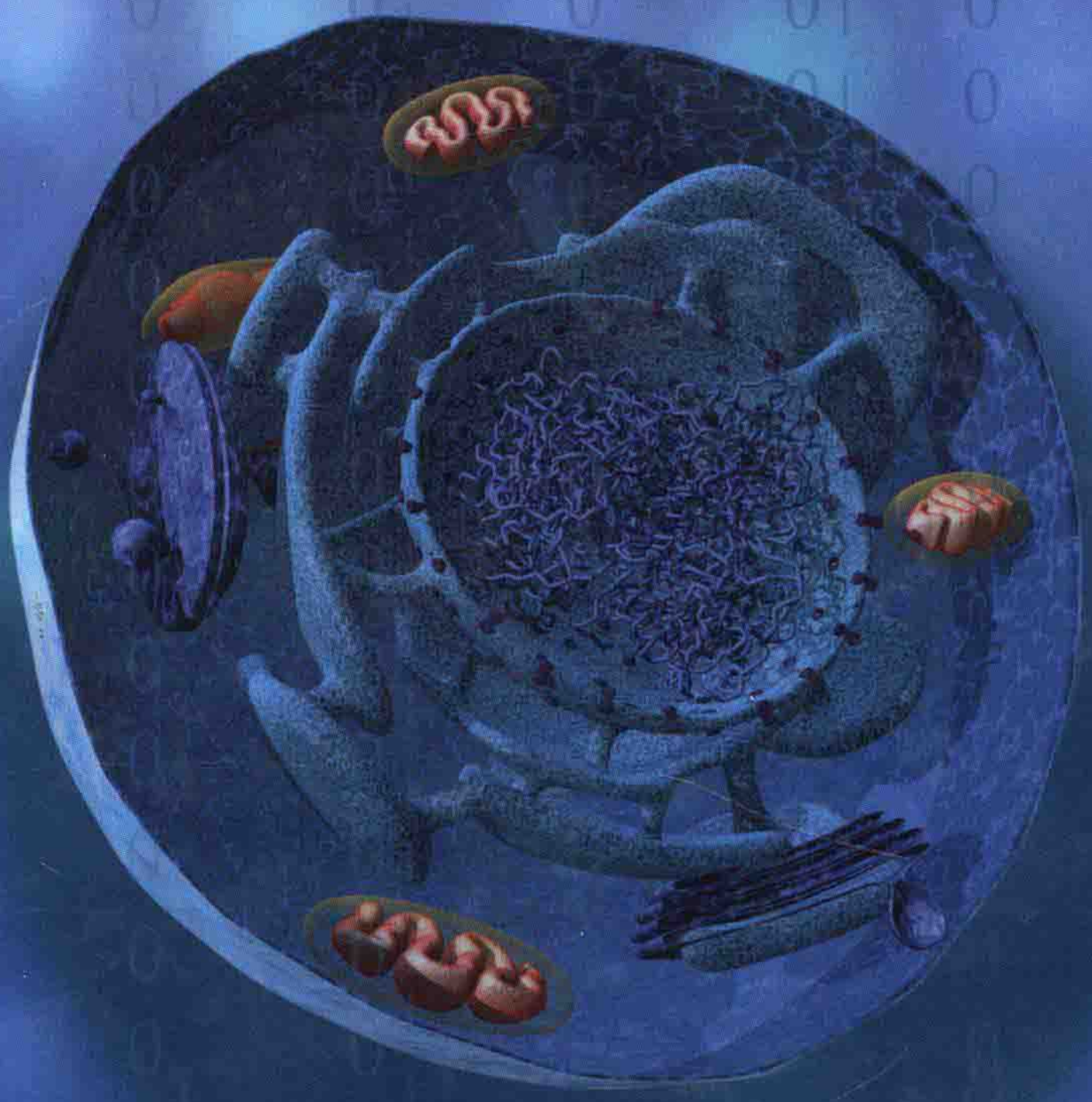
Systems Biology
Simulation of Dynamic Network States



