JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 9, Number 1, 2002
Mary Ann Liebert, Inc.
Pp. 67–103

Modeling and Simulation of Genetic Regulatory Systems: A Literature Review

HIDDE DE JONG

ABSTRACT

In order to understand the functioning of organisms on the molecular level, we need to know which genes are expressed, when and where in the organism, and to which extent. The regulation of gene expression is achieved through genetic regulatory systems structured by networks of interactions between DNA, RNA, proteins, and small molecules. As most genetic regulatory networks of interest involve many components connected through interlocking positive and negative feedback loops, an intuitive understanding of their dynamics is hard to obtain. As a consequence, formal methods and computer tools for the modeling and simulation of genetic regulatory networks will be indispensable. This paper reviews formalisms that have been employed in mathematical biology and bioinformatics to describe genetic regulatory systems, in particular directed graphs, Bayesian networks, Boolean networks and their generalizations, ordinary and partial differential equations, qualitative differential equations, stochastic equations, and rule-based formalisms. In addition, the paper discusses how these formalisms have been used in the simulation of the behavior of actual regulatory systems.

Key words: genetic regulatory networks, mathematical modeling, simulation, computational biology.

1. INTRODUCTION

THE GENOME OF AN ORGANISM PLAYS A CENTRAL ROLE in the control of cellular processes, such as the response of a cell to environmental signals, the differentiation of cells and groups of cells in the unfolding of developmental programs, and the replication of the DNA preceding cell division. Proteins synthesized from genes may function as transcription factors binding to regulatory sites of other genes, as enzymes catalyzing metabolic reactions, or as components of signal transduction pathways. With few exceptions, all cells in an organism contain the same genetic material. This implies that, in order to understand how genes are implicated in the control of intracellular and intercellular processes, the scope should be broadened from sequences of nucleotides coding for proteins to regulatory systems determining which genes are expressed, when and where in the organism, and to which extent.

Gene expression is a complex process regulated at several stages in the synthesis of proteins (Lewin, 1999). Apart from the regulation of DNA transcription, the best-studied form of regulation, the expression

Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de Recherche Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier CEDEX, France.

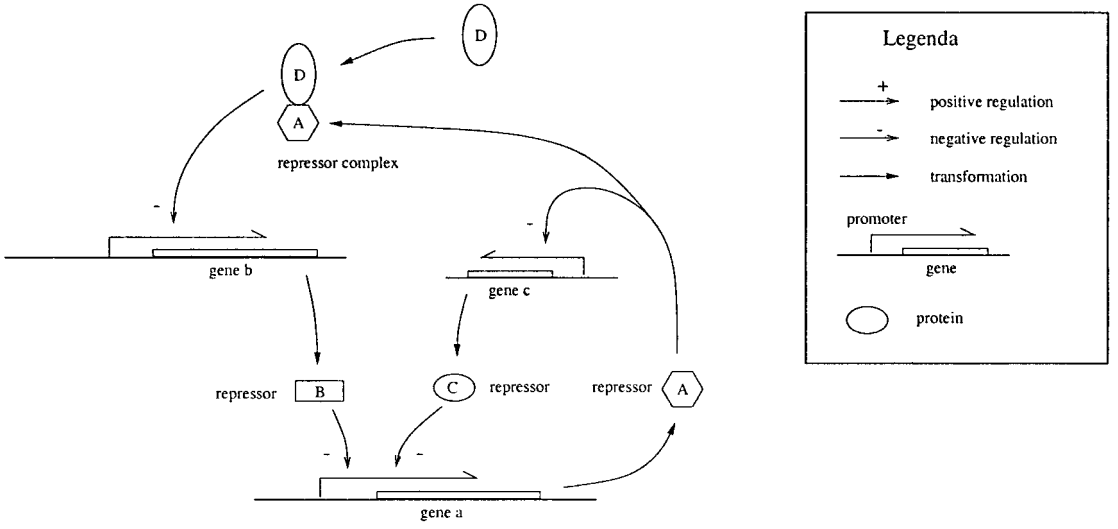


FIG. 1. Example of a genetic regulatory system, consisting of a network of three genes *a*, *b*, and *c*, repressor proteins A, B, C, and D, and their mutual interactions. The figure distinguishes several types of interaction. More complex graphical conventions to represent cellular networks are proposed by Kohn (1999, 2000).

of a gene may be controlled during RNA processing and transport (in eukaryotes), RNA translation, and the posttranslational modification of proteins. The degradation of proteins and intermediate RNA products can also be regulated in the cell. The proteins fulfilling the above regulatory functions are produced by other genes. This gives rise to *genetic regulatory systems* structured by networks of *regulatory interactions* between DNA, RNA, proteins, and small molecules. An example of a simple *regulatory network*, involving three genes that code for proteins inhibiting the expression of other genes, is shown in Fig. 1. Proteins B and C independently repress gene *a* by binding to different regulatory sites of the gene, while A and D interact to form a heterodimer that binds to a regulatory site of gene *b*.[1] Binding of the repressor proteins prevents RNA polymerase from transcribing the genes downstream.

Analyses of the huge amounts of data made available by sequencing projects have contributed to the discovery of a large number of genes and their regulatory sites. The KEGG database, for instance, contains information on the structure and function of about 110,000 genes for 29 species (Kanehisa and Goto, 2000). In some cases, the proteins involved in the control of the expression of these genes, as well as the molecular mechanisms through which regulation is achieved, have been identified. Much less is known, however, about the functioning of the regulatory systems of which the individual genes and interactions form a part (Brownstein *et al.*, 1998; Collado-Vides *et al.*, 1996; Fields *et al.*, 1999; Kanehisa, 2000; Lander, 1996; Loomis and Sternberg, 1995; Nowak, 1995; Palsson, 1997; Strohman, 1997; Thieffry, 1999). Gaining an understanding of the emergence of complex patterns of behavior from the interactions between genes in a regulatory network poses a huge scientific challenge with potentially high industrial pay-offs.

The study of genetic regulatory systems has received a major impetus from the recent development of experimental techniques like cDNA microarrays and oligonucleotide chips, which permit the spatiotemporal expression levels of genes to be rapidly measured in a massively parallel way (Brown and Botstein, 1999; Lipschutz *et al.*, 1999; Lockhart and Winzeler, 2000). Other techniques, such as the mass spectrometric identification of gel-separated proteins, allow the state of a cell to be characterized on the proteomic level as well (Kahn, 1995; Mann, 1999; Pandey and Mann, 2000; Zhu and Snyder, 2001). Although still in their infancy, these techniques have become prominent experimental tools, by opening up a window on the dynamics of gene expression.

In addition to experimental tools, formal methods for the modeling and simulation of gene regulation processes are indispensable. As most genetic regulatory systems of interest involve many genes connected

[1] As a notational convention, names of genes are printed in italic and names of proteins and other molecules start with a capital.

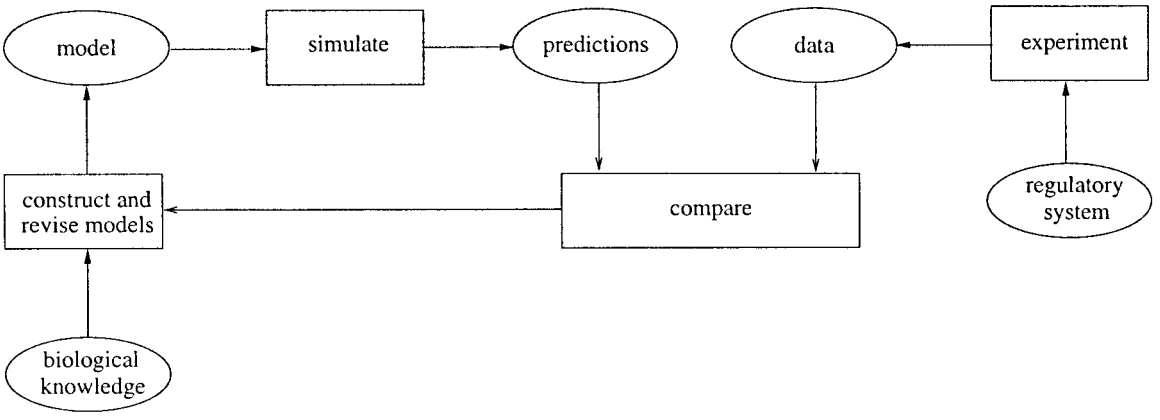


FIG. 2. Analysis of genetic regulatory systems. The boxes represent activities, the ovals information sources, and the arrows information flows.

through interlocking positive and negative feedback loops, an intuitive understanding of their dynamics is hard to obtain. Using formal methods, the structure of regulatory systems can be described unambiguously, while predictions of their behavior can be made in a systematic way. Especially when supported by user-friendly computer tools, modeling and simulation methods permit large and complex genetic regulatory systems to be analyzed.

Figure 2 shows the combined application of experimental and computational tools. Starting from an initial model, suggested by knowledge of regulatory mechanisms and available expression data, the behavior of the system can be simulated for a variety of experimental conditions. Comparing the predictions with the observed gene expression profiles gives an indication of the adequacy of the model. If the predicted and observed behavior do not match, and the experimental data is considered reliable, the model must be revised. The activities of constructing and revising models of the regulatory network, simulating the behavior of the system, and testing the resulting predictions are repeated until an adequate model is obtained.

The formal basis for computer tools supporting the modeling and simulation tasks in Fig. 2 lies in methods developed in mathematical biology and bioinformatics. Since the 1960s, with some notable precursors in the two preceding decades, a variety of mathematical formalisms for describing regulatory networks have been proposed. These formalisms are complemented by *simulation* techniques to make behavioral predictions from a model of the system, as well as *modeling* techniques to construct the model from experimental data and knowledge on regulatory mechanisms. Traditionally, the emphasis has been on simulation techniques, where the models are assumed to have been hand-crafted from the experimental literature. With more experimental data becoming available and easily accessible through databases and knowledge bases, modeling techniques are currently gaining popularity.

This paper gives an overview of formalisms to describe genetic regulatory networks and discusses their use in the modeling and simulation of regulatory systems. While it was being prepared, several other reviews on the modeling and simulation of genetic regulatory systems appeared (Endy and Brent, 2001; Hasty *et al.*, 2001; McAdams and Arkin, 1998; Smolen *et al.*, 2000) [see Glass (1977), Rosen (1968), Thomas (1979) for earlier reviews]. The present report differs from these reviews in that it focuses on the mathematical methods, evaluating their relative strengths and weaknesses, rather than on the biological results obtained through their application. Recently, a collection of introductory chapters covering some of the methods discussed below appeared (Bower and Bolouri, 2001).

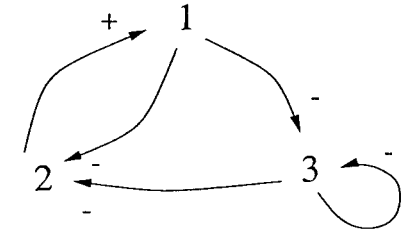
Sections 2 to 11 of this paper give an overview of formalisms proposed in the literature and discuss modeling and simulation techniques appropriate for each of the formalisms. Formalisms to be discussed include directed graphs, Bayesian networks, Boolean networks and their generalizations, ordinary and partial differential equations, qualitative differential equations, stochastic master equations, and rule-based formalisms. It will come as no surprise that the review is not meant to be exhaustive. As hinted at above, the computational study of gene regulation is a subject with a long history. Moreover, in the last few years the number of papers seems to be growing in an exponential fashion. To mention some omissions, no attention is paid to Petri nets (Goss and Peccoud, 1998; Hofestädt and Thelen, 1998; Matsuno *et al.*,

2000; Reddy *et al.*, 1996), transformational grammars (Collado-Vides, 1989, 1996; Collado-Vides *et al.*, 1998), and process algebra (Regev *et al.*, 2001). The impact of these approaches on the dynamic analysis of genetic regulatory networks has been limited thus far. Moreover, they are related to some of the other formalisms discussed below.

2. DIRECTED AND UNDIRECTED GRAPHS

Probably the most straightforward way to model a genetic regulatory network is to view it as a *directed graph.* A directed graph G is defined as a tuple $\langle V, E\rangle$, with V a set of vertices and E a set of edges. A directed edge is a tuple $\langle i, j\rangle$ of vertices, where i denotes the head and j the tail of the edge. The vertices of a directed graph correspond to genes or other elements of interest in the regulatory system, while the edges denote interactions among the genes. The graph representation of a regulatory network can be generalized in several ways. The vertices and edges could be labeled, for instance, to allow information about genes and their interactions to be expressed. By defining a directed edge as a tuple $\langle i, j, s\rangle$, with s equal to $+$ or $-$, it can be indicated whether i is activated or inhibited by j. Hypergraphs can be used to deal with situations in which proteins cooperatively regulate the expression of a gene, for instance by forming a heterodimer (Fig. 1). The edges are then defined by $\langle i, J, S\rangle$, where J represents a list of regulating genes and S a corresponding list of signs indicating their regulatory influence. Figure 3 shows simple regulatory networks of three genes.

Current databases and knowledge bases containing information about regulatory interactions can be viewed, to a large extent, as richly annotated graph representations. Examples of such databases and knowledge bases are aMAZE (van Helden *et al.*, 2000), EcoCyc (Karp *et al.*, 1999a), GeneNet (Kolpakov *et al.*, 1999), GeNet (Samsonova *et al.*, 1998), GIF-DB (Jacq *et al.*, 1997), KEGG (Kanehisa and Goto, 2000), KNIFE (Euzenat *et al.*, 1997), and RegulonDB (Salgado *et al.*, 2000). GeneNet describes cell types and lineages, genes with their regulatory features, proteins and protein complexes, regulatory interactions and other chemical reactions, physiological conditions under which the interactions have been observed, and primary literature sources. The databases and knowledge bases are usually supplemented by applications to compose and edit networks by selecting and manipulating individual interactions. At the very least, they

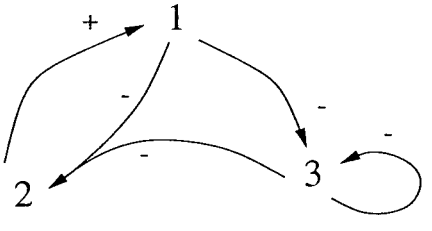


(a)

$V = \{1, 2, 3\}$

$E = \{\langle 2, 1, -\rangle, \langle 3, 1, -\rangle, \langle 1, 2, +\rangle, \langle 2, 3, -\rangle, \langle 3, 3, -\rangle\}$

(b)



(c)

$V = \{1, 2, 3\}$

$E = \{\langle 2, [1, 3], [-, -]\rangle, \langle 3, [1], [-]\rangle, \langle 1, [2], [+]\rangle, \langle 3, [3], [-]\rangle\}$

(d)

FIG. 3. **(a)** Directed graph representing a genetic regulatory network and **(b)** its definition. The plus and minus symbols in the pictorial representation can be omitted and replaced by $\rightarrow$ and $\dashv$ edges, respectively. **(c)–(d)** Directed hypergraph representation of a regulatory network with cooperative interactions.

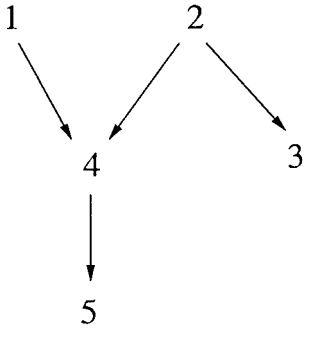
permit the user to visualize complete or partial networks, often on different levels of detail, and to navigate the networks (Samsonova and Serov, 1999; Sanchez *et al.*, 1999). Some current databases and knowledge bases integrate metabolic and signal transduction pathways with gene regulation processes (*e.g.*, Kanehisa and Goto [2000] and Karp *et al.* [1999a]).

A number of operations on graphs can be carried out to make biologically relevant predictions about regulatory systems. A search for paths between two genes, for instance, may reveal missing regulatory interactions or provide clues about redundancy in the network. Furthermore, cycles in the network point at feedback relations that are important for homeostasis and differentiation (Section 6). Global connectivity characteristics of a network, such as the average and the distribution of the number of regulators per gene, give an indication of the complexity of the network. Loosely connected subgraphs point at functional modules of the regulatory system of which the behavior could be considered in isolation. From a comparison of regulatory networks of different organisms, it might be possible to establish to which extent parts of regulatory networks have been evolutionary conserved.

The graph models to which the above operations are applied may be composed from information stored in databases and knowledge bases, but they can also be obtained from gene expression data through *inductive* or *reverse engineering* approaches. A variety of *clustering* algorithms have been used to group together genes with similar temporal expression patterns (Alon *et al.*, 1999; Ben-Dor *et al.*, 1999; Brown *et al.*, 2000; Cho *et al.*, 1998; Eisen *et al.*, 1998; Holter *et al.*, 2001; Michaels *et al.*, 1998; Spellman *et al.*, 1998; Tamayo *et al.*, 1999; Wen *et al.*, 1998). Several techniques for clustering expression data time-series have been proposed in the literature, based on measures like Euclidian distance, mutual information, linear correlation, and rank correlation (D'haeseleer *et al.*, 1998). The use of clustering algorithms is motivated by the idea that two genes exhibiting similar expression patterns over time may regulate each other or be coregulated by a third gene. Additional analyses may permit one to extract more information on putative regulatory connections between coexpressed genes in the graph, such as the analysis of time lags (Arkin and Ross, 1995; Arkin *et al.*, 1997; Chen *et al.*, 1999a), the performance of perturbation experiments (Holstege *et al.*, 1998; Hughes *et al.*, 2000; Laub *et al.*, 2000; Spellman *et al.*, 1998), the mapping of co-expressed genes to pathway functions (Kanehisa and Goto, 2000; Zien *et al.*, 2000), and the screening of upstream regions of coexpressed genes for shared promoters or regulatory sites (Laub *et al.*, 2000; Spellman *et al.*, 1998; Tavazoie *et al.*, 1999).

3. BAYESIAN NETWORKS

In the formalism of *Bayesian networks* (Friedman *et al.*, 2000; Pearl, 1988), the structure of a genetic regulatory system is modeled by a directed acyclic graph $G = \langle V, E \rangle$ (Fig. 4). The vertices $i \in V$, $1 \leq i \leq n$, represent genes or other elements and correspond to random variables X_i. If i is a gene, then X_i will describe the expression level of i. For each X_i, a conditional distribution $p(X_i \mid \text{parents}(X_i))$ is defined, where $\text{parents}(X_i)$ denotes the variables corresponding to the direct regulators of i in G. The



$$p(X_1),\ p(X_2),\ p(X_4 \mid X_1, X_2)$$
$$p(X_5 \mid X_4),\ p(X_3 \mid X_2)$$

$$p(\boldsymbol{X}) = p(X_5 \mid X_4)\, p(X_4 \mid X_1, X_2)\, p(X_3 \mid X_2)\, p(X_2)\, p(X_1)$$

$$i(X_1; X_2, X_3),\ i(X_2; X_1),\ i(X_4; X_3 \mid X_1, X_2)$$
$$i(X_3; X_1, X_4, X_5 \mid X_2),\ i(X_5; X_1, X_2, X_3 \mid X_4)$$

FIG. 4. Example Bayesian network consisting of a graph, conditional probability distributions for the random variables, the joint probability distribution, and conditional independencies (Friedman *et al.*, 2000).

graph G and the conditional distributions $p(X_i \mid \text{parents}(X_i))$, together defining the Bayesian network, uniquely specify a joint probability distribution $p(\boldsymbol{X})$.

Let a *conditional independency* $i(X_i; \boldsymbol{Y} \mid \boldsymbol{Z})$ express the fact that X_i is independent of $\boldsymbol{Y}$ given $\boldsymbol{Z}$, where $\boldsymbol{Y}$ and $\boldsymbol{Z}$ denote sets of variables. The graph encodes the *Markov assumption*, stating that for every gene i in G, $i(X_i; \text{nondescendants}(X_i) \mid \text{parents}(X_i))$. By means of the Markov assumption, the joint probability distribution can be decomposed into

$$p(\boldsymbol{X}) = \prod_{i=1}^{n} p(X_i \mid \text{parents}(X_i)). \tag{1}$$

The graph implies additional conditional independencies, as shown in Fig. 4 for the example network. Two graphs, and hence two Bayesian networks, are said to be equivalent, if they imply the same set of independencies. The graphs in an equivalence class cannot be distinguished by observation on $\boldsymbol{X}$. Equivalent graphs can be formally characterized as having the same underlying undirected graph, but may disagree on the direction of some of the edges (see Friedman *et al.* [2000] for details and references).

