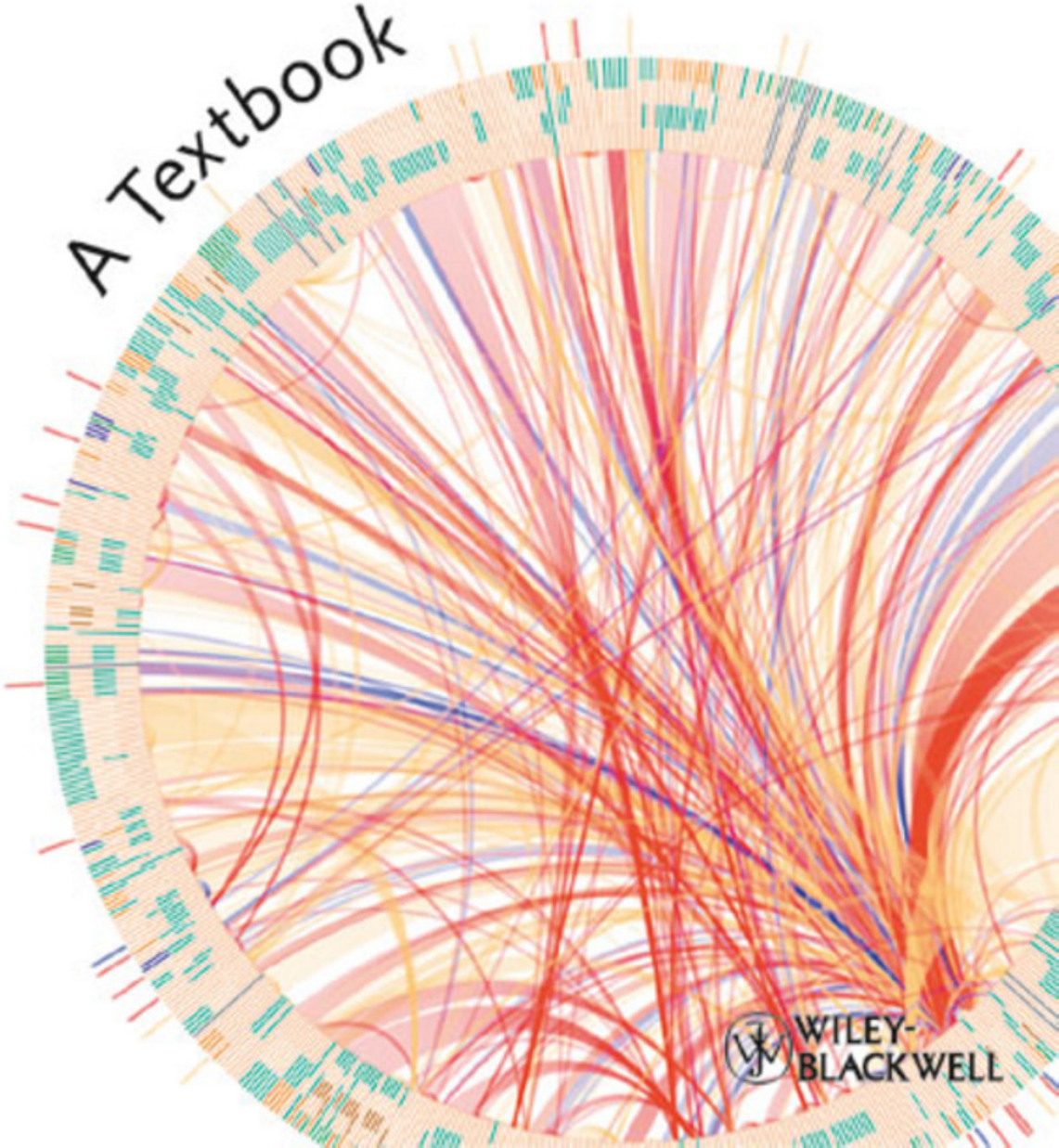
Systems Biology

Edda Klipp, Wolfram Liebermeister, Christoph Wierling,
Axel Kowald, Hans Lehrach, and Ralf Herwig

A Textbook



WILEY-
BLACKWELL

Edda Klipp, Wolfram Liebermeister,
Christoph Wierling, Axel Kowald,
Hans Lehrach, and Ralf Herwig
Systems Biology

Related Titles

Helms, V.

