

REVIEW ARTICLE

Toward Computational Systems Biology

*Lingchong You**

*Division of Chemistry and Chemical Engineering, California Institute of Technology,
Pasadena, CA 91125*

Abstract

The development and successful application of high-throughput technologies are transforming biological research. The large quantities of data being generated by these technologies have led to the emergence of systems biology, which emphasizes large-scale, parallel characterization of biological systems and integration of fragmentary information into a coherent whole. Complementing the reductionist approach that has dominated biology for the last century, mathematical modeling is becoming a powerful tool to achieve an integrated understanding of complex biological systems and to guide experimental efforts of engineering biological systems for practical applications. Here I give an overview of current mainstream approaches in modeling biological systems, highlight specific applications of modeling in various settings, and point out future research opportunities and challenges.

Index Entries: Systems biology; mathematical modeling; gene networks; deterministic simulation; stochastic simulation; biological databases.

INTRODUCTION

Advances in experimental and computational technologies for biosciences have been revolutionizing biological research. The last several decades have witnessed the development and maturation of several remarkable experimental techniques, such as DNA sequencing technique, DNA microarray (1), and large-scale two-dimensional protein gel electrophoresis (2). Emerging from the applica-

tion of these technologies is a new mode of biology, the systems biology, which emphasizes a holistic understanding of how biological systems function (3,4).

As demonstrated by the successful sequencing of more than 1000 genomes of natural plasmids, organelles, viruses and viroids, bacteria, plants, and animals, including mouse (5) and human (6,7), there are few, if any, technological hurdles in obtaining the genetic information of virtually any organism. In addition to its implications for practical applications, such information promises to bring us closer to a complete understanding on how the genetic information stored in a genome determines the behaviors or characteristics (i.e., the phenotype)

*Author to whom all correspondence and reprint requests should be addressed. E-mail: you@cheme.caltech.edu

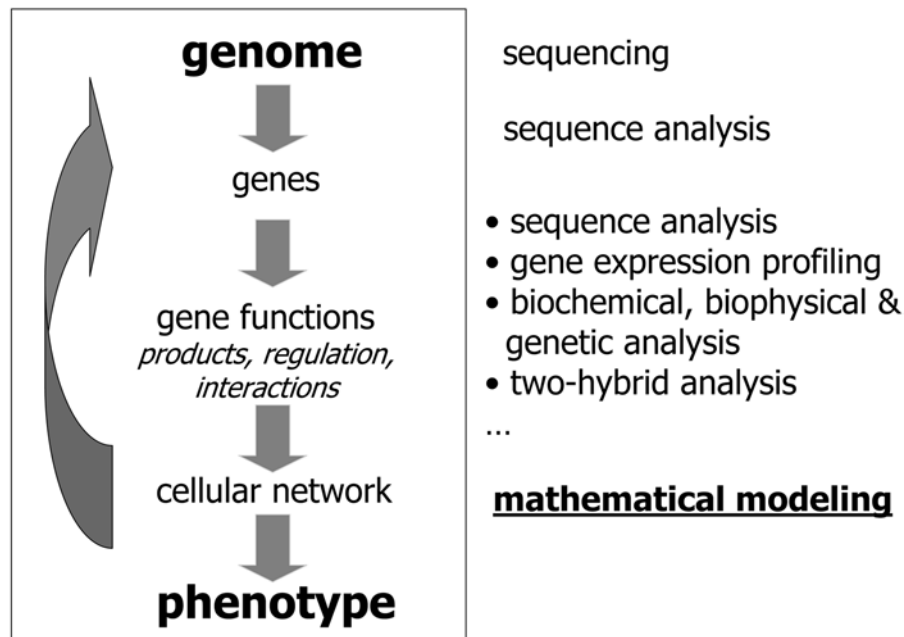


Fig. 1. A new mode of biology research. Sequencing of various genomes has led to the question as how these genomes “program” cellular behaviors, or the phenotype, in a particular environment. Different intermediate levels of understanding need to be established to make this connection (left panel). Many computational and experimental tools (right panel) are required in achieving such understanding. Mathematical modeling will play a key role in integrating fragmentary information into a coherent whole. A fundamental goal of modeling is to explain existing data and to predict system behaviors. Every model should be examined in the context of experiment whenever possible. Iterations of model prediction, comparison with experiment, and model revision in the light of new data and mechanisms will constantly improve our understanding of the underlying system. (See text for details.)

of an organism or a cell in a particular environment. As shown in Fig. 1, however, much remains to be done to really make the link between a genome and the resulting phenotype(s). Logical next steps in this odyssey are to identify the genes in a genome and to determine their functions, in particular, by elucidating what products these genes produce and how these products interact with one another. For any particular biological system, these downstream analyses can be orders of magnitude more complex than sequencing the genome. Indeed, they require a wide spectrum of tools to characterize individual gene products by using biochemical, biophysical, or genetic techniques or a large set

of such molecular components by profiling gene expression at the mRNA level—using DNA microarray (1)—or at the protein level—using two-dimensional protein gels (2) or mass spectrometry (8–10). In addition, other high-throughput techniques, such as yeast two-hybrid analysis (11), have been successfully applied to study interactions between gene products at a large scale (12,13).

Yet, large-scale characterization of the components of biological systems is not the ultimate goal, and it should not be. Supposing that someday we succeed in characterizing the function of all genes in a cell by identifying all the gene products and constructing the interaction

map among these gene products, will we have completed the link between the genome and cellular behaviors? Partly. Extensive as it is, all this information is still local and fragmentary. Knowing what a cell is composed of and how each component works does not necessarily mean understanding how the cell as a whole works. For example, understanding what parts a car is composed of and how each part works does not mean that we would understand how the car itself works. To be confident that we understand how the car works, we should be able to put the parts back together and demonstrate that the car works. Similarly, to understand how a cell function as a whole, we will need to integrate our understanding of the parts and see whether or to what extent these pieces of understanding will coherently predict cellular behaviors. For any system, the process of integrating and analyzing the information on individual pieces can be loosely defined as “modeling.”

Regardless of its formulation, a model by definition represents an integration of the knowledge of the underlying system. It is based on answers to questions such as: what components is the system composed of? And how do they interact? Note that the integration of fragmentary information is a critical step for achieving global, system-level understanding. As discussed in detail in the section Specific Applications, this integration process helps to reveal features not easily recognizable by examining the constituent parts. For example, information regarding some individual pieces may be inconsistent with other information, but we often will not know the inconsistency until we put these pieces together. The fundamental goal of a model is to explain existing data and to predict system behaviors. Oftentimes, model predictions will deviate from the experiment, which calls for re-evaluation of the knowledge integrated into the model or for additional experiments. Either way, our understanding will improve by going through iterations of model prediction, comparison between the model prediction and experiment, and model refinement (see Fig. 1).

The wide appreciation and interest in modeling biological systems is evidenced by the publication of not only countless modeling studies, but also many excellent reviews that discuss various aspects of modeling (4,14–23) (note “A Brief Guide to Recent Reviews” in ref. 17). Intending to give a comprehensive tutorial on modeling in biology, here is provided an overview of mainstream modeling approaches and discuss their application domains. To achieve this goal, many details are omitted of these modeling approaches but instead their connections are highlighted between each other so that they can be examined in a coherent manner. Focusing on kinetic models, recent applications are discussed of modeling in enhancing our understanding of natural or engineered biological systems. Although modeling is the focus issue of this review, I strive to put the discussions in a broader context of systems biology, emphasizing how modeling may facilitate system-level study or understanding of complex biological systems. Finally, this article closes with some future opportunities and challenges for the mathematical modeling community.

