Modeling and Simulation of Genetic Regulatory Systems: A Literature Review

Hidde De Jong

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Modeling and Simulation of Genetic Regulatory Systems: A Literature Review

Hidde de Jong

N° 4032

September 2000

THÈME 3

ISSN 0249-6399 ISRN INRIA/RR--4032--FR+ENG

INRIA
RHÔNE-ALPES

Modeling and Simulation of Genetic Regulatory Systems: A Literature Review

Hidde de Jong

Thème 3 — Interaction homme-machine,
images, données, connaissances
Projet HELIX

Rapport de recherche n° 4032 — September 2000 — 43 pages

Abstract: The spatiotemporal expression of genes in an organism is determined by regulatory systems that involve a large number of genes connected through a complex network of interactions. As an intuitive understanding of the behavior of these systems is hard to obtain, computer tools for the modeling and simulation of genetic regulatory networks will be indispensable. This report reviews formalisms that have been employed in mathematical biology and bioinformatics to describe genetic regulatory systems, in particular directed graphs, Bayesian networks, ordinary and partial differential equations, stochastic equations, Boolean networks and their generalizations, qualitative differential equations, and rule-based formalisms. In addition, the report discusses how these formalisms have been used in the modeling and simulation of regulatory systems.

Key-words: genetic regulatory networks, mathematical modeling, simulation, bioinformatics, mathematical biology

This work was funded in part by the CNRS in the framework of its 'Génome' research programme (project 'Bioinformatique des interactions moléculaires et des réseaux régulateurs à l'échelle des génomes', coordinator B. Jacq).

Ce travail a été partiellement financé par le CNRS dans le cadre de son programme 'Génome' (projet 'Bioinformatique des interactions moléculaires et des réseaux régulateurs à l'échelle des génomes', coordonnateur B. Jacq).

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 Montbonnot-St-Martin (France)
Téléphone : +33 4 76 61 52 00 — Télécopie +33 4 76 61 52 52

Modélisation et simulation des systèmes régulateurs géniques: Une revue de la littérature

Résumé : L'expression spatio-temporelle de gènes dans un organisme est déterminée par des systèmes régulateurs impliquant de nombreux gènes reliés par un réseau complexe d'interactions. Comme une compréhension intuitive du comportement du système est difficile à obtenir, des outils informatiques pour la modélisation et la simulation des réseaux régulateurs géniques sont indispensables. Ce rapport présente une revue des formalismes qui ont été employés en biologie mathématique et en bioinformatique pour décrire des systèmes régulateurs géniques, en particulier des graphes orientés, des réseaux Bayesiens, des équations différentielles ordinaires et aux dérivées partielles, des équations stochastiques, des réseaux booléens et leurs généralisations, des équations différentielles qualitatives, et des formalismes basés sur des règles de production. En outre, le rapport discute comment ces formalismes sont utilisés pour la modélisation et la simulation des systèmes régulateurs.

Mots-clés : réseaux régulateurs géniques, modélisation mathématique, simulation, bioinformatique, biologique mathématique

1 Introduction

The genome plays a central role in the control of cellular processes, like the response of cells to environmental signals, the differentiation of cells and groups of cells in the unfolding of developmental programs, and the replication of the genome preceding cell division. A protein synthesized from the information contained in a coding region of DNA may function as a transcription factor binding to regulatory sites elsewhere on the DNA, as an enzyme catalyzing a metabolic reaction, or as a component of a signal transduction pathway. 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 encoding 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 [114]. Apart from DNA transcription regulation, the best-studied form of regulation, the expression of a gene may be controlled during RNA processing and transport (in eukaryotes), RNA translation, and the post-translational modification of proteins (figure 1). The proteins fulfilling these 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 other molecules. An example of a simple *regulatory network*, involving three genes that code for proteins inhibiting the expression of other genes, is shown in figure 2. Repressors B and C bind to different regulatory sites of gene *a*, while A and D interact to form a heterodimer that binds to a regulatory site of gene *b*.[1] This prevents RNA polymerase from transcribing the genes downstream.

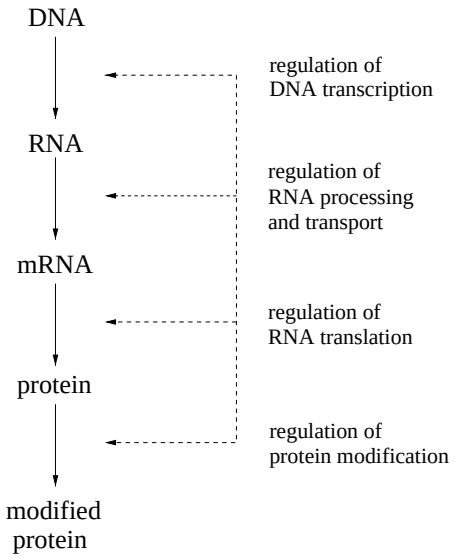


Figure 1: Regulation of gene expression at different stages of protein synthesis.

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, currently contains information on the structure and function of about 110,000 genes for 29 species [86]. In some cases, the proteins involved in the regulation of the expression of these genes have been identified, as well as the molecular mechanisms through which they achieve this. Much less is known, however, about the functioning of the regulatory systems of which the individual genes and

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

interactions form the components [18, 29, 40, 85, 113, 120, 154, 196, 204]. 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.

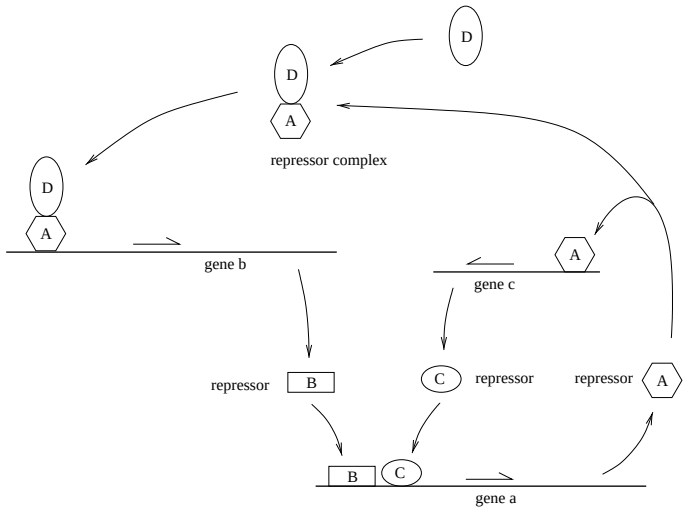


Figure 2: Example of a genetic regulatory system, consisting of a network of three genes a, b, and c encoding repressors A, B, and C, respectively. The arrows $\rightharpoonup$ denote the direction of transcription of the genes. The repressor proteins inhibit the transcription of the genes that they regulate.

Experimental techniques to dissect regulatory interactions on the molecular level are critical to this end. In fact, the study of genetic regulatory systems has received a major impetus from the recent development of techniques like cDNA microarrays and oligonucleotide chips, which permit the spatiotemporal expression levels of genes to be rapidly measured in a massively parallel way (for an overview see [17, 118] and other papers in that issue). 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, computer tools will be indispensable. As most genetic regulatory systems of interest involve many genes connected through interlocking feedback loops, an intuitive understanding of their behavior is hard to obtain. By explicating hypotheses on the topology of a regulatory network in the form of a computer model, the behavior of possibly large and complex regulatory systems can be predicted and explained in a systematic way.

Figure 3 shows the combined application of experimental and computational tools. Starting from an initial model, suggested by knowledge on regulatory mechanisms and expression data, the behavior of the system is simulated at a variety of experimental conditions. Comparing the predictions with the observed temporal evolution of gene expression levels 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 modeling a 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 figure 3 is provided by methods developed in mathematical biology and bioinformatics. Since the 1960s, with some notable precursors in the two preceding decades, a variety of mathematical formalisms to describe regulatory networks have been proposed. These formalisms are complemented by *simulation* techniques to make behavioral predictions from the model of a 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

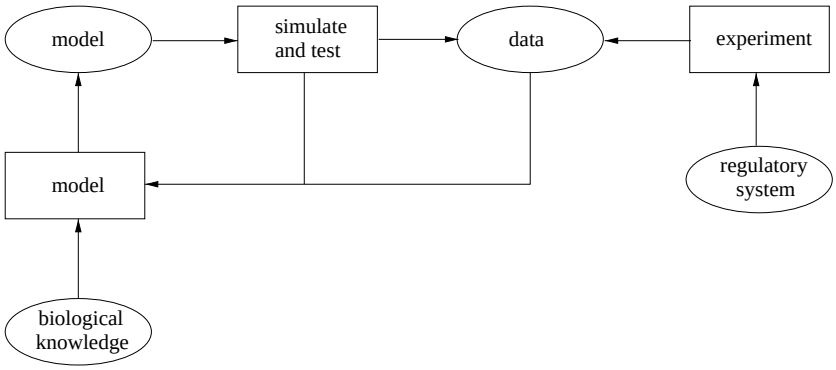


Figure 3: Analysis of a genetic regulatory system. The boxes represent activities, the ovals information sources, and the arrows information flows.

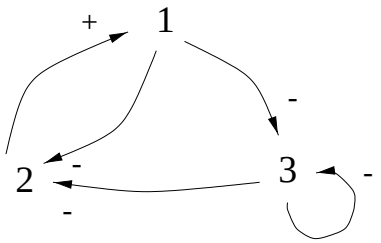
from the experimental literature. With more experimental data becoming available, easily accessible through databases and knowledge bases, modeling techniques are currently gaining popularity.

This report gives an overview of formalisms to describe genetic regulatory networks, and discusses their use in the modeling and simulation of regulatory systems. While preparing it, two other reviews on the modeling and simulation of regulatory systems appeared: one by MacAdams and Arkin [126], and one by Smolen, Baxter, and Byrne [189] (see [53, 174, 210] for earlier reviews). The present report differs from these reviews in that it focusses on the mathematical methods, evaluating their relative strengths and weaknesses, rather than on the biological results obtained through their application.

Sections 2 to 11 of this report give an overview of formalisms proposed in the literature and discusses modeling and simulation techniques appropriate for each of the formalisms. Formalisms to be discussed include directed graphs, Bayesian networks, ordinary and partial differential equations, stochastic equations, Boolean networks and their generalizations, qualitative differential 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 two omissions, no attention is paid to Petri nets [65, 78, 125, 166] and transformational grammars [26, 27, 28]. The impact of these approaches on the dynamic analysis of genetic regulatory networks has been limited so far. Moreover, they are related to some of the other formalisms discussed below.

2 Directed and undirected graphs

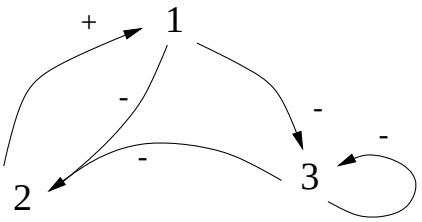
Probably the most straightforward way to model a genetic regulatory network is to view it as a *directed graph*. A directed graph G is defined as a tuple $\langle V, E\rangle$, with V a set of vertices and E a set of edges. A directed edge is a tuple $\langle i, j\rangle$ of vertices, where i denotes the head and j the tail of the edge. The vertices of a directed graph correspond to the genes or other elements of interest in the regulatory system, while the edges denote interactions among the genes. The graph representation of regulatory networks 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 dimer (figure 2). 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 4 shows simple regulatory networks of three genes.



