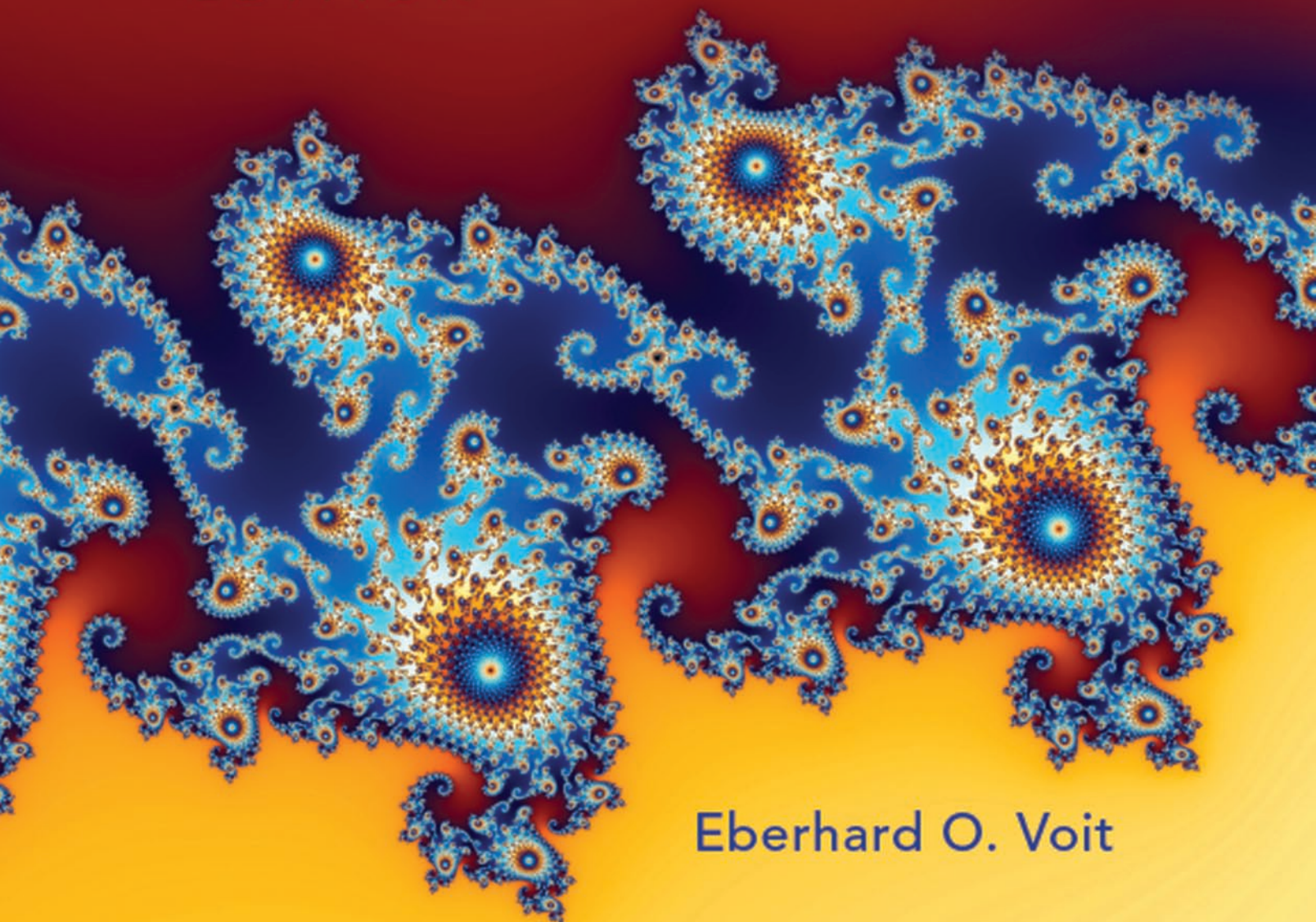


A FIRST COURSE IN
SYSTEMS
BIOLOGY

SECOND
EDITION



Eberhard O. Voit

A FIRST COURSE IN
SYSTEMS
BIOLOGY

SECOND
EDITION

*To Ann,
Still the Hub of my Support System*

A FIRST COURSE IN
SYSTEMS
BIOLOGY

SECOND
EDITION

Eberhard O. Voit

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://lccn.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>

Preface

Hard to believe, but it is already time for the second edition! I am happy to report that the first edition of *A First Course in Systems Biology* has met with great success. The book has been a required or recommended text for over 70 courses worldwide, and it has even been translated into Korean. So why should a new edition be necessary after only five short years? Well, much has happened. Systems biology has come out of the shadows with gusto. Research is flourishing worldwide, quite a few new journals have been launched, and many institutions now offer courses in the field.