MODELING WITH VARYING RESOLUTIONS

Qualitative Models

Depending on its specific objectives, a model may involve details at different levels. One of the most profound biological models is the central dogma of molecular biology, which describes the basic information transfer process from DNA to RNA and from RNA to protein (Fig. 2). Although omitting many details involved in the individual steps—for example, the binding of RNA polymerases to the DNA sequence during transcription—the central dogma summarizes decades of endeavor by biologists that led to the elucidation of this process (24). Reducing complex biological processes into one dimension, it has served as an elegant conceptual framework for the establishment of molecular biology. As



Fig. 2. The central dogma of molecular biology. The central dogma states that genetic information can be perpetuated or transferred, but the transfer of information into protein is irreversible (151).

with all other models, the central dogma as depicted in Fig. 2 only reflects partial truth, and it evolves. Decades of biological research has tremendously enriched the content of the model—a more accurate representation of the relationship, although still omitting details, would include the replication of DNA, reverse transcription of RNA to DNA, replication of RNA, and catalyzing role of the protein in all these processes.

The central dogma exemplifies probably the most intuitive models that biologists have been using: diagrams. Diagrams have been widely used and will probably continue to prevail in biology textbooks and the literature. Recently, there have been significant efforts to formalize the conventions of drawing diagram models (25). Diagrams have served as tremendous visual aids in understanding the biological processes and for formulating new hypotheses to be tested experimentally. In a closely related area, several groups have analyzed large-scale metabolic networks based solely on the connectivity of network components (26–33). Unlike conventional diagrams that involve perhaps dozens of components at most, wiring of thousands of components (often proteins) in these large systems makes little sense to the naked eye. Yet computational analysis has provided insights into some global, usually topological properties of metabolic networks, such as degree of network connectivity, clustering of network components, and lengths of biochemical pathways (26–31). Recently, this statistical approach has been applied to reveal potential cellular motifs in large metabolic networks (32,33). Despite their intuitiveness (excluding computer-generated diagrams of large metabolic networks), the

drawback of diagrams is that they are descriptive only, and cannot predict in a quantitative and sometimes not even in a qualitative fashion behaviors of a given system.

The simplest approach to characterizing the *dynamics* of biological networks is a Boolean model, which is often applied to studying gene networks (34–36). In this paradigm, each player (e.g., a gene) has two states, *on* and *off*; the system of interest is represented as a logic network, and the dynamics describe how genes interact to change one another's states over time. A Boolean model is advantageous in its simplicity and it does not require detailed data on how cellular components interact. Despite their apparent simplicity, Boolean models can provide many insights into the qualitative behavior of the underlying system. For instance, Kauffman has successfully employed Boolean models to explore self-organization phenomena and their implications in evolution (37). Yet, Boolean models are often overly simplified and tend to give ambiguous predictions on system behaviors (38). For example, consider a simple circuit in which a component *S* negatively regulates its own accumulation. In the Boolean formulation, this process can be modeled as: $S(T + 1) = \text{NOT } S(T)$, where $S(T)$ means the state of *S* at time step *T*. This model predicts that the level of *S* will oscillate between 0 and 1, irrespective of the initial condition. However, in reality, such a system often leads to a steady state (e.g., see ref. 39) unless there is a significant time delay in the self-regulation. In addition, some ambiguity is evident in the simple model: the function “NOT” could mean that *S* slows down its own synthesis or that *S* facilitates its own degradation.

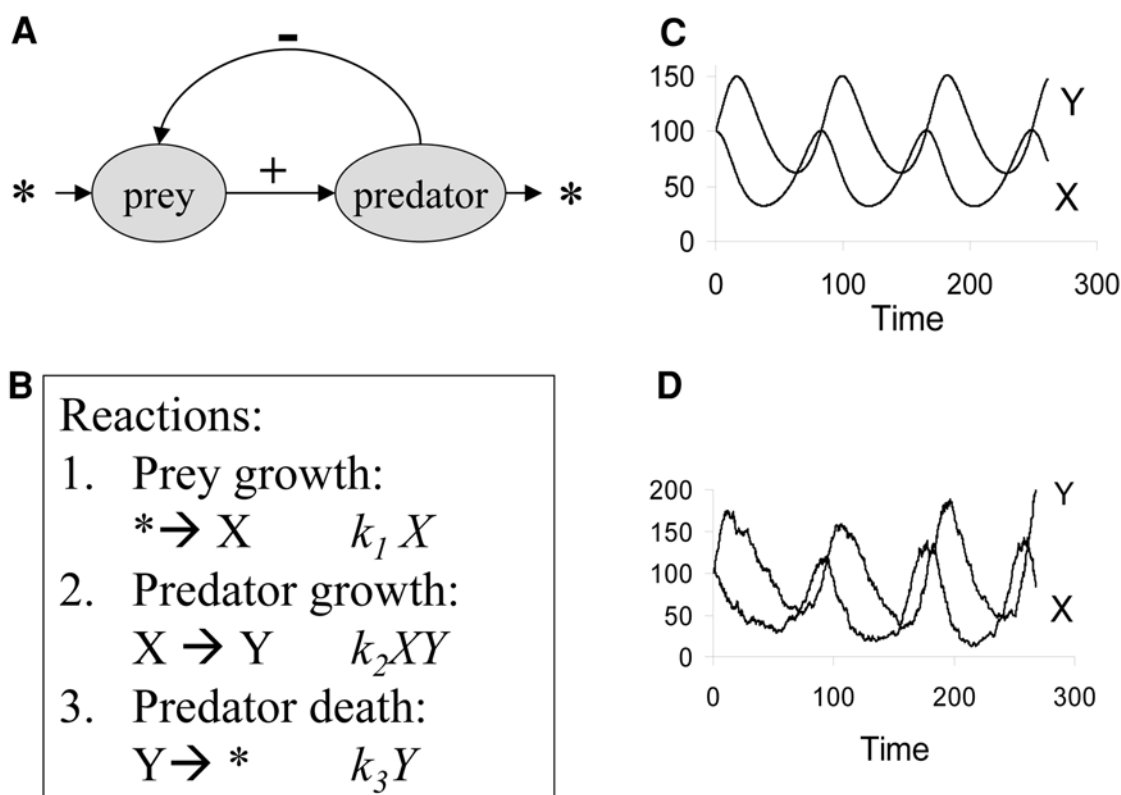


Fig. 3. A simple predator-prey system. **(A)** A diagram representation of the predator-prey system: The prey feeds on unspecified foods (represented by $*$) but is consumed by the predator. The predator dies and degenerates into unspecified products (represented by $*$). Thus the prey promotes the increase in the predator population (indicated by $+$) but the predator facilitates the decrease in the prey population (indicated by a $-$). **(B)** The system can be represented by a set of simple reactions ($X = \text{prey}$; $Y = \text{predator}$); each reaction is characterized by its stoichiometry and its kinetics. **(C)** A typical deterministic simulation result from an ODE model based on the reactions in **(B)**. The differential equations used for this simulation are:

$$\frac{dX}{dt} = k_1 X - k_2 XY; \quad \frac{dY}{dt} = k_2 XY - k_3 Y.$$

(D) A typical simulation result using the Gillespie algorithm. Results in **(C)** and **(D)** were generated using Dynetica (i.e., a simulator of dynamic networks) (143). Levels of prey and predator are expressed in numbers. Note how the stochastic simulation result resembles the deterministic simulation result in terms of qualitative behavior—oscillation—but drastically differs from the latter in numerical details. From the given initial condition, the stochastic simulation can generate completely different dynamics from the deterministic simulation (not shown).

Quantitative Models

A more detailed and precise approach of representing a biological system is to treat it as a network of chemical reactions (Fig. 3). To analyze the system, one needs the stoichiometry of

all the reactions involved. Using this formulation, one can gain deep insights into the steady-state system behavior and characteristics of the network topology by merely analyzing the underlying stoichiometric matrix (40–42). A number of computational tech-

niques, such as metabolic flux analysis (41,42) and flux balance analysis (43), have been successfully developed to assist such analysis. They have played an instrumental role in shaping the field of metabolic engineering, by providing theoretical guidance for experimental manipulation of metabolic networks (44). Recently, these techniques have been successfully employed to reveal the underlying structure of metabolic networks by determining the elementary flux modes (45) and the null space base vectors (46), and in predicting steady-state metabolic capabilities of several model organisms, such as *Escherichia coli* (47,48) and *Haemophilus influenzae* (49).