Given a set of expression data D in the form of a set of independent values for $\boldsymbol{X}$, learning techniques for Bayesian networks (Heckerman, 1998) allow one to induce the network, or rather the equivalence class of networks, that best matches D. Basically, the techniques rely on a matching score to evaluate the networks with respect to the data and search for the network with the optimal score. As this optimization problem is known to be NP-hard, heuristic search methods have to be used, which are not guaranteed to lead to a globally optimal solution. As an additional problem, currently available expression data underdetermines the network, since at best a few dozen of experiments provide information on the transcription level of thousands of genes.

Friedman and colleagues (Friedman *et al.*, 2000) have proposed a heuristic algorithm for the induction of Bayesian networks from expression data that is able to deal with this so-called dimensionality problem. Instead of looking for a single network, or a single equivalence class of networks, they focus on features that are common to high-scoring networks. In particular, they look at Markov relations and order relations between pairs of variables X_i and X_j. A Markov relation exists, if X_i is part of the minimal set of variables that shields X_j from the rest of the variables, while an order relation exists, if X_i is a parent of X_j in all of the graphs in an equivalence class. An order relation between two variables may point at a causal relationship between the corresponding genes. Statistical criteria to assess the confidence in the features have been developed. A recent extension of the method (Pe'er *et al.*, 2001) is able to deal with genetic mutations and considers additional features, like activation, inhibition, and mediation relations between variables.

Markov relations and order relations have been studied in an application of the algorithm to the cell cycle data set of Spellman and colleagues (1998) (see Pe'er *et al.* [2001] for another application). This data set contains 76 measurements of the mRNA expression level of 6,177 *S. cerevisiae* ORFs included in time-series obtained under different cell cycle synchronization methods. The Bayesian induction algorithm has been applied to the 800 genes whose expression level varied over the cell cycle. By inspecting the high-confidence order relations in the data, Friedman and colleagues found that only a few genes dominate the order, which indicates that they are potential regulators of the cell cycle process. Many of these genes are known to be involved in cell-cycle control and initiation. Of the high-confidence Markov relations, most pairs are functionally related. Some of these relations were not revealed by the cluster analysis of Spellman and colleagues.

A Bayesian network approach towards modeling regulatory networks is attractive because of its solid basis in statistics, which enables it to deal with the stochastic aspects of gene expression and noisy measurements in a natural way. Moreover, Bayesian networks can be used when only incomplete knowledge about the system is available. Although Bayesian networks and the graph models in the previous sections are intuitive representations of genetic regulatory networks, they have the disadvantage of leaving dynamical aspects of gene regulation implicit. To some extent, this can be overcome through generalizations like *dynamical Bayesian networks*, which allow feedback relations between genes to be modeled (Murphy and Mian, 1999). In the next sections, formalisms that explicitly take into account the dynamics of regulatory systems will be discussed.

4. BOOLEAN NETWORKS

As a first approximation, the state of a gene can be described by a Boolean variable expressing that it is active (on, 1) or inactive (off, 0) and hence that its products are present or absent. Moreover, interactions between elements can be represented by Boolean functions which calculate the state of a gene from the activation of other genes. The result is a *Boolean network*, an example of which is shown in Fig. 5. Modeling regulatory networks by means of Boolean networks has become popular in the wake of a groundbreaking study by Kauffman (1969b) (see Sugita [1961, 1963] and Walter *et al.* [1967] for early uses of the Boolean approximation). Recent reviews of the use of Boolean network models can be found in Kauffman's book and in a number of articles (Huang, 1999; Kauffman, 1993; Somogyi and Sniegoski, 1996).

Let the n-vector $\hat{\boldsymbol{x}}$ of variables in a Boolean network represent the state of a regulatory system of n elements. Each $\hat{x}_i$ has the value 1 or 0, so that the state space of the system consists of 2^n states. The state $\hat{x}_i$ of an element at time-point $t+1$ is computed by means of a Boolean function or *rule* $\hat{b}_i$ from the state of k of the n elements at the previous time-point t. (Notice that k may be different for each $\hat{x}_i$.) The variable $\hat{x}_i$ is also referred to as the *output* of the element and the k variables from which it is calculated the *inputs*. For k inputs, the total number of possible Boolean functions $\hat{b}_i$ mapping the inputs to the output is 2^{2^k}. This means that for $k = 2$ there are 16 possible functions, including the nand, or, and nor in Fig. 5. In summary, the dynamics of a Boolean network describing a regulatory system are given by

$$\hat{x}_i(t+1) = \hat{b}_i(\hat{\boldsymbol{x}}(t)), \quad 1 \le i \le n, \tag{2}$$

where $\hat{b}_i$ maps k inputs to an output value.

The structure of a Boolean network can be recast in the form of a wiring diagram (Fig. 5(c)). The upper row lists the state at t and the lower row the state at $t+1$, while the Boolean function calculating the output from the input is shown below each element. The wiring diagram is a convenient representation for computing transitions between states. The transition from one state to the next is usually determined in a parallel fashion, applying the Boolean function of each element to its inputs. For instance, given a state vector 000 at $t = 0$, the system in the example will move to a state 011 at the next time-point $t = 1$. That is, if all three genes are inactive at $t = 0$, the second and the third gene will become active at the next time-point. Hence, transitions between states in a network are *deterministic*, with a single output state for a given input, and *synchronous*, in the sense that the outputs of the elements are updated simultaneously.

A sequence of states connected by transitions forms a *trajectory* of the system. Because the number of states in the state space is finite, the number of states in a trajectory will be finite as well. More specifically, all initial states of a trajectory will eventually reach a steady state or a state cycle, also referred to as

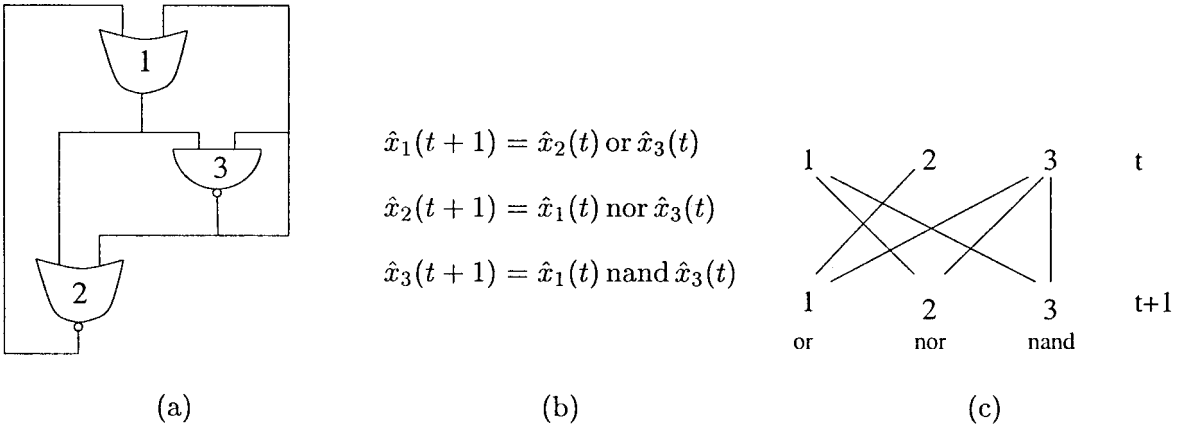


FIG. 5. **(a)** Example Boolean network and **(b)** the corresponding equations. In this case, $n = 3$ and $k = 2$. **(c)** Wiring diagram of the Boolean network.

point attractor or *dynamic attractor*, respectively. The states that are not part of an attractor are called *transient* states. The attractor states and the transient states leading to the attractor together constitute the *basin of attraction.* In the example, both attractors have a basin of attraction consisting of four states.

For simple networks, the attractors and their basins of attraction in the state space can be calculated by hand, but for larger systems computer programs become inevitable. An example of such a program is DDLab, developed by Wuensche (1998). Given sufficient memory, DDLab allows the basins of attraction of a network to be calculated up to $n = 31$. Individual basins of attraction can be determined for much larger networks. Another example is the program Gene-O-Matic (Platzer *et al.*, 2001), which permits differentiation in multicellular systems to be simulated (Jackson *et al.*, 1986).

Boolean networks were among the first formalisms for which model induction methods were proposed, the REVEAL algorithm developed by Liang *et al.* (1998) being an example (see Akutsu *et al.* [1998a, 1998b, 1999], Ideker *et al.* [2000], Karp *et al.* [1999b], Maki *et al.* [2001], Noda *et al.* [1998] for other examples and Thomas [1973] for seminal ideas). In outline, this algorithm uses information theory to establish how given elements are connected in the network and then determines the functions that specify the logic of the interactions from the data. For each of the n elements, all possible combinations of k inputs are considered ($1 \leq k \leq n$), until a set of inputs is found which fully determines the output of the element. In information-theoretic terms, this means that the *mutual information* of the output of the element and its inputs equals the information value of the output of the element alone. The function for the element is then found by comparing the state transitions in the observed trajectory with the Boolean function definitions. An implementation of the algorithm proved to be able to reliably reproduce networks with $n = 50$ and $k = 3$ given 100 state transition pairs (out of the 10^{15} possible pairs).

Examples of Boolean network models of simple regulatory systems can be found in Kauffman (1974). An interesting application of Boolean networks is their use in the study of global properties of large-scale regulatory systems (see Kauffman [1991], Kauffman [1993], Somogyi and Sniegoski [1996], Szallasi and Liang [1998], Weisbuch [1986] for reviews). The basic idea is to generate random Boolean networks with local properties that hold for all members of a class of systems. Properties of interest are, for example, the number k of regulators of a gene and the type of functions $\hat{b}_i$ through which the regulators influence gene expression. By locating the attractors, trajectories, and basins of attraction in the state space, one can systematically investigate the implications of the local properties for the global dynamics of the networks.

This approach has been explored by Kauffman (1969b) (see Kauffman [1969a, 1974, 1977, 1979] for later work, summarized in Kauffman [1991, 1993]).[2] He randomly connected each of the n elements in the network to k inputs and randomly selected the Boolean function $\hat{b}_i$ computing the output of the element from among the 2^{2^k} possible functions. Analysis of networks of up to 10,000 elements showed that, for low k and certain choices of regulatory functions, the systems exhibited highly-ordered dynamics. For example, the expected median number of attractors was empirically found to be about $\sqrt{n}$, the square root of the number of elements. This means that a network of 10,000 elements would be expected to have only 100 state cycles and steady states. Moreover, the length of the attractors was found to be restricted as well, also proportional to $\sqrt{n}$. Interpreting an attractor of the Boolean network as a pattern of gene expression corresponding to a cell type, Kauffman argued that the square root dependence on n of the number of attractors is in accordance with the observation that the number of cell types seems to grow with the square root of the number of genes. The high degree of order observed in large, random Boolean networks has led Kauffman to venture that, once living systems with certain local properties have evolved under the pressure of natural selection, ordered global dynamics necessarily follows (Kauffman, 1993).

Boolean networks allow large regulatory networks to be analyzed in an efficient way, by making strong simplifying assumptions on the structure and dynamics of a genetic regulatory system. In the Boolean network formalism, a gene is considered to be either on or off, and intermediate expression levels are neglected. Also, transitions between the activation states of the genes are assumed to occur synchronously. When transitions do not take place simultaneously, as is usually the case, certain behaviors may not be predicted by the simulation algorithm. There are situations in which the idealizations underlying Boolean networks are not appropriate, and more general methods are required.

[2]Similar ideas have been explored by means of continuous formalisms discussed in later sections (Alves and Savageau, 2000; Bagley and Glass, 1996; Glass and Hill, 1998; Kauffman, 1969a; Newman and Rice, 1971).

5. GENERALIZED LOGICAL NETWORKS

In this section, a generalized logical method developed by Thomas and colleagues will be discussed. The method is based on a formalism that generalizes upon Boolean networks in that it allows variables to have more than two values and transitions between states to occur asynchronously. Since its original conception, the method has undergone several extensions. The version that will be presented here in outline is more extensively described by Thomas (1991), Thomas and d'Ari (1990), Thomas *et al.* (1995) (see Thomas [1973, 1979] for earlier formulations).

The formalism of Thomas and colleagues uses discrete variables $\hat{x}_i$, so-called *logical variables*, that are abstractions of real concentrations variables x_i. The possible values of $\hat{x}_i$ are defined by comparing the concentrations x_i with the thresholds of the influence of i on other elements of the regulatory system. If an element i influences p other elements of the regulatory system, it may have as many as p distinct thresholds (van Ham, 1979):

$$\sigma_i^{(1)} < \sigma_i^{(2)} < \ldots < \sigma_i^{(p)}.$$

Given these thresholds, $\hat{x}_i$ has the possible values $\{0, \ldots, p\}$ and is defined as follows:

$$\hat{x}_i = 0, \text{ if } x_i < \sigma_i^{(1)}$$

$$\hat{x}_i = 1, \text{ if } \sigma_i^{(1)} < x_i < \sigma_i^{(2)}$$

$$\ldots$$

$$\hat{x}_i = p, \text{ if } x_i > \sigma_i^{(p)}.$$

The vector $\hat{\boldsymbol{x}}$ denotes the *logical state* of the regulatory system.

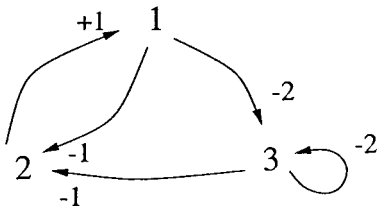
The pattern of regulatory interactions in the system is described by logical equations of the form

$$\hat{X}_i(t) = \hat{b}_i(\hat{\boldsymbol{x}}(t)), \quad 1 \leq i \leq n, \tag{3}$$

where $\hat{X}_i$ is called the *image* of $\hat{x}_i$. The image is the value towards which $\hat{x}_i$ tends when the logical state of the system is $\hat{\boldsymbol{x}}$. The image of $\hat{x}_i$ is not necessarily its successor value, and should therefore be distinguished from $\hat{x}(t+1)$ in (2). The logical function $\hat{b}_i$ is a generalization of the Boolean function in (2), since the logical variables now have more than two possible values. The logical function computes the image of $\hat{x}_i$ from the logical state of the system, more specifically from the value of k of its n elements.

In Fig. 6, an example regulatory network is shown. Gene 1 regulates genes 2 and 3, so that it has two thresholds and the corresponding logical variable $\hat{x}_1$ takes its value from $\{0, 1, 2\}$. Similarly, $\hat{x}_2$ and $\hat{x}_3$ have one and two thresholds, respectively, and hence possible values $\{0, 1\}$ and $\{0, 1, 2\}$. Each edge in the graph has been labeled with the rank number of the threshold as well as the sign of the regulatory influence. For instance, the label -2 from gene 1 to gene 3 means that 1 inhibits 3 above its second threshold, that is, when $\hat{x}_1 = 2$.

The graph in Fig. 6(a) can be transformed into logical equations, as shown in Fig. 6(b). The functions $\hat{b}_1, \ldots, \hat{b}_3$ need to be specified such as to be consistent with the threshold restrictions in the graph. Examples of logical functions allowed by the method are shown in Fig. 6(c), in a notation that emphasizes the correspondence with the Boolean functions in the previous section (see Thomas and d'Ari [1990], Thomas *et al.* [1995] for a full account of the specification of logical functions). Consider the case of $\hat{b}_2$. If $\hat{x}_1 > 0$ and $\hat{x}_3 > 0$, so that x_1 and x_3 have values above their first threshold, the inhibitory influences of genes 1 and 3 on gene 2 become operative. Figure 6(c) indicates that $\hat{x}_2$ will tend to 0, that is, below the first threshold of the protein produced by gene 2. If either $\hat{x}_1 = 0$ or $\hat{x}_3 = 0$, that is, if only one of the inhibitory influences is operative, then gene 2 is moderately expressed. This is here represented by the value 1 for the image of $\hat{x}_2$. In general, several logical functions will be consistent with the threshold restrictions. Exactly which logical function is chosen may be motivated by biological considerations or may be a guess reflecting uncertainty about the structure of the system being studied.



(a)

$$\hat{X}_1 = \hat{b}_1(\hat{x}_2)$$
$$\hat{X}_2 = \hat{b}_2(\hat{x}_1, \hat{x}_3)$$
$$\hat{X}_3 = \hat{b}_3(\hat{x}_1, \hat{x}_3)$$

(b)

$\hat{b}_1(\hat{x}_2)$	$\hat{x}_2$
0	0
2	1

$\hat{x}_1$	$\hat{b}_2(\hat{x}_1, \hat{x}_3)$	$\hat{x}_3$ = 0	1	2
	0	1	1	1
	1	1	0	0
	2	1	0	0

$\hat{x}_1$	$\hat{b}_3(\hat{x}_1, \hat{x}_3)$	$\hat{x}_3$ = 0	1	2
	0	2	2	1
	1	2	2	1
	2	0	0	0

(c)

FIG. 6. **(a)** Example regulatory network in graph notation, where the edges are labeled to express the rank number of the threshold as well as the sign of the influence (Thomas *et al.*, 1995). **(b)** Logical equations corresponding to the graph and **(c)** possible definitions of the logical functions.

The logical equations form the input for the analysis of the regulatory system, in particular the determination of its *logical steady states.* A logical steady state occurs when the logical state of the system equals its image, i.e.,

$$\hat{\boldsymbol{X}} = \hat{\boldsymbol{x}}. \tag{4}$$

Since the number of logical states is finite, due to the discretization achieved by the introduction of logical variables, one can exhaustively test for logical steady states. In the case of the logical equations in Fig. 6, the logical state $[2\ 0\ 1]'$ is the only steady state among the 18 possible logical states. The other states are *transient logical states.*

If the regulatory system is in a transient logical state, it will make a transition to another logical state (Thomas, 1991; Thomas and d'Ari, 1990). Since a logical value will move into the direction of its image, the possible successor states of a logical state can be deduced by comparing the value of a logical variable with that of its image. With the assumption that two logical variables will not change their value simultaneously, a maximum of n successor states can be reached from a current state. Of course, if a logical state is steady, it has no successor states, since the logical variables equal their image. The logical states and the transitions among them can be organized in a *state transition graph.* In more advanced analyses of state transitions, time delays occasioned by transcription, translation, and transport can be taken into account (Thomas, 1983; Thomas and d'Ari, 1990).

The generalized logical method outlined above has been generalized in several ways, notably by the introduction of logical parameters and threshold values for the logical variables. *Logical parameters* (Snoussi, 1989; Thomas and d'Ari, 1990) allow classes of logical functions, as compared with individual logical functions, to be characterized. This facilitates the identification of logical steady states. By introducing logical values that correspond to threshold values in a continuous description (Snoussi and Thomas, 1993; Thomas and d'Ari, 1990), it becomes possible to detect additional logical steady states. The resulting method integrates elements from the differential equation formalisms discussed in the next sections (Snoussi, 1989; Thomas and d'Ari, 1990).

The logical method of Thomas has been implemented (Thieffry *et al.*, 1993) and its effectiveness demonstrated in the study of a number of genetic regulatory systems of limited scale, like λ phage infection in *E. coli* (Thieffry and Thomas, 1995; Thomas and d'Ari, 1990; Thomas *et al.*, 1976) and dorso-ventral pattern formation and gap gene control in *Drosophila* (Sánchez and Thieffry, 2001; Sánchez *et al.*, 1997). Another example is a study of the regulatory network controlling flower morphogenesis in *Arabidopsis thaliana* (Mendoza *et al.*, 1999) (see also Mendoza and Alvarez-Buylla [1998]). In this case, a regulatory

network involving ten genes has been derived from published genetic and molecular data and reduced to two subnetworks of two and four genes, respectively. After choosing appropriate logical functions, the system can be shown to have six logical steady states. Four steady states correspond to the patterns of gene expression found in the floral organs of the plant (sepals, petals, stamens, carpels), whereas a fifth accounts for a nonfloral state. The sixth steady state has not been characterized experimentally thus far. Consideration of the state transition graph, and constraints on the logical parameters, leads to the prediction that the gene *LFY* must have at least one more regulator to account for the transition from the nonfloral steady state to one of the four flowering states.

