Qualitative Simulation of Genetic Regulatory Networks: Method and Application

Hidde de Jong,[1] **Michel Page,**[1,2] **Céline Hernandez,**[1] **and Johannes Geiselmann**[3]

[1]INRIA Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier CEDEX, France
[2]ESA, Université Pierre Mendès France, Grenoble [3]PEGM, Université Joseph Fourier, Grenoble, France

{Hidde.de-Jong, Céline.Hernandez, Michel.Page}@inrialpes.fr, Johannes.Geiselmann@ujf-grenoble.fr

Abstract

Computer modeling and simulation are indispensable for understanding the functioning of an organism on a molecular level. We present an implemented method for the qualitative simulation of large and complex genetic regulatory networks. The method allows a broad range of regulatory interactions between genes to be represented and has been applied to the analysis of a real network of biological interest, the network controlling the initiation of sporulation in the bacterium *B. subtilis.*

1 Introduction

It is now commonly accepted in biology that most interesting properties of an organism emerge from the interactions among its genes, proteins, metabolites, and other molecules. This implies that, in order to understand the functioning of an organism, the networks of interactions involved in gene regulation, metabolism, signal transduction, and other cellular and intercellular processes need to be elucidated.

A *genetic regulatory network* consists of a set of genes and their mutual regulatory interactions. The interactions arise from the fact that genes code for proteins that may control the expression of other genes, for instance by activating or inhibiting DNA transcription [Lewin, 1999]. The study of genetic regulatory networks has received a major impetus from the recent development of experimental techniques permitting the spatiotemporal expression levels of genes to be rapidly measured in a massively parallel way [Brown and Botstein, 1999]. However, in addition to experimental tools, computer tools for the modeling and simulation of gene regulation processes will be indispensable. As most genetic regulatory systems of interest involve many genes connected through interlocking positive and negative feedback loops, an intuitive understanding of their dynamics is hard to obtain.

Currently, only a few regulatory networks are well-understood on the molecular level, and quantitative information about the interactions is seldom available. This has stimulated an interest in modeling and simulation techniques developed within qualitative reasoning (QR) [Heidtke and Schulze-Kremer, 1998; Trelease *et al.*, 1999]. A major problem with these approaches, based on well-known methods like QSIM [Kuipers, 1994] and QPT [Forbus, 1984], is their lack of upscalability. Following approaches in mathematical biology, de Jong and Page [2000] have proposed a qualitative simulation method capable of handling large and complex networks.

The aim of this paper is to generalize the latter method and to demonstrate its applicability to real networks of biological interest. The generalization of the method allows a broader range of regulatory interactions between genes to be expressed. This enables more complex systems to be analyzed, such as the network of interactions controlling the initiation of sporulation in the bacterium *Bacillus subtilis.* We have simulated the sporulation network using a model constructed from published reports of experiments. The simulations reveal that an additional interaction, proposed in the literature before but not yet experimentally identified, may be involved.

In the next section, we will discuss the class of equations being used to model genetic regulatory networks. The third section describes the qualitative simulation algorithm, focusing on the representation of the qualitative state of a regulatory system and the determination of state transitions by the simulation algorithm. The subsequent sections present the results of the analysis of the sporulation network as well as a discussion of the method in the context of related work.

2 Modeling genetic regulatory networks

2.1 Approximations of regulatory interactions

In order to model a genetic regulatory network, we first have to describe the regulatory interactions in an empirically valid and mathematically rigorous way. Consider a DNA-binding protein encoded by gene j, activating the expression of a target gene i. The rate of transcription of i as a function of the concentration x_j of the regulatory protein follows a sigmoid curve [Yagil and Yagil, 1971]. Below a threshold concentration $\theta_j > 0$ the gene is hardly expressed at all, whereas above this threshold its expression rapidly saturates.

Sigmoid curves are also found in the case of more complex regulatory mechanisms. Consider the proteins J and K that form a dimer repressing the transcription of gene i (Fig. 2(b)). Analysis of a kinetic model of this regulatory mechanism reveals that the rate of expression of i depends in a sigmoidal fashion on the total concentrations x_j and x_k of J and K, respectively. That is, both J and K need to be available above