Bernhard Ø. Palsson

CAMBRIDGE

Systems Biology

SIMULATION OF DYNAMIC NETWORK STATES

Bernhard Ø. Palsson
University of California, San Diego

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town,
Singapore, São Paulo, Delhi, Tokyo, Mexico City

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9781107001596

© B. Ø. Palsson 2011

This publication is in copyright. Subject to statutory exception
and to the provisions of relevant collective licensing agreements,
no reproduction of any part may take place without the written
permission of Cambridge University Press.

First published 2011

Printed in the United Kingdom at the University Press, Cambridge

A catalogue record for this publication is available from the British Library

ISBN 978-1-107-00159-6 Hardback

Additional resources for this publication at http://systemsbiology.ucsd.edu

Cambridge University Press has no responsibility for the persistence or
accuracy of URLs for external or third-party internet websites referred to
in this publication, and does not guarantee that any content on such
websites is, or will remain, accurate or appropriate.

SYSTEMS BIOLOGY: SIMULATION OF DYNAMIC NETWORK STATES

Biophysical models have been used in biology for decades, but they have been limited in scope and size. In this book, Bernhard Ø. Palsson shows how network reconstructions that are based on genomic and bibliomic data, and take the form of established stoichiometric matrices, can be converted into dynamic models using metabolomic and fluxomic data. The Mass Action Stoichiometric Simulation (MASS) procedure can be used for any cellular process for which data is available and allows a scalable step-by-step approach to the practical construction of network models. Specifically, it can treat integrated processes that need explicit accounting of small molecules and proteins, which allows simulation at the molecular level. The material has been class-tested by the author at both the undergraduate and graduate level.

All computations in the text are available online in MATLAB® and MATHEMATICA® workbooks, allowing hands-on practice with the material.

Bernhard Ø. Palsson is The Galletti Professor of Bioengineering and Adjunct Professor of Medicine at the University of California, San Diego.

TO
ERNA AND PÁLL

此为试读,需要完整PDF请访问: www.ertongbook.com

Preface

(Molecular) Systems biology has developed over roughly the past 10 years. Its emergence has led to the development of broad genome-wide or network-wide viewpoints of organism functions that have developed against the context of whole genome sequences. Bottom-up approaches to network reconstruction have resulted in organism-specific networks that have a direct genetic and genomic basis. Such networks are now available for a growing number of organisms.

Genome-scale networks have been used to develop constraint-based reconstruction and analysis (COBRA) procedures that treat structural properties of networks, their physiological capabilities, optimal functional states of organisms, and studies of adaptive and long-term evolution. These topics are treated in the companion book that emphasizes that while biology is dynamic, it still functions under the constraints of the topological structure of the molecular networks that underlie its functions.

Events over the time scales associated with distal causation in biology, i.e., over multiple generations, can be studied within the COBRA framework. However, analysis of proximal or immediate dynamic responses of organisms is limited. The recent development of high-throughput technologies and the availability of omics data sets has opened up an alternative approach to building large-scale models that can compute the dynamic states of biological networks. Omics-based abundance measurements (i.e., for proteins, transcripts, and metabolites) can now be mapped onto network reconstructions. In addition, functional states can be determined from fluxomic, exo-metabolomic, and various physiological data types.

The combination of omics data sets and network reconstructions allows the generation of Mass Action Stoichiometric Simulation (MASS) models. Such models can, at this point in time, be formulated for metabolism and associated enzymes and other protein molecules. MASS models will be condition specific, as they use particular data sets. In principle, MASS

models can be formulated for any cellular phenomena for which reconstructions and omics data sets are available. Although the procedure is now established, some of the practical issues associated with its broad implementation will need additional experience that will call on further research in this field.