6. NONLINEAR ORDINARY DIFFERENTIAL EQUATIONS

Being arguably the most widespread formalism to model dynamical systems in science and engineering, *ordinary differential equations (ODEs)* have been widely used to analyze genetic regulatory systems. The ODE formalism models the concentrations of RNAs, proteins, and other molecules by time-dependent variables with values contained in the set of nonnegative real numbers. Regulatory interactions take the form of functional and differential relations between the concentration variables.

More specifically, gene regulation is modeled by *rate equations* expressing the rate of production of a component of the system as a function of the concentrations of other components. Rate equations have the mathematical form

$$\frac{dx_i}{dt} = f_i(\boldsymbol{x}),\ 1 \le i \le n, \tag{5}$$

where $\boldsymbol{x} = [x_1, \ldots, x_n]' \ge \boldsymbol{0}$ is the vector of concentrations of proteins, mRNAs, or small molecules, and $f_i : \mathbb{R}^n \to \mathbb{R}$ a usually nonlinear function. The rate of synthesis of i is seen to be dependent upon the concentrations $\boldsymbol{x}$, possibly including x_i, and f_i can be extended to include concentrations $\boldsymbol{u} \ge \boldsymbol{0}$ of input components, e.g., externally-supplied nutrients. Discrete time delays, arising from the time required to complete transcription, translation, and diffusion to the place of action of a protein, can also be represented:

$$\frac{dx_i}{dt} = f_i(x_1(t - \tau_{i1}), \ldots, x_n(t - \tau_{in})),\ 1 \le i \le n, \tag{6}$$

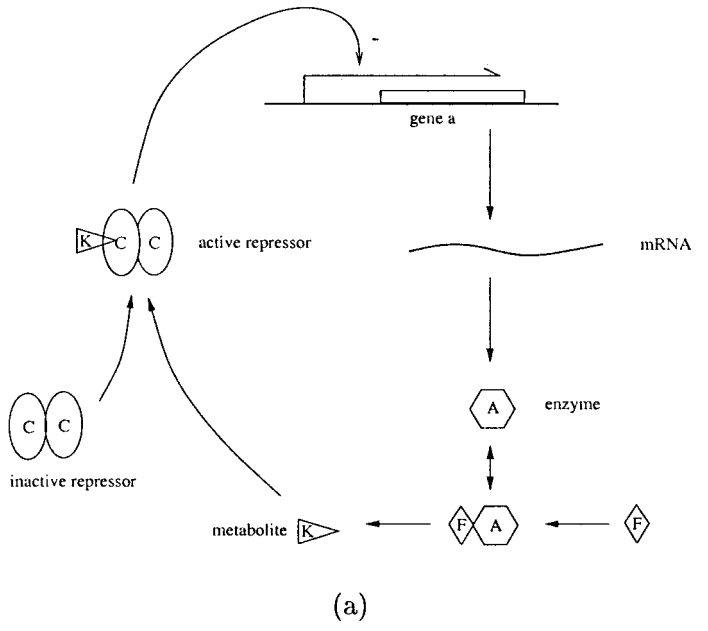
where $\tau_{i1}, \ldots, \tau_{in} > 0$ denote discrete time delays. Alternatively, integrals can be used to model distributed time delays (Landahl, 1969; Mahaffy, 1984; MacDonald, 1989; Smolen *et al.*, 2000).

Powerful mathematical methods for modeling biochemical reaction systems by means of rate equations have been developed in the past century, especially in the context of metabolic processes (see Cornish-Bowden [1995], Heinrich and Schuster [1996], and Voit [2000] for introductions). Using these methods, kinetic models of genetic regulation processes can be constructed by specifying the functions f_i.

In Fig. 7, a kinetic model of a simple genetic regulation process is shown, going back to early work by Goodwin (1963, 1965). The end-product of a metabolic pathway co-inhibits 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. More generally, equations of the form

$$\begin{aligned} \frac{dx_1}{dt} &= \kappa_{1n} r(x_n) - \gamma_1 x_1, \\ \frac{dx_i}{dt} &= \kappa_{i,i-1} x_{i-1} - \gamma_i x_i,\ 1 < i \le n, \end{aligned} \tag{7}$$

are employed (Tyson and Othmer, 1978). The parameters $\kappa_{1n}, \kappa_{21}, \ldots, \kappa_{n,n-1} > 0$ are production constants and $\gamma_1, \ldots, \gamma_n > 0$ degradation constants. The rate equations express a balance between the number of molecules appearing and disappearing per unit time. In the case of x_1, the production term involves a nonlinear *regulation function* $r : \mathbb{R} \to \mathbb{R}$, whereas the concentrations x_i, $1 < i \le n$, increase linearly with x_{i-1}. In order to express that the metabolic product is a co-repressor of the gene, r needs to be a decreasing



$$\dot{x}_1 = \kappa_1 r(x_3) - \gamma_1 x_1$$
$$\dot{x}_2 = \kappa_2 x_1 - \gamma_2 x_2$$
$$\dot{x}_3 = \kappa_3 x_2 - \gamma_3 x_3$$

(b)

FIG. 7. **(a)** Example of a genetic regulatory system involving end-product inhibition and **(b)** its ODE model (adapted from Goodwin [1963, 1965]). 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; $\kappa_1, \kappa_2, \kappa_3$ are production constants, $\gamma_1, \gamma_2, \gamma_3$ degradation constants, and $r : \mathbb{R} \to \mathbb{R}$ a decreasing, nonlinear regulation function ranging from 0 to 1.

function, i.e., $\partial r / \partial x_n < 0$. The terms $-\gamma_i x_i$, $1 \leq i \leq n$, state that the concentrations x_i decrease through degradation, diffusion, and growth dilution at a rate proportional to the concentrations themselves.

A regulation function often found in the literature is the so-called *Hill curve* (Fig. 8(a)):

$$h^+(x_j, \theta_j, m) = \frac{x_j^m}{x_j^m + \theta_j^m}, \tag{8}$$

with $\theta_j > 0$ the *threshold* for the regulatory influence of x_j on a target gene, and $m > 0$ a steepness parameter. The function ranges from 0 to 1 and increases as $x_j \to \infty$, so that an increase in x_j will tend to increase the expression rate of the gene (*activation*). In order to express that an increase in x_j decreases the expression rate (*inhibition*), as in (7), the regulation function $h^+(x_j, \theta_j, m)$ is replaced by $h^-(x_j, \theta_j, m) = 1 - h^+(x_j, \theta_j, m)$. For $m > 1$, Hill curves have a sigmoid shape, in agreement with experimental evidence (Yagil, 1975; Yagil and Yagil, 1971). Figure 8 shows two further examples of regulation functions that will be discussed in more detail in a later section.

Due to the nonlinearity of f_i, analytical solution of the rate equations (5) is not normally possible. In special cases, qualitative properties of the solutions, such as the number and the stability of steady

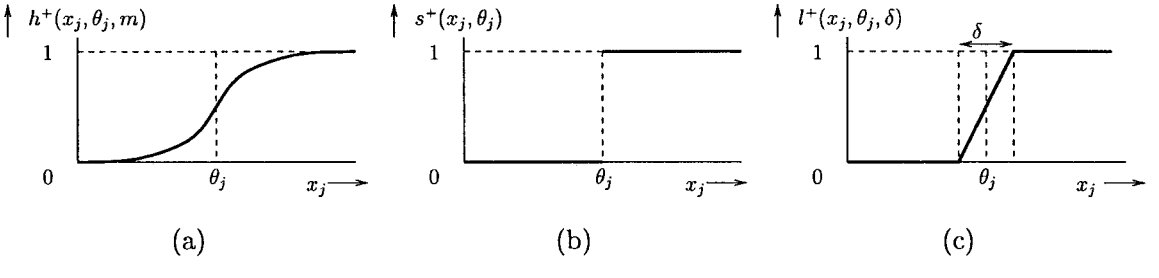


FIG. 8. Examples of regulation functions: **(a)** Hill function h^+, **(b)** Heaviside or step function s^+, and **(c)** logoid function l^+.

states and the occurrence of limit cycles, can be established. This is illustrated by studies that investigate the relationship between (single) feedback loops and the dynamics of regulatory systems (e.g., Cherry and Adler [2000], Goodwin [1965], Griffith [1968a, 1968b], Keller [1994], Othmer [1976], Tyson [1975], Walter [1970], Wolf and Eeckman [1998], reviewed in Smolen *et al.* [2000], Tyson and Othmer [1978]). In the case of a *negative feedback* loop, as in Fig. 7, the system may approach or oscillate around a single steady state. In the presence of a *positive feedback* loop, on the contrary, the system tends to settle in one of two stable states, depending on whether the initial state is on one or the other side of a separatrix.

In fact, a negative feedback loop is a necessary condition for stable periodicity and a positive feedback loop for multistationarity (Gouzé, 1998; Plahte *et al.*, 1995a; Snoussi, 1998; Thomas, 1981; Thomas and Kaufman, 2001). Following seminal ideas in Delbrück (1949) and Monod and Jacob (1961), Thomas has drawn attention to the relation between the feedback structure of a system and the biological phenomena of homeostasis and differentiation (Thieffry *et al.*, 1995; Thomas, 1981, 1998; Thomas and d'Ari, 1990). In particular, the stable periodicity of negative feedback loops can be interpreted as representing homeostasis, whereas the multistationarity occasioned by a positive feedback loop provides a suggestive parallel to differentiation processes observed in development.

One way to work around the refractoriness of nonlinear rate equations to mathematical analysis is to simplify the models, an approach that will be discussed in Section 7. Alternatively, one can take recourse to numerical techniques. In *numerical simulation*, the exact solution of the equations is approximated by calculating approximate values $\boldsymbol{x}_0, \ldots, \boldsymbol{x}_m$ for $\boldsymbol{x}$ at consecutive time-points $t_0, \ldots, t_m$ (Lambert, 1991). A variety of computer tools specifically adapted to the simulation and analysis of biochemical reaction systems are available, such as DBsolve (Goryanin *et al.*, 1999), GEPASI (Mendes, 1993), MIST (Ehlde and Zacchi, 1995), and SCAMP (Sauro, 1993).

Numerical simulations of a few well-studied regulatory systems have been carried out. For instance, Reinitz and Vaisnys (1990) discuss a numerical model of the *cro-cI* switch controlling the growth of phage λ in *E. coli*. They study the switch when certain genes have been mutated, causing it to be functionally isolated from the rest of the phage genome. Another simulation of the phage λ system is described by MacAdams and Shapiro (1995) (see also Ackers *et al.* [1997] and Shea and Ackers [1985] and the review in McAdams and Arkin [1998]). An interesting feature of this work is the parallel drawn between genetic regulatory networks and electrical circuits, resulting in a hybrid approach that integrates biochemical kinetic modeling within the framework of circuit simulation. Other well-known examples of genetic regulation processes for which numerical models have been developed include the induction of the *lac* operon in *E. coli* (Bliss *et al.*, 1982; Carrier and Keasling, 1999; Mahaffy, 1984; Sanglier and Nicolis, 1976; Wong *et al.*, 1997), the developmental cycle of bacteriophage T7 (Endy *et al.*, 2000; You and Yin, 2001), the synthesis of *trp* in *E. coli* (Koh *et al.*, 1998; Prokudina *et al.*, 1991; Santillán and Mackey, 2001; Xiu *et al.*, 1997), the expression of a human immunodeficiency virus (HIV) (Hammond, 1993), and circadian rhythms in *Drosophila* and other organisms (Goldbeter, 1995; Leloup and Goldbeter, 1998; Ruoff *et al.*, 2001; Ueda *et al.*, 2001).

Simulation of the functioning of a regulatory network is often complemented by *bifurcation analysis* tools to investigate the sensitivity of steady states and limit cycles to parameter values (Strogatz, 1994). Borisuk and Tyson (1998) have applied these techniques to a numerical model of a much-studied example of posttranslational modification, the control of mitosis in the *Xenopus* oocyte (Novak and Tyson, 1993b) (see Novak and Tyson [1993a] for a variant of this model). The model consists of ten ODEs describing the network of biochemical reactions regulating M-phase promoting factor (MPF), a protein triggering a number of key events of mitosis. By introducing additional assumptions about the biochemical properties of the system, the M-phase control model can be simplified to just two ODEs, for the concentrations of active MPF and its regulatory subunit Cyclin, without losing essential qualitative features of the solutions. A rich array of distinct behaviors were found when varying parameter values, corresponding to well-known physiological states of the cell as well as states that had never been recognized experimentally. Recently, similar models for yeast have been developed by the same group (Chen *et al.*, 2000; Novak *et al.*, 1998; Novak and Tyson, 1995) (see Goldbeter [1997], Obeyesekere *et al.* [1992], Tyson [1999], and Tyson *et al.* [1996] for reviews of cell cycle models).

A problem hampering the use of numerical techniques is the lack of *in vivo* or *in vitro* measurements of the kinetic parameters in the rate equations. Numerical parameter values are available for only a handful of well-studied systems, the λ phage switch being a rare example. In contrast, in the cell cycle models

mentioned above, in most cases the parameter values had to chosen such that the models are able to reproduce the observed qualitative behavior. For larger models, finding appropriate values may be difficult to achieve.

The growing availability of gene expression data could alleviate the problem to some extent. From measurements of the state variables $\boldsymbol{x}$ at several stages of a process of interest, possibly under a variety of experimental conditions, parameter values might be estimated with the help of system identification techniques (Ljung, 1999). In a later section, this approach will be examined in more detail. Some recent studies suggest another response to the absence of measured values of the kinetic parameters. For instance, in a simulation study of the segment polarity network in *Drosophila*, it was found that essential properties of the system were quite robust to variations in parameter values and initial conditions over sometimes several orders of magnitude (von Dassow *et al.*, 2000). A similar conclusion had been drawn earlier from analysis of a model of bacterial chemotaxis (Barkai and Leibler, 1997) (see Savageau [1971] for pioneering work). These robustness results suggest that it is the network structure rather than the precise value of the parameters that confers stability to the system.

7. PIECEWISE-LINEAR DIFFERENTIAL EQUATIONS

In this section, a special case of the rate equations (5) will be considered that is based on two simplifying assumptions. First, the models abstract from the biochemical details of regulatory interactions by directly relating the expression levels of genes in the network. Second, the switch-like behavior of genes whose expression is regulated by continuous sigmoid curves is approximated by discontinuous step functions (Fig. 8). The resulting differential equations are piecewise linear, and have favorable mathematical properties facilitating their qualitative analysis. *Piecewise-linear differential equation (PLDE)* models are related to the logical models in Sections 4 and 5, notwithstanding subtle differences caused by the discrete rather than continuous nature of the latter formalisms.

We will be concerned with piecewise-linear differential equations of the form

$$\frac{dx_i}{dt} = g_i(\boldsymbol{x}) - \gamma_i x_i, \quad 1 \leq i \leq n, \tag{9}$$

where x_i denotes the cellular concentration of the product of gene i and $\gamma_i > 0$ the degradation rate of x_i (Glass, 1975a; Mestl *et al.*, 1995a; Thomas and d'Ari, 1990). Regulation of degradation could be modeled by replacing γ_i with a function similar to g_i, but this will be omitted here.

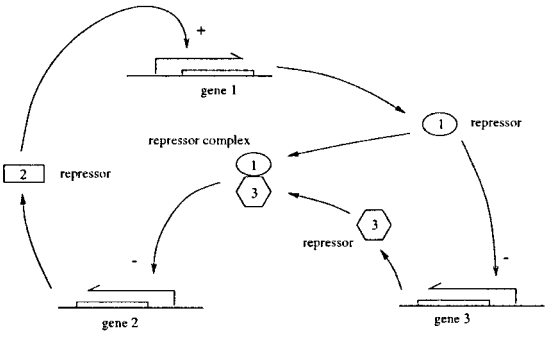
In its most general form, the function $g_i : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ is defined as

$$g_i(\boldsymbol{x}) = \sum_{l \in L} \kappa_{il}\, b_{il}(\boldsymbol{x}) \geq 0, \tag{10}$$

where $\kappa_{il} > 0$ is a rate parameter, $b_{il} : \mathbb{R}^n_{\geq 0} \to \{0, 1\}$ a function defined in terms of sums and multiplications of step functions, and L a possibly empty set of indices. The function b_{il} should be seen as the arithmetic equivalent of a Boolean function, expressing the conditions under which the gene is expressed at a rate κ_{il} (Glass, 1977; Plahte *et al.*, 1998; Snoussi, 1989). The conditions are specified by means of step functions s^+ and s^-, defined by

$$s^+(x_j, \theta_j) = \begin{cases} 1, & x_j > \theta_j \\ 0, & x_j < \theta_j \end{cases}, \quad \text{and} \quad s^-(x_j, \theta_j) = 1 - s^+(x_j, \theta_j). \tag{11}$$

In the example of Fig. 9, we see that for gene 1, $g_1(x_2) = \kappa_1\, s^+(x_2, \theta_{21})$, meaning that protein 1 is synthesized at a rate κ_1, if the concentration of protein 2 is below the threshold concentration θ_{21}. Gene 2 is repressed by a heterodimer composed of proteins 1 and 3, so that it is expressed at a rate κ_2, if the concentration of either protein remains below its respective threshold ($g_2(x_1, x_3) = \kappa_2\,(1 - s^+(x_1, \theta_{11})\, s^+(x_3, \theta_{31}))$). More complex combinations of step functions can be used to model the combined effects of several regulatory proteins.



(a)

$$\dot{x}_1 = \kappa_1\, s^+(x_2, \theta_{21}) - \gamma_1\, x_1$$
$$\dot{x}_2 = \kappa_2\, (1 - s^+(x_1, \theta_{11})\, s^+(x_3, \theta_{31})) - \gamma_2\, x_2$$
$$\dot{x}_3 = \kappa_3\, s^-(x_1, \theta_{12}) + \kappa_4\, s^-(x_3, \theta_{32}) - \gamma_3\, x_3$$

(b)

FIG. 9. **(a)** Example regulatory network of three genes and **(b)** corresponding piecewise-linear differential equations; x_1, x_2, and x_3 represent protein or mRNA concentrations, respectively, $\kappa_1, \ldots, \kappa_4$ production constants, $\gamma_1, \ldots, \gamma_3$ degradation constants, and $\theta_{11}, \theta_{12}, \theta_{21}, \theta_{31}, \theta_{32}$ threshold constants.

Equations (9) have been well-studied in mathematical biology and elsewhere (Edwards, 2000; Edwards and Glass, 2000; Edwards *et al.*, 2001; Glass, 1975a, 1975b, 1977; Glass and Hill, 1998; Glass and Pasternack, 1978a, 1978b; Lewis and Glass, 1991; Mestl *et al.*, 1995a, 1995b; Plahte *et al.*, 1994; Snoussi, 1989; Snoussi and Thomas, 1993; Thomas and d'Ari, 1990). Consider an n-dimensional (hyper)box of the phase space defined as follows:

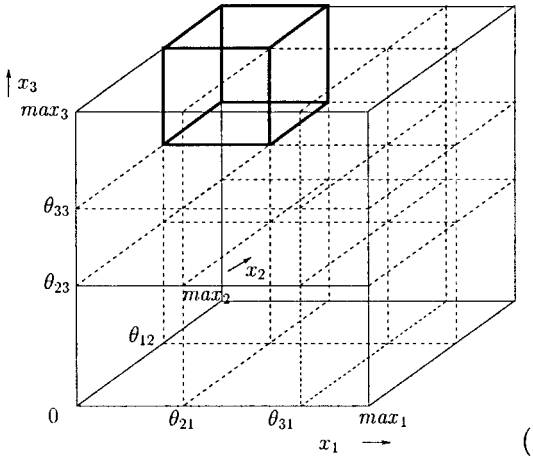
$$0 \le x_i \le max_i > \max_{\boldsymbol{x} \ge \mathbf{0}} g_i(\boldsymbol{x})/\gamma_i, \quad 1 \le i \le n, \tag{12}$$

where we assume that for all threshold concentrations θ_{ik} of the protein encoded by gene i it holds that $\theta_{ik} < max_i$. The $n-1$-dimensional threshold (hyper)planes $x_i = \theta_{ik}$ divide the n-box into orthants. Figure 10(a) displays the phase space box and orthants corresponding to the example system in Fig. 9.

