7

MATHEMATICAL MODELING OF GENETIC REGULATORY NETWORKS: STRESS RESPONSES IN *Escherichia coli*

Delphine Ropers[1], Hidde de Jong[1], and Johannes Geiselmann[2]

[1]*INRIA Grenoble-Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France*

[2]*Laboratoire Adaptation et Pathogénie des Microorganismes, Université Joseph Fourier, UMR CNRS 5163, Bâtiment Jean Roget, Faculté de Médecine-Pharmacie, Domaine de la Mercie, La Tronche 38700, France*

7.1 INTRODUCTION

Living organisms have to adapt and respond to an ever-changing environment. The genes of the organism are the basis of both immediate responses to these changes and long-term evolutionary adaptation. In fact, the functional capabilities of an organism are the result of complex interactions between the gene products encoded by its genome, and cellular functions are therefore tightly linked to the regulation of gene expression.

We call genetic regulatory network the set of genes of an organism and the molecular components controlling gene expression. This control is generally exerted by proteins

Systems Biology and Synthetic Biology Edited by Pengcheng Fu and Sven Panke

regulating the different stages of gene expression, that is, transcription, mRNA stability, translation, and protein degradation. The regulators not only include proteins but also signaling molecules allosterically modulating the activity of proteins and small RNAs regulating gene expression [1]. The ultimate effect of a regulator of gene expression is to modulate the concentration or activity of a gene product. The characteristic timescale of interactions within the genetic regulatory network is therefore generally determined by the speed of transcription and translation, ranging from several minutes for prokaryotes to several hours for higher eukaryotes.

The genetic regulatory network is connected to other cellular networks, such as signal-transduction cascades and metabolic pathways. The interactions within these networks are typically much faster than gene regulation: The average metabolic enzyme carries out a reaction cycle within milliseconds to microseconds, and most signal-transduction events also involve rapid covalent modifications, such as phosphorylation. Even though a complete description of the functioning of an organism will have to include all these networks and their interactions, the genetic regulatory network occupies a central position. Modifications of gene expression are at the very basis of developmental decisions and the response to a particular environment in the short term (adaptation) and long term (evolution). Moreover, due to differences in characteristic timescales, metabolic and signal-transduction pathways can often be seen as mediating indirect interactions on the genetic level. For instance, if a metabolite produced by a particular enzyme affects the activity of a transcriptional regulator, we can hide the molecular details of the (fast) enzymatic reactions and simply say that the gene coding for the enzyme indirectly regulates the activity of the regulator [2].

Genetic regulatory networks are the product of evolutionary processes that are better described as tinkering than engineering, in the words of François Jacob [3]. In fact, evolution does not work according to a preconceived plan, but achieves efficient performance by exploiting contingent events. It does not build an organism from scratch for a well-defined purpose, but modifies and reorganizes what is already available in order to meet arising environmental challenges. Notwithstanding these differences, many aspects of the structure and dynamics of biological systems can be compared with the principles governing man-made, engineered systems [4]. For instance, biological systems can be seen as being put together from reusable parts, assembled into modules, in much the same way as are man-made systems. Moreover, the question which aspects of the structure of a system allow it to reliably function over a range of environmental conditions, in the presence of noise, can be asked of biological and man-made systems alike. Not surprisingly therefore, living organisms have been fruitfully compared to airplanes [5] and genetic regulatory networks to electronic circuits [6]. In addition, the analogies between a signal-transduction pathway and a transistor radio have inspired some insightful comments on current biological research [7].

From the observation that the functioning of biological systems might be understood in much the same way as man-made systems, it is only a small step to applying traditional engineering methods to the study of cellular networks. This is one of the main inspirations of the emerging field of systems biology [8], and it also underlies our

contribution. More specifically, the aim of this chapter is to review different approaches toward the mathematical modeling of genetic regulatory networks, from both a theoretical and a practical point of view. Much emphasis will be put on a point that is familiar to engineers, but is often forgotten when it comes to biological modeling: A model is not a faithful copy of reality, but a simplified representation adapted to a particular type of biological questions. Instead of a single modeling approach, we therefore need a multiplicity of approaches, each capturing a different aspect of the biological system under study.

In addition to the observation that adopting an engineering approach might lead to fresh insights into the functioning of biological processes, the parallels between biological and man-made systems can be pushed further by applying engineering methods to the design of genetic regulatory networks and their actual implementation in living cells. Such networks could be useful in a variety of ways as a test bed for the study of naturally occurring networks or as a vector for biotechnological and medical applications. The rise of synthetic biology [9,10] is the second major theme of this book and is addressed in a number of other contributions, such as the chapter by Fussenegger. In this chapter, we will describe modeling approaches that can be used not only for the analysis of genetic regulatory networks but also for their design. Although, in our examples, we focus on applications in the field of systems biology, many of the theoretical and practical considerations carry over to network design in synthetic biology.

In order to illustrate the different kinds of modeling approaches, as well as the kinds of questions that can be addressed by each of these, we focus in this chapter on one particular model system: *Escherichia coli*. Although its ecological niche is the human colon, this enterobacterium has turned out to be an excellent model system for biological research as it is capable of persisting in diverse environments, easy to manipulate in the laboratory, and evolutionarily close to many pathogenic bacteria. In this chapter, we will show what the different modeling approaches have taught us about the stress responses of *E. coli*, that is, the adaptation of the bacterium to a variety of stresses, such as a lack of essential nutrients, overcrowding, and temperature shocks [11,12]. The capability to respond to challenges arising from its environment is essential to the survival of the bacterium in the short and long term.

The stress responses of *E. coli* are controlled at the molecular level by a genetic regulatory network integrating various environmental signals. The network involves the interplay of numerous signal-transduction cascades, metabolic pathways, and gene expression interactions, which together control the reorganization of the bacterial physiology and metabolism in response to a given stress [11,12]. Although many of the molecular components of the networks have been identified, currently not much is known about how the interactions between these components give rise to the cellular response to external stresses. It is clear that, when dealing with networks of this size and complexity, intuitive reasoning about the dynamical behavior of the system quickly becomes infeasible or fraught with error. This motivates the use of modeling and simulation approaches to better understand the survival strategies of the bacterium.

More generally, due to the enormous amount of information that has been accumulated about cellular interaction networks [13], *E. coli* has become a system

of choice for modeling and simulation studies. The first models on the molecular level of its response to particular nutrient shifts have appeared already in the early seventies (e.g., see Ref. [14]), while pioneering attempts to develop whole-cell models of *E. coli* adaptation appeared more than 20 years ago [15]. Recently, an International *E. coli* Alliance has been founded aiming at the coordination of modeling efforts so as to create an *in silico* cell corresponding to the bacterium [16,17].

In the remainder of this chapter, three different kinds of modeling formalisms are discussed: graphs, ordinary differential equations, and stochastic master equations. We summarize the mathematical basis of the formalisms as well as their application to the analysis of various *E. coli* stress–response networks. In particular, we investigate how these formalisms have helped address questions on (i) the structural decomposition of the stress–response network into modules and motifs, (ii) the existence of steady states and the dynamic response of the stress–response network to external perturbations, and (iii) the emergence of robust network behavior in the presence of intracellular and extracellular noise. In the concluding discussion we consider which questions are suitably addressed by each of the modeling formalisms and emphasize the point of model pluralism. For further information, the reader may wish to consult other reviews on the modeling of genetic regulatory networks [18–22].

7.2 GRAPH MODELS

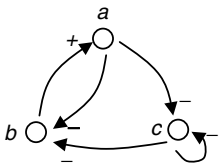
7.2.1 Model Formalism and Analysis Techniques

Probably the most straightforward way to model a genetic regulatory network is to view it as a graph. Formally, a *graph* is defined as a tuple (V, E), with V indicating a set of *vertices*, and $E \subseteq V \times V$ indicating a set of *edges* [23] as follows:

$$G = (V, E) \tag{7-1}$$

The edges represent the relation between vertices and may be directed or undirected. A *directed* edge is a pair $(i, j) \in E$ of vertices, where i denotes the head, and j denotes the tail of the edge. (i, j) is an *undirected* edge if the order of the vertices is of no importance. The vertices of a graph correspond to genes or other elements of interest in the cell, while the edges denote interactions among the genes. In the case of directed graphs, edges point from regulating to regulated genes, for example, from genes encoding transcription factors to the targets of the transcription factors. The graph representation of a genetic regulatory network can be generalized in several ways. For instance, the vertices and edges could be labeled, by adding information about genes and their interactions. Defining a directed edge as (i, j, s), with s equal to $+$ or $-$, allows one to indicate whether i is activated or inhibited by j, respectively.

An example of a simple directed graph model is shown in Figure 7-1. It consists of three genes, connected by labeled interactions that indicate whether a gene positively or negatively regulates the expression of its target. Many of the pictures of biological networks found in the literature can be mapped to some sort of graph representation. Two particularly impressive examples are the mammalian cell-cycle control



(*a*)

$$V = \{a, b, c\}$$
$$\mathrm{E} = \{(b, a, -), (c, a, -), (a, b, +), (b, c, -), (c, c, -), \}$$

(*b*)

Figure 7-1 (a) Directed, labeled graph representing a genetic regulatory network composed of three genes (*a*, *b*, and *c*) and their mutual regulatory interactions. The symbols + and − indicate whether the regulator gene activates or inhibits its target. (b) Formal definition of the graph in (a).

network [24] and the network regulating endomesoderm specification in the sea urchin [25].

The representation of a genetic regulatory network as a graph allows the analysis of its *topological* properties by means of graph–theoretical techniques [26,27]. The global connectivity properties of the network can, for instance, be described by the average degree and the degree distribution of the vertices. The *degree k* of a vertex indicates the number of edges to which it is connected (if necessary, incoming and outgoing edges can be distinguished). $\langle k \rangle$ denotes the *average degree* and $P(k)$ denotes the *degree distribution* of the graph. These properties give an indication of the complexity of the graph and allow different types of graphs, and therefore networks, to be distinguished (Fig. 7-2). In classical *random graphs* (Fig. 7-2a), also called Erdős-Rényi graphs, the probability that a given vertex has k edges follows a Poisson distribution $P(k)$. That is, the vertices typically have $\langle k \rangle$ edges, and the vertices having significantly more or less edges than $\langle k \rangle$ are extremely rare, as shown in part (c) of the figure. By contrast, in *scale-free graphs* (Fig. 7-2b), the vertex degrees obey a power-law distribution $P(k) \sim k^{-\gamma}$, shown in part (d) of the figure. Scale-free graphs are inhomogeneous, in the sense that most of the vertices have few edges, whereas some vertices, called *hubs*, have many edges and hold the graph together.

For values of the degree exponent γ between 2 and 3, scale-free graphs have a number of surprising properties. First, the average length of the path between two vertices of the graph is proportional to log log $|V|$, where $|V|$ denotes the number of vertices of the graph [26,27]. This is even shorter than the average path length in random graphs, which scales as log $|V|$ and confers them the *small-world property* [28]. The small-world property implies that local perturbations can quickly spread out through the entire network. Second, the presence of hubs makes scale-free graphs robust against accidental failures [29–31]. Whereas randomly removing a certain number of vertices disintegrates a random graph, in a scale-free graph this mainly affects the numerous low-degree vertices, the absence of which does not decompose the graph. Finally, a scale-free topology may also confer robustness to the dynamical

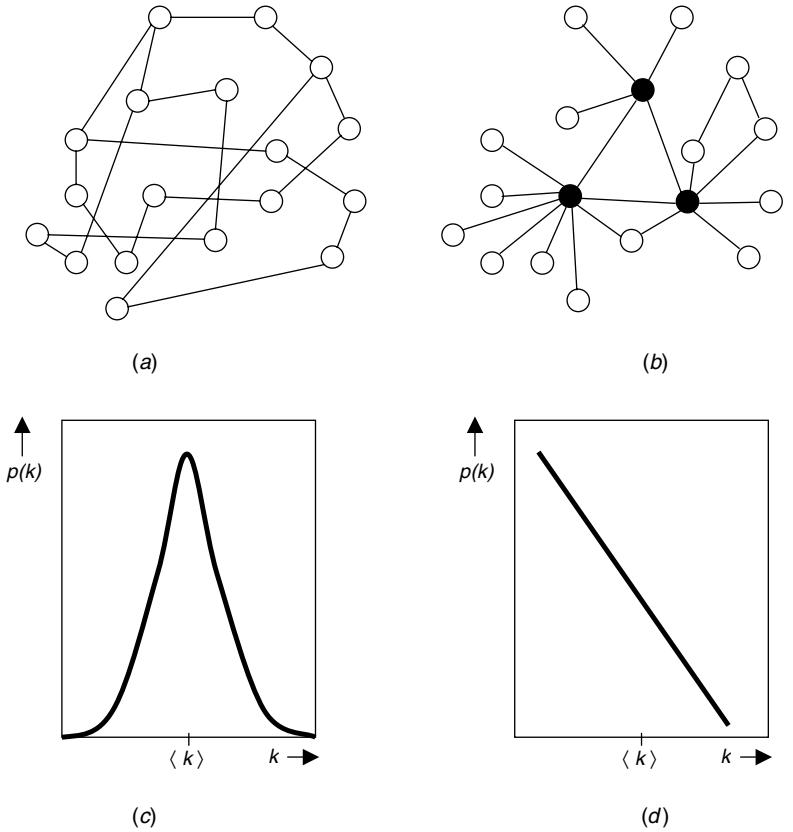


Figure 7-2 Schematic illustration of the architecture of (a) random and (b) scale-free undirected graphs [170]. The degree distribution follows (c) a Poisson distribution in random graphs and (d) a power-law distribution in scale-free graphs. k denotes the degree of a vertex and $P(k)$ denotes the degree distribution. The filled vertices in (b) are hubs.

properties of the network [32], suggesting that the latter are relatively insensitive to the precise values of the parameters (Section 3.1).