(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)

Figure 4: (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.

Current databases and knowledge bases containing information about regulatory interactions in genetic regulation systems can be viewed, to a large extent, as richly annotated graph representations. Examples of such databases and knowledge bases are EcoCyc [90], GeneNet [102], GeNet [178], GIF-DB [82], KEGG [86], KNIFE [39], and RegulonDB [176]. 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 enhanced with applications offering facilities to compose the network by selecting and manipulating individual interactions. At the very least, they permit the user to visualize complete or partial networks, often on different levels of detail, and to navigate through the networks [177, 179]. Some current databases and knowledge bases integrate metabolic and signal transduction pathways with gene regulation processes (e.g., [86, 90]).

A number of operations on graphs can be carried out to make biologically relevant predictions about a regulatory system. A search for paths between two genes, for instance, may reveal missing regulatory interactions or provide clues about redundancy in the network. Further, cycles in the network point at feedback relations that are important in homeostasis and differentiation (section 4). 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 the comparison of regulatory networks of different organisms it might be possible to establish to which extent parts of regulation networks have been evolutionary conserved.

The graph models to which the above operations are applied can be composed from information stored in databases and knowledge bases, but they may also be induced from gene expression data. A variety of *clustering* algorithms have been used to group together similar temporal expression patterns and thus reveal clusters of genes that seem to be co-regulated in experiments [13, 25, 36, 141, 194, 235] (see [43] for a review of clustering techniques). By itself, clustering does not provide much information on regulatory connections in a network, as it is not clear which of two genes in a cluster regulates

the other, or whether they have a common regulator. If the time resolution of the expression data time-series is sufficiently high, additional analyses may permit one to extract more information on the structure of a regulatory network. This *inductive* or *reverse engineering* approach towards model construction will be looked at in some detail in the next section, in the context of a probabilistic extension of the graph models discussed here (see [2, 5, 7, 23, 81] for other approaches).

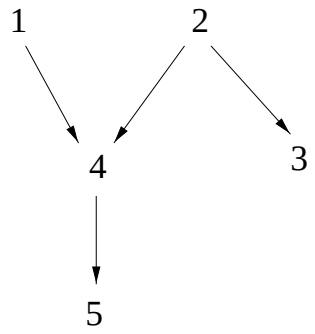
3 Bayesian networks

In the formalism of *Bayesian networks* [42, 159], the structure of a genetic regulatory system is modeled by a directed acyclic graph $G = \langle V, E \rangle$ (figure 5). The vertices $i \in V$, $1 \leq i \leq n$, represent genes or other state indicators 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 $p(X_i \mid \text{ancestors}(X_i))$ is defined, where $\text{ancestors}(X_i)$ denotes the variables corresponding to the direct regulators of i in G. The graph G and the conditional distributions $p(X_i \mid \text{ancestors}(X_i))$, 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, $i(X_i; \text{non-descendants}(X_i) \mid \text{ancestors}(X_i))$. By means of the Markov assumption, the joint probability distribution can be decomposed into

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

The graph implies additional conditional independencies, as shown in figure 5 for the example. 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 [42] for details and references).



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

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

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

Figure 5: Example Bayesian network consisting of a graph, conditional probability distributions for the random variables, the joint probability distribution, and conditional independencies [42].

Given a set of expression data D, in the form of a set of independent values for $\boldsymbol{X}$, learning techniques for Bayesian networks [71] allow one to induce the network, or rather the equivalence class of networks, that best matches D. Basically, the techniques relie 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 is not guaranteed to lead to a globally optimal solution. As an additional problem, currently available

expression data underdetermines the network, as only a few dozen of experiments provide information on the transcription level of thousands of genes.

Friedman and colleagues [42] 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 exist 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 an ancestor 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.

Markov relations and order relations have been studied in an application of the algorithm to the cell cycle data set of Spellman and colleagues [194]. This data set contains 76 measurements of the mRNA expression level of 6177 *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 do 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 may be overcome through generalizations like *dynamical Bayesian networks* [144]. In the next sections, formalisms that explicitly take into account the dynamics of regulatory system will be discussed.

4 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 elements of the system by time-dependent variables with values contained in the set of non-negative real numbers. Regulatory interactions take the form of functional and differential relations between the concentration variables.

More specifically, gene regulation is modeled by *reaction-rate equations* expressing the rate of production of a gene product – a protein or an mRNA – as a function of the concentrations of other elements of the system. Reaction-rate equations have the mathematical form

$$\frac{dx_i}{dt} = f_i(\boldsymbol{x}),\ x_i \geq 0,\ 1 \leq i \leq n, \tag{2}$$

where $\boldsymbol{x}$ is the vector of concentrations of proteins, mRNAs, or small metabolites, and f_i a usually nonlinear function. The rate of synthesis of i is seen to be dependent upon the concentrations $\boldsymbol{x}$, possibly including x_i. The equations can be extended to take into account concentrations $\boldsymbol{u} \geq \boldsymbol{0}$ of input elements, e.g. externally-supplied nutrients:

$$\frac{dx_i}{dt} = f_i(\boldsymbol{x}, \boldsymbol{u}),\ x_i \geq 0,\ 1 \leq i \leq n. \tag{3}$$

They may also take into account discrete time delays arising from the time required to complete transcription, translation, and diffusion to the place of action of a protein:

$$\frac{dx_i}{dt} = f_i(x_1(t-\tau_{i1}), \ldots, x_n(t-\tau_{in})),\ x_i \geq 0,\ 1 \leq i \leq n, \tag{4}$$

where $\tau_{i1}, \ldots, \tau_{in} > 0$ represent time delays [24, 112, 172, 189] (see [10, 189] for other ways to represent time delays).

Powerful mathematical tools for modeling biochemical reaction systems by reaction-rate equations have been developed in the past century, especially for metabolic processes (see [30, 74, 229] for introductions). Using these tools, it is possible to construct kinetic models of genetic regulation processes by specifying the functions f_i.

In figure 6 a simple example of a kinetic model of a genetic regulation process is shown, going back to early work by Goodwin [61, 62]. 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 protein concentration x_2, and the metabolite concentration x_3. More generally, equations of the form

$$\begin{aligned} \frac{dx_1}{dt} &= \kappa_{1n} r(x_n) - \gamma_1 x_1,\ x_1 \geq 0, \\ \frac{dx_i}{dt} &= \kappa_{i,i-1} x_{i-1} - \gamma_i x_i,\ x_i \geq 0,\ 1 < i \leq n, \end{aligned} \tag{5}$$

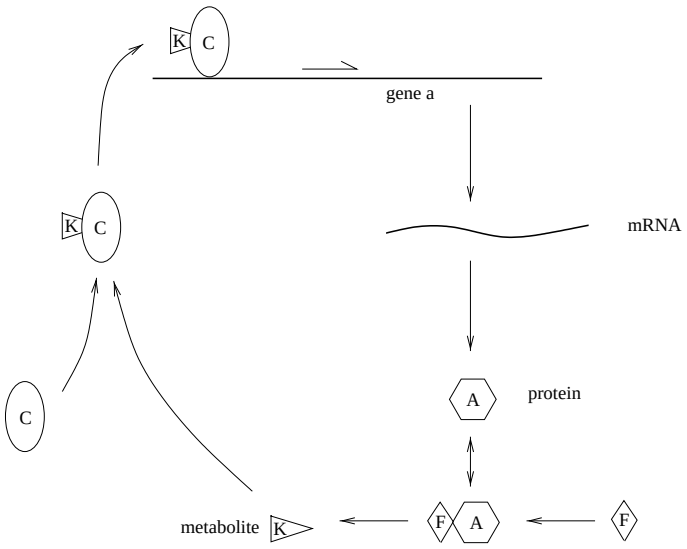
are employed [223]. The parameters $\kappa_{1n}, \kappa_{21}, \ldots, \kappa_{n,n-1} > 0$ are production constants and $\gamma_1, \ldots, \gamma_n > 0$ degradation constants. The reaction-rate equations express a balance between the number of molecules appearing and disappearing per unit time. In the case of x_1, the production term involves a nonlinear *regulation function* r, whereas the concentrations x_i, $1 < i \leq n$, 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 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* (figure 7(a)):

$$h^+(x_j, \theta_{ij}, m) = \frac{x_j^m}{x_j^m + \theta_{ij}^m}, \tag{6}$$

with $\theta_{ij} > 0$ the *threshold* for the influence of j on i, 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 i (*activation*). In order to express that an increase in x_j decreases the expression rate (*inhibition*), as in (5), the regulation function $h^+(x_j, \theta_{ij}, m)$ is replaced by $h^-(x_j, \theta, m) = 1 - h^+(x_j, \theta_{ij}, m)$. For $m > 1$, Hill curves have a sigmoid shape, in agreement with experimental evidence [237, 238].

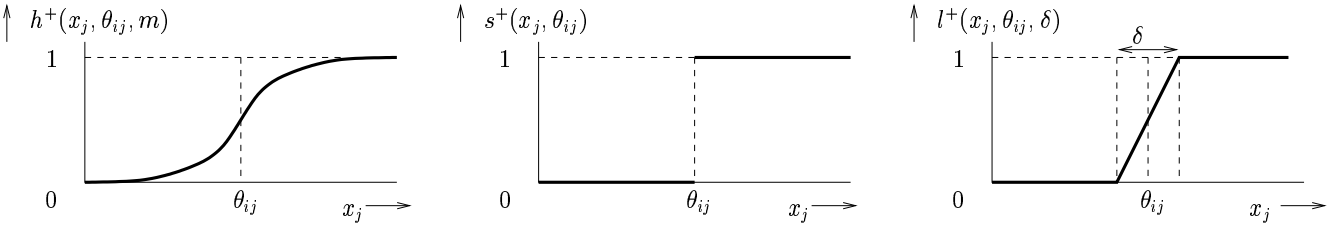
Due to the nonlinearity of f_i, analytical solution of the reaction-rate equations (2) is not normally possible. In special cases, qualitative properties of the solutions, such as the number and the stability of critical points and the occurrence of limit cycles, can be established. This is illustrated by studies that investigate the relationship between single feedback loops and the dynamics of regulatory systems (e.g., [62, 67, 68, 157, 220, 231], reviewed in [189, 223]). In the case of a *negative feedback* loop, as in figure 6, 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. It has been conjectured by Thomas



$$\dot{x}_1 = \kappa_{13} r(x_3) - \gamma_1 x_1$$
$$\dot{x}_2 = \kappa_{21} x_1 - \gamma_2 x_2$$
$$\dot{x}_3 = \kappa_{32} x_2 - \gamma_3 x_3$$

(a) (b)

Figure 6: (a) An example of a genetic regulatory system involving end-product inhibition and (b) its ODE model (adapted from [61, 62]). A and C represent proteins, and J and K metabolites. x_1, x_2, and x_3 represent the concentrations of mRNA a, protein A, and metabolite K, respectively. $\kappa_{13}, \kappa_{21}, \kappa_{32}$ are production constants, $\gamma_1, \gamma_2, \gamma_3$ degradation constants, and r a decreasing, nonlinear regulation function ranging from 0 to 1.