In each orthant of the n-box, by evaluation of the step functions, (9) reduces to ODEs with a constant production term μ_i composed of rate parameters in b_i:

$$\dot{x}_i = \mu_i - \gamma_i x_i, \quad 1 \le i \le n. \tag{13}$$

Figure 10(b) gives an example of the state equations corresponding to the orthant $0 \le x_1 < \theta_{21}$, $\theta_{12} < x_2 \le max_2$, and $\theta_{33} < x_3 \le max_3$.



(a)

$$\dot{x}_1 = \kappa_{12} - \gamma_1 x_1$$
$$\dot{x}_2 = -\gamma_2 x_2$$
$$\dot{x}_3 = \kappa_{31} - \gamma_3 x_3$$

(b)

FIG. 10. **(a)** The phase space box of the model in Fig. 9, divided into $2 \cdot 3 \cdot 3 = 18$ orthants by the threshold planes. **(b)** The state equations for the orthant $0 \le x_1 < \theta_{21}$, $\theta_{12} < x_2 \le max_2$, and $\theta_{33} < x_3 \le max_3$ (the orthant demarcated by bold lines).

Since Equations (13) are linear and orthogonal, the behavior inside an orthant is straightforward. In fact, all trajectories evolve towards a stable steady state $\boldsymbol{x}^* = \boldsymbol{\mu}/\boldsymbol{\gamma}$, the so-called *focal state*. The focal state of an orthant may lie inside or outside the orthant. If the focal state lies outside the orthant, the trajectories will tend towards one or several of the threshold planes bounding the orthant. The trajectories may be continued in an adjacent orthant, with a new focal state determined by the reduced state equations in that orthant. Whether a trajectory can be continued in an adjacent orthant will depend on the precise way in which the PLDEs (9) are defined in the threshold planes, where the step functions are discontinuous. Several alternatives have been proposed in the literature (Glass, 1977; Gouzé and Sari, 2001; Plahte *et al.*, 1994), ranging from restrictions on the function g_i to general solutions based on sophisticated mathematical techniques.

The PLDEs (9) have two types of steady states. In addition to *regular* steady states lying inside the orthants, there are *singular* steady states situated on one or more threshold planes separating the orthants (Snoussi and Thomas, 1993). Regular steady states can be readily identified, since they are the focal state of some orthant in the n-box. Singular steady states, important as they are as possible homeostatic points of the regulatory system, pose a difficulty for the piecewise-linear formalism, due to the discontinuities at the thresholds. Techniques to identify these steady states are presented by Plahte *et al.* (1977, 1994), Snoussi and Thomas (1993), and Gouzé and Sari (2001).

The global behavior of the PLDEs can be quite complex and is not well understood in general. Continuations of trajectories in several orthants may give rise to oscillations towards a singular steady state located at the intersection of threshold planes, cycles, or limit cycles (Glass and Pasternack, 1978a, 1978b; Lewis and Glass, 1991; Mestl *et al.*, 1995b; Plahte *et al.*, 1994). Numerical simulation studies have shown that aperiodic, chaotic dynamics can occur for $n \geq 4$ and become quite common for higher dimensions and particular characteristics of the functions g_i (Glass and Hill, 1998; Lewis and Glass, 1991, 1992; Mestl *et al.*, 1996, 1997). Mestl *et al.* (1996) analyzed the occurrence of chaos in some detail in an example network.

The use of PLDEs is exemplified by the method of *generalized threshold models* developed by Tchuarev and colleagues (Prokudina *et al.*, 1991; Tchuraev, 1991) (see also Kananyan *et al.* [1981], Ratner and Tchuraev [1978], and Tchuraev and Ratner [1983] for earlier formulations of threshold models). They use PLDEs that have been embedded in a finite-state machine framework, allowing regulatory interactions to be modeled in a more explicit and detailed manner than is possible by (9) alone. Among other things, the average time a regulatory protein is bound to the DNA can be taken into account, as well as the existence of several copies of a gene in different functional states. The resulting equations may contain time delays (Section 6). Generalized threshold models have been applied to study the dynamics of the regulation of tryptophan synthesis and arabinose catabolism in *E. coli*. The numerical simulation results suggest a number of hypotheses on the functioning of the regulatory systems, among other things the relative importance of three different mechanisms controlling the synthesis of tryptophan (Prokudina *et al.*, 1991). Other applications of PLDE models can be found in Alur *et al.* [2001], de Jong *et al.* [2001] and Ghosh and Tomlin [2001].

How much information is lost by the approximation of sigmoid regulation functions by step functions? Numerical simulation studies by Glass and Kauffman (1972, 1973) and Glass and Pérez (1974) show that in most cases there is no difference in the qualitative properties of the solutions when step functions instead of Hill functions are used (see also Plahte *et al.* [1995b, 1998], and Thomas and d'Ari [1990]; Plahte *et al.* [1994] discusses situations in which differences do occur). In the context of a study on the regulation of free iron level in mammalian cells, Omholt and colleagues have shown that as the step functions were relaxed to moderately steep sigmoids, the singular steady state predicted on the basis of an analysis of the PLDEs was preserved (Omholt *et al.*, 1998; Plahte *et al.*, 1998). The numerical and analytical results turned out to be within the same order of magnitude, comparable to the experimental accuracy.

A major advantage of the use of PLDEs is the simplicity of the mathematical analysis entailed by the form of the equations. In a similar vein, Savageau (1969, 1996a, 1996b) has proposed to define g_i in (5) by means of *power-law functions*, giving rise to nonlinear models of the form

$$\dot{x}_i = \sum_{k=1}^{p} \kappa_{ik} \prod_{j=1}^{n} x_j^{r_{ijk}} - \sum_{k=1}^{p} \gamma_{ik} \prod_{j=1}^{n} x_j^{s_{ijk}}, \quad 1 \leq i \leq n, \tag{14}$$

where r_{ijk} (s_{ijk}) are kinetic orders for elementary processes contributing to the production (degradation) of x_i, and κ_{ik} (γ_{ik}) are rate constants for these processes ($\kappa_{ik}, \gamma_{ik} > 0$). By changing to a logarithmic scale, (14) becomes a linear system that is much more tractable to analysis than the original nonlinear system (see Savageau [1996b], and Voit [1991, 2000] for reviews and a software package). Among other things, the power-law formalism has been applied to the study of the functional effectiveness of different types of coupling in regulatory networks (Savageau, 1989, 1996a).

8. QUALITATIVE DIFFERENTIAL EQUATIONS

Piecewise-linear differential equations of the form (9) can be analyzed qualitatively by defining qualitative states that correspond to the orthants of the phase space. Whenever trajectories can pass from one orthant to another, there is a transition between the corresponding qualitative states. This results in a transition graph summarizing the qualitative dynamics of the system (Edwards *et al.*, 2001; Glass, 1975a. A similar idea was encountered in the logical method discussed in Section 5. In fact, Snoussi (1989) has demonstrated that the generalized logical formalism of Thomas and colleagues can be seen as an abstraction of a special case of (9).

The idea of abstracting a discrete description from a continuous model and analyzing the discrete instead of continuous equations to draw conclusions about the dynamics of the system, is central to work on qualitative reasoning in artificial intelligence. One of the best-known formalisms developed for this purpose are the *qualitative differential equations (QDEs)* used in the simulation method QSIM (Kuipers, 1986, 1994). QDEs are abstractions of ODEs of the form

$$\frac{dx_i}{dt} = f_i(\boldsymbol{x}), \;\; 1 \leq i \leq n, \tag{15}$$

where $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ can be any linear or nonlinear function. The variables $\boldsymbol{x}$ take a *qualitative value* composed of a qualitative magnitude and direction. The qualitative magnitude of a variable x_i is a discrete abstraction of its real value, while the qualitative direction is the sign of its derivative. The function f_i is abstracted into a set of *qualitative constraints* which restrict the possible qualitative values of the variables. Given an initial *qualitative state* consisting of the qualitative values for $\boldsymbol{x}$ at the initial time-point, the QSIM algorithm generates a tree of *qualitative behaviors.* Each behavior in the tree describes a possible sequence of state transitions from the initial state. It has been proven that every qualitatively distinct behavior of the ODE corresponds to a behavior in the tree generated from the QDE, although the reverse may not be true.

The incomplete understanding of many genetic regulatory mechanisms on the molecular level and the general absence of quantitative knowledge have stimulated an interest in qualitative simulation techniques. An example of their application is given by Heidtke and Schulze-Kremer (1998b). Their case study concerns λ phage growth control in *E. coli.* The model consists of seven QDEs representing the different stages of λ phage growth as it follows either the lytic or lysogenic pathway. Each QDE describes in detail the infected bacterium including explicit representation of λ phages inside and outside the cell, viral DNA, ribosomes, mRNA, and proteins, and the way in which they interact to control major cellular events. For other examples of the application of qualitative reasoning concepts to gene regulation, see Akutsu *et al.* [2000], de Jong *et al.* [2001], Koile and Overton [1989], and Trelease *et al.* [1999].

A problem with qualitative simulation approaches is their limited upscalability. As a consequence of the weak nature of qualitative constraints and the difficulty to identify implicit constraints, behavior trees quickly grow out of bounds. This causes the range of application of the methods to be limited to regulatory systems of modest size and complexity. Systems of even a few genes related by positive and negative feedback loops cannot be handled, unless these systems have been so well-studied already that behavior predictions can be tightly constrained. An attractive feature of the class of PLDEs discussed in Section 7 is that they put strong constraints on local behavior in the phase space. By reframing the mathematical analysis of these models in a qualitative simulation framework, using a simulation algorithm tailored to the equations, larger networks with complex feedback loops can be treated. The resulting method has been applied to the simulation of the initiation of sporulation in *B. subtilis* (de Jong *et al.*, 2001).

Extensions of qualitative simulation methods allow weak numerical information in the form of interval bounds on the qualitative magnitude and direction of the variables, and numerical envelopes around the functions f_i, to be integrated in the simulation process (Berleant and Kuipers, 1997). This presents another way to restrict the number of behaviors, while it may also refine the predictions of the remaining behaviors. Heidtke and Schulze-Kremer (1998b) show how the conclusions of their model can be made more precise by adding semiquantitative information. The integration of numerical information is more difficult to achieve in the logical approaches discussed above, in which the relation between the discrete and the underlying continuous models is less direct.

Much work in qualitative reasoning has focused on the automated composition of an appropriate model of a system, given a user query, from a situation description and a domain knowledge base (Nayak, 1995). One of the basic approaches towards automated modeling, *qualitative process theory* (Forbus, 1984), has been used by Karp in his method for the construction and revision of gene regulation models (Karp, 1991, 1993a, 1993b). From a taxonomic knowledge base describing biological objects like strains of bacteria, genes, enzymes, and amino acids, and a process knowledge base comprising a theory about biochemical reactions, the program GENSIM predicts the outcome of a proposed experiment. If the predictions do not match with the observations made while actually carrying out the experiment, then the program HYPGENE generates hypotheses to explain the discrepancies. In particular, it revises assumptions about the experimental conditions that GENSIM used to derive its predictions. The programs have been able to partially reproduce the experimental reasoning underlying the discovery of the attenuation mechanism regulating the synthesis of tryptophan in *E. coli*. Other examples of the use of automated modeling techniques in the context of gene regulation can be found in Heidtke and Schulze-Kremer [1998a] and Koile and Overton [1989] (see also de Jong and Rip [1997] and Kohn [2001]).

9. PARTIAL DIFFERENTIAL EQUATIONS AND OTHER SPATIALLY DISTRIBUTED MODELS

Differential equations of the form (5) describe genetic regulatory processes while abstracting from spatial dimensions. The regulatory systems of interest are assumed, implicitly, to be spatially homogeneous. There are situations in which these assumptions are not appropriate. It might be necessary, for instance, to distinguish between different compartments of a cell, say the nucleus and the cytoplasm, and to take into account the diffusion of regulatory proteins or metabolites from one compartment to another. Moreover, gradients of protein concentrations across cell tissues are a critical feature in embryonal development. The introduction of time delays for diffusion effects allows some aspects of spatial inhomogeneities to be dealt with, while preserving the basic form of the rate equations (Busenberg and Mahaffy, 1985; Smolen *et al.*, 1999) (Section 6). However, in the case that multiple compartments of a cell, or multiple cells, need to be explicitly modeled, a more drastic extension of (5) becomes necessary.

Suppose that a multicellular regulatory system is considered, where the p cells are arranged in a row, as shown in Figure 11(a). We introduce a vector $\boldsymbol{x}^{(l)}(t) \geq \mathbf{0}$, which denotes the time-varying concentration of gene products in cell l, with l a discrete variable ranging from 1 to p. *Within* each cell, regulation of gene

FIG. 11. Examples of spatial configurations of multicellular genetic regulatory systems.

expression occurs in the manner described by equation (5). *Between* pairs of adjacent cells l and $l+1$, $1 \leq l < p$, diffusion of gene products is assumed to occur proportionally to the concentration differences $x_i^{(l+1)} - x_i^{(l)}$, $x_i^{(l)} - x_i^{(l-1)}$, and a diffusion constant δ_i. Taken together, this leads to a system of coupled ODEs, so-called *reaction-diffusion equations*

$$\frac{dx_i^{(l)}}{dt} = f_i(\boldsymbol{x}^{(l)}) + \delta_i \left(x_i^{(l+1)} - 2x_i^{(l)} + x_i^{(l-1)} \right), \quad 1 \leq i \leq n, \quad 1 < l < p. \tag{16}$$

Notice that f_i is the same for all l, in order to account for the fact that the genetic regulatory network is the same in every individual cell.

The reaction-diffusion equations are valid for the case that cells are arranged in a row, but they can be generalized to other one-dimensional and higher-dimensional spatial configurations (Fig. 11). In addition, the diffusion constants can be made to vary at different locations. Although (16) has been introduced here with a multicellular regulatory system in mind, it is valid for compartmental unicellular systems as well, perhaps with some small adaptations (see Glass and Kauffman [1972] for an example).

If the number of cells is large enough, the discrete variable l in (16) can be replaced by a continuous variable ranging from 0 to λ, where λ represents the size of the regulatory system. The concentration variables $\boldsymbol{x}$ are now defined as functions of both l and t, and the reaction-diffusion equations become *partial differential equations (PDEs)*:

$$\frac{\partial x_i}{\partial t} = f_i(\boldsymbol{x}) + \delta_i \frac{\partial^2 x_i}{\partial l^2}, \quad 0 \leq l \leq \lambda, \quad 1 \leq i \leq n. \tag{17}$$

If it is assumed that no diffusion occurs across the boundaries $l = 0$ and $l = \lambda$, the boundary conditions become

$$\frac{\partial^2}{\partial l^2} x_i(0, t) = 0 \text{ and } \frac{\partial^2}{\partial l^2} x_i(\lambda, t) = 0. \tag{18}$$

Reaction-diffusion equations have been widely used in mathematical biology to study pattern formation in development (see Gierer [1981], Kauffman [1993], Maini *et al.* [1997], Meinhardt [1982], and Nicolis and Prigogine [1977] for reviews). The equations are also referred to as Turing systems, after the mathematician who suggested their applicability to developmental phenomena (Turing, 1951). Turing considered a special case of Equations (16) and (17), with two concentration variables x_1 and x_2 (also called *morphogens*). Most of the work on reaction-diffusion equations is concerned with this special case, although higher-dimensional systems have been investigated as well (e.g., Gierer [1981] and Lacalli [1990]).

In what follows, continuous equations (17) will be considered for the case $n = 2$. Not surprisingly, direct analytical solution of this system of nonlinear PDEs is not possible in general. However, suppose that there exists a unique, spatially homogeneous steady state (x_1^*, x_2^*), such that $x_1^*, x_2^* > 0$ (Fig. 12). The spatial homogeneity of the steady state ensures that it is consistent with boundary conditions (18). By linearizing around (x_1^*, x_2^*), the behavior of the system in response to a small perturbation of the steady state can be predicted. Let Δx_i represent the deviation of x_i from the homogeneous steady-state concentration x_i^* on $[0, \lambda]$. Given the above boundary conditions,

$$\Delta x_i(l, t) = x_i(l, t) - x_i^* = \sum_{k=0}^{\infty} c_{ik}(t) \cos\left(\frac{\pi k}{\lambda} l\right), \quad i = 1, 2, \tag{19}$$

with $\cos(\pi kl/\lambda)$ the *modes* or *eigenfunctions* of the Laplacian operator $\partial^2/\partial l^2$ on $[0, L]$, and $c_{ik}(t)$ the *mode amplitudes* (Fig. 12) (Britton, 1986; Nicolis and Prigogine, 1977; Segel, 1984). The steady state (x_1^*, x_2^*) is stable to perturbations, if the mode amplitudes c_{ik} all decay exponentially in response to a fluctuation decomposable into a large number of weighted modes. However, if at least one mode amplitude c_{ik}, $k \neq 0$, causes the corresponding mode $\cos(\pi kl/\lambda)$ to grow exponentially, then the steady state is

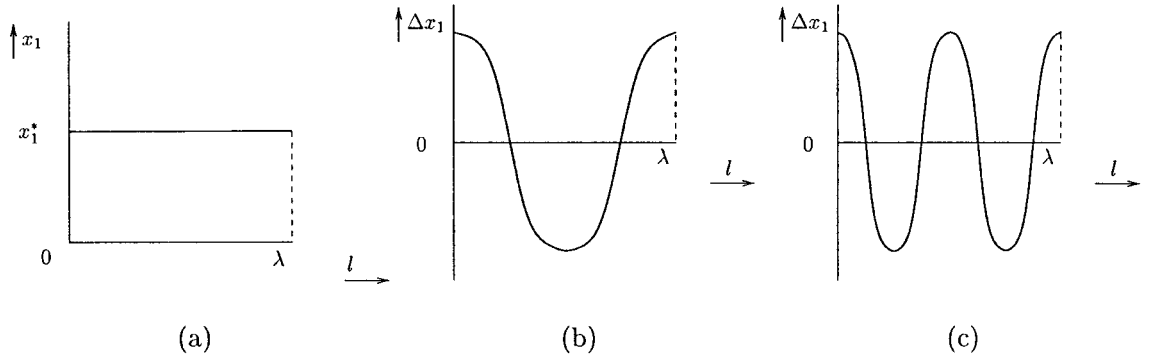


FIG. 12. **(a)** Homogeneous steady state of the reaction-diffusion system. **(b)–(c)** Two modes of the solution of the linearized system around the steady state. In **(b)** $k = 2$ and in **(c)** $k = 4$.

unstable. In this case, diffusion between adjacent cells does not have a homeogenizing influence, but instead entails spatially heterogeneous gene expression patterns. Details on the mathematics of the stability analysis can be found in the references listed above.

Gierer and Meinhardt have formulated constraints on the functions and parameters in (17) such that an initial steady state can be destabilized by a small fluctuation (Gierer, 1981; Gierer and Meinhardt, 1972; Meinhardt, 1982, 2000). Among other things, the product of gene 1, the *activator*, must positively regulate itself, *i.e.*, $\partial f_1/\partial x_1 > 0$, while the product of gene 2, the *inhibitor*, must negatively regulate gene 1, *i.e.*, $\partial f_1/\partial x_2 < 0$. Further, the inhibitory effect must be sufficiently strong and be relatively fast compared to the activating effect. Both the activator and inhibitor diffuse through the system ($\delta_1, \delta_2 > 0$), but the latter has to do so more rapidly than the former to achieve the necessary combination of short-range activation and long-range inhibition ($\delta_2/\delta_1 > 1$). Another prerequisite for diffusion-driven instability is that the size λ of the domain is larger than some minimum domain size (or, what comes to the same thing, the diffusivity of the inhibitor is smaller than a certain minimum diffusion coefficient, or the reactions proceed at a rate faster than a certain maximum reaction rate).

