Preface

System Modeling: Why and Why Now?

Vipul Periwal, Zoltan Szallasi and Jörg Stelling

Introduction

Biology is the study of self-replicating chemical processes. Biology is the study of systems accurately transmitting a genetic blueprint. Biology is the study of complex adaptive reproducing systems.

What is systems biology if all definitions of biology implicitly or explicitly refer to the study of a whole object, whether it is a virus, a cell, a bacterium, a protozoan or a metazoan? We treat systems biology as the *quantitative* study of biological systems, aided (or hindered) by technological advances that both permit molecular observations on far more inclusive scales than possible even 15 years ago, and permit computational analysis of such observations. Thus, for the purposes of this book, systems biology is the promise of biology on a larger and quantitatively rigorous scale, a marriage of molecular biology and physiology. Concretely, this defines the focus of the book: data-centric quantitative modeling of biological processes and systems.

Biology is an experimentally driven science simply because evolutionary processes are not understood well enough to allow theoretical advances to rest on terra firma. Systems biology is experimentally driven, computationally driven, *and* knowledge driven. It is experimentally driven because the complexity of biological systems is difficult to penetrate without large-scale coverage of the molecular underpinnings; it is computationally driven because the data obtained from experimental investigations of complex systems need extensive quantitative analysis to be informative; and it is knowledge driven because it is not computationally feasible to analyze the data without incorporating all that is already known about the biology in question. Furthermore, the use of data, computation and knowledge must be concurrent. Available knowledge guides experiment design, novel knowledge is generated by the computational analysis of new data in light of available knowledge, and the cycle repeats.

The difference between knowledge and data is central to understanding the underpinnings of systems biology. The sequencing of whole genomes is a good example. Any given genome is data. Without extensive analysis, it is just as uninformative about biological processes as a photograph of the night sky. First steps in transforming a genome into knowledge include identifying genes, identifying transcription factor binding sites, finding the transcription factor complexes that control the expression of the genes, and finding the chromatin structure in the cell being studied, to determine which genes are accessible for transcription. While this is wildly optimistic in terms of the knowledge that can be extracted from the genome data, it is still nowhere close to the level of understanding required to make predictions about the response of an organism to a specific stimulus. A reductionist approach to biology is bootless because complex adaptive systems are inherently *nonlinear*, so their behavior is well summarized by the statement: the whole is more than the sum of the components.

Handicapping the Bout

From a quantitative perspective, there are striking features of biological dynamics that make analysis challenging:

1. Large range of spatial scales
2. Large range of temporal scales
3. A lack of separation between responses to external stimuli versus internal programs
4. Multiple functionalities of constituents
5. Multiple levels of signal processing
6. Incomplete evolutionary record
7. Wide range of sensitivities to perturbations
8. Genotypic variation

None of these challenges is an absolute barrier to progress. Nevertheless, these challenges must be addressed to make real progress.

From an experimental perspective, the challenges of biology are better understood:

1. Coverage in terms of components and interactions
2. Reproducibility
3. Spatial resolution
4. Temporal resolution
5. Cross-validation
6. Combinatorial perturbations
7. Accuracy

From a knowledge perspective, there are four central problems:

1. Find an appropriate level of abstraction for a given analytic problem.
2. Find a common basis to relate knowledge gained using different experimental techniques on the same system.
3. Find a common basis to relate knowledge gained from the same experiment on different model systems.
4. Incorporate knowledge incrementally as new data is analyzed.

Taking all these difficulties together, it is not surprising that researchers traditionally have considered the study of biological systems rather resistant to quantitative approaches. It is, therefore, worth pointing out to skeptics that in some cases thorough quantitative analysis has produced insights into or explanations of biological phenomena that would have been impossible without the application of advanced mathematical tools. Various chapters in this book will discuss a great variety of, often counterintuitive, examples. For instance, the advantages of a more extensive mathematical analysis over simpler approaches are emphasized in chapter 8 (pp. 170–173). When circadian oscillators are analyzed by formal logic, the traditional analytical tool of molecular biology, or by macroscopic descriptors such as differential equations, the experimentally observed behavior cannot be reconstructed from the molecular machinery. Stochastic analysis, however, demonstrates how, by random fluctuations, the system escapes the macroscopic point-attractor and thus oscillatory behavior is maintained. Examples such as this will probably contribute to the long-awaited common ground for discussions between biologists and quantitative scientists. The mutual suspicion on both sides, which has been difficult to overcome by intellectual curiosity alone, will probably be eliminated by the mutual need for each other's expertise.

"My Complications Had Complications"

The goal of systems biology is a predictive understanding of the whole. If the whole is more than the sum of its parts, it follows that acquiring a catalog of all the parts is *not* necessarily the first order of business. In a caricature, there are two avenues of attack possible: either one focuses on subsystems governing a specific function in arbitrary conditions and gains a predictive understanding of the system, one subsystem at a time, or one focuses on the system in a restricted set of conditions and gains an understanding by gradually increasing the set of conditions and, as required, the level of detail in the model of the system. The analogy is with molecular biology in the former approach and with physiology in the latter.

The modeling associated with each approach is distinct. In the molecular biology type approach, the aim is to go beyond traditional pathway-centric points of view and deal with the challenges of feedback loops formed either directly or indirectly due to interactions with other pathways. In the physiology type approach, interactions between the components in the model are added as needed to maintain contact with the experimental data. The components in this approach are not necessarily directly related to biochemical species. Eventually, these bottom-up and top-down approaches should meet. However, each has its own strengths and weaknesses and they complement each other.

