and a look into the future.

The modeling approaches parallel two fundamental properties of biological systems, namely their static structure and their dynamics, that is, their changes over time. For static analyses, we will characterize and rationalize how nature put particular systems together and which parts are directly or loosely connected with each other. We will see that there are distinct types of connections and interactions. One important difference is that some connections allow the flow of material from a source to a target, while others serve the sole purpose of signaling the state of the system. In the latter case, no material changes location. Like a billboard that is not changed whether hundreds of people look at it or nobody at all, a signaling component may not be changed when it sends a signal. It is also to be expected that some connections are crucial, while others are of secondary importance. Finally, there is the very challenging question of how we can even determine the structure of a system. What types of data do we need to infer the structure of a system, and how reliable is such an inference?

The dynamics of a system is of the utmost importance, because all biological systems change over time. Organisms traverse a life cycle during which they undergo tremendous changes. Even a lowly yeast cell fills up with scars where it has given birth to daughter cells, and once its surface is full, the cell slides into the sunset of life. We can easily see these changes under a microscope, but an incomparably larger number of changes remain hidden from our sight. Gene expression patterns, amounts of proteins, profiles of metabolites, all of these change dramatically between birth and death. In addition to normal changes throughout its lifetime, every organism responds to fast and slow changes in the environment and adapts rather quickly to new situations. Today we may observe and characterize the gene expression network in a bacterium, but tomorrow it may already have changed in response to some environmental pressures. Indeed, bacteria evolve so quickly that the commonly used term “wild type” no longer has much meaning. Even more than static aspects, dynamic aspects of biological systems mandate the use of computational models. These models help us reveal how fascinating living systems are, with respect both to their overall efficiency and to the ingenuity with which dynamic responses and adaptations are coordinated.

Whether static or dynamic, some model designs and analyses will be performed in the bottom-up and others in the top-down direction. However, since we seldom

start at the very bottom, that is, with individual atoms, or at the very top with models of complete organisms as they interact with their environment, most modeling strategies in systems biology are in truth “middle-out,” to use Nobel Laureate Sydney Brenner’s expression (cited in [26]). They begin somewhere in between the extremes, maybe with pathways or with cells. Over time, they may be incorporated into larger models, or they may become more and more refined in detail.

The second section of the book addresses the molecular inventories of biological systems. Paralleling the biological organization of organisms, one chapter is devoted to gene systems, one to proteins, and one to metabolites. A further chapter discusses signal transduction systems, and the final chapter of this section describes features of populations. Clearly, all these chapters are woefully incomplete and should not be thought of as substitutes for real biology books. Their purpose is merely to provide brief overviews of the main classes of biological components that are the foundation of all modeling strategies.

The third section contains case studies that in one way or another highlight aspects of biological systems that are in some sense representative. One chapter describes the coordinated stress response system in yeast, which operates simultaneously at the gene, protein, and metabolite levels. Another chapter provides a presentation of how very different models can be useful to focus attention on selected aspects of a multiscale system, the heart. A [third chapter](#) indicates how systems biology can contribute to medicine and drug development. The fourth chapter illuminates aspects of the natural design of biological systems and of the artificial design of synthetic biological systems. Finally, the last chapter discusses emerging and future trends in systems biology.

EXERCISES

- 1.1. Search the Internet, as well as different dictionaries, for definitions of a “system.” Extract commonalities among these definitions and formulate your own definition.
- 1.2. Search the Internet for definitions of “systems biology.” Extract commonalities among these definitions and formulate your own definition.
- 1.3. List 10 systems within the human body.
- 1.4. Exactly what features make the system in [Figure 1.10](#) so much more complicated than the system in [Figure 1.8](#)?
- 1.5. Imagine that [Figure 1.10](#) represents a system that has become faulty owing to disease. Describe the consequences of its complexity for any medical treatment strategy.
- 1.6. In [Figure 1.2](#), are there control paths along which ABA either activates or inhibits closure of stomata? If so, list at least one path each. If both paths exist in parallel, discuss what you expect the effect of an increase in ABA to be on stomata closure?
- 1.7. Imagine a control system like that in [Figure 1.2](#), but much simpler. Specifically, suppose there is only one activating and one inhibiting path in parallel. Suppose further that the activating path reacts much faster than the inhibiting path. What would be the consequence with respect to the effect of an increase in input (ABA) on output (stomata closure)? How could a difference in speed be implemented in a natural cell? Does it matter how strong or weak the activation and inhibition are? Discuss!
- 1.8. We have discussed that it is often difficult to infer the structure of a biological system from data. Is it possible that two different systems produce exactly the same input–output data? If you think it is impossible, discuss and defend your conclusion. If you think the answer is affirmative, construct a conceptual example.
- 1.9. List and discuss features supporting the claim that reductionism alone is not sufficient for understanding biological systems.
- 1.10. List those challenges of systems biology that cannot be solved with intuition alone.
- 1.11. Discuss why it is important to create terminology and tools for communicating systems biology.
- 1.12. Assemble a “to-do” list for the future of systems biology.

REFERENCES

- [1] Li S, Assmann SM & Albert R. Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biol.* 4 (2006) e312.
- [2] USNO. United States Naval Observatory Cesium Fountain. <http://tycho.usno.navy.mil/clockdev/cesium.html> (2010).

- [3] Voit EO, Alvarez-Vasquez F & Hannun YA. Computational analysis of sphingolipid pathway systems. *Adv. Exp. Med. Biol.* 688 (2010) 264–275.
- [4] Vodovotz Y, Constantine G, Rubin J, et al. Mechanistic simulations of inflammation: current state and future prospects. *Math. Biosci.* 217 (2009) 1–10.
- [5] Savageau MA. *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology*. Addison-Wesley, 1976.
- [6] Kim HJ & Triplett BA. Cotton fiber growth *in planta* and *in vitro*. Models for plant cell elongation and cell wall biogenesis. *Plant Physiol.* 127 (2001) 1361–1366.
- [7] Kumar A. *Understanding Physiology*. Discovery Publishing House, 2009.
- [8] Wolkenhauer O, Kitano H & Cho K-H. An introduction to systems biology. *IEEE Control Syst. Mag.* 23 (2003) 38–48.
- [9] Westerhoff HV & Palsson BO. The evolution of molecular biology into systems biology. *Nat. Biotechnol.* 22 (2004) 1249–1252.
- [10] Strange K. The end of “naive reductionism”: rise of systems biology or renaissance of physiology? *Am. J. Physiol. Cell Physiol.* 288 (2005) C968–C974.
- [11] Voit EO & Schwacke JH. Understanding through modeling. In *Systems Biology: Principles, Methods, and Concepts* (AK Konopka ed.), pp 27–82. Taylor & Francis, 2007.
- [12] Bernard C. *Introduction à l'étude de la médecine expérimentale*. JB Baillière, 1865 (reprinted by Éditions Garnier-Flammarion, 1966).
- [13] Noble D. Claude Bernard, the first systems biologist, and the future of physiology. *Exp. Physiol.* 93 (2008) 16–26.
- [14] Bertalanffy L von. Der Organismus als physikalisches System betrachtet. *Naturwissenschaften* 33 (1940) 521–531.
- [15] Bertalanffy L von. *General System Theory: Foundations, Development, Applications*. G Braziller, 1969.
- [16] Mesarović MD. Systems theory and biology—view of a theoretician. In *Systems Theory and Biology* (MD Mesarović, ed.), pp 59–87. Springer, 1968.
- [17] Rosen R. A means toward a new holism: “Systems Theory and Biology. Proceedings of the 3rd Systems Symposium, Cleveland, Ohio, Oct. 1966. M. D. Mesarović, Ed. Springer-Verlag, New York, 1968. xii + 403 pp” [Book review]. *Science* 161 (1968) 34–35.
- [18] Yamamoto N, Nakahigashi K, Nakamichi T, et al. Update on the Keio collection of *Escherichia coli* single-gene deletion mutants. *Mol. Syst. Biol.* 5 (2009) 335.
- [19] Savageau MA. The challenge of reconstruction. *New Biol.* 3 (1991) 101–102.
- [20] Freitas RAJ. *Nanomedicine, Volume I: Basic Capabilities*. Landis Bioscience, 1999.
- [21] Lee LA & Wang Q. Adaptations of nanoscale viruses and other protein cages for medical applications. *Nanomedicine* 2 (2006) 137–149.
- [22] Martel S, Mathieu J-B, Felfoul O, et al. Automatic navigation of an untethered device in the artery of a living animal using a conventional clinical magnetic resonance imaging system. *Appl. Phys. Lett.* 90 (2007) 114105.
- [23] SBML. The systems biology markup language. http://sbml.org/Main_Page (2011).
- [24] Stanislaus R, Jiang LH, Swartz M, et al. An XML standard for the dissemination of annotated 2D gel electrophoresis data complemented with mass spectrometry results. *BMC Bioinformatics* 5 (2004) 9.
- [25] Voit EO. *The Inner Workings of Life*. Cambridge University Press, 2016.
- [26] Noble D. *The Music of Life: Biology Beyond Genes*. Oxford University Press, 2006.

FURTHER READING

- Alon U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall/CRC, 2006.
- Kitano H (ed.). *Foundations of Systems Biology*. MIT Press, 2001.
- Klipp E, Herwig R, Kowald A, et al. *Systems Biology in Practice: Concepts, Implementation and Application*. Wiley-VCH, 2005.
- Noble D. *The Music of Life: Biology Beyond Genes*. Oxford University Press, 2006.
- Szallasi Z, Periwál V & Stelling J (eds). *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. MIT Press, 2006.
- Voit EO. *The Inner Workings of Life*. Cambridge University Press, 2016.
- Voit EO. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*. Cambridge University Press, 2000.

Introduction to Mathematical Modeling

2

When you have read this chapter, you should be able to:

- Understand the challenges of mathematical modeling
- Describe the modeling process in generic terms
- Know some of the important types of mathematical models in systems biology
- Identify the ingredients needed to design a model
- Set up simple models of different types
- Perform basic diagnostics and exploratory simulations
- Implement changes in parameters and in the model structure
- Use a model for the exploration and manipulation of scenarios

At the core of any computational analysis in systems biology is a mathematical **model**. A model is an artificial construct in the language of mathematics that represents a process or phenomenon in biology. **Modeling** is the process of creating such a construct and squeezing new insights out of it. Mathematical models in systems biology can take many forms. Some are small and simple, others large and complicated. A model can be intuitive or very abstract. It may be mathematically elegant in its streamlined simplicity or account for very many details with a large number of equations and specifications that require sophisticated computer **simulations** for their analysis. The key commonality among all good models in biology is their ability to offer us insights into processes or systems that we would not be able to gain otherwise. Good models can make sense of a large number of isolated facts and observations. They explain why natural systems have evolved into the particular organizational and regulatory structures that we observe. They allow us to make **predictions** and **extrapolations** about experimentally untested situations and future trends. They can lead to the formulation of new hypotheses. Good models have the potential to guide new experiments and suggest specific recipes for manipulating biological systems to our advantage. Models can tell us how cell cycles are controlled and why plaques appear in certain sections of our arteries and not in others. Future models may help us understand how the brain works and prescribe specific interventions in tumor formation that will be infinitely more subtle and less crude than our current methods of radio- or chemotherapy.

Modeling in systems biology is really a special way of looking at the world. Whether we just conceptualize a situation in the form of components and interactions, or whether we actually set up equations and analyze them, we begin to see how complicated phenomena can be conceptually simplified and dissected into manageable submodels or **modules**, and we develop a feel for how such modules interact with each other and how they might respond to changes. Be aware! Modeling may become a life-changing adventure.

An analogy for a mathematical model is a geographical map. This map is obviously very different from reality, and many details are missing. What do the houses look like in some remote village? Are the lakes clean enough for swimming? Are the people friendly? Then again, it is exactly this simplification of the complicated real world that makes the map useful. Without being distracted by secondary details, the map allows us to look at the network of major roads throughout an entire country while sitting at home on our living room floor. The map provides us with distances and highway types that let us estimate travel times. Altitudes on the map indicate whether the area we want to drive through is hilly and could be cold even during the summer months. Once we learn how to interpret the map, we can infer what the countryside might look like. Empty spaces, like the Badlands in South Dakota, suggest that living in the area might be difficult or undesirable for humans. The density of cities and villages provides us with some idea of how easy it might be to find lodging, food, or gas. Clusters of cities along the ocean probably imply that living there is desirable. Mathematical models have the same generic properties: they reflect some aspects quite well, but barely touch on others or ignore them altogether. They allow us to infer certain properties of reality but not others. The inclusion or exclusion of certain features in models has a direct, important implication for the modeling process. Namely, the key task in designing a suitable model is a crisp focus on the aspects of primary interest and a representation of these aspects that retains their most important properties. At the same time, extraneous or distracting information must be minimized, and the model must be limited in complexity to permit analysis with reasonable efficiency.

Modeling is a mathematical and computational endeavor that in some sense resembles the fine arts ([Figure 2.1](#)). Basic math problems in calculus or linear algebra have the correct solutions in the back of the book, or at least somebody,

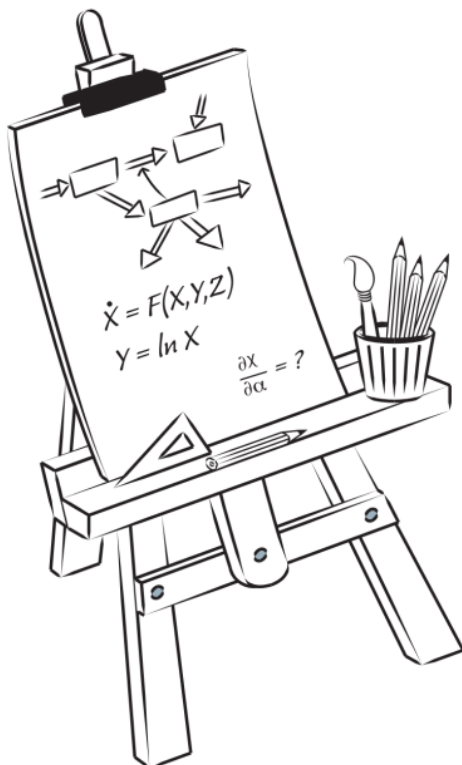


Figure 2.1 Modeling resembles a fine art. Modeling requires both mastery of techniques and creativity.

somewhere, knows these unique, correct solutions. In contrast to this rigor and crispness, many aspects of modeling allow a lot more flexibility. They permit different approaches, and often there are multiple, quite distinct solutions that are all valuable and correct in their own right. Who is to say that an oil painting is better than a watercolor? Or that an opera has higher artistic value than a symphony? Similarly, a mathematical model has aspects and properties in different dimensions and its value is not always easy to judge. For instance, is a very complicated, detailed model better than a simple, yet elegant model?

The answers to these questions depend critically on the goals and purposes of the model, and it is not just beneficial but indeed mandatory to specify the reasons for setting up a model and to ask a lot of questions before we embark on actually designing it. These questions must clarify the potential utilization of the model, its desired accuracy, the degree of complexity we are willing to accept, and many other details that we will touch on in this chapter. Indeed, it is prudent to invest quite a bit of time and exploration in pondering these aspects before one begins selecting a specific model structure and defining variables, parameters, and other model features, because technical challenges often distract from the big picture.

As in the fine arts, the creation of a good model requires two ingredients: technical expertise and creativity. The technical aspects can and must be learned through the acquisition of a solid repertoire of math and computer science skills. To make theoretical and numerical analyses meaningful, one also has to dig sufficiently deeply into biology in order to understand the background and terminology of the phenomenon of interest and to interpret the model results in a correct and sensible manner. As with the fine arts, the creative aspect of the process is more difficult to learn and teach. For instance, it requires more experience than one might expect to ask the right questions and reformulate them in such a way that they become amenable to mathematical and computational analysis. For a novice, all data may look equally valuable, but they really are not. In fact, it takes time and experience to develop a feel for what is doable with a particular set of data and what is probably out of reach. It is also important to learn that some data are simply not well suited for mathematical modeling. This insight is sometimes disappointing, but it can save a lot of effort and frustration.