Activator-inhibitor systems have been extensively used to study the emergence of segmentation patterns in the early *Drosophila* embryo. In the early stage of embryogenesis, the embryo forms a syncytial blastoderm, a single cell with many nuclei. This permits spatial interactions to be conveniently treated in terms of diffusion of gene products. Goodwin and Kauffman (1989, 1990) and Kauffman (1993), following earlier suggestions in Kauffman *et al.* (1978), have noted that the observed spatial and temporal expression patterns of genes involved in the segmentation process much resemble the modes of the linearization of (17) around its equilibrium. By taking parameters slowly time-varying to mimic developmental effects, sequences of sigmoids of increasing frequency (increasing k, see Fig. 12) are generated that conform to the expression of the pair-rule genes in the mid-embryo. Numerical simulation studies have demonstrated that some aspects of stripe formation in *Drosophila* can indeed be reproduced in this way (e.g., Bunow *et al.* [1980], Lacalli [1990], Nagorcka [1988]). Other examples of modeling *Drosophila* embryogenesis by means of reaction-diffusion systems can be found in Hunding [1989], Hunding *et al.* [1990], Lacalli *et al.* [1988], Meinhardt [1977, 1986, 1988], Russell [1985] (see Kauffman [1993] and Meinhardt [1982] for reviews).

The use of reaction-diffusion equations in modeling spatially-distributed gene expression is compromised by the observation that the predictions are quite sensitive to the shape of the spatial domain, the initial and boundary conditions, and the chosen parameter values (Bard and Lauder, 1974; Bunow *et al.*, 1980). This affects the credibility of the models, as biological variation among individual members of a species or small fluctuations in environmental conditions do not generally lead to large deviations of developmental trajectories.

Even more seriously, the activator-inhibitor models commonly used suffer from the fact that scarce experimental evidence has been found for the hypothesis that two reacting and diffusing morphogens underlie pattern formation. From what is known from experimental studies of *Drosophila* development, it appears that the situation is much more complex. The regulatory system consists of layered networks of

tens or hundreds of developmental genes that mutually interact and whose products diffuse through the embryo. It seems, therefore, preferable to work with instances of the reaction-diffusion Equations (16) and (17) that reflect the underlying regulatory networks more directly.

The *gene circuit method* proposed by Reinitz and colleagues (Mjolsness *et al.*, 1991; Reinitz *et al.*, 1995, 1998; Reinitz and Sharp, 1995), reviewed in Reinitz and Sharp (1996) is an example of such an approach (see Burstein [1995], Edgar *et al.* [1989], Hamahashi and Kitano [1998], Kyoda *et al.* [2000], Sánchez *et al.* [1997], and von Dassow *et al.*]2000] for related work). The method employs a variant of (16):

$$\frac{dx_i^{(l)}}{dt} = \rho_i r_i \left(\sum_{j=1}^{n} \lambda_{ij} x_j + \mu_j u_1^{(l)} + \kappa_i) - \gamma_i x_i^{(l)} + \delta_i(k)(x_i^{(l+1)} - 2x_i^{(l)} + x_i^{(l-1)} \right),$$
$$1 \leq i \leq n, \ 1 < l < p, \tag{20}$$

where $\boldsymbol{x}^{(l)} \geq \mathbf{0}$ is a state vector of protein concentrations in nucleus l, and $u_1^{(l)} \geq \mathbf{0}$ an input variable. The real parameters λ_{ij} and μ_j describe the contribution of the state and input variables, respectively, to the expression rate of $x_i^{(l)}$. The sum of the regulatory influences, including a basal expression term $\kappa_i > 0$, is modulated by a sigmoid regulation function $r_i : \mathbb{R} \to \mathbb{R}$ and a constant $\rho_i > 0$ indicating the maximum expression rate of gene i. The diffusion parameter is the same for every protein, but depends on the number k of nuclear divisions that have taken place.[3]

Equations (20) have been used to model the formation of striped patterns of the products of gap genes and pair-rule genes in the middle region of the *Drosophila* blastoderm (see Marnellos *et al.* [2000] and Marnellos and Mjolsness [1998] for other applications). State variables are introduced for the gap genes *Kruppel*, *knirps*, *giant*, and *hunchback*, and for the pair-rule gene *even-skipped* ($n = 5$). The input variable represents the maternal Bicoid concentration. The values of the parameters in (20) can be estimated by means of measurements of protein concentrations in the nuclei at a sequence of time-points (Kosman *et al.*, 1998; Myasnikova *et al.*, 2001; Reinitz and Sharp, 1995). Given that the sign of a λ_{ij} specifies the nature of the interaction between two genes i and j (positive, negative, or no interaction), the parameter values thus obtained yield the regulatory network underlying the formation of the striped Eve pattern. On the basis of a least-square fit to expression data, Reinitz and Sharp demonstrate that each border in the stripes 2 to 5 is controlled by one of the four gap genes and that for the pattern to emerge, the diffusivity of Eve should be orders of magnitude lower than the diffusivity of the gap gene proteins (Reinitz *et al.*, 1995; Reinitz and Sharp, 1995). Using the estimated parameter values, mutant patterns have also been predicted (Sharp and Reinitz, 1998).

The fact that parameters of the model may represent interactions in the regulatory network, like the λ_{ij}s in equation (20), deserves special mention. As noted above, it implies that an estimation of parameter values simultaneously fixes the regulatory structure of the system. Models of the system could thus be induced from expression data, without relying on prior knowledge on the existence of regulatory interactions between genes. This idea has already been encountered in Sections 3 and 4, of course, but is here generalized to the case of continuous dynamical models. Related work taking ordinary differential equations and difference equations as their point of departure has come up recently (Chen *et al.*, 1999b; D'haeseleer *et al.*, 2000; Noda *et al.*, 1998; van Someren *et al.*, 2000; Weaver *et al.*, 1999; Wessels *et al.*, 2001).

The induction of models from measurements of $\boldsymbol{x}$ at a sequence of time-points is made attractive by the growing availability of gene expression data. However, precise measurements of absolute expression levels are currently difficult to achieve. In addition, as a consequence of the dimensionality problem referred to in Section 3, the models need to be simple and are usually strong abstractions of biological processes (Erb and Michaels, 1999). For larger and more complex models, the computational costs of finding an optimal fit between the parameter values and the data may be prohibitively high.

[3]The definition of $\delta_i(k)$ reveals a particularity of Equation (20), namely that it only holds in interphase, the period between two nuclear divisions in which the gene products are synthesized. Mjolsness, Reinitz, and Sharp supplement the model by a transition rule that describes how the model and its parameters need to be changed after division (see Mjolsness *et al.* [1991] for details).

10. STOCHASTIC MASTER EQUATIONS

In principle, differential equations allow gene regulation to be described in great detail, down to the level of individual reaction steps like the binding of a transcription factor to a regulatory site, or the transcription of a gene by the stepwise advancement of DNA polymerase along the DNA molecule (Fig. 13) (Drew, 2001). However, a number of implicit assumptions underlying DE formalisms are no longer valid on the molecular level.

Differential equations presuppose that concentrations of substances vary *continuously* and *deterministically*, both of which assumptions may be questionable in the case of gene regulation (Gibson and Mjolsness, 2001; Gillespie, 1977; Ko, 1991, 1992; McAdams and Arkin, 1999; Nicolis and Prigogine, 1977; Rigney, 1979; Szallasi, 1999). In the first place, the small numbers of molecules of some of the components of the regulatory system compromise the continuity assumption. There may be only a few tens of molecules of a transcription factor in the cell nucleus, and a single DNA molecule carrying the gene. Second, deterministic change presupposed by the use of the differential operator d/dt may be questionable due to fluctuations in the timing of cellular events, such as the delay between start and finish of transcription. As a consequence, two regulatory systems having the same initial conditions may ultimately settle into different states, a phenomenon strengthened by the small numbers of molecules involved.

Instead of taking a continuous and deterministic approach, some authors have proposed to use *discrete* and *stochastic* models of gene regulation (Arkin *et al.*, 1998; Gillespie, 1977; McAdams and Arkin, 1997; Morton-Firth and Bray, 1998; Nicolis and Prigogine, 1977; Rigney, 1979). Discrete amounts $\boldsymbol{X}$ of molecules are taken as state variables, and a joint probability distribution $p(\boldsymbol{X}, t)$ is introduced to express the probability that at time t the cell contains X_1 molecules of the first species, X_2 molecules of the second species, etc. The time evolution of the function $p(\boldsymbol{X}, t)$ can now be specified as follows:

$$p(\boldsymbol{X}, t + \Delta t) = p(\boldsymbol{X}, t)\left(1 - \sum_{j=1}^{m} \alpha_j \Delta t\right) + \sum_{j=1}^{m} \beta_j \Delta t, \tag{21}$$

where m is the number of reactions that can occur in the system, $\alpha_j \Delta t$ the probability that reaction j will occur in the interval $[t, t + \Delta t]$ given that the system is in the state $\boldsymbol{X}$ at t, and $\beta_k \Delta t$ the probability that reaction j will bring the system in state $\boldsymbol{X}$ from another state in $[t, t + \Delta t]$ (Gillespie, 1977, 1992). Rearranging (21), and taking the limit as $\Delta t \to 0$, gives the *master equation* (van Kampen, 1997):

$$\frac{\partial}{\partial t} p(\boldsymbol{X}, t) = \sum_{j=1}^{m} (\beta_j - \alpha_j p(\boldsymbol{X}, t)). \tag{22}$$

Compare this equation with the rate equations (5) in Section 6. Whereas the latter determine how the state of the system changes with time, the former describes how the probability of the system being in a certain state changes with time. Notice that the state variables in the stochastic formulation can be reformulated as concentrations by dividing the number of molecules X_i by a volume factor.

Although the master equation provides an intuitively clear picture of the stochastic processes governing the dynamics of a regulatory system, it is even more difficult to solve by analytical means than the

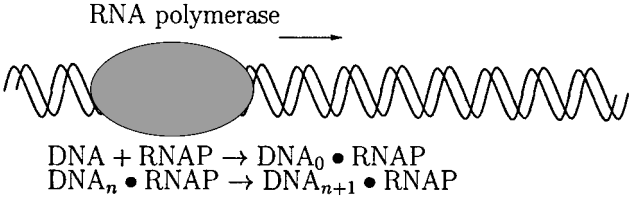


FIG. 13. Example of two reaction steps involved in DNA transcription. The first reaction step describes the binding of RNA polymerase (RNAP) to the transcription initiation site, while the second reaction step describes the advancement of RNAP along the DNA.

deterministic rate equation. Moreover, numerical simulation of the system is complicated by the form of (22), which contains $n + 1$ independent variables: n discrete variables $\boldsymbol{X}$ and a continuous variable t. Under certain conditions, the master equation can be approximated by stochastic differential equations, so-called Langevin equations, that consist of a deterministic ODE of the type (5) extended with a noise term (Gillespie, 2000; Nicolis and Prigogine, 1977; van Kampen, 1997). The conditions under which the approximation is valid may not always be possible to satisfy in the case of genetic regulatory systems.

An alternative approach would be to disregard the master equation altogether and directly simulate the time evolution of the regulatory system. This idea underlies the *stochastic simulation* approach developed by Gillespie (1977). Basically, the stochastic simulation algorithm (i) determines *when* the next reaction occurs and of *which* type it will be, given that the system is originally in state $\boldsymbol{X}$ 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 interval 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 (21). This guarantees that when a large number of stochastic simulations are carried out, the resulting distribution for $\boldsymbol{X}$ at t will approach the distribution implied by the master equation. In other words, whereas the master equation deals with behavior averages, obtained by calculating the averages and the variances of the X_is at t from $p(\boldsymbol{X}, t)$, stochastic simulation provides information on individual behaviors.

Gibson and Bruck (2000) have proposed various improvements of the Gillespie algorithm directed at reducing the computational complexity of the procedure. Morton-Firth and Bray (1998, 1999) discuss another algorithm that is more efficient at simulating systems with a large number of reactions. The algorithm does so by following individual molecules over time, and storing the way they combine into complexes that can have multiple phosphorylration, methylation, and other activation states. The algorithm, which is particularly suitable for simulating signal transduction pathways, has been implemented in the simulator StochSim.

Stochastic simulation has been used by McAdams and Arkin to analyze the interactions controlling the expression of a single prokaryotic gene (McAdams and Arkin, 1997) (see also Barkai and Leibler [1999]). In particular, they investigate how the time interval between the activation of one gene and the regulatory action of its product on another gene, the so-called switching delay, is affected by the stochastic nature of the transcription initiation intervals and the numbers of protein molecules produced per transcript. Using parameter values that approximate those for the expression of the *cro* gene from the P_R promotor in phage λ, they find that the protein is produced in sharp bursts occurring at random time intervals, leading to strong fluctuations in switching delays. Interestingly, this may provide an explanation of phenotypic variations in isogenic populations, when it is assumed that the variations arise from switching mechanisms which decide between alternative developmental paths depending on the effective concentrations of competitive regulatory proteins.

The above hypothesis has been pursued in later work, which investigates the influence of fluctuations in the rate of gene expression on the choice between lytic and lysogenic growth in phage λ (Arkin *et al.*, 1998). The predicted fraction of infected cells selecting the lysogenic pathway, at different infection levels, is consistent with experimental observations. The predictions of the models have been shown to be relatively insensitive to changes in the translation rate, protein dimerization rates, and protein degradation rates. However, they are somewhat sensitive to the transcription rate and quite sensitive to the average number of proteins per mRNA transcript (Gibson and Bruck, 1998).

Stochastic simulation results in closer approximations to the molecular reality of gene regulation, but its use is not always evident. In the first place, the approach requires detailed knowledge of the reaction mechanisms to be available, including estimates of the probability density function $p(\tau, \rho)$. Moreover, stochastic simulation is costly, due to the large number of simulations that need to be carried out to calculate an approximate value of $p(\boldsymbol{x}, t)$ and due to the large number of reactions that need to be traced in each of these simulations. Whether the costs always balance the expected benefits depends on the level of granularity at which one wishes to study regulatory processes. On a larger time-scale, stochastic effects may level out, so that continuous and deterministic models form a good approximation (see Gillespie [2000] for a discussion).

11. RULE-BASED FORMALISMS

The model formalisms discussed so far, whether they are continuous or discrete, static or dynamic, deterministic or stochastic, all use a restricted notion of state. In fact, the state of a regulatory system is equated with the concentration or the number of molecules of proteins, mRNAs, metabolites, and other species at a certain time-point. *Knowledge-based* or *rule-based simulation formalisms*, developed in the field of artificial intelligence, permit a richer variety of knowledge about the system to be expressed in a single formalism. For instance, the logical and physical structure of a gene—such as the relative position of the regulatory sites at which transcription is initiated, prevented, and aborted—can be conveniently represented in the rule-based approach.

Basically, rule-based formalisms consist of two components, a set of *facts* and a set of *rules*, that are stored in a knowledge base (Hayes-Roth *et al.*, 1983). Facts express knowledge about the objects of a regulatory system. Each object is described by properties that take their value from well-defined domains. The objects are usually structured in a hierarchy of classes. For example, the class `DNA-Sequence` could be defined as consisting of objects with properties that include `topology` and `strandedness`. A specific DNA sequence might have the value `circular` for `topology` and `single-stranded` for `strandedness`. The class `DNA-Sequence` could be defined as a subclass of the class `Sequences`. Other objects of interest in a regulatory system include RNAs, proteins, cells and cell boundaries, but also experimental conditions and processes like transcription, splicing, and translation. The rules in the knowledge base consist of two parts, a *condition part* and an *action part.* The condition part expresses conditions in terms of properties of objects, while the action part operates upon the objects, e.g., by changing a property of an existing object. An example of a rule describing the conditions under which conditions DNA polymerase I will bind to bacterial DNA, adapted from Brutlag *et al.* (1991) and Galper *et al.* (1993), is

```
IF (AND (temperature-range of Exp-conditions is 0-to-45)
        (ionic-strength-range of Exp-conditions is 0.001-to-.3)
        (pH-range of Exp-conditions is 6.0-to-9.5))
THEN (activity of DNA-polymerase-I is DNA-binding)
```

Rule-based simulation consists of the repeated process of matching the facts in the knowledge base against the condition parts of the rules and carrying out the action part of the rules whose conditions are satisfied. A control strategy determines the order in which the rules are evaluated and resolves the conflicts that arise when several rules match at the same time. Control strategies vary from random selection to complicated priority schemes, for instance, procedures that take into account estimates of the *a priori* probability that a rule will fire. Rule-based simulation can be used in a *forward-chaining* or *backward-chaining* mode. In the case of forward chaining, sketched above, one deduces all facts implied by the set of rules and the initial facts in the knowledge base. In the case of backward chaining, the facts are conclusions to be explained and simulation proceeds in the reverse direction. The facts are recursively matched against the action parts of the rules in order to find all possible combinations of facts that can lead to the conclusions.

Meyers and Friedland have applied rule-based simulation techniques to model λ phage growth control (Meyers and Friedland, 1984). They have performed simulations to study the decision between the lytic and lysogenic pathway for wild-type λ phage and several single and double mutants. Their simulations produced correct results, except in two cases in which the assumption of deterministic behavior proved deleterious. The λ phage example is also used by Shimada and colleagues (Shimada *et al.*, 1995), a novel aspect of their work being that the regulation process is considered on two levels of abstraction. The system uses rule-based simulation for the "noncritical" parts of the decision between the lytic and lysogenic pathway and a qualitative phase-space analysis of rate equations for the "critical" parts.

The major advantage of rule-based formalisms, their capability to deal with a richer variety of biological knowledge in an intuitive way, is counteracted by difficulties in maintaining the consistency of a (revised) knowledge base and the problem to incorporate quantitative information. Although attempts have been made to integrate symbolic and numerical knowledge into a single formalism, for instance in the *genetic grammar* underlying the Metabolica system developed by Hofestädt and colleagues (Hofestädt, 1993; Hofestädt and Meineke, 1995), it remains true that in this respect rule-based formalisms cannot compete with the continuous formalisms discussed in previous sections.

12. CONCLUSIONS

The formalisms discussed in this review allow genetic regulatory systems to be modeled in quite different ways. In Table 1 the different approaches are compared on a number of dimensions, such as whether the models are discrete or continuous, static or dynamic, deterministic or stochastic, and qualitative or quantitative. In addition, the approaches are compared with respect to the level of granularity on which they describe regulatory processes. Apart from constraining the capability of the models to answer certain biological questions, the level of granularity to a large extent determines the amount of computer power required to simulate a regulatory network.

Until recently, modeling and simulation studies have predominantly used deterministic, coarse- to average-grained models, such as logical models and simple differential equation models. Although quantitative models have been used, the conclusions drawn from them have been mostly qualitative. This situation can be explained by two major difficulties hampering the modeling and simulation of genetic regulatory networks. In the first place, the biochemical reaction mechanisms underlying regulatory interactions are usually not known or are incompletely known. This means that detailed kinetic models cannot be built and more approximate models are needed. In the second place, quantitative information on kinetic parameters and molecular concentrations is only seldom available. As a consequence, traditional methods for numerical analysis are difficult to apply.

Not surprisingly, the few modeling and simulation studies using fine-grained, quantitative, and stochastic models have been restricted to regulatory networks of small size and modest complexity that have been well-characterized already by experimental means. For such model systems, the paradigm example being phage λ discussed in Section 6, we do have detailed knowledge on regulatory mechanisms and at least partial quantitative information. Besides a lack of relevant information to build models and test predictions, the use of fine-grained, stochastic models is also hampered by the inherent computational complexity of the simulation problems. The expression of a single eukaryotic gene may involve a dozen of transcription factors interacting in complex ways (Arnone and Davidson, 1997; Yuh *et al*., 1998). Each of these transcription factors may be modified by many other proteins, thus giving rise to large and complex networks of interactions.

It can be reasonably expected that at least some of the problems mentioned above will be considerably relieved in the near future. The emergence of new experimental techniques, along with the development of databases and other infrastructural provisions giving access to published and unpublished experimental data, promise to relieve the data bottleneck. Together with the continuing increase of computer power, this might allow hitherto impractical approaches to modeling and simulation to be tried.

Table 1. Summary of Properties of Different Modeling Formalisms: Directed Graphs (DG), Bayesian Networks (BYN), Boolean Networks (BNN), Generalized Logical Networks (GLN), Nonlinear Differential Equations (NLDE), Piecewise-Linear Differential Equations (PLDE), Partial Differential Equations (PDE), Stochastic Master Equations (SME), and Rule-Base Formalisms (R)

	Static (s), dynamic (d)	*Discrete (d), continuous (c)*	*Deterministic (d), stochastic (s)*	*Qualitative (ql), quantitative (qn)*	*Coarse (c), average (a), fine (f) grained*
DG	s		d	ql	c
BYN	s[a]	d,c	s	qn	c
BNN	d	d	d	ql	c
GLN	d	d	d	ql	a
NLDE	d	c	d	qn	a,f
PLDE	d	c	d	ql,qn[c]	a
QDE	d	d	d	ql	a,f
PDE	d	c[b]	d	qn	a,f
SME	d	d	s	qn	f
R	d	d	d	ql	a,f

[a]Generalization to dynamic Boolean networks is possible.
[b]Spatial dimension is often discretized.
[c]Qualitative analysis of models is possible.