The relation between the scale-freeness of a graph and such fundamental properties of living systems as reactivity and robustness makes this type of graph interesting as a model of genetic regulatory networks. In recent years quite some evidence has accumulated, showing that genetic regulatory networks, and many other biological and nonbiological networks, are indeed scale-free [30,31,33–39]. The results should be interpreted with some care though. Because current data on regulatory interactions are incomplete, only subnetworks of the actual networks can be analyzed, which may have a different degree distribution [40,41]. Moreover, the particular graph representation chosen to model the network may bias the results, as shown by Arita for the *E. coli* metabolic network [42]. Further, in the case of genetic regulatory networks, graph models are usually restricted to direct transcriptional regulatory interactions,

thus ignoring indirect interactions that are mediated by metabolites binding to transcription regulators [43].

Another example of the use of graph analysis is to study the structural decomposition of a graph into subgraphs. Here, we focus on two kinds of subgraphs: modules and motifs. Informally speaking, a *module* is a (possibly hierarchically structured) cluster of vertices, such that the vertices within a module are strongly connected, while the connections between modules are much looser. The formalization of this intuition can be achieved by means of different graph–theoretical concepts, for instance clustering coefficients [44], shortest path distances [45], and edge betweenness, denoting the number of shortest paths between pairs of vertices that run through an edge [46]. On a different level of granularity, *motifs* are small subgraphs, consisting of a few vertices only, which frequently recur in the graph [47–49]. More precisely, motifs are defined with respect to a statistical background consisting of a randomized version of the graph: a small subgraph is called a motif if it occurs significantly more often in the original graph than in the randomized graphs.

The interest of the structural decomposition of a graph into modules and motifs is that the latter may correspond to a particular function of the genetic regulatory networks. Some results validating this intuition will be presented below, in the case of the *E. coli* transcriptional regulatory network [50,51]. One should be careful in interpreting the results of such graph analyses though. As mentioned above, currently available data are incomplete and specific modeling choices may introduce a bias. Moreover, this point needs to be emphasized, the relation between topological concepts like modules and motifs on the one hand, and the functioning of biological systems on the other, is far from straightforward. Consider the example of a module.

Even if some of the genes in the module are known to play a role in a particular biological function, this may not be sufficient for concluding that the module is responsible for the function. For instance, some of the interactions between the genes in the network may not be operative at all under the physiological conditions where the function is called upon. One could counter this objection to some extent by integrating other kinds of data in the process of module identification, such as transcriptome data, phylogenetic profiles, and biological sequences (e.g., see Refs. [52–59]). This certainly allows for more refined answers to the question *which* genes and interactions are relevant for a particular biological function. However, by itself it does not explain *how* the function emerges from the genes and interactions in the module. In order to deal with the latter questions, we need dynamical models of the kind discussed in later sections of this chapter.

In summary, graph models of genetic regulatory networks allow one to address questions concerning the network topology, giving insights into global structural properties like the edge distribution. In addition, they enable the identification of local substructures like modules and motifs that may be related to the functioning of the biological system. Graph models are applicable to genome-scale models, and the computer tools to support the analysis exist, such as mfinder [60] and TopNet [41]. However, in order to clarify the relation between the topological properties of graph models and the functioning of genetic regulatory networks, more powerful dynamical

models are necessary. The capabilities and limitations of graph models will be illustrated in the next section.

7.2.2 Modules and Motifs in the Transcriptional Regulatory Network of *E. coli*

Transcription factors are key components in the control of the *E. coli* stress responses, in that they link the sensing of environmental changes to the reorganization of the pattern of gene expression, and thus to the control of metabolic pathways. Depending on the environmental conditions, different sets of transcription factors are used by the bacterium. We would therefore like to ask such questions as: Can we relate the stress adaptation capabilities of *E. coli* to the topological organization of its transcriptional regulatory network? More precisely, are the different sets of genes organized in modules, for example, can we define a carbon utilization module and a nitrogen assimilation module? And more generally, how can we define such modules and detect them from a graph model of the regulatory dependencies of the genes of the bacterium?

The topological analysis of the *E. coli* network has been much facilitated by the rich store of information about the components of the network and their interactions, which are published in the literature or stored in databases, like RegulonDB [61] or EcoCyc [62]. Hence, several studies aiming at the analysis of structural properties of genetic regulatory networks by means of the approaches mentioned in the previous section have exploited the information on the *E. coli* network stored in RegulonDB [33,49,63–65].

An example of the search for modules is the study by Resendis-Antonio and collaborators [65]. Considering only genes for which experimental evidence on their involvement in regulatory interactions is available, the authors analyzed a network composed of 55 transcription factors controlling the expression of 747 genes. The relations between the genes in the network of transcription factors were determined by computing the shortest path distance for every pair of genes. Based on this information, eight topological modules were identified using a clustering approach. Further analysis revealed that the modules are composed of functionally related genes, for example, involved in (i) respiration, (ii) stress response, and (iii) chemotaxis, motility, and biofilm formation. The largest module (iv) gathers genes involved in the assimilation of the various carbon sources. The remaining modules are composed of genes involved in various cellular responses, like (v) sulphur assimilation, (vi) nitrogen metabolism, (vii) fermentative conditions, and (viii) chromosome replication.

The topological analysis of Resendis-Antonio *et al.* suggests that the *E. coli* network possesses a modular structure, where each module consists of genes that perform one or more tasks in response to particular environmental conditions. For instance, the *nac* and *asnC* genes, coding for transcription factors known to be involved in the control of nitrogen assimilation, are found inside the same module [65]. Further, the carbon assimilation module includes transcription factor genes like *crp*, *araC*, *malF*, *fruR*, which regulate the utilization of carbon sources. Interestingly, the latter module can be further decomposed into submodules, each submodule being specialized in the use of a

different carbon source. One can guess that this supplementary internal organization makes *E. coli* cells able to easily grow on various carbon sources and switch from the use of one carbon source to another. Other analyses have found a similar modular structure of the *E. coli* transcriptional regulatory network, though using different approaches and arriving at a different number of modules [59,63].

Several questions regarding the *E. coli* network structure remain open. For instance, how is the global coordination of cellular responses to be explained? Most often, a stress response does not involve a single module but rather a combination of modules. For instance, *E. coli* uses its motility, controlled by module (iii), to seek optimal oxygen concentrations, required for the respiration task performed by module (i). It seems obvious that accomplishing this function requires a connection between the two modules, which agrees with the fact that, generally, the modules are not clearly separated from the rest of the network but tend to overlap [65]. Can these interconnections be characterized by means of certain topological properties?

To address this question, we need to take into account the local topology of networks, defined in terms of motifs. Using information from RegulonDB and the literature, Shen-Orr et al. [49] have analyzed the transcriptional regulatory network of *E. coli*. They found that in this network, consisting of 855 genes and 1330 regulatory interactions, three different motifs occur more frequently than expected: the *feedforward loop*, in which a transcription factor regulates a second transcription factor and both regulate together a target gene; the *single-input* motif, in which a group of genes is controlled by a single transcription factor; and the *dense overlapping regulons*, in which genes and the transcription factors controlling their expression form a highly overlapping structure. The feedforward loop is the motif occurring most frequently (40 times) in the *E. coli* network. This has been subsequently confirmed by means of an extended version of the same network, in which an even higher number of feedforward loop motifs were found [63]. The different motifs are not equally distributed in the network of *E. coli*. In the above-mentioned study, Resendis-Antonio and collaborators found that the feedforward loop motifs are mainly located inside modules (71 percent of the cases), whereas the *bifan* motif (which forms the basic building block of the above-mentioned dense overlapping regulons) is the main motif connecting modules (65 percent of the cases).

What is the advantage for the cell of conserving certain network motifs? Do they have a functional role, in addition to their structural role? The group of Alon demonstrated both theoretically and experimentally the information-processing task carried out by the coherent feedforward loop. Using a differential equation model of the feedforward loop motif, they showed that its role might be to filter out fluctuations in input stimuli and allow a rapid response when the stimuli disappear [66,67]. Consider the coherent feedforward loop motif in Figure 7-3, where the transcription factors X and Y together activate the gene *z*. When X is active and above a threshold concentration, the input signal activating X is transmitted to the output Z through a direct path from X and an indirect path from X through Y. Hence, a transient signal is not transmitted to Z, since it does not allow the concentration of Y to reach a threshold level high enough to stimulate the expression of gene *z* (Fig. 7-3). On the contrary, a persistent input signal enables the concentration of Y to rise and eventually

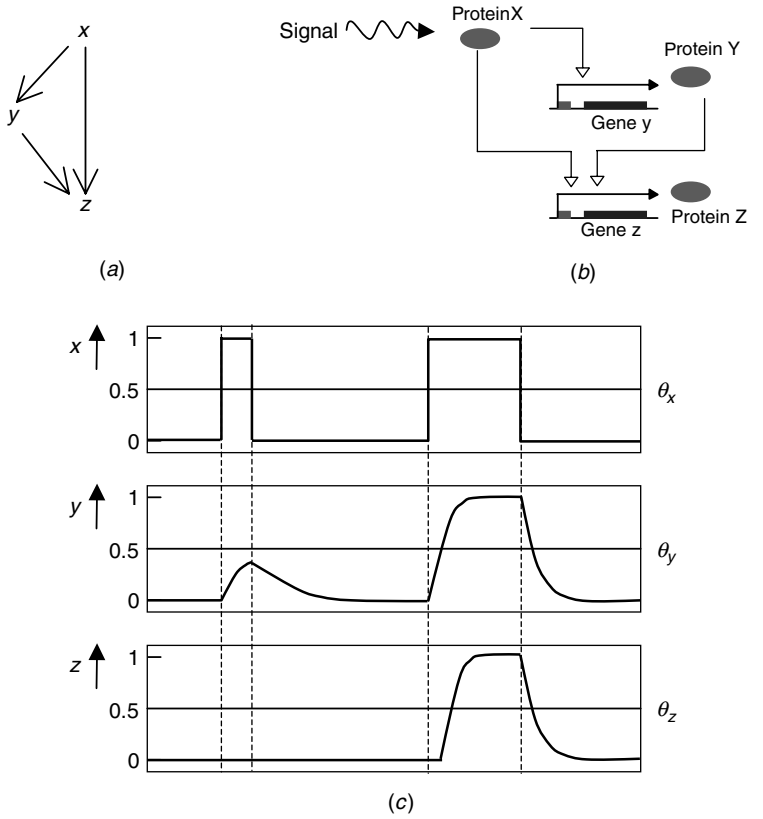


Figure 7-3 (a) Coherent feedforward loop motif in a graph representation. (b) Feedforward loop in a genetic regulatory network, where it is assumed that both X and Y are necessary for expression of *z*. (c) Dynamic properties of the feedforward loop [49]; *x*, *y*, and *z* denote the concentrations of X, Y, and Z, respectively, while θ_x, θ_y, and θ_z denote their threshold levels. The input signal activates X.

allows Z to pass its threshold level. The functioning of the feedforward loop motif is asymmetric, since the inactivation of X leads to the rapid downregulation of *z*. The above predictions have been experimentally verified for the L-arabinose utilization system in *E. coli* using reporter genes [67]. In this feedforward loop motif, CRP corresponds to the general transcription factor X and AraC to the specific transcription factor Y, while *z* is the operon *araBAD*.