However, it is often difficult to formulate in stoichiometric models regulatory interactions (50), which are clearly ubiquitous in biological systems. Moreover, because stoichiometric models lack the time-domain in their formulation, they are unable to predict the temporal evolution of biological systems. To make such predictions, a stoichiometric model needs to be supplemented with detailed kinetic information. That is, one needs to specify how fast each reaction is occurring, in addition to its stoichiometry (Fig. 3B). Compared with stoichiometric models, the drawback of kinetic models appears obvious: construction of each model will require much more information, and kinetic data are usually much more difficult to acquire than the stoichiometry of a reaction network. But the payoff of the added complexity is significant: with an appropriate kinetic model, the modeler can gain much deeper insight into the behavior of the system than with a stoichiometric model.

Usually kinetic models are represented as a set of ordinary differential equations (ODE) describing the rates of change for the interacting components (Fig. 3B). Solving these ODEs, often numerically, generates time courses of the interacting components. This ability to predict dynamical behaviors is essential for analyzing systems with rich temporal behaviors, such as circadian clocks. ODE-based kinetic models are also termed *deterministic* because a given initial condition will completely deter-

mine the temporal behavior of the underlying systems (within the errors of the numerical solutions) (Fig. 3C). As to be discussed in the following section, kinetic models can be formulated in a stochastic framework, which may give finer details of system dynamics by revealing random fluctuations in the numbers of each interacting component.

Application of kinetic models in biology has a long history. For example, Lotka and Volterra independently developed a kinetic model nearly 80 years ago to describe dynamics of a predator-prey ecosystem (as cited in (51); see also Fig. 3). Today, kinetic modeling continues to play a major role in modern ecology (51,52). In cellular biology, decades of genetic, biochemical, and biophysical studies have generated a large amount of experimental data that can be used to deduce reaction mechanisms and rate constant parameters, which in turn makes kinetic modeling a preferable choice for describing cellular processes (16,19,23). This point is particularly obvious for many well-characterized systems, including bacterial chemotaxis signaling networks (53,54), developmental pattern formation in *Drosophila* (55–58), aggregation stage network of *Dictyostelium* (59), viral infection (60–65), circadian rhythms (66,67), *E. coli* stress response circuit (68), single *E. coli* cell growth (69), yeast cell-cycle control (70,71), and physiological processes (72–74).

Kinetic models may be coupled with equations describing mass transport processes, leading to more complicated mathematical models. For example, in modeling some biological processes, it is necessary and feasible to account for not only reaction kinetics, but also transport of interacting components by diffusion. This approach is most commonly adopted in models describing the spatiotemporal dynamics of small chemicals such as calcium (75–79) and cyclic adenosine monophosphate (80–83), or larger components (84–89) whose transport can be treated as diffusion. Such models often involve using partial differential equations (PDE) in addition to ODEs. Current efforts are under way to acquire detailed information on the distribution of molecules in the

cell. Yet, for most intracellular gene expression processes, experimental data on the spatial domain are often too sparse to make sensible PDE models. In addition, solving a PDE model often invokes much higher computational cost for the same number of interacting components. Thus, most current mathematical models of gene regulation networks have ignored spatial heterogeneity in a system. When it is essential to describe transport processes, these processes may be approximated by first-order reactions, which fall into the framework of ODE models (55,56,63).

Stochastic Formulation of Reaction Networks

Despite their broad applications, ODE-based kinetic models are criticized for their implicit assumption of continuity in the concentrations of interacting cellular species, particularly for intracellular processes. In particular, many proteins are expressed at nanomolar levels, which correspond to only tens or hundreds of molecules per cell. In this scenario, the small numbers of the interacting species can lead to significant random fluctuations in the levels of these species. A good deal of experimental data has indeed demonstrated significant noise in gene expression processes (90–94). Deterministic in nature, ODE models will fail to predict such fluctuations. For this reason, some researchers have questioned the use of deterministic simulations in characterizing the behaviors of biological systems, and suggested using stochastic simulations instead (95–97).

Several algorithms are available for carrying out stochastic simulations (98); the Gillespie algorithm (99) is by far the most popular. Following a Monte Carlo procedure, the Gillespie algorithm predicts the time evolution of the system by determining when and in what order the next reaction is going to occur. This algorithm has a rigorous theoretical foundation, and is shown to give *exact* solution for a network of elementary reactions occurring in a well-stirred environment (99,100). It often generates dynamics drastically different from

the prediction by deterministic simulations, particularly when some reactions have nonlinear terms in their rate expressions (101; see also Fig. 3D). The application of stochastic simulations has led to speculations regarding the implications of intracellular noise, in particular, how it may be exploited in nature and how it can be effectively controlled (98). For example, it has been suggested that the intrinsic noise might be exploited in generating phenotypic diversity in a clonal population so that the population is more capable of surviving different environments (102).

Although the Gillespie algorithm reveals stochastic fluctuations resulting from small molecular numbers, several outstanding questions make it unclear whether or to what extent it is more appropriate than a deterministic approach in modeling cellular reaction networks. From a practical perspective, well-polished computational techniques (e.g., bifurcation analysis—analysis of qualitative changes in the dynamics of a system caused by the variation of some system parameters (103)) and software tools (e.g., Xppaut at <http://www.math.pitt.edu/~bard/xpp/xpp.html>) are available for high-level analysis of system dynamics using ODEs, although such analysis is far from practical using the stochastic formulation (18,98). Also, stochastic simulations by the Gillespie algorithm are often much more time consuming than deterministic simulations. In fact, the computation time of this algorithm approximately scales with the frequency of the reaction events: the more reactions there are, or the more molecules there are, the longer the computation will take for a given simulated time span (99). Recent efforts have been made to improve the computation efficiency of stochastic simulations, either by directly improving the efficiency of the Gillespie algorithm while keeping its rigor (104), or by approximating the computation for fast reactions when separation of time scales is justifiable (105,106). Another alternative to exact simulation by the Gillespie algorithm is to incorporate fluctuations by explicitly including random variables in the differential equations describing the system (107–111). This approach

results in a Langevin equation or a stochastic differential equation, leading to substantial gain in computation speed at the acceptable loss of computation accuracy (109,110). A major advantage of this approach is that it ties the basis of Gillespie algorithm—a chemical master equation—to a conventional, deterministic formulation of chemical kinetics. In fact, a chemical master equation is shown to be asymptotically equivalent to the corresponding Langevin equation under certain conditions (109). For more detailed discussion on this topic, the reader is directed to an excellent recent review (98).

Aside from the issue of computational speed, a fundamental question is to what extent the simulated noise reflects the true noise. For the Gillespie algorithm to be exact, the system must consist of elementary reactions only and it must be well-stirred (i.e., spatially homogeneous) (99,100). These criteria are rarely met when modeling intracellular processes: the intracellular environment is highly heterogeneous and many a reaction lumps multiple steps. For example, a reaction as basic as transcription of a single gene consists of several more basic steps, such as binding of RNA polymerase to promoter region in forming a closed complex, formation of an open complex, and eventually transcriptional elongation. Moreover, in addition to small numbers of interacting molecules, randomness may stem from other factors, such as the cellular components not specified in the model and even conformational changes of biological macromolecules. From this standpoint, the Gillespie algorithm, as well as other stochastic algorithms, gives *empirical* predictions (when applied to describing intracellular processes), just as its ODE counterpart does. In fact, recent experiments by Elowitz et al. demonstrated that the total noise of gene expression consists of both intrinsic noise and extrinsic noise, and that the total noise does not correlate with the intrinsic noise (91). It is probable, but not proven, that the noise generated by the Gillespie algorithm accounts for a major part of the intrinsic noise but not the extrinsic noise for cellular processes. In all, one needs to take

caution in interpreting the fluctuations generated by a stochastic simulation.