The shaping and refining of specific questions mandates that we learn how to match the biological problem at hand with appropriate mathematical and computational techniques. This matching requires knowledge of what techniques are available, and it also means that the modeler must make an effort to understand how biologists, mathematicians, and computer scientists are trained to think. The differences in approach are quite pronounced in these three disciplines. Mathematicians are encouraged to simplify and abstract, and there is hardly anything more appealing than a tough problem that can be scribbled on the back of a napkin. Computer scientists look at problems algorithmically, that is, by dissecting them into very many, very tiny steps. Contrast that with biologists, who have learned to marvel at the complexity and intricacies of even the smallest biological item, and you can imagine the mental and conceptual tensions that may develop during the modeling process.

The formation of good questions requires decisions about what is really important in the model, which features can be ignored, and what inaccuracies we are willing to tolerate when we simplify or omit details. It is this complex of decisions on the inclusion and exclusion of components, facts, and processes that constitutes the artistic aspect of modeling. As in the fine arts, a good model may give an outsider the impression "Oh sure, I could have done that," because it is well designed and does not reveal the hard work and failed attempts that went into it. A good degree of proficiency with respect to these decision processes can be learned, but, like everything else, it requires practice with simple and then increasingly complex modeling tasks. This learning by doing can be exhilarating in success and depressing in failure, and while we have every right to enjoy our successes, it is often the failures that in the end are more valuable, because they point to aspects we do not truly understand.

The computational side of modeling draws from two classes of methods: mathematical analysis and computer simulation. Most practical applications use a combination of the two, because both have their own strengths. A purely mathematical analysis leads to general solutions that are always true, if the given assumptions and prerequisites are satisfied. This is different from even a large set of computer

simulations, which can never confer the same degree of certainty and generality. For instance, we can mathematically prove that every real number has a real inverse (4 has 0.25, -0.1 has -10), with the notable exception of 0. But if we were to execute a mega-simulation on a computer, randomly picking a real number and checking whether its inverse is real, we would come to the conclusion that indeed all real numbers have a real inverse. We would miss zero, because the probability of randomly picking zero from an infinite range of numbers is nil. Thus, in principle, mathematics should always be preferred. However, in practical terms, the mathematics needed to analyze biological systems becomes very complicated very quickly, and even relatively simple-looking tasks may not have an explicit analytical solution and therefore require computational analysis.

Independent of the specific biological subject area and the ultimate choice of a particular mathematical and computational framework, the generic modeling process consists of several phases. Each phase is quite distinct both in input requirements and in techniques. Broadly categorized, the phases are shown in **Figure 2.2**, together with a coarse flow chart of the modeling process. Further details are presented in **Table 2.1**. The process starts with conceptual questions about the biological problem, becomes more and more technical, and gradually returns to the biological phenomenon under study. It is very important to devote sufficient effort to each of these phases, especially the early ones, because they ultimately determine whether the modeling effort has a good chance of succeeding.

One of the early choices addresses the explanatory character of the model. One may choose a comparatively simple regression model, which correlates two or more quantities to each other, without, however, suggesting a rationale for the correlation. As an example, one may find that cardiovascular disease is often associated with high blood pressure, but a regression model does not offer an explanation. As an alternative, one could develop a much more complex mechanistic model of cardiovascular disease, in which blood pressure is a component. If formulated validly, the model would explain the mechanistic connection, that is, the network of causes and effects leading from high blood pressure to the disease. Even if we disregard the difficulties in formulating a comprehensive mechanistic model, the choice of model type is not black and white. Regression models do not provide an explanation, but they often make correct, reliable predictions. An explanatory model may also be quite simple, but if it accounts for a large number of processes, it often becomes so

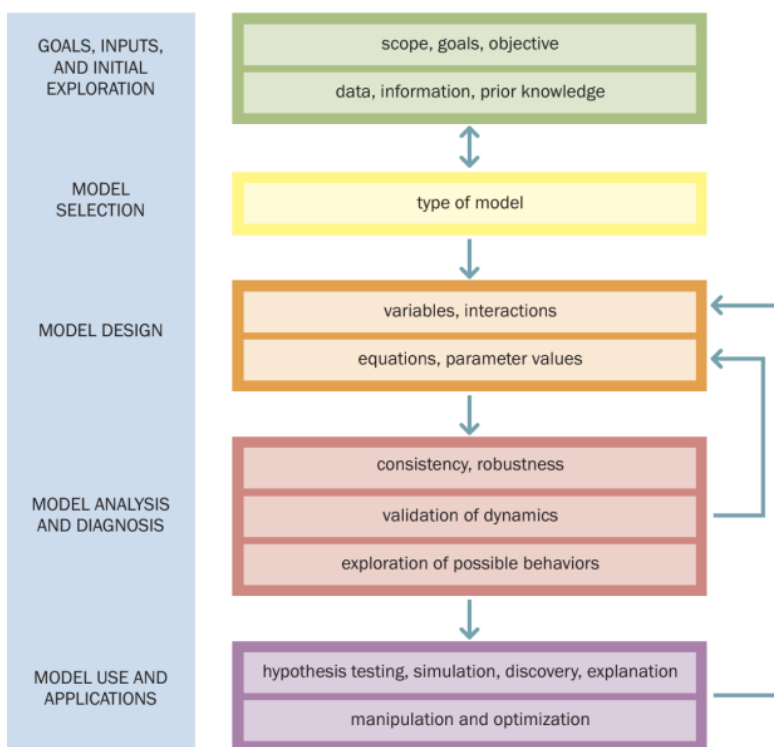


Figure 2.2 Flow chart of the modeling process. The modeling process consists of distinct phases, but is often iterative and involves refinements in model design and repeated diagnosis. See also **Table 2.1**. (Adapted from Voit EO, Qi Z & Miller GW. *Pharmacopsychiatry* 41(Suppl 1) [2008] S78–S84. With permission from Thieme Medical Publishers.)

TABLE 2.1: ISSUES TO PONDER DURING THE MODELING PROCESS

<p>Goals, Inputs, and Initial Exploration</p> <ul style="list-style-type: none"> • Scope of the model, including goals, objectives, and possible applications • Data needs and availability (types, quantity, quality) • Other available information (nonquantitative heuristics, qualitative input from experts) • Expected feasibility of the model • Relevance and degree of interest within the scientific community
<p>Model Selection</p> <ul style="list-style-type: none"> • Model structure <ul style="list-style-type: none"> ◦ Explanatory (mechanistic) versus correlative (black box) ◦ Static versus dynamic ◦ Continuous versus discrete ◦ Deterministic versus stochastic ◦ Spatially distributed versus spatially homogeneous ◦ Open versus closed ◦ Most appropriate, feasible approximations
<p>Model Design</p> <ul style="list-style-type: none"> • Model diagram and lists of components <ul style="list-style-type: none"> ◦ Dependent variables ◦ Independent variables ◦ Processes and interactions ◦ Signals and process modulations ◦ Parameters • Design of symbolic equations • Parameter estimation <ul style="list-style-type: none"> ◦ Bottom-up ◦ Top-down
<p>Model Analysis and Diagnosis</p> <ul style="list-style-type: none"> • Internal consistency (for example, conservation of mass) • External consistency (for example, closeness to data and expert opinion) • Reasonableness of steady state(s) • Stability analysis • Sensitivity and robustness analysis • Structural and numerical boundaries and limitations • Range of behaviors that can or cannot be modeled for (example, oscillations, multiple steady states)
<p>Model Use and Applications</p> <ul style="list-style-type: none"> • Confirmation, validation of hypotheses • Explanation of causes and effects; causes of failure, counterintuitive responses • Best case, worst case, most likely scenarios • Manipulation and optimization (for example, treatment of disease; avoidance of undesired states or dynamics; yield optimization in biotechnology) • Discovery of design principles

complicated and requires so many assumptions that it may no longer describe or predict new scenarios with high reliability. In reality, most models are intermediates—explanatory in some aspects and correlative in others.

In the following sections, we discuss each phase of the modeling process and illustrate it with a variety of small, didactic “sandbox models” as well as a famous, very old model describing the spread of infectious diseases. The process begins with an exploration of the available information and the selection of a model that can utilize this information and ultimately yield new insights (see [Figure 2.2](#) and

Table 2.1). Often, one will go back and forth between these two aspects before a suitable model design is chosen. Once an appropriate model type is determined, the actual model design phase begins. Components of the model are identified and characterized. The resulting model is diagnosed, and, should troublesome aspects emerge, one returns to the model design. A model that fares well in the diagnostic phase is used for explanations, simulations, and manipulations. It often turns out that model features are not optimal after all, requiring a return to the model design phase.

GOALS, INPUTS, AND INITIAL EXPLORATION

The initial phase of model development consists of defining the purpose of the model and of surveying and screening input data and contextual information. This setting of goals, combined with an exploration of available information and the context for the model, is arguably the most crucial of all phases. This assessment may sound counterintuitive, because one might think that the most effort would go into some complicated math. However, if insufficient consideration is given to this initial exploration, the modeling process may derail right away, and one may not even notice that something is not quite the way it was intended until the final phases of the modeling process.

The main task of the input and exploration phase is to establish the specific purposes and goals of the model, and to feel out whether the model may have a chance to achieve its goals, given the available data and information. “Feeling out” does not sound like rigorous mathematical terminology, but actually describes the task quite well. It is often here where experience is needed the most: right at the beginning of the effort!

2.1 Questions of Scale

The first important aspect of the exploration phase is to ask what are the most appropriate scales for the model, with respect to both time and space, and also with respect to the organizational level (Figure 2.3). These scales are typically tied to each other, because processes at the organizational level of a cell occur on a small spatial scale and usually run faster than processes at the scales of organisms or ecological systems. Suppose we are interested in trees and want to develop a mathematical model describing their **dynamics**. While that might sound like a fine idea, it is far too unspecific, because trees may be studied in many different ways. To model the growth of a forest or a tree plantation, the scale of organization is typically the individual tree, the spatial scale is in hectares or square kilometers, and the corresponding timescale is most likely in years. Contrast this with the photosynthesis in a leaf or needle, which truly drives tree growth. That organizational

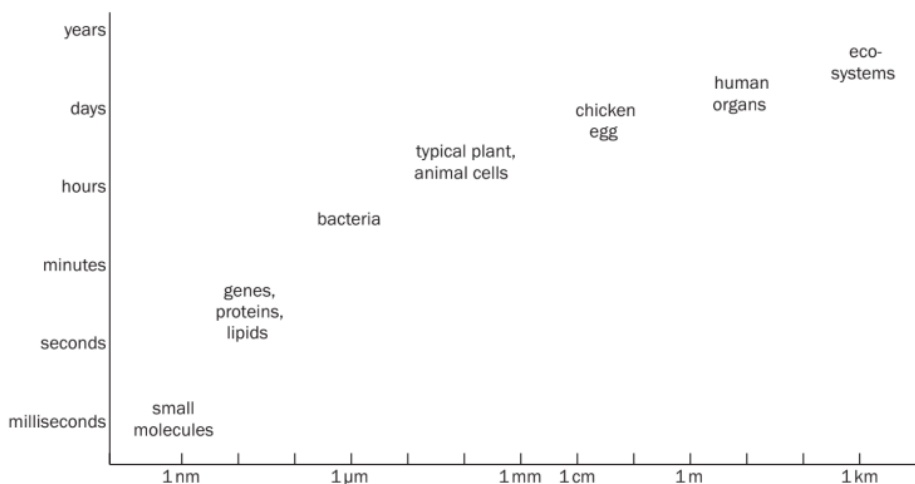


Figure 2.3 Scales in biology span several orders of magnitude in size and time. This plot shows typical sizes and time ranges. Of course, human organs and ecosystems contain small molecules and genes, so that they really span multiple size- and time-scales, which makes multiscale modeling challenging and fascinating.

level is biochemical or physiological, and the corresponding timescale is that of seconds or minutes. It is easy to imagine other time and organizational scales between or outside these two scenarios.

While multiscale modeling is one of the challenging frontiers of systems biology (see [Chapter 15](#)), the attempt to accommodate too many different scales at once is very difficult and, at least at the beginning of model design, dangerous. Although there are exceptions (see for instance [Chapter 12](#)), it is usually not really feasible to account in the same model for the entire spectrum of processes, from the details of photosynthesis or wood formation to the long-term growth of forests. Nonetheless, the real strength of mathematical models lies in their ability to explain phenomena at one level through the analysis and integration of processes at a lower level. Thus, a good model, even of small size, may span two or maybe three levels of biological organization. For instance, it may describe the formation of a leaf in terms of genetic and biochemical processes. A model may also skip some levels. For instance, we could try to associate long-term tree growth and final size with alterations in gene expression. In this case, one might be able to translate the effects of gene modulations into the tree's growth rate, but a model of this type would not capture all the biochemical and physiological causes and effects that occur between gene expression and the total amount of wood in the forest after 50 years.

The determination of suitable scales depends very significantly on the available data that are needed to construct, support, and validate the model. If only yearly biomass measurements of the tree stand are available, we should not even attempt to model the biochemistry that deep down controls tree growth. Conversely, if the model uses light conductance as main input data, good as they may be, this input information would hardly support a model of the growth of trees in the forest over many years.

As a different example, suppose it is our task to improve the yield of a fermentation process in which yeast cells produce alcohol. Typical experimental techniques for this task are selections of mutants and changes in the growth medium. Thus, if we intend to use a model to guide and optimize this process, the model must clearly account for gene activities and their biochemical consequences, as well as for concentrations of nutrients in the medium and their uptake by cells.

2.2 Data Availability

The quantity of available data provides important clues for exploring options of model selection and design. If a complicated model with many **variables** and parameters is supported only by very scarce data, the model may capture these data very nicely, but as soon as we try to apply it to a slightly different scenario, such as wider spacing of the trees or a richer fertilization regimen, the model predictions may become unreliable or inaccurate. This over-fitting problem (that is, using a model that is too complicated for the given data) always looms when only one or a few datasets are available. A similar problem arises if the data show a lot of **variability**, or noise, as it is often called in the field of data analysis. In this case, many models may yield a good representation of the particular data, but predictions regarding new data may be unreliable. Counterintuitive as it may sound, it is often easier to model bad data than good data, because bad data grant us a lot more latitude and potentially let us get away with an inferior model. Of course, this inferior model may turn round and bite us later when the model is expected to fit new data and simply doesn't!

While crisp, numerical data are the gold standard for mathematical modeling, we should not scoff at qualitative data or information that contains a fair amount of uncertainty. For instance, if we are told that experiments show some output increasing over time (even though the exact degree is not given), this piece of information can provide a valuable constraint that the model must satisfy. Especially in a clinical setting, this type of qualitative and semiquantitative information is prevalent, and the more of it that is available, the more any model of the phenomenon is constrained. Expressed differently, this information can be extremely helpful for weeding out models that could otherwise appear plausible. In a study of Parkinson's disease, we asked clinicians for concentrations of dopamine and other brain metabolites associated with the disease. While they could not give us exact values, they

provided us with rough, relative estimates of concentrations and metabolic fluxes through the system, and this information, combined with literature data, allowed us to set up a model that made surprisingly accurate predictions [1].

Finally, the exploration phase should ascertain whether experimental biologists or clinicians are really interested in the project. If so, their interest will help drive the project forward, and their direct insights and contacts with other experts can be invaluable, because much biological knowledge is not published, especially if it is semiquantitative or not yet one hundred percent supported by data and controls. It is, of course, possible to base models solely on information found in the literature, but if no experts are at hand, many pitfalls loom.

MODEL SELECTION AND DESIGN

To illustrate the process of model design, let's consider one specific example that we will carry through this chapter and revisit again in [Chapter 4](#). This example concerns an infectious disease, such as tuberculosis or influenza, that easily spreads throughout a human population ([Figure 2.4](#)). Down the road, we would like to determine whether it is beneficial to vaccinate all or some individuals, whether infected persons should be quarantined, or whether the disease will eventually go away by itself because people recover and acquire immunity. Immediately, many questions come to mind. Do people die from the disease? If the death rate is small, can we ignore it in the model? How does the disease get started in the first place? For how long are people sick? How infectious are they? Are there individuals who show no symptoms yet are infectious? Do all people in our target population have potential contact with all other people? Do we need to account for air travel? Should we attempt to characterize waves of infection that start at the point of the first contact with an infected person and move through certain geographical areas or the entire population? Does our disease model have to be accurate enough to predict whether a specific person will get sick, or is our target the average infection level within a community or the nation? Clearly, questions are unlimited in number and type, and some of them may be important to ponder now while others are probably less relevant, at least at the beginning.