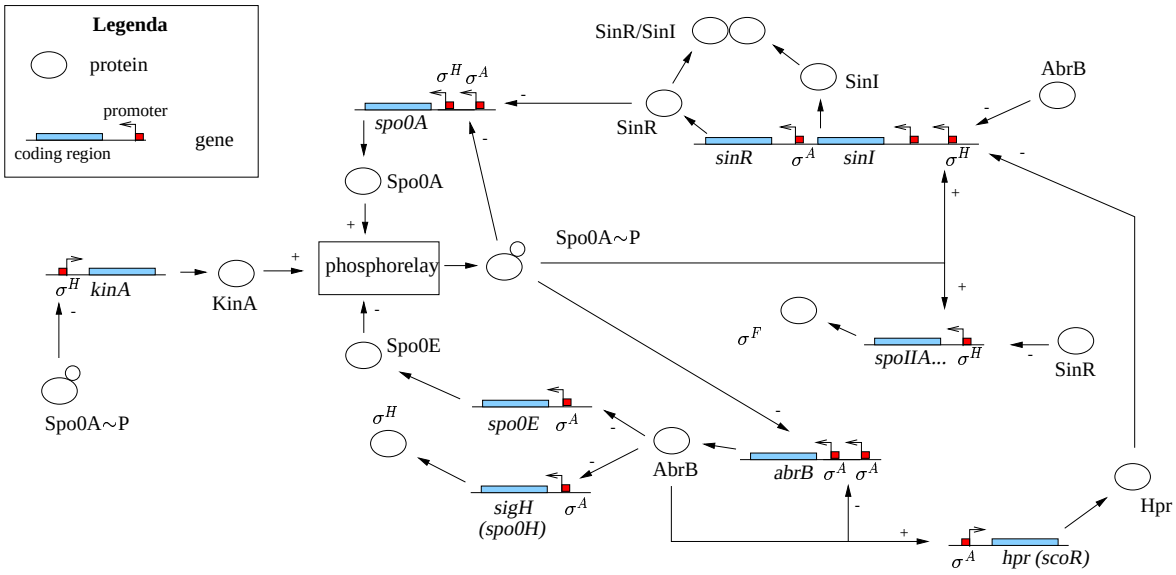


Figure 1: Genetic regulatory network underlying the inititation of sporulation in *B. subtilis.* For every gene, the coding region and the promoters are shown. Promoters are distinguished by the specific σ factor directing DNA transcription. The regulatory action of a protein tending to activate (inhibit) expression is indicated by a '+' ('-'). As a notational convention, names of genes are printed in italic and names of proteins start with a capital.

their threshold concentrations for i to be repressed.

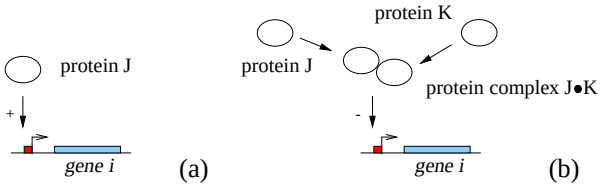


Figure 2: (a) Activation of a target gene i by a regulatory protein J. (b) Inhibition of i by a protein complex J•K.

In the case of steep sigmoids, the combined effect of regulatory proteins on gene expression can be approximated by means of Boolean functions [Kauffman, 1993; Thomas and d'Ari, 1990]. Here we will rewrite a Boolean function in terms of step functions and arithmetic sums and multiplications, following the procedure of Plahte *et al.* [1998]. For the activator protein in Fig. 2(a), we thus obtain a *regulation function* $r(x_j) = \kappa\, s^+(x_j, \theta_j)$, where $s^+()$ is a step function (Fig. 5) and $\kappa > 0$ a rate parameter. Similarly, the effect of a regulatory protein repressing gene i can be described by $r(x_j) = \kappa\, s^-(x_j, \theta_j)$, with $s^-(x_j, \theta_j) = 1 - s^+(x_j, \theta_j)$. The dimer repressor example in Fig. 2(b) leads to the function $r_i(x_j, x_k) = \kappa_i\,(1 - s^+(x_j, \theta_j)\, s^+(x_k, \theta_k))$.

Although the above discussion has focused on the representation of interactions regulating the synthesis of proteins, it also applies to the degradation of proteins. Sigmoid relations are observed in the latter case as well, so that the logical approximations are valid. In order to formally distinguish protein degradation rates from protein synthesis rates, we will denote the former by γ_i instead of κ_i.

2.2 State equations

The dynamics of genetic regulatory networks can be modeled by a system of differential equations suggested by Mestl *et al.* [1995], extending earlier proposals by Glass and Kauffman [1973] and Thomas *et al.* [1990].

$$\dot{x}_i = f_i(\boldsymbol{x}) - g_i(\boldsymbol{x})\, x_i, \quad x_i \geq 0,\ 1 \leq i \leq n, \qquad (1)$$

where $\boldsymbol{x}$ is a vector of cellular protein concentrations. The *state equations* (1) define the rate of change of the concentration x_i as the difference of the rate of synthesis $f_i(\boldsymbol{x})$ and the rate of degradation $-g_i(\boldsymbol{x})x_i$ of the protein. Exogenous variables can be defined by setting $\dot{x}_i = 0$.

The function $f_i : \mathbb{R}^n_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ is defined as

$$f_i(\boldsymbol{x}) = \sum_{l \in L} r_{il}(\boldsymbol{x}) > 0, \qquad (2)$$

where $r_{il}()$ is a regulation function and L a possibly empty set of indices of regulation functions. The function $g_i()$ is defined analogously to (2), except that for reasons that will become clear below, we demand that $g_i(\boldsymbol{x})$ is strictly positive. Notice that for the above definitions of $f_i()$ and $g_i()$, the state equations (1) are *piecewise-linear*. Equations of this form, and their logical abstractions, have been well-studied in mathematical biology (e.g., [Lewis and Glass, 1991; Mestl *et al.*, 1995; Thomas and d'Ari, 1990]).