While the landscape of systems biology has evolved rapidly, the fundamental topics covered by the first edition are as important as they were five years ago and probably will be several decades from now. Thus, I decided to retain the structure of the first edition but have rearranged some items and added a few topics, along with new examples. At Georgia Tech we have used the book to teach well over 1000 students, mostly at the undergraduate level, but also for an introductory graduate-level course. Most of the additions and amendments to this new edition respond to feedback from these students and their instructors, who have pointed out aspects of the material where more or better explanations and illustrations would be helpful. New topics in this edition include: default modules for model design, limit cycles and chaos, parameter estimation in Excel, model representations of gene regulation through transcription factors, derivation of the Michaelis-Menten rate law from the original conceptual model, different types of inhibition, hysteresis, a model of differentiation, system adaptation to persistent signals, nonlinear nullclines, PBPK models, and elementary modes.

I would like to thank three undergraduates from my classes who helped me with the development of some of the new examples, namely Carla Kumbale, Kavya Muddukumar, and Gautam Rangavajla. Quite a few other students have helped me with the creation of new practice exercises, many of which are available on the book's support website. I also want to express my gratitude to David Borrowdale, Katie Laurentiev, Georgina Lucas, Denise Schanck, and Summers Scholl at Garland Science for shepherding this second edition through the review and production process.

It is my hope that this new edition retains the appeal of the original and has become even better through the alterations and twists it has taken, large and small.

Eberhard Voit
Georgia Tech
2017

Instructor Resources Website

The images from *A First Course in Systems Biology, Second Edition* are available on the Instructor Site in two convenient formats: PowerPoint® and JPEG. They have been optimized for display on a computer. Solutions to end-of-chapter exercises are also available. The resources may be browsed by individual chapters and there is a search engine. Figures are searchable by figure number, figure name, or by keywords used in the figure legend from the book.

Accessible from www.garlandscience.com, the Instructor's Resource Site requires registration and access is available only to qualified instructors. To access the Instructor Resource site, please email science@garland.com.

Acknowledgments

The author and publisher of *A First Course in Systems Biology, Second Edition* gratefully acknowledge the contributions of the following reviewers in the development of this book:

Guy Grant, University of Bedfordshire

Princess Imoukhuede, University of Illinois at Urbana-Champaign

Dimitrios Morikis, University of California at Riverside

Oliver Schildgen, University of Witten

Manuel Simões, University of Porto

Mark Speck, Chaminade University

Marios Stavridis, Ninewells Hospital & Medical School

Geraint Thomas, University College London

Floyd Wittink, Leiden University

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

8.10 Case Study 1: Analyzing Metabolism in an Incompletely Characterized Organism 250

8.11 Case Study 2: Metabolic Network Analysis 251

8.12 Case Study 3: Extraction of Dynamic Models from Experimental Data 251

Exercises 252

References 254

Further Reading 255

Chapter 9: Signaling Systems 257

Static Models of Signal Transduction Networks 259

9.1 Boolean Networks 259

9.2 Network Inference 261

Signal Transduction Systems Modeled with Differential Equations 261

9.3 Bistability and Hysteresis 261

9.4 Two-Component Signaling Systems 266

9.5 Mitogen-Activated Protein Kinase Cascades 270

9.6 Adaptation 273

9.7 Other Signaling Systems 274

Exercises 278

References 279

Further reading 281

Chapter 10: Population Systems 283

Population Growth 283

10.1 Traditional Models of Population Growth 284

10.2 More Complex Growth Phenomena 286

Population Dynamics Under External Perturbations 288

Analysis of Subpopulations 289

Interacting Populations 292

10.3 General Modeling Strategy 292

10.4 Phase-Plane Analysis 292

10.5 More Complex Models of Population Dynamics 297

Exercises 299

References 301

Further reading 302

Chapter 11: Integrative Analysis of Genome, Protein, and Metabolite Data: A Case Study in Yeast 303