For our purposes of studying a very complex phenomenon with as simple a model as possible, let's suppose the following:

- The infectious disease is spread by human-to-human contact.
- Although bacteria or viruses are the true vectors of the disease, we ignore them.
- Individuals may acquire and lose immunity.
- The primary questions are how fast the disease progresses, whether it disappears without intervention due to immunity, and whether vaccination or quarantine would be effective.

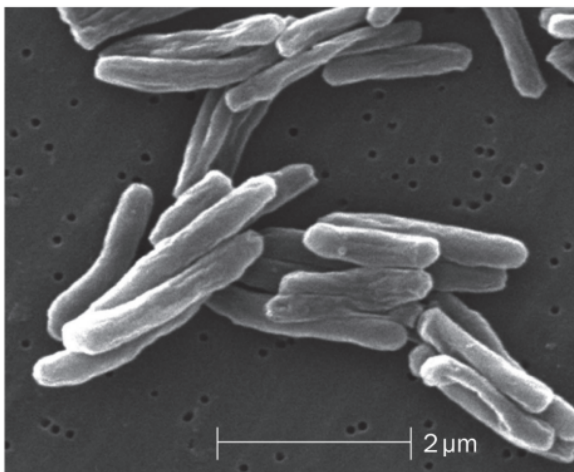


Figure 2.4 Mathematical modeling can be a powerful tool for studying the spread of bacterial diseases. Shown here is the bacterium *Mycobacterium tuberculosis*, which causes tuberculosis, one of the ongoing scourges of humankind.

- We do not intend to account for the spatial distribution of cases of infection throughout some geographical area.

2.3 Model Structure

Once we have decided what the model is supposed to accomplish, we need to determine its structure. Some decision points are given in Table 2.1. None of these are really associated with different degrees of sophistication, usefulness, or quality of the model. We are simply required to make decisions, again based on the data and the purposes of the model.

The first decision targets the question of how much detail the model is expected to explain. As mentioned before, the two extremes in this dimension are **correlative** and **explanatory models**; most models have aspects of both. A correlative model simply and directly relates one quantity to another. For instance, data may show that the infectious disease is particularly serious among the elderly; specifically, the number of infected individuals per 100,000 may increase with age (Figure 2.5). A **linear** or **nonlinear** regression could even quantify this relationship and allow us to make predictions about the average disease risk for a person of a certain age. However, even if the prediction were quite reliable, the model would not give us a hint as to why older people are more likely to become infected. The model does not account for health status, genetic predisposition, or lifestyle. Sometimes a prediction would be wrong, but the model would not be able to explain why.

An explanatory model, by contrast, relates an observation or outcome of an experiment to biological processes and mechanisms that drive the phenomenon under investigation. To illustrate, imagine an explanatory model of cancer. Cancer is associated with cells that are not dying when they should. Cell death normally occurs through a process called apoptosis, whose control involves the tumor suppressor protein p53, which in turn acts as a transcription factor for a number of pertinent genes. The p53 protein becomes activated under many stress conditions, such as ultraviolet radiation or oxidative stress. Pulling the pieces of information together allows us to compose a conceptual or mathematical model in the form of a chain of causes and effects: Suzie gets baked on the beach. She is exposed to enough ultraviolet radiation to activate the p53 proteins in her skin. The proteins change the normal activation of some genes. The genes cause some cells to escape apoptosis and proliferate out of control. Suzie is diagnosed with melanoma. The model explains to some degree (although not completely, of course) the development of cancer. This explanatory potential does not come for free: explanatory, mechanistic models are usually much more complicated than correlative models and require vastly more input in the form of data and assumptions. While explanatory models offer more insight, we should not discard correlative models offhand. They are much easier to construct and often make more reliable predictions than explanatory models, especially when assumptions are uncertain and data are scarce.

The second decision refers to the question of whether something associated with the phenomenon of interest changes over time. If the data show trends changing with time, we will probably prefer a dynamical model, which allows us to quantify these changes. By contrast, in a static model, such as a regression model, only the dependencies of one variable on others are explored. Box 2.1 contrasts static

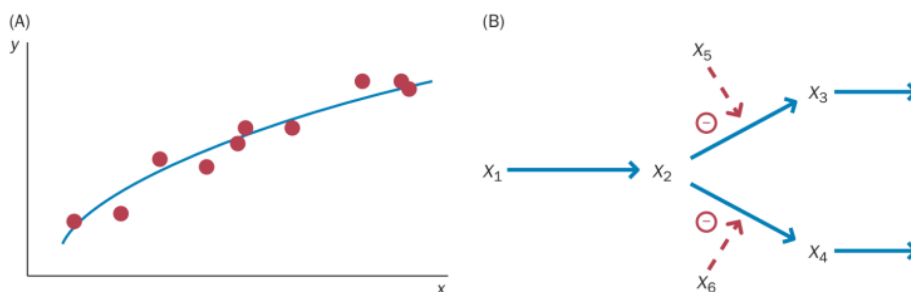


Figure 2.5 Correlative and explanatory models serve different purposes.

The correlation between x and y in (A), whether linear or nonlinear, permits robust predictions of a value of y , given a new value of x , although the model does not explain the connection between x and y . An example could be disease prevalence (y) and age (x) among the elderly. The simple explanatory model in (B) could describe a metabolic pathway with two inhibitors (X_5 and X_6). It not only captures the (positive) correlation between X_3 and X_6 , but also provides an explanation for why X_3 increases if the inhibitor X_6 of the other branch of the pathway is increased.

BOX 2.1: STATIC AND DYNAMIC MODELS

Suppose it is our task to compute the volume of an *Escherichia coli* cell. Studying a photograph of an *E. coli* culture (Figure 1), we see that each bacterium looks more or less like a hot dog with variable length and flattened ends, and since the volume doesn't change if the bacterium curves up, we can just study the shape when it is stretched out. Thus, a simple static model could be a cylinder with two "caps" at the end. The caps look a little flatter than hemispheres and we decide to represent them mathematically by halved ellipsoids that have the same axis radius in two directions (a circular cross section) and a shorter axis radius in the third direction (Figure 2). The formula for the volume of a whole ellipsoid is $\frac{4}{3}\pi r_1 r_2 r_3$, where r_1, r_2, r_3 are the axis radii. Thus, if the "straight" part of the cell is x μm long and has a diameter of y μm , and if each cap has a height of z μm , we can formulate a static model for the model of an *E. coli* cell (Figure 3) as

$$V(x, y, z) = \underbrace{x\pi\left(\frac{y}{2}\right)^2}_{\text{cylinder}} + 2\underbrace{\left(\frac{4}{3}\pi\frac{y}{2}\frac{y}{2}\frac{z}{2}\right)}_{\text{two half-ellipsoidal caps}} \quad (1)$$

$$= \left(\frac{x}{4} + \frac{z}{3}\right)\pi y^2. \quad (2)$$

Our static model allows us to pick any length, width, and cap height and to compute the corresponding volume of the bacterium.

Let's now suppose the bacterium has just divided and is growing. The size parameters x , y , and z become variables that depend on time t , and we may write them as $x(t)$, $y(t)$, and $z(t)$. It may even happen that these variables do not develop independently from each other, but that the bacterium maintains some degree of proportionality in its shape, which would mathematically connect $x(t)$, $y(t)$, and $z(t)$ through constraint relationships. The main consequence is that the volume is no longer static, but a function of time. The model has become a dynamic or **dynamical** model.

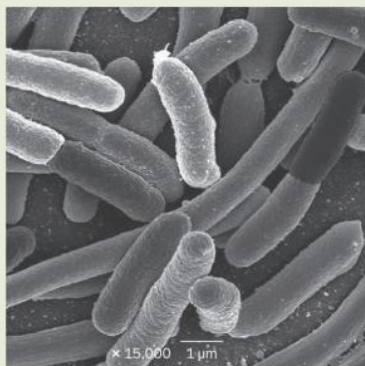


Figure 1 The shape of *E. coli* bacteria resembles cylinders with caps. The shape may be modeled approximately with the geometrical shapes shown in Figures 2 and 3.

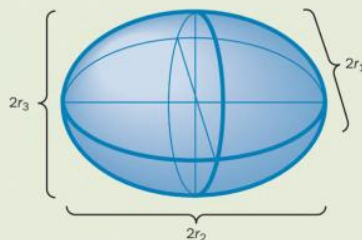


Figure 2 To model the shape of a bacterium (see Figure 1), we need a cylinder and "caps" at both ends. The caps can be modeled with flat ellipsoids. Here, such an ellipsoid is shown with two longer radii of the same length $r_1 = r_2$ and one shorter radius r_3 .

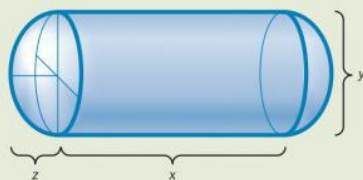


Figure 3 Static model of the volume of *E. coli*. The model is inspired by the actual cell shape (Figure 1) and composed of a cylinder with caps at both ends (Figure 2).

In the typical understanding of dynamical models, one expects that the current state of the object or system will influence its further development. In the case of a bacterium, the growth in volume very often depends on how big the bacterium already is. If it is small, it grows faster, and as it reaches its final size before it divides, growth slows down. This "self-dependence" makes the mathematics more complicated. Specifically, one formulates the change in volume over time as a function of the current volume. The change in volume over time is given in mathematical terminology as its time derivative dV/dt . Thus, we come to the conclusion that volume-dependent growth should be formulated as

$$\frac{dV}{dt} = \text{some function of volume } V \text{ at time } t \quad (3)$$

$$= f(V(t)) = f(V). \quad (4)$$

This type of formulation, in which the derivative of a quantity (dV/dt) is expressed as a function of the quantity (V) itself, constitutes an ordinary differential equation, which is also lovingly called a diff. eq., o.d.e., or ODE. Very many models in systems biology are based on differential equations. The function f is not limited to a dependence on V ; in fact, it most often contains other variables as well. For example, f could depend on the temperature and the substrate concentration in the medium, which will certainly affect the speed of growth, and on a lot of "internal" variables that describe the uptake mechanisms with which the bacterium internalizes the substrate. Looking more closely, many biochemical and physiological processes are involved in the conversion of substrate into bacterial volume and growth, so that a very detailed dynamical model can quickly become quite complicated. Setting up, analyzing, and interpreting these types of dynamical models is at the core of computational systems biology.

It is no secret that biological data always contain variability, uncertainties, measurement inaccuracies, and other types of noise. Thus, as a final point to ponder when choosing a model: should we—must we—always use a model that somehow captures this randomness? Not necessarily. If we are primarily interested in average model responses, a deterministic model will generally do just fine. However, if we need to explore all possible outcomes, the worst and best cases, or the likelihood of specific outcomes from a system with a lot of uncertainty, we may be best advised to choose a stochastic model.

Because this is an important and subtle issue, let's study similarities and differences between deterministic and stochastic models in more detail. As we discussed, each trial in a stochastic system is unpredictable, but we can make very accurate predictions about average or long-term behaviors. By comparison, deterministic models present us with situations where everything is exactly and unambiguously defined. Yet, there are strange cases where the behavior of the system over time is so complicated that we cannot forecast it. A quite simple, but very instructive example is the blue sky catastrophe, which is discussed in **Box 2.2**. This system is completely determined and does not contain any noise or stochastic features. All information about the system is fixed in its definitions and settings, and yet the system is able to generate chaotic time courses that are impossible to anticipate. **Chapter 4** discusses more cases of this nature.

As an example of a stochastic process, suppose we irradiate a culture of microorganisms with ultraviolet light (see, for example, [4]) in order to introduce random mutations. This method of random mutagenesis is often used to create strains that out-compete the wild type in a specific situation of interest. For instance, we may want to find a strain with improved tolerance to acid in its environment. We know from genetics that most mutations are neutral or decrease rather than increase fitness. Nonetheless, one mutation in a million or a billion may improve the microorganism's survival or growth rate, and this particular mutation will rise to the top and multiply in a properly designed random mutagenesis experiment. For simplicity, let's suppose that the mutations occur randomly throughout the genome with a rate of between three and four mutations per kilobase (1000 base pairs of DNA) [5].

A typical question in this scenario is: What is the probability of actually finding three or four mutations in a given DNA segment that is one kilobase long? Our first inclination might be to claim (wrongly) that the probability is 1, because isn't that what we just assumed? Not quite. We assumed that the *average* number of mutations is between three and four, and need to realize that it could happen that there are only one or two mutations, or five or six, owing to the randomness of the mutation process. Models from probability theory allow us to assess situations of this type. For instance, we can forecast the probability of encountering one, two, or three mutations in a given piece of DNA. Or we can make statements like "with p percent probability, we will find at least four mutations." Would it be possible to find 200 random mutations within 1000 base pairs? In a true stochastic process, this outcome is indeed theoretically possible, although the probability is extremely low.

Even in this clear-cut scenario with precise conditions and well-defined questions, different models are available. For instance, we could use the binomial model or the Poisson model. Both are applicable in principle. The binomial model would be better suited for small numbers (here, short segments of DNA), while it would be very cumbersome for large numbers (long segments of DNA). In fact, a pocket calculator would not be able to manage the numbers that would have to be computed for 1000 kilobases. The Poisson model uses a clever approximation, which is not very accurate for small numbers (short segments of DNA), but is incomparably easier to use and misses the correct results for large numbers (long segments of DNA) by only very small errors. **Boxes 2.3** and **2.4** discuss these two models as typical representatives of simple stochastic models.

2.4 System Components

Biological systems models contain different classes of components, which we discuss in this section. The most prominent components are variables, which represent biological entities of interest, such as genes, cells, or individuals. Because biological

BOX 2.3: THE BINOMIAL PROCESS: A TYPICAL STOCHASTIC MODEL

In the text, we asked: What is the probability of finding a certain number of mutations in a DNA segment, if one typically observes three or four mutations per 1000 base pairs? Let's do some math to find out. We formulate the problem as a binomial process, for which we need two ingredients, namely the average mutation rate, which we assume to be 3.5 (instead of "3 or 4") mutations per kilobase, and the number of nucleotides within the DNA segment of interest. From this information, we can compute how likely it is to find 3, 4, 1, 20, or however many mutations within some stretch of DNA.

In the language of probability theory, a mutation in our experiment is called a success and its (average) probability is usually termed p ; in our case, p is 3.5 in 1000, that is, 0.0035. A failure means that the base is not mutated; its probability is called q . Because nothing else can happen in our experiment (either a base is mutated or it is not; let's ignore deletions and insertions), we can immediately infer $p + q = 1$. Put into words: "for a given nucleotide, the probability of a mutation plus the probability of no mutation together are a sure bet, because nothing else is allowed to happen."

In order to deal with more manageable numbers, let's first assume that the DNA piece contains only $n = 10$ bases. Then the probability P to find k mutations is given as

$$P(k; n, p) = \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}, \quad (1)$$

according to the binomial formula of probability theory, which we will not discuss here further. In this formulation, the term $n!$ (pronounced " n factorial") is shorthand for the product $1 \cdot 2 \cdot 3 \cdots n$. The semicolon within the function indicates that k is the independent variable, while n and p are constant parameters, which may, however, be changed from one example to the next. So, what is the probability of finding exactly one mutation in the 10-nucleotide segment? The answer results from substituting values for the parameters and for k , namely,

$$P(1; 10, 0.0035) = \frac{10!}{1!9!} 0.0035^1 (1-0.0035)^9 = 0.0339. \quad (2)$$

That is not a very big probability, so the result is rather unlikely: roughly 3%. What about two mutations? The answer is

$$P(2; 10, 0.0035) = \frac{10!}{2!8!} 0.0035^2 (1-0.0035)^8 = 0.000537, \quad (3)$$

which is even smaller. It is not difficult to see that the probabilities become smaller and smaller for higher numbers of mutations. So, what's the most likely event? It is that there is no mutation at all. This probability is given by the binomial **probability distribution** (1) just like the others, namely,

$$P(0; 10, 0.0035) = \frac{10!}{0!10!} 0.0035^0 (1-0.0035)^{10} = 0.9655. \quad (4)$$

In over 96% of all 10-nucleotide DNA segments, we expect to find no mutation at all!