SPECIFIC APPLICATIONS

Because the majority of current mathematical models of biological systems are kinetic models, I will focus here on applications of kinetic models only. The categorization of different applications is somewhat arbitrary. It should also be noted that some of these applications may be achieved by other types of the modeling as well, with potentially different resolutions and outcomes.

Revealing Gaps in Current Understanding

By integrating current understanding of the underlying system, a kinetic model can be used to test the consistency in the experimental data or mechanisms. Often times, model predictions will deviate from the experiment. This discrepancy is not necessarily a negative thing; in fact, it can be very informative. It indicates the gap or hole in our knowledge (or at least the knowledge integrated into the model) and may provide guidance for future experimentation.

von Dassow et al. recently developed a mathematical model of the segmentation network of *Drosophila* embryo development (55). Their initial model based on known network connectivity was unable to predict the segmentation pattern observed in experiments. To test whether this discrepancy was due to inaccurate parameters, the authors randomly changed the kinetic parameters over a wide range of plausible values and carried out simulations for each parameter set. However, none of these parameter sets could generate the desired pattern. This discrepancy between simulation and experiment, the authors argued, indicates a gap in the understanding of the segmentation network. Specifically, if the network connectivity as implemented in the initial model had been correct, then at least some parameter sets should have generated the desired pattern. To account for the discrepancy, the authors hypothesized

additional connections, based on experimental observations in the literature, in their network model and repeated their computational analysis. Interestingly, the added links indeed seemed to be an important missing piece: their model now could generate the desired pattern for a significant portion of the parameter sets they generated. Although these predictions await experimental verification, this work highlights a key use of modeling: to identify inconsistency in our knowledge and to suggest experimentally testable hypothesis. It is worth noting that a similar model was developed by Reinitz et al. to describe pattern formation in *Drosophila* embryos (56–58). The strength of these studies lies in the extensive use of detailed experimental data to fit the parameters in the dynamic model and subsequent application of the model to explore specific roles of different genes in forming embryonic patterns.

Based on more than four decades of literature data, Endy et al. developed a detailed model of phage T7 (62). This model describes the entire intracellular life cycle of the virus from the entry of viral genome to the production of viral progeny. This model was later used to predict how the viral growth rate would depend on the organization of genetic elements (such as genes, promoters, and transcription terminators) (112,113). Among other results, the simulation predicted that the T7 growth rate overall would decrease as the T7 polymerase gene (gene 1) was moved away from the entering end of the T7 genome. This makes intuitive sense: the further downstream T7 gene 1 is, the more delayed its expression. The delay in gene 1 expression subsequently would lead to a delay in the expression of the majority of T7 genes, thus slowing down T7 growth. Interestingly, the simulation also predicted that when gene 1 is immediately downstream the early T7 promoters, the T7 growth rate would be higher than the wild-type value because of the establishment of a positive feedback for the production of T7 RNA polymerase. However, experimental results with three ectopic gene 1 mutants only verified the optimality of the wild-type T7. The mutant

ecto1.7, which was predicted to grow faster than the wild type, turned out to grow much more slowly. One possible reason for this mismatch, the authors found, was that disruption of the seemingly nonessential gene 1.7 (in ecto1.7, gene 1 was inserted within the coding region of gene 1.7) actually had significant quantitative effects on T7 growth due to unknown mechanisms.

Another possible reason for the discrepancy is the assumption that the host cell offers unlimited translation resources, such as ribosomes and amino acids. In fact, with a more realistic representation of the *E. coli* host (65), an extended T7 model was able to give more accurate predictions for the growth of ectopic gene 1 mutants (114). The revised simulation showed that the positive feedback for producing the T7 RNA polymerase was detrimental to overall viral growth by causing unfavorable distribution of limited protein synthesis resources. Moreover, the revised model also predicted that the phage T7 growth rate would be faster in faster growing host cells, a prediction that was soon confirmed experimentally (65). In addition to suggesting possible mechanisms for the mismatch between the original model and experiment, these series of work based on the revised model highlighted the importance of the host environment in the development of an organism, a factor ignored in the original T7 model.

Characterizing Emergent Properties

Mathematical models are beginning to reveal the so-called emergent properties, which are often difficult to grasp intuitively by examining the constituent parts of these systems. An often-characterized emergent property is robustness of a system. For experimental biologists, robustness describes the stability of a phenotype in the presence of genetic and environmental variations, but the term often comes with some ambiguity in its exact meaning and is difficult to quantify (115). In a well-defined mathematical model, however, robustness can be quantified in a straightforward fashion: it

can be measured by the sensitivity of a system function to variations in parameters.

Barkai and Leibler (53) carried out a detailed analysis of how the output of a chemotaxis network responds to perturbations in the kinetic parameters that define the network behavior. Their simulations demonstrated that key properties of the network are robust to parametric perturbations. Admirably, key findings of this computational study were later verified by experimental work from the same group (116). The authors went on arguing that such robust behavior might be a generic feature that is necessary to ensure proper functioning of a wide variety of biological systems (53). This argument was echoed by Morohashi et al. (117), who contended that robustness to variations could be used as a measure of the plausibility of the models of a given system. Moreover, others have argued that many complex systems, including biological systems, often demonstrate "robust yet fragile" features (118,119). These arguments are gaining support from several recent modeling studies. In the previously mentioned work by von Dassow et al. (55), the "remedied" version of the *Drosophila* segmentation network model demonstrated robust features: the model was able to predict the correct segmentation pattern for a wide range of parameter settings (55). Later, the same group found that another system—the *Drosophila* neurogenic network—also demonstrated significant robustness with respect to network parameters (120). Using an extended phage T7 model, You and Yin recently found that the ability of phage T7 to survive was overall robust to perturbations to its kinetic parameters that define its physiology, but sensitive to perturbations to the organization of genetic elements in the genome (121).

Analysis of robustness by modeling is still a relatively new area. Despite impressive progresses made so far, open questions remain as whether and to what extent results generated from simulations reflect reality (122). A major challenge in addressing these questions is to properly map kinetic parameters to the genotype of a given biological system. Ideally, we

should be able to answer the question: what mutations are required to achieve a certain amount of change in a given parameter? However, this mapping is difficult for many parameters, especially when a mutation in one gene may affect multiple phenotypic traits (112,123).

Testing Complex Hypotheses

As models become more "realistic" by incorporating increasingly detailed data and mechanisms, they may be treated as *in silico* organisms and used to explore applied or fundamental questions that are beyond the underlying system per se. In particular, such complex models can also be called upon to test hypotheses or theories that are difficult, expensive, or even impossible to explore experimentally with current technology. Note that the ability to quantitatively test complex hypotheses is an important feature of sophisticated kinetic models. Models of further abstraction, such as diagrams or Boolean models, lack this ability, largely because of detachment between system behaviors and physical parameters.

Thanks to the ease of using the phage T7 model to create thousands of T7 mutants *in silico* and to efficiently evaluate their fitness, You and Yin were able to systematically characterize the nature of the interactions among deleterious mutations in terms of their effects on fitness at the population level (123). Such genetic interactions play a major role in a variety of fundamental biological phenomena, including evolution of recombination, dynamics of fitness landscapes, and buffering of genetic variations, but their experimental characterization has been hindered by the difficulty in generating and quantifying a large number of mutants. From their simulation, You and Yin found that the nature of genetic interactions depended on the growth environment for the organism, as well as the severity of the deleterious mutations. Their results offered an intuitive explanation for the seemingly conflicting conclusions on the nature of genetic interactions from prior experimental studies: the mutations tested in

experiments may have been of differing severity, and their interactions may have been tested in different environments.