This book is focused on the process and the issues associated with the generation of MASS models. Their foundational concepts are described and they are applied to specific cases. Once the reader has mastered these concepts and gone through the details of their application to familiar cellular processes, you should be able to build MASS models for cellular phenomena of interest.

One should be aware of the fact that dynamic models have been constructed to describe biological phenomena for many decades. At the biochemical level, such models have been largely based on biophysical principles, heavily focused in particular on the use of *in-vitro*-derived rate laws. Given the scarcity of such rate laws, this approach to building kinetic models has limited the scope and size of dynamic models built in this fashion. The omics data-driven MASS procedure provides an alternative condition-dependent approach that is scalable.

This book, in a sense, brings my career full circle. My first love in graduate school was building complex dynamic models in biology based on the contents of the graduate curriculum in chemical engineering. However, as stated above, the application of these methods to biology was necessarily limited due to data availability and due to the "absolute" characteristics of biophysical models. The path through stoichiometric models from the biochemical to the genome scale based on full genome sequences, to large-scale dynamic models based on omics data sets has been an interesting one. Given the impending onslaught of genetic data and associated potential for biological variation, this field might be just in its infancy.

This text is constructed to teach how to build complex dynamic models of biochemical networks and how to simulate their responses. The material has been taught both at the undergraduate and graduate level at UC San Diego since 2008. Teaching the material at these two levels has led to the development of a set of homework problems (Appendix B) and a collection of Mathematica workbooks. It is my intent to make these available through an on-line source, initially on http://systemsbiology.ucsd.edu. I hope both will be helpful to instructors.

The path to this book has had many influences. Reich and Selkov's 1982 book, *Energy Metabolism of the Cell*, certainly contains many foundational and influential concepts. The *Color Atlas of Biochemistry* by Koolman and Roehm provides succinct representation of biochemical knowledge that has been useful in developing the material. All the computations in the

text were done in Mathematica. Throughout my entire career, LaTeX has been an essential resource, as it was for writing this book.

There are special thanks due to two individuals. Neema Jamshidi has been an MD/PhD student in my lab over the past 6 or 7 years. He has been a fantastic colleague and friend. He educated me about the use of Mathematica and tirelessly answered my repeated and often naive questions. He has also been a source of great intellectual stimulation and discussions. He was a major influence in completing this book. As with the companion book, Marc Abrams made the writing, preparation, editing, and production of this book possible. He supervised, coordinated, and implemented the construction of the LaTeX document and the preparation of many figures in the text. Special thanks to these two gentlemen.

In addition, three PhD students in my lab were of invaluable help in getting this book to the state of completion that it has reached. Aarash Bordbar helped me with the formulation of the complicated Mathematica workbooks for Part IV of the text. In addition, he has played a notable role in developing the work flow for MASS models. Daniel Zielinski not only helped build the Mathematica workbooks for Part III, but proofread the text with his impeccable eye for detail and logical flow of material. Addiel U. de Alba Solis helped with the Mathematica workbooks for Parts I and II of the text. All three were very helpful in reviewing, correcting, and providing solutions to the homework sets given in Appendix B.

Others have helped with this text either indirectly or directly through thoughtful comments or the preparation of illustrations. For their assistance, I am grateful: Kenyon Applebee, Tom Conrad, Markus Herrgard, Joshua Lerman, Vasiliy Portnoy, Jan Schellenberger, Paolo Vicini and Michael Zager.

This book is dedicated to my parents, who enabled, allowed, supported, and encouraged me to pursue my studies and interests in integrated biological processes. Without them I would not have reached this level of professional development and would not have written this book. Kærar þakkir.

