

Preface

Underlying every living cell are billions of molecules interacting in a beautifully concerted network of pathways such as metabolic, signalling, and regulatory pathways. The complexity of such biological systems has intrigued scientists from many disciplines and has given birth to the highly influential field of *systems biology* wherein a wide array of mathematical techniques, such as flux balance analysis, and technology platforms, such as next generation sequencing, is used to understand, elucidate, and predict the functions of complex biological systems. This field traces its roots to the general systems theory of Ludwig von Bertalanffy and effectively started in 1952 with a mathematical model of the neuronal action potential for which Alan Hodgkin and Andrew Huxley received the Nobel Prize in 1963. More recently, the field of *synthetic biology*, i.e., *de novo* engineering of biological systems, has emerged. Here, the phrase ‘biological system’ can assume a vast spectrum of meanings: DNA, protein, genome, cell, cell population, tissue, organ, ecosystem, and so on. Scientists from various fields are focusing on how to render this *de novo* engineering process more predictable, reliable, scalable, affordable, and easy. Systems biology and synthetic biology are essentially two facets of the same entity. As was the case with electronics research in the 1950s, a large part of synthetic biology research, such as the *BioFab* project, has focused on reusable macromolecular “parts” and their standardization so that composability can be guaranteed. Recent breakthroughs in DNA synthesis and sequencing combined with newly acquired means to synthesize plasmids and genomes have enabled major advances in science and engineering and marked the true beginning of the era of synthetic biology. Significant industrial investments are already underway. For example, in 2009, Exxon Mobil set up a collaboration worth \$600 million with Synthetic Genomics to develop next generation biofuels.

Recent advances in systems and synthetic biology clearly demonstrate the benefits of a rigorous and systematic approach rooted in the principles of systems and control theory—not only does it lead to exciting insights and discoveries but it also reduces the inordinately lengthy trial-and-error process of wet-lab experimentation, thereby facilitating significant savings in human and financial resources. So far, state-of-the-art systems-and-control-theory-inspired results in systems and synthetic biology have been scattered across various books and journals from various disciplines. Hence, we felt the need for an edited book that provides a

panoramic view and illustrates the potential of such systematic and rigorous mathematical methods in systems and synthetic biology.

Systems and control theory is a branch of engineering and applied sciences that rigorously deals with the complexities and uncertainties of interconnected systems with the objective of characterizing fundamental systemic properties such as stability, robustness, communication capacity, and other performance metrics. Systems and control theory also strives to offer concepts and methods that facilitate the design of systems with rigorous guarantees on these fundamental properties. For more than 100 years, the insights and techniques provided by systems and control theory have enabled outstanding technological contributions in diverse fields such as aerospace, telecommunication, storage, automotive, power systems, and others. Notable examples include Lyapunov's theorems, Bellman's theory of dynamic programming, Kalman's filter, H^∞ control theory, Nyquist-Shannon sampling theorem, Pontryagin's minimum principle, and Bode's sensitivity integral. Can systems and control theory have, or evolve to have, a similar impact in biology? The chapters in this book demonstrate that, indeed, systems and control theoretic concepts and techniques can be useful in our quest to understand how biological systems function and/or how they can be (re-)designed from the bottom up to yield new biological systems that have rigorously characterized robustness and performance properties.

Several barriers must be overcome to contribute significantly in this exciting journey. One of these is the language barrier, e.g., what a systems theorist means by the word *sensitivity* is different from what a biologist means by it. Another one is the knowledge barrier as, traditionally, systems and control theorists and biologists are not well versed with each other's knowledge base (although that scenario is now fast changing for the better with the introduction of bioengineering courses in systems and control theory at the undergraduate and graduate levels). A third barrier is due to the sheer volume of *big data*: the European Bioinformatics Institute in Hixton, UK, which is one of the world's largest biological data repositories, currently stores 20 petabytes of data and backups about genes, proteins and small molecules, and this number is more than doubling every year. Finally, a fourth barrier comes from the effort required to produce timely contributions based on currently available models. As an example of this last barrier, the systems and control theory community could have played a greater role than it did in two of the most significant technological advances of the last 50 years: VLSI and Internet. In retrospect, besides the fact that the systems and control theorists caught on the Internet too late, by which time infrastructures based on TCP/IP were already in place, the main difficulty posed by the Internet for the systems and control theory community was a lack of *good* models of the underlying networked system. This lack-of-good-models barrier is even more daunting in biology since some of the currently available *big data* are not guaranteed to be reproducible. As Prof. M. Vidyasagar illustrates and observes in the September 2012 issue of IEEE Lifesciences, one of the major challenges to the application of systems and control theory concepts in biology comes from "the fact that many biological experiments are not fully repeatable, and thus the resulting data sets are not readily amenable to

the application of methods that people like us [i.e., systems and control theorists] take for granted.”