Principles of Computational Cell Biology

From Protein Complexes to Cellular Networks

2008
Softcover
ISBN: 978-3-527-31555-0

Baxevanis, A. D., Ouellette, B. F. F. (eds.)

Bioinformatics

A Practical Guide to the Analysis of Genes and Proteins

560 pages
2004
Hardcover
ISBN: 978-0-471-47878-2

Edda Klipp, Wolfram Liebermeister, Christoph Wierling,
Axel Kowald, Hans Lehrach, and Ralf Herwig

Systems Biology

A Textbook

WILEY-VCH Verlag GmbH & Co. KGaA

The Authors

Prof. Edda Klipp
Humboldt-Universität Berlin
Institut für Biologie
Theoretische Biophysik
Invalidenstr. 42
10115 Berlin

Dr. Wolfram Liebermeister
Humboldt-Universität Berlin
Institut für Biologie
Theoretische Biophysik
Invalidenstr. 42
10115 Berlin

Dr. Christoph Wierling
MPI für Molekulare Genetik
Ihnestr. 73
14195 Berlin
Germany

Dr. Axel Kowald
Protagen AG
Otto-Hahn-Str. 15
44227 Dortmund

Prof. Hans Lehrach
MPI für Molekulare Genetik
Ihnestr. 73
14195 Berlin
Germany

Prof. Ralf Herwig
MPI für Molekulare Genetik
Ihnestr. 73
14195 Berlin
Germany

Cover
The cover pictures were provided with kind permission by Santiago Ortiz and Dr. Michael Erlowitz

Library of Congress Card No.: applied for

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available on the Internet at http://dnb.d-nb.de.

Typesetting Thomson Digital, Noida, India
Printing Strauss GmbH, Mörlenbach
Binding Litges & Dopf GmbH, Heppenheim
Cover Design Adam-Design, Weinheim

Printed in the Federal Republic of Germany
Printed on acid-free paper

ISBN: 978-3-527-31874-2

Contents

Systems Biology: A Textbook. Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald, Hans Lehrach, and Ralf Herwig

ISBN: 978-3-527-31874-2

Preface

Life is probably the most complex phenomenon in the universe. We see kids growing, people aging, plants blooming, and microbes degrading their remains. We use yeast for brewery and bakery, and doctors prescribe drugs to cure diseases. But can we understand how life works? Since the 19th century, the processes of life have no longer been explained by special "living forces," but by the laws of physics and chemistry. By studying the structure and physiology of living systems more and more in detail, researchers from different disciplines have revealed how the mystery of life arises from the structural and functional organization of cells and from the continuous refinement by mutation and selection.

In recent years, new imaging techniques have opened a completely new perception of the cellular microcosm. If we zoom into the cell, we can observe how structures are built, maintained, and reproduced while various sensing and regulation systems help the cell to respond appropriately to environmental changes. But along with all these fascinating observations, many open questions remain. Why do we age? How does a cell know when to divide? How can severe diseases such as cancer or genetic disorders be cured? How can we convince – i.e., manipulate – microbes to produce a desirable substance? How can the life sciences contribute to environmental safety and sustainable technologies?

This book provides you with a number of tools and approaches that can help you to think in more detail about such questions from a theoretical point of view. A key to tackle such questions is to combine biological experiments with computational modeling in an approach called systems biology: it is the combined study of biological systems through (i) investigating the components of cellular networks and their interactions, (ii) applying experimental high-throughput and whole-genome techniques, and (iii) integrating computational methods with experimental efforts.