Bernhard Palsson
La Jolla, CA
April 2010

Contents

此为试读,需要完整PDF请访问: www.ertongbook.com

CHAPTER 1

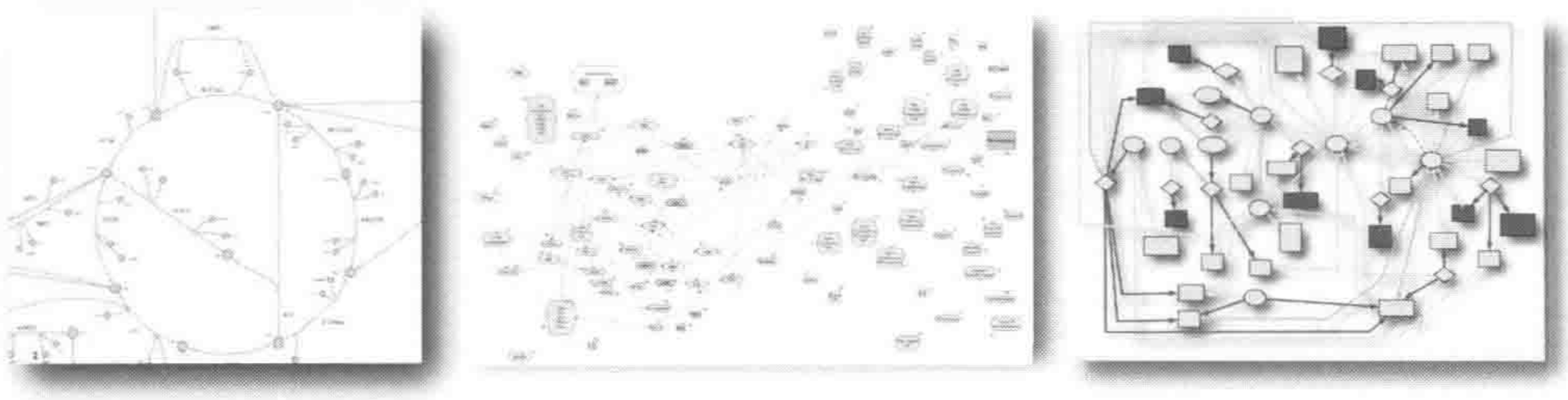
Introduction

Systems biology has been brought to the forefront of life-science-based research and development. The need for systems analysis is made apparent by the inability of focused studies to explain whole network, cell, or organism behavior, and the availability of component data is what is fueling and enabling the effort. This massive amount of experimental information is a reflection of the complex molecular networks that underlie cellular functions. Reconstructed networks represent a common denominator in systems biology. They are used for data interpretation, comparing organism capabilities, and as the basis for computing their functional states. The companion book [89] details the topological features and assessment of functional states of biochemical reaction networks and how these features are represented by the stoichiometric matrix. In this book, we turn our attention to the kinetic properties of the reactions that make up a network. We will focus on the formulation of dynamic simulators and how they are used to generate and study the dynamic states of biological networks.

1.1 Biological networks

Cells are made up of many chemical constituents that interact to form networks. Networks are fundamentally comprised of *nodes* (the compounds) and the *links* (chemical transformations) between them. The networks take on functional states that we wish to compute, and it is these physiological states that we observe. This text is focused on dynamic states of networks.

There are many different kinds of biological network of interest, and they can be defined in different ways. One common way of defining networks is based on a preconceived notion of what they do. Examples include metabolic, signaling, and regulatory networks; see Figure 1.1. This



(a) Metabolic (b) Signaling (c) Regulatory

Figure 1.1 Three examples of networks that are defined by major function. (a) Metabolism. (b) Signaling. From Arisi *et al. BMC Neuroscience* 2006 7(Suppl 1):S6 DOI: 10.1186/1471-2202-7-S1-S6. (c) Transcriptional regulatory networks. Image courtesy of Christopher Workman, Center for Biological Sequence Analysis, Technical University of Denmark.

approach is driven by a large body of literature that has grown around a particular cellular function.

