Chapter 15
Abstraction Methods for Analysis of Gene Regulatory Networks

Hiroyuki Kuwahara
Carnegie Mellon University, USA; Microsoft Research - University of Trento CoSBi, Italy

Chris J. Myers
University of Utah, USA

ABSTRACT

With advances in high throughput methods of data collection for gene regulatory networks, we are now in a position to face the challenge of elucidating how these genes coupled with environmental stimuli orchestrate the regulation of cell-level behaviors. Understanding the behavior of such complex systems is likely impossible to achieve with wet-lab experiments alone due to the amount and complexity of the data being collected. Therefore, it is essential to integrate the experimental work with efficient and accurate computational methods for analysis. Unfortunately, such analysis is complicated not only by the sheer size of the models of interest but also by the fact that gene regulatory networks often involve small molecular counts making discrete and stochastic analysis necessary. To address this problem, this chapter presents a model abstraction methodology which systematically performs various model abstractions to reduce the complexity of computational biochemical models resulting in substantial improvements in analysis time with limited loss in accuracy.

INTRODUCTION

Thanks to advances in technologies, in genetic regulatory networks—where, for instance, high-throughput gene expression analysis methods are available and a vast amount of quantitative data has been collected—the information required for building quantitative models of gene regulatory networks can be obtained. The most exact way to simulate a quantitative model of a molecular system is *molecular dynamics* where movements of every molecule in the system are tracked (Gillespie, 2005, 2007). The system state of molecular dynamics is the positions and velocities of every molecule in the system where the dynamics

DOI: 10.4018/978-1-60566-685-3.ch015

of the system state are described by capturing every movement and every collision of molecules in the system. While this approach can show the time evolution of species' populations as well as the spatial distribution of each species, acquiring such detailed knowledge and performing such computationally expensive simulations is typically infeasible. By making the well-stirred assumption, the spatial property of a system can be abstracted away, overriding the system state to be simply the populations of species in the system. While this assumption greatly simplifies the complexity of models, it adds uncertainty in the time evolution of the system owing to the insufficient knowledge of the system descriptions that the very assumption precludes.

Traditionally, gene regulatory networks are modeled and analyzed within the continuous-deterministic, *classical chemical kinetics* (CCK) framework based on the *law of mass action* where the dynamics of a well-stirred system are described by a set of *ordinary differential equations* (ODEs). Although such treatment can be justified when the molecular populations are very large—and hence a CCK analysis may provide one of the most efficient approaches to estimate the time evolution of a system—the limitations of the CCK analysis have been broadly accepted (Arkin et al., 1998; Gillespie, 1992a, 2000; Elowitz et al., 2002; Rao et al., 2002; Samoilov et al., 2005; Samoilov and Arkin, 2006). In particular, given the same initial condition, the CCK analysis of biochemical systems always produces the same results as it neglects uncertainty in system dynamics. Furthermore, many regulatory components (e.g., DNA, RNA, proteins) in biological systems can be present in amounts too small to simply neglect the effects of inherent fluctuations (McAdams and Arkin, 1997; Golding et al., 2005; Raser and O'Shea, 2005; Pedraza and van Oudenaarden, 2005; Cai et al., 2006; Newman et al., 2006).

In order to more accurately predict the temporal behavior of gene regulatory networks, the *stochastic chemical kinetics* (SCK) framework can be used (Gillespie, 2005, 2007). Assuming that the system is spatially homogeneous, this SCK approach describes the time evolution of a biochemical system at the individual reaction level by exactly tracking the quantities of each molecular species and by treating each reaction as a separate random event. One consequence of SCK is a stochastic process description of the system that is analytically governed by the *chemical master equation* (CME) (McQuarrie, 1967; Gillespie, 1992b). However, directly obtaining the solution of the CME of any realistic system, either analytically or numerically, is not feasible due to its intrinsic complexity.

Instead of attempting to solve the CME, exact numerical realizations of a SCK model via Gillespie's *stochastic simulation algorithm* (SSA) (Gillespie, 1976, 1977), which is derived from the same premise as the CME, are often used to infer the temporal system behavior with a much smaller memory footprint. Unfortunately, the computational requirements of the SSA can be substantial due largely to the fact that it not only requires a potentially large number of simulation runs in order to estimate the system behavior at a reasonable degree of statistical confidence, but it also requires every single reaction event to be simulated one at a time.

Ultimately, given the substantial computational requirements of stochastic simulations, abstraction is absolutely essential for efficient computational analysis of complex gene regulatory networks. For such networks, any applications of the all-inclusive, low-level, quantitative models are largely impractical because of high computational demands, while the use of entirely high-level qualitative representations is typically inadequate owing to the substantial dynamical and functional complexity they can manifest. Therefore, a search for some intermediate level of abstraction becomes necessary. This, however, frequently presents a problem: while most abstractions used in modeling of biochemical networks have traditionally been implemented manually on a mechanism-by-mechanism basis, doing so accurately in general settings is a tedious and time-consuming process, which is highly susceptible to errors during model translation and transformation.