(a) (b) (c)

Figure 7: Examples of regulation functions: (a) Hill function h^+, (b) Heaviside or step function s^+, and (c) logoid function l^+.

that a negative feedback loop is a necessary condition for stable periodicity and a positive feedback loop for multistationarity [211, 214]. These conjectures have been formally proven in later work [66, 161, 191]. Following seminal ideas in [33, 143], Thomas has drawn attention to the relation between the feedback structure of a system and the biological phenomena of homeostasis and differentiation [207, 211, 213, 214]. 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 reaction-rate equations to mathematical analysis is to simplify the models, an approach that will be discussed in section 5. Alternatively, one can take recourse to numerical techniques. In *numerical simulation*, the exact solution of the equations is approximated by calculating approximate values $\boldsymbol{x_0}, \ldots, \boldsymbol{x_n}$ for $\boldsymbol{x}$ at consecutive time-points $t_0, \ldots, t_n$ [111]. Computer tools specifically adapted to the simulation and analysis of biochemical reaction systems have become available, such as GEPASI [133] and SCAMP [181].

Numerical simulations of a few well-studied regulatory systems have been carried out. For instance, Reinitz and Vaisnys [171] discuss a numerical model of the *cro-cI* switch controlling the growth of λ phage 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 λ phage system is described by MacAdams and Shapiro [127] (see also [187] and the review in [126]). 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.

Simulation of the functioning of a regulatory network is often complemented by techniques from system dynamics, for instance *bifurcation analysis* tools to investigate the sensitivity of steady states and limit cycles to parameter values [195]. Borisuk and Tyson [15] have applied these techniques to a numerical model of a much-studied example of post-translational modification, the control of mitosis in the *Xenopus* oocyte [152] (see [151] 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 [22, 153] (see [60, 155, 221, 222] 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 parameters in (2). Numerical parameter values are available for only a handful of systems, the λ phage switch being the prime example, and then usually only for part of the system. In contrast, in the cell cycle models mentioned above, the parameter values were chosen such that the models are able to reproduce the observed qualitative behavior. For larger models, finding such values may require many simulations. The growing availability of gene expression data suggests another approach to obtain numerical values for the rate constants and other parameters in (2). 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 can be estimated with the help of system identification techniques [119]. In a later section, this approach will be examined in more detail.

5 Piecewise-linear differential equations

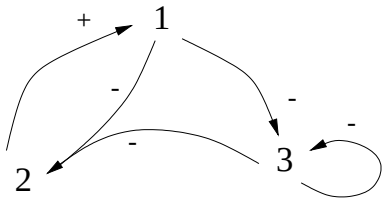
Building complete kinetic models of genetic regulatory systems requires detailed knowledge on reaction mechanisms. In many situations, this information is not available and one needs to take recourse to more approximate models. In this section, one such class of models will be considered, a class of *piecewise-linear differential equations (PLDEs)* obtained from the reaction-rate equations (2) by making a number of simplifying assumptions. The PLDEs have favorable mathematical properties that facilitate their analysis.

Consider reaction-rate equations of the form

$$\frac{dx_i}{dt} = g_i(\boldsymbol{x}) - \gamma_i x_i, \;\; x_i \geq 0, \;\; 1 \leq i \leq n, \tag{7}$$

where x_i denotes the cellular concentration of the product of gene i and $\gamma_i > 0$ the decay rate of x_i [31, 51, 138, 214]. The functions g_i are defined as a sum of interaction terms corresponding to the interactions in the regulatory network. More precisely, for each interaction involving a regulated gene i and a set of regulating genes J, the sum contains a term $\kappa_{iJ} \prod_{j \in J} r(x_j)$, where $r(x_j)$ is a regulation function and κ_{iJ} a rate constant specifying the maximum expression level of i under the influence of J

[31, 138]. In the example of figure 8, the two genes regulating gene 2 lead to a product of two regulation functions, indicating that both regulatory proteins are necessary for the inhibition of gene 2.



(a)

$$\dot{x}_1 = \kappa_1 + \kappa_{12} h^+(x_2, \theta_{12}, m) - \gamma_1 x_1$$
$$\dot{x}_2 = \kappa_2 + \kappa_{213} h^-(x_1, \theta_{21}, m) h^-(x_3, \theta_{23}, m) - \gamma_2 x_2$$
$$\dot{x}_3 = \kappa_3 + \kappa_{31} h^-(x_1, \theta_{31}, m) + \kappa_{33} h^-(x_3, \theta_{33}, m) - \gamma_3 x_3$$

(b)

Figure 8: (a) Example regulatory network of three genes and (b) reaction-rate equations (adapted from [216]). x_1, x_2, and x_3 represent protein or mRNA concentrations, respectively, the κ parameters are production constants, and $\gamma_1, \gamma_2, \gamma_3$ degradation constants.

Kinetic models (7) abstract from the biochemical details of genetic interactions by directly relating the expression levels of gene products through sigmoid regulation functions. In order to simplify the mathematical analysis, the continuous Hill functions are replaced by discontinuous step functions (figure 7(b)):

$$s^+(x_j, \theta_{ij}) = \begin{cases} 1, & x_j > \theta_{ij} \\ 0, & x_j < \theta_{ij} \end{cases}, \quad \text{and} \quad s^-(x_j, \theta_{ij}) = 1 - s^+(x_j, \theta_{ij}). \tag{8}$$

As a consequence, the nonlinearities in g_i are eliminated. The approximation of a continuous sigmoid by a discontinuous step function has been proposed by several authors (e.g., [51, 56, 93, 197, 198, 209, 232]), and has been justified on the ground of the switch-like character displayed by genes whose expression is regulated by steep sigmoid curves: below (above) a certain concentration, the regulator tends to have little influence, whereas above (below) this concentration its influence rapidly reaches a maximum level.

The resulting reaction-rate equations are piecewise-linear differential equations (PLDEs) of the form

$$\frac{dx_i}{dt} = b_i(\boldsymbol{x}) - \gamma_i x_i, \; x_i \geq 0, \; 1 \leq i \leq n, \tag{9}$$

where b_i is a piecewise-constant function. In particular, b_i is a sum of products of step functions weighted by a rate constant.

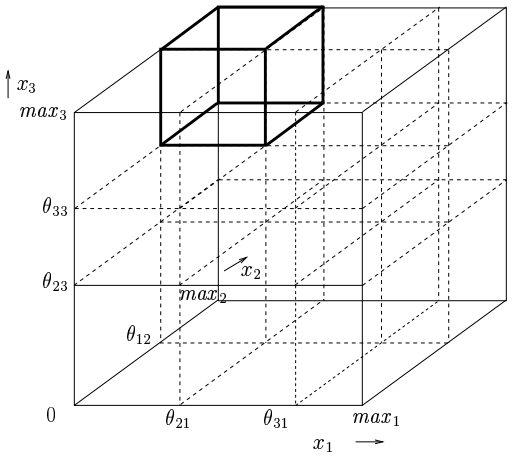
Equations (9) have been well-studied in mathematical biology [51, 52, 53, 54, 57, 58, 115, 138, 139, 160, 190, 192, 214]. Consider an n-dimensional (hyper)box of the phase space defined as follows:

$$0 \leq x_i \leq max_i = \max_{\boldsymbol{x} \geq \mathbf{0}} b_i(\boldsymbol{x})/\gamma_i, \; 1 \leq i \leq n. \tag{10}$$

Assuming that $\theta_{ji} < max_i$ for all genes j regulated by the product of gene i, the $n-1$-dimensional threshold (hyper)planes $x_i = \theta_{ji}$ divide the n-box into volumes. Figure 9(a) displays the phase space box and volumes corresponding to the example system of figure 8.

In each volume 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:

$$\dot{x}_i = \mu_i - \gamma_i x_i, \; x_i \geq 0, \; 1 \leq i \leq n. \tag{11}$$



(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)

Figure 9: (a) The phase space box of the model in figure 8, divided into $2 \cdot 3 \cdot 3 = 18$ volumes by the threshold planes. (b) The state equations for the volume $0 \leq x_1 < \theta_{21}$, $\theta_{12} < x_2 \leq max_2$, and $\theta_{33} < x_3 \leq max_3$ (the volume demarcated by bold lines).

Figure 9(b) gives an example of the state equations corresponding to the volume $0 \leq x_1 < \theta_{21}$, $\theta_{12} < x_2 \leq max_2$, and $\theta_{33} < x_3 \leq max_3$.

Since equations (11) are linear and orthogonal, the behavior inside a volume is straightforward. In particular, all trajectories evolve towards the stable *focal state* $\boldsymbol{\mu}/\boldsymbol{\gamma}$ of (11). The focal state of a volume may lie inside or outside the volume. If the focal state lies outside the volume, the trajectories will tend towards one or several of the threshold planes bounding the volume. Since (9) is not defined in the threshold planes, special attention should be given to the behavior of the system as it approaches the threshold planes. Following Plahte, Mestl and Omholt [163], the behavior of the piecewise-linear equations (9) at the threshold planes can be defined as the behavior of the original nonlinear equations (7) when the sigmoid function (6) approaches the step function (8), in the limit $m \to \infty$ (see Plahte, Mestl and Omholt [160] for a different definition in terms of the logoid regulation function).

The PLDEs (9) have two types of steady states. In addition to *regular* steady states lying inside the volumes of the n-box, there are *singular* steady states situated on one or more threshold planes separating the volumes [192]. Regular steady states can be readily identified, since they are the focal state of some volume 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 Glass [53], Snoussi and Thomas [192] and Plahte, Mestl and Omholt [160].

If the focal state of a volume lies outside the volume, a trajectory approaching a threshold plane may be continued by a trajectory in an adjacent volume.[2] The resulting global behavior of the PLDEs can be quite complex and is not well understood in general. Continuations of trajectories in several volumes may give rise to oscillations towards a singular steady state located at the intersection of threshold planes, cycles, or limit cycles [57, 58, 115, 139, 160]. 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 $\boldsymbol{b}$ [54, 115, 116, 136, 137]. Mestl and colleagues [137] 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 ([164, 202]; see also [84, 165, 203] 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.

[2] In some cases, though, trajectories approaching a threshold plane may not be able to cross it. This is particularly so for systems involving autoregulation, such as gene 3 in figure 8 [52, 160, 163, 190, 214].

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 4). 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 [164].

How much information is lost by the piecewise-linear approximation of the original nonlinear system (7)? Numerical simulation studies by Glass and Kauffman [55, 56] and Glass and Pérez [59] show that in many cases there is no difference in the qualitative properties of the solutions of the nonlinear and piecewise-linear DEs (see also [162, 163, 214]; [160] discusses situations in which differences do occur). In the context of a study on the regulation of free iron level in mammalian cells, a post-transcriptional regulatory systems for which quantitative data are available, Omholt and colleagues [156, 163] 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. The numerical and analytical results turned out to be within the same order of magnitude, comparable to the experimental accuracy.

One of the main advantages of the use of PLDEs is the simplicity of the mathematical analysis entailed by the form of the equations. In a similar vein, Savageau [182, 184] has proposed to define $f_i(\boldsymbol{x})$ in (2) by means of *power-law functions*, giving rise to nonlinear models of the form

$$\dot{x}_i = \sum_{k=1}^{p} \kappa_{ik} \prod_{j=1}^{n} x_j^{g_{ijk}} - \sum_{k=1}^{p} \gamma_{ik} \prod_{j=1}^{n} x_j^{h_{ijk}}, \;\; x_i \geq 0, \;\; 1 \leq i \leq n, \tag{12}$$

where g_{ijk} (h_{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, (12) becomes a linear system that is much more tractable to analysis than the original nonlinear system (see [228, 229] for overviews 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 a regulatory network [183, 184].