Metabolic networks Metabolism is ubiquitous in living cells and is involved in essentially all cellular functions. It has a long history – glycolysis was the first pathway elucidated in the 1930s – and is thus well known in biochemical terms. Many of the enzymes and the corresponding genes have been discovered and characterized. Consequently, the development of dynamic models for metabolism is the most advanced at the present time.

A few large-scale kinetic models of metabolic pathways and networks now exist. Genome-scale reconstructions of metabolic networks in many organisms are now available. With the current developments in metabolomics and fluxomics, there is a growing number of large-scale data sets becoming available. However, there are no genome-scale dynamic models yet available for metabolism.

Signaling networks Living cells have a large number of sensing mechanisms to measure and evaluate their environment. Bacteria have a surprising number of two-component sensing systems that inform the organism about its nutritional, physical, and biological environment. Human cells in tissues have a large number of receptor systems in their membranes to which specific ligands bind, such as growth factors or chemokines. Such signaling influences the cellular fate processes: differentiation, replication, apoptosis, and migration.

The functions of many of the signaling pathways that is initiated by a sensing event are presently known, and this knowledge is becoming more detailed. Only a handful of signaling networks are well known,

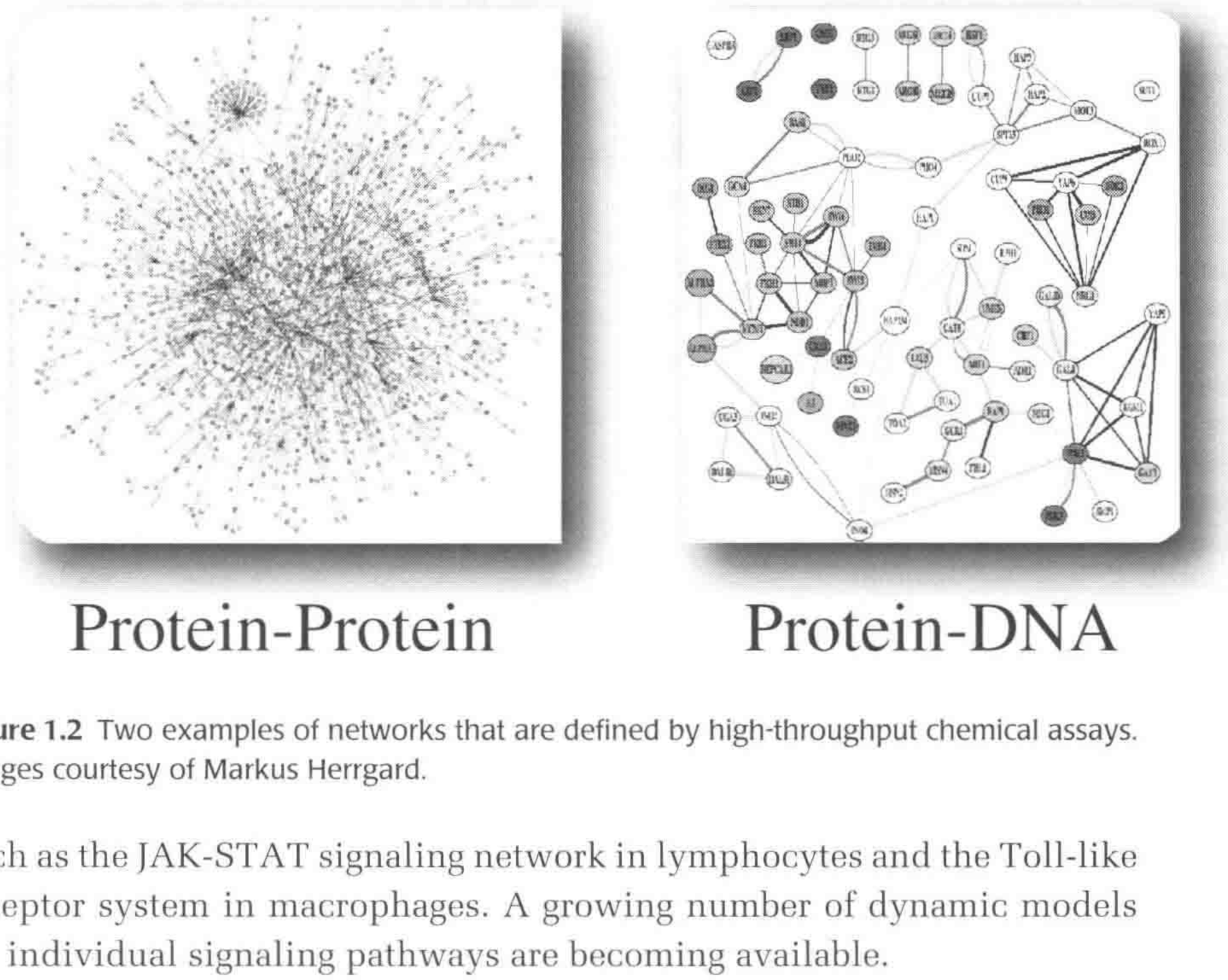


Figure 1.2 Two examples of networks that are defined by high-throughput chemical assays. Images courtesy of Markus Herrgard.

such as the JAK-STAT signaling network in lymphocytes and the Toll-like receptor system in macrophages. A growing number of dynamic models for individual signaling pathways are becoming available.

Regulatory networks There is a complex network of interactions that determine the DNA binding state of most proteins, which in turn determine whether genes are being expressed. The RNA polymerase must bind to DNA, as do transcription factors and various other proteins. The details of these chemical interactions are being worked out, but in the absence of such information, most of the network models that have been built are discrete, stochastic, and logistical in nature.