Why Read This?

The importance of feedback loops and crosstalk in almost all facets of biological systems has been apparent for several decades. The cell cycle control circuitry or the developmental programs in bilaterians are prime examples of this. The ability of cancerous cells to evade targeted therapies results largely from biological systems having evolved in ways that place a premium on robustness and adaptability. Such properties, as yet only nebulously defined, are not localizable to a small set of interactions. They reside in the network as a whole, as has been clearly demonstrated in predictions on metabolic networks.

Modeling biological systems faces the challenge of appropriate abstractions—levels on which to focus, and details to be left out. For instance, molecular biology abounds with mechanistic analogies, but on a more detailed level often the underlying interactions are driven by chemistry. This makes modeling subtle since statistical biases are often the driving force in what superficially appears to be a mechanical process, for example, chemotaxis. At what level does such detail become relevant, and at what level can one ignore it? This is not *a priori* obvious, and one needs rigorous approaches to model parsimony to answer such questions. Indeed, the answer to the model selection question depends to a great extent on the predictions required. This is an important point in all biological modeling: The model, its purpose, and the experimental data are intimately related. A model that predicts hepatic glucose uptake precisely but insulin levels with greater uncertainty is not a useful selection if the only measurement available is insulin levels.

There are two main approaches to computational analysis of biological data. The *causal* approach makes concrete deterministic or stochastic models (differential equations, stochastic differential equations, Boolean networks, et cetera) of biological processes. The *probabilistic* view is associated with probabilistic inference approaches, using pattern recognition or learning algorithms (such as neural networks and graphical models) for analysis of data from large-scale experimental methods. These two approaches rest on a large part of applied mathematics (including numerical integration, optimization, interpolation, and control theory) and computer science (search theory, coding theory, and database design). This breadth necessitates collaborations between people with diverse backgrounds, but an inadequate understanding of the limitations and applicability of techniques and concepts from different fields hinders such collaborations. The background information required makes biological modeling a difficult task, but the real challenge remains that of making computational models *effective* and *efficient* representations of biological systems.

What's Included and What's Not

This book starts with generalities and progresses towards practicalities. Thus, the first section is conceptual, with attempts to define the role of modeling in biology, as well as attempts to cut through the miasma that surrounds the use of the terms *robust*, *complex*, *adaptive*, and *module* in the systems biology literature. As will be evident, these are important notions that need much further work to crystallize to the point where they can be assigned the honorific *concept*. Nevertheless, these terms may ultimately be quantitatively used as concrete guiding principles in modeling.

The next section provides introductions to general approaches to making models of biology: qualitative models, constraint-based models, dynamical systems based on differential equations, and stochastic models, as well as models with spatial structure. The other side of the modeling coin, probabilistic inference aimed at inference from large-scale data sets, is also introduced. The section proceeds from relatively simple towards mathematically more demanding approaches. Although each chapter tries to convey its central messages in an intuitive as well as in a mathematically rigorous way, readers arriving from biology will have to realize that each method has a certain minimum difficulty level associated with it. While ordinary differential equation–based or qualitative models can be quite readily introduced in an intuitive manner, stochastic or spatial modeling cannot be described in simple terms and require an appropriate level of background in quantitative sciences. Key applications of the various modeling approaches are also widely covered. Taken together, this section will provide the reader with an overall impression of the relationship between the potential utility of quantitative approaches and their associated analytical cost.

Reality bites. And models model biological reality. The section that follows next contains introductions to the data that is available for systems biology and the caveats that go with the data. It also contains introductions to inferring model architecture from data, using control theory in models, and studying synthetic gene networks. The antidote to these computational limitations is multi-level modeling, and this is also introduced in this section. Limitations in observability, accuracy, and coverage of biological data are widely recognized. One of the goals of this section is to guide the readers through various data interpretation methods while emphasizing what the data will or will not allow in terms of quantitative analysis.

The last section of the book contains the computational issues and techniques for practical application of the preceding approaches: numerical methods for simulating biochemical systems, and the software infrastructure for representing models in a reusable and exchangeable manner. Biological data quality is not the only obstacle systems biology is facing. The various numerical methods also have their well known strengths and limitations and these should be considered when designing experiments and their associated models. For instance, computational limitations form barriers to increasing model size arbitrarily.

The book ends with an eclectic list of the software tools that the contributing authors of this book find useful.

While this book contains a plethora of approaches to biological modeling, we are keenly aware that there are many that we have not covered. For instance, we have eschewed much discussion of pattern recognition because this is only really useful when combined with domain specific biological knowledge—for which no general technique exists. Likewise, we do not cover approaches such as neural networks or Petri nets that have either limited application in systems biology so far, or are problematic regarding model interpretation. Our attempt has been to provide broad basic coverage of fundamental approaches and techniques. In our view, picking some of the techniques introduced in this book and combining them artfully leads to almost complete coverage of modeling in systems biology.

Enjoy

Systems biology is an approach to quantitatively understand biological systems that attempts to embrace the complexity of life as a fact of life. There is no hope of understanding biological systems at the predictive level required for disease detection, prevention, or cure other than by this means. Nevertheless, it would serve us well to temper Burnham's maxim of grand thinking, "Make no little plans . . ." with the story of the emperor's new clothes.