On the Origin of Models 304

A Brief Review of the Heat Stress Response in Yeast 306

11.1 The Trehalose Cycle 308

Modeling Analysis of the Trehalose Cycle 310

11.2 Design and Diagnosis of a Metabolic Pathway Model 310

11.3 Analysis of Heat Stress 312

11.4 Accounting for Glucose Dynamics 314

11.5 Gene Expression 315

Multiscale Analysis 318

11.6 *In Vivo* NMR Profiles 318

11.7 Multiscale Model Design 320

11.8 The Trehalase Puzzle 324

Concluding Comments 327

Exercises 328

References 329

Further reading 330

Chapter 12: Physiological Modeling: The Heart as an Example 331

Hierarchy of Scales and Modeling Approaches 332

12.1 Basics of Heart Anatomy 333

12.2 Modeling Targets at the Organ Level 334

12.3 Modeling Targets at the Tissue Level 335

12.4 Modeling Targets at the Cell Level 337

Simple Models of Oscillations 339

12.5 Black-Box Models of Oscillations 339

12.6 Summary of Black-Box Oscillation Models 342

12.7 From a Black Box to Meaningful Models 343

Electrochemistry in Cardiomyocytes 345

12.8 Biophysical Description of Electrochemical Processes at the Membrane of Cardiomyocytes 347

12.9 Resting Potentials and Action Potentials 348

12.10 Models of Action Potentials 350

12.11 Repeated Heartbeats 354

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

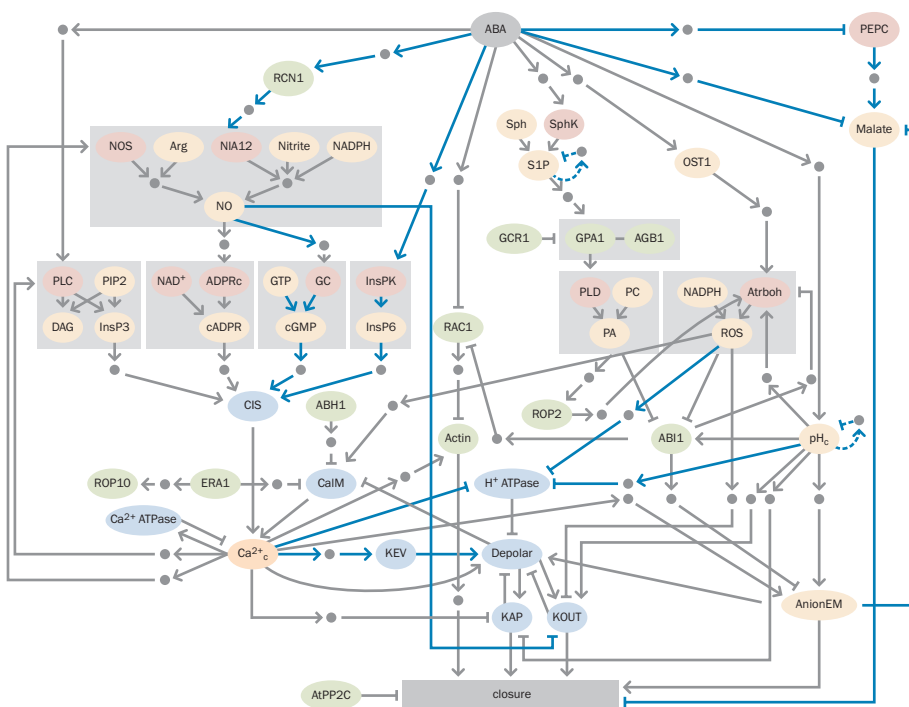


Figure 1.2 Diagram of a complicated system of molecules that coordinate the response of plants to drought. While the details are not important here, we can see that a key hormone, called abscisic acid (ABA), triggers a cascade of reactions that ultimately promote the closure of stomata and thereby reduce 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