6 Stochastic 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 (figure 10). As noted in the previous section, this may not always be possible because the information required to build such models is lacking. A more fundamental objection against the use of DEs is that a number of implicit assumptions underlying the formalism are no longer valid on the molecular level.

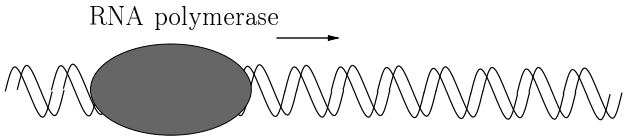


Figure 10: Example of two reaction steps involved in DNA transcription. The first reaction step describes the binding of RNA polymerase (RNAP) to the transcription initiation site, while the second reaction step describes its advancement along the DNA molecule.

Differential equations presuppose that concentrations of substances vary *continuously* and *deterministically*, both of which assumptions may be questionable in the case of gene regulation [49, 148, 173, 199]. In the first place, the small numbers of molecules of some of the elements 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. 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 [6, 49, 128, 148, 173]. Discrete amounts $\boldsymbol{X}$ of molecules are taken as state variables, and a joint probability distribution $p(\boldsymbol{X}, t)$ is introduced expressing the probability that at time t the cell contains X_1 molecules of the first substance, X_2 molecules of the second substance, etc. The time evolution of the function $p(\boldsymbol{X}, t)$ can now be specified as follows:

$$p(\boldsymbol{X}, t + \Delta t) = p(\boldsymbol{X}, t)(1 - \sum_{j=1}^{m} \alpha_j \Delta t) + \sum_{j=1}^{m} \beta_j \Delta t, \tag{13}$$

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 $[t, t + \Delta t]$ [49, 50]. Rearranging (13), and taking the limit as $\Delta t \to 0$, gives the *master equation* [226]:

$$\frac{\partial}{\partial t} p(\boldsymbol{X}, t) = \sum_{j=1}^{m} (\beta_j - \alpha_j p(\boldsymbol{X}, t)). \tag{14}$$

Compare this equation with the reaction-rate equations (2) in section 4. Whereas the latter determines 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 deterministic reaction-rate equation. Moreover, numerical simulation of the system is complicated by the form of (14), which contains $n+1$ independent variables: n discrete variables $\boldsymbol{X}$ and a continuous variable t. Under certain conditions the master equation can be transformed into stochastic differential equations, where the values of the state variables are to be interpreted as average concentrations [148, 226]. These conditions, however, are rather restrictive and often do not hold in the case of genetic regulatory systems.

An alternative approach would be to disregard the master equation altogether and directly simulate the stochastic time evolution of the regulatory system. This idea underlies the *stochastic simulation* approach developed by Gillespie [49]. 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 (13). 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 [46] discuss various improvements of the Gillespie algorithm, directed at reducing the computational complexity of the procedure.

Stochastic simulation has been used by McAdams and Arkin to analyze the interactions controlling the expression of a single prokaryotic gene [128]. 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. The influence of fluctuations in the rate of gene expression on the choice between lytic and lysogenic growth in phage λ has been investigated in later work [6]. 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 [46].

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 $p(\boldsymbol{x}, t)$. 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.

7 Partial differential equations and other spatially distributed models

Reaction-rate equations (2) 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 reaction-rate equations (section 4). However, in the case that multiple compartments of a cell, or multiple cells, need to be explicitly modeled, a more drastic extension of (2) 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 $\boldsymbol{x}^{(l)}(t)$, 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 expression occurs in the manner described by equation (2). *Between* pairs of adjacent cells l and $l+1$, $1 \leq l \leq p-1$, diffusion of gene products is assumed to take place proportional 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*

$$\frac{dx_i^{(l)}}{dt} = f_i(\boldsymbol{x}^{(l)}) + \delta_i \left(x_i^{(l+1)} - 2x_i^{(l)} + x_i^{(l-1)} \right), \; x_i^{(l)} \geq 0, \; 1 \leq i \leq n, \; 1 < l < p. \tag{15}$$

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 (figure 11). In addition, the diffusion constants can be made to vary at different locations. Although (15) 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 [55] for an example).

(a) (b)

Figure 11: Examples of spatial configurations of multicellular genetic regulatory systems.

If the number of cells is large enough, the discrete variable l in (15) 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)*:

$$\frac{\partial x_i}{\partial t} = f_i(\boldsymbol{x}) + \delta_i \frac{\partial^2 x_i}{\partial l^2},\ x_i^{(l)} \geq 0,\ 0 \leq l \leq \lambda,\ 1 \leq i \leq n. \tag{16}$$

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

$$\frac{\partial^2}{\partial l^2} x_i(0,t) = 0 \text{ and } \frac{\partial^2}{\partial l^2} x_i(\lambda,t) = 0. \tag{17}$$

Reaction-diffusion equations have been widely used in mathematical biology to study pattern formation in development (see [47, 98, 122, 130, 149] for reviews). The equations are also referred to as Turing systems, after the mathematician who suggested their applicability to developmental phenomena [219]. Turing considered a special case of equations (15) and (16), 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., [47]).

In what follows, continuous equations (16) 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 $x_1^*, x_2^* > 0$ (figure 12). The spatial homogeneity of the steady state ensures that it is consistent with boundary conditions (17). 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,

$$\Delta x_i(l,t) = x_i(l,t) - x_i^* = \sum_{k=0}^{\infty} c_{ik}(t) \cos(\frac{\pi k}{\lambda} l),\ i = 1, 2, \tag{18}$$

with $\cos(\pi k l/\lambda)$ the *modes* or *eigenfunctions* of the Laplacian operator $\partial^2/\partial l^2$ on $[0, L]$, and $c_{ik}(t)$ the *mode amplitudes* [16, 149, 185] (figure 12). 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 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.

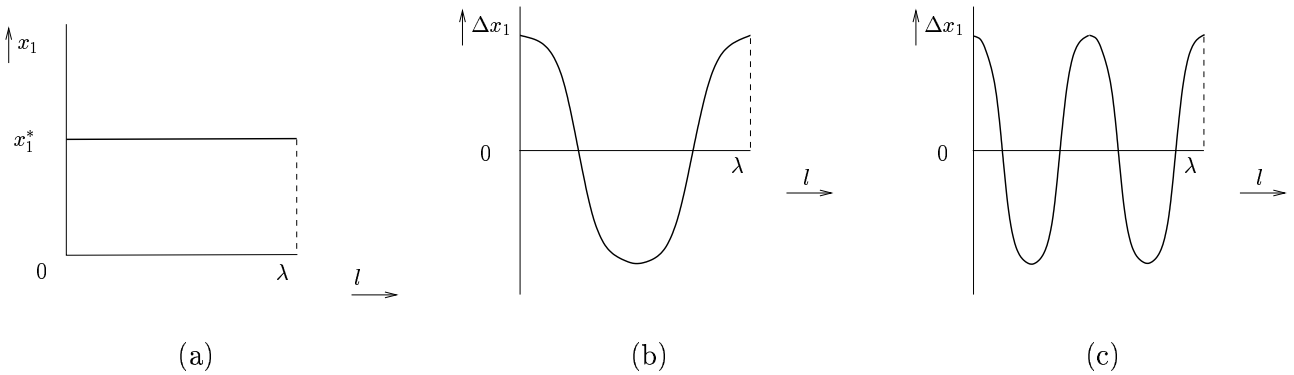


Figure 12: (a) Homogeneous steady state of the reaction-diffusion system. (b)-(c) Two modes of the solution of the linearized system around the steady state. In (b) $k = 2$ and in (c) $k = 4$.

Gierer and Meinhardt have formulated constraints on the functions and parameters in (16) such that an initial steady state can be destabilized by a small fluctuation [47, 48, 130]. 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 [64, 63, 98], following earlier suggestions in [99], 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 (16) around its equilibrium. By taking parameters slowly time-varying to mimic developmental effects, sequences of sigmoids of increasing frequency (increasing k, see figure 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., [20, 109, 145]). Other approaches towards modeling *Drosophila* embryogenesis by means of reaction-diffusion systems can be found in [79, 80, 110, 129, 131, 132, 175] (see [98, 130] 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 [11, 20]. 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 (15) and (16) that reflect the underlying regulatory networks more directly.

The *gene circuit method* proposed by Mjolsness and colleagues ([142, 167, 168, 169], reviewed in [170]) is an example of such an approach (see [21, 35, 69, 106, 107, 108, 180, 224, 230] for related work). The method employs a variant of (15):

$$\frac{dx_i^{(l)}}{dt} = \rho_i r_i(\sum_{j=1}^{n} \lambda_{ij} x_j + \mu_j u_1^{(l)} + \kappa_i) - \gamma_i x_i^{(l)} + \delta_i(k)(x_i^{(l+1)} - 2x_i^{(l)} + x_i^{(l-1)}), \qquad (19)$$
$$x_i^{(l)} \geq 0,\ 1 \leq i \leq n,\ 1 < l < p.$$

where $\boldsymbol{x}^{(l)}$ is a state vector of protein concentrations in nucleus l, and $u_1^{(l)}$ an input variable. The 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 κ_i, is modulated by a sigmoid regulation function r_i and a constant ρ_i 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]

Model (19) has 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 [124, 123] 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. 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 give an indication of 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 [168, 169]. Using the estimated parameter values, mutant patterns have also been predicted [186].

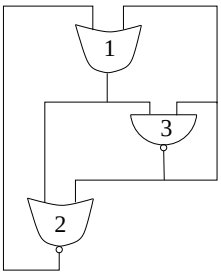
The fact that parameters of the model may represent interactions in the regulatory network, like the λ_{ij}s in equation (19), deserves special mention. As noted above, it implies that an estimation of parameter values simultaneously fixes the regulatory structure of the system. Models of the system could thus be induced from expression data, without relying on prior knowledge on the existence of regulatory interactions between the genes. This idea has already been encountered in section 3, of course, but is here generalized to the case of dynamical models. Similar work taking ordinary differential equations and difference equations as their point of departure has been presented recently [24, 34, 150, 227, 233].

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 [38]. 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 (19), 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 [142] for details).

8 Boolean networks

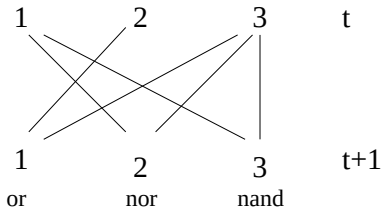
The use of discontinuous step functions in piecewise-linear reaction-rate equations suggests a more radical idealization of the elements of a genetic regulatory network and their interactions. The state of a gene can be described by a Boolean variable, expressing that it is active (on, 1) or inactive (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 figure 13. Modeling regulatory networks by means of Boolean networks has followed in the wake of a groundbreaking study by Kauffman [93] (see [197, 198] for early use of the Boolean approximation). Recent reviews of the use of Boolean network models are Kauffman's book [98] and an article by Somogyi and Sniegoski [193].