The study by the group of Alon assigns a function to a pattern of interactions, the coherent feedforward loop, which is overrepresented in the transcriptional regulatory network of *E. coli*. However, the relation between structure and function is not straightforward, given that motifs do not usually occur in isolation, but rather overlap to generate *motif clusters* [33]. Does the function of a motif change when it is embedded within a network and interacting with many other components? The group of Alon partially answered this question in a subsequent study on the *incoherent feedforward loop*, that is, a feedforward loop in which the transcription factor X activates genes *y* and *z*, while Y represses *z* [68]. The motif was experimentally shown

to perform response acceleration, as predicted by a differential equation model [66], despite the fact that it participates in additional interactions that were not included in the model. However, it is not sure that this will turn out to be true in general.

7.3 ORDINARY DIFFERENTIAL EQUATION MODELS

7.3.1 Model Formalism and Analysis Techniques

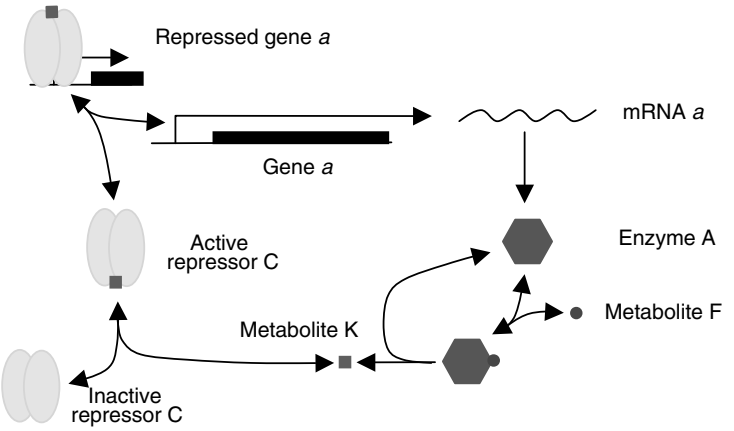
As concluded in the previous section, a better comprehension of the relation between the structure and functioning of a regulatory system requires the use of dynamical models. *Ordinary differential equations* [69] are probably the most-widespread formalism for modeling the dynamical behavior of cellular interaction networks. In this formalism, the concentrations of gene products (mRNAs or proteins) are represented by continuous, time-dependent variables, $x(t)$, $t \in T$ and T being a closed time interval ($T \subseteq \mathbf{R}_{\geq 0}$). The variables take their values from the set of nonnegative real numbers (x: $T \rightarrow \mathbf{R}_{\geq 0}$), reflecting the constraint that a concentration cannot be negative.

The regulatory interactions between genes are modeled by a system of ordinary differential equations having the following general form:

$$\mathrm{d}x_i/\mathrm{d}t = f_i(x), \quad i \in \{1, \ldots, n\}, \tag{7-2}$$

where $x = (x_1, \ldots, x_n)'$ is the vector of concentration variables of the system, and the usually highly nonlinear function f_i: $\mathbf{R}_{\geq 0}{}^n \rightarrow \mathbf{R}_{\geq 0}$ represents the regulatory interactions. The above system of equations describes how the temporal derivative of the concentration variables depends on the values of the concentration variables themselves. Several variants of Equation 7-2 can be imagined [22]. For instance, by taking into account input variables u, it becomes possible to express the dependence of the temporal derivative on external factors, such as the presence of nutriments. In order to account for the delays resulting from the time it takes to complete transcription, translation, and the other stages of the synthesis and the transport of proteins, Equation 7-2 could be changed into a system of delay differential equations.

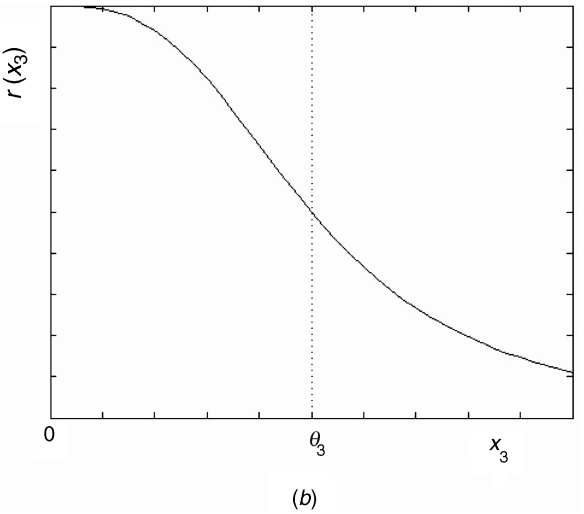
In Figure 7-4, an example of a simple genetic regulatory network and its associated differential equation model is shown, based on early work by Goodwin [70,71]. The end product of a metabolic pathway coinhibits the expression of a gene coding for an enzyme that catalyzes a reaction step in the pathway. This gives rise to a negative feedback loop involving the mRNA concentration x_1, the enzyme concentration x_2, and the metabolite concentration x_3. The equations each express a balance between the increase and decrease of the molecular concentration per unit time due to the occurrence of the various reactions. More precisely, the equations describe the rate of synthesis of the enzyme (k_2x_1) and the metabolite (k_3x_2), as well as the rate of synthesis of mRNA ($k_1r(x_3)$). The nonlinear, sigmoidal Hill function r expresses that the rate of synthesis of mRNA depends in a cooperative way on the concentration of the metabolite, which binds and thereby activates a repressor of the gene (Fig. 7-4b). The terms $-g_1x_1$, $-g_2x_2$, and $-g_3x_3$ indicate that the concentrations x_1, x_2, and x_3 decrease through degradation and growth dilution, at a rate proportional to the concentrations themselves.



(*a*)

$$\mathrm{d}x_1/\mathrm{d}t = k_1\, r(x_3) - g_1\, x_1$$
$$\mathrm{d}x_2/\mathrm{d}t = k_2\, x_1 - g_2\, x_2$$
$$\mathrm{d}x_3/\mathrm{d}t = k_3\, x_2 - g_3\, x_3$$

$$r(x_3) = \theta_3^{\,n} / (\theta_3^{\,n} + x_3^{\,n})$$



(*b*)

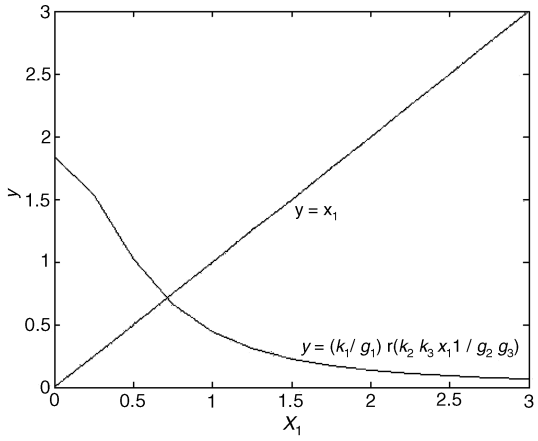
Figure 7-4 (a) Simple example of gene regulation involving end-product inhibition and (b) the corresponding differential equation model. A is an enzyme and C a repressor protein, while K and F are metabolites. x_1, x_2, and x_3 represent the concentrations of mRNA *a*, protein A, and metabolite K, respectively, k_1, k_2, k_3 are production constants, g_1, g_2, g_3 degradation constants, and $r\colon \mathrm{R}_{\geq 0} \rightarrow \mathrm{R}_{\geq 0}$ is a decreasing Hill function ranging from 0 to 1, with threshold parameter θ and exponent n. All parameter values are positive and $n > 1$, in order to obtain the sigmoidal shape characteristic of cooperative interactions.

A first dynamical property that can be studied by means of ordinary differential equation models is the *asymptotic behavior* of the system, notably the occurrence of equilibrium points and limit cycles, as well as their stability and basin of attraction. The equilibrium points and limit cycles may correspond to functional modes of the systems, for instance a particular growth stage or a particular response of the cell to an external stress. The equilibrium points are simply determined by setting every dx_i/dt given in Equation 7-2 to 0 and solving for x_i. In the example of the end-product inhibition network, we thus obtain a single equilibrium point [72]. This follows from Equation 7-2 by noting that at equilibrium $x_1 = (k_1/g_1)\, r(k_2k_3x_1/g_2g_3)$, and bearing in mind that r is a monotonically decreasing function (Fig. 7-5a).

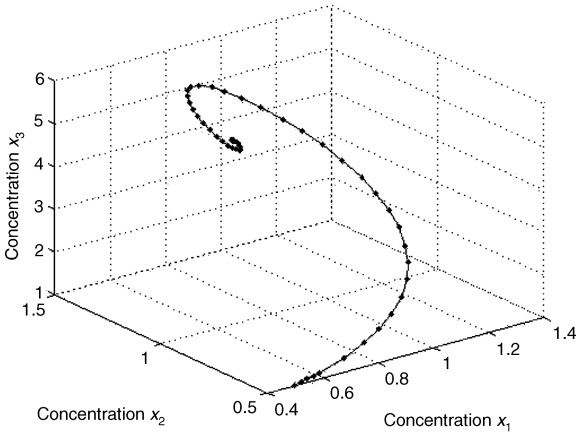
The stability of the equilibrium point $\boldsymbol{x}^*$ can be determined by linearizing the system of differential equations given in Equation 7-2 around $\boldsymbol{x}^*$, computing the characteristic equation, and solving for the eigenvalues. The sign of the (real part of the) eigenvalues then determines the stability of the system [69,73]. The characteristic equation for the end-product inhibition network is given by $(\lambda + g_1)\ (\lambda + g_2)\ (\lambda + g_3) - k_1k_2k_3\ \partial r(x_3^*)/\partial x_3 = 0$, where $x^* = (x_1^*, x_2^*, x_3^*)'$. The equation can be rewritten as a third-order polynomial whose roots λ are the eigenvalues. Depending on the exact numerical values of the parameters, different configurations of eigenvalues are found, notably (i) three negative real eigenvalues, (ii) a negative real eigenvalue and two conjugate complex eigenvalues with negative real part, or (iii) a negative real eigenvalue and two conjugate complex eigenvalues with positive real part. In the former two cases, the equilibrium point is asymptotically stable, meaning that after a (small and temporary) perturbation the system will eventually return to the equilibrium point. In contrast, in the third case the equilibrium point is unstable: a perturbation will cause the system to diverge from the equilibrium point and approach a stable limit cycle, corresponding to sustained oscillations in the protein concentrations. Figure 7-5b illustrates case (ii) for arbitrary but not unrealistic parameter values.

A second dynamical property of interest is the *transient behavior* of the system. The transient behavior provides information on the manner in which the genetic regulatory network controls the response of the system to an external perturbation, for example, by switching from one functional mode to another. In order to predict the transient behavior, we need to compute the solutions of the system of ordinary differential equations 7-2. Since the models of most genetic regulatory networks of practical interest are nonlinear, it is usually not possible to find an analytical solution. This means that in all but the simplest cases we have to resort to *numerical simulations* [74], which yield approximations of the exact solutions. The solutions obtained by simulation can be visualized by plotting their trajectories in the phase space, for two or three-dimensional systems, or by simply plotting the solutions as a function of time. This is illustrated in Figure 7-5b and c for the model of the end-product inhibition network. The plots show how the system adapts to a perturbation from its steady state, by returning to this state through damped oscillations.

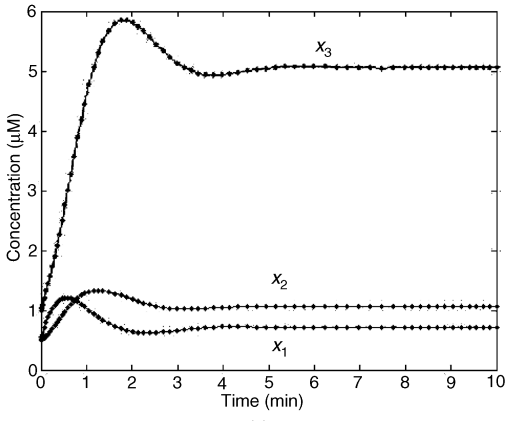
The analysis of the feedback inhibition network shows that it is a homeostatic system, with a tendency to maintain a stable steady state or stable oscillations. The negative feedback loop, arising from the (indirect) inhibition of the expression of the gene by its own product, tends to compensate for a transient perturbation. Examples of


y = x1
y = (k1/ g1) r(k2 k3 x1 1 / g2 g3)
y
X1

(*a*)


Concentration x3
Concentration x2
Concentration x1

(*b*)


x3
x2
x1
Concentration (μM)
Time (min)

(*c*)

negative feedback loops abound in biological systems and play an important role in gene expression, metabolism, and signal transduction (e.g., see Refs. [75–77]). More generally, Thomas conjectured that negative feedback loops are a prerequisite for homeostasis [78,79]. In a similar vein, he proposed that positive feedback loops are a necessary condition for the occurrence of multiple steady states, corresponding to different functional modes of the system. Several proofs of the latter conjecture have been given under increasingly general conditions [80–84]. These results illustrate the potential of mathematical models to highlight fundamental relations between the topology and the dynamics of regulatory networks.