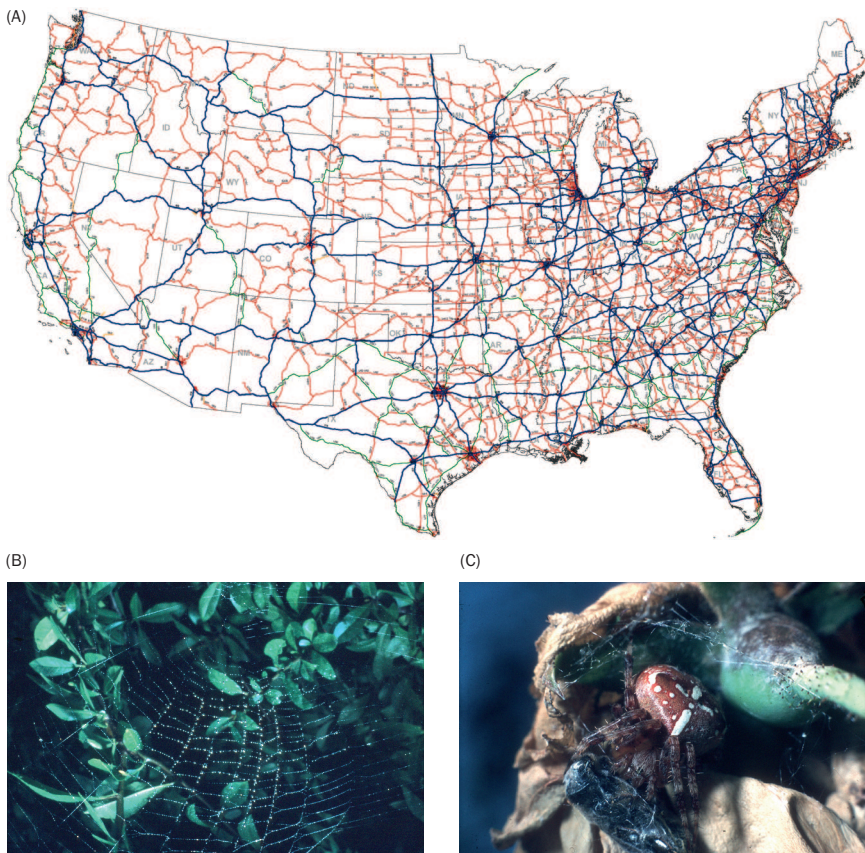


Figure 1.4 The size of a network or system is not necessarily correlated with its complexity. (A) The network of major highways in the continental United States covers over 3 million square miles. Nonetheless, its functionality is easy to grasp, and problems with a particular road are readily ameliorated 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.)

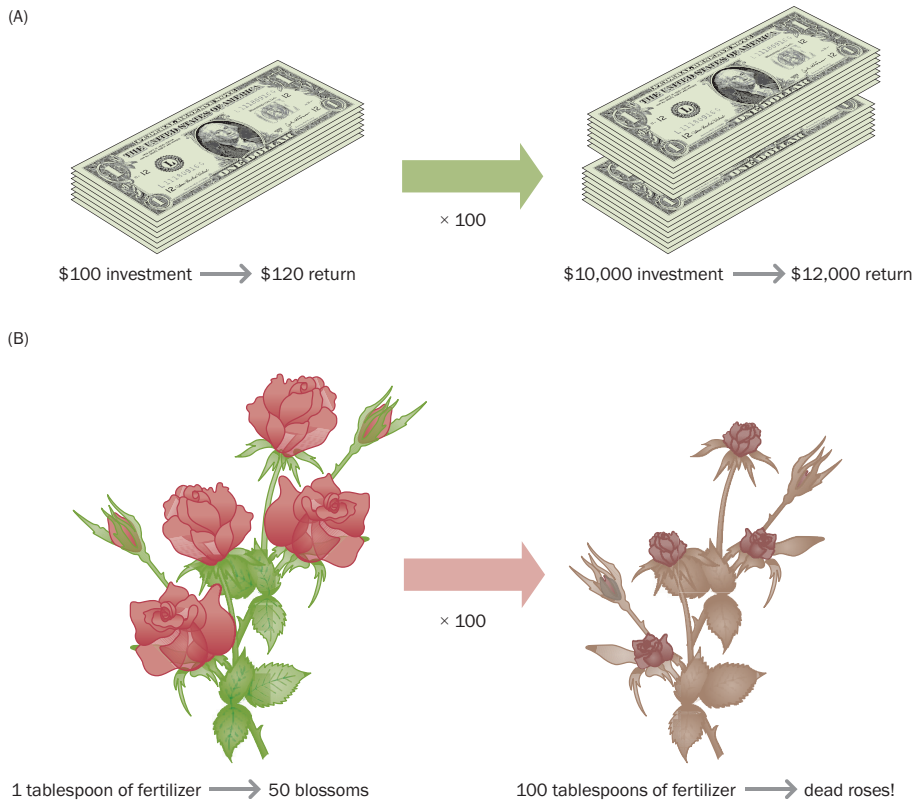


Figure 1.5 Biological phenomena are often difficult to understand, because our minds are trained to think linearly. (A) The return on an investment grows (or decreases) linearly with the amount invested. (B) In biology, more is not necessarily better. Biological responses often scale within a modest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