The systemic approach in biology is not new, but it recently gained new thrust due to the emergence of powerful experimental and computational methods. It is based on the accumulation of an increasingly detailed biological knowledge, on the emergence of new experimental techniques in genomics and proteomics, on a tradition of mathematical modeling of biological processes, on the exponentially growing computer power (as prerequisite for databases and the calculation of large

Systems Biology: A Textbook. Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald, Hans Lehrach, and Ralf Herwig

ISBN: 978-3-527-31874-2

systems), and on the Internet as the central medium for a quick and comprehensive exchange of information.

Systems Biology has influenced modern biology in two major ways: on the one hand, it offers computational tools for analyzing, integrating and interpreting biological data and hypotheses. On the other hand, it has induced the formulation of new theoretical concepts and the application of existing ones to new questions. Such concepts are, for example, the theory of dynamical systems, control theory, the analysis of molecular noise, robustness and fragility of dynamic systems, and statistical network analysis. As systems biology is still evolving as a scientific field, a central issue is the standardization of experiments, of data exchange, and of mathematical models.

In this book, we attempt to give a survey of this rapidly developing field. We will show you how to formulate your own model of biological processes, how to analyze such models, how to use data and other available information for making your model more precise – and how to interpret the results. This book is designed as an introductory course for students of biology, biophysics and bioinformatics, and for senior scientists approaching Systems Biology from a different discipline. Its nine chapters contain material for about 30 lectures and are organized as follows.

Chapter 1 – Introduction (E. Klipp, W. Liebermeister, A. Kowald, 1 lecture)

Introduction to the subject. Elementary concepts and definitions are presented. Read this if you want to start right from the beginning.

Chapter 2 – Modeling of Biochemical Systems (E. Klipp, C. Wierling, 4 lectures)

This chapter describes kinetic models for biochemical reaction networks, the most common computational technique in Systems Biology. It includes kinetic laws, stoichiometric analysis, elementary flux modes, and metabolic control analysis. Introduces tools and data formats necessary for modeling.

Chapter 3 – Specific Biochemical Systems (E. Klipp, C. Wierling, W. Liebermeister, 5 lectures)

Using specific examples from metabolism, signaling, and cell cycle, a number of popular modeling techniques are discussed. The aim of this chapter is to make the reader familiar with both modeling techniques and biological phenomena.

Chapter 4 – Model Fitting (W. Liebermeister, A. Kowald, 4 lectures)

Models in systems biology usually contain a large number of parameters. Assigning appropriate numerical values to these parameters is an important step in the creation of a quantitative model. This chapter shows how numerical values can be obtained from the literature or by fitting a model to experimental data. It also discusses how model structures can be simplified and how they can be chosen if several different models can potentially describe the experimental observations.

Chapter 5 – Analysis of High-Throughput Data (R. Herwig, 2 lectures)

Several techniques that have been developed in recent years produce large quantities of data (e.g., DNA and protein chips, yeast two-hybrid, mass spectrometry). But such large quantities often go together with a reduced quality of the individual measurement. This chapter describes techniques that can be used to handle this type of data appropriately.

Chapter 6 – Gene Expression Models (R. Herwig, W. Liebermeister, E. Klipp, 3 lectures)

Thousands of gene products are necessary to create a living cell, and the regulation of gene expression is a very complex and important task to keep a cell alive. This chapter discusses how the regulation of gene expression can be modeled, how different input signals can be integrated, and how the structure of gene networks can be inferred from experimental data.

Chapter 7 – Stochastic Systems and Variability (W. Liebermeister, 4 lectures)

Random fluctuations in transcription, translation and metabolic reactions make mathematics complicated, computation costly and interpretation of results not straight forward. But since experimentalists find intriguing examples for macroscopic consequences of random fluctuation at the molecular level, the incorporation of these effects into the simulations becomes more and more important. This chapter gives an overview where and how stochasticity enters cellular life.