Functional genomics has yielded experimental techniques allowing interactions between genes and gene products to be elucidated in a large-scale manner. An example is the use of cDNA microarrays to monitor protein–DNA interactions (Ren *et al.*, 2000; Iyer *et al.*, 2001). Key contributions to the study of protein–protein interactions have been made by mass spectrometric protein identification and by the yeast two-hybrid system (Pandey and Mann, 2000; Tucker *et al.*, 2001; Zhu and Snyder, 2001). Increasing knowledge on the molecular mechanisms underlying gene regulation will eventually allow regulatory systems to be modeled on a finer level of granularity than is currently possible.

Other techniques will promote the use of quantitative models of gene regulation. For example, the ability to synthesize regulatory networks *in vivo*, as described by Becskei and Serrano (2000), Elowitz and Leiber (2000) and Gardner *et al.* (2000), might facilitate the direct measurement of model parameters. Although the expression profiles provided by cDNA microarrays, oligonucleotide chips, and other tools to measure the evolution of the state of a cell are currently effectively qualitative in nature, expected improvements in the (re)production and the statistical interpretation of the measurements may allow veritable quantitative approaches to take hold. The use of quantitative models permits larger systems to be studied at higher precision (Koshland, 1998; Maddox, 1992).

The above-mentioned advances in biology and computer technology may bring us nearer to what seems to be the ultimate goal of modeling and simulation efforts: to study the behavior of entire prokaryotic or even eukaryotic organisms *in silico*. That is, the use of models that integrate gene regulation with metabolism, signal transduction, replication and repair, and a variety of other cellular processes. Several initiatives to achieve this have been launched in recent years, including the simulation of a virtual cell using the E-CELL or initial cell environments (Tomita *et al.*, 1999; Tomita, 2001; Loew and Schaff, 2001), the simulation of the development of the *Drosophila* embryo in the framework of the Virtual Drosophila project (Kitano *et al.*, 1997), and the analysis of metabolic networks in *E. coli* and *H. influenzae* (Edwards and Palsson, 2000; Schilling *et al.*, 2000; Schilling and Palsson, 2000).

Even if computer technology would develop to the point that whole cells and organisms can be simulated on the molecular level, by tracing hundreds of thousands of reactions occurring in parallel, such brute-force strategies are not guaranteed to yield insight into the functioning of living systems. In fact, it may be far from straightfoward to find meaningful patterns in the masses of data generated by whole-cell simulations. As an alternative, one could focus on smaller subsystems first, that is, start with modules that can be studied in relative isolation because they are loosely connected through regulatory interactions or act on different time-scales. In a second step, the analysis of the individual modules could be supplemented by an investigation of the interactions between the modules, possibly on a more abstract and hence computationally less demanding level. The modular organization of genetic regulatory networks has been put forward in studies of *Drosophila* segmentation (von Dassow *et al.*, 2000) and flower morphogenesis in *Arabidopsis* (Mendoza *et al.*, 1999) (see also Hartwell *et al.* [1999] and Thieffry and Romero [1999]).

The overall picture that is emerging may thus not show a big super model covering all aspects of cellular dynamics in great detail, but rather a hierarchy of models accounting for different aspects of the cell on different levels of abstraction. Switching from one level to another, one can follow up an analysis of the global regulatory structure of the system by an in-depth look at the fine structure of a regulatory module. Whereas, for the first problem, coarse-grained models of the type discussed in Sections 3, 5, or 7 may be sufficient, the latter problem may require detailed models of the form encountered in Sections 6 and 10.

One of the consequences of this view is the emphasis it puts on modeling, a task that has received less systematic attention than simulation thus far. The construction of the right model for the job, describing relevant aspects of the system on the desired level of granularity, demands methods for computer-supported modeling that are only in the process of emerging. Two burgeoning approaches towards computer-supported modeling have been encountered in this paper. On the one hand, models can be *composed* from knowledge on regulatory interactions stored in databases and knowledge bases, as illustrated in Section 8. On the other hand, models can be *induced* from expression data by methods like those discussed in Sections 2, 4, and 7. Each of these approaches has its merits, but neither of them seems sufficient in itself. Rather, it can be expected that a combination of the two approaches, exploiting a wide range of structural and functional information on regulatory networks, will be most effective.

Eventually, these efforts might result in computer-supported modeling environments (de Jong and Rip, 1997; Thieffry *et al.*, 1998) that integrate a variety of experimental and computational tools to assist the biologist in unraveling the structure and functioning of genetic regulatory networks.

13. ACKNOWLEDGMENTS

The author would like to thank Olivier Bernard, Olivier Martin, Michel Page, François Rechenmann, Denis Thieffry, and two anonymous referees for their comments on earlier versions of this paper.

REFERENCES

Ackers, G.K., Johnson, A.D., and Shea, M.A. 1997. Quantitative model for gene regulation by λ phage repressor. *Proc. Natl. Acad. Sci. USA*, 79, 1129–1133.

Akutsu, T., Kuhara, S., Maruyama, O., and Miyano, S. 1998a. Identification of gene regulatory networks by strategic gene disruptions and gene overexpressions. In H. Karloff, ed., *Proc. 9th Ann. ACM-SIAM Symp. Discr. Algorithms (SODA-98)*, 695–702, Society for Industrial and Applied Mathematics, Philadelphia.

Akutsu, T., Kuhara, S., Maruyama, O., and Miyano, S. 1998b. A system for identifying genetic networks from gene expression patterns produced by gene disruptions and overexpressions. In *Genome Informatics*, vol. 9, 151–160.

Akutsu, T., Miyano, S., and Kuhara, S. 1999. Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds., *Proc. Pac. Symp. Biocomput. (PSB'99)*, vol. 4, 17–28, World Scientific Publishing, Singapore.

Akutsu, T., Miyano, S., and Kuhara, S. 2000. Inferring qualitative relations in genetic networks and metabolic pathways. *Bioinformatics* 16(8), 727–734.

Alon, U., Barkai, N., Notterman, D.A., Gish, K., Ybarra, S., Mack, D., and Levine, A.J. 1999. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl. Acad. Sci. USA*, 96, 6745–6750.

Alur, R., Belta, C., Ivančíc, F., Kumar, V., Mintz, M., Pappas, G.J., Rubin, H., and Schlug, J. 2001. Hybrid modeling and simulation of biomolecular networks. In M.D. Di Benedetto and A. Sangiovanni-Vincentelli, eds., *Hybrid Syst. Comput. Control (HSCC 2001)*, vol. 2034 of *Lecture Notes in Computer Science*, 19–32, Springer-Verlag, Berlin.

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

Arkin, A., and Ross, J. 1995. Statistical construction of chemical reaction mechanisms from measured time-series. *J. Phys. Chem.* 99, 970–979.

Arkin, A., Ross, J., and McAdams, H.A. 1998. Stochastic kinetic analysis of developmental pathway bifurcation in phage λ-infected *Escherichia coli* cells. *Genetics* 149, 1633–1648.

Arkin, A., Shen, P., and Ross, J. 1997. A test case of correlation metric construction of a reaction pathway from measurements. *Science* 277, 1275–1279.

Arnone, M.I., and Davidson, E.H. 1997. The hardwiring of development: Organization and function of genomic regulatory systems. *Development* 124, 1851–1864.

Bagley, R.J., and Glass, L. 1996. Counting and classifying attractors in high dimensional dynamical systems. *J. Theor. Biol.* 183, 269–284.

Bard, J., and Lauder, I. 1974. How well does Turing's theory of morphogenesis work? *J. Theor. Biol.* 45, 501–531.

Barkai, N., and Leibler, S. 1997. Robustness in simple biochemical networks. *Nature* 387, 913–917.

Barkai, N., and Leibler, S. 1999. Circadian clocks limited by noise. *Nature* 403, 267–268.

Becskei, A., and Serrano, L. 2000. Engineering stability in gene networks by autoregulation. *Nature* 405, 590–591.

Ben-Dor, A., Shamir, R., and Yakhini, Z. 1999. Clustering gene expression patterns. *J. Comp. Biol.* 6(3-4), 281–297.

Berleant, D., and Kuipers, B. 1997. Qualitative and quantitative simulation: Bridging the gap. *Artif. Intell.* 95, 215–256.

Bliss, R.D., Painter, P.R., and Marr, A.G. 1982. Role of feedback inhibition in stabilizing the classical operon. *J. Theor. Biol.* 97, 177–193.

Borisuk, M.T., and Tyson, J.J. 1998. Bifurcation analysis of a model of mitotic control in frog eggs. *J. Theor. Biol.* 195, 69–85.

Bower, J.M., and Bolouri, H., eds. 2001. *Computational Modeling of Genetic and Biochemical Networks*, MIT Press, Cambridge, MA.

Britton, N.F. 1986. *Reaction-Diffusion Equations and Their Applications to Biology*, Academic Press, London.

Brown, P.A., and Botstein, D. 1999. Exploring the new world of the genome with DNA microarrays. *Nature Genet.* 21(suppl.), 33–37.

Brown, M.P.S., Grundy, W.N., Lin, D., Cristianini, N., Sugnet, C.W., Furey, T.S., Ares Jr, M., and Haussler, D. 2000. Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA* 97(1), 262–267.

Brownstein, M.J., Trent, J.M., and Boguski, M.S. 1998. Functional genomics. In M. Patterson and M. Handel, eds. *Trends Guide to Bioinformatics*, 27–29, Elsevier, Oxford.

Brutlag, D.L., Galper, A.R., and Millis, D.H. 1991. Knowledge-based simulation of DNA metabolism: Prediction of enzyme action. *CABIOS* 7(1), 9–19.

Bunow, B., Kernevez, J.-P., Joly, G., and Thomas, D. 1980. Pattern formation by reaction-diffusion instabilities: Application to morphogenesis in *Drosophila*. *J. Theor. Biol.* 84, 629–649.

Burstein, Z. 1995. A network model of developmental gene hierarchy. *J. Theor. Biol.* 174, 1–11.

Busenberg, S., and Mahaffy, J. 1995. Interaction of spatial diffusion and delays in models of genetic control by repression. *J. Math. Biol.* 22, 313–333.

Carrier, T.A., and Keasling, J.D. 1999. Investigating autocatalytic gene expression systems through mechanistic modeling. *J. Theor. Biol.* 201, 25–36.

Chen, K.C., Csikász-Nagy, A., Győrffy, B., Val, J., Novak, B., and Tyson, J.J. 2000. Kinetic analysis of a molecular model of the budding yeast cell cycle. *Mol. Biol. Cell* 11(1), 369–391.

Chen, T., Filkov, V., and Skiena, S.S. 1999. Identifying gene regulatory networks from experimental data. In *Proc. 3rd Ann. Int. Conf. Comp. Mol. Biol. (RECOMB'99)*, 94–103, New York, ACM Press.

Chen, T., He, H.L., and Church, G.M. 1999. Modeling gene expression with differential equations. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomuting (PSB'99)*, vol. 4, 29–40, Singapore, World Scientific Publishing.

Cherry, J.L., and Adler, F.R. 2000. How to make a biological switch. *J. Theor. Biol.* 203, 117–133.

Cho, R.J., Campbell, M.J., Winzeler, E.A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T.G., Gabrielian, A.E., Landsman, D., Lockhart, D.J., and Davis, R.W. 1998. A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* 2, 65–73.

Collado-Vides, J. 1989. A transformational-grammar approach to the study of the regulation of gene expression. *J. Theor. Biol.* 136, 403–425.

Collado-Vides, J. 1996. Integrative representations of the regulation of gene expression. In J. Collado-Vides, B. Magasanik, and T.F. Smith, eds. *Integrative Approaches to Molecular Biology*, 179–203, MIT Press, Cambridge, MA.

Collado-Vides, J., Gutièrrez-Ríos, R.M., and Bel-Enguix, G. 1998. Networks of transcriptional regulation encoded in a grammatical model. *BioSystems* 47, 103–118.

Collado-Vides, J., Magasanik, B., and Smith, T.F., eds. 1996. *Integrative Approaches to Molecular Biology*, MIT Press, Cambridge, MA.

Cornish-Bowden, A. 1995. *Fundamentals of Enzyme Kinetics*, rev. ed., Portland Press, London.

de Jong, H., Page, M., Hernandez, C., and Geiselmann, J. 2001. Qualitative simulation of genetic regulatory networks: Method and application. In B. Nebel, ed. *Proc. 17th Int. Joint Conf. Artif. Intell. (IJCAI-01)*, 67–73, San Mateo, CA. Morgan Kaufmann.

de Jong, H., and Rip, A. 1997. The computer revolution in science: Steps towards the realization of computer-supported discovery environments. *Artif. Intell.* 91, 225–256.

Delbrück, M. 1949. Discussion. In *Unités Biologiques Douées de Continuité Génétique*, 33, Éditions du CNRS, Lyon.

D'haeseleer, P., Liang, S., and Somogyi, R. 2000. Genetic network inference: From co-expression clustering to reverse engineering. *Bioinformatics* 16(8), 707–726.

D'haeseleer, P., Wen, X., Fuhrman, S., and Somogyi, R. 1998. Mining the gene expression matrix: Inferring gene relationships from large-scale gene expression data. In R.C. Paton and M. Holcombe, eds. *Information Processing in Cells and Tissues*, 203–212. Plenum Press, New York.

Drew, D.A. 2001. A mathematical model for prokaryotic protein synthesis. *Bull. Math. Biol.* 63, 329–351.

Edgar, B.A., Odell, G.M., and Schubiger, G. 1989. A genetic switch, based on negative regulation, sharpens stripes in *Drosophila* embryos. *Dev. Genet.* 10, 124–142.

Edwards, R. 2000. Analysis of continuous-time switching networks. *Physica D* 146, 165–199.

Edwards, R., and Glass, L. 2000. Combinatorial explosion in model gene networks. *Chaos* 10(3), 691–704.

Edwards, J.S., and Palsson, B.O. 2000. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* 97(10), 5528–5533.

Edwards, R., Siegelmann, H.T., Aziza, K., and Glass, L. 2001. Symbolic dynamics and computation in model gene networks. *Chaos* 11(1), 160–169.

Ehlde, M., and Zacchi, G. 1995. MIST: A user-friendly metabolic simulator. *CABIOS* 11(2), 201–207.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. 1998. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* 95, 14863–14868.

Elowitz, M.B., and Leiber, S. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403, 335–338.

Endy, D., and Brent, R. 2001. Modelling cellular behavior. *Nature* 409, 391–395.

Endy, D., You, L., Yin, J., and Molineux, I.J. 2000. Computation, prediction, and experimental tests of fitness for bacteriophage T7 mutants with permuted genomes. *Proc. Natl. Acad. Sci. USA* 97(10), 5375–5380.

Erb, R.S., and Michaels, G.S. 1999. Sensitivity of biological models to errors in parameter estimates. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'99)*, vol. 4, 53–64, Singapore. World Scientific Publishing.

Euzenat, J., Chemla, C., and Jacq, B. 1997. A knowledge base for *D. melanogaster* gene interactions involved in pattern formation. In T. Gaasterland, P. Karp, K. Karplus, C. Ouzounis, C. Sander, and A. Valencia, eds. *Proc. Fifth Int. Conf. Intell. Syst. Mol. Biol. (ISMB-97)*, 108–119, Menlo Park, CA. AAAI Press.

Fields, S., Kohara, Y., and Lockhart, D.J. 1999. Functional genomics. *Proc. Natl. Acad. Sci. USA* 96, 8825–8826.

Forbus, K.D. 1984. Qualitative process theory. *Artif. Intell.* 24, 85–168.

Friedman, N., Linial, M., Nachman, I., and Pe'er, D. 2000. Using Bayesian networks to analyze expression data. *J. Comput. Biol.* 7, 601–620.

Galper, A.R., Brutlag, D.L., and Millis, D.H. 1993. Knowledge-based simulation of DNA metabolism: Prediction of action and envisionment of pathways. In L. Hunter, ed. *Artificial Intelligence and Molecular Biology*, 365–395. AAAI Press/MIT Press, Menlo Park, CA/Cambridge, MA.

Gardner, T.S., Cantor, C.R., and Collins, J.J. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403, 339–342.

Ghosh, R., and Tomlin, C.J. 2001. Lateral inhibition through Delta-Notch signaling: A piecewise affine hybrid model. In M.D. Di Benedetto and A. Sangiovanni-Vincentelli, eds. *Hybrid Syst. Comput. Control (HSCC 2001)*, vol. 2034 of *Lecture Notes in Computer Science*, 232–246. Springer-Verlag, Berlin.

Gibson, M., and Bruck, J. 1998. An efficient algorithm for generating trajectories of stochastic gene regulation reactions. Technical Report ETR026, Parallel and Distributed Computing Group, CalTech, Pasadena, CA.

Gibson, M.A., and Bruck, J. 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A* 104, 1876–1889.

Gibson, M.A., and Mjolsness, E. 2001. Modeling the activity of single genes. In J.M. Bower and H. Bolouri, eds. *Computational Modeling of Genetic and Biochemical Networks*, 1–48. MIT Press, Cambridge, MA.

Gierer, A. 1981. Generation of biological patterns and form: Some physical, mathematical, and logical aspects. *Prog. Biophys. Mol. Biol.* 37, 1–47.

Gierer, A., and Meinhardt, H. 1972. A theory of biological pattern formation. *Kybernetik* 12, 30–39.

Gillespie, D.T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81(25), 2340–2361.

Gillespie, D.T. 1992. A rigorous derivation of the chemical master equation. *Physica D* 188, 404–425.

Gillespie, D.T. 2000. The chemical Langevin equation. *J. Chem. Phys.* 113(1), 297–306.

Glass, L. 1975a. Classification of biological networks by their qualitative dynamics. *J. Theor. Biol.* 54, 85–107.

Glass, L. 1975b. Combinatorial and topological methods in nonlinear chemical kinetics. *J. Chem. Phys.* 63(4), 1325–1335.

Glass, L. 1977. Global analysis of nonlinear chemical kinetics. In B. Berne, ed. *Statistical Mechanics, Part B: Time Dependent Processes*, 311–349, Plenum Press, New York.

Glass, L., and Hill, C. 1998. Ordered and disordered dynamics in random networks. *Europhys. Lett.* 41(6), 599–604.

Glass, L., and Kauffman, S.A. 1972. Co-operative components, spatial localization and oscillatory cellular dynamics. *J. Theor. Biol.* 34, 219–237.

Glass, L., and Kauffman, S.A. 1973. The logical analysis of continuous non-linear biochemical control networks. *J. Theor. Biol.* 39, 103–129.

Glass, L., and Pasternack, J.S. 1978. Prediction of limit cycles in mathematical models of biological oscillations. *B. Math. Biol.* 40, 27–44.

Glass, L., and Pasternack, J.S. 1978. Stable oscillations in mathematical models of biological control systems. *J. Math. Biol.* 6, 207–223.

Glass, L., and Pérez, R. 1974. Limit cycle oscillations in compartmental chemical systems. *J. Chem. Phys.* 61(12), 5242–5249.

Goldbeter, A. 1995. A model for circadian oscillations in the *Drosophila* Period protein (PER). *Proc. R. Soc. Lond. B* 261, 319–324.

Goldbeter, A. 1997. *Biochemical Oscillations and Cellular Rhythms : The Molecular Bases of Periodic and Chaotic Behaviour*, Cambridge University Press, Cambridge.

Goodwin, B.C. 1963. *Temporal Organization in Cells*, Academic Press, New York.

Goodwin, B.C. 1965. Oscillatory behavior in enzymatic control processes. In G. Weber, ed. *Advances in Enzyme Regulation*, 425–438. Pergamon Press, Oxford.

Goodwin, B.C., and Kauffman, S.A. 1989. Bifurcations, harmonics, and the four color wheel model of *Drosophila* development. In A. Goldbeter, ed. *Cell to Cell Signalling: From Experiments to Theoretical Models*, 213–227, Academic Press, London.