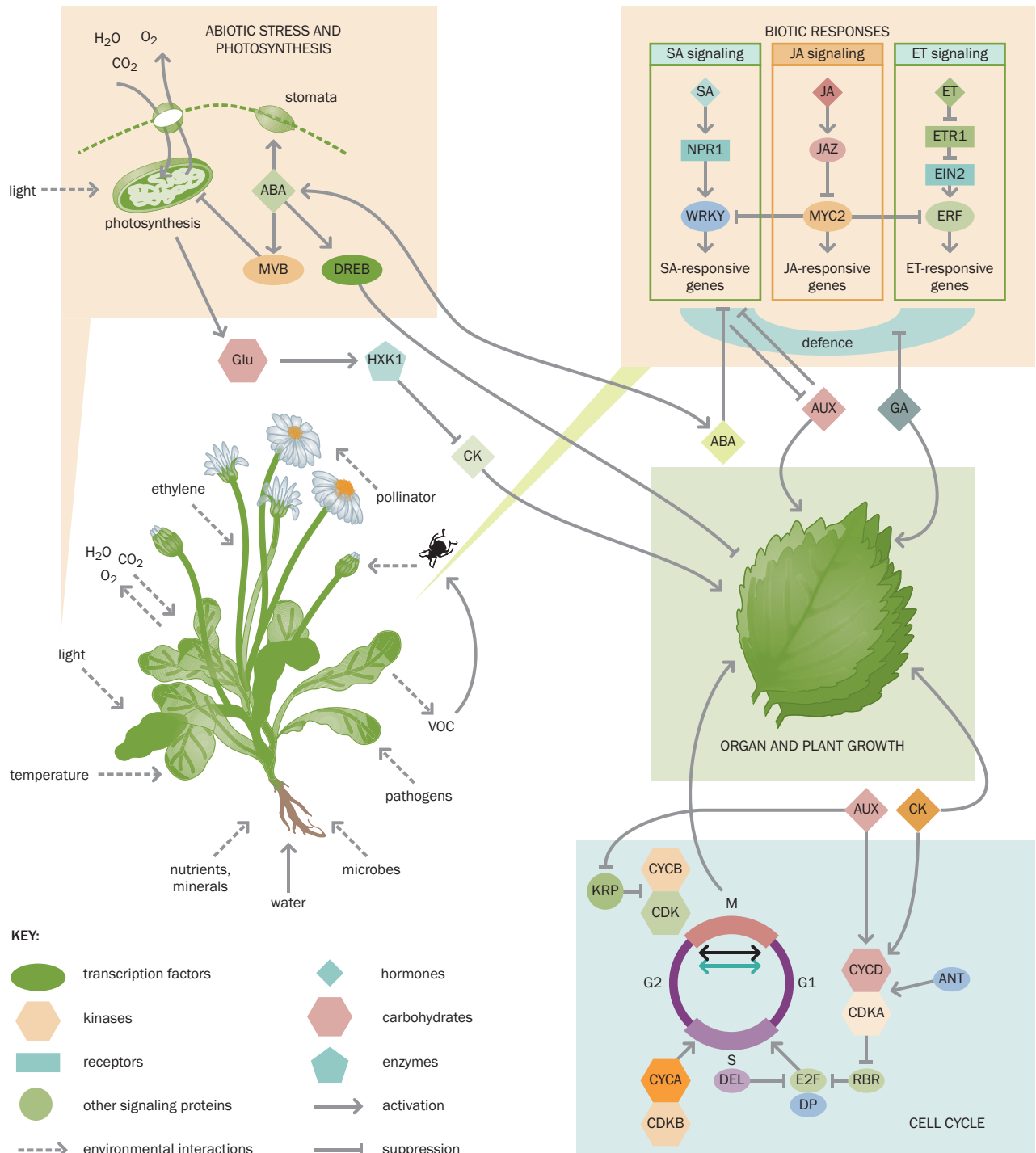


Figure 1.7 Stress responses are coordinated by systems at different levels of organization (cf. Figure 1.2). At the physiological level, the stress response system in plants includes changes at the cellular, organ, and whole-plant levels and also affects interactions of the plant with other species. (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