(a)

$$\hat{x}_1(t+1) = \hat{x}_2(t) \text{ or } \hat{x}_3(t)$$
$$\hat{x}_2(t+1) = \hat{x}_1(t) \text{ nor } \hat{x}_3(t)$$
$$\hat{x}_3(t+1) = \hat{x}_1(t) \text{ nand } \hat{x}_3(t)$$

(b)



(c)

Figure 13: (a) Example Boolean network and (b) the corresponding equations. In this case, $n = 3$ and $k = 2$. (c) Wiring diagram of the Boolean network.

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 figure 13. In summary, the dynamics of a Boolean network describing a regulatory system is given by

$$\hat{x}_i(t+1) = \hat{b}_i(\hat{\boldsymbol{x}}(t)), \ 1 \leq i \leq n, \tag{20}$$

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 (figure 13(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 *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 [236]. 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.

Boolean networks were among the first formalisms for which model induction methods were proposed, the REVEAL algorithm developed by Liang and colleagues being an example ([117]; see [1, 2, 3, 81, 91, 150] for other examples and [209] 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 [94]. An interesting application of Boolean networks is their use in the study of global properties of large-scale regulatory systems (see [97, 98, 193, 234] 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. Similar ideas have been explored by means of continuous formalisms [9, 54, 92, 147].

The above approach has been explored by Kauffman ([93]; see [92, 94, 95, 96] for later work, summarized in [97, 98]. 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 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 [98].

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 these idealizations are not appropriate, and more general methods are required.

9 Generalized logical formalisms

In this section, a generalized logical method developed by Thomas and colleagues will be discussed. The method is based on a formalism taking a middle ground between Boolean networks and differential equations. The formalism is discrete in that it abstracts from the continuous description of differential equations. But at the same time it 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 underwent several extensions. The version that will be presented here in outline is more extensively described in [212, 214, 216] (see [209, 210] for earlier formulations). Related ideas have been developed in a series of papers by Glass and colleagues [51, 52, 53, 115], using hybrid models with logical and continuous aspects.

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 [225]:

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

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

$$\begin{aligned}
\hat{x}_i &= 0, \text{ if } x_i < \sigma_i^{(1)} \\
\hat{x}_i &= 1, \text{ if } \sigma_i^{(1)} < x_i < \sigma_i^{(2)} \\
&\ldots \\
\hat{x}_i &= p, \text{ if } x_i > {}^{p}\sigma_i.
\end{aligned}$$

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

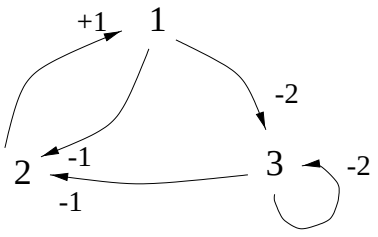
The pattern of regulatory interactions in the system is described by logical equations of the form

$$\hat{X}_i(t) = \hat{b}_i(\hat{\boldsymbol{x}}(t)),\ 1 \leq i \leq n, \tag{21}$$

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 $\hat{\boldsymbol{x}}$ at t. It should be distinguished from the time derivative $\dot{x}_i$ in (2) and the successor value $\hat{x}(t+1)$ in (20).[4] The logical function $\hat{b}_i$ is a generalization of the Boolean function in (20), 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 figure 14, 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 $\hat{x}_1 = 2$.

[4] In fact, the image of $\hat{x}_i$ can be shown to correspond to the focal state in the PLDEs of section 5 [190, 214].



(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

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

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

(c)

Figure 14: (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 [216]. (b) Logical equations corresponding to the graph, and (c) possible definitions of the logical functions.

The graph in (a) can be transformed into logical equations, as shown in (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 (c), in a notation that emphasizes the correspondence with the Boolean functions in the previous section (see [214, 216] 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. The table 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.

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.,

$$\hat{\boldsymbol{X}} = \hat{\boldsymbol{x}}. \tag{22}$$

Since the number of logical states is finite, due 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 figure 14, 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 [212, 214]. 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. Assuming 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 [214].

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* [190, 214] allow classes of logical functions, as compared with individual logical functions, to be conveniently characterized. This facilitates the identification of logical steady states. By introducing logical values that correspond to threshold values in the continuous description [192, 214], it becomes possible to detect logical steady states that are located on the thresholds, related to the singular steady states in the continuous formalism (section 5).

The logical method of Thomas has been implemented [205] and proven its effectiveness in the study of a number of genetic regulatory systems of limited scale, like λ phage infection in *E. coli* [208, 214, 215] and dorso-ventral pattern formation in *Drosophila* [180]. Another examples is a study of the regulatory network controlling flower morphogenesis in *Arabidopsis thaliana* ([135], see also [134]). 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 non-floral 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 non-floral steady state to one of the four flowering states. More generally, the results support the claim, referred to above, that positive feedback loops play a dominant role in developmental processes, and they point at the modularity of genetic regulatory networks involved in development.

10 Qualitative differential equations

Snoussi [190] has demonstrated that the generalized logical formalism of the previous section can be seen as an abstraction of a special case of (9), allowing additions but not multiplications of step functions. In addition, the logical steady states can be shown to correspond to the steady states of the piecewise-linear differential equations. In some cases, cycles in the state transition graph correspond to limit cycles in the phase space. 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 used for this purpose are the *qualitative differential equations (QDEs)* used in the simulation method QSIM [104, 105]. QDEs are abstractions of ODEs of the form

$$\frac{dx_i}{dt} = f_i(\boldsymbol{x}),\ x_i \geq 0,\ 1 \leq i \leq n, \tag{23}$$

where f_i 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, has stimulated an interest in qualitative simulation

techniques. An example of their application is given by Heidtke and Schulze-Kremer [73]. 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 [4, 101, 218]

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 5 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, it has been shown possible to treat regulatory networks of up to 18 genes involved in complex feedback loops [31].

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 [14]. 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 show how the conclusions of their model can be made more precise by adding semi-quantitative information [73]. The integration of numerical information is more difficult to achieve in the logical approach of the previous section, 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 [146]. One of the basic approaches towards automated modeling, *qualitative process theory* [41], has been used by Karp in his method for the construction and revision of gene regulation models [87, 88, 89]. 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 [72, 101] (see also [32]).

11 Rule-based formalisms

The discrete formalisms discussed in sections 8 to 10 allow one to deal with situations in which not much is known about the numerical value of the parameters and conditions. Although the resulting models are weaker than their continuous counterparts, they preserve the basic form of the latter. Whereas DEs model the elements of a regulatory system by continuous variables, and regulatory interactions by functions relating the variables, the logical formalisms and their generalizations employ discrete variables and finite relations defined over their domains. *Knowledge-based* or *rule-based simulation formalisms*, developed in the field of artificial intelligence, present another approach to make up for the lack of quantitative knowledge. They permit a rich variety of knowledge about the regulatory system to be expressed in an intuitively appealing way. The logical and physical structure of a gene – such

as the 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 [70]. 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 [19, 44], 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 [140]. 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 [188], 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 "non-critical" parts of the decision between the lytic and lysogenic pathway, and a qualitative phase-space analysis of reaction-rate equations for the "critical" parts.

The major advantage of rule-based formalisms, their capability to deal with a richer variety of biological knowledge, 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 ([75, 76, 77], it remains true that in this respect rule-based formalisms cannot compete with DE formalisms.

12 Outlook

Figure 3 shows the integration of experimental and computational methods in the analysis of genetic regulatory systems. Until recently, modeling and simulation studies have been mostly constrained

to regulatory systems of modest size and complexity that have been well-characterized already by experimental means. The analysis of larger systems, falling outside the restricted canon of model systems, is hampered by difficulties in obtaining the necessary information to build models and test predictions.

In the first place, the biochemical reaction mechanisms underlying regulatory interactions are usually not or incompletely known. This means that detailed kinetic models cannot be built and more approximate models are needed, such as the PLDEs of section 5 and the Boolean networks and their generalizations in sections 8 and 9. The price being paid for the simplifying assumptions embodied in the formalisms is a potential loss of biological relevance of the predictions derived from the models. In the second place, quantitative information about the regulatory system is only seldom available. This means so that simulation methods have to be able to deal with qualitative models. Sections 8 to 11 discuss different ways to formalize genetic regulatory systems in a qualitative way. Unfortunately, except when introducing strong simplifying assumptions, qualitative approaches experience upscaling problems due to the weak nature of qualitative information. In addition, their predictions are less precise than those obtained with numerical models.

The emergence of new experimental techniques, as well as the development of databases and other infrastructural provisions to give access to experimental data, promise to relieve the data bottleneck. Together with ever-increasing computer power, this allows new approaches to modeling and simulation to be tried.

First of all, one can expect models of gene regulation to become more quantitative in the future (see [103, 121] for more general statements on the quantification of cellular dynamics). The ability to engineer regulatory networks *in vivo*, as described in [37, 45], might facilitate the direct measurement of model parameters. In addition, if the expected improvements in the quantification of gene expression data materialize, reliable numerical estimates of parameter values can be obtained by means of the techniques described in section 7. The use of quantitative models would allow larger systems to be studied at higher precision. More attention to the quantitative aspects of gene regulation does not mean that qualitative models lose their value. On the contrary, they may guide the application of numerical methods and the interpretation of their results (see [156] for an example).

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. In the limit, this may result in detailed models breaking regulatory processes down to individual reaction steps, like in the λ phage model mentioned in section 6. It should be remarked, though, that testing the validity of such models requires measurements of a wider range of biological parameters than usually available [199, 200]. In order to get a grasp on the biochemical details of post-transcriptional regulation, for example, measurements of mRNA expression levels need to be supplemented by measurements of cellular protein concentrations, preferably distinguishing between the different forms in which a protein occurs. The measurement of protein concentrations is currently much more difficult to achieve, although recent advances in proteomics promise to improve this situation [83, 158].

Even a cursory look at the examples of eukaryotic transcriptional regulation in a review paper by Arnone and Davidson [8] cannot fail to make clear that models of gene regulation may need to take into account tens of transcription factors for a single gene, interacting in complex ways (see also [239]). The scale and complexity of realistic regulatory systems will impose limits on modeling and simulation, even if the increasing quantification of experimental data is taken into account. Computer power has grown enormously in the past decades, but the inherent complexity of optimization problems like the estimation of model parameters is sure to drain computational resources rapidly for even relatively small networks.

Interestingly, the biological systems being studied may impose constraints that considerably relieve upscaling problems in practice. For instance, in a simulation study of the segment polarity network in *Drosophila*, it was found that the behavior of the system was quite robust to variations in parameter values and initial conditions over sometimes several orders of magnitude [230] (see also [12]). This sug-

gests that it is the network structure rather than the strength of the interactions that confers stability to the system. Moreover, there are indications that regulatory networks are organized in a modular fashion. The above-mentioned study suggested that *Drosophila* segmentation genes constitute a relatively autonomous module responding to varying inputs (see also [135, 206]). Modular organization of regulatory networks would make it possible to consider smaller subsystems, loosely connected through regulatory interactions or acting on different time-scales, in relative isolation from each other. Of course, this reductionist strategy needs to be complemented by an investigation of interactions among the modules, possibly on a more abstract and hence computationally less demanding level.

Within the broader context of a cell, gene regulation is interacting with metabolism, signal transduction, replication, recombination and repair, and a variety of other processes. In figure 15 an example of a self-regulatory mechanism is displayed, in which gene *a* regulates its own expression by encoding an enzyme that catalyzes a reaction step in a metabolic pathway. The resulting metabolite K induces the regulatory action of protein C on *a* by modifying C's conformation. This simple example of end-product inhibition illustrates that in order to understand the functioning of cells, and by extensions organisms, the scope of current models needs to be enlarged by integrating gene regulation with other cellular processes. Some initiatives towards modeling and simulating entire prokaryote and even eukaryote systems have gone into this direction [86, 90, 100, 217].