Guiding Experimental Design of De Novo Gene Circuits

With the elucidation of a wide variety of biological components, including genes and gene expression regulation elements, there has been soaring enthusiasm for building *de novo* or synthetic gene networks in the last several years (39,124–131). Designing and constructing such synthetic gene networks has great potential in generating genetic “gadgets” with novel applications, as well as in improving our understanding of how cellular components function. As discussed in detail in a recent review (126), the basic strategy of constructing such networks follows roughly the same procedure:

1. Outline the basic network connectivity.
2. Analyze the network behavior by mathematical modeling.
3. Implement the network experimentally.

Usually the adopted models are purposefully simplified to capture the qualitative behavior of the designed system. In addition to time course simulations, bifurcation analysis is often employed to examine how qualitative characteristics of the system dynamics may depend on model parameters (103). So far, this deceptively simple procedure has yielded only a few successful products, including a negative feedback control circuit (39), a toggle switch (128), a ring oscillator (or the “repressilator”) (129), a relaxation-type oscillator (132), and a genetic inverter (130). One of the major challenges of constructing these engineered circuits is that the behavior of the implemented circuit often deviated significantly from the predicted behavior. This is probably more an indication of our lack of detailed understanding of how cellular components function than a failure of the approach. In fact, in a sense the success of this series of seminal work was safeguarded by the proper use of modeling. For example, in designing the repressilator, a simplified kinetic

model was used to highlight several design goals for achieving the desired function—oscillations in protein concentrations (129). In constructing the genetic inverter, Weiss used a mathematical model to guide his efforts to “rationally debug” an initially nonfunctional circuit. When modeling fails, powerful experimental methods can come into play. As demonstrated by recent seminal work (133), directed-evolution techniques, which have proven to be particularly powerful for optimizing enzymes (134,135), are equally valuable for fine-tuning the function of *de novo* gene circuits. In addition to fine-tuning circuit functions, this “design-then-mutate” approach may also play an important role in the development and refinement of quantitative models by offering additional structure-function insights (136).

OUTLOOK: A MATURING INFRASTRUCTURE FOR MODELING

With increasing appreciation of the merit of modeling for deeper understanding of biological systems, I anticipate that two lines of related research efforts—namely, the development of biologist-friendly modeling tools and high-level databases harboring information on biomolecular interactions (including pathway databases)—will dramatically facilitate broader application of modeling in biology by providing an streamlined platform for storage and communication of data and integrated models.

Software Tools

Despite its potential benefits for fundamental and applied biological research, the application of mathematical modeling in biology has been hindered by the lack of software tools to build, analyze, and visualize models, particularly for researchers unfamiliar with programming and numerical methods. But the situation is changing. To address this issue, a number of programs that aim to facilitate the model construction and analysis have been developed in the last several

years. A partial list of these programs includes Gepasi (137), DBsolve (138), E-Cell (139), Virtual Cell (140,141), StochSim (97), Jarnac/Jdesigner (<http://www.cds.caltech.edu/~hsauro/index.htm>), Cellerator (142), and Dynetica (121,143).

Although different implementation strategies are adopted, all these programs strive to provide an intuitive interface and versatile simulation capabilities for the user. It is as yet difficult to give an unbiased evaluation of all these programs without detailed third-party benchmark comparisons. Briefly, Gepasi and DBsolve focus on the analysis of biochemical and metabolic networks. In addition to basic time-course simulations, these programs provide additional modules to explore the properties of metabolic networks. E-Cell aims to construct whole-cell models, and it has been applied to model a self-sustaining hypothetical cell (144) and a human erythrocyte (139). Virtual Cell is advantageous in that it accounts for the diffusion of molecules in addition to their reactions in describing cellular processes. StochSim simulates the system dynamics using an approximate stochastic algorithm, which is more efficient but less accurate than the Gillespie algorithm. Jarnac is an interactive and interpreted language for describing and modeling cellular networks, and it can interact with JDesigner, a tool for visual construction of these networks. Implemented using Mathematica as the back end, Cellerator introduces palette-driven, arrow-based notations to represent biochemical reaction networks, and provides a mechanism to translate these representations into ODEs that can be solved by Mathematica. Dynetica integrates model construction, analysis, and network visualization into a unified modeling framework; it is distinct from others in that (1) it allows time-course simulations using both deterministic algorithm (ODE-based) and the Gillespie algorithm (e.g., see Fig. 3), and (2) it facilitates the construction of genetic networks, where the majority of reactions revolve around gene expression (121,143).

Although encouraging for biologists, an immediate issue raised by these diverse programs is the lack of compatibility among them. To address this issue, there have been many efforts toward developing standard modeling languages, particularly the SBML (Systems Biology Markup Language) (145) (see also <http://www.cds.caltech.edu/erato>) and the CellML (Cell Markup Language) (<http://www.cellml.org>), both of which are based on XML (<http://www.xml.org>), a popular structural data representation language. It is foreseeable that in the near future, many of these programs can share models via such standard modeling languages. If this is realized, the user can use any of the modeling software to build models and conveniently switch to other tools if needed. As such, the user can take full advantage of different software packages with minimum efforts.

Databases of Biological Interactions

The usefulness of a model is often determined by the quality of the underlying experimental data and mechanisms. This point highlights the importance of managing such information. But what kind of information should we document? A good starting point is probably the data on biomolecular interactions and reaction pathways, because these types of data can be mapped into a model in a straightforward fashion. In fact, an amazing number of databases or knowledge environments have been developed along this line, many being accessible from the Internet. Notably among these are the AfCS-Nature Signaling Gateway (<http://www.signalinggateway.org>), the Biomolecular Interaction Network Database (BIND, <http://www.bind.ca/>) (146), the Database of Interacting Proteins (DIP, <http://dip.doe-mbi.ucla.edu/>) (147), the EcoCyc (<http://www.ecocyc.org/>) (148), the Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.genome.ad.jp/kegg/>) (149), and the Signaling Transduction Knowledge Environment (<http://stke.sciencemag.org/>).

The blooming efforts to collect and document experimental data on cellular processes

reflect encouraging recent progresses in acquiring such data in large quantities. More importantly, the documented data are beginning to form an information infrastructure for high-level data compilation and analysis. But much challenge is still ahead, especially if one wishes to make efficient use of such information to create mathematical models. In contrast to gene or protein sequences, whose data structure is straightforward to define, complexity of biological interactions as well as different ways that researchers perceive as interactions have hindered data representation (150). The diversity of data format, data quality, notations, and access interfaces will probably pose a hurdle even greater than incompatible modeling languages for communicating these interaction data. To maximally benefit the research community, the databases eventually need to converge into using common standard data formats. Again, XML or XML-based standard languages may prove useful for addressing the compatibility issue. In fact, XML has already been adopted in KEGG to represent metabolic pathways. A potential advantage of databases built upon a standard representation language is that the stored data can be readily mapped into an integrated mathematical model. In the long run, standardization of data representation languages may not only facilitate the construction of the information technology infrastructure per se, but also provide guidance and incentive for experimental biologists to document data in a more consistent fashion.

ACKNOWLEDGMENT

I thank Edward Massaro for inviting me to write this review. I gratefully acknowledge John Yin and Frances Arnold for their advice and guidance. Chris Rao and two anonymous reviewers provided many valuable comments to the manuscript. Financial support is provided by the Defense Advanced Research Projects Agency (DARPA) under Award No. N66001-02-1-8929.