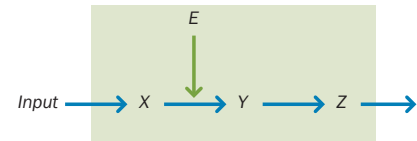


Figure 1.8 The human brain handles linear chains of causes and events very well. In this simple pathway, an external input is converted sequentially into X , Y , and Z , which leaves the system. The conversion of X into Y is catalyzed by an enzyme E . It is easy to imagine that any increase in *Input*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

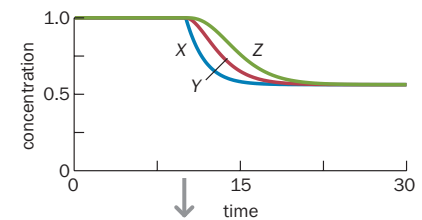


Figure 1.9 Simulations with the system in (1.1) confirm our intuition: X , Y , and Z reflect changes in *Input*. For instance, reducing *Input* in (1.1) to 75% at time 10 (arrow) leads to permanent decreases in X , Y , and Z .

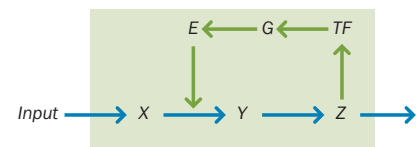


Figure 1.10 Even simple systems may not allow us to make reliable predictions regarding their responses to stimuli.

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

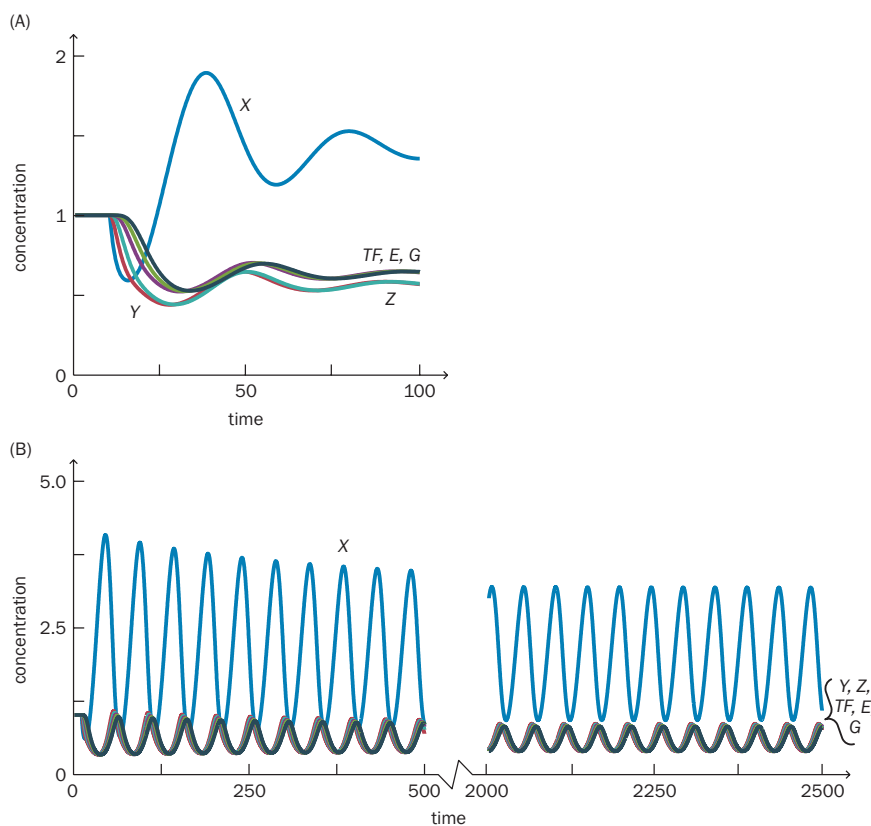


Figure 1.11 Simulation results demonstrate that the looped system in Figure 1.10 may exhibit drastically different responses. If the effect of *Z* on *TF* is very small, the response is essentially like that in Figure 1.9 (results not shown). (A) If the effect of *Z* on *TF* is relatively small, the 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.