Goodwin, B.C., and Kauffman, S.A. 1990. Spatial harmonics and pattern specification in early *Drosophila* development: Part I. Bifurcation sequences and gene expression. *J. Theor. Biol.* 144, 303–319.

Goryanin, I., Hodgman, T.C., and Selkov, E. 1999. Mathematical simulation and analysis of cellular metabolism and regulation. *Bioinformatics* 15(9), 749–758.

Goss, P.J.E., and Peccoud, J. 1998. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proc. Natl. Acad. Sci. USA* 95, 6750–6755.

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

Gouzé, J.-L., and Sari, T. 2001. A class of piecewise linear differential equations arising in biological models. Technical Report RR-4207, INRIA Sophia-Antipolis, Sophia-Antipolis.

Griffith, J. 1968. Mathematics of cellular control processes: I. Negative feedback to one gene. *J. Theor. Biol.* 20, 202–208.

Griffith, J. 1968b. Mathematics of cellular control processes: II. Positive feedback to one gene. *J. Theor. Biol.* 20, 209–216.

Hamahashi, S., and Kitano, H. 1998. Simulation of *Drosophila* embryogenesis. In C. Adami, R.K. Belew, H. Kitano, and C. Taylor, eds. *Proc. 6th Int. Conf. Artif. Life (Artificial Life VI)*, 151–160, Cambridge, MA. MIT Press.

Hammond, B.J. 1993. Quantitative study of the control of HIV-1 gene expression. *J. Theor. Biol.* 163, 199–221.

Hartwell, L.H., Hopfield, J.J., Leibler, S., and Murray, A.W. 1999. From molecular to modular cell biology. *Nature* 402(suppl.), C47–C52.

Hasty, J., McMillen, D., Isaacs, F., and Collins, J.J. 2001. Computational studies of gene regulatory networks: *In numero* molecular biology. *Nat. Rev. Genet.* 2, 268–279.

Hayes-Roth, F., Waterman, D.A., and Lenat, D.B. 1983. *Building Expert Systems.* Addison-Wesley, London.

Heckerman, D. 1998. A tutorial on learning with Bayesian networks. In M.I. Jordan, ed. *Learning in Graphical Models*, Kluwer, Dordrecht.

Heidtke, K.R., and Schulze-Kremer, S. 1998a. BioSim: A new qualitative simulation environment for molecular biology. In J. Glasgow, T. Littlejohn, and F. Major, eds. *Proc. 6th Int. Conf. Intell. Syst. Mol. Biol. (ISMB-98)*, 85–94, Menlo Park, CA. AAAI Press.

Heidtke, K.R., and Schulze-Kremer, S. 1998b. Design and implementation of a qualitative simulation model of λ phage infection. *Bioinformatics* 14(1), 81–91.

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

Hofestädt, R. 1993. A simulation shell to model metabolic pathways. *SAMS* 11, 253–262.

Hofestädt, R., and Meineke, F. 1995. Interactive modelling and simulation of biochemical networks. *Comput. Biol. Med.* 25(3), 321–334.

Hofestädt, R., and Thelen, S. 1998. Quantitative modeling of biochemical networks. *In Silico Biol.* 1(0006).

Holstege, F.C., Jennings, E.G., Wyrick, J.J., Lee, T.I., Hengartner, C.J., Green, M.R., Golub, T.R., Lander, E.S., and Young, R.A. 1998. Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* 95, 717–728.

Holter, N.S., Maritan, A., Cieplak, M., Fedoroff, N.V., and Banavar, J.R. 2001. Dynamic modeling of gene expression data. *Proc. Natl. Acad. Sci. USA* 98(4), 1693–1698.

Huang, S. 1999. Gene expression profiling, genetic networks, and cellular states: An integrating concept for tumorigenesis and drug discovery. *J. Mol. Med.* 77, 469–480.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, D., Lum, P.Y., Stepaniants, S.B., Shoemaker, D.D., Gachotte, D., Chakrabortty, K., Simon, J., Bard, M., and Friend, S.H. 2000. Functional discovery via a compendium of expression profiles. *Cell* 102(1), 109–126.

Hunding, A. 1989. Turing prepatterns of the second kind simulated on supercomputers in three curvilinear coordinates and time. In A. Goldbeter, ed. *Cell to Cell Signalling: From Experiments to Theoretical Models*, 229–236, Academic Press, London.

Hunding, A., Kauffman, S.A., and Goodwin, B.C. 1990. *Drosophila* segmentation: Supercomputer simulation of prepattern hierarchy. *J. Theor. Biol.* 145, 369–384.

Ideker, T.E., Thorsson, V., and Karp, R.M. 2000. Discovery of regulatory interactions through perturbation: Inference and experimental design. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB 2000)*, vol. 5, 302–313, Singapore, World Scientific Publishing.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. 2001. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409, 533–538.

Jackson, E.R., Johnson, D., and Nash, W.G. 1986. Gene networks in development. *J. Theor. Biol.* 119(4), 379–396.

Jacq, B., Horn, F., Janody, F., Gompel, N., Serralbo, O., Mohr, E., Leroy, C., Bellon, B., Fasano, L., Laurenti, P., and Röder, L. 1997. GIF-DB: A WWW database on gene interactions involved in *Drosophila melanogaster* development. *Nucl. Acids Res.* 25(1), 67–71.

Kahn, P. 1995. From genome to proteome: Looking at cell's proteins. *Science* 270, 369–370.

Kananyan, G.K., Ratner, V.A., and Tchuraev, R.N. 1981. Enlarged model of lambda phage ontogenesis. *J. Theor. Biol.* 90, 301–315.

Kanehisa, M. 2000. *Post-Genome Informatics*, Oxford University Press, Oxford.

Kanehisa, M., and Goto, S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucl. Acids Res.* 28(1), 27–30.

Karp, P.D. 1991. Hypothesis formation as design. In J. Shrager and P. Langley, eds. *Computational Models of Scientific Discovery and Theory Formation*, 275–317, Morgan Kaufmann, San Mateo, CA.

Karp, P.D. 1993a. Design methods for scientific hypothesis formation and their application to molecular biology. *Mach. Learn.* 12, 89–116.

Karp, P.D. 1993b. A qualitative biochemistry and its application to the regulation of the tryptophan operon. In L. Hunter, ed. *Artificial Intelligence and Molecular Biology*, 289–324, AAAI Press/MIT Press, Menlo Park, CA.

Karp, P.D., Riley, M., Paley, S.M., Pellegrini-Toole, A., and Krummenacker, M. 1999a. EcoCyc: Encyclopedia of *Escherichia coli* genes and metabolism. *Nucl. Acids Res.* 27(1), 55–58.

Karp, R.M., Stoughton, R., and Yeung, K.Y. 1999b. Algorithms for choosing differential gene expression experiments. In *Proc. 3rd Ann. Int. Conf. Computational Mol. Biol. (RECOMB'99)*, 208–217, New York, ACM Press.

Kauffman, S.A. 1969a. Homeostasis and differentiation in random genetic control networks. *Nature* 224, 177–178.

Kauffman, S.A. 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22, 437–467.

Kauffman, S.A. 1974. The large-scale structure and dynamics of gene control circuits: An ensemble approach. *J. Theor. Biol.* 44, 167–190.

Kauffman, S.A. 1977. Gene regulation networks: A theory for their global structure and behaviors. In *Current Topics in Developmental Biology*, vol. 6, 145–182, Academic Press, New York.

Kauffman, S.A. 1979. Assessing the probable regulatory structures and dynamics of the metazoan genome. In R. Thomas, ed. *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, vol. 29 of *Lecture Notes in Biomathematics*, 30–60, Springer-Verlag, Berlin.

Kauffman, S.A. 1991. Antichaos and adaptation. *Sci. Am.* 265, 78–84.

Kauffman, S.A. 1993. *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York.

Kauffman, S.A., Shymko, R.M., and Trabert, K. 1978. Control of sequential compartment formation in *Drosophila*. *Science* 199, 259–270.

Keller, A.D. 1994. Specifying epigenetic states with autoregulatory transcription factors. *J. Theor. Biol.* 170, 175–181.

Kitano, H., Hamahashi, S., Kitazawa, J., Takao, K., and Imai, S. 1997. The Virtual Biology Laboratories: A new approach to computational biology. In P. Husbands and I. Harvey, eds. *Proc. 4th Eur. Conf. Artif. Life (ECAL-97)*, 274–283, Cambridge, MA, MIT Press.

Ko, M.S.H. 1991. A stochastic model for gene induction. *J. Theor. Biol.* 153, 181–194.

Ko, M.S.H. 1992. Induction mechanism of a single gene molecule: Stochastic or deterministic? *BioEssays* 14(5), 341–346.

Koh, B.T., Tan, R.B.H., and Yap, M.G.S. 1998. Genetically structured mathematical modeling of *trp* attenuator mechanism. *Biotechnol. Bioeng.* 58, 502–509.

Kohn, K.W. 1999. Molecular interaction map of the mammalian cell cycle control and DNA repair systems. *Mol. Biol. Cell* 10, 2703–2734.

Kohn, K.W. 2001. Molecular interaction maps as information organizers and simulation guides. *Chaos* 11(1), 1–14.

Koile, K., and Overton, C.G. 1989. A qualitative model for gene expression. In *Proc. 1989 Summer Comp. Simul. Conf.*, 415–421, Society for Computer Simulation.

Kolpakov, F.A., Ananko, E.A., Kolesov, G.B., and Kolchanov, N.A. 1999. GeneNet: A gene network database and its automated visualization. *Bioinformatics* 14(6), 529–537.

Koshland, Jr., D.E. 1998. The era of pathway quantification. *Science* 280, 852–853.

Kosman, D., Reinitz, J., and Sharp, D.H. 1998. Automated assay of gene expression at cellular resolution. In *Proc. Pac. Symp. Biocomput. (PSB'98)*, vol. 3, 6–17, Singapore, World Scientific Publishing.

Kuipers, B. 1986. Qualitative simulation. *Artif. Intell.* 29, 289–388.

Kuipers, B. 1994. *Qualitative Reasoning: Modeling and Simulation with Incomplete Knowledge*. MIT Press, Cambridge, MA.

Kyoda, K.M., Muraki, M., and Kitano, H. 2000. Construction of a generalized simulator for multi-cellular organisms and its application to Smad signal transduction. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB 2000)*, vol. 5, 314–325, Singapore, World Scientific Publishing.

Lacalli, T.C. 1990. Modeling the *Drosophila* pair-rule pattern by reaction diffusion: Gap input and pattern control in a 4-morphogen system. *J. Theor. Biol.* 144, 171–194.

Lacalli, T.C., Wilkinson, D.A., and Harrison, L.G. 1988. Theoretical aspects of stripe formation in relation to *Drosophila* segmentation. *Development* 103, 105–113.

Lambert, J.D. 1991. *Numerical Methods for Ordinary Differential Equations*. Wiley, Chichester.

Landahl, H.D. 1969. Some conditions for sustained oscillations in biochemical chaines. *B. Math. Biophys.* 31, 775–787.

Lander, E.S. 1996. The new genomics: Global views of biology. *Science* 274, 536–539.

Laub, M.T., McAdams, H.H., Feldblyum, T., Fraser, C.M., and Shapiro, L. 2000. Global analysis of the genetic network controlling a bacterial cell cycle. *Science* 290, 2144–2148.

Leloup, J.-C., and Goldbeter, A. 1998. A model for the circadian rhythms in *Drosophila* incorporating the formation of a complex between the PER and TIM proteins. *J. Biol. Rhythms* 13(1), 70–87.

Lewin, B. 1999. *Genes VII*. Oxford University Press, Oxford.
Lewis, J.E., and Glass, L. 1991. Steady states, limit cycles, and chaos in models of complex biological networks. *Int. J. Bifurcat. Chaos* 1(2), 477–483.
Lewis, J.E., and Glass, L. 1992. Nonlinear dynamics and symbolic dynamics of neural networks. *Neural Comput.* 4, 621–642.
Liang, S., Fuhrman, S., and Somogyi, R. 1998. REVEAL: A general reverse engineering algorithm for inference of genetic network architectures. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'98)*, vol. 3, 18–29, Singapore, 1998, World Scientific Publishing.
Lipschutz, R.J., Fodor, S.P.A., Gingeras, T.R., and Lockhart, D.J. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* 21(suppl.), 20–24.
Ljung, L. 1999. *System Identification: Theory for the User*, 2nd ed., Prentice Hall, Upper Saddle River, NJ.
Lockhart, D.J., and Winzeler, E.A. 2000. Genomics, gene expression and DNA arrays. *Nature* 405, 827–836.
Loew, L.M., and Schaff, J.C. 2001. The virtual cell: A software environment for computational cell biology. *Trends Biotechnol.* 19(10), 401–406.
Loomis, W.F., and Sternberg, P.W. 1995. Genetic networks. *Science* 269, 649.
MacDonald, N. 1989. *Biological Delay Systems: Linear Stability Theory*, Cambridge University Press, Cambridge.
Maddox, J. 1992. Is molecular biology yet a science? *Nature* 355, 201.
Mahaffy, J.M. 1984. Cellular control models with linked positive and negative feedback and delays: I. The models. *J. Math. Biol.* 106, 89–102.
Maini, P.K., Painter, K.J., and Nguyen Phong Chau, H. 1997. Spatial pattern formation in chemical and biological systems. *J. Chem. Soc., Faraday Trans.* 93(20), 3601–3610.
Maki, Y., Tominaga, D., Okamoto, M., Watanabe, S., and Eguch, Y. 2001. Development of a system for the inference of large scale genetic networks. In R.B. Altman, A.K. Dunker, L. Hunter, K. Lauderdale, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'01)*, vol. 6, 446–458, Singapore, World Scientific Publishing.
Mann, M. 1999. Quantitative proteomics. *Nat. Biotechnol.* 17, 954–955.
Marnellos, G., Deblandre, G.A., Mjolsness, E., and Kintner, G. 2000. Delta-Notch lateral inhibitory patterning in the emergence of ciliated cells in *Xenopus*: Experimental observations and a gene network model. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB 2000)*, vol. 5, 326–337, Singapore, World Scientific Publishing.
Marnellos, G., and Mjolsness, E., 1998. A gene network approach to modeling early neurogenesis in *Drosophila*. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'98)*, vol. 3, 30–41, Singapore, World Scientific Publishing.
Matsuno, H., Doi, A., Nagasaki, M., and Miyano, S. 2000. Hybrid Petri net representation of gene regulatory network. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB 2000)*, vol. 5, 338–349, Singapore, World Scientific Publishing.
McAdams, H.M., and Arkin, A. 1997. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA* 94, 814–819.
McAdams, H.H., and Arkin, A. 1998. Simulation of prokaryotic genetic circuits. *Ann. Rev. Biophys. Biomol. Struct.* 27, 199–224.
McAdams, H.H., and Arkin, A. 1999. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends Genet.* 15(2), 65–69.
McAdams, H.H., and Shapiro, L. 1995. Circuit simulation of genetic networks. *Science* 269, 650–656.
Meinhardt, H. 1977. A model of pattern formation in insect embryogenesis. *J. Cell Sci.* 23, 117–139.
Meinhardt, H. 1982. *Models of Biological Pattern Formation*, Academic Press, London.
Meinhardt, H. 1986. Hierarchical inductions of cell states: A model for segmentation in *Drosophila*. *J. Cell Sci.* 4(suppl.), 357–381.
Meinhardt, H. 1988. Models for maternally supplied positional information and the activation of segmentation genes in *Drosophila* embryogenesis. *Development* 104(suppl.), 95–110.
Meinhardt, H. 2000. Pattern formation by local self-activation and lateral inhibition. *BioEssays* 22(8), 753–760.
Mendes, P. 1993. GEPASI: A software package for modelling the dynamics, steady states and control of biochemical and other systems. *CABIOS* 9(5), 563–571.
Mendoza, L., and Alvarez-Buylla, E.R. 1998. Dynamics of the genetic regulatory network for *Arabidopsis thaliana*. *J. Theor. Biol.* 193, 307–319.
Mendoza, L., Thieffry, D., and Alvarez-Buylla, E.R. 1999. Genetic control of flower morphogenesis in *Arabidopsis thaliana*: A logical analysis. *Bioinformatics* 15(7–8), 593–606.
Mestl, T., Bagley, R.J., and Glass, L. 1997. Common chaos in arbitrarily complex feedback networks. *Phys. Rev. Lett.* 79(4), 653–656.
Mestl, T., Lemay, C., and Glass, L. 1996. Chaos in high-dimensional neural and gene networks. *Physica D* 98, 33–52.

Mestl, T., Plahte, E., and Omholt, S.W. 1995a. A mathematical framework for describing and analysing gene regulatory networks. *J. Theor. Biol.* 176, 291–300.

Mestl, T., Plahte, E., and Omholt, S.W. 1995b. Periodic solutions in systems of piecewise-linear differential equations. *Dynam. Stabil. Syst.* 10(2), 179–193.

Meyers, S., and Friedland, P. 1984. Knowledge-based simulation of genetic regulation in bacteriophage lambda. *Nucleic Acids Res.* 12(1), 1–9.

Michaels, G.S., Carr, D.B., Askenazi, M., Fuhrman, S., Wen, X., and Somogyi, R. 1998. Cluster analysis and data visualization of large-scale gene expression data. In *Proc. Pac. Symp. Biocomput. (PSB'98)*, vol. 3, 42–53, Singapore, World Scientific Publishing.

Mjolsness, E., Sharp, D.H., and Reinitz, J. 1991. A connectionist model of development. *J. Theor. Biol.* 152, 429–453.

Monod, J., and Jacob, F. 1961. General conclusions: Teleonomic mechanisms in cellular metabolism, growth, and differentiation. In *Cold Spring Harbor Symp. Quant. Biol.*, vol. 26, 389–401, Cold Spring Harbor, NY.

Morton-Firth, C.J., and Bray, D. 1998. Predicting temporal fluctuations in an intracellular signalling pathway. *J. Theor. Biol.* 192, 117–128.

Morton-Firth, C.J., Shimizu, T.S., and Bray, D. 1990. A free-energy-based stochastic simulation of the Tar receptor complex. *J. Mol. Biol.* 286, 1059–1074.

Murphy, K., and Mian, S. 1999. Modelling gene expression data using dynamic Bayesian networks. Technical report, University of California at Berkeley, Berkeley, CA. Unpublished report.

Myasnikova, E., Samsonova, A., Kozlov, K., Samsonova, M., and Reinitz, J. 2001. Registration of the expression patterns of *Drosophila* segmentation genes by two independent methods. *Bioinformatics* 17(1), 3–12.

Nagorcka, B.N. 1988. A pattern formation mechanism to control spatial organization in the embryo of *Drosophila melanogaster*. *J. Theor. Biol.* 132, 277–306.

Nayak, P.P. 1995. *Automated Modeling of Physical Systems*, Springer-Verlag, Berlin.

Newman, S.A., and Rice, S.A. 1971. Model for constraint and control in biochemical networks. *Proc. Natl Acad. Sci. USA* 68(1), 92–96.

Nicolis, G., and Prigogine, I. 1977. *Self-Organization in Nonequilibrium Systems: From Dissipative Structures to Order Through Fluctuations*, Wiley-Interscience, New York.

Noda, K., Shinohara, A., Takeda, M., Matsumoto, S., Miyano, S., and Kuhara, S. 1988. Finding genetic network from experiments by weighted network model. In *Genome Informatics*, vol. 9, 141–150.

Novak, B., and Tyson, J.J. 1993a. Modeling the cell division cycle: M-phase trigger, oscillations, and size control. *J. Theor. Biol.* 165, 101–134.

Novak, B., Csikasz-Nagy, A., Gyorffy, B., Chen, K.C., and Tyson, J.J. 1998. Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. *Biophys. Chem.* 72, 185–200.

Novak, B., and Tyson, J.J. 1993b. Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. *J. Cell Sci.* 106, 1153–1168.

Novak, B., and Tyson, J.J. 1995. Quantitative analysis of a molecular model of mitotic control in fission yeast. *J. Theor. Biol.* 173, 283–305.

Nowak, R. 1995. Entering the postgenome era. *Science* 270, 368–371.

Obeyesekere, M.N., Tucker, S., and Zimmerman, S.O. 1992. Mathematical models for the cellular concentrations of Cyclin and MPF. *Biochem. Biophys. Res. Co.* 184(2), 782–789.

