Preface

*The intention and the result of a scientific inquiry is to obtain an understanding and a control of some part of the Universe - Arturo Rosenblueth and Norbert Wiener*

While the advances in technology can help in obtaining more and better quality data about the cellular processes and signals, the design and practical applications of predictive computational models of genomic regulation continue to face major challenges. This edited book draws attention to several of them and outlines important future research directions.

One can think of a *Gene Regulatory Network (GRN)* as a network of relations among strands of DNA (genes) and the regulatory activities associated with those genes (Dougherty & Braga-Neto, 2006). This general definition allows for many mathematical (usually dynamical) systems to be called GRNs. The goodness of each such model is evaluated using several important criteria: the level of description of the biochemical reactions involved, complexity of the model, model parameter estimation, and its predictive power. There have been many attempts to model the structure and dynamical behavior of GRNs, ranging from deterministic with discrete state and time space to fully stochastic with continuous state and time space (de Jong, 2002). The so called ‘central dogma’ of molecular biology (Crick, 1970) implies that genes communicate via the proteins they encode. All of the stages of protein production, transcription and translation, are controlled by a multitude of biochemical reactions, and are influenced by both internal and external to the cell factors. This perspective suggests that the expression of a given gene $i$, i.e. the quantity of either protein or messenger RNA, should be considered as a random function $X_i(t)$ of the cell’s internal and external environments. Thus, if one wants to study the dynamical behavior of a GRN, one must design a mathematical model for the gene-expression vector $X(t) = (X_1(t), X_2(t), \ldots, X_n(t))$ for the $n$ genes that are selected to constitute the network. The stochastic master differential equation model appears to provide the most detailed description of the dynamics of $X(t)$. In principle, it could include all of the information about the biochemical processes involved in gene regulation. At the same time, the estimation of its parameters can-
not be done without large amount of reliable time-series data. As a result, one is often forced to take a more pragmatic approach and look for simpler models for the dynamics of the gene-expression vector. One of the most extreme simplifications is the Boolean Network (BN) model, originally proposed by Kauffman (1969). The BN model had been successfully used in physics before attracting the attention of the biology community. The initial application of the model was to study the evolution of ensembles of networks that were restricted to a specific type of fitness landscape. The BN model is based on the observation that during the regulation of its functional states the cell often exhibits switch-like behavior. Work using the NCI 60 Anti-Cancer Drug Screen has demonstrated that Boolean logic type interactions can be detected in gene expression data (Pal, Ivanov, & Dougherty, 2005). While there are instances in gene regulation where the Boolean logic is the appropriate level of description of the interactions – for instance, when transcription factors have to form a complex that binds to the cis-regulatory DNA to activate transcription, one should keep in mind that coarse discrete models cannot capture the details of the biochemical reactions involved in those processes, and models like the Chemical Master Equation (CME) should be used, instead.

In recent years, it has become increasingly clear that systems biology in general and genomics in particular, underscores the role played by stochastic nonlinear dynamical systems which in turn, precludes uncritical reductionist realism. Moreover, systems biology and genomic regulation cannot be considered without the appreciation of their relation to translational science. Translational science transforms a mathematical model, whose purpose is to provide a predictive conceptualization of some portion of the physical world, into an intervention (action) strategy, again in the physical world. This ultimately leads to an appropriate framework addressing the specific challenges associated with the goals for controlling cellular regulatory activities. Thus, a properly designed translational mathematical system guides:

1. The scientist in building a model applicable to the specific problem;
2. The engineer in studying costs and benefits of actions;
3. The technologist in devising devices or treatments.

This book provides a compilation of recent and emerging research topics that address these important points. The first chapter concerns a broad overview of different GRN models with emphasis on model inference and prior knowledge incorporation. The problem of prior knowledge incorporation into the inference procedures for GRNs is a fundamental one in the field of genomic signal processing. The modern high-throughput technologies for acquiring genomic data and the intrinsic high-dimensional space of variables involved in genomic regulation leave one with a virtually unbounded model space where it is often insufficient to configure
mathematical theories in the absence of prior knowledge. Therefore, a scientist must come to the table with sufficient knowledge of the problem to formulate a small class of models for which it remains only necessary to utilize data to estimate some set of parameters to instantiate the model or the model class. This is imperative for the typically ill-posed GRN inference problem.

The second and third chapters of the book discuss the relationships between various flavors of GRN models and the basic Chemical Master Equation (CME) model. The topics of model approximations, their properties, advantages and disadvantages as well as the means to infer them from data, including rare-event sampling are presented in detail. Several examples of applications to real biological GRNs are provided to illustrate the general discussion.

