Systems Biology
Constraint-based Reconstruction and Analysis

Recent technological advances have enabled comprehensive determination of the molecular composition of living cells. The chemical interactions between many of these molecules are known, giving rise to genome-scale reconstructed biochemical reaction networks underlying cellular functions. Mathematical descriptions of the totality of these chemical interactions lead to genome-scale models that allow the computation of physiological functions.

Reflecting these recent developments, this textbook explains how such quantitative and computable genotype–phenotype relationships are built using a genome-wide basis of information about the gene portfolio of a target organism. It describes how biological knowledge is assembled to reconstruct biochemical reaction networks, the formulation of computational models of biological functions, and how these models can be used to address key biological questions and enable predictive biology.

Developed through extensive classroom use, the book is designed to provide students with a solid conceptual framework and an invaluable set of modeling tools and computational approaches.

Detailed lecture slides, along with MATLAB^{TM} and Mathematica^{TM} workbooks, are available for download at www.cambridge.org/sb.

Bernhard O. Palsson is the Galletti Professor of Bioengineering and Professor of Pediatrics at the University of California, San Diego. For almost 30 years, his research has focused on the development of large-scale models of biological functions and their use to solve basic and applied problems in the life sciences. He has authored three previous textbooks.

Systems Biology

Constraint-based Reconstruction and Analysis

BERNHARD O. PALSSON
Department of Bioengineering,
University of California at San Diego, USA

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107038851

First published 2015

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

Library of Congress Cataloging-in-Publication data
Palsson, Bernhard, author.
Systems biology : constraint-based reconstruction and analysis / Bernhard O. Palsson.
p. ; cm.
Includes bibliographical references.
ISBN 978-1-107-03885-1 (Hardback)
1. Title.
[DNLM: 1. Models, Biological. 2. Systems Biology. 3. Metabolic Networks and
Pathways–physiology. QU 26.5]
QH508
571.7–dc23 2014031793

ISBN 978-1-107-03885-1 Hardback

Cambridge University Press & Assessment 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.

To SHIREEN and SIRUS

Contents

Cambridge University Press & Assessment
978-1-107-03885-1 — Systems Biology: Constraint-based Reconstruction and Analysis
Bernhard Ø. Palsson
Frontmatter
More Information

Cambridge University Press & Assessment
978-1-107-03885-1 — Systems Biology: Constraint-based Reconstruction and Analysis
Bernhard Ø. Palsson
Frontmatter
More Information

Preface

The genesis of the bottom-up approach to systems biology was the availability of the first full genome sequences. In principle, these sequences had information about all the genetic elements that underlie the function of the sequenced organism. Enough information was available about the function of subsets of these genes – namely the genes encoding metabolic functions – that an organized assembly of all the biochemical, genetic, and genomic information was achievable. Such an organized assembly is *de facto* a knowledge base, or a k-base, that gives rise to a network reconstruction at the genome-scale. Since such reconstructions are represented with accurate chemical equations, they can be mathematically described. A mathematical description can be used to compute functional states of a network that correspond to observable phenotypes and biological functions. With these elements in place, a new genome-scale science was born that focused on mechanistic genotype–phenotype relationships.

The first genome-scale models of metabolism appeared in 1999 and 2000. In the next half-decade or so, an enthusiastic group of investigators developed many fundamental concepts, *in silico* methods, and algorithms to analyze their properties. At times, and to many, these initial efforts seemed mostly exploratory. Fortunately, in the mid-2000s an abundance of data sets and data types became available to validate and demonstrate the utility of genome-scale models for research and discovery. At the end of the decade several highly curated models were available for model organisms. These models gained predictive power and over the next 5 years or so, a number of prospective uses of genome-scale models appeared. In other words, predictive genotype–phenotype relationships had appeared. These predictions were somewhat limited in scope, but proved useful for a series of applications. Currently the range of possible predictions of biological properties and functions is growing rapidly and it appears that this approach to genome-scale science is in its early stages of development, with a bright future ahead of it.

For most of this fifteen-year history, the focus of genome-scale models has been metabolism. After initial successes with metabolic genome-scale models, it became clear that the same approach that led to their genesis could be applied to any other cellular process reconstructed in biochemically accurate detail. Thus, a vision was laid out in 2003 that the path to whole-cell models was conceptually possible and that such models could be used as a context for mechanistically integrating disparate omic data types. Ten years later, this vision started to be realized and a rapidly growing number of cellular functions are being reconstructed and addressed computationally. Given the fact that the genotype–phenotype relationship is fundamental to biology, this development has a broad transformative potential for the life sciences.