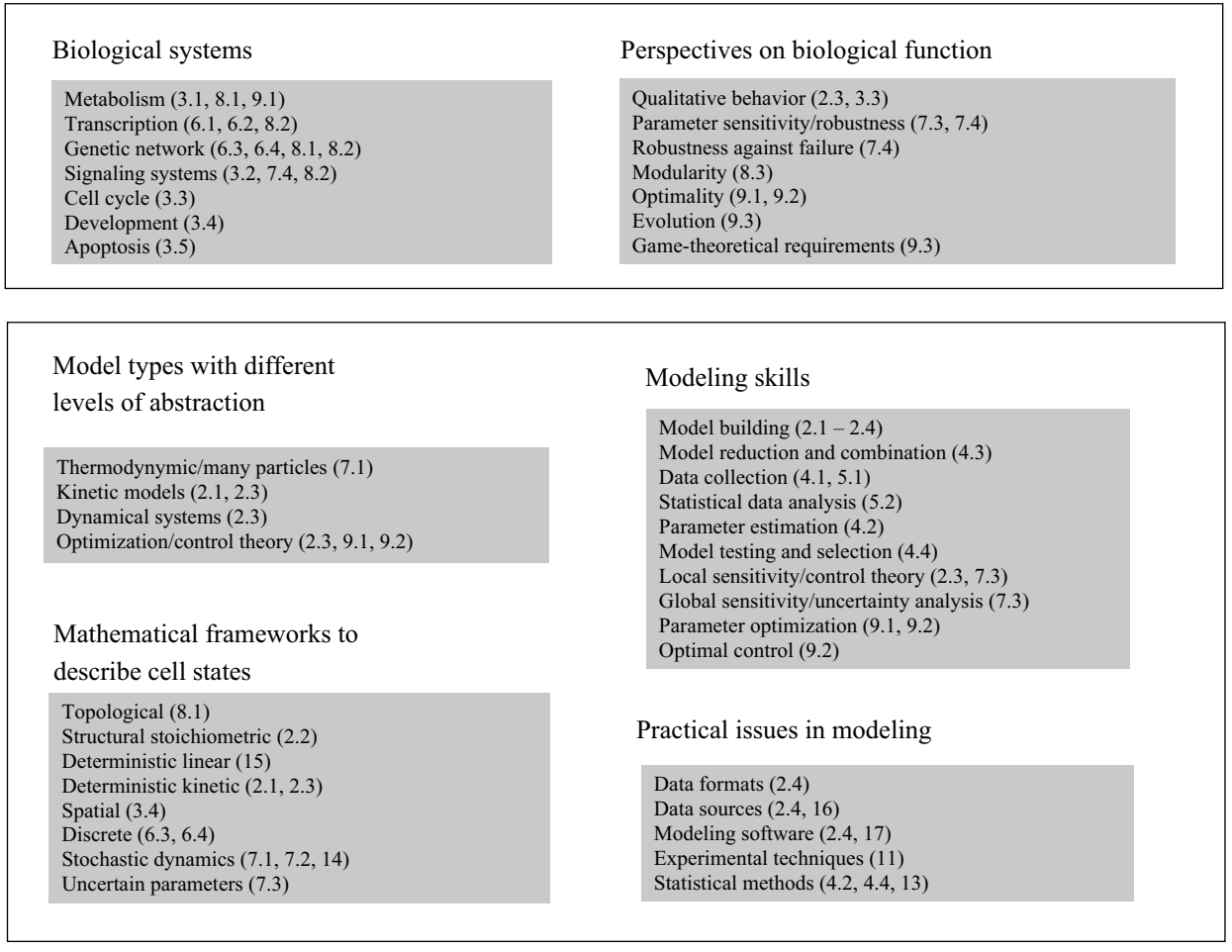
Chapter 8 – Network Structures, Dynamics and Function (W. Liebermeister, 3 lectures)

Many complex systems in biology can be represented as networks (reaction networks, interaction networks, regulatory networks). Studying the structure, dynamics, and function of such networks helps to understand design principles of living cells. In this chapter, important network structures such as motifs and modules as well as the dynamics resulting from them are discussed.