With the rapid development of experimental methods that measure expression states, the binding sites, and their occupancy, we may soon see large-scale reconstructions of transcriptional regulatory networks. Once these are available, we can begin to plan the process to build models that will describe their dynamic states.

Unbiased network definitions An alternative way to define networks is based on chemical assays. Measuring all protein–protein interactions regardless of function provides one such example; see Figure 1.2. Another example is a genome-wide measurement of the binding sites of a DNA-binding protein. This approach is driven by data-generating capabilities. It does not have an *a priori* bias about the function of molecules being examined.

Table 1.1 Web resources that contain information about biological networks (prepared by Jan Schellenberger)

		Metabolic	Protein–protein	Regulatory/signaling	Organisms	Curation[a]
KEGG	http://www.genome.jp/kegg/	x			many	C
BiGG	http://bigg.ucsd.edu/	x			many	M
BioCyc[b]	http://biocyc.org/	x		x	many	C/M
MetaCyc	http://metacyc.org/	x			many	C/M
Reactome	http://reactome.org/	x	x	x	many	M
BIND	http://www.bindingdb.org/		x		many	E/M
DIP	http://dip.doe-mbi.ucla.edu/		x		many	M
HPRD	http://www.hprd.org/		x		human	M
MINT	http://mint.bio.uniroma2.it/		x		many	M
Biogrid	http://www.thebiogrid.org/		x		many	E
UniHI	http://theoderich.fb3.mdc-berlin.de:8080/unihi/		x		human	E/M
Yeastract	http://www.yeastract.com/			x	yeast	M
TRANSFAC	http://www.gene-regulation.com			x	many	M
TRANSPATH	http://www.gene-regulation.com			x	many	M
RegulonDB	http://regulondb.ccg.unam.mx/			x	many	C/E
NetPath	http://www.netpath.org/			x	human	M

[a] M = manual/literature, C = computational, E = experimental.

[b] Links to other *Cyc databases.

Network reconstruction Metabolic networks are currently the best-characterized biological networks for which the most detailed reconstructions are available. The conceptual basis for their reconstruction has been reviewed [100], the workflows used detailed [35], and a detailed standard operating procedure (SOP) is available [117]. Some of the fundamental issues associated with the generation of dynamic models describing their functions have been articulated [52].