REFERENCES

1. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.
2. O'Farrell, P. H. (1975) High resolution two-dimensional electrophoresis of proteins. *J. Biol. Chem.* **250**, 4007–4021.
3. Kitano, H. (2002) Systems biology: a brief overview. *Science* **295**, 1662–1664.
4. Ideker, T., Galitski, T., and Hood, L. (2001) A new approach to decoding life: systems biology. *Annu. Rev. Genomics Hum. Genet.* **2**, 343–372.
5. Gregory, S. G., Sekhon, M., Schein, J., Zhao, S., Osoegawa, K., Scott, C. E., et al. (2002) A physical map of the mouse genome. *Nature* **418**, 743–750.
6. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001) Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.
7. Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., et al. (2001) The sequence of the human genome. *Science* **291**, 1304–1351.
8. Eckerskorn, C., Strupat, K., Karas, M., Hillenkamp, F., and Lottspeich, F. (1992) Mass spectrometric analysis of blotted proteins after gel electrophoretic separation by matrix-assisted laser desorption/ionization. *Electrophoresis* **13**, 664–665.
9. Henzel, W. J., Billeci, T. M., Stults, J. T., Wong, S. C., Grimley, C., and Watanabe, C. (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc. Natl. Acad. Sci. USA* **90**, 5011–5015.
10. Mann, M., Hendrickson, R. C., and Pandey, A. (2001) Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **70**, 437–473.
11. Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* **340**, 245–246.
12. Uetz, P., Giot, L., Cagney, G., Mansfield, T. A., Judson, R. S., Knight, J. R., et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627.
13. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive

- two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98**, 4569–4574.
14. Asthagiri, A. R. and Lauffenburger, D. A. (2000) Bioengineering models of cell signaling. *Annu. Rev. Biomed. Eng.* **2**, 31–53.
 15. Arkin, A. P. (2001) Synthetic cell biology. *Curr. Opin. Biotechnol.* **12**, 638–644.
 16. Endy, D. and Brent, R. (2001) Modelling cellular behaviour. *Nature* **409 Suppl**, 391–395.
 17. Gilman, A. and Arkin, A. P. (2002) Genetic “code”: representations and dynamical models of genetic components and networks. *Annu. Rev. Genomics Hum. Genet.* **3**, 341–369.
 18. Rao, C. V. and Arkin, A. P. (2001) Control motifs for intracellular regulatory networks. *Annu. Rev. Biomed. Eng.* **3**, 391–419.
 19. Tyson, J. J., Chen, K., and Novak, B. (2001) Network dynamics and cell physiology. *Nat. Rev. Mol. Cell. Biol.* **2**, 908–916.
 20. Palsson, B. (2000) The challenges of in silico biology. *Nat. Biotechnol.* **18**, 1147–1150.
 21. Bailey, J. E. (1998) Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities. *Biotechnol. Prog.* **14**, 8–20.
 22. Steven Wiley, H., Shvartsman, S. Y., and Lauffenburger, D. A. (2003) Computational modeling of the EGF-receptor system: a paradigm for systems biology. *Trends Cell Biol.* **13**, 43–50.
 23. Hasty, J., McMillen, D., Isaacs, F., and Collins, J. J. (2001) Computational studies of gene regulatory networks: in numero molecular biology. *Nat. Rev. Genet.* **2**, 268–279.
 24. Thieffry, D. and Sarkar, S. (1998) Forty years under the central dogma. *Trends Biochem. Sci.* **23**, 312–316.
 25. Kohn, K. W. (1999) Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell.* **10**, 2703–2734.
 26. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabasi, A. L. (2000) The largescale organization of metabolic networks. *Nature* **407**, 651–654.
 27. Jeong, H., Mason, S. P., Barabasi, A. L., and Oltvai, Z. N. (2001) Lethality and centrality in protein networks. *Nature* **411**, 41–42.
 28. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., and Barabasi, A. L. (2002) Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551–1555.
 29. Rives, A. W. and Galitski, T. (2003) Modular organization of cellular networks. *Proc. Natl. Acad. Sci. USA* **100**, 1128–1133.
 30. Wagner, A. and Fell, D. A. (2001) The small world inside large metabolic networks. *Proc. R. Soc. Lond. B. Biol. Sci.* **268**, 1803–1810.
 31. Strogatz, S. H. (2001) Exploring complex networks. *Nature* **410**, 268–276.
 32. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002) Network motifs: simple building blocks of complex networks. *Science* **298**, 824–827.
 33. Shen-Orr, S. S., Milo, R., Mangan, S., and Alon, U. (2002) Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64–68.
 34. Thomas, R. (1973) Boolean formalization of genetic control circuits. *J. Theor. Biol.* **42**, 563–585.
 35. Glass, L., and Kauffman, S. A. (1973) The logical analysis of continuous, non-linear biochemical control networks. *J. Theor. Biol.* **39**, 103–129.
 36. Glass, L. (1975) Classification of biological networks by their qualitative dynamics. *J. Theor. Biol.* **54**, 85–107.
 37. Kauffman, S. A. (1993) *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University, New York.
 38. Kuipers, B. (1986) Qualitative simulation. *Artif. Intell.* **29**, 289–338.
 39. Becskei, A. and Serrano, L. (2000) Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593.
 40. Clarke, B. L. (1988) Stoichiometric network analysis. *Cell Biophys.* **12**, 237–253.
 41. Fell, D. A. (1992) Metabolic control analysis: a survey of its theoretical and experimental development. *Biochem. J.* **286**, 313–330.
 42. Kacser, H., and Burns, J. A. (1995) The control of flux. *Biochem. Soc. Trans.* **23**, 341–366.
 43. Varma, A. and Palsson, B. O. (1994) Metabolic flux balancing-basic concepts, scientific and practical use. *Bio-Technology* **12**, 994–998.
 44. Stephanopoulos, G., Aristidou, A. A., and Nielsen, J. (1998) *Metabolic Engineering. Principles and Methodologies*, Academic, San Diego.
 45. Schuster, S., Fell, D. A., and Dandekar, T. (2000) A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* **18**, 326–332.
 46. Schilling, C. H. and Palsson, B. O. (1998) The underlying pathway structure of biochemical