The qualitative dynamics of the end-product inhibition network, the stability of the equilibrium point and the occurrence of a stable limit cycle, are determined by the values for parameters in the differential equation model in Figure 7-4. For large ranges of parameter values, the qualitative dynamics of the system remains invariant, that is, the qualitative dynamics is *robust* to fluctuations in the parameter values. This robustness is an essential property of living systems, which have to cope with continuous fluctuations in physiological and environmental conditions as well as with genetic variability. Following pioneering work by Savageau (1971), the study of the robustness of dynamical properties of regulatory networks to changes in the parameter values has been an active research area, demonstrating robust behavior of the chemotaxis system of *E. coli* [85,86], the development of the *Drosophila* embryo [87,88], the *Xenopus* cell cycle [89], and the circadian clock of *Drosophila* [90,91]. In the case of synthetic networks, the ability of the system to reliably function in the presence of noise is an important design objective. Control theory provides a range of methods that could be used to assess the robustness of naturally occurring networks and improve the robustness of synthetic networks (e.g., see Refs. [91–94]).

Although differential equation models allow making precise, quantitative predictions on the dynamics of large and complex genetic regulatory networks, they may be difficult to apply in practice. Most regulatory networks of interest are large and complex, possibly involving hundreds of genes, proteins, and other molecules. If these networks were to be modeled in the same way as the simple autoinhibition network in Figure 7-4, we would obtain huge models that cannot be analyzed other than by massive numerical simulations. Apart from the fact that such simulations may be difficult to carry out, given that numerical values for the parameters are often not available (see below), it is not sure that the generation of time-course predictions of hundreds of molecular components will be of much help in gaining a better understanding of the functioning of the system. This has stimulated an interest in strategies for model simplification, often based on indications that the networks have a modular structure (Section 7.2). In order to study large and complex networks, it may be more

Figure 7-5 (a) The differential equation model of the end-product inhibition network in Fig. 7-4 has a single equilibrium point. The stability of this equilibrium point varies with the parameter values. (b) Asymptotically stable equilibrium point with a trajectory spiraling toward this point. (c) Time-series representation of the solution. The parameter values are as follows: $k_1 = 4.6\,\mu$M/min, $k_2 = 1.8$/min, $k_3 = 10$/min, $g_1 = 2.5$/min, $g_2 = 1.2$/min, $g_3 = 2.1$/min, $\theta = 4\,\mu$M, and $n = 2$. The simulations have been carried out with Matlab.

judicious to first analyze the modules individually, and only afterward the question of how they are woven together, preferably using simpler and more abstract models for this second step. The definition of network modules may be based on topological criteria, not unlike those used for graph models but usually more directly relevant to the dynamics of the system, such as the feedback structure of the network [95–97]. Another way to define modules is based on the distinction between rapid and slow processes in the system, for example, allowing the separation of metabolism and gene expression in separate modules [98,99].

Even after model simplification, for most networks we will be left with large and complex models. Their analysis requires quantitative information on the values of kinetic constants and molecular concentrations, but unfortunately this information is only rarely available, especially when modeling systems on the forefront of experimental research. Several ways to deal with this problem have been proposed in the literature. First of all, pushing the robustness argument further, one could argue that important dynamical properties of actual regulatory networks do not so much depend on particular molecular mechanisms or precise values for the parameters, but rather on the network topology. A second strategy is to try to estimate the parameter values from experimental data [100]. The use of these techniques has been shown to work well on small to medium-sized systems, in cases where the interactions are well described by linear or quasilinear functions (e.g., see Refs. [101–104]). A third way out would be turn to simplified models, having a particular mathematical form that simplifies their analysis [105,106]. Examples of such models are the piecewise-linear differential equation models proposed by Glass and Kaufmann [107] or the logical models proposed by Kauffman [108] and Thomas [109,110].

In conclusion, differential equation models allow questions related to the transient or asymptotic dynamics of genetic regulatory networks to be answered. Many examples of their application exist, some of which will be discussed later in the context of the *E. coli* stress response. Techniques for the mathematical analysis and numerical solution of differential equation models are standard engineering tools, and a large variety of computer programs are available, ranging from general-purpose mathematical problem solvers like Matlab to tools specifically adapted to the analysis of cellular interaction networks such as Copasi [111], ProMoT/DIVA [112], Virtual Cell [113], and XPPAUT [114].[1] Due to the size and complexity of networks of practical interest as well as the lack of precise, quantitative information on the molecular mechanisms and kinetic constants, standard techniques for numerical analysis may be difficult to apply in practice. Several strategies to cope with this problem have been proposed, some of which will be illustrated in the next section.

7.3.2 Response of *E. coli* to Carbon-Source Availability

As seen in Section 2.2, the transcriptional regulatory network of *E. coli* contains a carbon assimilation module, allowing the bacterium to use a large range of carbon

[1] The SBML format [115] allows models to be exchanged between different computer tools. A list of computer tools compatible with the SBML format is available at http://www.sbml.org.

sources under a variety of conditions. For instance, when several carbon sources are available, the bacteria choose the "best" nutrient, meaning the nutrient sustaining fastest growth. Hence, if *E. coli* is presented with two carbon sources, for example glucose and lactose, it starts using glucose until this preferred nutrient is depleted from the medium. Growth then temporarily arrests while the bacterium modifies its pattern of gene expression so as to produce the enzymes necessary for the uptake and metabolism of lactose. This physiological response is referred to as *diauxic growth*. When all carbon sources in the growth medium have become depleted, *E. coli* bacteria are no longer able to sustain fast growth rates and enter into a *stationary phase* of growth, characterized by no net change of the size of the bacterial population. In response to carbon starvation, the bacteria completely modify their physiology to cope with the absence of nutrients. This implies that they conserve energy by shutting off most biosynthetic functions and protect their DNA from potential damage, while at the same time maintaining a minimal metabolism in order to explore potential alternative nutrient sources and "be ready" as soon as nutrients become available again.

Given the numerous tasks ensured by the carbon assimilation module, several questions arise regarding its functioning: How does the module coordinate the different responses of *E. coli* cells to carbon-source availability? How does the reorganization of gene expression and metabolism emerge from the interactions between the many components making up the regulatory network of *E. coli*? Can we develop comprehensive dynamic models of these interactions that account for the bacterial responses to carbon-source availability?

In the remainder of this section, we give two examples of differential equation models, describing, respectively, the diauxic growth of *E. coli* and its response to carbon starvation. Even though *E. coli* is a well-studied system, the development of the models has been limited by the lack of quantitative information on most of the molecular concentrations and the kinetic parameters characterizing the interactions inside the carbon assimilation module. To overcome these constraints, different strategies were chosen, one based on the estimation of parameter values from experimental data [116] and the other on a more abstract description of the network [117].

The group of Gilles has developed a dynamical model describing the successive assimilation of different carbon sources in *E. coli* (glucose, lactose, etc.), leading to diauxic growth [116]. The central part of the regulatory network controlling this process is a large, membrane-bound enzyme complex called *Phospho-Transferase System (PTS)*. The PTS transfers a phosphate onto the carbohydrates (e.g., glucose), which makes the transport irreversible and prepares the carbohydrate for metabolic breakdown and conversion into cellular energy. In the absence of glucose, the same complex activates another membrane-bound enzyme, adenylate cyclase (Cya). Cya produces a signaling molecule, cyclic adenosine mono-phosphate (cAMP), which in turn binds a transcription factor, CRP (cAMP receptor protein) and enables the latter to activate or inhibit transcription. The promoter of the lactose operon is one of the targets activated by cAMP–CRP. The same promoter is also under the negative control of the *lac* repressor. This transcription factor is inactivated by a metabolite, allolactose, which is produced in the presence of lactose. This allows derepression of the transcription of the operon and the subsequent use of lactose as a carbon source.

The model by Bettenbrock and colleagues is the last in a series of detailed models of the carbon assimilation module developed by the group of Gilles [118–121]. While other models of the same system are available in the literature (e.g., see Refs. 122–124), the Bettenbrock model provides the most comprehensive picture to date. It describes the PTS and its interactions with several uptake systems and metabolic pathways, accounting for the growth of *E. coli* on different carbohydrates. The network is composed of different types of interactions, involving genetic regulatory interactions, metabolic reactions, and reactions involved in the signal-transduction pathway. These were modeled by ordinary differential equations of the form described in Section 3.1, using kinetic rate laws appropriate for each type of interaction, and algebraic equations expressing conservation relations among the different molecular components of the system. In total, the model is composed of 50 differential equations and 14 algebraic equations.

Even though the network controlling diauxic growth is a well-characterized system, it was not always possible to include parameter values reported in the literature in the model, as they are often obtained under different experimental conditions and with different strains. To circumvent this problem, Bettenbrock and colleagues have carried out their own experiments, measuring the concentration of the various metabolites over time and have used the resulting data to estimate the value of the model parameters by means of the ProMoT/Diva environment [112]. In this way, some fifty uncertain or unknown parameter values could be obtained.

By means of the resulting numerical model, *E. coli* growth on various carbohydrates was simulated (see Fig. 7-6a, for example). The confrontation of these predictions with time-series measurements performed under the experimental conditions corresponding to the simulations revealed a number of contradictions that required model revision. For instance, the model could not account for the behavior of the system during disturbed batch experiments, consisting of the exponential growth of the cell on a carbon source (glycerol or lactose), followed by the application of a pulse of glucose. Although the simulations showed glucose uptake, as observed experimentally, the process was predicted to proceed too fast (Fig. 7-6b). The inclusion into the model of the regulation of the *pts* operon allowed a much better fit of the experimental data with the model predictions. In this instance, the model not only confirmed what is currently known about the accumulation of carbon sources by *E. coli* but also provided novel explanations of the role of certain network components in the process. Hence, the cAMP metabolite appears to play a key role in the short-term adaptation to a new carbon source during diauxic growth, whereas the complex cAMP–CRP seems to be more important for long-term adaptation.

The model of Bettenbrock and colleagues provides a detailed and rigorous description of the molecular events underlying diauxic growth. However, the model does not address the functioning of the carbon assimilation module in the broader context of the genetic regulatory network of *E. coli*. For instance, it is known that the PTS is closely connected to some major transcription regulators of the bacterium called *global regulators*. These transcription factors control the expression of large sets of genes in response to environmental stimuli [125,126]. More precisely, the PTS transfers information on the lack of carbon source to the global regulators, which reorganize gene expression and allow the bacteria to stop exponential growth and enter

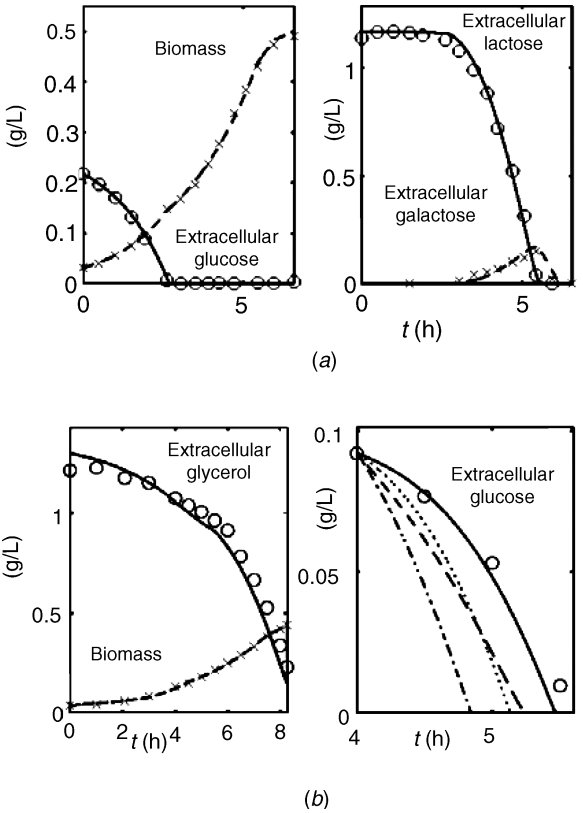


Figure 7-6 Differential equation model of the carbon assimilation module: confrontation of model predictions and experimental measurements [116]. The circles denote measurements, and the lines denote simulation results. The biomass and the extracellular concentrations of carbohydrates are mentioned on the curves. (a) Diauxic growth on glucose and lactose. Galactose is a product of lactose metabolism. (b) Disturbed batch experiment with application of a pulse of glucose on bacteria growing on glycerol. Dashed lines denote simulation results from different versions of the model that do not take into account regulation of the *pts* operon expression. Simulation of the model including this additional regulatory interaction results in the continuous line.

stationary phase upon carbon starvation. How does the growth adaptation of *E. coli* emerge from the interactions between the global regulators in response to a carbon starvation signal transmitted by the PTS?