There is much interest in reconstructing signaling and regulatory networks in a similar way. The prospects for reconstruction of large-scale signaling networks have been discussed [49]. Given the development of new omics data types and other information, it seems likely that we will be able to obtain reliable reconstructions of these networks in the not too distant future.

Public information about pathways and networks There is a growing number of networks that underlie cellular functions that are being unraveled and reconstructed. Many publicly available sources contain this information; see Table 1.1. We wish to study the dynamic states of such networks. To do so, we need to describe them in chemical detail and

incorporate thermodynamic information and formulate a mathematical model.

1.2 Why build and study models?

Mathematical modeling is practiced in various branches of science and engineering. The construction of models is a laborious and detailed task. It also involves the use of numerical and mathematical analysis, both of which are intellectually intensive and unforgiving undertakings. So why bother?

Bailey's five reasons The purpose and utility of model building has been succinctly summarized and discussed [15]:

1. *"To organize disparate information into a coherent whole."* The information that goes into building models is often found in many different sources and the model builder has to look for these, evaluate them, and put them in context. In our case, this comes down to building data matrices (see Table 1.3) and determining conditions of interest.
2. *"To think (and calculate) logically about what components and interactions are important in a complex system."* Once the information has been gathered it can be mathematically represented in a self-consistent format. Once equations have been formulated using the information gathered and according to the laws of nature, the information can be mathematically interrogated. The interactions among the different components are evaluated and the behavior of the model is compared with experimental data.
3. *"To discover new strategies."* Once a model has been assembled and studied, it often reveals relationships among its different components that were not previously known. Such observations often lead to new experiments, or form the basis for new designs. Further, when a model fails to reproduce the functions of the process being described, it means there is either something critical missing in the model or the data that led to its formulation is inconsistent. Such an occurrence then leads to a re-examination of the information that led to the model formulation. If no logical flaw is found, the analysis of the discrepancy may lead to new experiments to try to discover the missing information.
4. *"To make important corrections to the conventional wisdom."* The properties of a model may differ from the governing thinking about

process phenomena that is inferred based on qualitative reasoning. Good models may thus lead to important new conceptual developments.

5. *"To understand the essential qualitative features."* Since a model accounts for all interactions described among its parts, it often leads to a better understanding of the whole. In the present case, such qualitative features relate to multi-scale analysis in time and an understanding of how multiple chemical events culminate in coherent physiological features.

1.3 Characterizing dynamic states

The dynamic analysis of complex reaction networks involves the tracing of time-dependent changes of concentrations and reaction fluxes over time. The concentrations typically considered are those of metabolites, proteins, or other cellular constituents. There are three key characteristics of dynamic states that we mention here right at the outset, and they are described in more detail in Section 2.1.

Time constants Dynamic states are characterized by change in time; thus, the rate of change becomes the key consideration. The rate of change of a variable is characterized by a *time constant*. Typically, there is a broad spectrum of time constants found in biochemical reaction networks. This leads to time-scale separation, where events may be happening on the order of milliseconds all the way to hours, if not days. The determination of the spectrum of time constants is thus central to the analysis of network dynamics.

Aggregate variables An associated issue is the identification of the biochemical, and ultimately physiological, events that are unfolding on every time scale. Once identified, one begins to form *aggregate concentration variables*, or *pooled variables*. These variables will be combinations of the original concentration variables. For example, two concentration variables may interconvert rapidly, on the order of milliseconds and thus on every time scale longer than milliseconds these two concentrations will be "connected." They can, therefore, be "pooled" together to form an aggregate variable. An example is given in Figure 1.3.

The determination of such aggregate variables becomes an intricate mathematical problem. Once solved, it allows us to determine the *dynamic structure of a network*. In other words, we move hierarchically away from the original concentration variables to increasingly interlinked aggregate

此为试读,需要完整PDF请访问: www.ertongbook.com