Eqs. (1) and (2) generalize upon the formalism employed in [de Jong and Page, 2000] in two respects. First, the regulation functions $r_{il}()$ may be the mathematical equivalent of any Boolean function, whereas in the earlier paper it was restricted to logical functions composed of Boolean products. Second, the regulation of protein degradation can now

State equation for Spo0E:

$$\dot{x}_{se} = \kappa_{se}\, s^-(x_{ab}, \theta_{ab_1}) s^+(x_a, \theta_{a_1}) - \gamma_{se} x_{se},$$

Threshold inequalities:

$$0 < \theta_{se_1} < \theta_{se_2} < \theta_{se_3} < max_{se}$$

Equilibrium inequalities: $0 < \kappa_{se}/\gamma_{se} < \theta_{se_1}$ (a)

State equation for AbrB:

$$\dot{x}_{ab} = \kappa_{ab}\, s^-(x_{ab}, \theta_{ab_2})\, s^+(x_a, \theta_{a_1}) \cdot (1 - s^+(x_{sa}, \theta_{sa_1})\, s^+(x_{ka}, \theta_{ka_1})\, s^-(x_{se}, \theta_{se_3})) - \gamma_{ab}\, x_{ab}$$

Threshold inequalities:

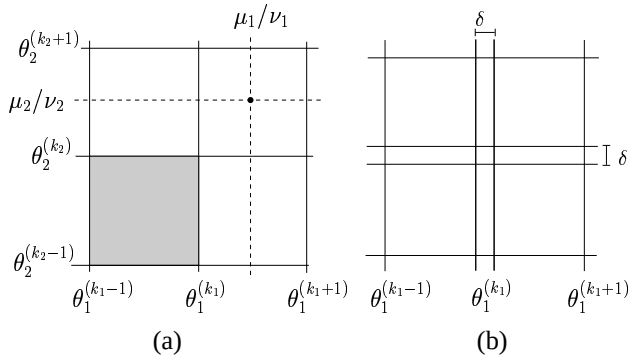
$$0 < \theta_{ab_1} < \theta_{ab_2} < max_{ab}$$

Equilibrium inequalities: $\theta_{ab_2} < \kappa_{ab}/\gamma_{ab} < max_{ab}$ (b)

Figure 3: State equations, threshold inequalities, and equilibrium inequalities for the genes (a) *spo0E* and (b) *abrB* in Fig. 1. The subscripts in the equations refer to the sporulation genes *kinA* (ka), *spo0E* (se), *spo0A* (sa), *sigA* (a), *sigH* (h), and *abrB* (ab).

be modeled, whereas before the degradation rate was set to $g_i(\boldsymbol{x}) = \gamma_i$. These extensions allow the structure of complex regulatory networks to be formalized in a convenient way.

In Fig. 3 the state equations corresponding to two of the genes in the sporulation network of Fig. 1 are shown. The differential equation in (a) states that *spo0E* is transcribed at a rate κ_{se} from a σ^A-promoter when its repressor AbrB is below its threshold concentration θ_{ab_1} (i.e., $s^-(x_{ab}, \theta_{ab_1}) = 1$). In addition, for transcription to commence, the sigma factor σ_A encoded by *sigA* needs to be available at a concentration above the threshold θ_{sa_1} (i.e., $s^+(x_{sa}, \theta_{sa_1}) = 1$). Spo0E degrades at a rate proportional to its own concentration.



(a) (b)

Figure 4: (a) Two-dimensional phase space divided into regulatory domains by the threshold planes. The shaded regulatory domain defined by $\theta_1^{(k_1-1)} < x_1 < \theta_1^{(k_1)}$ and $\theta_2^{(k_2-1)} < x_2 < \theta_2^{(k_2)}$ has a target equilibrium in an adjacent regulatory domain. (b) The same phase space with switching zones around the threshold planes.

2.3 Threshold and equilibrium inequalities

In general, a protein encoded by a gene will be involved in different interactions at different threshold concentrations. Although exact numerical values will not usually be available, we can order the p_i threshold concentrations of gene i, which gives the *threshold inequalities*

$$0 < \theta_i^{(1)} < \ldots < \theta_i^{(p_i)} < max_i. \tag{3}$$

The parameter max_i denotes a maximum concentration for the protein denoted by i.

For the sporulation gene *abrB* two thresholds are defined, θ_{ab_1} and θ_{ab_2}. AbrB has two threshold concentrations: θ_{ab_1} and θ_{ab_2}. The first threshold corresponds to the repression of *spo0E*, *sigH*, and other early sporulation genes by AbrB. The second threshold corresponds to the autoregulation of *abrB* during vegetative growth, when AbrB levels are at their highest. This motivates the ordering of the AbrB thresholds in Fig. 3(b): $\theta_{ab_1} < \theta_{ab_2}$.