In order to address these questions, we have developed an initial, simple model of the network of global regulators, including six genes believed to play a key role in the carbon starvation response (Fig. 7-7) [117]. The network includes genes that are targets of the PTS (the global regulator *crp* and the adenylate cyclase *cya*), genes involved in the control of metabolism (the global regulator *fis*), cellular growth (the *rrn* genes coding for stable RNAs), and DNA supercoiling, an important modulator of gene expression (the topoisomerase *topA* and the gyrase *gyrAB*). Although these genes have been the focus of intensive study over the last few decades, the development of a model

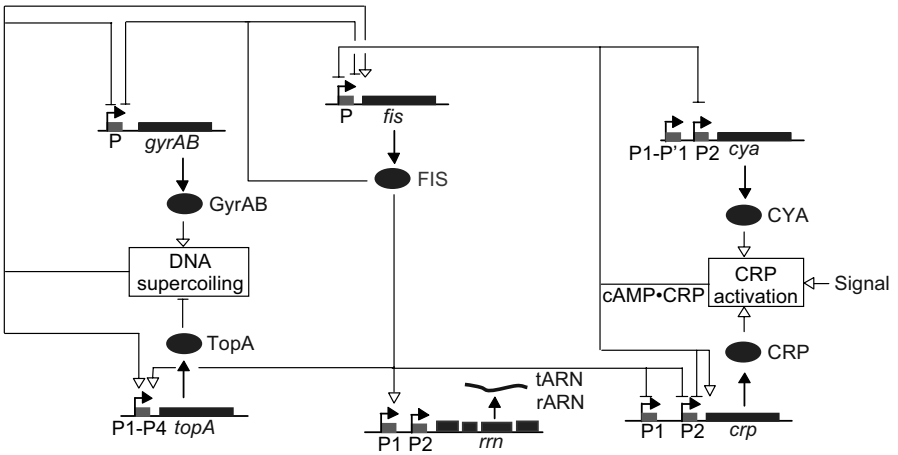


Figure 7-7 Key genes, proteins, and regulatory interactions making up the network involved in the response of *E. coli* bacteria to carbon-source availability [117].

of the network is limited by the lack of quantitative information about the concentrations of the network components and the parameters characterizing their interactions.

To overcome the lack of quantitative information, we have used a qualitative modeling and simulation method to analyze the network [127,128]. This method is based on piecewise-linear differential equations of the regulatory interactions and employs inequality constraints on the parameters to make predictions of the qualitative dynamics of the system. The piecewise-linear models of genetic regulatory networks are based on the use of step-function approximations of the sigmoidal functions describing the regulatory interactions (Fig. 7-4b). This approximation simplifies the analysis of the dynamics in that it allows the phase space to be subdivided into hyperrectangular regions where the system behaves in a qualitatively homogeneous way. The continuous dynamics of the system in the phase space can be discretized into a state transition graph, that is, a graph composed of states corresponding to the phase-space regions and transitions between these states. The state transition graph describes the possible qualitative behaviors of the system and allows the attractors of the system and their reachability to be determined.

Based on the qualitative simulations, two regulatory feedback loops were hypothesized to play a key role in the response of *E. coli* cells to carbon starvation. A positive feedback loop, involving *fis* and *crp*, seems to function as a switch controlling the transition of *E. coli* cells between the exponential and the stationary growth phase in response to a carbon starvation signal transmitted by the PTS. The other loop is a negative feedback loop, a homeostatic mechanism involving *fis* and DNA supercoiling, which regulates the resumption of cellular growth when a carbon source is available again, causing damped oscillations in certain protein concentrations. The qualitative simulations provide a description of the ordering of qualitative events (such as the upregulation and downregulation of key genes), which can be tested by monitoring gene expression over time, for instance through the use of gene reporter systems.

The assimilation of carbon sources by *E. coli*, in particular lactose, has been the subject of a large number of modeling studies (e.g., see Refs. 14,129–133). However, differential equation models have also been used to model the response of *E. coli* to other stresses, such as the response to a heat shock [96], bacteriophage infection [6,134,135], phosphate starvation [136], or the SOS response [103,136].

A common assumption underlying these models is that individual bacteria, under identical conditions, respond to a stress in exactly the same way. However, it is known that, whereas most bacteria enter a nongrowth state in response to carbon starvation, some of them continue to grow and divide. These different behaviors of individual cells arise from the stochasticity of the underlying processes that is not accounted for by the differential equation models. The next section will elaborate this point and introduce a modeling approach capable of dealing with the stochastic aspects of gene expression.

7.4 STOCHASTIC MASTER EQUATION MODELS

7.4.1 Model Formalism and Analysis Techniques

Ordinary differential equations provide a deterministic view on genetic regulatory networks, in the sense that, for given parameter values and initial conditions, Equation 7-2 has a unique solution and consequently predicts a single behavior of the system. Real biological systems, however, are not deterministic since noise arises inside and outside the system, due to fluctuations in the synthesis and degradation of proteins—strengthened by the low number of molecules of each species—and fluctuations in the environmental conditions [138–140]. As a consequence, genetically identical cells evolving under the same conditions may display different phenotypic characteristics [141,142]. In order to capture the stochastic aspects of cellular processes on the molecular level, different types of models can be used [140,143,144]. Here we focus on *stochastic master equations*, which give a detailed description of the biochemical reactions occurring in a cell.

Instead of continuous concentrations x_i, the variables in a stochastic master equation denote discrete numbers of molecules $X_i \in \boldsymbol{N}$. For each different species in the system—proteins, RNA, DNA, or metabolites—a separate variable X_i is introduced. The continuous rates of change $f_i(\boldsymbol{x})$ in ordinary differential equations are replaced by discrete reaction events occurring with a certain probability per time interval. We can write the following equation for the time evolution of the system:

$$p[X(t+\Delta t) = V, t+\Delta t] = p[X(t) = V, t](1-\sum\nolimits_{j=1,\ldots,m}\alpha_j\,\Delta t) + \sum\nolimits_{j=1,\ldots,m} p[X(t) = V-\nu_j, t]\beta_j\Delta t, \tag{7-3}$$

where $\boldsymbol{X} = (X_1,\ldots,X_n)'$, m is the number of reactions that can occur in the system, $\alpha_j\,\Delta t$ is the probability that reaction j will occur in the time interval $[t,\ t+\Delta t]$ given that $\boldsymbol{X}(t) = \boldsymbol{V}$, and $\beta_j\,\Delta t$ is the probability that reaction j will bring the system from a state $\boldsymbol{X}(t) = \boldsymbol{V} - \boldsymbol{\nu}_j$ to a state $\boldsymbol{X}(t+\Delta t) = \boldsymbol{V}$ in $[t, t+\Delta t]$, where ν_j represents the stoichiometry of the reaction. In other words, Equation 7-3 expresses that the probability of having

$\boldsymbol{V}$ molecules at time $t + \Delta t$ equals the sum of the probability of having already $\boldsymbol{V}$ molecules at t with no reaction occurring on $[t, t + \Delta t]$, and the probability of having $\boldsymbol{V}-\boldsymbol{\nu}_j$ molecules at t and reaction j occurring on $[t, t + \Delta t]$. Rearranging Equation 7-3 and taking the limit $\Delta t \rightarrow 0$ yield the stochastic master equation (see [145] and [146] for details):

$$\partial p[X(t) = V, t]/\partial t = \sum\nolimits_{j=1,\ldots,m} p[X(t) = V-\nu_j, t]\beta_j - p[X(t) = V, t]\alpha_j. \qquad (7\text{-}4)$$

Compare this equation with the ordinary differential equation given in Equation 7-2. Whereas the latter specifies how the state of the system evolves over time, Equation 7-4 describes how the probability that the system is in a certain state evolves over time. Notice that the variables in Equation 7-4 can be reformulated as concentrations by dividing the number of molecules $\boldsymbol{X}$ by the cell volume.

Figure 7-8 gives an example of a negative feedback loop that is even simpler than the one shown in Figure 7-4. It consists of a single gene a coding for a protein A that forms a dimer capable of binding to the promoter region of a, thus inhibiting the expression of the gene. The reactions involving the different molecular species of the system are shown in the figure. For instance, the dimerization of the repressor is represented by the reaction $\mathrm{A} + \mathrm{A} \rightarrow \mathrm{A}_2$. Even for this simple system, the stochastic master Equation 7-3 cannot be solved analytically. Under certain conditions, however, it can be approximated by stochastic differential equations, so-called Langevin equations, which consist of a differential equation Equation 7-2 extended with a noise term [140,146,147]. The conditions under which the approximation is valid may not always be possible to satisfy in the case of genetic regulatory networks.

An alternative way to proceed would be to disregard the stochastic master equation altogether and directly simulate the time evolution of the regulatory system. This idea underlies the stochastic simulation approach developed by Gillespie [145]. Basically, the stochastic simulation algorithm (i) determines *when* the next reaction occurs and of *which* type it will be, given that the system is in a state $\boldsymbol{X}(t) = \boldsymbol{V}$ at t, (ii) revises the state of the system in accordance with this reaction, and (iii) continues at the resulting next state. The stochastic variables τ and ρ are introduced, which represent the time that has passed until the next reaction occurs and the type of reaction, respectively. At each state a value for τ and ρ is randomly chosen from a set of values whose joint probability density function $p[\tau, \rho]$ has been derived from the same principles as those underlying the master equation 7-4. This guarantees that when a large number of stochastic simulations are carried out, the computed distribution for $\boldsymbol{X}$ at t will approach the distribution implied by the master equation.

It is obvious that stochastic simulation is a computationally intensive process, especially when dealing with species involving a large number of molecules and/or with reactions occurring at high frequency. Examples are metabolic reactions, which may occur millions of times on the timescale of one generation of a bacterial cell [148]. Another reason is that a large number of different molecular species may need to be taken into account, for instance when a protein has a large number of phosphorylation or methylation states, each of which participates in different reactions and therefore needs to be treated as a separate species [149,150]. Various improvements of the

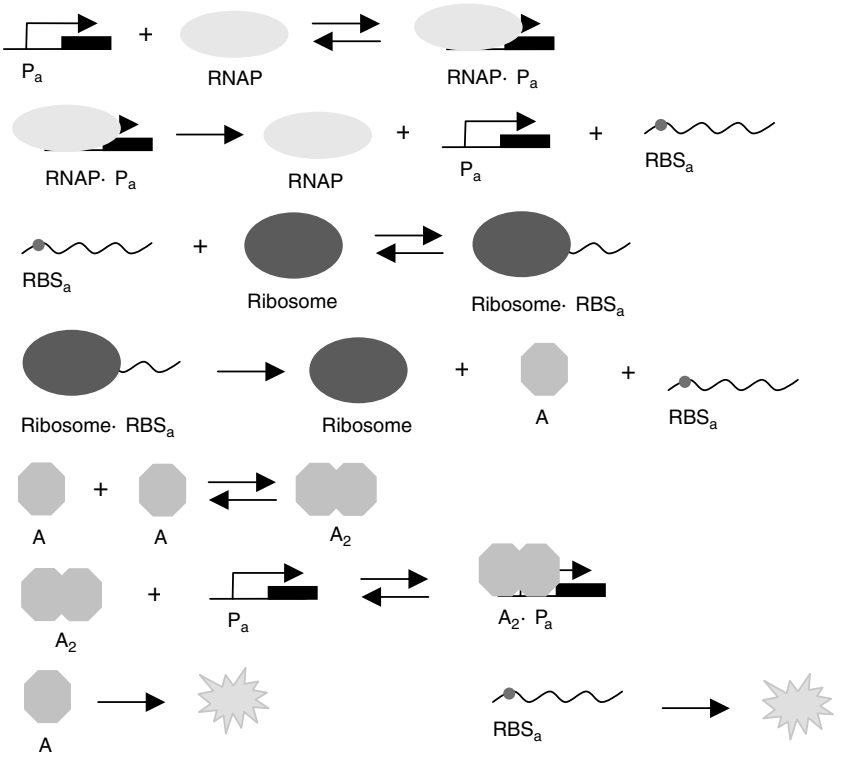


Figure 7-8 List of biochemical reactions in a simple autoinhibition network. The gene *a* encodes a repressor protein A that forms a dimer A_2. Gene expression consists of a transcription step (involving an RNA polymerase that binds to the promoter P_a on the DNA) and a translation step (involving a ribosome that binds to the ribosome binding site RBS_a on the mRNA). The promoter region P_a contains a binding site for A_2, which allows the protein to inhibit gene expression. Both the protein A and mRNA *a* are degraded.

original Gillespie algorithm have been proposed, directed at reducing the computational complexity of the procedure. For instance, Gibson and Bruck [151] have proposed a modification that reduces the number of random numbers to be generated. Whereas this improved algorithm remains exact, in the sense that it yields results consistent with the stochastic master equation, other algorithms address the performance problems by exploiting approximations that lower the accuracy but improve the computational complexity. A popular approximation is the τ-*leap method*, which chooses the time τ between two states such that the algorithm "leaps over" a large number of frequently occurring reactions [152,153]. This speeds up the simulation in that only a single random number needs to be generated for the latter reactions. Another approximation is to explicitly distinguish fast and slow reactions and to use composite reaction mechanisms based on quasi-steady-state approximations for the fast reactions [154] or simulate the latter by means of ordinary or stochastic differential equations [155].