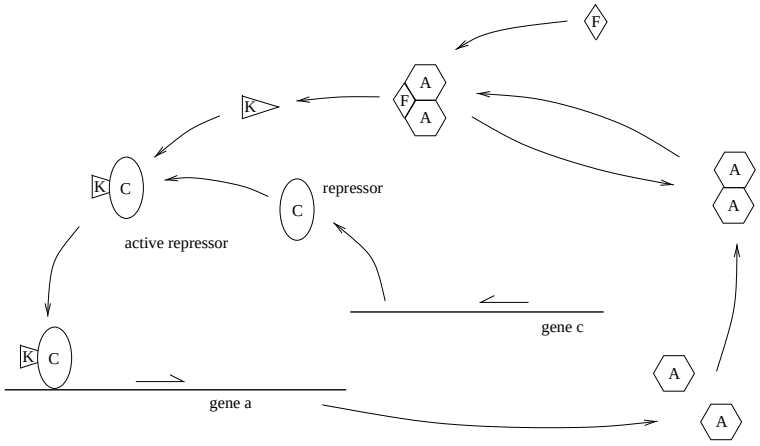


Figure 15: An example of a genetic regulatory systems intersecting with a metabolic pathway. A and C denote proteins and F and K metabolites. For the notation, see figure 2.

As remarked in the introductory section, computer support for the construction of models of regulatory systems is attracting increasing attention. Two different approaches can be distinguished, corresponding to the two arrows entering the modeling box in figure 3. On the one hand, models can be *composed* from knowledge on regulatory interactions stored in databases and knowledge bases, as illustrated in section 10. On the other hand, models can be *induced* from expression data by methods like those discussed in sections 2, 5, and 8. Each of these approaches has its merits, but neither of them is sufficient in itself. Model composition is thwarted by the fact that current knowledge of regulatory mechanisms is incomplete at best, whereas model induction struggles with the problem that models are underdetermined by available expression data.

It seems that a combination of the two approaches would be most effective, in that knowledge on regulatory mechanisms may guide the search for a fit between model and data, and play a role in the validation of the networks proposed by induction algorithms. Because of the problems mentioned at the end of section 5, spurious interactions are bound to occur in these models and independent validation will be necessary. Apart from designing further experiments to test the models [2, 81], validation could be achieved by mapping the genetic interactions supposed to underlie gene expression patterns to biochemical pathways in the cell, assigning a pathway function to each of the genes in the

network [86, 240]. In addition, it might involve the analysis of DNA sequences to check whether a regulated gene has upstream targets for the hypothesized regulatory proteins [201]. These strategies are illustrative of a more general tendency to integrate different experimental and computational tools in the analysis of genetic regulatory networks [204].

Acknowledgements The author would like to thank Olivier Martin, Michel Page, Fran cois Rechenmann, and Denis Thieffry for their comments on earlier versions of this report.

References

[1] T. Akutsu, S. Kuhara, O. Maruyama, and S. Miyano. Identification of gene regulatory networks by strategic gene disruptions and gene overexpressions. In H. Karloff, editor, *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA-98)*, pages 695–702, Philadelphia, 1998. Society for Industrial and Applied Mathematics.

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

[3] T. Akutsu, S. Miyano, and S. Kuhara. 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, editors, *Pacific Symposium on Biocomputing, PSB'99*, volume 4, pages 17–28, Singapore, 1999. World Scientific Publishing.

[4] T. Akutsu, S. Miyano, and S. Kuharu. Algorithms for inferring qualitative models of biological networks. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB 2000*, volume 5, pages 290–301, Singapore, 2000. World Scientific Publishing.

[5] A. Arkin and J. Ross. Statistical construction of chemical reaction mechanisms from measured time-series. *Journal of Physical Chemistry*, 99:970–979, 1995.

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

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

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

[9] R.J. Bagley and L. Glass. Counting and classifying attractors in high dimensional dynamical systems. *Journal of Theoretical Biology*, 183:269–284, 1996.

[10] H.T. Banks and J.M. Mahaffy. Stability of cyclic gene models for systems involving repression. *Journal of Theoretical Biology*, 74:323–334, 1978.

[11] J. Bard and I. Lauder. How well does Turing's theory of morphogenesis work? *Journal of Theoretical Biology*, 45:501–531, 1974.

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

[13] A. Ben-Dor and Z. Yakhini. Clustering gene expression patterns. In *Proceedings of the Third Annual International Conference on Computational Molecular Biology, RECOMB'99*, pages 33–42, New York, N.Y., 1999. ACM Press.

[14] D. Berleant and B. Kuipers. Qualitative and quantitative simulation: Bridging the gap. *Artificial Intelligence*, 95:215–256, 1997.

[15] M.T. Borisuk and J.J. Tyson. Bifurcation analysis of a model of mitotic control in frog eggs. *Journal of Theoretical Biology*, 195:69–85, 1998.

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

[17] P.A. Brown and D. Botstein. Exploring the new world of the genome with DNA microarrays. *Nature Genetics*, 21(supplement):33–37, 1999.

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

[19] D.L. Brutlag, A.R. Galper, and D.H. Millis. Knowledge-based simulation of DNA metabolism: Prediction of enzyme action. *Computer Applications in the Biosciences*, 7(1):9–19, 1991.

[20] B. Bunow, J.-P. Kernevez, G. Joly, and D. Thomas. Pattern formation by reaction-diffusion instabilities: Application to morphogenesis in *Drosophila*. *Journal of Theoretical Biology*, 84:629–649, 1980.

[21] Z. Burstein. A network model of developmental gene hierarchy. *Journal of Theoretical Biology*, 174:1–11, 1995.

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

[23] T. Chen, V. Filkov, and S.S. Skiena. Identifying gene regulatory networks from experimental data. In *Proceedings of the Third Annual International Conference on Computational Molecular Biology, RECOMB'99*, pages 94–103, New York, N.Y., 1999. ACM Press.

[24] T. Chen, H.L. He, and G.M. Church. Modeling gene expression with differential equations. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB'99*, volume 4, pages 29–40, Singapore, 1999. World Scientific Publishing.

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

[26] J. Collado-Vides. A transformational-grammar approach to the study of the regulation of gene expression. *Journal of Theoretical Biology*, 136:403–425, 1989.

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

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

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

[30] A. Cornish-Bowden. *Fundamentals of Enzyme Kinetics*. Portland Press, London, revised edition, 1995.

[31] H. de Jong and M. Page. Qualitative simulation of large and complex genetic regulatory systems. In W. Horn, editor, *Proceedings of the 14th European Conference on Artificial Intelligence, ECAI 2000*, pages 141–145, Amsterdam, 2000. IOS Press.

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

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

[34] P. d'Haeseleer, X. Wen, S. Fuhrman, and R. Somogyi. Linear modeling of mRNA expression levels during CNS development and injury. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB'99*, volume 4, pages 41–52, Singapore, 1999. World Scientific Publishing.

[35] B.A. Edgar, G.M. Odell, and G. Schubiger. A genetic switch, based on negative regulation, sharpens stripes in *Drosophila* embryos. *Developmental Genetics*, 10:124–142, 1989.

[36] M.B. Eisen, P.T. Spellman, P.O. Brown, and D. Botstein. Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the USA*, 95:14863–14868, 1998.

[37] M.B. Elowitz and S. Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[38] R.S. Erb and G.S. Michaels. Sensitivity of biological models to errors in parameter estimates. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB'99*, volume 4, pages 53–64, Singapore, 1999. World Scientific Publishing.

[39] J. Euzenat, C. Chemla, and B. Jacq. 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, editors, *Proceedings of the Fifth International Conference on Intelligent Systems for Molecular Biology, ISMB-97*, pages 108–119, Menlo Park, CA, 1997. AAAI Press.

[40] S. Fields, Y. Kohara, and D.J. Lockhart. Functional genomics. *Proceedings of the National Academy of Sciences of the USA*, 96:8825–8826, 1999.

[41] K.D. Forbus. Qualitative process theory. *Artificial Intelligence*, 24:85–168, 1984.

[42] N. Friedman, M. Linial, I. Nachman, and D. Pe'er. Using Bayesian networks to analyze expression data. In *Proceedings of the Fourth Annual International Conference on Computational Molecular Biology, RECOMB 2000*, New York, N.Y., 2000. ACM Press.

[43] S. Fuhrman, X. Wen, G.S. Michaels, and R. Somogyi. Genetic network inference. *InterJournal*, 104, 1997.

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

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

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

[47] A. Gierer. Generation of biological patterns and form: Some physical, mathematical, and logical aspects. *Progress in Biophysics and Molecular Biology*, 37:1–47, 1981.

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

[49] D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.

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

[51] L. Glass. Classification of biological networks by their qualitative dynamics. *Journal of Theoretical Biology*, 54:85–107, 1975.

[52] L. Glass. Combinatorial and topological methods in nonlinear chemical kinetics. *Journal of Chemical Physics*, 63(4):1325–1335, 1975.

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

[54] L. Glass and C. Hill. Ordered and disordered dynamics in random networks. *Europhysics Letters*, 41(6):599–604, 1998.

[55] L. Glass and S.A. Kauffman. Co-operative components, spatial localization and oscillatory cellular dynamics. *Journal of Theoretical Biology*, 34:219–237, 1972.

[56] L. Glass and S.A. Kauffman. The logical analysis of continuous non-linear biochemical control networks. *Journal of Theoretical Biology*, 39:103–129, 1973.

[57] L. Glass and J.S. Pasternack. Prediction of limit cycles in mathematical models of biological oscillations. *Bulletin of Mathematical Biology*, 40:27–44, 1978.

[58] L. Glass and J.S. Pasternack. Stable oscillations in mathematical models of biological control systems. *Journal of Mathematical Biology*, 6:207–223, 1978.

[59] L. Glass and R. Pérez. Limit cycle oscillations in compartmental chemical systems. *Journal of Chemical Physics*, 61(12):5242–5249, 1974.

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

[61] B.C. Goodwin. *Temporal Organization in Cells*. Academic Press, New York, N.Y., 1963.

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

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

[64] B.C. Goodwin and S.A. Kauffman. Spatial harmonics and pattern specification in early *Drosophila* development: Part I. Bifurcation sequences and gene expression. *Journal of Theoretical Biology*, 144:303–319, 1990.

[65] P.J.E. Goss and J. Peccoud. Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. *Proceedings of the National Academy of Sciences of the USA*, 95:6750–6755, 1998.

[66] J.-L. Gouzé. Positive and negative circuits in dynamical systems. *Journal of Biological Systems*, 6(1):11–15, 1998.

[67] J. Griffith. Mathematics of cellular control processes: I. Negative feedback to one gene. *Journal of Theoretical Biology*, 20:202–208, 1968.

[68] J. Griffith. Mathematics of cellular control processes: II. Positive feedback to one gene. *Journal of Theoretical Biology*, 20:209–216, 1968.