From these three results, we can estimate the chance of finding more than two mutations. Because all probabilities taken together must sum to 1, the probability of more than two mutations is the same as the probability of *not* having zero, one, or two mutations. Thus, we obtain

$$P(k > 2) = 1 - 0.9655 - 0.0339 - 0.000537 \approx 0.00002, \quad (5)$$

which is 2 in 100,000!

If we increase the length of the DNA piece, we should intuitively expect the probabilities of finding mutations to increase. Using the same formula, the probability of exactly one mutation in a 1000-nucleotide segment can be written as

$$P(1; 1000, 0.0035) = \frac{1000!}{1!999!} 0.0035^1 (1-0.0035)^{999} \quad (6)$$

This is a perfectly fine formulation, but our pocket calculator goes on strike. 1000! is a huge number. To ameliorate the situation, we can play some tricks. For instance, we see from the definition of factorials ($n! = 1 \cdot 2 \cdot 3 \cdots n$) that the first 999 terms in 1000! and 999! are exactly the same and cancel out from the fraction, leaving us simply with 1000 for the first term. With that, we can compute the probability of 1 mutation as 0.1054 and the probability of no mutations at all as 0.0300 (0! is defined as 1). Indeed, in contrast to the scenario with 10 bases above, we are now about three times more likely to encounter one mutation than no mutation at all. Is it imaginable that we would find 10, 20, or 30 mutations? The probability of finding 10 mutations is about 0.00226. Formulate the corresponding probabilities for 20 and 30. It is not hard to do, but even with the tricks above we run into problems computing the numerical results from the formulae. **Box 2.4** offers a solution!

chemical reaction, which is the number of molecules the reaction can process within a given amount of time. The particular values of the parameters depend on the system and its environment. They are constant during a given computer experiment, but may assume different values for the next experiment. Finally, there are (universal) constants, such as π , e , and Avogadro's number, which never change.

In order to design a specific formulation of a model, we begin with a diagram and several lists, which are often developed in parallel and help us with our book-keeping. The diagram contains all entities of the biological system that are of interest and indicates their relationships. These entities are represented by variables and drawn as nodes. Connections between nodes, called edges, represent the flow of material from one node to another. For instance, a protein in a signaling cascade may be unphosphorylated or phosphorylated in one or two positions. Within a short timespan, the total amount of the protein is constant, but the distribution among the three forms changes in response to some signal. Thus, a diagram of this small protein phosphorylation system might look like **Figure 2.7**.

In contrast to the flow of material, a diagram may also contain the flow of information. For instance, the end product of a pathway may signal—through **feedback** inhibition—that no further substrate should be used. This signal is distinctly

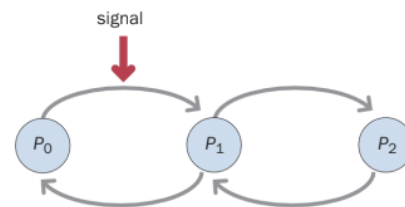


Figure 2.7 Different states of a protein in a signaling cascade. The protein may be unphosphorylated (P_0) or singly (P_1) or doubly (P_2) phosphorylated. The total amount of the protein remains the same, but, during signal transduction, material flows among the pools, driven by kinases and phosphatases, which are not shown here (see **Chapter 9**).

experiment and could therefore be considered independent. This replacement would make the model simpler, but might preclude analyses of other model settings where the variable does change. It actually happens more frequently that independent variables are replaced by dependent variables in later model extensions. A reason could be that the extended system affects the dynamics of a formerly independent variable. Second, if an independent variable is constant throughout an experiment, we could replace it with a parameter. However, it is often beneficial to keep parameters and independent variables separate. First, variables and parameters have a different biological meaning, which may help us with the estimation of appropriate values for them. And second, the computational effort of including an independent variable or a parameter is the same.

The third list is actually closer to a table or spreadsheet. It shows which of the (dependent or independent) variables have a direct effect on any of the processes in the system. As in the example of [Figure 2.8](#), an effect may be positive (enlarging or activating) or negative (diminishing or inhibiting). The list accounts for all edges (flow of material between pools), as well as all signals that affect any of the edges.

The final list contains parameters that contribute to the external or internal conditions of the system, such as pH, temperature, and other physical and chemical determinants, which will ultimately require numerical specification.

Throughout the model design process, we will add to these lists quantitative information about pool sizes, magnitudes of fluxes, strengths of signals, and normal values and ranges of parameters.

The lists are the basis for establishing a diagram of the system. For the design of this diagram, it is beneficial to begin by representing each variable with one of the core modules in [Figure 2.9](#). In the simpler case ([Figure 2.9A](#)), each variable is displayed as a box with one process entering and one exiting. In the end, there may be more or fewer arrows, but using one influx and one efflux is a good default, since it reminds us later to include these processes when we set up equations. Indeed, a dependent variable tends to deplete without influx and to keep accumulating without efflux. Of course there are situations where depletion or accumulation is the desired behavior, and, if so, we can secondarily remove the corresponding process. However, these situations are quite rare. We may also use the default in [Figure 2.9\(B\)](#), which reminds us that many processes are regulated by other variables and that the variable in question may send out signals that could affect the influxes or effluxes of other variables. Independent variables typically do not have an influx, but they have an efflux if they feed material into the system. Once all variables are defined, they need to be functionally connected: the efflux of one variable may become the influx of another, and the signal sent by one variable may affect the influxes or effluxes of other variables.

Let's illustrate this construction of a diagram with the example of a population that is experiencing the outbreak of an infection. The model is intended to answer some of the questions mentioned earlier. Following some old ideas of Kermack and McKendrick [7], we use a similar terminology and keep things as simple as possible; one should, however, note that thousands of variations on this model have been proposed since Kermack and McKendrick's days. We begin by defining just three dependent variables, namely, the number of individuals susceptible to the disease (S), the number of infected individuals (I), and the number of individuals that are "removed" (R) from the two pools, because they have acquired immunity. We suppose that all individuals could be in contact with each other, at least in principle, and assume that a certain percentage of the immune individuals (R) lose their immunity and become susceptible again. We also allow for the possibility that individuals are born or immigrate and that individuals may die while infected. A diagram summarizing this population dynamics is shown in [Figure 2.10](#). Processes



Figure 2.9 Default core modules. When designing a diagram of a system, it is beneficial to augment the display of each variable with an influx and efflux (A) and possibly with modification signals that affect the variable or are sent out by the variable (B).

Toggle switch (genetics)

A phenomenon or **model** in which a gene is either fully expressed or silent, but does not allow intermediates. (Chapter 14)

Toll-like receptor (TLR)

A member of a class of **signaling** proteins that recognize a variety of foreign macromolecules, especially those on the cell walls of bacteria, and trigger the initiation of an immune response. See also **Receptor** and **Transmembrane protein**. (Chapter 7)

Training set

A portion of a dataset used by a **machine learning algorithm** to distinguish different classes of outputs and to establish classification rules. See also **Validation**. (Chapter 5)

Trajectory

The collective change in the **state** of a **system** over time; usually between a **stimulus** and some later state, such as a steady state. (Chapter 10)

Transcription

The process of creating a matching messenger RNA (mRNA) from a DNA sequence. (Chapter 6)

Transcription factor

A protein affecting the expression of a specific gene by binding to the corresponding DNA. (Chapter 6)

Transcriptome

The totality of RNAs transcribed from the **genome** of an organism. See also **-ome**. (Chapter 6)

Transfer RNA (tRNA)

One of several specific RNAs that facilitate the attachment of the correct amino acid to a growing **peptide** or protein, while it is being translated from mRNA. This **translation** process occurs within a ribosome. (Chapter 6)

Transient

A collective term for behaviors of a **system** between the time of a stimulus and the time the system reaches a **steady state**, a **stable oscillation**, or some other **attractor** or point of interest. (Chapter 4)

Translation

The process of generating a protein or **peptide** whose **amino acid** sequence corresponds to the **genetic code** of an mRNA. The process occurs within a ribosome and uses **transfer RNA**. See also **Post-translational modifications**. (Chapter 6)

Transmembrane protein

A member of a large class of proteins that span the membrane of a cell and thereby allow communication between the outside and inside of the cell. Transmembrane proteins are involved in **signal transduction** and a large number of diseases. See also **Gap junction** and **G-protein-coupled receptor**. (Chapter 7)

Transposon

A stretch of DNA that can move from one location in a **genome** to another. (Chapter 6)

Trial (stochastic process)

One of many evaluations of a **stochastic model**. (Chapter 2)

Two-component (signaling) (TCS) system

While not exclusively limited to this definition, usually a specific type of **signaling** mechanism, with which microorganisms sense their environment. See also **Receptor** and **Signal transduction**. (Chapter 9)

Two-dimensional gel electrophoresis

An experimental method for separating proteins by their size and electrical charge. (Chapter 7)

Ubiquitin

A small protein that, when attached to other proteins, tags them for disassembly in the **proteasome** and later recycling of the resulting **peptides** and **amino acids**. The tagging process is called ubiquitination. (Chapter 7)

Uncertainty

Degree to which an experimental result is not exactly known, owing to experimental inaccuracies or other challenges; see also **Variability**. Advanced measurement techniques might reduce uncertainty, but not variability. (Chapter 5)

Validation

The process of testing (and confirming) the appropriateness of a **model** with data that had not been used for the construction of the model. See also **Training set**, **Identifiability**, and **Sloppiness**. (Chapter 2)

Variability

Degree to which an experimental result is affected by differences among individual cells, organisms, or other items under investigation; see also **Uncertainty**. Advanced measurement techniques might reduce uncertainty, but not variability. (Chapter 2)

Variable

A symbol representing a component of a **system**, which may or may not change over time. See also **Dependent** and **Independent variable**. (Chapter 2)

Vector

1. In mathematics, a convenient arrangement of variables or numbers in a one-dimensional array. Often used in conjunction with a **matrix**. Many aspects within the fields of linear algebra, **multivariate** calculus, and statistics are based on vectors and matrices. (Chapter 4)
2. In genetic engineering, an artificially constructed virus or plasmid used to introduce foreign DNA into a bacterium or host cell. (Chapter 14)

Ventricle

One of the two larger chambers of the heart. See also **Atrium**. (Chapter 12)

Vertex (pl. Vertices)

Synonym of **node**. (Chapter 3)

Voltage-gated

Property of a membrane channel to open and close in response to the membrane potential. (Chapter 12)

Western blot

A technique for measuring gene expression based on produced protein. See also **Northern blot**. (Chapter 6)

XML

Abbreviation for eXtensible Mark-up Language. Computer language with a defined vocabulary that permits the structuring of data and facilitates translation into different languages and software packages. See also **Ontology** and **SBML**. (Chapters 1 and 15)

X-ray crystallography

A technique for determining the structure of a protein. The protein must first be crystallized. An X-ray shot through the crystal scatters in a characteristic fashion that allows the deduction of the molecular structure of the protein. (Chapter 7)

Index

Note: Page numbers followed by F refer to figures, those followed by T refer to tables, and those followed by B refer to information boxes

A

ABM (agent-based modeling) 110, 437–438, 438F
 gene regulation 191
ACO (ant colony optimization) 148
Actin 217, 338, 338F
Action potentials 337, 348–350
 cardiomyocytes 337, 348–350, 349F, 350F
 lengthening during repolarization 355
 models 350–354, 353F
 phases 349–350, 350F
 prolongation 361
 repeated 354, 354F
 SA and AV node cells 349
 sodium–calcium exchanger 347
Activator(s), metabolic systems 238–239
Activator module 417
Acute lymphoblastic leukemia (ALL) 373–374
Adaptation 273–274, 273F, 274F, 274T
Adaptive evolution 407
Adaptive immune system 428–429, 429F
Adenine 171, 172F, 173F
Ad hoc models 101–102, 101F, 102F
Adjacency matrices 55, 55F
Adrenaline 336
Affinity 139
AFM (atomic force microscopy) 174F
Agent-based modeling (ABM) 110, 437–438, 438F
 gene regulation 191
Aging
 heart failure 355–356
 inflammation 429
 lens 217
Alanine (Ala) 203F, 204T
Albumin, serum 212–213
Algae 400, 418
Algorithms for parameter estimation 135
 nonlinear systems 142–143
 comprehensive grid search 142, 143–145, 144F
 evolutionary 142–143, 146–148, 147F
 genetic 142–143, 146–148, 147F
 gradient 142, 143F
 other stochastic 148–149
ALL (acute lymphoblastic leukemia) 373–374
Allostasis 373
Allosteric inhibition 211, 239
Allosteric regulation 239
 α -helices 221, 221F
Alzheimer's disease 14
Amino acids 202–204, 203T, 204T
 codons 174–175
Amorpha-4,11-diene 414, 415, 415F
Amorphadiene 414
AND gate 402
Anfinsen, Christian 204
Animal breeding 407
ANN (artificial neural network) 374, 426
ANS (autonomic nervous system) 336
Antagonistic pressures, heat stress response 308
Ant colony optimization (ACO) 148
Antibodies 215, 215F
 monoclonal 382F–385F, 383–385, 384T
Antibody binding 382F–385F, 383–385, 384T
Antigen, immune cascade 404–406
Antimalarial drug 414–415
Anti-parallel strand 171, 172F
Antiporter 346–347
Aplysia californica, neural networks 427, 427F
Apolipoproteins 217
Apoptosis, differentiation 306
Applications, emerging 426–434
Approximation(s) 29, 29F
 canonical 101

 continuum 336
 linearized models 95–98, 96B, 97B, 97F
 power-law 108B, 109B, 239
Aptamers 410
Aquaporins 217
Aquatic systems 3
 oceans 433–434, 434F, 435F
Archaea 266
Arginine (Arg) 203F, 204T
Arrhythmias 355
Artemisinic acid production 414–415, 415F
Artemisinin production 414–415, 415F
Artificial neural network (ANN) 374, 426
Ascorbate pathway 250
Asparagine (Asn) 203F, 204T
Aspartic acid (Asp) 203F, 204T
ATHH gene 310, 325
Atomic force microscopy (AFM) 174F
Atrial fibrillation 356
Atrioventricular (AV) node 336
Atrioventricular (AV) valves 333F, 334
Atrium(a) 333–334, 333F
Attractors 122–128
 chaotic 126–128, 127F, 128F
 defined 122
 limit cycles 123–125, 123F, 125F, 126F
 Lorenz 126–127, 127F
 strange 126–128, 127F, 128F
Automaticity 349
Autonomic nervous system (ANS) 336
Auto-regulation 402, 402F
AV (atrioventricular) node 336
AV (atrioventricular) valves 333F, 334
Axioms 439–440