The $n-1$-dimensional threshold hyperplanes $x_i = \theta_i^{(k_i)}$, $1 \le k_i \le p_i$ divide the phase space box into regions, called *regulatory domains* (Fig. 4). Within each regulatory domain, the step function expressions in (2) can be evaluated, which reduces $f_i()$ and $g_i()$ to sums of rate constants. That is, $f_i()$ simplifies to some $\mu_i \in M_i \equiv \{f_i(\boldsymbol{x}) \mid \boldsymbol{0} \le \boldsymbol{x} \le \boldsymbol{max}\}$, and $g_i()$ to some $\nu_i \in N_i \equiv \{g_i(\boldsymbol{x}) \mid \boldsymbol{0} \le \boldsymbol{x} \le \boldsymbol{max}\}$. The sets M_i and N_i collect the different synthesis and degradation rates of the protein in different domains of the phase space.

It can be easily shown that all trajectories in a regulatory domain tend towards a single, stable steady state $\boldsymbol{x} = \boldsymbol{\mu/\nu}$, the *target equilibrium*, lying at the intersection of the hyperplanes $x_i = \mu_i/\nu_i$ [Glass and Kauffman, 1973; Mestl *et al.*, 1995; Thomas and d'Ari, 1990]. The target equilibrium level μ_i/ν_i of the protein concentration x_i gives an indication of the strength of gene expression in the regulatory domain.

As in the case of threshold parameters, exact numerical values for the rate constants will not usually be available. However, it is possible order the possible target equilibrium levels of x_i in different regions of the phase space with respect to the threshold concentrations.[1] The resulting *equilibrium inequalities* define the strength of gene expression in a regulatory domain in a qualitative way, on the scale of ordered threshold concentrations. More precisely, for every $\mu_i \in M_i$, $\nu_i \in N_i$, we specify some l_i, $1 \le l_i < p_i$, such that

$$\theta_i^{(l_i)} < \mu_i/\nu_i < \theta_i^{(l_i+1)}, \tag{4}$$

with special cases $0 < \mu_i/\nu_i < \theta_i^{(1)}$ and $\theta_i^{(p_i)} < \mu_i/\nu_i < max_i$.

Inspection of the state equations of AbrB shows that $M_{ab} = \{0, \kappa_{ab}\}$ and $N_{ab} = \{\gamma_{ab}\}$. The equilibrium level

[1] The equilibrium inequalities have also been called *nullcline inequalities* [de Jong and Page, 2000], because the equilibrium levels correspond to nullcline hyperplanes in the phase space.

κ_{ab}/γ_{ab} is placed above the highest AbrB threshold, since otherwise the concentration of AbrB would never be able to reach or maintain a level at which negative autoregulation takes place. This leads to the equilibrium inequalities $\theta_{ab_2} < \kappa_{ab}/\gamma_{ab} < max_{ab}$.

The threshold and equilibrium inequalities determine the sign of $\dot{x}_i$ in a regulatory domain, and hence the local dynamics of the system. In fact, given a regulatory domain defined in dimension i by $\theta_i^{(k_i)} < x_i < \theta_i^{(k_i+1)}$, it can be shown that, if $\mu_i/\nu_i < \theta_i^{(k_i)}$, then $\dot{x}_i < 0$ everywhere inside the regulatory domain. Similarly, if $\mu_i/\nu_i > \theta_i^{(k_i+1)}$, then $\dot{x}_i > 0$. On the other hand, if $\theta_i^{(k_i)} < \mu_i/\nu_i < \theta_i^{(k_i+1)}$, then the sign of $\dot{x}_i$ in the regulatory domain is not unique, written as $\dot{x}_i \lesseqgtr 0$. In particular, $\dot{x}_i < 0$ on one side of the hyperplane $x_i = \mu_i/\nu_i$, $\dot{x}_i > 0$ on the other side of the plane, and $\dot{x}_i = 0$ inside the plane.

2.4 Discontinuities in state equations

The question must be raised what happens on the threshold hyperplanes $x_i = \theta_i^{(k_i)}$ separating the regulatory domains, where the step functions, and hence the state equations, are not defined. Several solutions are possible for this non-trivial problem. In this paper we follow the approach of Plahte *et al.* [1994] and replace the discontinuous step functions by continuous ramp functions (Fig. 5). The solution of the PLDEs with step functions is then defined to be the solution of the DEs with ramp functions considered in the limit $\delta \to 0$.

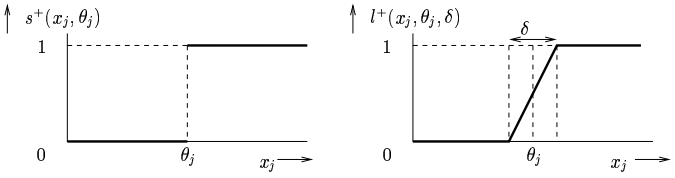


Figure 5: Step function $s^+(x_j,\theta_j)$ and ramp function $l^+(x_j,\theta_j,\delta)$. l^+ approaches s^+ as $\delta \to 0$.

The use of ramp functions divides the phase space into regulatory domains separated by *switching zones*, regions in which one or more x_i have a value in the δ-interval $[\theta_i^{(k_i)} - \delta/2, \theta_i^{(k_i)} + \delta/2]$ around a threshold $\theta_i^{(k_i)}$. An example phase space with switching zones is shown in Fig. 4. Inside the regulatory domains, the DEs with ramp functions are equivalent to the PLDEs with step functions. Outside the regulatory domains, in the switching zones, the DEs with ramp functions may be nonlinear functions of the concentration variables. The switching zones separating the regulatory domains vanish and approach (intersections of) the threshold hyperplanes as $\delta \to 0$.