Chapter 9 – Optimality and Evolution (W. Liebermeister, E. Klipp, 3 lectures)

Theoretical research suggests that constraints of the evolutionary process should have left their marks in the construction and regulation of genes and metabolic pathways. In some cases, the function of biological systems can be well understood by models based on an optimality hypothesis. This chapter discusses the merits and limitations of such optimality approaches.

Various aspects of systems biology – the biological systems themselves, types of mathematical models to describe them, and practical techniques – reappear in different contexts in various parts of the book. The following diagram, which shows the contents of the book sorted by a number of different aspects, may serve as an orientation.


Biological systems
Metabolism (3.1, 8.1, 9.1)
Transcription (6.1, 6.2, 8.2)
Genetic network (6.3, 6.4, 8.1, 8.2)
Signaling systems (3.2, 7.4, 8.2)
Cell cycle (3.3)
Development (3.4)
Apoptosis (3.5)
Perspectives on biological function
Qualitative behavior (2.3, 3.3)
Parameter sensitivity/robustness (7.3, 7.4)
Robustness against failure (7.4)
Modularity (8.3)
Optimality (9.1, 9.2)
Evolution (9.3)
Game-theoretical requirements (9.3)
Model types with different levels of abstraction
Thermodynymic/many particles (7.1)
Kinetic models (2.1, 2.3)
Dynamical systems (2.3)
Optimization/control theory (2.3, 9.1, 9.2)
Mathematical frameworks to describe cell states
Topological (8.1)
Structural stoichiometric (2.2)
Deterministic linear (15)
Deterministic kinetic (2.1, 2.3)
Spatial (3.4)
Discrete (6.3, 6.4)
Stochastic dynamics (7.1, 7.2, 14)
Uncertain parameters (7.3)
Modeling skills
Model building (2.1 – 2.4)
Model reduction and combination (4.3)
Data collection (4.1, 5.1)
Statistical data analysis (5.2)
Parameter estimation (4.2)
Model testing and selection (4.4)
Local sensitivity/control theory (2.3, 7.3)
Global sensitivity/uncertainty analysis (7.3)
Parameter optimization (9.1, 9.2)
Optimal control (9.2)
Practical issues in modeling
Data formats (2.4)
Data sources (2.4, 16)
Modeling software (2.4, 17)
Experimental techniques (11)
Statistical methods (4.2, 4.4, 13)

At the end of the regular course material, you will find a number of additional chapters that summarize important biological and mathematical methods. The first chapters deal with to cell biology (chapter 10, C. Wierling) and molecular biological methods (chapter 11, A. Kowald). For looking up mathematical and statistical definitions and methods, turn to chapters 12 and 13 (R. Herwig, A. Kowald). Chapters 14 and 15 (W. Liebermeister) concentrate on random processes and control theory. The final chapters provide an overview over useful databases (chapter 16, C. Wierling) as well as a huge list of available software tools including a short description of their purposes (chapter 17, A. Kowald).

Further material is available on an accompanying website

(www.wiley-vch.de/home/systemsbiology)

Beside additional and more specialized topics, the website also contains solutions to the exercises and problems presented in the book.

We give our thanks to a number of people who helped us in finishing this book. We are especially grateful to Dr. Ulrich Liebermeister, Prof. Dr. Hans Meinhardt, Dr. Timo Reis, Dr. Ulrike Baur, Clemens Kreutz, Dr. Jose Egea, Dr. Maria Rodriguez-Fernandez, Dr. Wilhelm Huisinga, Sabine Hummert, Guy Shinar, Nadav Kashtan, Dr. Ron Milo, Adrian Jinich, Elad Noor, Niv Antonovsky, Bente Kofahl, Dr. Simon Borger, Martina Fröhlich, Christian Waltermann, Susanne Gerber, Thomas Spießer, Szymon Stoma, Christian Diener, Axel Rasche, Hendrik Hache, Dr. Michal Ruth Schweiger, and Elisabeth Maschke-Dutz for reading and commenting on the manuscript.