[69] S. Hamahashi and H. Kitano. Simulation of *Drosophila* embryogenesis. In C. Adami, R.K. Belew, H. Kitano, and C. Taylor, editors, *Proceedings of the Sixth International Conference on Artificial Life, Artificial Life VI*, pages 151–160, Cambridge, MA, 1998. MIT Press.

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

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

[72] K.R. Heidtke and S. Schulze-Kremer. BioSim: A new qualitative simulation environment for molecular biology. In J. Glasgow, T. Littlejohn, and F. Major, editors, *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology, ISMB-98*, pages 85–94, Menlo Park, CA, 1998. AAAI Press.

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

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

[75] R. Hofestädt. A simulation shell to model metabolic pathways. *Systems Analysis, Modeling, and Simulation*, 11:253–262, 1993.

[76] R. Hofestädt and F. Meineke. Interactive modelling and simulation of biochemical networks. *Computers in Biology and Medicine*, 25(3):321–334, 1995.

[77] R. Hofestädt and U. Scholz. Information processing for the analysis of metabolic pathways and inborn errors. *BioSystems*, 47:91–102, 1998.

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

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

[80] A. Hunding, S.A. Kauffman, and B.C. Goodwin. *Drosophila* segmentation: Supercomputer simulation of prepattern hierarchy. *Journal of Theoretical Biology*, 145:369–384, 1990.

[81] T.E. Ideker, V. Thorsson, and R.M. Karp. Discovery of regulatory interactions through perturbation: Inference amd experimental design. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB 2000*, volume 5, pages 302–313, Singapore, 2000. World Scientific Publishing.

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

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

[84] G.K. Kananyan, V.A. Ratner, and R.N. Tchuraev. Enlarged model of lambda phage ontogenesis. *Journal of Theoretical Biology*, 90:301–315, 1981.

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

[86] M. Kanehisa and S. Goto. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research*, 28(1):27–30, 2000.

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

[88] P.D. Karp. Design methods for scientific hypothesis formation and their application to molecular biology. *Machine Learning*, 12:89–116, 1993.

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

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

[91] R.M. Karp, R. Stoughton, and K.Y. Yeung. Algorithms for choosing differential gene expression experiments. In *Proceedings of the Third Annual International Conference on Computational Molecular Biology, RECOMB'99*, pages 208–217, New York, N.Y., 1999. ACM Press.

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

[93] S.A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.

[94] S.A. Kauffman. The large-scale structure and dynamics of gene control circuits: An ensemble approach. *Journal of Theoretical Biology*, 44:167–190, 1974.

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

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

[97] S.A. Kauffman. Antichaos and adaptation. *Scientific American*, 265(august):78–84, 1991.

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

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

[100] H. Kitano, S. Hamahashi, J. Kitazawa, K. Takao, and S. Imai. The Virtual Biology Laboratories: A new approach to computational biology. In P. Husbands and I. Harvey, editors, *Proceedings of the Fourth European Conference on Artificial Life*, pages 274–283, Cambridge, MA, 1997. MIT Press.

[101] K. Koile and C.G. Overton. A qualitative model for gene expression. In *Proceedings of the 1989 Summer Computer Simulation Conference*, pages 415–421. Society for Computer Simulation, 1989.

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

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

[104] B. Kuipers. Qualitative simulation. *Artificial Intelligence*, 29:289–388, 1986.

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

[106] K.M. Kyoda and H. Kitano. A model of axis determination for the *Drosophila* wing disc. In D. Floreano, J.-D. Nicoud, and F. Mondada, editors, *Advances in Artificial Life. Proceedings of the Fifth European Conference, ECAL'99*, volume 1674 of *Lecture Notes in Computer Science*, pages 472–476, Berlin, 1999. Springer-Verlag.

[107] K.M. Kyoda and H. Kitano. Simulation of genetic interactions for *Drosophila* leg formation. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB'99*, volume 4, pages 77–89, Singapore, 1999. World Scientific Publishing.

[108] K.M. Kyoda, M. Muraki, and H. Kitano. 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, editors, *Pacific Symposium on Biocomputing, PSB 2000*, volume 5, pages 314–325, Singapore, 2000. World Scientific Publishing.

[109] T.C. Lacalli. Modeling the *drosophila* pair-rule pattern by reaction diffusion: Gap input and pattern control in a 4-morphogen system. *Journal of Theoretical Biology*, 144:171–194, 1990.

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

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

[112] H.D. Landahl. Some conditions for sustained oscillations in biochemical chaines. *Bulletin of Mathematical Biophysics*, 31:775–787, 1969.

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

[114] B. Lewin. *Genes VII.* Oxford University Press, Oxford, 1999.

[115] J.E. Lewis and L. Glass. Steady states, limit cycles, and chaos in models of complex biological networks. *International Journal of Bifurcation and Chaos*, 1(2):477–483, 1991.

[116] J.E. Lewis and L. Glass. Nonlinear dynamics and symbolic dynamics of neural networks. *Neural Computation*, 4:621–642, 1992.

[117] S. Liang, S. Fuhrman, and R. Somogyi. 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, editors, *Pacific Symposium on Biocomputing, PSB'98*, volume 3, pages 18–29, Singapore, 1998. World Scientific Publishing.

[118] R.J. Lipschutz, S.P.A. Fodor, T.R. Gingeras, and D.J. Lockhart. High density synthetic oligonucleotide arrays. *Nature Genetics*, 21(supplement):20–24, 1999.

[119] L. Ljung. *System Identification: Theory for the User*. Prentice Hall, Upper Saddle River, NJ, 2nd edition, 1999.

[120] W.F. Loomis and P.W. Sternberg. Genetic networks. *Science*, 269:649, 1995.

[121] J. Maddox. Is molecular biology yet a science? *Nature*, 355:201, 1992.

[122] P.K. Maini, K.J. Painter, and H. Nguyen Phong Chau. Spatial pattern formation in chemical and biological systems. *Journal of the Chemical Society, Faraday Transactions*, 93(20):3601–3610, 1997.

[123] G. Marnellos, G.A. Deblandre, E. Mjolsness, and G. Kintner. 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, editors, *Pacific Symposium on Biocomputing, PSB 2000*, volume 5, pages 326–337, Singapore, 2000. World Scientific Publishing.

[124] G. Marnellos and E. Mjolsness. A gene network approach to modeling early neurogenesis in *Drosophila*. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB'98*, volume 3, pages 30–41, Singapore, 1998. World Scientific Publishing.

[125] H. Matsuno, A. Doi, M. Nagasaki, and S. Miyano. Hybrid Petri net representation of gene regulatory network. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB 2000*, volume 5, pages 338–349, Singapore, 2000. World Scientific Publishing.

[126] H.H. McAdams and A. Arkin. Simulation of prokaryotic genetic circuits. *Annual Review of Biophysics and Biomolecular Structure*, 27:199–224, 1998.

[127] H.H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269:650–656, 1995.

[128] H.M. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the USA*, 94:814–819, 1997.

[129] H. Meinhardt. A model of pattern formation in insect embryogenesis. *Journal of Cell Science*, 23:117–139, 1977.

[130] H. Meinhardt. *Models of Biological Pattern Formation*. Academic Press, London, 1982.

[131] H. Meinhardt. Hierarchical inductions of cell states: A model for segmentation in *Drosophila*. *Journal of Cell Science*, 4(supplement):357–381, 1986.

[132] H. Meinhardt. Models for maternally supplied positional information and the activation of segmentation genes in *Drosophila* embryogenesis. *Development*, 104(supplement):95–110, 1988.

[133] P. Mendes. GEPASI: A software package for modelling the dynamics, steady states and control of biochemical and other systems. *Computer Applications in the Biosciences*, 9(5):563–571, 1993.

[134] L. Mendoza and E.R. Alvarez-Buylla. Dynamics of the genetic regulatory network for *arabidopsis thaliana*. *Journal of Theoretical Biology*, 193:307–319, 1998.

[135] L. Mendoza, D. Thieffry, and E.R. Alvarez-Buylla. Genetic control of flower morphogenesis in *arabidopsis thaliana*: A logical analysis. *Bioinformatics*, 15(7-8):593–606, 1999.

[136] T. Mestl, R.J. Bagley, and L. Glass. Common chaos in arbitrarily complex feedback networks. *Physical Review Letters*, 79(4):653–656, 1997.

[137] T. Mestl, C. Lemay, and L. Glass. Chaos in high-dimensional neural and gene networks. *Physica D*, 98:33–52, 1996.

[138] T. Mestl, E. Plahte, and S.W. Omholt. A mathematical framework for describing and analysing gene regulatory networks. *Journal of Theoretical Biology*, 176:291–300, 1995.

[139] T. Mestl, E. Plahte, and S.W. Omholt. Periodic solutions in systems of piecewise-linear differential equations. *Dynamics and Stability of Systems*, 10(2):179–193, 1995.

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

[141] G.S. Michaels, D.B. Carr, M. Askenazi, S. Fuhrman, X. Wen, and R. Somogyi. Cluster analysis and data visualization of large-scale gene expression data. In *Pacific Symposium on Biocomputing, PSB'98*, volume 3, pages 42–53, Singapore, 1998. World Scientific Publishing.

[142] E. Mjolsness, D.H. Sharp, and J. Reinitz. A connectionist model of development. *Journal of Theoretical Biology*, 152:429–453, 1991.

[143] J. Monod and F. Jacob. General conclusions: Teleonomic mechanisms in cellular metabolism, growth, and differentiation. In *Cold Spring Harbor Symposium on Quantitative Biology*, volume 26, pages 389–401, Cold Spring Harbor, NY, 1961.

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

[145] B.N. Nagorcka. A pattern formation mechanism to control spatial organization in the embryo of *Drosophila melanogaster*. *Journal of Theoretical Biology*, 132:277–306, 1988.

[146] P.P. Nayak. *Automated Modeling of Physical Systems*. Springer Verlag, Berlin, 1995.

[147] S.A. Newman and S.A. Rice. Model for constraint and control in biochemical networks. *Proceedings of the National Academy of Sciences of the USA*, 68(1):92–96, 1971.

[148] G. Nicolis. Mathematical aspects of biological regulatory processes. In R. Thomas, editor, *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, volume 29 of *Lecture Notes in Biomathematics*, pages 178–211. Springer-Verlag, Berlin, 1979.

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

[150] K. Noda, A. Shinohara, M. Takeda, S. Matsumoto, S. Miyano, and S. Kuhara. Finding genetic network from experiments by weighted network model. In *Genome Informatics*, volume 9, pages 141–150, 1998.

[151] B. Novak and J.J. Tyson. Modeling the cell division cycle: M-phase trigger, oscillations, and size control. *Journal of Theoretical Biology*, 165:101–134, 1993.

[152] B. Novak and J.J. Tyson. Numerical analysis of a comprehensive model of M-phase control in Xenopus oocyte extracts and intact embryos. *Journal of Cell Science*, 106:1153–1168, 1993.

[153] B. Novak and J.J. Tyson. Quantitative analysis of a molecular model of mitotic control in fission yeast. *Journal of Theoretical Biology*, 173:283–305, 1995.

[154] R. Nowak. Entering the postgenome era. *Science*, 270:368–371, 1995.

[155] M.N. Obeyesekere, S. Tucker, and S.O. Zimmerman. Mathematical models for the cellular concentrations of Cyclin and MPF. *Biochemical and Biophysical Research Communications*, 184(2):782–789, 1992.