3 Qualitative simulation of genetic regulatory networks

The goal of *qualitative simulation* is to exploit the qualitative constraints on parameter values in order to predict the qualitative dynamics of a regulatory network. More specifically, we would like to know which regulatory domains can be reached by some solution trajectory starting in the initial regulatory domain, for parameter values consistent with the specified threshold and equilbrium inequalities. A sequence of regulatory domains thus generated gives an indication of the evolution of the functional state of the system, as transitions between regulatory domains reflect changes in the synthesis and degradation rates of proteins.

3.1 Qualitative values, states, and behaviors

The analysis of PLDEs of the form (1) motivates the introduction of the *qualitative value* of a state variable and its derivative, as well as definitions of the *qualitative state* and *qualitative behavior* of the system. Let $\boldsymbol{\xi}(t) = \boldsymbol{\xi}(\boldsymbol{x}, \boldsymbol{x}_0, \boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{\gamma}, t)$ be the solution of PLDEs with step functions describing a regulatory network on the time-interval $[t_0, t_\infty[$ for given parameter values and initial conditions $\boldsymbol{x}(t_0) = \boldsymbol{x}_0$.

Def. 1 (Qualitative value) Suppose that at some $t > t_0$ it holds that $\boldsymbol{\xi}(t)$ lies in a regulatory domain, such that $\theta_i^{(k_i)} < \xi_i(t) < \theta_i^{(k_i+1)}$ for every i ($1 \le i \le n$ and $1 \le k_i < p_i$). The qualitative value $qv(\xi_i, t)$ of $\xi_i(t)$ is given by the inequalities $\theta_i^{(k_i)} < x_i < \theta_i^{(k_i+1)}$, while the qualitative value $\dot{qv}(\xi_i, t)$ of $\dot{\xi}_i(t)$ is given by one of the inequalities $\dot{x}_i < 0$, $\dot{x}_i > 0$, $\dot{x}_i \lesseqgtr 0$, depending on the sign of $\dot{\xi}_i(t)$ in the regulatory domain.

The definition can be straightforwardly generalized to the case of regulatory domains bounded by $x_i = 0$ or $x_i = max_i$. Notice that the qualitative value is not defined in the switching zones around the threshold hyperplanes.

Def. 2 (Qualitative state) The qualitative state $qs(\boldsymbol{\xi}, t)$ for $\boldsymbol{\xi}$ at t is given by the vectors $qv(\boldsymbol{\xi}, t)$ and $\dot{qv}(\boldsymbol{\xi}, t)$ of qualitative values.

A qualitative state associated with a regulatory domain can be interpreted as representing a functional state of the regulatory system. Each protein concentration has a value lying between two consecutive thresholds, and is either tending towards one of the threshold values, or evolving towards a value between the thresholds.

Def. 3 (Qualitative behavior) The qualitative behavior $qb(\boldsymbol{\xi})$ of $\boldsymbol{\xi}$ is given by the sequence of qualitative states $qs(\boldsymbol{\xi}, t)$ on $[t_0, t_\infty[$.

A qualitative behavior defines a succession of qualitative states of the regulatory system. It is not difficult to show that every solution $\boldsymbol{\xi}$ of (1) can be abstracted into a unique qualitative behavior.

3.2 Qualitative simulation algorithm

In terms of the above definitions, the qualitative simulation procedure can be formulated as follows [Kuipers, 1994]. Given initial qualitative values $\boldsymbol{qv}_0$, describing the initial protein concentrations $\boldsymbol{x}$, the simulation algorithm computes the initial qualitative state qs_0, and then determines all possible *transitions* from qs_0 to successor qualitative states. The generation of successor states is repeated in a recursive manner until all qualitative states reachable from the initial qualitative state have been found.

The possible transitions from a qualitative state are determined by the rule below.

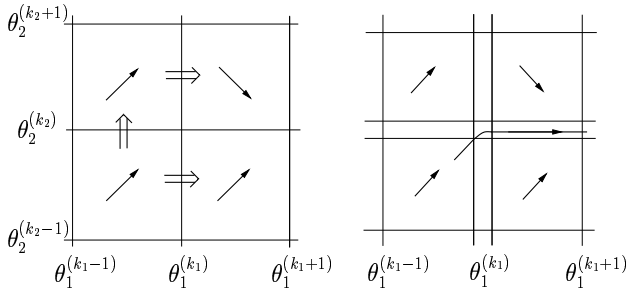
qv_i and $\dot{qv}_i$	qv'_i and $\dot{qv}'_i$
$\theta_i^{(k_i)} < x_i < \theta_i^{(k_i+1)}, \dot{x}_i > 0$	$\theta_i^{(k_i+1)} < x_i < \theta_i^{(k_i+2)}, \dot{x}_i > 0$
$\theta_i^{(k_i)} < x_i < \theta_i^{(k_i+1)}, \dot{x}_i > 0$	$\theta_i^{(k_i+1)} < x_i < \theta_i^{(k_i+2)}, \dot{x}_i \leqq 0$
$\theta_i^{(k_i)} < x_i < \theta_i^{(k_i+1)}, \dot{x}_i < 0$	$\theta_i^{(k_i-1)} < x_i < \theta_i^{(k_i)}, \dot{x}_i < 0$
$\theta_i^{(k_i)} < x_i < \theta_i^{(k_i+1)}, \dot{x}_i < 0$	$\theta_i^{(k_i-1)} < x_i < \theta_i^{(k_i)}, \dot{x}_i \leqq 0$