We thank the Max Planck Society for support and encouragement. We are grateful to the European Commission for funding via different European projects (MEST-CT2004-514169, LSHG-CT-2005-518254, LSHG-CT-2005-018942, LSHG-CT-2006-037469, LSHG-CT-2006-035995-2 NEST-2005-Path2-043310, HEALTH-F4-2007-200767, and LSHB-CT-2006-037712). Further funding was obtained from the Sysmo project "Translucent" and from the German Research Foundation (IRTG 1360) E.K. thanks with love her sons Moritz and Richard for patience and incentive and the Systems Biology community for motivation. W.L. wishes to thank his daughters Hannah and Marlene for various insights and inspiration. A.K. likes to thank Prof. Dr. H.E. Meyer for support and hospitality. This book is dedicated to our teacher Prof. Dr. Reinhart Heinrich (1946–2006), whose works on metabolic control theory in the 1970s paved the way to systems biology and who greatly inspired our minds.

Part One
Introduction to Systems Biology

Systems Biology: A Textbook. Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald, Hans Lehrach, and Ralf Herwig

ISBN: 978-3-527-31874-2

1
Introduction

1.1
Biology in Time and Space

Biological systems like organisms, cells, or biomolecules are highly organized in their structure and function. They have developed during evolution and can only be fully understood in this context. To study them and to apply mathematical, computational, or theoretical concepts, we have to be aware of the following circumstances.

The continuous reproduction of cell compounds necessary for living and the respective flow of information is captured by the central dogma of molecular biology, which can be summarized as follows: genes code for mRNA, mRNA serves as template for proteins, and proteins perform cellular work. Although information is stored in the genes in form of DNA sequence, it is made available only through the cellular machinery that can decode this sequence and can translate it into structure and function. In this book, this will be explained from various perspectives.

A description of biological entities and their properties encompasses different levels of organization and different time scales. We can study biological phenomena at the level of populations, individuals, tissues, organs, cells, and compartments down to molecules and atoms. Length scales range from the order of meter (e.g., the size of whale or human) to micrometer for many cell types, down to picometer for atom sizes. Time scales include millions of years for evolutionary processes, annual and daily cycles, seconds for many biochemical reactions, and femtoseconds for molecular vibrations. Figure 1.1 gives an overview about scales.

In a unified view of cellular networks, each action of a cell involves different levels of cellular organization, including genes, proteins, metabolism, or signaling pathways. Therefore, the current description of the individual networks must be integrated into a larger framework.

Many current approaches pay tribute to the fact that biological items are subject to evolution. The structure and organization of organisms and their cellular machinery has developed during evolution to fulfill major functions such as growth, proliferation, and survival under changing conditions. If parts of the organism or of the cell fail to perform their function, the individual might become unable to survive or replicate.

Systems Biology: A Textbook. Edda Klipp, Wolfram Liebermeister, Christoph Wierling, Axel Kowald, Hans Lehrach, and Ralf Herwig