In Figure 7-9 we show the results of applying the original Gillespie algorithm to the autoregulatory feedback network of Figure 7-8, for a maximum of 8000 reaction steps. In comparison with the simulations of the deterministic ordinary differential equation models in Figure 7-5, the number of A_2 molecules fluctuates due to the stochastic nature of the underlying reaction events. The figure illustrates that expression of the gene occurs in bursts [138,142], associated with the binding of RNA polymerase to the promoter, which initiates the transcription of mRNA molecules, in turn translated into proteins. It can be seen in the figure that, as the number of A_2 molecules increases, transcription initiation becomes less frequent due to the occupation of the promoter region by the repressor protein. As a consequence, A_2 reaches a stationary level of about 65 molecules. Not surprisingly, a much higher level is reached in a variant of the above model in which autoregulation has been disabled, for instance due to a mutation in the promoter region that prevents the repressor from binding to the DNA (figure not shown).

A network with the same autoregulatory feedback structure as in Figure 7-5, as well as its mutant variant, has been designed and constructed on a plasmid by Becskei and Serrano [156]. Measurements of the repressor protein concentration in the two networks, by means of a fluorescent reporter, show that the negative feedback loop

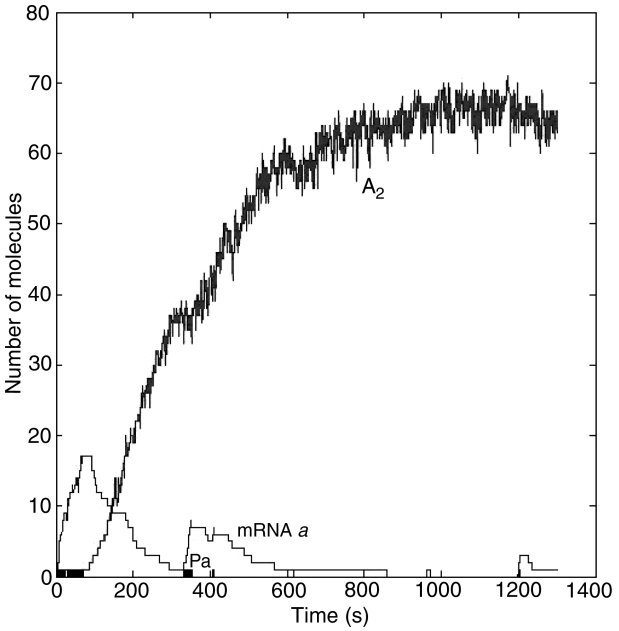


Figure 7-9 Example of a stochastic simulation of the autoinhibition network shown in Fig. 7-8, using a Matlab implementation of the Gillespie algorithm. The figure shows the temporal evolution of the number of molecules of three molecular species (A_2, mRNA *a*, and promoter P_a occupied by RNA polymerase) over 8000 steps. The values of the kinetic constants used in the simulation have been adapted from [141] and [148].

has the effect of decreasing fluctuations in the concentration. This illustrates how topological properties of the regulatory network may reduce the effect of noise arising from the stochasticity of the reaction events [138,140]. Besides mechanisms to *reduce* the effect of noise, the network may also include mechanisms to *amplify* fluctuations so as to increase differences between individual cells in a population. For instance, Isaacs et al. [157] have constructed an autoregulatory feedback network with an activator rather than a repressor protein. The positive feedback loop leads to bistability with states of high and low expression of the gene. Due to stochasticity in gene expression, cells may switch from a high to low expression state, giving rise to a bimodal distribution of protein concentrations in the cell population (see also Refs. [158,159]). The resulting population heterogeneity may have important phenotypic consequences, as discussed in the next section for *E. coli* cells. Pedraza and van Oudenaarden [160] have shown that noise attenuation and amplification can also arise from other mechanisms, for instance gene cascades propagating noise through the network.

In summary, stochastic models of genetic regulatory networks focus on aspects of gene expression that are not taken into account by the deterministic models discussed in Section 7.3. In particular, stochastic models allow the effects of noise on the dynamical behavior of the cell to be studied, by analyzing the way in which fluctuations are filtered out or exploited by means of different molecular mechanisms. A number of computer tools for the stochastic simulation of molecular reaction systems are available, for instance Copasi [111], STOCKS [148], and StochSim [161]. Although stochastic simulation results in closer approximations of the molecular reality than can possibly be obtained by means of the other model types reviewed in this chapter, it is also more difficult to put in practice. Apart from the fact that it requires detailed knowledge of the reactions occurring in the system, notably the value of the kinetic parameters that specify the probability density function $p[\tau, \rho]$ [145], stochastic simulation is a computationally intensive process. In many cases, conventional deterministic models may provide an adequate description of the dynamics of genetic regulatory networks [90].

7.4.2 Effects of Noise in the Carbon Assimilation of *E. coli*

We know experimentally that, even when genetically completely identical, not all individuals of a bacterial population behave in the same way. For example, a long time ago already, it has been observed that in a population of *E. coli* cells the activity of the *lac* operon is not homogeneous [162,163]. That is, under conditions favoring the use of lactose, the *lac* operon is expressed in most but not all cells. An intuitive explanation of this phenomenon relies on the observation that certain kinds of molecules are present at very low numbers in the cell. For example, only about ten copies of the *lac* repressor protein are present in an *E. coli* cell. If these proteins are distributed randomly during cell division, about one cell in a thousand will not contain any *lac* repressor just after cell division. This would lead to derepression of the *lac* operon even in the absence of lactose in the growth medium. What are the consequences of such stochastic phenomena on the behavior of cells and their progeny? Has the cell developed

compensatory mechanisms to cope with the fluctuations, or are they propagated throughout the entire network?

To address these questions, several studies have analyzed the role of stochasticity in the control of carbon assimilation in *E. coli* [148,164,165]. We will focus here on the stochastic model of the growth of bacteria on various carbon sources (glucose, lactose, and glycerol) developed by Puchałka and Kierzek [165]. Using the approach presented in Section 4.1, they have described the PTS and the metabolic pathways involved in the assimilation of these three carbohydrates. In particular, a list of more than 80 molecular species and 120 reactions has been compiled, as well as kinetic parameter values characterizing these reactions.

Stochastic simulation of such a large biochemical reaction system is extremely computationally intensive, given that the interactions in the carbon assimilation module take place on quite different timescales. For instance, the breakdown of carbohydrates and signal transduction by the PTS are fast reactions (less than 1 sec) involving large numbers of molecules, whereas the regulation of gene expression is a slow process (several minutes) involving a very small number of molecules. As discussed in Section 4.1, this prevents the simulation of individual reaction events by means of the basic Gillespie algorithm. Puchałka and Kierzek have therefore used a variant, the *maximal time-step method*, which dynamically partitions the reactions into fast and slow reactions. Whereas the slow reactions are simulated using the Gillespie algorithm, the fast reactions are treated by the τ-leap method (Section 4.1; [165]).

Stochastic simulation of the assimilation of various carbon sources by *E. coli* reproduced expected and well-known phenomena, like the use of glucose as a preferred nutrient. In addition, the simulations showed that stochastic fluctuations in reactions involving a small number of molecules may propagate through the network and influence the time course of other processes in the system, even metabolic pathways processing large numbers of molecules. For instance, during the transition from glucose to a mixture of lactose and glycerol, random delays in the expression of the *lac* operon may favor the use of glycerol (even though lactose is the preferred nutrient). This results in an almost complete shutdown of the glycolytic pathway, fuelled by glucose and lactose but not by glycerol. A striking effect of these time delays is the heterogeneity in the induction of the *lac* operon within the cell population switching from glucose to a mixture of lactose and glycerol (Fig. 7-10). Moreover, this heterogeneity in the use of carbon sources is conserved throughout consecutive cell divisions.

The model of Puchałka and Kierzek is probably the most extensive stochastic model to date, describing an integrated network of gene expression regulation, signal transduction, and metabolism. There exist a few other examples of stochastic models in *E. coli*, notably the model of the lysis-lysogeney decision following λ phage infection [141] and the model of the regulation of the *pap* operon in a pathogenic strain of *E. coli* [166]. However, as noted in Section 4.1, the development of such models requires precise knowledge about the molecular mechanisms underlying the biological processes, and their analysis involves high computational costs. As a consequence, the development of stochastic models has been limited to rather small, well-characterized systems.

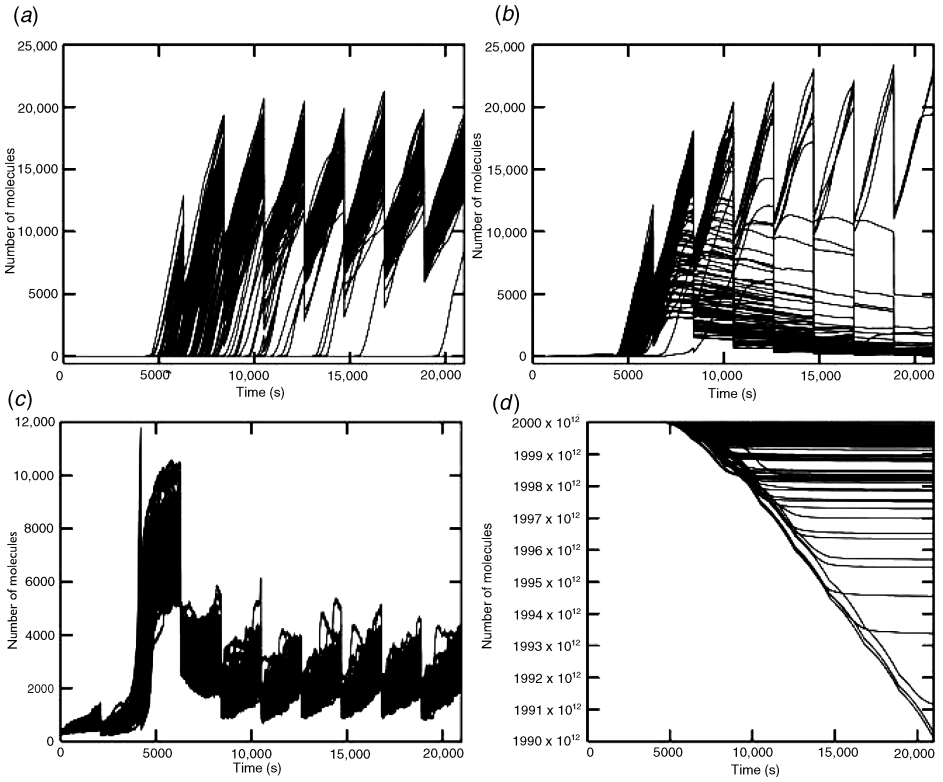


Figure 7-10 Stochastic simulation of *E. coli* growth on a mixture of glucose, lactose, and glycerol, using a model of the carbon assimilation module [164]. Glucose is depleted during the first 5000 s of the simulation. Plots A, B, C, and D show the time-course predictions for LacZ (expressed from the lactose operon), GlpF (expressed from the glycerol operon), cAMP, and external glycerol, respectively. The sawtooth patterns arise from the occurrence of a cell division every 2100 s, during which the molecules present in the cell are distributed over two daughter cells. Plots B and D reveal that there exists a small subpopulation of cells expressing proteins allowing the consumption of glycerol, instead of lactose, after the depletion of glucose.

7.5 DISCUSSION

In this chapter, we have reviewed three different approaches toward the modeling of genetic regulatory networks, based on graphs, ordinary differential equations, and stochastic master equations, respectively. The approaches make different modeling assumptions. Whereas graph models provide a static description of the network, formalizing the structure of interactions between genes, proteins, and other network components, ordinary differential equation and stochastic master equation models describe the dynamic behavior of the system. However, they do so in quite different ways. Differential equations are deterministic models, whereas master equations take into account the stochastic nature of the underlying biochemical reaction processes.