Figure 6: Continuity constraints for the qualitative values $qv_i, \dot{qv}_i$ and $qv'_i, \dot{qv}'_i$ of two qualitative states qs and qs' defined on adjacent regulatory domains. Valid for $1 < k_i < p_i - 1$, the constraints can be easily generalized to the case of qualitative values $0 \leq x_i < \theta_i^{(1)}$ and $\theta_i^{(p_i)} < x_i \leq max_i$.

Def. 4 (State transition) Let qs and qs' be two qualitative states associated with adjacent regulatory domains. A transition from qs to qs' is possible, if for every i, $1 \leq i \leq n$, such that $qv_i \neq qv'_i$, the qualitative values $qv_i, \dot{qv}_i$ and $qv'_i, \dot{qv}'_i$ satisfy the continuity constraints in Fig. 6.

Fig. 7(a) illustrates the application of the rule. Intuitively formulated, the rule says that a transition from one qualitative state to another is possible, if a trajectory may cross the switching zone separating the regulatory domains of the qualitative states.

A simulation algorithm based on Def. 4 is described in [de Jong and Page, 2000]. The qualitative states and transitions generated by the algorithm form a *state transition graph*. The graph may contain cycles and states without successors, which are together referred to as *attractors*. Since the number of possible qualitative states is finite, every path in the state transition graph will reach an attractor at some point. Each path running from the initial qualitative state qs_0 to an attractor forms a possible qualitative behavior of the regulatory system.



(a) (b)

Figure 7: (a) Phase space with derivative vectors $\dot{\boldsymbol{x}}$ in the regulatory domains ($\rightarrow$) and the state transitions ($\Rightarrow$) permitted by Def. 4. (b) Solution trajectory escaping through switching zones.

3.3 Properties of simulation algorithm

Given a model and initial qualitative values $\boldsymbol{qv}_0$, what can be said about the correctness of the behaviors produced by qualitative simulation? We define a set Ξ of possible solutions of (1) on $[t_0, t_\infty[$, such that for every $\boldsymbol{\xi} \in \Xi$ the numerical values of $\boldsymbol{\theta}$, $\boldsymbol{\kappa}$, and $\boldsymbol{\gamma}$ satisfy (3)-(4), and $qv(\boldsymbol{\xi}, t_0) = \boldsymbol{qv}_0$. The ultimate aim of qualitative simulation is to determine the set QB of qualitative behaviors, such that (1) for every $\boldsymbol{\xi} \in \Xi$ there is a $b \in QB$, such that $qb(\boldsymbol{\xi}) = b$ (*soundness*), and (2) for every $b \in QB$ there is a $\boldsymbol{\xi} \in \Xi$, such that $qb(\boldsymbol{\xi}) = b$ (*completeness*).

Unfortunately, the simulation algorithm based on Def. 4 is not sound. A trajectory may enter a switching zone from a regulatory domain, and then escape through other switching zones to enter another, possibly non-adjacent regulatory domain (Fig. 7(b)). As a consequence, the simulation algorithm may overlook qualitative state transitions. We found that the practical consequences of such omissions are limited, since qualitative states not directly reachable by a transition are often indirectly reachable by a sequence of transitions.

Completeness of the simulation algorithm has neither been proven nor disproven, but seems difficult to guarantee given the behavioral complexity that can be attained by models of the form (1) [Lewis and Glass, 1991].

3.4 Genetic Network Analyzer

The simulation method has been implemented in Java 1.2, in a program called *GNA (Genetic Network Analyzer)*.[2] The program reads and parses input files specifying the model of the system (state equations, threshold and equilibrium inequalities) and the initial state. From this information it produces a state transition graph. Extensions of the simulation algorithm allow all qualitative states and their transitions to be generated, as well as the completion and simulation of models with unspecified threshold and equilibrium inequalities.