ISBN: 978-3-527-31874-2

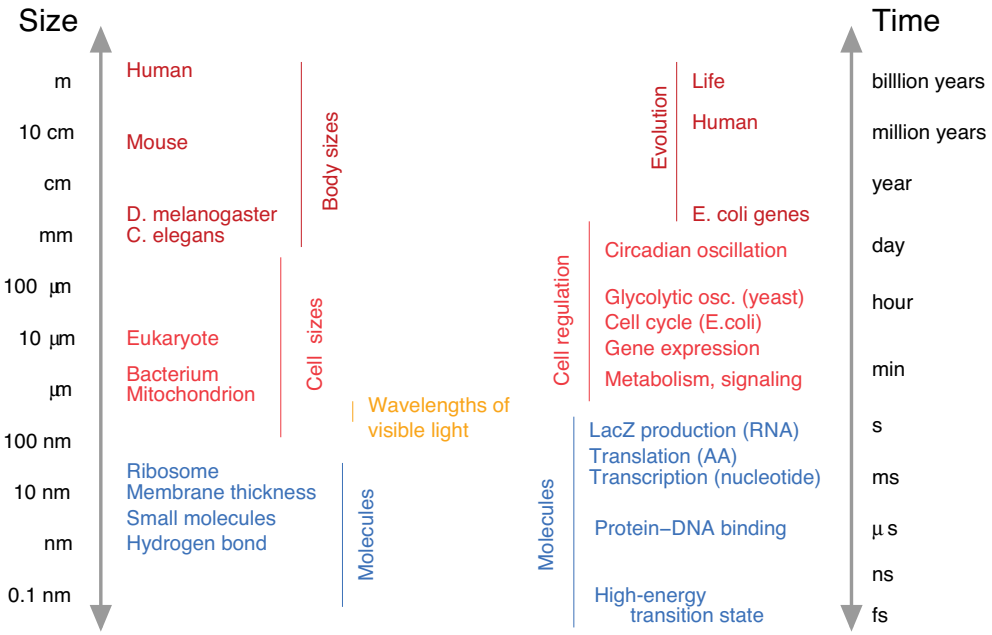


Figure 1.1 Length and time scales in biology. Data from the BioNumbers database http://bionumbers.hms.harvard.edu.

One consequence of evolution is the similarity of biological organisms from different species. This similarity allows for the use of model organisms and for the critical transfer of insights gained from one cell type to other cell types. Applications include, e.g., prediction of protein function from similarity, prediction of network properties from optimality principles, reconstruction of phylogenetic trees, or the identification of regulatory DNA sequences through cross-species comparisons. But the evolutionary process also leads to genetic variations within species. Therefore, personalized medicine and research is an important new challenge for biomedical research.

1.2 Models and Modeling

If we observe biological processes, we are confronted with various complex processes that cannot be explained from first principles and the outcome of which cannot reliably be foreseen from intuition. Even if general biochemical principles are well established (e.g., the central dogma of transcription and translation, the biochemistry of enzyme-catalyzed reactions), the biochemistry of individual molecules and systems is often unknown and can vary considerably between species. Experiments lead to biological hypotheses about individual processes, but it often remains unclear if these hypotheses can be combined into a larger coherent picture because it is often

difficult to foresee the global behavior of a complex system from knowledge of its parts. Mathematical modeling and computer simulations can help us understand the internal nature and dynamics of these processes and to arrive at predictions about their future development and the effect of interactions with the environment.

1.2.1 What is a Model?

The answer to this question will differ among communities of researchers. In a broad sense, a model is an abstract representation of objects or processes that explains features of these objects or processes (Figure 1.2). A biochemical reaction network can be represented by a graphical sketch showing dots for metabolites and arrows for reactions; the same network could also be described by a system of differential equations, which allows simulating and predicting the dynamic behavior of that network. If a model is used for simulations, it needs to be ensured that it faithfully predicts the system's behavior – at least those aspects that are supposed to be covered by the model. Systems biology models are often based on well-established physical laws that justify their general form, for instance, the thermodynamics of chemical reactions; besides this, a computational model needs to make specific statements about a system of interest – which are partially justified by experiments and biochemical knowledge, and partially by mere extrapolation from other systems. Such a model can summarize established knowledge about a system in a coherent mathematical formulation. In experimental biology, the term "model" is also used to denote a species that is especially suitable for experiments, for example, a genetically modified mouse may serve as a model for human genetic disorders.