When going from graphs to stochastic master equations, the models take into account increasingly more aspects of physical reality. The counterpart is that this makes them increasingly more difficult to treat in practice. In many cases, even for a well-studied model system like *E. coli*, the information required for building a model on the level of individual reactions is not available. Moreover, the computational burden of simulating such detailed systems is high and does not currently allow its application to large reaction systems, although technical improvements are expected to push the limit further up [167].

In this context, it is crucial to stress that more detailed models are not necessarily *better* models. A model is by its very nature a simplified representation, based on assumptions that ignore certain aspects of reality so as to better bring out others. This is even true for the most detailed models discussed in this chapter, stochastic master equations, which implicitly assume that the reaction volumes are spatially homogeneous, an assumption that is not generally true [168]. In the end, what counts is whether a certain type of model, and thus certain types of simplifying assumptions, are adequate for answering the biological questions at hand. As some authors put it, the art of modeling consists in choosing the ''right model for the job" [169].

The discussion of *E. coli* stress response models in this chapter has confirmed that different kinds of models are appropriate for different kinds of biological questions. The graph models are able to answer questions about the structure of the transcriptional regulatory network of the bacterium, such as the manner in which the network is composed of building blocks like modules and motifs. However, in order to study the dynamics of the building blocks, for instance the carbon assimilation module, one has to resort to ordinary differential equations and stochastic master equations. The former are well adapted for studying the steady states of a regulatory module and the way in which the system may evolve from one steady state to another in response to a perturbation. The latter are especially appropriate for questions about the way the network deals with noise arising from intracellular and extracellular processes, which in the case of noise amplification may give rise to heterogeneous phenotypes in a genetically identical population. This was illustrated by the differential induction of the *lac* operon in *E. coli* cells when glucose in the medium is depleted.[2]

The most effective strategy for studying a complex biological system therefore relies on *model plurality*, using different kinds of models that look at the system from different angles. Instead of building one large supermodel, describing the entire system on the most detailed level possible, it is more fruitful to build a hierarchy of models, accounting for different aspects of the system on different levels of abstraction.

[2] Of course, there is no one-to-one correspondence between biological questions and model formalisms. For instance, the robustness of a dynamic property of the system to fluctuations in the environment can be studied by means of a stochastic model of the biochemical reaction system, but also by varying parameter values in a differential equation model.

ACKNOWLEDGMENTS

The authors acknowledge financial support from the ACI IMPBio initiative of the French Ministry of Research (project BacAttract) and the NEST Adventure and Pathfinder programs of the European Commission (projects Hygeia and Cobios).

REFERENCES

1. Majdalani N, Vanderpool CK, Gottesman S. Bacterial small RNA regulators. *Crit Rev Biochem Mol Biol* 2005;40(2):93–113.
2. Brazhnik P, de la Fuente A, Mendes P. Gene networks: how to put the function in genomics. *Trends Biotechnol* 2002;20(11):467–472.
3. Jacob F. Evolution and thinkering. *Science* 1977;196(4295):1161–1166.
4. Alon U. Biological networks: the tinkerer as engineer. *Science* 2003;301(5641):1866.
5. Csete ME, Doyle JC. Reverse engineering of biological complexity. *Science* 2002;295 (5560):1664–1669.
6. McAdams HH, Shapiro L. Circuit simulation of genetic networks. *Science* 1995;269 (5224):650–656.
7. Lazebnik Y. Can a biologist fix a radio? Or, what I learned while studying apoptosis. *Biochemistry (Mosc)* 2004;69(12):1403–1406.
8. Kitano H. Systems biology: a brief overview. *Science* 2002;295(5560):1662–1664.
9. Endy D. Foundations for engineering biology. *Nature* 2005;438(7067):449–453.
10. Sprinzak D, Elowitz MB. Reconstruction of genetic circuits. *Nature* 2005;438(7067): 443–448.
11. Hengge-Aronis R. Signal transduction and regulatory mechanisms involved in control of the σ^S (RpoS) subunit of RNA polymerase. *Microbiol Mol Biol Rev* 2002;66(3): 373–395.
12. Wick LM, Egli T. Molecular components of physiological stress responses in *Escherichia coli*. *Adv Biochem Eng/Biotechnol* 2004;89:1–45.
13. Neidhardt FC, et al., editors. *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 2nd ed. Washington, DC: ASM Press, 1996.
14. Babloyantz A, Sanglier M. Chemical instabilities of 'all-or-none' type in β-galactosidase induction and active transport. *FEBS Lett* 1972;23(3):364–366.
15. Domach MM, et al. Computer model for glucose-limited growth of a single cell of *Escherichia coli* B/r-A. *Biotechnol Bioeng* 1984;26(3):203–216.
16. Holden C. Alliance launched to model *E. coli*. *Science* 2002;297(5586):1459–1460.
17. Wanner BL, Finney A, Hucks M. Modeling the *E. coli* cell: the need for computing, cooperation, and consortia. *Top Curr Genet* 2005;13:163–189.
18. Bolouri H, Davidson EH. Modeling transcriptional regulatory networks. *BioEssays* 2002;24(12):1118–1129.
19. de Jong H. Modeling and simulation of genetic regulatory systems: a literature review. *J Comput Biol* 2002;9(1):67–103.

20. Hasty J, et al. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nat Rev Genet* 2001;2(4):268–279.
21. McAdams HH, Arkin A. Simulation of prokaryotic genetic circuits. *Annu Rev Biophys Biomol Struct* 1998;27:199–224.
22. Smolen P, Baxter DA, Byrne JH. Modeling transcriptional control in gene networks: methods, recent results, and future directions. *Bull Math Biol* 2000;62(2):247–292.
23. Berge C. *The Theory of Graphs*. Mineola, NY: Dover Publications, 2001.
24. Kohn KW. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol Biol Cell* 1999;10:2703–2734.
25. Davidson EH, et al. A genomic regulatory network for development. *Science* 2002;295 (5560):1669–1678.
26. Barabási AL, Oltvai ZN. Network biology: understanding the cell's functional organization. *Nat Rev Genet* 2004;5(2):101–113.
27. Newman MEJ. The structure and function of complex networks. *SIAM Rev* 2003;45(2): 167–256.
28. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature* 1998;393 (6684):440–442.
29. Albert R, Jeong H, Barabási AL. Error and attack tolerance of complex networks. *Nature* 2000;406(6794):378–382.
30. Jeong H, et al. Lethality and centrality in protein networks. *Nature* 2001;411(6833): 41–42.
31. Jeong H, et al. The large-scale organization of metabolic networks. *Nature* 2000;407 (6804):651–654.
32. Aldana M, Cluzel P. A natural class of robust networks. *Proc Natl Acad Sci USA* 2003;100 (15):8710–8714.
33. Dobrin R, et al. Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network. *BMC Bioinform* 2004;5(1):10.
34. Featherstone DE, Broadie K. Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *BioEssays* 2002;24(3):267–274.
35. Guelzim N, et al. Topological and causal structure of the yeast transcriptional regulatory network. *Nat Genet* 2002;31(1):60–63.
36. Lee TI, et al. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 2002;298(5594):799–804.
37. Maslov S, Sneppen K. Specificity and stability in topology of protein networks. *Science* 2002;296(5569):910–913.
38. Tong AH, et al. Global mapping of the yeast genetic interaction network. *Science* 2004;303(5659):808–813.
39. Wagner A, Fell DA. The small world inside large metabolic networks. *Proc R Soc Lond, B*, 2001;268(1478):1803–1810.
40. Stumpf MPH, Wiuf C, May RM. Subnets of scale-free networks are not scale-free: sampling properties of networks. *Proc Natl Acad Sci USA* 2005;102(12):4221–4224.
41. Yu H, et al. TopNet: a tool for comparing biological sub-networks, correlating protein properties with topological statistics. *Nucleic Acids Res* 2004;32(1):328–337.

42. Arita M. The metabolic world of *Escherichia coli* is not small. *Proc Natl Acad Sci USA* 2004;101(6):1543–1547.

43. Alm E, Arkin AP. Biological networks. *Curr Opin Struct Biol* 2003;13(2):193–202.

44. Ravasz E, et al. Hierarchical organization of modularity in metabolic networks. *Science* 2002;297(5586):1551–1555.

45. Rives AW, Galitski T. Modular organization of cellular networks. *Proc Natl Acad Sci USA* 2003;100(3):1128–1133.

46. Girvan M, Newman M. Community structure in social and biological networks. *Proc Natl Acad Sci USA* 2002;99(12):7821–7826.

47. Milo R, et al. Superfamilies of evolved and designed networks. *Science* 2004;303(5663): 1538–1542.

48. Milo R, et al. Network motifs: simple building blocks of complex networks. *Science* 2002;298(5594):824–827.

49. Shen-Orr SS, et al. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 2002;31(1):64–68.

50. Wuchty S, Oltvai ZN, Barabási A-L. Evolutionary conservation of motif constituents in the yeast protein interaction network. *Nat Genet* 2003;35(2):118–119.

51. Conant GC, Wagner A. Convergent evolution of gene circuits. *Nat Genet* 2003;34(3): 264–266.

52. Bar-Joseph Z, et al. Computational discovery of gene modules and regulatory networks. *Nat Biotechnol* 2003;21:1337–1342.

53. Ihmels J, et al. Revealing modular organization in the yeast transcriptional network. *Nat Genet* 2002;31(4):370–377.

54. Luscombe NM, et al. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 2004;431(7006):308–311.

55. Segal E, et al. Module networks: identifying regulatory modules and their condition-specific regulators from gene expression data. *Nat Genet* 2003;34(2):166–176.

56. Snel B, Bork P, Huynen MA. The identification of functional modules from the genomic association of genes. *Proc Natl Acad Sci USA* 2002;99(9):5890–5895.

57. Tanay A, et al. Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous genomewide data. *Proc Natl Acad Sci USA* 2004;101(9):2981–2986.

58. Mering CV, et al. Genome evolution reveals biochemical networks and functional modules. *Proc Natl Acad Sci USA* 2003;100:15428–15433.

59. Wu H, et al. Prediction of functional modules based on comparative genome analysis and gene ontology application. *Nucleic Acids Res* 2005;33(9):2822–2837.

60. Kashtan N, et al. Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs. *Bioinformatics* 2004;20(11):1746–1758.

61. Salgado H, et al. RegulonDB (version 5.0): *Escherichia coli* K-12 transcriptional regulatory network, operon organization, and growth conditions. *Nucleic Acids Res* 2006;34:D394

62. Keseler IM, et al. EcoCyc: a comprehensive database resource for *Escherichia coli*. *Nucleic Acids Res* 2005;33:D334–D337

63. Ma H-W, et al. An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. *Nucleic Acids Res* 2004;32(22): 6643–6649.

64. Madan Babu M, Teichmann SA. Evolution of transcription factors and the gene regulatory network in *Escherichia coli*. *Nucleic Acids Res* 2003;31(4):1234–1244.

65. Resendis-Antonio O, et al. Modular analysis of the transcriptional regulatory network of *E. coli*. *Trends Genet* 2004;21(1):16–20.

66. Mangan S, Alon U. Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* 2003;100(21):11980–11985.

67. Mangan S, Zaslaver A, Alon U. The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol* 2003;334(2):197–204.

68. Mangan S, et al. The incoherent feed-forward loop accelerates the response time of the *gal* system of *Escherichia coli*. *J Mol Biol* 2006;356(5):1073–1081.

69. Boyce WE, DiPrima RC. *Elementary Differential Equations and Boundary Value Problems*, 5th ed. John Wiley & Sons, 1992.

70. Goodwin BC. Oscillatory behavior in enzymatic control processes. *Adv Enzyme Regul* 1965;425–438.

71. Goodwin BC. *Temporal Organization in Cells*. New York: Academic Press, 1963.

72. Tyson JJ, Othmer HG. The dynamics of feedback control circuits in biochemical pathways. *Prog Theor Biol* 1978;5:1–62.

73. Hirsch MW, Smale S, editors. *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, 1974.

74. Lambert JD. *Numerical Methods for Ordinary Differential Equations*. Chichester, UK: Wiley, 1991.

75. Freeman M. Feedback control of intercellular signalling in development. *Nature* 2000;408(6818):313–319.

76. Perrimon N, McMahon AP. Negative feedback mechanisms and their roles during pattern formation. *Cell* 1999;97(1):13–16.

77. Thieffry D, et al. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. *BioEssays* 1998;20(5):433–440.

78. Thomas R. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. In: Lacolle B et al., editors. *Numerical Methods in the Study of Critical Phenomena*. Berlin: Springer-Verlag, 1981, pp. 180–193.

79. Thomas R, d'Ari R. *Biological Feedback*. Boca Raton, FL: CRC Press, 1990.

80. Cinquin O, Demongeot J. Positive and negative feedback: striking a balance between necessary antagonists. *J Theor Biol* 2002;216(2):229–241.