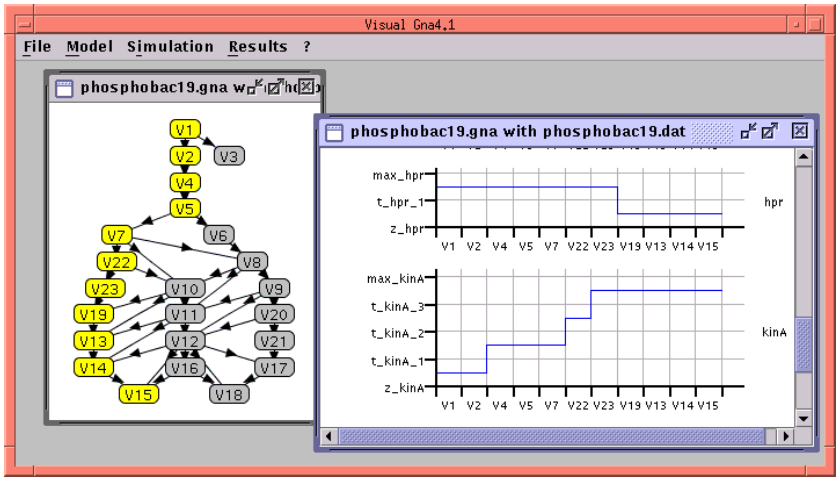
GNA is accessible through a graphical user-interface, which allows the network of interactions between genes to be displayed, as well as the state transition graph resulting from the simulation. In addition, the user can analyze the attractors with their basins of attraction, and focus on qualitative behaviors to study the temporal evolution of protein concentrations in more detail (Fig. 8).

4 Application: sporulation in *B. subtilis*

The method and its implementation have been used to study the regulatory network underlying the initiation of the *sporulation* process in the Gram-positive soil bacterium *Bacillus subtilis* [Grossman, 1995; Hoch, 1993]. Under conditions of nutrient deprivation, *B. subtilis* can decide not to divide and form a dormant, environmentally-resistant spore instead. The decision to either divide or sporulate is made by a complex regulatory network integrating various environmental, cell-cycle, and metabolic signals.

A schematic representation of the core of this network, displaying key genes and their regulatory interactions, is shown in Fig. 1. The central component of the network is a phosphorylation pathway, a *phosphorelay*, which transfers phosphates from the KinA kinase to the Spo0A regulator. Above a certain threshold, the phosphorylated form of Spo0A (Spo0A$\sim$P) activates various genes that commit the bacterium to sporulation. An example is the *spoIIA* operon, which encodes the transcription factor σ^F, essential for the development of the forespore. The flux of phosphate across the phosphorelay,

[2] GNA is available from the authors upon request.



(a)

$0 < \kappa_{se}/\gamma_{se} < \theta_{se_1}$ 23 states attractors: sporulation and division
$\theta_{se_1} < \kappa_{se}/\gamma_{se} < \theta_{se_2}$ 44 states attractors: division
$\theta_{se_2} < \kappa_{se}/\gamma_{se} < \theta_{se_3}$ 65 states attractors: division
$\theta_{se_3} < \kappa_{se}/\gamma_{se} < max_{se}$ 1155 states attractors: division

(b)

Figure 8: (a) GNA output for the equilibrium inequalities $0 < \kappa_{se}/\gamma_{se} < \theta_{se_1}$. The left window shows the state transition graph with a qualitative behavior running from the initial qualitative state V1 to the attractor state V15. In the right window the temporal evolution of the qualitative value of Hpr and KinA for this behavior can be followed. (b) Summary of simulation results for an initial state expected to induce sporulation, while varying the equilibrium inequalities for Spo0E. The simulations take between 0.5 and 3 seconds to complete on a Sun Ultra 10 workstation.

and hence the concentration of Spo0A∼P, is controlled by various external signals influencing the activity of pathway components. In addition, the flux of phosphate is regulated by Spo0A∼P itself, through a number of direct and indirect feedback loops involving *abrB*, *sinI*, *sinR*, and other genes.

Using the extensive literature on *B. subtilis* sporulation, and the Subtilist database at the Institut Pasteur, we have formulated state equations and appropriate parameter inequalities for every gene in the network. In total, the mathematical model consists of 10 state equations, 10 threshold inequalities, and 30 equilibrium inequalities. In the rare cases that the literature did not unambiguously determine the parameter inequalities, we have systematically explored the alternatives and selected those that permit the observed behavior of the bacterium to be reproduced.

The behavior of *B. subtilis* has been simulated from a variety of initial states, reflecting different physiological conditions. For example, fig. 8(a) shows the simulation results for an initial state reflecting a perturbation of the vegetative growth conditions, when the protein kinase KinA autophosphorylates in response to an external signal indicating a state of nutritional deprivation. Under these conditions, a state transition graph with two attractors is produced, corresponding to states in which the bacterium continues to divide (V3) or initiates spore formation (V15 and V18). Both states may be reached, depending on the exact values of the parameters satisfying the threshold and equilibrium inequalities.

In order to obtain result consistent with experimental data, we found that the target equilibrium concentrations of Spo0E have to be placed below the lowest threshold concentrations ($0 < \kappa_{se}/\gamma_{se} < \theta_{se_1}$). That is, we need to assume that *spo0E* expression levels are quite weak. When other equilibrium inequalities are chosen, the simulations predict that sporulation cannot be initiated under appropriate conditions, contrary to what is observed (Fig. 8(b)). In fact, Spo0E mediates the negative autoregulation of Spo0A∼P, and thus prevents a critical concentration of Spo0A∼P to accumulate.