[156] S.W. Omholt, X. Kefang, Ø. Andersen, and E. Plahte. Description and analysis of switchlike regulatory networks exemplified by a model of cellular iron homeostasis. *Journal of Theoretical Biology*, 195:339–350, 1998.

[157] H.G. Othmer. The qualitative dynamics of a class of biochemical control circuits. *Journal of Mathematical Biology*, 3:53–78, 1976.

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

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

[160] E. Plahte, T. Mestl, and S.W. Omholt. Global analysis of steady points for systems of differential equations with sigmoid interactions. *Dynamics and Stability of Systems*, 9(4):275–291, 1994.

[161] E. Plahte, T. Mestl, and S.W. Omholt. Feedback loops, stability and multistationarity in dynamical systems. *Journal of Biological Systems*, 3(2):409–413, 1995.

[162] E. Plahte, T. Mestl, and S.W. Omholt. Stationary states in food web models with threshold relationships. *Journal of Biological Systems*, 3(2):569–577, 1995.

[163] E. Plahte, T. Mestl, and S.W. Omholt. A methodological basis for description and analysis of systems with complex switch-like interactions. *Journal of Mathematical Biology*, 36:321–348, 1998.

[164] E.I. Prokudina, R.Y. Valeev, and R.N. Tchuraev. 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. *Journal of Theoretical Biology*, 151:89–110, 1991.

[165] V.A. Ratner and R.N. Tchuraev. Simplest genetic systems controlling ontogenesis: Organization principles and models of their function. In R. Rosen, editor, *Progress in Theoretical Biology*, volume 5, pages 81–127. Academic Press, New York, 1978.

[166] V.N. Reddy, M.N. Liebman, and M.L. Mavrovouniotis. Qualitative analysis of biochemical reaction systems. *Computers in Biology and Medicine*, 26(1):9–24, 1996.

[167] J. Reinitz, D. Kosman, C.E. Vanario-Alonso, and D.H. Sharp. Stripe forming architecture of the gap gene system. *Developmental Genetics*, 23:11–27, 1998.

[168] J. Reinitz, E. Mjolsness, and D.H. Sharp. Model for cooperative control of positional information in *Drosophila* by bicoid and maternal hunchback. *Journal of Experimental Zoology*, 271:47–56, 1995.

[169] J. Reinitz and D.H. Sharp. Mechanism of *eve* stripe formation. *Mechanisms of Development*, 49:133–158, 1995.

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

[171] J. Reinitz and J.R. Vaisnys. Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of co-operativity. *Journal of Theoretical Biology*, 145:295–318, 1990.

[172] J. Richelle. Comparative analysis of negative loops by continuous, boolean and stochastic approaches. In R. Thomas, editor, *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, volume 29 of *Lecture Notes in Biomathematics*, pages 281–325. Springer-Verlag, Berlin, 1979.

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

[174] R. Rosen. Recent developments in the theory of control and regulation of cellular processes. *International Review of Cytology*, 23:25–88, 1968.

[175] M.A. Russell. Positional information in insect segments. *Developmental Biology*, 108:269–283, 1985.

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

[177] M. Samsonova and V.N. Serov. 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, editors, *Pacific Symposium on Biocomputing, PSB'99*, volume 4, pages 102–111, Singapore, 1999. World Scientific Publishing.

[178] M.G. Samsonova, E.G. Savostyanova, V.N. Serov, A.V. Spirov, and J. Reinitz. GeNet: A database of genetic networks. In *Proceedings of the First International Conference on Bioinformatics of Genome Regulation and Structure, BGRS'98*, Novosibirsk, 1998. ICG.

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

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

[181] H.M. Sauro. SCAMP: A general-purpose simulator and metabolic control analysis program. *Computer Applications in the Biosciences*, 9(4):441–450, 1993.

[182] M.A. Savageau. Biochemical systems analysis. II The steady state solutions for an n-pool system using a power-law approximation. *Journal of Theoretical Biology*, 25:370–379, 1969.

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

[184] M.A. Savageau. 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, editors, *Integrative Approaches to Molecular Biology*, pages 115–146. MIT Press, Cambridge, MA, 1996.

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

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

[187] M.A. Shea and G.K. Ackers. The O_R control system of bacteriophage lambda: A physical-chemical model for gene regulation. *Journal of Molecular Biology*, 181:211–230, 1985.

[188] T. Shimada, M. Hagiya, M. Arita, S. Nishizaki, and C.L. Tan. Knowledge-based simulation of regulatory action in lambda phage. *International Journal of Artificial Intelligence Tools*, 4(4):511–524, 1995.

[189] P. Smolen, D.A. Baxter, and J.H. Byrne. Modeling transcriptional control in gene networks: Methods, recent results, and future directions. *Bulletin of Mathematical Biology*, 62:247–292, 2000.

[190] E.H. Snoussi. Qualitative dynamics of piecewise-linear differential equations: A discrete mapping approach. *Dynamics and Stability of Systems*, 4(3-4):189–207, 1989.

[191] E.H. Snoussi. Necessary conditions for multistationarity and stable periodicity. *Journal of Biological Systems*, 6(1):3–9, 1998.

[192] E.H. Snoussi and R. Thomas. Logical identification of all steady states: The concept of feedback loop characteristic states. *Bulletin of Mathematical Biology*, 55(5):973–991, 1993.

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

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

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

[196] R.C. Strohman. The coming Kuhnian revolution in biology. *Nature Biotechnology*, 15:194–200, 1997.

[197] M. Sugita. Functional analysis of chemical systems *in vivo* using a logical circuit equivalent. *Journal of Theoretical Biology*, 1:179–192, 1961.

[198] M. Sugita. Functional analysis of chemical systems *in vivo* using a logical circuit equivalent: II. The idea of a molecular automaton. *Journal of Theoretical Biology*, 4:179–192, 1963.

[199] Z. Szallasi. 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, editors, *Pacific Symposium on Biocomputing, PSB'99*, volume 4, pages 5–16, Singapore, 1999. World Scientific Publishing.

[200] Z. Szallasi and S. Liang. 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, editors, *Pacific Symposium on Biocomputing, PSB'98*, volume 3, Singapore, 1998. World Scientific Publishing.

[201] S. Tavazoie, J.D. Hughes, M.J. Campbell, R.J. Cho, and G.M. Church. Systematic determination of genetic network architecture. *Nature Genetics*, 22:281–285, 1999.

[202] R.N. Tchuraev. A new method for the analysis of the dynamics of the molecular genetic control systems: I. Description of the method of generalized threshold models. *Journal of Theoretical Biology*, 151:71–87, 1991.

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

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

[205] D. Thieffry, M. Colet, and R. Thomas. Formalisation of regulatory nets: A logical method and its automatization. *Mathematical Modelling and Scientific Computing*, 2:144–151, 1993.

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

[207] D. Thieffry, E.H. Snoussi, J. Richelle, and R. Thomas. Positive loops and differentiation. *Journal of Biological Systems*, 3(2):457–466, 1995.

[208] D. Thieffry and R. Thomas. Dynamical behaviour of biological networks: II. Immunity control in bacteriophage lambda. *Bulletin of Mathematical Biology*, 57(2):277–297, 1995.

[209] R. Thomas. Boolean formalization of genetic control circuits. *Journal of Theoretical Biology*, 42:563–585, 1973.

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

[211] R. Thomas. 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, editors, *Numerical Methods in the Study of Critical Phenomena*, pages 180–193. Springer Verlag, Berlin, 1981.

[212] R. Thomas. Regulatory networks seen as asynchronous automata: A logical description. *Journal of Theoretical Biology*, 153:1–23, 1991.

[213] R. Thomas. Laws for the dynamics of regulatory networks. *International Journal of Developmental Biology*, 42:479–485, 1998.

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

[215] R. Thomas, A.-M. Gathoye, and L. Lambert. A complex control circuit: Regulation of immunity in temperate bacteriophages. *European Journal of Biochemistry*, 71:211–227, 1976.

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

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

[218] R.B. Trelease, R.A. Henderson, and J.B. Park. A qualitative process system for modeling NF-κB and AP-1 gene regulation in immune cell biology research. *Artifical Intelligence in Medicine*, 17:303–321, 1999.

[219] A.M. Turing. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London, Series B*, 237:37–72, 1951.

[220] J.J. Tyson. On the existence of oscillatory solutions in negative feedback cellular control processes. *Journal of Mathematical Biology*, 1:311–315, 1975.

[221] J.J. Tyson. Models of cell cycle control in eukaryotes. *Journal of Biotechnology*, 71(1-3):239–244, 1999.

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

[223] J.J. Tyson and H.G. Othmer. The dynamics of feedback control circuits in biochemical pathways. *Progress in Theoretical Biology*, 5:1–62, 1978.

[224] H. Ueda and H. Kitano. A generalized reaction-diffusion simulator for studying the molecular basis of pattern formation in biological systems. In C. Adami, R.K. Belew, H. Kitano, and C. Taylor, editors, *Proceedings of the Sixth International Conference on Artificial Life, Artificial Life VI*, pages 462–466, Cambridge, MA, 1998. MIT Press.

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

[226] N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, revised edition, 1997.

[227] E.P. van Someren, L.F.A. Wessels, and M.J.T. Reinders. 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, editors, *Proceedings of the Eight International Conference on Intelligent Systems for Molecular Biology, ISMB 2000*, pages 355–366, Menlo Park, CA, 2000. AAAI Press.

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

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

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

[231] C. Walter. The occurrence and the significance of limit cycle behavior in controlled biochemical systems. *Journal of Theoretical Biology*, 27:259–272, 1970.

[232] C. Walter, R. Parker, and M. Yčas. A model for binary logic in biochemical systems. *Journal of Theoretical Biology*, 15:208–217, 1967.

[233] D.C. Weaver, C.T. Workman, and G.D. Stormo. Modeling regulatory networks with weight matrices. In R.B. Altman, K. Lauderdale, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB'99*, volume 4, pages 112–123, Singapore, 1999. World Scientific Publishing.

[234] G. Weisbuch. Networks of automata and biological organization. *Journal of Theoretical Biology*, 121:255–267, 1986.

[235] X. Wen, S. Fuhrman, G.S. Michaels, D.B. Carr, S. Smith, J.L. Barker, and R. Somogyi. Large-scale temporal gene expresion mapping of central nervous system development. *Proceedings of the National Academy of Sciences of the USA*, 95:334–339, 1998.

[236] A. Wuensche. Genomic regulation modeled as a network with basins of attraction. In R.B. Altman, A.K. Dunker, L. Hunter, and T.E. Klein, editors, *Pacific Symposium on Biocomputing, PSB'98*, volume 3, pages 89–102, Singapore, 1998. World Scientific Publishing.

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

[238] G. Yagil and E. Yagil. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophysical Journal*, 11:11–27, 1971.

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

[240] A. Zien, R. Küffner, R. Zimmer, and T. Lengauer. 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, editors, *Proceedings of the Eight International Conference on Intelligent Systems for Molecular Biology, ISMB 2000*, pages 407–417, Menlo Park, CA, 2000. AAAI Press.

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe - 38330 Montbonnot-St-Martin (France)

Unité de recherche INRIA Lorraine : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex (France)
Unité de recherche INRIA Rennes : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex (France)
Unité de recherche INRIA Rocquencourt : Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)
Unité de recherche INRIA Sophia Antipolis : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex (France)

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399