Writing this book was hard. It represents an attempt to summarize the concepts that have been developing over the past 15 years or so, that underlie what has become a true genome-scale science. Looking at the history of the field after the writing process, it is quite remarkable to see its rapid emergence, development, and maturation. Furthermore, looking forward, it appears that numerous areas of microbiology, cell biology, and developmental biology will be influenced by the approach and methods described in this book.

To master this field one needs familiarity with an unusual range of disciplines. One needs to understand the basics of life sciences: biochemistry, molecular biology, genetics, microbiology, and cell biology. High-throughput measurements call for an understanding of basic technological characteristics, such as multiplexing, miniaturization, and automation. The large data sets generated call for proficiency in bioinformatics and a comfort level with big data. Mathematically modeling such data sets from a fundamental standpoint requires familiarity with the mathematical language of linear algebra and logistical relationships. Simulations require the use of constraint-based optimization and an understanding of the evolutionary principles of generation of diversity and selection. Bottom-up systems biology is thus a field with a broad conceptual basis. This book attempts to bring all these concepts from the expert level to the general senior or first-year graduate student level in bioengineering, bioinformatics, and life sciences.

As with all major undertakings, this project could not have been completed without the help of several individuals.

Marc Abrams managed all aspects of the preparation of the manuscript. He tirelessly helped me with preparing the text and the illustrations, assembling the references, correcting LATEX scripts, and interacting with the publisher. Without him this book would not have been completed.

Nathan Lewis and Adam Feist were responsible for the challenging task of managing the original illustrations in the book. Their contribution was immense, making the concepts in the text and the material as a whole more accessible.

The following people were generous with their time and expertise, improving the manuscript with their contributions to the text, figures, or proofreading of the final manuscript. I am very grateful to these individuals:

Ramy Aziz, Aarash Bordbar, Roger Chang, Addiel U. de Alba Solis, Andreas Drger, Juan Nogales Enrique, Gabriela Guzman, Hooman Hefzi, Daniel Hyduke, Neema Jamshidi, Ryan LaCroix, Haythem Latif, Josh Lerman, Douglas McCloskey, Jonathan Monk, Harish Nagarajan, Jeff Orth, Troy Sandberg, Nikolaus Sonnenschein, Alex Thomas, and Daniel Zielinski.

The conceptual framework that this book describes has been under development since the birth of my two children, to whom it is dedicated.

Bernhard Palsson
On the Oracle, August 2014

Abbreviations

ALE	adaptive laboratory evolution
AOS	alternative optimal solutions
BiGG	biochemical genetic and genomic
BOF	biomass objective function
CDS	coding sequence
COBRA	constraint-based reconstruction and analysis
CoSy	community systems
DIET	direct interspecies electron transfer
DIP	di-*myo*-inositol 1,1′-phosphate
DMMM	dynamic multi-species metabolic modeling
EnMe	endo-metabolome
ETS	electron-transport system
ExME	exo-metabolome
FA	fraction of agreement
FBA	flux balance
FCF	flux coupling finder
FIG	Fellowship for Interpretation of Genomes
FVA	flux variability analysis
GAM	growth-associated maintenance
GDLS	genetic design through local search
GEM	genome-scale model
GENRE	genome-scale reconstruction
GOF	gain of function
GPR	gene–to–protein–to–reaction
GUR	glucose uptake rate
HGP	human genome project
HMDB	Human Metabolome Database
HT	high-throughput
I/O	input/output
IDV	isotopomer distribution vector
IEM	inborn error of metabolism
IOFA	input-output feasibility array
k-base	knowledge base
KI	knock-in
KO	knock-out
LIMS	laboratory information management system
LO	line of optimality
LOF	loss of function
LPR	ligand to protein to reaction
LPS	lipid polysaccharide

MDV	mass distribution vector
MILP	mixed-integer linear programming
MOMA	minimization of metabolic adjustment
MS	mass spectrometry
MU	modular unit
NGAM	non-growth-associated maintenance
NMR	nuclear magnetic resonance
NTP	nucleotide triphosphate
ORF	open reading frame
PCA	principal component analysis
PDB	Protein Data Bank
PFL	pyruvate formate lyase
PhPP	phenotypic phase plane
POR	pyruvate oxidoreductase
PPS	pentose phosphate shunt
PVT	pressure volume temperature
QA	quality-assured
QC	quality-controlled
RBR	RNA polymerase binding region
RBS	ribosome binding site
rFBA	regulated flux balance
ROOM	regulation off/on modification
RTS	RNAP-guided transcript segment
SKI	species knowledge index
SNP	single nucleotide polymorphism
SOP	standard operating procedure
SVD	singular value decomposition
TCA	tricarboxylic acid
TF	transcription factor
Tr/Tr	transcription/translation
TRN	transcriptional regulatory network
TSS	transcription start site
TU	transcription unit