The above choice of parameter constraints is troublesome, because it implies that Spo0E cannot exert any influence on the decision to sporulate, since its concentration will not reach the threshold levels above which it can block the phosphate flux through the phosphorelay. The simulation results thus suggest that the network in Fig. 1, based on interactions reported in the literature, may be incomplete. As a remedy, we could postulate that an unknown signal decreases the activity of Spo0E at the onset of sporulation. Molecular studies of the interaction of Spo0E with components of the phosporelay suggest the existence of such a cellular factor which remains as of yet unidentified [Ohlsen *et al.*, 1994].

5 Discussion

We have presented an implemented method for the qualitative simulation of genetic regulatory systems that can handle large and complex networks of genes and interactions. The method is a generalization of the method in [de Jong and Page, 2000], in that it allows a larger class of regulatory relationships between genes to be modeled. In the first place, there are no restrictions on the logical functions that can be represented by (2). This permits complex regulatory interactions to be included in the models, as illustrated by the state equation for *abrB* in Fig. 3(b). In the second place, the regulation of protein degradation can be taken into account. Although this feature of the method has not been used in the sporulation example, it turned out to be crucial in modeling the network controlling the induction of the lytic cycle following phage μ infection of *E. coli* (results not shown here).

The applicability of the method to actual regulatory networks has been demonstrated by an analysis of the large

and complex network underlying the initiation of sporulation in *B. subtilis.* The analysis has resulted in a suggestion to complete the model compiled from the sporulation literature, which shows the potential of the method as a tool to focus further exprimentation. To our knowledge, qualitative simulation of genetic regulatory networks of the size and complexity considered in this paper has not been undertaken thus far.

Upscaling of the simulation method is achieved by modeling genetic regulatory systems by a class of piecewise-linear differential equations imposing strong constraints on the local dynamics of the system. Besides in molecular genetics, PLDEs of this form have been used in other biological domains, for instance in population biology. In order to effectively apply the constraints, the representation of the qualitative state of the system and the simulation algorithm are adapted to the mathematical structure of the equations.

Adaptation to a specific class of models is the principal respect in which the method presented in this paper differs from well-known QR methods like QPT and QSIM [Forbus, 1984; Kuipers, 1994]. A major difference with QSIM is that the qualitative state of a regulatory system is described on a higher level of abstraction. In particular, the behavior inside a regulatory domain is abstracted into a single qualitative state, making use of the fact that inside a regulatory domain either $\dot{x}_i < 0$, $\dot{x}_i > 0$, or $\dot{x}_i \lessgtr 0$. In QSIM one would have to distinguish an exponentially growing number of qualitative states inside and on the boundary of a regulatory domain.

Approximating step functions by infinitely steep ramp functions allows a precise definition of the behavior of the system in the threshold planes, and hence of the possible successors of a qualitative state. The state transition rule in Def. 4 is simple and intuitively clear, but does not preserve soundness of the algorithm. Even though the practical limitations of this may be limited, the development of state transition rules guaranteeing soundness is an important topic for further research.

The *B. subtilis* example suggest an approach to validate hypothesized models of genetic regulatory networks. Given temporal gene expression patterns observed under certain physiological conditions in wild-type or mutant strains of the bacterium, one can develop algorithms to systematically search the space of freely adjustable parameter inequalities to find constraints for which the model is able to account for the observations. Extensions of this type would allow the simulation method presented in this paper to evolve into a more general modeling tool.

Acknowledgments The authors would like to thank I. Vatcheva and three anonymous referees for comments on a previous version of this paper.

References

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

[de Jong and Page, 2000] H. de Jong and M. Page. Qualitative simulation of large and complex genetic regulatory systems. In W. Horn, editor, *Proc. 14th Europ. Conf. Artif. Intell. (ECAI 2000)*, pages 141–145. IOS Press, 2000.

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

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

[Grossman, 1995] A.D. Grossman. Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis. Annu. Rev. Genet.*, 29:477–508, 1995.

[Heidtke and Schulze-Kremer, 1998] 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.

[Hoch, 1993] J.A. Hoch. *spo0* genes, the phosphorelay, and the initiation of sporulation. In A.L. Sonenshein *et al.*, editor, *Bacillus subtilis and other Gram-Positive Bacteria*, pages 747–756. AMS, 1993.

[Kauffman, 1993] S.A. Kauffman. *The Origins of Order*. Oxford University Press, 1993.

[Kuipers, 1994] B. Kuipers. *Qualitative Reasoning*. MIT Press, 1994.

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

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

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

[Ohlsen *et al.*, 1994] K.L. Ohlsen, J.K. Grimsley, and J.A. Hoch. Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase. *Proc. Natl Acad. Sci. USA*, 91:1756–1760, 1994.

[Plahte *et al.*, 1994] E. Plahte, T. Mestl, and S.W. Omholt. Global analysis of steady points for systems of differential equations with sigmoid interactions. *Dyn. Stab. Syst.*, 9(4):275–291, 1994.

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

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

[Trelease *et al.*, 1999] 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. *Artif. Intell. Med.*, 17:303–321, 1999.

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