B

Bacillus anthracis 267
Bacillus subtilis 410B
Bacteria
 actin 217
 barrels 222F
 biofilms 433, 433F
 carrier proteins 346
 cyano- 208, 434, 434F
 cytokines 215
 expressed sequence tag 193
 feedforward loop 403–404
 flagellar motor 213–214, 213F
 gene regulation 406
 genetic engineering 409, 418
 genome
 organization 188, 401
 size 174
 glycoproteins 218
 inflammation 431
 keratin 217
 lipo-oligosaccharides 209F
 metapopulations 432
 operons 185, 188
 plasmids 170, 178
 population growth 284, 284F, 285, 286, 292
 predation 288
 quorum sensing 278
 riboswitches 410B
 stress signal sensing 403
 Toll-like receptors 215
 transcription factors 188–189
 transposons 170
 trehalose cycle 310
 tryptophan 170
 two-component signaling systems 266–269,
 266F–269F, 268T
 vectors 409
 “wild type” 16
Bacteriophages 409
Barcoding 175, 175F
Barrels 221, 222F
Base(s), DNA and RNA 171, 172F, 173F
Base pairs 171, 172F, 173F
Basin of attraction 262
Bassingthwaite, James 332
Bayes' formula 65
Bayesian inference 64–69
Bayesian reconstruction, interaction networks 63–69,
 65F, 67F
Bayes' rule 65
Bayes' theorem 65
Behavior(s), of discrete vs continuous models 29
Behavioral conditioning 440
Bernard, Claude 10
Bertalanffy, Ludwig von 10
Berzelius, Jöns Jacob 201
 β -sheets 221, 222F
Bidirectional graphs 53, 55F
Bi-fan motifs 402, 402F
Bifurcations 119
Bimolecular metabolic reactions 234–235
Binomial model 30, 32B
Biochemical reactions, metabolic systems 232–240
 background 232–234, 233B
 mathematical formulation of elementary reactions
 234–235
 rate laws 235–240, 235F, 237F, 238F
Biochemical systems theory (BST) 103
 Generalized Mass Action (GMA) systems 103–104,
 104F
 metabolic pathways 239, 242
 S-system 104–105, 114–115
 trehalose cycle 311
BioCyc 74, 242, 244B
Bioengineering 11
Biofilms 278, 433, 433F
Biofuel 400, 411
Bioinformatics 12, 74
 DNA sequences 178
 Markov models 89
Biological components, variations 9–10
Biological process, gene products 176
Biological Revolution 3
Biological systems 1–18
 analyzing 3–4, 4F
 capability to manipulate 2–3
 complicated 1, 2F
 confusing simple 8–10, 8F, 9F
 conversion into computational models 16
 design 399–423
 case studies 411–418
 design principles 399, 404–406, 404F, 406F
 drug development 414–415, 415F
 elementary mode analysis 411–414, 413F
 future 419
 gene circuits 415–418, 416F–418F
 goal-oriented manipulations 407–408
 metabolic engineering 407–408, 411–414, 413F
 natural 400–407
 network motifs 399, 402–404, 402F, 403F
 operating principles 406–407, 406F
 optimization 400
 overview 399–400
 structural patterns 400–402, 401F
 synthetic 399, 408–410, 410B, 411–418
 emergent properties 11, 12
 engineering of 3

- examples [1](#), [2F](#)
 goal of deeper understanding of [3](#)
 history of dealing with [1-2](#)
 mathematics [83-134](#)
 nonlinear processes in [5](#), [5F](#), [8](#)
 size vs complexity of [4-5](#), [4F](#)
 static structure and dynamics of [16](#)
 Biological target, drug development [379](#)
 Biological theory [439-441](#)
 Biological traits [176](#)
 Biological variability, and disease [369-370](#)
 modeling [370-372](#), [371F](#), [371T](#), [372F](#)
 Bioluminescence [278](#)
 Biomarkers [373-374](#), [374F](#)
 disease [218](#), [373-376](#), [374F](#)
 normal range [377](#), [377F](#)
 Biomass [411](#), [413-414](#)
 Bionanotechnology [11](#), [13F](#)
 Biotechnology [11](#), [13F](#)
 Bipartite graphs [54](#)
 Bistability
 gene circuits [416](#)
 signaling systems [261-263](#), [262F](#), [262T](#), [263F](#)
 Black box models, oscillations [339-343](#), [340F-342F](#)
 Blood sugar level [373](#)
 Blue Brain Project [426](#), [426F](#), [427F](#)
 Blue sky catastrophe [30](#), [31B](#)
 Bolus experiments [120](#), [120F](#)
 Boolean networks [259-261](#), [259F](#), [260F](#)
 Bootstrap method [152-153](#)
 Bottom-up strategy [16-17](#)
 parameter estimation [136](#)
 Bradycardia [355](#)
 Brady-tachy syndrome [355](#)
 Brain, neurons [362](#)
 Brain tumor [189](#), [193](#), [224F](#)
 Branch-and-bound methods [144-145](#)
 Branched pathway [406](#), [406F](#)
 Breeding [407](#)
 BRENDA [240](#), [241B](#)
 pathway screening [392-393](#)
 Brenner, Sydney [17](#)
 BST *see* [Biochemical systems theory \(BST\)](#)
 Bundle of His [336](#), [337-338](#)
Burkholderia pseudomallei [175F](#)
- C**
 CAD (computer-aided design) [410](#)
Caenorhabditis elegans, neural networks [427](#), [427F](#)
 Calcium adenosine triphosphatase (Ca²⁺ ATPase),
 plasma membrane [347](#)
 Calcium (Ca²⁺) channels
 L-type [349](#), [356F](#), [357-361](#), [357F](#)
 T-tubules [356](#)
 T-type [349](#)
 Calcium dynamics, dyads [356-361](#), [356F](#), [357F](#),
 [359F](#), [360F](#)
 Calcium fluxes
 dyads [356F](#), [358](#)
 slowing down [355](#)
 Somogyi-Stucki model [344](#), [344F](#)
 Calcium-induced calcium release (CICR) [343](#), [357](#)
 Calcium release unit (CaRU) [361](#)
 cAMP *see* [Cyclic adenosine monophosphate \(cAMP\)](#)
 Cancer [431-432](#), [431F](#)
 Canonical models [101](#), [102-110](#)
 concept-map modeling [438](#)
 generalized mass action (GMA) systems
 [103-104](#), [104F](#)
 limit-cycle oscillations [102](#)
 lin-log model [107](#), [115](#)
 Lotka-Volterra [101](#), [102-103](#), [103F](#), [114](#)
 power-law approximation [107](#), [108B](#), [109B](#)
 S-systems [104-105](#), [114-115](#)
 toggle switch model for SOS pathway of *E. coli*
 [105-107](#), [105F](#), [107F](#)
 trehalose cycle [310-311](#)
 CAP (catabolite activator protein) [186](#), [186F](#), [187](#)
 Capacitance [351](#)
 Capillary electrophoresis (CE) [247](#)
 Caprioli, Richard [222](#)
 Carcinogenesis [431-432](#), [431F](#)
 Cardiac cells *see* [Cardiomyocytes](#)
 Cardiac cycle [356F](#)
 blood flow [333F](#), [335](#)
 ECG [339F](#)
 metabolic responses to oxygen supply [342](#)
 movement of ions [343](#), [347](#)
 shape of heart [336](#)
 Cardiac output [334](#)
 Cardiomyocytes [334](#), [335-336](#)
 action potentials [337](#), [348-350](#), [349F](#), [350F](#)
 models [350-354](#), [353F](#)
 aging [356](#)
 automaticity [349](#)
 contraction [332](#)
 electrochemistry [345-354](#)
 biophysical description [347-348](#)
 models of action potentials [350-354](#), [353F](#)
 repeated heartbeats [354](#), [354F](#)
 resting potentials [348-350](#), [349F](#), [350F](#)
 excitable [336](#), [337](#), [338-339](#), [338F](#)
 funny currents [349](#)
 heart failure [355-356](#)
 modeling based on molecular events [356-361](#),
 [356F](#), [357F](#), [359F](#), [360F](#)
 oxidative stress [355-356](#)
 resting potentials and action potentials [337](#),
 [348-350](#), [349F](#), [350F](#)
 sarcolemma [356-357](#), [357F](#)
 sinoatrial node [337](#)
 β-Carotene [415](#)
 Carrier proteins [346-347](#)
 Carriers [211-214](#), [212F](#), [213F](#)
 Carrying capacity [284](#), [288](#), [288F](#)
 CaRU (calcium release unit) [361](#)
 Case studies [16](#), [17](#)
 Catabolite activator protein (CAP) [186](#), [186F](#), [187](#)
 Causality analysis [62](#)
 cDNA (complementary DNA) [192](#)
 CE (capillary electrophoresis) [247](#)
 Cell cycle [60](#)
 genomic regulation [274](#)
 heat stress response [306](#)
 oscillator [276](#)
 regulatory network [260](#)
 subpopulation analysis [292](#)
 Cell membrane
 depolarization [337](#)
 gated channels [353-354](#), [355](#)
 heart function and failure [356](#)
 ion movements [345](#), [346](#)
 resting potential [348](#)
 T-tubules [358](#)
 Cell-to-cell communication [211](#)
 Cellular automata [437](#), [438](#)
 Cellular component, gene products [176](#)
 Cellular stress response [305](#)
 Central Dogma [169-171](#), [170F](#), [440](#)
 Central limit theorem [358](#)
 Central nervous system [11](#)
 Chalfie, Martin [222](#)
 Channel proteins [346](#)
 Chaotic attractors [126-128](#), [127F](#), [128F](#)
 Chaotic dynamics [298](#)
 Chaotic process [93](#)
 Chaperones [205](#), [207F](#)
 Chaperonins [205](#)
 Charge gradient [345](#)
 CheY protein [266](#), [266F](#)
 Children, growth curves [47](#), [48F](#)
 Chitin [217](#)
 Chloroplasts [170](#), [188](#)
 Christou, Paul [415](#)
 Chromatin [178](#)
 Chromatography [193](#), [219](#), [247](#)
 Chromosomes [178](#)
 CICR (calcium-induced calcium release) [343](#), [357](#)
 Ciechanover, Aaron [205](#)
 Cilia [213](#)
 Circadian clocks [275-278](#), [276F](#), [277F](#), [278T](#)
cis-regulatory motifs [185](#)
 Citric acid [399](#), [408](#)
 Clinical phase, drug development [380](#)
 Clinical trials [380](#)
 Cliques [55](#), [60](#)
 Closed system model [29](#)
 Clustering, heat stress response [317](#)
 Clustering coefficient [55-56](#), [56F](#)
 Coding regions [177](#)
 Codon(s) [174-175](#), [202-204](#), [204T](#)
 Codon bias [175](#)
 Codon law [440](#)
 Codon usage bias [204](#)
 Coefficients [94](#)
 Cohen, Joel [60](#)
 Collagen [216](#), [217F](#)
 Collins, James [417-418](#)
 Communication [13-15](#), [14F](#)
 Community proteomics [433](#)
 Compartment models [94](#)
 pharmacokinetics [385-387](#), [385F-387F](#)
 Competitive inhibition [239](#)
 Complementary DNA (cDNA) [192](#)
 Complexities, Lotka-Volterra model [103](#)
 Comprehensive grid search [143-145](#), [144F](#)
 Comprehensive screening experiments [121](#)
 Computational analysis [241-244](#), [243B](#), [244B](#)
 Computational models [16](#)
 Computational power [435](#)
 Computational systems biology [375](#), [375F](#)
 Computer-aided approximation [12](#)
 Computer-aided design (CAD) [410](#)
 Computer-based reasoning systems [433](#)
 Computer languages [427](#), [435](#)
 Computer simulations [19](#), [38](#), [38F](#)
 personalized medicine [376-377](#)
 population growth [286](#)
 Computing environment [12-13](#), [13F](#)
 Concatamers [193](#)
 Concentration gradient [345](#)
 Concept-map modeling [437](#), [438-439](#), [439F](#)
 Conceptual model [15](#)
 Conditional probabilities [64-65](#), [65F](#)
 Conductance [336](#), [351-353](#)
 Conformation [187](#), [204](#)
 Congestive heart failure [349F](#), [355](#), [356](#)
 Conjugation [409](#)
 Connections [16](#)
 Connectivity relationships [76-77](#)
 trehalose cycle [308](#), [308F](#)
 Connexins [211](#), [212F](#)
 Consistency [38-40](#), [39F](#)
 Constrained optimization [73](#)
 Constraints [70](#), [73](#)
 Continuous linear systems [93-100](#)
 linear differential equations [94](#)
 linearized models [95-100](#), [95F](#), [96B](#), [97B](#), [97F](#),
 [99F](#), [100F](#)
 Continuous models [29](#), [84](#)
 Continuous nonlinear systems [100-110](#)
 ad hoc models [101-102](#), [101F](#), [102F](#)
 canonical models [101](#), [102-110](#), [103F-105F](#),
 [107F](#), [108B](#), [109B](#)
 more complicated dynamical systems
 descriptions [110](#)
 Continuum approximation [336](#)
 Control coefficients [74-75](#)
 Controlled comparison [404-406](#), [406F](#)
 Conway, John [260](#)
 Cook, James [6F](#)
 Cooperativity [211](#)
 Coral growth [434](#), [435F](#)
 Co-regulating genes [188](#)
 Corn [415](#)
 Coronaries [335](#)
 Correlative models [27](#), [27F](#)
Corynebacterium glutamicum [189](#), [251](#), [408](#)
 Crick, Francis [169](#), [171](#)
 Critical point analyses [121-122](#), [122F](#)
 Cross-bridges [338](#), [338F](#)
 Crosstalk, MAPK cascade [273](#)
 Crowding effect [285](#)
 Cryo-electron microscopy (cryo-EM) [206](#), [207F](#)
 Curved surface, linearization of [95](#), [95F](#)
 Customized mark-up languages (XML) [12](#)
 Cyanobacteria [208](#), [434](#), [434F](#)
 Cyclic adenosine monophosphate (cAMP) [187](#)
 funny currents [349](#)
 heat stress response [307](#)
 signaling systems [258](#)
 Cysteine (Cys) [203F](#), [204T](#)
 Cystine [202](#)
 Cytochromes [213](#)

Cytokine(s) 215, 428–429, 429F
 Cytokine storm 428–429, 429F
 Cytoscape 60, 61F
 Cytosine 171, 172F, 173F

D

DAG (directed acyclic graph) 67–68
 Damped oscillations 103
 Data availability 25–26
 Data mining 250, 381
 Data-modeling pipeline 437–439, 438F, 439F
 Data warehouses 433
 DD (duplication-divergence) model 56
 Degree distribution 57, 58F
 Degrees associated with nodes 54–55, 55F
 Demand theory 406, 440
 Denaturation 306
 Deoxyribonucleic acid *see* DNA (deoxyribonucleic acid)
 Deoxyribose 172F
 Dependencies among network components 62–63, 63F
 Dependent variables 33–34, 33f
 Depolarization 337, 348, 350, 350F
 Design principles 399, 404–406, 404F, 406F
 Detection, metabolite analysis 247–249, 248F
 Deterministic models 29, 30, 31B
 population dynamics 291–292
 recursive 85–88, 85T, 86F
 Diabetes 29, 373, 381, 428
 Diagram, model development 34–35, 34F, 35F
 Diastole 334
 Dicer 184
 Difference equations 91
 Difference gel electrophoresis 219
 Differential equations
 linear 94
 models of gene regulation 190
 ordinary 91, 110
 heart 358
 metabolic systems 234
 models of gene regulation 190
 population growth 284, 285–286
 signaling systems modeled with 258, 261–278
 parameter estimation for systems of 153–159, 153F, 154T, 156F–159F, 156T–158T
 partial 110
 heart 334, 358
 models of gene regulation 191
 signaling systems modeled with 261–278
 adaptation 273–274, 273F, 274F, 274T
 bistability and hysteresis 261–265, 262F–264F, 262T, 264T, 265B–266B
 mitogen-activated protein kinase cascades 270–273, 270F–273F, 271T
 other 274–278, 276F, 277F, 278T
 stochastic 258, 261–278
 two-component 266–269, 266F–270F, 268T
 stochastic 258, 261–278
 Differentiation 84
 Differentiation apoptosis 306
 Diffusion process 234
 Digitalis 347
Digitalis purpurea 347
 Directed acyclic graph (DAG) 67–68
 Directed evolution 209F, 407
 Directed graphs 53, 53F, 55–56, 55F, 56F, 62
 Discrete linear systems models 85–91
 recursive deterministic 85–88, 85T, 86F
 recursive stochastic 84F, 88–91
 Discrete models 29, 84
 Discrete nonlinear systems 91–93, 92F
 Discrete process 41
 Disease(s)
 biological variability and 369–370
 modeling 370–372, 371F, 371T, 372F
 complex 369, 428
 delineation 377, 377F
 see also Biomarkers; Infectious disease modeling
 Disulfide bond 202
 DNA (deoxyribonucleic acid)
 adducts 178
 barcoding 175, 175F