The reader can notice that there are clear overlaps in the topics discussed in the first three chapters. However, the editorial team decided not to attempt to reduce those because on one hand, each one of these chapters presents the authors' unique perspective of the modeling issues and on the other hand, that allows for preserving the consistency within each separate chapter. In addition, the reader might find it useful and easier to read self-contained chapters.

The fourth and fifth chapters focus on the second major translational point: the control of the model dynamics and the associated costs and benefits of the control actions. Structural stability is one of the central concepts in the theory of dynamical systems. It describes persistent behavior that cannot be destroyed by small changes to the system. As real GRNs are capable of maintaining metabolic homeostasis and stable developmental program in the face of a changing environment, they certainly possess structural stability. These two chapters discuss these important topics in the setting of Markovian GRNs. While this specific class of models is clearly a simplification of the detailed stochastic differential equation GRN model, it is rich enough to model such important notions as cellular states (Huang, 1999, 2001) and cellular types (Kauffman, 1993). Moreover, the stochastic nature of the Markovian GRNs allows for incorporation of uncertainty on various levels. First, due to biological variability, gene expression is inherently stochastic. Second, the complex measurement process, e.g. the microarray or next-generation sequencing signal acquisition and processing, create experimental noise that has to be taken into account during the inference of the model. All of these combined with the presence of latent or unobservable variables such as proteins or environmental conditions present one with the problem to infer deterministic predictor functions under uncertainty which is one of the main reasons why the Markovian GRN models have attracted significant attention among the research community. As the main objective of GRN modeling is to predict the dynamical behavior of the underlying real regulatory system and ultimately provide the effective intervention strategies for prevention and control of disease that is associated with altered genomic regulation, the control policies are often evaluated by their effects on the steady-state probability distribution of the model.
Chapter 6 presents a convincing case why one cannot expect to successfully use uncritical reductionist realism when modeling real gene regulatory interactions. The author argues convincingly why multiple regulators are required for the transcription of each gene. The implications of the presented evolutionary and computational considerations are clear: modeling of gene regulation leads, with some notable exceptions, to highly complex models. It is important to point out that the term complexity is overloaded with many different meanings depending on the field of study. Chapter 6 emphasizes that GRNs are composed of many parts that interact with each other and those interactions involve feedback loops and are stochastic from the observer’s point of view. Thus, GRN satisfy the definition of a complex system and this observation brings up the need for complexity-reducing strategies. One of the aspects of such strategies in this setting is to find subsystems with a fewer number of interacting genes and with simpler rules of interactions between them that preserve, to a degree, critical properties of the whole system. Unfortunately, this important topic – how to optimally compress a large and complex GRN model remained outside the scope of this book.

Chapter 7 considers a specific application of GRN modeling – deciphering of the regulatory machinery in early embryos using whole-transcriptome data. This particular application of GRNs is especially important in terms of translational science given the well documented influence of epigenetic and environmental factors on the regulatory processes guiding the developmental stages starting from an embryo to an adult organism.

Chapter 8 complements the discussion presented in the previous two chapters and proposes a framework for prioritizing the transcription factor binding sites for multiple co-expressed gene sets. The framework uses Lasso multinomial regression modeling and ultimately leads to constraints that reduce the size of the search space for the reverse engineering of GRNs.

Chapter 9 provides a description of an impressive combination of specialized genomic, image acquisition and processing devices and techniques that allow for observing and computational modeling of drug efficacy on specific cell regulatory pathways over time and for the same cell population. The presented technology and modeling approaches underscore the importance of close collaborative interactions between scientists and engineers of various backgrounds. The discussion in the chapter serves as an example of a properly functioning relationship, where the scientist does not hand the engineer or statistician a set of data and asks them to find something in it; instead, assuming a translational goal, the enterprise is guided by that goal and it leads to a carefully designed experiment. Moreover, the authors present a convincing case for the need of time series data on the same cell population or even individual cells in that population if one would hope for a GRN model with sufficient predictive capability.
Predictive computational modeling of genomic regulation has the potential to unravel the mechanisms of cellular functions from a systemic perspective. There are several important goals of the modeling process:

- To infer optimal models from data using prior knowledge as a constraint in the optimization process.
- To characterize the dynamical behavior of the real GRN in terms of long-run probability distributions of the model and the structure of its attractors and their basins.
- To investigate how structure determines dynamics and the restrictions imposed by a specific dynamical behavior on the structure of the GRN.
- To characterize classes of intervention strategies for control of the dynamical behavior of the system.

We hope that this edited book will provide the reader with a good mix of theoretical and practical topics representing the goals of mathematical and computational modeling of genomic regulation in the context of translational science. The book could serve as a guide to some of the important and highly relevant current and emerging research trends in the analysis and modeling of GRNs.

REFERENCES


