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