base pairs 171, 172F, 173F
 chemical and physical features 171–172, 172F–174F
 codons 174–175, 202–204, 204T
 double-stranded 170, 171, 174, 188, 193
 epigenetics 170, 178
 eukaryotic packing 178, 179F, 180F
 exons and introns 177, 177F
 genetic variation 170
 hybridization 171
 “junk” 176–177
 key properties 171–178
 noncoding 174, 175–178, 177F
 nucleotides 171, 172F
 origin-binding domains (OBDs) 173F
 random mutations 30
 replication 171–172, 173F
 sequencing 174–175
 size and organization 174–175, 175F
 supercoiled 174, 174F
 synthetic biology 408–410
 transcription 169, 172, 176, 180
 hierarchical control 189, 189F
 translation 176
 DNA adducts 178
 DNA chips 193–194, 193F
 heat stress response 315
 DNA methyltransferases (DNMTs) 179F
 Dose 379
Drosophila 194, 195F
 Drought, plant response 2F
 Drug development 378–393
 antimalarial 414–415
 biological systems designs 414–415, 415F
 biological target 379
 biologics 380–381, 382F
 clinical phase 380
 compartment models 385–387, 385F–387F
 lead identification 381
 lead optimization 379
 models
 compartment 385–387, 385F–387F
 dynamic 390–393, 391F
 pharmacokinetic 385–390, 386F, 387F, 388B–389B
 receptor dynamics 382–385, 382F–385F, 384T
 new chemical entities (NCEs) 378–379
 pipeline 378–380, 379F, 381, 391
 post-approval phase 380
 preclinical phase 378–379, 379F
 process 378–380, 379F
 quantitative structure-activity relationships (QSARs) 379
 role of systems biology 380–393
 computational target and lead identification 381
 emerging 393
 pathway screening with dynamic models 390–393, 391F
 pharmacokinetic modeling 385–390, 386F, 387F, 388B–389B
 receptor dynamics 382–385, 382F–385F, 384T
 toxicity testing 379
 Drug sensitivity 373–374
 Duplication-divergence (DD) model 56
 Dyads, calcium dynamics 356–361, 356F, 357F, 359F, 360F
 Dynamical features, exploration and validation 38, 40–42, 41B, 42F, 43F
 Dynamic(al) models 83–134
 continuous linear 93–100
 linear differential equations 94
 linearized models 95–100, 95F, 96B, 97B, 97F, 99F, 100F
 continuous nonlinear 100–110
 ad hoc models 101–102, 101F, 102F
 canonical models 101, 102–110, 103F–105F, 107F, 108B, 109B
 more complicated dynamical systems descriptions 110
 defined 27–29, 28B
 discrete linear 85–91
 recursive deterministic 85–88, 85T, 86F
 recursive stochastic 84F, 88–91
 discrete nonlinear 91–93, 92F
 other attractors 122–128
 chaotic 126–128, 127F, 128F
 limit cycles 123–125, 123F, 125F, 126F

overview 83–85
 pathway screening 390–393, 391F
 standard analyses 110–122
 analysis of systems dynamics 119–122, 120F, 122F
 parameter sensitivity 118–119
 stability analysis 115–117, 116F, 117F
 steady-state analysis 110–115, 112F, 113F
 Dynamics
 of biological systems 16
 model development 24–25
 Dystrophin 232

E

EC (excitation-contraction) coupling 338
 EC (Enzyme Commission) number 241F
 Ecoinformatics 433–434
 Edelstein-Keshet, Leah 86
 Edges 53, 55
 Effector cells 404–406, 404F, 406F
 Efficacy 379
 Eigenvalues
 linear differential equations 94
 stability analysis 116–117
 trehalose cycle 311
 Elasticities 74, 76
 Elasticity coefficients 76
 Elastins 216
 Electrical charge 350–351
 Electrical conductance 336, 351–353
 Electrical conductor 351
 Electrical gradient 345
 Electrical potential 351
 Electrical resistance 351
 Electric current 351
 Electrocardiogram (ECG, EKG) 339, 339F
 Electrochemical gradients 347–348
 Electrochemical processes, heart 332
 Electrochemistry, cardiomyocytes 345–354
 biophysical description 347–348
 models of action potentials 350–354, 353F
 repeated heartbeats 354, 354F
 resting potentials and action potentials 348–350, 349F, 350F
 Electromotive force (EMF) 351
 Electrophysiology, heart 332
 Elementary mode analysis (EMA) 411–414, 413F
 Ellington, Andrew 418
 Embryonic cell cycle oscillator 276
 Emergent properties 11
 EMF (electromotive force) 351
 Entropy 63, 63F
 Environment
 interactions of organisms with 432–434, 433F–435F
 model development 29
 Environmental pressures 16
 EnvZ two-component sensing system 268, 406
 ENZYME 240
 Enzyme(s) 209–211, 209F, 210F
 activity profile 52
 catalyzing 72F, 73, 102, 102F, 139, 139F, 235, 235F
 control of pathway systems 245–246
 dicer 184
 drug development 379, 392, 414, 415F
 flux analysis 249
 functional loop 8–9, 8F
 gene circuits 418
 gene expression 191–193, 315–318
 generalized mass action systems 107
 gene regulation 185
 heat stress response 306–309, 307F, 312–318, 313F, 314F
 hexokinase 231, 232F
 trehalose cycle 308–309
 immune system 214
 incompletely characterized organism 250
 independent variables 33, 121
 kinases 214
 kinetics 102, 102F, 160, 240–241
 lac operon 186, 186F, 187
 metabolic control analysis 75–78, 75F
 metabolic engineering 407–408
 metabolic network analysis 251
 metabolic systems 232, 233, 233B, 243B

- reconstruction 74, 74F
 - Michaelis-Menten rate law 236–238, 237F, 238F
 - mitogen-activated protein kinase cascades 270–273
 - multiscale analysis 318–326, 324F, 324T, 362
 - personalized medicine 373, 375
 - perturbations 371–372, 392
 - phospholipase A 222F
 - population growth 283
 - power-law approximations of multivariate functions 109B
 - protein activity and dynamics 224
 - protein folding 205
 - proteomics 220
 - rate laws 236–239
 - ribozymes 183, 183F
 - sensitivity analysis 118
 - signal transduction 258, 266
 - small RNAs 184
 - substrate conversion 36–37
 - synthetic biology 408, 410, 412–413
 - transcription factors 189
 - tryptophan pathway 170
 - Enzyme-catalyzed reaction 72F, 73, 102, 102F, 139, 139F, 235, 235F
 - Enzyme Commission (EC) number 241F
 - Epidemics 297
 - Epidermal growth factor receptor 214, 214F
 - Epigenetics 170, 178, 179F, 370
 - Epinephrine 336
 - ERG9* gene 415
 - ERGO 242
 - ERK (extracellular-signal-regulated kinase) 270
 - Escherichia coli*
 - analyzing metabolism in incompletely characterized organism 250, 251
 - barcode 175F
 - barrel structure 222F
 - computation of volume 28B
 - CheY signal transduction 266
 - design of biological systems 400, 402, 406
 - elementary mode analysis 411–414
 - gene regulation 185, 190
 - lac* operon 186
 - metabolic engineering 407, 411, 412
 - metabolic flux distribution 72, 72F
 - modeling variability and disease 371
 - network motifs 403
 - resources for computational pathway analysis 242
 - single mutants 11
 - SOS pathway 105–107, 105F, 107F
 - supercoiled DNA 174F
 - synthetic biology 408, 410
 - toggle switch model 105, 105F
 - transcription factors 189
 - two-component signaling system 266, 268
 - EST (expressed sequence tag) 193
 - Ethanol 399, 400, 411, 412, 414
 - Euler, Leonard 54, 54F
 - Evolutionary algorithms, parameter estimation 142–143, 146–148, 147F
 - Evolutionary programming 148
 - Excitable cardiomyocytes 336, 337, 338–339, 338F
 - Excitation-contraction (EC) coupling 338
 - Exons 177, 177F
 - ExPASy database 251
 - Experimental systems biology 374–375, 375F
 - Explanation 16
 - Explanatory models 27, 27F
 - Exploration, mathematical model 24, 38, 40–42, 41B, 42F, 43F
 - Exponential growth 283, 286
 - Exposome 372
 - Expressed sequence tag (EST) 193
 - External validation 40
 - Extracellular matrix 208
 - Extracellular-signal-regulated kinase (ERK) 270
 - Extraction, metabolite analysis 247, 251–252
 - Extrapolations 19
- F**
- Factor VIII 381
 - False-negative outcomes 377–378
 - False-positive outcomes 377, 378
 - Farnesyl pyrophosphate (FPP) 414
 - FBA (flux balance analysis) 73, 411
 - FDA (Food and Drug Administration) 378, 379F, 380, 382F
 - Feature selection, structure identification 160
 - Feedback
 - metabolic pathways 245, 245F, 246F
 - nested 245, 246F
 - Feedback inhibition 32–33, 33F
 - metabolic pathways 245, 245F, 246F
 - Feedback loops
 - design of biological systems 403, 403F
 - positive 428
 - Feedforward loops, design of biological systems 403–404, 404F
 - Fermentation 25, 400
 - Fibroblasts 216
 - Fibronectins 213
 - Filaments, heart 334
 - Finite element approaches, heart 335
 - First-degree heart block 355
 - Fish populations 434
 - Fitness 142
 - Fitting 137
 - FitzHugh, Richard 332
 - FitzHugh-Nagumo equations 354
 - Flagellar motor 213–214, 213F
 - Flow chart 22, 22F
 - Flow of information 32–33, 33F
 - Flow of material 32, 33B
 - Fluid mechanics, heart 335
 - Flux(es) 52, 70
 - metabolic systems 240, 242–244
 - trehalose cycle 311
 - Flux analysis, metabolic 249
 - Flux balance analysis (FBA) 73, 411
 - Flux network 70–73, 70F–72F
 - Food(a), transgenic 415
 - Food and Drug Administration (FDA) 378, 379F, 380, 382F
 - Force-directed layout 60, 61F
 - F6P (fructose 6-phosphate) 72, 242, 309
 - FPP (farnesyl pyrophosphate) 414
 - Fractals 93, 122, 432
 - Franklin, Rosalind 171
 - Frank-Starling law 334
 - Free radicals 355
 - Fructose 1,6-bisphosphate 8, 318
 - Fructose 6-phosphate (F6P) 72, 242, 309
 - Fugu rubripes* 174
 - Functional loop 8–9, 8F, 9F
 - Funny currents 349
 - Future developments 16, 17
 - Fuzzy laws 440
- G**
- GA(s) (genetic algorithms) 142–143, 146–148, 147F
 - Gain analysis 118
 - trehalose cycle 311–312
 - Galactose 186
 - Galactose transacetylase 186, 186F
 - Galactose utilization gene network 60, 61F
 - β -Galactosidase 186, 186F
 - β -Galactoside permease 186, 186F
 - Galileo 62
 - Game of Life* (Conway) 260
 - Gap junctions 211, 212F, 337
 - Gas chromatography with mass spectrometry (GC-MS) 247
 - Gates 402
 - GC-MS (gas chromatography with mass spectrometry) 247
 - Gel electrophoresis
 - difference 219
 - sodium dodecyl sulfate polyacrylamide 191
 - two-dimensional (2D) 219, 219F
 - Gene(s)
 - biological variability 369–370
 - modeling 370–372, 371F, 371T, 372F
 - Central Dogma 169–171, 170F
 - coding 176
 - defined 176
 - human 6, 176, 177
 - information transfer 170
 - noncoding DNA 174, 175–178, 177F
 - open reading frame 177
 - phenotype 176
 - transcription network 402
 - Gene annotation 171
 - Gene circuits 415–418, 416F–418F
 - Gene expression
 - control of metabolic pathways 245–246
 - localization 194, 195F
 - measuring 191–194, 192F, 193F
 - structural patterns 401
 - trehalose cycle 308, 309F, 315–318, 316F, 316T, 317F
 - Gene Expression Atlas 194
 - Gene Expression Omnibus 194
 - Gene interaction network 190
 - Gene Ontology (GO) Consortium 176
 - Generalized mass action (GMA) systems 103–104, 104F
 - structure identification 160
 - trehalose cycle 311
 - see also* Biochemical Systems Theory
 - Gene regulation 185–191
 - lac* operon 186–187, 186F, 187T
 - models 190–191, 190F, 191F
 - modes 187–188
 - transcription factors 169, 188–190, 189F
 - Gene regulatory networks 69
 - as signaling systems 274
 - Gene systems 169–200
 - Central Dogma 169–171, 170F
 - DNA
 - chemical and physical features 171–172, 172F–174F
 - epigenetics 170, 178
 - eukaryotic packing 178, 179F, 180F
 - genes and noncoding 175–178, 177F
 - key properties 171–178
 - size and organization 174–175, 175F
 - gene expression
 - localization 194, 195F
 - measuring 191–194, 192F, 193F
 - gene regulation 185–191
 - lac* operon 186–187, 186F, 187T
 - models 190–191, 190F, 191F
 - modes 187–188
 - transcription factors 169, 188–190, 189F
 - outlook 196
 - RNA
 - chemical and physical features 172F, 178–179, 181F
 - key properties 178–185
 - messenger 179
 - ribosomal 182–183, 183F
 - small 183–184, 184F
 - transfer 170, 182, 182F
 - viruses 184–185, 185F
- Genetic algorithms (GAs) 142–143, 146–148, 147F
- Genetic code 174, 440
- Genetic engineering 409
- Genetics, Mendelian 440
- Genetic variation 69, 370
- Genome(s)
 - bacterial 188, 401
 - human 6, 171, 184
 - integrative analysis 303–330
 - size and organization 174–175, 175F
 - structural patterns 401, 401F
 - transposons 170–171
- Genome comparison 74, 74F
- Genome engineering, multiplex automated 407
- Genomic networks 69
- Genomic regulation, cell cycle 274
- Genomics
 - meta- 172, 433
 - pharmaco- 373–374
- GFP (green fluorescent protein) 194, 222, 223F
- Gibbs sampler 68
- Gilman, Alfred 258
- Gln (glutamic acid) 203F, 204T
- Global optimum 144
- Glu (glutamine) 203F, 204T
- Glucokinase, trehalose cycle 308–309