1.2.2 Purpose and Adequateness of Models

Modeling is a subjective and selective procedure. A model represents only specific aspects of reality but, if done properly, this is sufficient since the intention of modeling is to answer particular questions. If the only aim is to predict system outputs from given input signals, a model should display the correct input–output relation, while its interior can be regarded as a black box. But if instead a detailed biological mechanism has to be elucidated, then the system's structure and the relations between its parts must be described realistically. Some models are meant to be generally applicable to many similar objects (e.g., Michaelis–Menten kinetics holds for many enzymes, the promoter–operator concept is applicable to many genes, and gene regulatory motifs are common), while others are specifically tailored to one particular object (e.g., the 3D structure of a protein, the sequence of a gene, or a model of deteriorating mitochondria during aging). The mathematical part can be kept as simple as possible to allow for easy implementation and comprehensible results. Or it can be modeled very realistically and be much more complicated. None of the characteristics mentioned above makes a model wrong or right, but they determine whether a model is appropriate to the problem to be solved. The phrase "essentially,

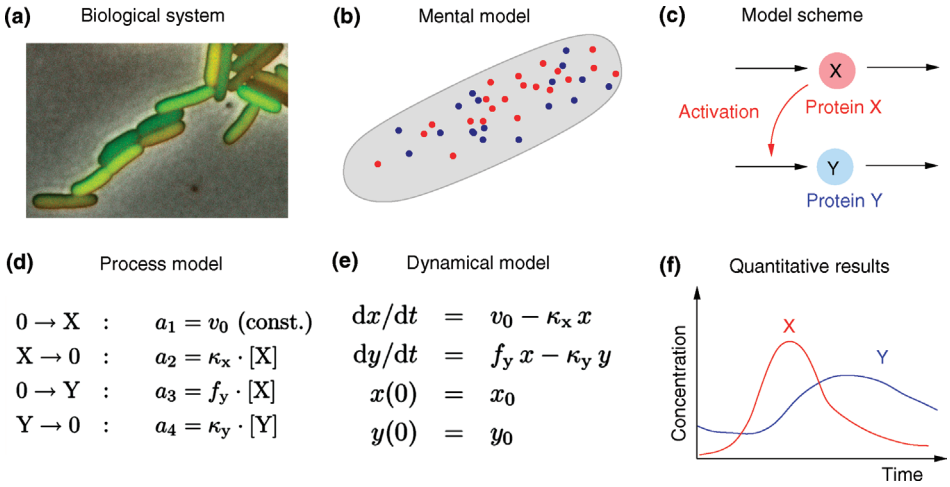


Figure 1.2 Typical abstraction steps in mathematical modeling. (a) *Escherichia coli* bacteria produce thousands of different proteins. If a specific protein type is fluorescently labeled, cells glow under the microscope according to the concentration of this enzyme (Courtesy of M. Elowitz). (b) In a simplified mental model, we assume that cells contain two enzymes of interest, X (red) and Y (blue) and that the molecules (dots) can freely diffuse within the cell. All other substances are disregarded for the sake of simplicity. (c) The interactions between the two protein types can be drawn in a wiring scheme: each protein can be produced or degraded (black arrows). In addition, we assume that proteins of type X can increase the production of protein Y. (d) All individual processes to be considered are listed together with their rates a (occurrence per time). The mathematical expressions for the rates are based on a simplified picture of the actual chemical processes. (e) The list of processes can be translated into different sorts of dynamic models; in this case, deterministic rate equations for the protein concentrations x and y. (f) By solving the model equations, predictions for the time-dependent concentrations can be obtained. If these predictions do not agree with experimental data, it indicates that the model is wrong or too much simplified. In both cases, it has to be refined.

all models are wrong, but some are useful" coined by the statistician George Box is indeed an appropriate guideline for model building.

1.2.3 Advantages of Computational Modeling

Models gain their reference to reality from comparison with experiments, and their benefits therefore depend on the quality of the experiments used. Nevertheless, modeling combined with experimentation has a lot of advantages compared to purely experimental studies:

- Modeling drives conceptual clarification. It requires verbal hypotheses to be made specific and conceptually rigorous.
- Modeling highlights gaps in knowledge or understanding. During the process of model formulation, unspecified components or interactions have to be determined.