The chapters in this book serve to propose ways to overcome such barriers and to illustrate that biologists as well as systems and control theorists can make deep and timely contributions in life sciences by collaborating with each other to solve important questions such as how to devise experiments to obtain models of biological systems, how to obtain predictive models using information extracted from experimental data, how to choose components for (re-)engineering biological networks, how to adequately interconnect biological systems, and so on. Furthermore, and as Prof. Mustafa Khammash observes in his foreword, this research will fundamentally enrich systems and control theory as well by forcing it to investigate currently open questions that are specific to living biological systems, e.g., Why do biological systems naturally evolve the way they do? Can the evolvability of biological systems be consciously exploited for (re-)design and optimization purposes?

This book is intended for (1) systems and control theorists interested in molecular and cellular biology, and (2) biologists interested in rigorous modeling, analysis, and control of biological systems. We believe that research at the intersection of these disciplines will foster exciting discoveries and will stimulate mutually beneficial developments in systems & control theory and systems & synthetic biology.

The book consists of 12 chapters contributed by leading researchers from the fields of systems and control theory, systems biology, synthetic biology, and computer science. Chapters 1–6 focus on general mathematical concepts, methods, and tools that are currently used to answer important questions in biology. Chapters 7–12 describe various biological network modeling approaches used to untangle biological complexity and reverse-engineer biological networks from data.

- **Part I—Mathematical Analysis:** Chapters 1–6 present core mathematical concepts and methods that can be used and further adapted for solving specific problems in biology. As an example, consider the law of mass action. It has been widely used in chemistry since Guldberg and Waage formulated it in 1864. But does it have a deeper significance that is applicable outside chemistry? Likewise, reaction-diffusion systems feature in all pattern formation problems which, in turn, are significant in neuronal networks and disease phenotypes. Under which conditions is spatial uniformity guaranteed? The chapters in this part provide rigorous mathematical foundations that can be used to resolve such questions. A brief summary of each chapter is as follows.

- Chapter 1: The law of mass action is used in (bio-)chemistry to characterize and predict the behavior of interacting (bio-)chemical species. Guldberg and Waage formulated it in 1864 and it has since been built upon and widely used in (bio-)chemistry and cellular biology. To make it available for consideration by researchers in areas other than chemistry, Adleman et al. present it in a new form, viz., in the context of event systems, after solidifying its mathematical foundations.

- Chapter 2: Molecular systems often have a mathematical representation with uncertainties embedded in it. These uncertainties make predictions of the system's behavior harder. Nonetheless, it is still possible, in some scenarios, to obtain certain qualitative behavioral results that are fairly parameter independent and, instead, are a property of the system structure. Blanchini and Franco use a parameter-free qualitative modeling framework and show under which conditions behaviors such as oscillations and multi-stability are only structure dependent.
 - Chapter 3: Reaction-diffusion systems are of central importance in all applications that feature pattern formations. Aminzare et al. present conditions that guarantee the spatial uniformity of the solutions of reaction-diffusion partial differential equations. They demonstrate that these conditions can be verified using linear matrix inequalities and outline the applicability of these results in analyzing biological oscillations and enzymatic signalling pathways.
 - Chapter 4: Biologists often rely on linearized models to examine stability and on phase-plane analysis to understand the effect of parameter variations. Although useful, phase-plane analysis cannot be used to address simultaneous variations in more than two parameters. Kulkarni et al. show how multiplier theory can be used to overcome these limitations and illustrate its use via a case study of the celebrated Elowitz–Leibler oscillator.
 - Chapter 5: Modularity possibly emerged at the cellular level through natural selection and evolution. But do modules make sense in the context of metabolic networks? Goelzer and Fromion present a framework that allows a modular decomposition of steady-state metabolic networks, and show how this framework can also be used for a qualitative predictive modeling based on omics datasets.
 - Chapter 6: Biological network modeling often encounters the problem of how to deal with hidden state dynamics. Santiello et al. address the problem of predicting hidden state transitions from temporal sequential datasets (for example, EEG, EMG, MER) by developing a Bayesian detection paradigm that combines optimal control and Markov processes.
- **Part II—Biological Network Modeling:** Chapters 7–12 focus on certain techniques that can be used to obtain predictive models of biological networks. Here, the limitations of the perturbation methods used to generate the data, the vast amount of available data (which does not necessarily correlate with the amount of *information* they contain), hidden states, measurement noise, and other factors combine to render this broad area of research one of the greatest scientific and technological challenges of today. The chapters in this section summarize some of these challenges and present architectures that constitute an important step in arriving at a definitive solution. Somewhat similar, but less complex system identification problems have been encountered and resolved in systems theory and computer science over the last decades. Can these techniques and the insight they provide be useful in biology? To answer this