- Glucose 231, 232F
 bacteria 153
¹³C-labeled 250, 325, 325F
 diabetes 29
 enzymes 209
 heat stress response 306, 308, 308F
 accounting for dynamics 314–315, 315F
 analysis 312–314
 gene expression 317
 metabolic pathway model 311–313, 312F
 multiscale analysis 318–323, 319F–321F
 trehalose puzzle 325
lac operon 186–187, 186F
 metabolic control analysis 242
 metabolic engineering 414
 metabolic flux distribution 72, 72F, 73
 metabolic networks 70, 231, 232F, 233
 metabolomic data generation 247
 simple system 8, 10
 small metabolite 169–170
- Glucose 6-phosphate (G6P)
Escherichia coli 72, 72F
 glycolysis 8
 trehalose cycle 308, 309–310, 311, 314, 316, 317, 318, 320F
- Glutamic acid (Gln) 203F, 204T
 Glutamine (Glu) 203F, 204T
 Glycans 217
 Glycine (Gly) 203F, 204T
 Glycogen 308
 Glycogen synthase 323
 Glycolysis
 computational pathway analysis 242
 enzymes 209
 in incompletely characterized organism 250–251
 simple systems 8
 static metabolic network 70
 time-series data 252
 trehalose cycle 308, 308F, 309–310, 311, 316, 317, 318
- Glycoproteins 217–218
 GMA systems *see* Generalized mass action (GMA) systems
- Goal-oriented manipulations 407–408
 Goals, model development 24
 GO (Gene Ontology) Consortium 176
 Golden Rice 415
 Goldman-Hodgkin-Katz equation 348, 360
 Gompertz growth law 284–286
 Gout 391
 G6P *see* Glucose 6-phosphate (G6P)
 G-protein(s) 257–258
 G-protein-coupled receptor (GPCR) 257–258, 258F, 349
- Gradient methods, parameter estimation 142, 145–146, 147F
- Granger causality 62
- Graphs
 adjacency matrices 55, 55F
 bidirectional 53, 55F
 bipartite 54
 clique 55, 60
 clustering coefficient 55–56, 56F
 defined 53
 degree distribution 57, 58F
 degrees 55, 55F
 directed acyclic 67–68
 directed vs. undirected 53, 53F, 55–56, 55F, 56F, 62
 edges 53, 55
 hubs 56, 57F, 60
 interaction 53–61
 nodes 53, 55, 55F, 56B
 properties 54–58, 55F–58F, 56B
 random 57–58
 signaling networks as 258, 259–261
 theory 54, 54F
 vertex 53
- Green fluorescent protein (GFP) 194, 222, 223F
 Grid search, comprehensive 143–145, 144F
 GroEL/ES protein complex 205
 Growth curves, children 47, 48F
 Growth dynamics 284
 Growth laws 284–286
 Growth rate 284
 Guanine 171, 172F, 173F
 Guare, John 56
 Gut microbiome 432–433
- H**
- Half-life 94, 379
 HDACs (histone deacetylases) 179F
 HDLs (high-density lipoproteins) 217
- Health
 delineation 377, 377F
 predictive 372–378
- Heart 331–367
 aging effects 355–356
 anatomy 331, 332F, 333–334, 333F
 electrochemistry in cardiomyocytes 345–354
 biophysical description 347–348
 models of action potentials 350–354, 353F
 repeated heartbeats 354, 354F
 resting potentials and action potentials 348–350, 349F, 350F
 failing 355–361
 modeling based on molecular events 356–361, 356F, 357F, 359F, 360F
 hierarchy of scales and modeling approaches 332–339
 basics of anatomy 333–334, 333F
 at cell level 337–339, 338F, 339F
 at organ level 334–335, 335F
 at tissue level 335–337
 outlook for physiological multiscale modeling 361–362
 overview 331–332, 332F
 pacemaker 342, 342F, 354–355
 simple models of oscillations 339–345, 339F
 black box 339–343, 340F–342F
 meaningful 343–345, 344F, 345F
- Heartbeat(s)
 repeated 354, 354F
 simulation of perturbed 341–342, 341F
- Heart block 355
- Heart failure 355–361
 modeling based on molecular events 356–361, 356F, 357F, 359F, 360F
- Heat-shock proteins (HSPs) 205, 306–307
- Heat stress
 analysis 312–314, 313F, 313T, 314F
 glucose depletion 314–315, 315F
- Heat stress response 303–330
 brief review 306–310, 307F–309F
 gene expression 308, 309F, 315–318, 316F, 316T, 317F
 modeling analysis 310–318
 analysis of heat stress 312–314, 313F, 313T, 314F
 design and diagnosis of metabolic pathway model 310–312, 312F
 gene expression 315–318, 316F, 316T, 317F
 glucose dynamics 314–315, 315F
 multiscale analysis 318–326
 model design 320–324, 323T, 324F, 324T
 trehalose puzzle 215F–327F, 324–326
in vivo NMR profiles 318–320, 319F–321F
 origin of models 304–305
- Heme 212F
 Hemoglobin 212, 212F
 Henri-Michaelis-Menten rate law *see* Michaelis-Menten rate law
- Hepatocytes 362
 Heraclitus of Ephesus 52
 Hershko, Avram 205
 Heterogeneity, medicine 369–370
 modeling 370–372, 371F, 371T, 372F
- Heterogeneous information merged into conceptual model 15
- Hexokinase 231, 232F
 trehalose cycle 308–309
- Hexose 321F, 411, 412, 414
 Hexose transporters (HXTs) 308
- HGPRT (hypoxanthine-guanine phosphoribosyl-transferase) 392
- High-density lipoproteins (HDLs) 217
- High-performance liquid chromatography (HPLC) 247
- High-throughput methods, of molecular biology 11, 12F
- Hill-climbing methods 142, 145–146, 147F
- Hill function
 heart 344
 metabolic systems 240
 parameter estimation 36, 37, 37F, 160
 signaling systems 262F, 265B
- Hill rate law (HRL) 238, 238F
 pharmacokinetics 387
 trehalose cycle 311
- His, Wilhelm 336
- Histidine (His) 203F, 204T
 Histidine kinase 267, 268
 Histone(s) 178, 179F
 Histone deacetylases (HDACs) 179F
 Hits, small-molecule 379
 Hodgkin, Alan Lloyd 332, 350, 353
 Hodgkin-Huxley model 350, 353, 353F, 354, 354F
- Homeostasis 373
- Homogeneous population 289
- Hormones 214, 217
- HPLC (high-performance liquid chromatography) 247
- HRL *see* Hill rate law (HRL)
- HSPs (heat-shock proteins) 205, 306–307
- Hubs 56, 57F, 60
- Human genome sequencing 6
 Human Protein Atlas 201
- Hunter, Peter 332
- Huxley, Andrew 332, 338, 350, 353
 HXTs (hexose transporters) 308
- Hybridization 171
- Hyperuricemia 391–393
- Hypotheses 15
- Hypoxanthine-guanine phosphoribosyltransferase (HGPRT) 392
- Hysteresis 263–266, 264F, 264T, 265B–266B
- I**
- Identifiability 160
- Ile (isoleucine) 203F, 204T
- Immune cascade 404–406, 404F, 406F
- Immune system(s)
 innate and adaptive 428–429, 429F
 proteins 215, 215F, 216F
- Immunoglobulins 215, 215F
- Immunoprecipitation 225
- Independent variables 33–34, 33F
 trehalose cycle 312, 313, 313F
- Inducers 185–186
- Inducible promoters 409
- Infection 428–429, 429F
- Infectious disease modeling 23
 asymptomatics 43–45, 44F, 45F
 canonical 102, 103
 consistency and robustness 39F, 40
 diagram 35F
 model selection 26
 model structure 27
 parameter estimation 37
- Inflammation 428–432, 429F, 430F
- Information
 coarse vs fine-grained 437
 mutual 62–63, 63F
- Information theory 62
- Inhibition
 allosteric 211, 239
 competitive, uncompetitive, and noncompetitive 239
- Inhibitors, metabolic systems 238–240
- Initial value 94
- Innate immune system 428–429, 429F
- Inputs 8–9, 8F, 9F
 model development 24
 persistent changes 121
- Insulin
 crystal structure 208F
 design of biological systems 399, 408
 diabetes modeling 428
 drug development 381
 metabolic network analysis 251
 model design 29, 31
- Integrative analysis, genome, protein, and metabolite data 303–330
- Interacting populations 292–299
 general modeling strategy 292, 292F
 more complex models 297–299, 298F, 298T
 phase-plane analysis 292–297, 293F–297F, 293T
- Interaction(s) 16

- Interaction graphs 53–61
 edges 53, 55
 nodes 53, 55, 55F, 56B
 properties 54–58, 55F–58F, 56B
 theory 54, 54F
 vertex 53
- Interaction networks, Bayesian reconstruction 63–69, 65F, 67F
- Introns 177, 177F
- Inverse parameter estimation 439
- Inversion recombination elements 410
- Ion(s) 332, 345
- Ion channels
 heart 346, 348, 350, 353, 361
 protein systems 221, 221F
- Ion transport 211, 337, 346–347
- Irreversible step, trehalose cycle 308
- Isobaric tags for relative and absolute quantitation (iTRAQ) technique 219
- Isoleucine (Ile) 203F, 204T
- Isoprenoid pathway 407, 414
- Isozymes
 control of metabolic pathways 246
 trehalose cycle 308, 317
- Iterations 46
- iTRAQ (isobaric tags for relative and absolute quantitation) technique 219
- J**
- Jackknife method, parameter estimation 152–153
- Jacob, François 186
- Jacobian of system 98–100, 99F, 100F
- Julia sets 93
- “Junk DNA” 176–177
- c-Jun N-terminal kinase (JNK) 270
- K**
- Kai system 277, 277F
- Kauffman, Stuart 260
- Keasling, Jay 414
- KEGG database
 metabolic network reconstruction 74
 metabolic systems 241F, 242, 243B, 250, 251
 trehalose cycle 308, 311
- Kekulé von Stradonitz, Friedrich August 304
- Keratin 216–217
- Kinases 214
- Kinetic models 160, 251, 277
- Kinetic order 103
 trehalose cycle 311
- KInfer 242
- Kirchhoff’s laws 351
- Kitano, Hiroaki 440
- Kyoto Encyclopedia of Genes and Genomes *see* KEGG database
- L**
- lacI* regulator gene 186, 186F
- lac* operon 186–187, 186F, 187T
- Lactococcus lactis* 252
- Lactose 185–187
- Laplace transform 94
- Large-scale model assessments 45–46, 47F
- Latin hypercube technique 121, 144
- Latin square technique 121
- Laws, biological 29, 46, 440
- LCCs (L-type Ca²⁺ channels) 349, 356F, 357–361, 357F
- LDLs (low-density lipoproteins) 217
- Lead(s) 379
- Lead identification, computational 381
- Lead optimization 379
- Least-squares method, parameter estimation 135
- Left ventricular hypertrophy 356
- Lens 217, 218F
- Leslie model, population dynamics 289–292, 291B
- Leucine (Leu) 203F, 204T
- Leukemia, acute lymphoblastic 373–374
- Levinthal’s Paradox 204–205
- Lewin, Kurt 440
- Life cycle 16
- Ligand binding 384
- Limit-cycle oscillations 102
- Limit cycles 123–125, 123F, 125F, 126F
- Limiting distribution 90
- Linear chains 8, 8F
- Linear differential equations 94
- Linearization 95
- Linearized models 95–100
 approximation 95–98, 96B, 97B, 97F
 curved surface 95, 95F
 Jacobian 98–100, 99F, 100F
 operating point 95
 perturbation 95
- Linear models 84
- Linear regression 27, 27F
 misuse 137–138, 137F
 parameter estimation 136–141
 several variables 138–141, 138F–141F, 139T–141T
 single variable 136–138, 137F
- Linear relationship, analysis of heat stress 312
- Linear systems 84
 continuous 93–100
 linear differential equations 94
 linearized models 95–100, 95F, 96B, 97B, 97F, 99F, 100F
 discrete 85–91, 85T, 86F
 recursive deterministic models 85–88, 85T, 86F
 recursive stochastic models 84F, 88–91
 parameter estimation 136–141
 several variables 138–141, 138F–141F, 139T–141T
 single variable 136–138, 137F
- Lin-log model 107, 115
- Lipid Maps 242
- Lipo-oligosaccharides 209F
- Lipoproteins 217
- Liver
 compartment model 385–386, 386F, 387F
 drug development 379, 385–387, 388B, 390
 enzyme activity profile 52
 lipoproteins 217
 multiscale modeling 362
 Virtual Liver Network 333, 356, 362
- Local minimum 146
- Local motifs 221
- Logistic function, graph 101, 101F
- Logistic growth function 283, 284
- Logistic map 91–93, 92F
- Lorenz, Edward 126
- Lorenz attractor 126–127, 127F
- Lotka, Alfred 296
- Lotka-Volterra (LV) model 101, 102–103, 103F, 114
 interacting populations 296–298, 298F, 298T
- Low-density lipoproteins (LDLs) 217
- Low-diameter networks 56B
- L-type calcium (Ca²⁺) channels (LCCs) 349, 356F, 357–361, 357F
- Lung(s)
 and heart 333, 334, 362
 infection and inflammation 428, 431
 pharmacokinetic modeling 385, 389, 390
 pneumonia 285
- Lung cancer 372
- Lung Physiome 333
- Lwoff, André Michel 186
- Lymphocytes, suppressor 404, 405
- Lysine (Lys) 203F, 204T, 251
- M**
- Machine learning 12, 148
- Malaria 414–415
- MALDI (matrix-assisted laser desorption/ionization) 222, 223F
- MALDI-TOF (matrix-assisted laser desorption/ionization-time of flight) 220
- Malthus, Thomas Robert 283
- Mandelbrot sets 93
- Map(s)
 concept 437, 438–439, 439F
 logistic 91–93, 92F
- metabolic 74, 74F, 250
 stoichiometric connectivity 242
- MAPK (mitogen-activated protein kinase) cascades 270–273, 270F–273F, 271T
- Markov chain models 84F, 88–91
- Markov chain Monte Carlo (MCMC) algorithms 68
- Markov models 84F, 88–91
- Mark-up languages 12
- Mass action kinetics 235
 heart function 359–360, 360F
- Mass-action models 103, 268
see also Generalized Mass Action systems
- Mass action systems, generalized 103–104, 104F
 structure identification 160
 trehalose cycle 311
- Mass spectrometry (MS) 220, 220F, 223F
 metabolite analysis 247–248, 248F
- Mathematica 435
- Mathematical model(s) 19–50
 analysis 37–42
 approximations 29, 29F
 binomial 30, 32B
 blue sky catastrophe 30, 31B
 computer simulations 19, 38, 38F
 consistency and robustness 38–40, 39F
 correlative vs explanatory 27, 27F
 data availability 25–26
 defined 19
 deterministic 29, 30, 31B
 diagnostics 37–40, 39F
 diagram 34–35, 34F, 35F
 discrete vs continuous 29
 equations 35–36
 exploration and validation of dynamical features 38, 40–42, 41B, 42F, 43F
 extensions and refinements 43–45, 44F, 45F
 external validation 40
 flow chart of process 22, 22F
 flow of information 32–33, 33F
 flow of material 32, 32F
 goals, inputs, and initial exploration 24–26, 24F
 issues to address 22–23, 23T
 large-scale model assessments 45–46, 47F
 open vs closed 29
 overview of 19–24, 20F, 22F, 23T
 overfitting 40, 160
 parameters 31–32, 34, 36–37, 37F
 Poisson 30, 33B
 probabilistic or stochastic 29–30, 32B, 33B
 processes and interactions 31
 questions of design 46–47
 questions of scale 24–25, 24F
 selection and design 26–37, 26F
 sensitivity 39–40, 39F
 simplicity vs complexity 47–49, 48F
 spatial aspects 29
 static vs dynamic 27–29, 28B
 steady state 38–39
 structure 27–30
 system components 30–35
 use and applications 43–49
 variables 30–31, 33–34, 33F
- Mathematics of biological systems 83–134
 continuous linear 93–100
 linear differential equations 94
 linearized models 95–100, 95F, 96B, 97B, 97F, 99F, 100F
 continuous nonlinear 100–110
 ad hoc models 101–102, 101F, 102F
 canonical models 101, 102–110, 103F–105F, 107F, 108B, 109B
 more complicated dynamical systems
 descriptions 110
 discrete linear 85–91
 recursive deterministic models 85–88, 85T, 86F
 recursive stochastic models 84F, 88–91
 discrete nonlinear 91–93, 92F
 other attractors 122–128
 chaotic 126–128, 127F, 128F
 limit cycles 123–125, 123F, 125F, 126F
 overview 83–85
 standard analyses 110–122
 analysis of systems dynamics 119–122, 120F, 122F
 parameter sensitivity 118–119
 stability analysis 115–117, 116F, 117F
 steady-state analysis 110–115, 112F, 113F