- reaction networks. *Proc. Natl. Acad. Sci. USA* **95**, 4193–4198.
47. Edwards, J. S., Ibarra, R. U., and Palsson, B. O. (2001) In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* **19**, 125–130.
 48. Schilling, C. H., Edwards, J. S., and Palsson, B. O. (1999) Toward metabolic phenomics: analysis of genomic data using flux balances. *Biotechnol. Prog.* **15**, 288–295.
 49. Edwards, J. S. and Palsson, B. O. (1999) Systems properties of the Haemophilus influenzae Rd metabolic genotype. *J. Biol. Chem.* **274**, 17410–17416.
 50. Varner, J. and Ramkrishna, D. (1999) Mathematical models of metabolic pathways. *Curr. Opin. Biotechnol.* **10**, 146–150.
 51. May, R. M. (1974) *Stability and Complexity in Model Ecosystems*, ed. 2, Princeton University, Princeton, NJ.
 52. Levin, S. A., Grenfell, B., Hastings, A., and Perelson, A. S. (1997) Mathematical and computational challenges in population biology and ecosystems science. *Science* **275**, 334–343.
 53. Barkai, N. and Leibler, S. (1997) Robustness in simple biochemical networks. *Nature* **387**, 913–917.
 54. Spiro, P. A., Parkinson, J. S., and Othmer, H. G. (1997) A model of excitation and adaptation in bacterial chemotaxis. *Proc. Natl. Acad. Sci. USA* **94**, 7263–7268.
 55. von Dassow, G., Meir, E., Munro, E. M., and Odell, G. M. (2000) The segment polarity network is a robust developmental module. *Nature* **406**, 188–192.
 56. Reinitz, J., Kosman, D., Vanario-Alonso, C. E., and Sharp, D. H. (1998) Stripe forming architecture of the gap gene system. *Dev. Genet.* **23**, 11–27.
 57. Reinitz, J., Mjolsness, E., and Sharp, D. H. (1995) Model for cooperative control of positional information in Drosophila by bicoid and maternal hunchback. *J. Exp. Zool.* **271**, 47–56.
 58. Reinitz, J. and Sharp, D. H. (1995) Mechanism of eve stripe formation. *Mech. Dev.* **49**, 133–158.
 59. Laub, M. T. and Loomis, W. F. (1998) A molecular network that produces spontaneous oscillations in excitable cells of Dictyostelium. *Mol. Biol. Cell.* **9**, 3521–3532.
 60. McAdams, H. H. and Shapiro, L. (1995) Circuit simulation of genetic networks. *Science* **269**, 650–656.
 61. Shea, M. A. and Ackers, G. K. (1985) The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation. *J. Mol. Biol.* **181**, 211–230.
 62. Endy, D., Kong, D., and Yin, J. (1997) Intracellular kinetics of a growing virus: a genetically structured simulation for bacteriophage T7. *Biotech. Bioeng.* **55**, 375–389.
 63. Reddy, B. and Yin, J. (1999) Quantitative intracellular kinetics of HIV type 1. *AIDS Res. Hum. Retroviruses.* **15**, 273–283.
 64. Eigen, M., Biebricher, C. K., Gebinoga, M., and Gardiner, W. C. (1991) The hypercycle. Coupling of RNA and protein biosynthesis in the infection cycle of an RNA bacteriophage. *Biochemistry* **30**, 11005–11018.
 65. You, L., Suthers, P. F., and Yin, J. (2002) Effects of Escherichia coli physiology on growth of phage T7 in vivo and in silico. *J. Bacteriol.* **184**, 1888–1894.
 66. Barkai, N. and Leibler, S. (2000) Circadian clocks limited by noise. *Nature* **403**, 267–268.
 67. Smolen, P., Baxter, D. A., and Byrne, J. H. (2001) Modeling circadian oscillations with interlocking positive and negative feedback loops. *J. Neurosci.* **21**, 6644–6656.
 68. Srivastava, R., Peterson, M. S., and Bentley, W. E. (2001) Stochastic kinetic analysis of the Escherichia coli stress circuit using sigma(32)-targeted antisense. *Biotechnol. Bioeng.* **75**, 120–129.
 69. Shuler, M. L., Leung, S., and Dick, C. C. (1979) A mathematical model for the growth of a single bacterial cell. *Ann. NY Acad. Sci.* **326**, 35–55.
 70. Novak, B., Csikasz-Nagy, A., Gyorfy, B., Chen, K., and Tyson, J. J. (1998) Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. *Biophys. Chem.* **72**, 185–200.
 71. Chen, K. C., Csikasz-Nagy, A., Gyorfy, B., Val, J., Novak, B., and Tyson, J. J. (2000) Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell.* **11**, 369–391.
 72. Quick, D. J. and Shuler, M. L. (1999) Use of in vitro data for construction of a physiologically based pharmacokinetic model for naphthalene in rats and mice to probe species differences. *Biotechnol. Prog.* **15**, 540–555.
 73. Winslow, R. L., Scollan, D. F., Holmes, A., Yung, C. K., Zhang, J., and Jafri, M. S. (2000) Electrophysiological modeling of cardiac

- ventricular function: from cell to organ. *Annu. Rev. Biomed. Eng.* **2**, 119–155.
74. Noble, D. (2002) Modeling the heart—from genes to cells to the whole organ. *Science* **295**, 1678–82.
 75. Dupont, G., Pontes, J., and Goldbeter, A. (1996) Modeling spiral Ca²⁺ waves in single cardiac cells: role of the spatial heterogeneity created by the nucleus. *Am. J. Physiol.* **271**, C1390–C1399.
 76. Schuster, S., Marhl, M., and Hofer, T. (2002) Modelling of simple and complex calcium oscillations. From single-cell responses to intercellular signalling. *Eur. J. Biochem.* **269**, 1333–1355.
 77. Hofer, T., Politi, A., and Heinrich, R. (2001) Intercellular Ca²⁺ wave propagation through gap-junctional Ca²⁺ diffusion: a theoretical study. *Biophys. J.* **80**, 75–87.
 78. Hofer, T., Venance, L., and Giaume, C. (2002) Control and plasticity of intercellular calcium waves in astrocytes: a modeling approach. *J. Neurosci.* **22**, 4850–4859.
 79. Fink, C. C., Slepchenko, B., Moraru, II, Watras, J., Schaff, J. C., and Loew, L. M. (2000) An image-based model of calcium waves in differentiated neuroblastoma cells. *Biophys. J.* **79**, 163–183.
 80. Palsson, E., Lee, K. J., Goldstein, R. E., Franke, J., Kessin, R. H., and Cox, E. C. (1997) Selection for spiral waves in the social amoebae *Dictyostelium*. *Proc. Natl. Acad. Sci. USA* **94**, 13719–13723.
 81. Palsson, E. and Cox, E. C. (1996) Origin and evolution of circular waves and spirals in *Dictyostelium discoideum* territories. *Proc. Natl. Acad. Sci. USA* **93**, 1151–1155.
 82. Halloy, J., Lauzeral, J., and Goldbeter, A. (1998) Modeling oscillations and waves of cAMP in *Dictyostelium discoideum* cells. *Biophys. Chem.* **72**, 9–19.
 83. Tyson, J. J. and Murray, J. D. (1989) Cyclic AMP waves during aggregation of *Dictyostelium amoebae*. *Development* **106**, 421–426.
 84. Smith, A. E., Slepchenko, B. M., Schaff, J. C., Loew, L. M., and Macara, I. G. (2002) Systems analysis of Ran transport. *Science* **295**, 488–491.
 85. You, L. and Yin, J. (1999) Amplification and spread of viruses in a growing plaque. *J. Theor. Biol.* **200**, 365–373.
 86. Yin, J. and McCaskill, J. S. (1992) Replication of viruses in a growing plaque: a reaction-diffusion model. *Biophys. J.* **61**, 1540–1549.
 87. Shvartsman, S. Y., Wiley, H. S., Deen, W. M., and Lauffenburger, D. A. (2001) Spatial range of autocrine signaling: modeling and computational analysis. *Biophys. J.* **81**, 1854–1867.
 88. Pribyl, M., Muratov, C. B., and Shvartsman, S. Y. (2003) Long-range signal transmission in autocrine relays. *Biophys. J.* **84**, 883–896.
 89. Shvartsman, S. Y., Muratov, C. B., and Lauffenburger, D. A. (2002) Modeling and computational analysis of EGF receptor-mediated cell communication in *Drosophila* oogenesis. *Development* **129**, 2577–2589.
 90. van Roon, M. A., Aten, J. A., van Oven, C. H., Charles, R., and Lamers, W. H. (1989) The initiation of hepatocyte-specific gene expression within embryonic hepatocytes is a stochastic event. *Dev. Biol.* **136**, 508–516.
 91. Elowitz, M. B., Levine, A. J., Siggia, E. D., and Swain, P. S. (2002) Stochastic gene expression in a single cell. *Science* **297**, 1183–1186.
 92. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., and van Oudenaarden, A. (2002) Regulation of noise in the expression of a single gene. *Nat. Genet.* **31**, 69–73.
 93. Ross, I. L., Browne, C. M., and Hume, D. A. (1994) Transcription of individual genes in eukaryotic cells occurs randomly and infrequently. *Immunol. Cell Biol.* **72**, 177–185.
 94. Zlokarnik, G., Negulescu, P. A., Knapp, T. E., Mere, L., Burres, N., Feng, L., et al. (1998) Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* **279**, 84–88.
 95. McAdams, H. H. and Arkin, A. (1998) Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* **27**, 199–224.
 96. Goss, P. J. E. and Peccoud, J. (1998) Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. USA* **95**, 6750–6755.
 97. Morton-Firth, C. J. and Bray, D. (1998) Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.* **192**, 117–128.
 98. Rao, C. V., Wolf, D. M., and Arkin, A. P. (2002) Control, exploitation and tolerance of intracellular noise. *Nature* **420**, 231–237.
 99. Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361.
 100. Gillespie, D. T. (1992) A rigorous derivation of the chemical master equation. *Physica A.* **188**, 404–425.