question, it is crucial to understand the advantages and limitations of each particular technique. The set of chapters collated in this part aim to highlight the current state of the art for biological network modeling and the advantages and limitations of the presented approaches. A brief summary of each chapter is as follows.

- Chapter 7: In metabolic networks, the metabolite dynamics evolve on much shorter timescales than their catalytic enzymes. Kuntz et al. show how such timescale separation can be exploited using Tikhonov’s theorem for singularly perturbed systems to derive reduced models whose behaviors are guaranteed to remain quantifiably close to those of the non-reduced models. They illustrate this approach by applying it to an example of genetic feedback control for branched metabolic pathways.
- Chapter 8: A central theme in complex network theory, popularized by the study of small-world and scale-free networks at the turn of the last century, is the study of biological networks using various metrics. In this chapter, Roy discusses the utility of various network metrics as well as the need to go beyond fundamental metrics, such as node degree, to better understand how an organism’s phenotype is encoded by its network topology.
- Chapter 9: Even though most of the complex real-world systems exhibit nonlinearities, linear models serve as a useful first order approximation. Carignano et al. present a detailed exposition on how linear system identification techniques can be used to obtain causal relationships between biomolecular entities.
- Chapter 10: Fisher and Piterman discuss how ideas from computer science can be useful for *model checking* in systems biology. They present a methodology to analyze biochemical networks, and specifically a method to test for a faithful reproduction of biological interactions that are known a priori as well as to identify interactions that are not known a priori.
- Chapter 11: Bussetto et al. discuss objective-specific strategies for designing informative experiments in systems biology. Following a formal description of the task of experimental design, they illustrate the use of Bayesian and information-theoretic approaches to design experiments in systems biology.
- Chapter 12: Today, there is a critical need for new methods that rapidly transform high-throughput genomics, transcriptomics, and metabolomics data into predictive network models for metabolic engineering and synthetic biology. In this chapter, Chandrasekaran describes the state of the art of these methods and explains an approach for this purpose called Probabilistic Regulation of Metabolism (PROM).

The burgeoning fields of systems biology and synthetic biology have thrown up a very large number of interesting research problems. As the pre-eminent computer scientist Donald Knuth put it, “biology easily has 500 years of exciting problems to work on.” The chapters in this book address but a small fraction of these interesting challenges. Nevertheless, we believe this book can serve as a

good introduction on some of the currently open problems and on some of the state-of-the-art concepts and techniques available to propose solutions to such problems.

We are very grateful to all authors for their invaluable time and contributions and to Prof. Mustafa Khammash (ETH Zürich) for his stimulating foreword. We are also grateful to our institutions: University of Minnesota (Minneapolis, USA), Imperial College (London, UK), and Indian Institute of Technology Madras (Chennai, India) for their support and for providing a stimulating work environment. Finally, we thank and acknowledge the financial support of our respective funding agencies: the National Science Foundation, the UK Engineering and Physical Sciences Research Council, and the Ministry of Human Resource and Development of the Government of India.

Minneapolis, MN, USA, September 2013
London, UK
Chennai, India

Vishwesh V. Kulkarni
Guy-Bart Stan
Karthik Raman

A Systems Theoretic Approach to Systems and
Synthetic Biology I: Models and System
Characterizations

Kulkarni, V.; Stan, G.-B.; Raman, K. (Eds.)

2014, XVIII, 332 p. 84 illus., 41 illus. in color., Hardcover

ISBN: 978-94-017-9040-6