- Mating process, genetic algorithms 142–143, 147
 MATLAB 435
 Matrices 84
 Matrix-assisted laser desorption/ionization (MALDI) 222, 223F
 Matrix-assisted laser desorption/ionization-time of flight (MALDI-TOF) 220
 Matrix equation, population dynamics 289–290
 Matrix proteins 208
 MBDs (methyl-CG-binding domains) 179F
 MCA (metabolic control analysis) 74–78, 75F, 77F, 107
 metabolic pathways 242
 MCMC (Markov chain Monte Carlo) algorithms 68
 Medicine 369–378
 biological variability and disease 369–370
 modeling 370–372, 371F, 371T, 372F
 personalized (precision) 372–378
 data needs and biomarkers 373–374, 374F
 mathematical models 374–378, 377F
 traditional vs. 374, 375F
 predictive health 372–378
 Membrane-bound receptors, drug development 382, 382F
 Membrane potential 346–349
 Mendel, Gregor 176, 440
 Mendelian genetics 440
 Mesarović, Mihajlo 10
 Messenger proteins 214, 214F
 Messenger RNA (mRNA) 179
 Metabolic analysis
 flux 249
 incompletely characterized organism 250–251
 network 251
 resources 241–244, 243B, 244B
 Metabolic burden 408
 Metabolic control analysis (MCA) 74–78, 75F, 77F, 107
 lin-log model 107, 115
 metabolic pathways 242
 Metabolic engineering 407–408, 411–414, 413F
 elementary mode analysis 411–414, 413F
 Metabolic flux analysis 249
 Metabolic map 74, 74F, 250
 Metabolic network(s) 52, 69–78
 Bayesian methods 69
 metabolic control analysis 74–78, 75F, 77F
 reconstruction 73–74, 74F
 stoichiometric 70–73, 70F–72F
 variants of stoichiometric analysis 73
 Metabolic network analysis 251
 Metabolic pathway(s), structural patterns 401
 Metabolic pathway model, design and diagnosis 310–312, 312F
 Metabolic profile 240
 Metabolic reactions
 mathematical formulation 234–235
 rate laws 235–240, 235F, 237F, 238F
 Metabolic syndrome 428
 Metabolic systems 231–255
 biochemical reactions 232–240
 background 232–234, 233B
 mathematical formulation of elementary reactions 234–235
 rate laws 235–240, 235F, 237F, 238F
 data generation methods 246–249
 detection methods 247–249, 248F
 flux analysis 249
 sampling, extraction, and separation methods 247, 251–252
 from data to systems models 250–252
 extraction of dynamic models from experimental data 251–252
 incompletely characterized organism 250–251
 metabolic network analysis 251
 metabolites 231–232, 232F
 overview 231–232
 pathways and pathway systems 240–246
 biochemistry and metabolomics 240–241, 241B
 control 244–246, 245F, 246F
 resources for computational analysis 241–244, 243B, 244B
 Metabolism, defined 232
 Metabolites 231–232, 232F
 integrative analysis 303–330
 Metabolomics 232, 240–241
 data generation methods 246–249, 248F, 249F
 MetaCyc database 74
 Metadata 433
 Metagenomics 172, 433
 Metapopulations 286, 432–433, 433F
 Methicillin-resistant *Staphylococcus aureus* (MRSA) 170
 Methionine (Met) 203F, 204T
 Method of mathematically controlled comparison (MMCC) 404–406, 406F
 Methotrexate (MTX) 388B–389B
 Methyl-CG-binding domains (MBDs) 179F
 Methyl groups 178
 Metropolis-Hastings method 68
 Mevalonate pathway 414, 415D
 MI (myocardial infarction) 339
 Michaelis-Menten function
 ad hoc 102
 MAPK cascade 270F, 271
 parameter estimation 36, 139, 140
 PBPK models 390
 power-law approximations 109B
 Michaelis-Menten process
 biological variability and disease 370–372, 371F, 371T, 372F
 mathematical model 235–236
 pharmacokinetics 387
 power-law approximations 109B
 Michaelis-Menten rate law (MMRL)
 metabolic systems 236–238, 237F, 238F
 parameter estimation
 linear systems 139–140, 139F–141F, 140T, 141T
 nonlinear systems 143–144, 144F
 trehalose cycle 311
 Microarrays 193–194, 193F
 drug sensitivity 374
 heat stress response 315
 Microbial biofilms 433, 433F
 Microbial metapopulations 432–433
 Microbial systems 399, 418, 419
 Microbiome 432–433
 Micro-RNA (miRNA) 177, 183, 184, 409
 Middle-out analyses 17
 Mill, John Stuart 62
 miRNA (micro-RNA) 177, 183, 184, 409
 Mitochondria 1
 Mitochondrial oxidative phosphorylation 77
 Mitogen-activated protein kinase (MAPK) cascades 270–273, 270F–273F, 271T
 MMCC (method of mathematically controlled comparison) 404–406, 406F
 MMRL *see* Michaelis-Menten rate law (MMRL)
 Model(s)
 dynamic(al) *see* Dynamic(al) models
 mathematical *see* Mathematical model(s)
 origin of 304–305
 static network *see* Static network models
 Model extensions 43–45, 44F, 45F
 Modeling
 defined 19
 needs 435–439
 Modeling approach 16
 Modeling tools 16–17
 Modules 20
 Modulons 188
 Molecular biology
 DNA sequences 178
 high-throughput methods of 11, 12F
 Molecular chaperones 205, 207F
 Molecular dynamics 222
 Molecular function, gene products 176
 Molecular inventories 16, 17
 Molecular systems 53, 376
 Monoclonal antibody 382F–385F, 383–385, 384T
 Monod, Jacques 186
 Monte Carlo simulation 45–46, 47F
 personalized medicine 376
 Mori, Hirota 11
 Motifs, network 53
 design of biological systems 399, 402–404, 402F, 403F
 mRNA (messenger RNA) 179
 MRSA (methicillin-resistant *Staphylococcus aureus*) 170
 MS (mass spectrometry) 220, 220F, 223F
 metabolite analysis 247–248, 248F
 MTX (methotrexate) 388B–389B
 Mucins 217
 Multidimensional protein identification technology (MudPIT) 219
 Multidisciplinary team 15
 Multiple linear regression, parameter estimation 138–141, 138F–141F, 139T–141T
 Multiple organ failure 431
 Multiplex automated genome engineering 407
 Multiscale analysis, heat stress response 318–326
 model design 320–324, 323T, 324F, 324T
 trehalose puzzle 215F–327F, 324–326
in vivo NMR profiles 318–320, 319F–321F
 Multiscale modeling 25, 361–362, 436–437
 Multistage cancer model 432
 Multivariate functions, power-law approximations 109B
 Multivariate systems 31
 Muscle cells 211, 217, 335, 338
 Mutations
 analysis of system dynamics 121
 biologic variability 370
 cyanobacteria 434
 drug development 415
 evolutionary programming 148
 genetic algorithms 146, 147, 147F, 148
 heat stress response 312
 immune system 215
 metabolic engineering 407
 metabolic systems 231
 probability 32B, 33B
 random 30
 rate 33B
 sensitivity analysis 118
 sickle cell disease 104, 372
 signaling systems 277
 simulated annealing 149
 small-world networks 59–60
 transposons 170–171
 vectors 142–143
 Mutual information 62–63, 63F
Mycoplasma pneumoniae 250
 Myocardial cells *see* Cardiomyocytes
 Myocardial infarction (MI) 339
 Myocardium 335
 Myocytes, heart *see* Cardiomyocytes
 Myofibrils 338
 Myofilaments 338, 338F
 Myosin 338, 338F
- ## N
- Na⁺-Ca²⁺ (sodium-calcium) exchanger 347
 Nagumo, Jin-Ichi 332
 Na⁺-K⁺ ATPase (sodium-potassium adenosine triphosphatase) 347
 Na⁺-K⁺ (sodium-potassium) pump 346–347
 Nanomedicine 11, 13F
 NCEs (new chemical entities) 378–379
 Negatively inducible operons 187
 Negatively repressible operons 187
 Nernst equation 347–348
 Nested feedback 245, 246F
 Network(s) 51
 Bayesian reconstruction 63–69, 65F, 67F
 flux 70–73, 70F–72F
 gene regulatory 69
 low-diameter 56B
 metabolic 52, 69–78
 Bayesian methods 69
 metabolic control analysis 74–78, 75F, 77F
 reconstruction 73–74, 74F
 stoichiometric 70–73, 70F–72F
 variants of stoichiometric analysis 73
 motifs 53, 402–404, 402F
 neural 426–427, 426F, 427F
 protein-protein interaction (PPI) 53, 60, 69
 random 56B
 regular 56B
 regulated 54
 scale-free 57–58, 58F
 signaling, Bayesian analysis 66–69, 67F
 small-world 58–60, 59B, 61F
 biological variability and disease 372
 static 51–82
 Bayesian reconstruction 63–69

- dependencies among components 62–63, 63F
 - interaction graphs 53–60
 - metabolic 52, 69–78
 - small-world 58–60, 59B, 61F
 - stoichiometric 70–73, 70F–72F
 - strategies of analysis 52–53
 - stoichiometric 70–73, 70F–72F
 - types of generic 51–52, 51F
 - undirected 62
 - Network interference 261
 - Network motifs, design of biological systems 399, 402–404, 402F, 403F
 - Neural networks 426–427, 426F, 427F
 - Neurobiology 426–427, 426F, 427F
 - Neuron(s)
 - action potential 11, 350
 - multiscale modeling 362, 436
 - neural networks 426–427, 427F
 - signaling systems 261
 - Neuronal signal transduction 426
 - Neurotransmitters 426
 - New chemical entities (NCEs) 378–379
 - NMR spectroscopy *see* Nuclear magnetic resonance (NMR) spectroscopy
 - Noble, Denis 332
 - Nodes 53
 - degrees associated with 54–55, 55F
 - number of paths between 55, 56B
 - Noise, parameter estimation 149, 150F, 156–157, 156F, 157F, 157T
 - Noisy channels, reliability of communication through 62–63
 - Noncoding DNA 174, 176–178, 177F
 - Noncoding RNA 409
 - Noncompetitive inhibition 239
 - Nonlinear models 84
 - Nonlinear partial differential equation, heart 334
 - Nonlinear processes 5, 5F, 8
 - Nonlinear regression 27, 27F
 - parameter estimation 142, 145–146, 145F
 - Nonlinear systems 84
 - continuous 100–110
 - ad hoc models 101–102, 101F, 102F
 - canonical models 101, 102–110, 103F–105F, 107F, 108B, 109B
 - more complicated dynamical systems descriptions 110
 - discrete 91–93, 92F
 - parameter estimation 141–153
 - comprehensive grid search 142, 143–145, 144F
 - genetic algorithms 142–143, 146–148, 147F
 - nonlinear regression 142, 145–146, 145F
 - other stochastic algorithms 148–149
 - search algorithms 142–143, 143F
 - typical challenges 149–153, 150F–152F, 150T, 152T
 - Noradrenaline 336
 - Northern blot 192, 192F
 - NTH1/2* gene 310, 325
 - Nuclear magnetic resonance (NMR) spectroscopy 206, 206F
 - heat stress response 318–320, 319F–321F
 - metabolite analysis 247, 248–249, 248F, 249F
 - Nucleo-bases, DNA and RNA 171, 172F, 173F
 - Nucleosides 171, 172F
 - Nucleosome 178, 179F, 180F
 - Nucleotides 171, 172F
 - Nullcline, interacting populations 294–295, 294F, 295F
 - Numerical mathematics 12
- O**
- OBDs (origin-binding domains) 173F
 - Objective function, parameter estimation 135
 - Oceans 433–434, 434F, 435F
 - ODEs *see* Ordinary differential equations (ODEs)
 - Ohm's law 351
 - Oligonucleotides 193–194, 381
 - Open reading frame (ORF) 177
 - Open system model 29
 - Operating point (*OP*) 95
 - Operating principles 406–407, 406F
 - see also* Design principles
 - Operons 185
 - lac* 186–187, 186F, 187T
 - regulation 187–188
 - structural patterns 401, 401F
 - über- 188, 401, 401F
 - Optimization 54
 - constrained 73
 - design of biological systems 400
 - lead 379
 - metabolic engineering 407
 - parameter estimation 135
 - Orbit 123
 - Ordinary differential equations (ODEs) 91, 110
 - heart 358
 - metabolic systems 234
 - models of gene regulation 190
 - population growth 284, 285–286
 - signaling systems modeled with 258, 261–278
 - ORF (open reading frame) 177
 - Organism interactions with environment 432–434, 433F–435F
 - OR gate 402
 - Origin-binding domains (OBDs) 173F
 - Oscillation patterns
 - analysis of system dynamics 119, 121
 - chaotic 128
 - circadian clocks 275–278, 276F, 277F
 - damped 9, 9F, 103, 155
 - gene circuits 417, 418F
 - heart 332, 336, 338, 339, 339F, 345, 345F, 347
 - limit-cycle 102, 123–125, 123F, 125F, 126F, 438
 - parameter sensitivity 118
 - periodicity 93
 - predator-prey 296, 296F
 - sine-cosine 93
 - stable, S-system 340, 340F
 - unpredictable deterministic model 31B
 - Oscillations models 339–345, 339F
 - black box 339–343, 340F–342F
 - meaningful 343–345, 344F, 345F
 - repeated heartbeats 354
 - Overfitting, structure identification 160
 - Oxidative stress 355
- P**
- p53 protein 27
 - Pacemaker, effectiveness of simple 342, 342F
 - Pacemaker activity 336
 - Pacemaker cells 337
 - Pacemaker system abnormalities 355
 - Parameter(s)
 - defined 31–32
 - model development 34
 - population 284
 - sensitivity analysis 118–119
 - Parameter estimation 135–167
 - bottom-up strategy 136
 - identifiability 160
 - branch-and-bound method 144–145
 - evolutionary algorithms 142–143, 146–148, 147F
 - linear systems 136–141
 - several variables 138–141, 138F–141F, 139T–141T
 - single variable 136–138, 137F
 - in model selection and design 36–37, 37F
 - noise 149, 150F, 156–159, 156F–159F, 157T, 158T
 - nonlinear systems 141–153
 - comprehensive grid search 142, 143–145, 144F
 - genetic algorithms 142–143, 146–148, 147F
 - nonlinear regression 145–146, 145F
 - other stochastic algorithms 148–149
 - search algorithms 142–143, 143F
 - typical challenges 149–153, 150F–152F, 150T, 152T
 - overview 135–136
 - simulated annealing 148
 - sloppiness 151
 - SSE (sum of squared errors) 135, 137
 - structure identification 160
 - systems of differential equations 153–159, 153F, 154T, 156F–159F, 156T–158T
 - time-series-based 136, 153
 - uncertainty 151–152
 - Partial differential equations (PDEs) 110
 - heart 334, 358
 - models of gene regulation 191
 - Particle swarm optimization (PSO) 148
 - Path lengths 59
 - Pathogen(s)
 - antibodies 215
 - immune system 428, 429, 430, 431
 - lipo-oligosaccharides 209F
 - quorum sensing 278, 418
 - Toll-like receptors 216F
 - Pathogenicity 418
 - Pathway screening, dynamic models 390–393, 391F
 - Pathway Studio 242
 - Pathway systems
 - gene expression 317
 - generic 33F
 - metabolic 232, 240–246
 - analysis 52, 53
 - design 407, 408, 411–412, 413F
 - genome size 174
 - reconstruction 74
 - screening with dynamic models 390–393
 - simple 8
 - Pavlov, Ivan Petrovich 440
 - PBD (platform-based design) 410
 - PBPK (physiologically based pharmacokinetic) models 385, 387–390, 387F, 388B–389B
 - PCR (polymerase chain reaction), reverse transcriptase 192–193, 192F
 - PDEs *see* Partial differential equations (PDEs)
 - PDIs (protein disulfide isomerases) 205
 - Pentose(s) 411, 412
 - Pentose phosphate pathway (PPP) 309, 316F
 - Peptide(s) 204
 - Peptide bonds 202, 202F
 - Period doubling 93
 - Permeases 346–347
 - Persistent changes in structure or input 121
 - Personalized medicine 372–378
 - data needs and biomarkers 373–374, 374F
 - mathematical models 374–378, 377F
 - traditional vs. 374, 375F
 - Perturbation(s)
 - analyses of biological systems models 110
 - analysis of systems dynamics 121
 - Bayesian reconstruction of interaction networks 63
 - bistability and hysteresis 262
 - data-modeling pipeline 439
 - design of biological systems 405, 416, 417
 - drug development 391, 392
 - heart 339, 345, 360
 - heat stress response 312
 - inflammation 429
 - limit cycles 123, 124, 125F
 - linearized models 95, 98
 - metabolic control analysis 74, 75, 76
 - modeling variability and disease 371–372
 - other biological networks 69
 - parameter sensitivity 118, 119
 - phase-plane analysis 296, 297
 - population dynamics under external 288–289, 288F, 289F
 - recovery from 38
 - signaling networks 69, 277, 277F
 - stability analysis 115, 116, 116F, 117
 - static network models 51
 - steady-state analysis 111
 - Peskin, Charles 332
 - Petri net analysis 54
 - PEK1/2* genes 310, 317
 - Pharmacogenomics 373–374
 - Pharmacokinetic modeling 385–390, 386F, 387F, 388B–389B
 - Pharmacokinetics 388B–389B
 - Pharmacokinetic tests 379
 - Phase-plane analysis 292–297, 293F–297F, 293T
 - Phenotype 176
 - Phenylalanine (Phe) 203F, 204T
 - Phosphate-deoxyribose backbone 172, 173F
 - Phosphofructokinase 310, 317
 - Phospholipase A 222F
 - Phosphoribosylpyrophosphate synthase (PRPPS) 392
 - Photosynthesis 24, 25, 73, 208, 434, 436
 - Phylogeny 171
 - Physiologically based pharmacokinetic (PBPK) models 385, 387–390, 387F, 388B–389B