101. Srivastava, R., You, L., Summers, J., and Yin, J. (2002) Stochastic vs. deterministic modeling of intracellular viral kinetics. *J. Theor. Biol.* **218**, 309–321.
102. McAdams, H. H. and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* **94**, 814–819.
103. Seydel, R. (1994) *Practical Bifurcation and Stability Analysis: From Equilibrium to Chaos*, ed. 2, Springer-Verlag, New York.
104. Gibson, M. A. and Bruck, J. (2000) Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* **104**, 1876–1889.
105. Haseltine, E. L. and Rawlings, J. B. (2002) Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J. Chem. Phys.* **117**, 6959–6969.
106. Rao, C. V. and Arkin, A. P. (2003) Stochastic chemical kinetics and the quasi steady-state assumption: application to the Gillespie algorithm. *J. Chem. Phys.* **118**, 4999–5010.
107. Hasty, J., Pradines, J., Dolnik, M., and Collins, J. J. (2000) Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. USA* **97**, 2075–2080.
108. Kepler, T. B. and Elston, T. C. (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136.
109. Gillespie, D. T. (2000) The chemical Langevin equation. *J. Chem. Phys.* **113**, 297–306.
110. Gillespie, D. T. (2001) Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys.* **115**, 1716–1733.
111. Gillespie, D. T. (2002) The chemical Langevin and Fokker-Planck equations for the reversible isomerization reaction. *J. Phys. Chem. A* **106**, 5063–5071.
112. Endy, D., You, L., Yin, J., and Molineux, I. J. (2000) Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl. Acad. Sci. USA* **97**, 5375–5380.
113. Endy, D. (1997) *Development and Application of a Genetically Structured Simulation for Bacteriophage T7*, Ph.D. thesis, Dartmouth College.
114. You, L. and Yin, J. (2001) Simulating the growth of viruses. *Pac. Symp. Biocomput.* 532–543.
115. Nijhout, H. F. (2002) The nature of robustness in development. *Bioessays* **24**, 553–563.
116. Alon, U., Surette, M. G., Barkai, N., and Leibler, S. (1999) Robustness in bacterial chemotaxis. *Nature* **397**, 168–171.
117. Morohashi, M., Winn, A. E., Borisuk, M. T., Bolouri, H., Doyle, J., and Kitano, H. (2002) Robustness as a measure of plausibility in models of biochemical networks. *J. Theor. Biol.* **216**, 19–30.
118. Calson, J. M. and Doyle, J. (2000) Highly optimized tolerance: robustness and design in complex systems. *Phys. Rev. Lett.* **84**, 2529–2532.
119. Csete, M. E. and Doyle, J. C. (2002) Reverse engineering of biological complexity. *Science* **295**, 1664–1669.
120. Meir, E., von Dassow, G., Munro, E., and Odell, G. M. (2002) Robustness, flexibility, and the role of lateral inhibition in the neurogenic network. *Curr. Biol.* **12**, 778–786.
121. You, L. (2002), *The Extension, Application, and Generalization of a Phage T7 Intracellular Growth Model*, Ph.D. thesis, University of Wisconsin-Madison, Madison, WI.
122. Gibson, G. (2002) Developmental evolution: getting robust about robustness. *Curr. Biol.* **12**, R347–R349.
123. You, L. and Yin, J. (2002) Dependence of epistasis on environment and mutation severity as revealed by in silico mutagenesis of phage T7. *Genetics* **160**, 1273–1281.
124. Kobayashi, T., Chen, L., and Aihara, K. (2003) Modeling genetic switches with positive feedback loops. *J. Theor. Biol.* **221**, 379–399.
125. Hasty, J., Dolnik, M., Rottschäfer, V., and Collins, J. J. (2002) Synthetic gene network for entraining and amplifying cellular oscillations. *Phys. Rev. Lett.* **88**, 148101-1–148101-4.
126. Hasty, J., McMillen, D., and Collins, J. J. (2002) Engineered gene circuits. *Nature* **420**, 224–230.
127. Becskei, A., Seraphin, B., and Serrano, L. (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* **20**, 2528–2535.
128. Gardner, T. S., Cantor, C. R., and Collins, J. J. (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342.
129. Elowitz, M. B. and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338.
130. Weiss, R. (2001) *Cellular Computation and Communications Using Engineered Genetic Regulatory Networks*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

131. Weiss, R. and Knight, T. (2000) Engineered communications for microbial robotics, in *6th International Workshop on DNA-Based Computers, DNA 2000* (Condon, A., and Rozenberg, G., eds.) Leiden, The Netherlands, pp. 1–16.
132. Atkinson, M. R., Savageau, M. A., Myers, J. T., and Ninfa, A. J. (2003) Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell* **113**, 597–607.
133. Yokobayashi, Y., Weiss, R., and Arnold, F. H. (2002) Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA* **99**, 16587–16591.
134. Arnold, F. H. and Moore, J. C. (1997) Optimizing industrial enzymes by directed evolution. *Adv. Biochem. Eng. Biotechnol.* **58**, 1–14.
135. Arnold, F. H. and Volkov, A. A. (1999) Directed evolution of biocatalysts. *Curr. Opin. Chem. Biol.* **3**, 54–59.
136. Hasty, J. (2002) Design then mutate. *Proc. Natl. Acad. Sci. USA* **99**, 16516–16518.
137. Mendes, P. (1997) Biochemistry by numbers: simulation of biochemical pathways with Gepasi 3. *Trends Biochem. Sci.* **22**, 361–363.
138. Goryanin, I., Hodgman, T. C., and Selkov, E. (1999) Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics* **15**, 749–758.
139. Tomita, M. (2001) Whole-cell simulation: a grand challenge of the 21st century. *Trends Biotechnol.* **19**, 205–210.
140. Schaff, J., Fink, C. C., Slepchenko, B., Carson, J. H., and Loew, L. M. (1997) A general computational framework for modeling cellular structure and function. *Biophys. J.* **73**, 1135–1146.
141. Schaff, J. C., Slepchenko, B. M., and Loew, L. M. (2000) Physiological modeling with virtual cell framework. *Methods Enzymol.* **321**, 1–23.
142. Shapiro, B. E., Levchenko, A., Meyerowitz, E. M., Wold, B. J., and Mjolsness, E. D. (2003) Cellerator: extending a computer algebra system to include biochemical arrows for signal transduction simulations. *Bioinformatics* **19**, 677–678.
143. You, L., Hoonlor, A., and Yin, J. (2003) Modeling biological systems using Dynetica—a simulator of dynamic networks. *Bioinformatics* **19**, 435–436.
144. Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T. S., Matsuzaki, Y., Miyoshi, F., et al. (1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics* **15**, 72–84.
145. Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J., and Kitano, H. (2002) The ERATO Systems Biology Workbench: enabling interaction and exchange between software tools for computational biology. *Pac. Symp. Biocomput.* 450–461.
146. Bader, G. D., Donaldson, I., Wolting, C., Ouellette, B. F., Pawson, T., and Hogue, C. W. (2001) BIND—The Biomolecular Interaction Network Database. *Nucleic Acids Res.* **29**, 242–245.
147. Xenarios, I., Salwinski, L., Duan, X. J., Higney, P., Kim, S. M., and Eisenberg, D. (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* **30**, 303–305.
148. Karp, P. D., Riley, M., Saier, M., Paulsen, I. T., Collado-Vides, J., Paley, S. M., et al. (2002) The EcoCyc database. *Nucleic Acids Res.* **30**, 56–58.
149. Wixon, J. and Kell, D. (2000) The Kyoto encyclopedia of genes and genomes—KEGG. *Yeast* **17**, 48–55.
150. Xenarios, I. and Eisenberg, D. (2001) Protein interaction databases. *Curr. Opin. Biotechnol.* **12**, 334–339.
151. Lewin, B. (1997) *Genes*, ed. 6, Oxford University, New York.