81. Gouzé J-L. Positive and negative circuits in dynamical systems. *J Biol Syst* 1998;6(1): 11–15.

82. Plahte E, Mestl T, Omholt SW. Feedback loops, stability and multistationarity in dynamical systems. *J Biol Syst* 1995;3(2):409–413.

83. Snoussi EH. Necessary conditions for multistationarity and stable periodicity. *J Biol Syst* 1998;6(1):3–9.

84. Soulé C. Graphic requirements for multistationarity. *ComPlexUs* 2003;1(3):123–133.

85. Alon U, et al. Robustness in bacterial chemotaxis. *Nature* 1999;397(6715):168–171.

86. Barkai N, Leibler S. Robustness in simple biochemical networks. *Nature* 1997;387 (6636):913–917.

87. Barkai N, Shilo B-Z. Modeling pattern formation: counting to two in the Drosophila Egg. *Curr Biol* 2002;12(14):R493

88. von Dassow G, et al. The segment polarity network is a robust developmental module. *Nature* 2000;406(6792):188–192.

89. Morohashi M, et al. Robustness as a measure of plausibility in models of biochemical networks. *J Theor Biol* 2002;216(1):19–30.

90. Gonze D, Halloy J, Goldbeter A. Robustness of circadian rhythms with respect to molecular noise. *Proc Natl Acad Sci USA* 2002;99(2):673–678.

91. Stelling J, et al. Robustness of cellular functions. *Cell* 2004;118(6):675–686.

92. Ma L, Iglesias PA. Quantifying robustness of biochemical network models. *BMC Bioinform* 2002;3:38.

93. Alves R, Savageau MA. Comparing systemic properties of ensembles of biological networks by graphical and statistical methods. *Bioinformatics* 2000;16(6):527–533.

94. Yi T-M, et al. Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc Natl Acad Sci USA* 2000;97(9):4649–4653.

95. Bhalla US, Iyengar R. Emergent properties of networks of biological signaling pathways. *Science* 1999;283(5400):381–387.

96. El-Samad H, et al. Surviving heat shock: control strategies for robustness and performance. *Proc Natl Acad Sci USA* 2005;102(8):2736–2741.

97. Saez-Rodriguez J, et al. Modular analysis of signal transduction networks. *IEEE Control Systems Magazine* 2004;24(4):35–52.

98. Heinrich R, Schuster S. *The Regulation of Cellular Systems*. New York: Chapman & Hall, 1996.

99. Pécou E. Splitting the dynamics of large biochemical interaction networks. *J Theor Biol* 2005;232(3):375–384.

100. Ljung L. *System Identification: Theory for the User*. Upper Saddle River, NJ: Prentice Hall, 1999.

101. Gardner TS, et al. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 2003;301(5629):102–105.

102. Lemeille S, Latifi A, Geiselmann J. Inferring the connectivity of a regulatory network from mRNA quantification in *Synechocystis* PCC6803. *Nucleic Acids Res* 2005;33(10): 3381–3389.

103. Ronen M et al. Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proc Natl Acad Sci USA* 2002;99(16): 10555–10560.

104. van Someren EP, Wessels LFA, Reinders MJT. Linear modeling of genetic networks from experimental data. *Proceedings of the Eight International Conference on Intelligent Systems for Molecular Biology, ISMB 2000*, Menlo Park, CA: AAAI Press, 2000.

105. de Jong H, Ropers D. Qualitative approaches towards the analysis of genetic regulatory networks. In: Stelling J, editor. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. Cambridge, MA: MIT Press, 2006, pp. 125–148.

106. Gagneur J, Casari G. From molecular networks to qualitative cell behavior. *FEBS Lett* 2005;579(8):1867–1871.

107. Glass L, Kauffman SA. The logical analysis of continuous non-linear biochemical control networks. *J Theor Biol* 1973;39(1):103–129.

108. Kauffman SA. Homeostasis and differentiation in random genetic control networks. *Nature* 1969;224:177–178.

109. Thomas R. Regulatory networks seen as asynchronous automata: a logical description. *J Theor Biol* 1991;153:1–23.

110. Thomas R. Boolean formalization of genetic control circuits. *J Theor Biol* 1973;42:563–585.

111. Hoops S, et al. COPASI—a COmplex PAthway SImulator. *Bioinformatics* 2006;22(24): 3067–3077.

112. Ginkel M, et al. Modular modelling of cellular systems with ProMoT/Diva. *Bioinformatics* 2003;19:1169–1176.

113. Slepchenko BM, et al. Quantitative cell biology with the virtual cell. *Trends Cell Biol* 2003;13(11):570–576.

114. Ermentrout B. *Simulating, Analyzing, and Animating Dynamical Systems: A Guide to XPPAUT for Researchers and Students*. Philadelphia: SIAM, 2002.

115. Hucka M, et al. The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 2003;19 (4):524–531.

116. Bettenbrock K, et al. A quantitative approach to catabolite repression in *Escherichia coli*. *J Biol Chem* 2006;281(5):2578–2584.

117. Ropers D, et al. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *BioSystems* 2006;84(2):124–152.

118. Kremling A, et al. The organization of metabolic reaction networks: III. Application for diauxic growth on glucose and lactose. *Metab Eng* 2001;3(4):362–379.

119. Kremling A, et al. Time hierarchies in the *Escherichia coli* carbohydrate uptake and metabolism. *BioSystems* 2003;73(1):57–72.

120. Sauter T, Gilles ED. Modeling and experimental validation of the signal transduction via the *Escherichia coli* sucrose phosphotransferase system. *J Biotechnol* 2004;110(2): 181–199.

121. Wang J, et al. Modeling of inducer exclusion and catabolite repression based on a PTS-dependent sucrose and non-PTS-dependent glycerol transport systems in *Escherichia coli* K-12 and its experimental verification. *J Biotechnol* 2001;92(2):133–158.

122. Chassagnole C, et al. Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol Prog* 2002;79:53–73.

123. Covert MW, Palsson BO. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol* 2003;221(3):309–326.

124. Wong P, Gladney S, Keasling JD. Mathematical model of the lac operon: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose. *Biotechnol Prog* 1997;13:132–143.

125. Hengge-Aronis R. Regulation of gene expression during entry into stationary phase. In: Umbarger HE, editor. *Escherichia coli* and Salmonella: Cellular and Molecular Biology. Washington, DC: ASM Press, 1996, pp. 1497–1512.

126. Martinez-Antonio A, Collado-Vides J. Identifying global regulators in transcriptional regulatory networks in bacteria. *Curr Opin Microbiol* 2003;6(5):482–489.

127. de Jong H, et al. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull Math Biol* 2004;66(2):301–340.

128. Batt G, et al. Validation of qualitative models of genetic regulatory networks by model checking: analysis of the nutritional stress response in *Escherichia coli*. *Bioinformatics* 2005;21(Suppl. 1):i19–i28

129. Laffend L, Schuler ML. Structured model of genetic control via the lac promoter in *E. coli*. *Biotechnol Bioeng* 1994;43:399–410.

130. Ozbudak EM, et al. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 2004;427:737–740.

131. Santillán M, Mackey MC. Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. *Biophys J* 2004;86(3):1282–1292.

132. Yildirim N, Mackey MC. Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data. *Biophys J* 2003;84(5): 2841–2851.

133. Sanglier M, Nicolis G. Sustained oscillations and threshold phemomena in an operon control circuit. *Biophys Chem* 1976;4:113–121.

134. Endy D, et al. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc Natl Aca Sci USA* 2000;97(10):5375–5380.

135. Ranquet C, et al. Control of bacteriophage Mu lysogenic repression. *J Mol Biol* 2005;353 (1):186–195.

136. Van Dien SJ, Keasling JD. A dynamic model of the *Escherichia coli* phosphate-starvation response. *J Theor Biol* 1998;190(1):37–49.

137. Gardner TS, et al. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 2003; 301(5629):102–105.

138. Kaern M, et al. Stochasticity in gene expression: from theories to phenotypes. *Nat Rev Genet* 2005;6(6):451–464.

139. McAdams HH, Arkin A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet* 1999;15(2):65–69.

140. Rao CV, Wolf DM, Arkin AP. Control, exploitation and tolerance of intracellular noise. *Nature* 2002;420(6912):231–237.

141. Arkin A, Ross J, McAdams HA. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics* 1998;149(4): 1633–1648.

142. McAdams HH, Arkin A. Stochastic mechanisms in gene expression. *Proc Natl Acad Sci USA* 1997;94(3):814–819.

143. Gillespie DT, Petzold LR. Numerical simulation for biochemical kinetics. In: Szallasi VPZ, Stelling J, editors. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*, Cambridge, MA: MIT Press, 2006.

144. Paulsson J, Elf J. Stochastic modelling of intracellular kinetics. In: Szallasi VPZ, Stelling J, editors. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. Cambridge, MA: MIT Press, 2006.

145. Gillespie DT. Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 1977;81(25):2340–2361.

146. van Kampen NG. *Stochastic Processes in Physics and Chemistry.* Amsterdam: Elsevier, 1997.

147. Gillespie DT. The chemical Langevin equation. *J Chem Phys* 2000;113(1):297–306.

148. Kierzek AM. STOCKS: STOChastic Kinetic Simulations of biochemical systems with Gillespie algorithm. *Bioinformatics* 2002;18(3):470–481.

149. Morton-Firth CJ, Bray D. Predicting temporal fluctuations in an intracellular signalling pathway. *J Theor Biol* 1998;192(1):117–128.

150. Morton-Firth CJ, Shimizu TS, Bray D. A free-energy-based stochastic simulation of the tar receptor complex. *J Mol Biol* 1999;286:1059–1074.

151. Gibson MA, Bruck J. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J Phys Chem A* 2000;104(9):1876–1889.

152. Gillespie DT. Approximate accelerated stochastic simulation of chemically reacting systems. *J Chem Phys* 2001;115(4):1716–1733.

153. Gillespie DT, Petzold LR. Improved leap-size selection for accelerated stochastic simulation. *J Chem Phys* 2003;119(16):8229–8234.

154. Rao CV, Arkin AP. Stochastic chemical kinetics and the quasi steady-state assumption: application to the Gillespie algorithm. *J Chem Phys* 2003;118(11):4999–5010.

155. Haseltine EL, Rawlings JB. Approximate simulation of coupled fast and slow reactions for stochastic chemical kinetics. *J Chem Phys* 2002;117(15):6959–6969.

156. Becskei A, Serrano L. Engineering stability in gene networks by autoregulation. *Nature* 2000;405(6786):590–591.

157. Isaacs FJ, et al. Prediction and measurement of an autoregulatory genetic module. *Proc Natl Acad Sci USA* 2003;100(13):7714–7719.

158. Becskei A, Séraphin B, Serrano L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J* 2001;20(10):2528–2535.

159. Hasty J, et al. Noise-based switches and amplifiers for gene expression. *Proc Natl Acad Sci USA* 2000;97(5):2075–2080.

160. Pedraza JM, van Oudenaarden A. Noise propagation in gene networks. *Science* 2005;307(5717):1965–1969.

161. Le Novère N, Shimizu TS. STOCHSIM: modelling of stochastic biomolecular processes. *Bioinformatics* 2001;17(6):575–576.

162. Benzer S. Induced synthesis of enzymes in bacteria analyzed at the cellular level. *Acta Biochim Biophys* 1953;11:383–395.

163. Novick A, Weiner M. Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci USA* 1957;43(7):553–566.

164. Carrier TA, Keasling JD. Investigating autocatalytic gene expression systems through mechanistic modeling. *J Theor Biol* 1999;201:25–36.

165. Puchałka J, Kierzek AM. Bridging the gap between stiochastic and deterministic regimes in the kinetic simulations of the biochemical reaction networks. *Biophys J* 2004;86(3):1357–1372.

166. Jarboe LR, Beckwith D, Liao JC. Stochastic modeling of the phase-variable *pap* operon regulation in uropathogenic *Escherichia coli*. *Biotechnol Bioeng* 2004;88(2):189–203.

167. Endy D, Brent R. Modelling cellular behaviour. *Nature* 2000;409(6818):391–396.

168. Kruse K, Elf J. Kinetics in spatially-extended systems. In: Szallasi VPZ, Stelling J, editors. *System modeling in cellular biology: From concepts to nuts and Bolts*, Cambridge, MA: MIT Press, 2006.

169. Falkenhainer B, Forbus KD. Compositional modeling: finding the right model for the job. *Artif Intell* 1991;51(1–3):95–143.

170. Bray D. Molecular networks: the top–down view. *Science* 2003;301(5641):1864–1865.