Omholt, S.W., Kefang, X., Andersen, Ø., and Plahte, E. 1998. Description and analysis of switchlike regulatory networks exemplified by a model of cellular iron homeostasis. *J. Theor. Biol.* 195, 339–350.

Othmer, H.G. 1976. The qualitative dynamics of a class of biochemical control circuits. *J. Math. Biol.* 3, 53–78.

Palsson, B.O. 1997. What lies beyond bioinformatics? *Nat. Biotechnol.* 15, 3–4.

Pandey, A., and Mann, M. 2000. Proteomics to study genes and genomes. *Nature* 405, 837–846.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, San Francisco, CA.

Pe'er, D., Regev, A., Elidan, G., and Friedman, N. 2001. Inferring subnetworks from perturbed expression profiles. *Bioinformatics* 17(S1), S215–S224.

Plahte, E., Mestl, T., and Omholt, S.W. 1994. Global analysis of steady points for systems of differential equations with sigmoid interactions. *Dynam. Stabil. Syst.* 9(4), 275–291.

Plahte, E., Mestl, T., and Omholt, S.W. 1995a. Feedback loops, stability and multistationarity in dynamical systems. *J. Biol. Syst.* 3(2), 409–413.

Plahte, E., Mestl, T., and Omholt, S.W. 1995b. Stationary states in food web models with threshold relationships. *J. Biol. Syst.* 3(2), 569–577.

Plahte, E., Mestl, T., and Omholt, S.W. 1998. A methodological basis for description and analysis of systems with complex switch-like interactions. *J. Math. Biol.* 36, 321–348.

Platzer, U., Gumbel, M., and Meinzer, H.-P. 2001. Simulation of genetic networks controlling cell differentiation. Poster at ISMB Satellite Meeting on Computer Modeling of Cellular Processes, SIGSIM01.

Prokudina, E.I., Valeev, R.Y., and Tchuraev, R.N. 1991. A new method for the analysis of the dynamics of the molecular genetic control systems. II. Application of the method of generalized threshold models in the investigation of concrete genetic systems. *J. Theor. Biol.* 151, 89–110.

Ratner, V.A., and Tchuraev, R.N. 1978. Simplest genetic systems controlling ontogenesis: Organization principles and models of their function. In R. Rosen, ed. *Prog. Theor. Biol.*, vol. 5, 81–127, Academic Press, New York.

Reddy, V.N., Liebman, M.N., and Mavrovouniotis, M.L. 1996. Qualitative analysis of biochemical reaction systems. *Comput. Biol. Med.* 26(1), 9–24.

Regev, A., Silverman, W., and Shapiro, E. 2001. Representation and simulation of biochemical processes using the π-calculus process algebra. In R.B. Altman, A.K. Dunker, L. Hunter, K. Lauderdale, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'01)*, vol. 6, 459–470, Singapore, World Scientific Publishing.

Reinitz, J., Kosman, D., Vanario-Alonso, C.E., and Sharp, D.H. 1998. Stripe forming architecture of the gap gene system. *Dev. Genet.* 23, 11–27.

Reinitz, J., Mjolsness, E., and Sharp, D.H. 1995. Model for cooperative control of positional information in *Drosophila* by bicoid and maternal hunchback. *J. Exp. Zool.* 271, 47–56.

Reinitz, J., and Sharp, D.H. 1995. Mechanism of *eve* stripe formation. *Mech. Develop.* 49, 133–158.

Reinitz, J., and Sharp, D.H. 1996. Gene circuits and their uses. In J. Collado-Vides, B. Magasanik, and T.F. Smith, eds. *Integrative Approaches to Molecular Biology*, 253–272. MIT Press, Cambridge, MA.

Reinitz, J., and Vaisnys, J.R. 1990. Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of co-operativity. *J. Theor. Biol.* 145, 295–318.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., Volkert, T.L., Wilson, C.J., Bell, S.P., and Young, R.A. 2000. Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309.

Rigney, D.R. 1979. Stochastic models of cellular variability. In R. Thomas, ed. *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, vol. 29 of *Lecture Notes in Biomathematics*, 237–280. Springer-Verlag, Berlin.

Rosen, R. 1968. Recent developments in the theory of control and regulation of cellular processes. *Int. Rev. Cytol.* 23, 25–88.

Ruoff, P., Vinsjevik, M., Monnerjahn, C., and Rensing, L. 2001. The Goodwin model: Simulating the effect of light pulses on the circadian sporulation rhythm of *Neurospora crassa*. *J. Theor. Biol.* 209, 29–42.

Russell, M.A. 1985. Positional information in insect segments. *Dev. Biol.* 108, 269–283.

Salgado, H., Santos, A., Garza-Ramos, U., van Helden, J., Díaz, E., and Collado-Vides, J. 2000. RegulonDB (version 2.0): A database on transcriptional regulation in *Escherichia coli*. *Nucl. Acids Res.* 27(1), 59–60.

Samsonova, M.G., Savostyanova, E.G., Serov, V.N., Spirov, A.V., and Reinitz, J. 1998. GeNet: A database of genetic networks. In *Proc. First Int. Conf. Bioinformatics Genome Regul. Struct., BGRS'98*, Novosibirsk, ICG.

Samsonova, M., and Serov, V.N. 1999. NetWork: An interactive interface to the tools for analysis of genetic network structure and dynamics. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'99)*, vol. 4, 102–111, Singapore, World Scientific Publishing.

Sanchez, C., Lachaize, C., Janody, F., Bellon, B., Röder, L., Euzenat, J., Rechenmann, F., and Jacq, B. 1999. Grasping at molecular interactions and genetic networks in *Drosophila melanogaster*. *Nucleic Acids Res.* 27(1), 89–94.

Sánchez, L., and Thieffry, D. 2001. A logical analysis of the *Drosophila* gap genes. *J. Theor. Biol.* 211, 115–141.

Sánchez, L., van Helden, J., and Thieffry, D. 1997. Establishment of the dorso-ventral pattern during embryonic development of *Drosophila melanogaster*: A logical analysis. *J. Theor. Biol.* 189, 377–389.

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

Santillán, M., and Mackey, M.C. 2001. Dynamic regulation of the tryptophan operon: A modeling study and comparison with experimental data. *Proc. Natl. Acad. Sci. USA* 98(4), 1364–1369.

Sauro, H.M. 1993. SCAMP: A general-purpose simulator and metabolic control analysis program. *CABIOS* 9(4), 441–450.

Savageau, M.A. 1969. Biochemical systems analysis. II The steady state solutions for an n-pool system using a power-law approximation. *J. Theor. Biol.* 25, 370–379.

Savageau, M.A. 1971. Parameter sensitivity as a criterion for evaluating and comparing the performance of biochemical systems. *Nature* 229, 542–544.

Savageau, M.A. 1989. Are there rules governing patterns of gene regulation? In B.C. Goodwin and P.T. Saunders, eds. *Theoretical Biology: Epigenetic and Evolutionary Order*, 42–66. Edinburgh University Press, Edinburgh.

Savageau, M.A. 1996a. A kinetic formalism for integrative molecular biology: Manifestation in biochemical systems theory and use in elucidating design principles for gene circuits. In J. Collado-Vides, B. Magasanik, and T.F. Smith, eds. *Integrative Approaches to Molecular Biology*, 115–146, MIT Press, Cambridge, MA.

Savageau, M.A. 1996b. Power-law formalism: A canonical nonlinear approach to modeling and analysis. In V. Lakshmikantham, ed. *Proc. 1st World Congr. Nonlinear Anal.*, vol. 4, 3323–3334, Berlin, De Gruyter.

Schilling, C.H., Letscher, D., and Palsson, B.O. 2000. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J. Theor. Biol.* 203, 229–248.

Schilling, C.H., and Palsson, B.O. 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* rd through a genome-scale pathway analysis. *J. Theor. Biol.* 203, 249–283.

Segel, L.A. 1984. *Modeling Dynamic Phenomena in Molecular and Cellular Biology*, Cambridge University Press, Cambridge.

Sharp, D.H., and Reinitz, J. 1998. Prediction of mutant expression patterns using gene circuits. *BioSystems* 47, 79–90.

Shea, M.A., and Ackers, G.K. 1985. The O_R control system of bacteriophage lambda: A physical-chemical model for gene regulation. *J. Mol. Biol.* 181, 211–230.

Shimada, T., Hagiya, M., Arita, M., Nishizaki, S., and Tan, C.L. 1995. Knowledge-based simulation of regulatory action in lambda phage. *Int. J. Artif. Intell. Tools* 4(4), 511–524.

Smolen, P., Baxter, D.A., and Byrne, J.H. 1999. Effects of macromolecular transport and stochastic fluctuations on dynamics of genetic regulatory systems. *Am. J. Phys.* 277, C777–C790.

Smolen, P., Baxter, D.A., and Byrne, J.H. 2000. Modeling transcriptional control in gene networks: Methods, recent results, and future directions. *Bull. Math. Biol.* 62, 247–292.

Snoussi, E.H. 1989. Qualitative dynamics of piecewise-linear differential equations: A discrete mapping approach. *Dynam. Stabil. Syst.* 4(3–4), 189–207.

Snoussi, E.H. 1998. Necessary conditions for multistationarity and stable periodicity. *J. Biol. Syst.* 6(1), 3–9.

Snoussi, E.H., and Thomas, R. 1993. Logical identification of all steady states: The concept of feedback loop characteristic states. *Bull. Math. Biol.* 55(5), 973–991.

Somogyi, R., and Sniegoski, C.A. 1996. Modeling the complexity of genetic networks: Understanding multigenic and pleiotropic regulation. *Complexity* 1(6), 45–63.

Spellman, P., Sherlock, G., Zhang, M., Iyer, V., Anders, K., Eisen, M., Brown, P., Botstein, D., and Futcher, B. 1998. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* 9, 3273–3297.

Strogatz, S.H. 1994. *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering*, Perseus Books, Reading, MA.

Strohman, R.C. 1997. The coming Kuhnian revolution in biology. *Nat. Biotechnol.* 15, 194–200.

Sugita, M. 1961. Functional analysis of chemical systems *in vivo* using a logical circuit equivalent. *J. Theor. Biol.* 1, 179–192.

Sugita, M. 1963. Functional analysis of chemical systems *in vivo* using a logical circuit equivalent: II. The idea of a molecular automaton. *J. Theor. Biol.* 4, 179–192.

Szallasi, Z. 1999. Genetic network analysis in light of massively parallel biological data acquisition. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'99)*, vol. 4, 5–16, Singapore, World Scientific Publishing.

Szallasi, Z., and Liang, S. 1998. Modeling the normal and neoplastic cell cycle with 'realistic Boolean genetic networks': Their application for understanding carcinogenesis and assessing therapeutic strategies. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'98)*, vol. 3, 66–76, Singapore, World Scientific Publishing.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S., and Golub, T.R. 1999. Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* 96, 2907–2912.

Tavazoie, S., Hughes, J.D., Campbell, M.J., Cho, R.J., and Church, G.M. 1999. Systematic determination of genetic network architecture. *Nat. Genet.* 22, 281–285.

Tchuraev, R.N. 1991. A new method for the analysis of the dynamics of the molecular genetic control systems: I. Description of the method of generalized threshold models. *J. Theor. Biol.* 151, 71–87.

Tchuraev, R.N., and Ratner, V.A. 1983. A continuous approach with threshold characteristics for simulation of gene expression. In K. Bellmann, ed. *Molecular Genetic Information Systems: Modelling and Simulation*, 64–80. Akademie-Verlag, Berlin.

Thieffry, D. 1999. From global expression data to gene networks. *BioEssays* 21(11), 895–899.

Thieffry, D., Colet, M., and Thomas, R. 1993. Formalisation of regulatory nets: A logical method and its automatization. *Math. Modelling Sci. Computing* 2, 144–151.

Thieffry, D., Huerta, A.M., Pérez-Rueda, E., and Colladio-Vides, J. 1998. From specific gene regulation to genomic networks: A global analysis of transcriptional regulation in *Escherichia coli*. *BioEssays* 20, 433–440.

Thieffry, D., and Romero, D. 1999. The modularity of biological regulatory networks. *BioSystems* 50, 49–59.

Thieffry, D., Snoussi, E.H., Richelle, J., and Thomas, R. 1995. Positive loops and differentiation. *J. Biol. Syst.* 3(2), 457–466.

Thieffry, D., and Thomas, R. 1995. Dynamical behaviour of biological networks: II. Immunity control in bacteriophage lambda. *Bull. Math. Biol.* 57(2), 277–297.

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

Thomas, R., ed. 1979. *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, vol. 29 of *Lecture Notes in Biomathematics*, Springer-Verlag, Berlin.

Thomas, R. 1981. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. In J. Della Dora, J. Demongeot, and B. Lacolle, eds. *Numerical Methods in the Study of Critical Phenomena*, 180–193, Springer-Verlag, Berlin.

Thomas, R. 1983. Logical vs. continuous description of systems comprising feedback loops: The relation between time delays and parameters. *Stud. Phys. Theor. Chem.* 28, 307–321.

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

Thomas, R. 1998. Laws for the dynamics of regulatory networks. *Int. J. Dev. Biol.* 42, 479–485.

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

Thomas, R., Gathoye, A.-M., and Lambert, L. 1976. A complex control circuit: Regulation of immunity in temperate bacteriophages. *Eur. J. Biochem.* 71, 211–227.

Thomas, R., and Kaufman, M. 2001. Multistationarity, the basis of cell differentiation and memory: I. Structural conditions of multistationarity and other nontrivial behavior. *Chaos* 11(1), 170–179.

Thomas, R., Thieffry, D., and Kaufman, M. 1995. Dynamical behaviour of biological regulatory networks: I. Biological rool of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.* 57(2), 247–276.

Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., and Hutchison, C.A. 1999. E-CELL: Software environment for whole-cell simulation. *Bioinformatics* 15(1), 72–84.

Tomita, M. 2001. Whole-cell simulation: A grand challenge of the 21st century. *Trends Biotechnol.* 19(6), 205–210.

Trelease, R.B., Henderson, R.A., and Park, J.B. 1999. A qualitative process system for modeling NF-κB and AP-1 gene regulation in immune cell biology research. *Artif. Intell. Med.* 17, 303–321.

Tucker, C.L., Gera, J.F., and Uetz, P. 2001. Towards an understanding of complex protein networks. *Trends in Cell Biology* 11(3), 102–106.

Turing, A.M. 1951. The chemical basis of morphogenesis. *Philos. T. Roy. Soc. B.* 237, 37–72.

Tyson, J.J. 1975. On the existence of oscillatory solutions in negative feedback cellular control processes. *J. Math. Biol.* 1, 311–315.

Tyson, J.J. 1999. Models of cell cycle control in eukaryotes. *J. Biotechnol.* 71(1–3), 239–244.

Tyson, J.J., and Othmer, H.G. 1978. The dynamics of feedback control circuits in biochemical pathways. *Prog. Theor. Biol.* 5, 1–62.

Tyson, J.J., Novak, B., Odell, G.M., Chen, K., and Thron, C.D. 1996. Chemical kinetic theory: Understanding cell-cycle regulation. *Trends Biochem. Sci.* 21(3), 89–96.

Ueda, H.R., Hagiwara, M., and Kitano, H. 2001. Robust oscillations within the interlocked feedback model of *Drosophila* circadian rhythm. *J. Theor. Biol.* 210, 401–406.

van Ham, P. 1979. How to deal with more than two levels. In R. Thomas, ed. *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, vol. 29 of *Lecture Notes in Biomathematics*, 326–343, Springer-Verlag, Berlin.

van Helden, J., Naim, A., Mancuso, R., Eldridge, M., Wernisch, L., Gilbert, D., and Wodak, S.J. 2000. Representing and analysing molecular and cellular function using the computer. *Biol. Chem.* 381, 921–935.

van Kampen, N.G. 1997. *Stochastic Processes in Physics and Chemistry*, rev. ed., Elsevier, Amsterdam.

van Someren, E.P., Wessels, L.F.A., and Reinders, M.J.T. 2000. Linear modeling of genetic networks from experimental data. In R. Altman, T.L. Bailey, P. Bourne, M. Gribskov, T. Lengauer, I.N. Shindyalov, L.F. Ten Eyck, and H. Weissig, eds. *Proc. 8th Int. Conf. Intell. Syst. Mol. Biol. (ISMB 2000)*, 355–366, Menlo Park, CA. AAAI Press.

Voit, E.O. 1991. *Canonical Nonlinear Modeling: S-System Approach to Understanding Complexity*, Van Nostrand Reinhold, New York.

Voit, E.O. 2000. *Computational Analysis of Biochemical Systems: A Practical Guide for Biochemists and Molecular Biologists*, Cambridge University Press, Cambridge.

von Dassow, G., Meir, E., Munro, E.M., and Odell, G.M. 2000. The segment polarity network is a robust developmental module. *Nature* 406, 188–192.

Walter, C. 1970. The occurrence and the significance of limit cycle behavior in controlled biochemical systems. *J. Theor. Biol.* 27, 259–272.

Walter, C., Parker, R., and Yčas, M. 1967. A model for binary logic in biochemical systems. *J. Theor. Biol.* 15, 208–217.

Weaver, D.C., Workman, C.T., and Stormo, G.D. 1999. Modeling regulatory networks with weight matrices. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'99)*, vol. 4, 112–123, Singapore, World Scientific Publishing.

Weisbuch, G. 1986. Networks of automata and biological organization. *J. Theor. Biol.* 121, 255–267.

Wen, X., Fuhrman, S., Michaels, G.S., Carr, D.B., Smith, S., Barker, J.L., and Somogyi, R. 1998. Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl. Acad. Sci. USA* 95, 334–339.

Wessels, L.F.A., van Someren, E.P., and Reinders, M.J.T. 2001. A comparison of genetic network models. In R.B. Altman, A.K. Dunker, L. Hunter, K. Lauderdale, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'01)*, vol. 6, 508–519, Singapore, World Scientific Publishing.

Wolf, D.M., and Eeckman, F.H. 1998. On the relationship between genomic regulatory element organization and gene regulatory dynamics. *J. Theor. Biol.* 195, 167–186.

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

Wuensche, A. 1998. Genomic regulation modeled as a network with basins of attraction. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'98)*, vol. 3, 89–102, Singapore, World Scientific Publishing.

Xiu, Z.-L., Zeng, A.-P., and Deckwer, W.-D. 1997. Model analysis concerning the effects of growth rate and intracellular tryptophan level on the stability and dynamics of tryptophan biosynthesis in bacteria. *J. Biotechnol.* 58, 125–140.

Yagil, G. 1975. Quantitative aspects of protein induction. In B.L. Horecker and E.R. Stadtman, eds. *Current Topics in Cell Regulation*, 183–237, Academic Press, New York.

Yagil, G., and Yagil, E. 1971. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.* 11, 11–27.

You, L., and Yin, J. 2001. Simulating the growth of viruses. In R.B. Altman, A.K. Dunker, L. Hunter, K. Lauderdale, and T.E. Klein, eds. *Proc. Pac. Symp. Biocomput. (PSB'01)*, vol. 6, 532–543, Singapore, World Scientific Publishing.

Yuh, C.-H., Bolouri, H., and Davidson, E.H. 1998. Genomic *cis*-regulatory logic: Experimental and computational analysis of a sea urchin gene. *Science* 279, 1896–1902.

Zhu, H., and Snyder, M. 2001. Protein arrays and microarrays. *Curr. Opin. Chem. Biol.* 5, 40–45.

Zien, A., Küffner, R., Zimmer, R., and Lengauer, T. 2000. Analysis of gene expression data with pathway scores. In R. Altman, T.L. Bailey, P. Bourne, M. Gribskov, T. Lengauer, I.N. Shindyalov, L.F. Ten Eyck, and H. Weissig, eds. *Proc. 8th Int. Conf. Intell. Syst. Mol. Biol. (ISMB 2000)*, 407–417, Menlo Park, CA, AAAI Press.

Address correspondence to:
Hidde de Jong
Institut National de Recherche en Informatique et en Automatique (INRIA)
Unité de Recherche Rhône-Alpes
655 avenue de l'Europe
Montbonnot, 38334 Saint Ismier CEDEX, France

E-mail: Hidde.de-Jong@inrialpes.fr