Qualitative Simulation of the Initiation of Sporulation in *Bacillus subtilis*

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

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

Abstract. The study of genetic regulatory networks has received a major impetus from the recent development of experimental techniques allowing the measurement of patterns of gene expression in a massively parallel way. This experimental progress calls for the development of appropriate computer tools for the modeling and simulation of gene regulation processes. We present Genetic Network Analyzer (GNA), a computer tool for the modeling and simulation of genetic regulatory networks. The tool is based on a qualitative simulation method that employs coarse-grained models of regulatory networks. The use of GNA is illustrated by a case study of the network of genes and interactions regulating the initiation of sporulation in *Bacillus subtilis*.

1 Introduction

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

The study of *genetic regulatory networks* has taken a qualitative leap through the use of modern genomic techniques that allow simultaneous measurement of the expression levels of all genes of an organism. In addition to experimental tools, computer tools for the *modeling* and *simulation* of gene regulation processes will be indispensable. As most networks of interest involve many genes connected through interlocking positive and negative feedback loops, an intuitive understanding of their dynamics is difficult to obtain and may lead to erroneous conclusions. Modeling and simulation tools, with a solid foundation in mathematics and computer science, allow the behavior of large and complex systems to be predicted in a systematic way [2].

Several computer tools for the simulation of biochemical reaction networks by means of differential equations are currently available. These tools can be used to simulate genetic, metabolic, and signal transduction networks

described by differential equations. In addition, they allow the user to perform tasks like the analysis of steady states and the estimation of parameter values. The currently-available tools are essentially restricted to *quantitative* models of reaction networks, in the sense that numerical values for the kinetic parameters and molecular concentrations need to be specified. However, since this information is usually absent, especially in the case of systems that are not well-understood, the above-mentioned tools may be difficult to apply.

This paper presents *Genetic Network Analyzer (GNA)*, a computer tool for the qualitative simulation of genetic regulatory networks. GNA employs piecewise-linear (PL) differential equation models that have been well-studied in mathematical biology [6,10,11]. While abstracting from the precise molecular mechanisms involved, the PL models capture essential aspects of gene regulation. Their simple mathematical form permits a qualitative analysis of the dynamics of the genetic regulatory systems to be carried out. Instead of numerical values for parameters and initial conditions, GNA asks the user to specify qualitative constraints on these values in the form of algebraic inequalities. Unlike precise numerical values, these constraints can usually be inferred from the experimental literature.

The use of GNA will be illustrated in the context of a regulatory network of biological interest, consisting of the genes and interactions regulating the initiation of sporulation in the Gram-positive soil bacterium *Bacillus subtilis* [1,7,8]. 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 controlled by a regulatory network integrating various environmental, cell-cyle, and metabolic signals. The aim of the example is to show that GNA is able to reproduce experimental findings in the case of a large and complex network that is well-understood by molecular biologists.

2 Qualitative simulation of genetic regulatory networks

The dynamics of genetic regulatory networks can be modeled by a class of piecewise-linear differential equations of the following general form [6,10,11]:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{x})\,\boldsymbol{x},\ \boldsymbol{x} \geq \boldsymbol{0}, \tag{1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_n)'$ is a vector of cellular protein concentrations, and $\boldsymbol{f} = (f_1, \ldots, f_n)'$, $\boldsymbol{g} = \mathrm{diag}(g_1, \ldots, g_n)$. The rate of change of each concentration x_i, $1 \leq i \leq n$, is defined 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. The function $f_i : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{\geq 0}$ consists of a sum of step function expressions, each weighted by a rate parameter, which expresses the logic of gene regulation [10,12]. The function $g_i : \mathbb{R}^n_{\geq 0} \to \mathbb{R}_{>0}$ is defined analogously. On a formal level, the PL models are related to a class of asynchronous logical models proposed by Thomas and colleagues [12].

Figure 1 gives an example of a simple genetic regulatory network. Genes a and b, transcribed from separate promoters, encode proteins A and B, each of which controls the expression of both genes. More specifically, proteins A and B repress gene a as well as gene b at different concentrations. Repression of the genes is achieved by binding of the proteins to regulatory sites overlapping with the promoters.

The network in figure 1 can be described by means of the following pair of state equations:

$$\dot{x}_a = \kappa_a \, s^-(x_a, \theta_a^2) \, s^-(x_b, \theta_b^1) - \gamma_a \, x_a \tag{2}$$

$$\dot{x}_b = \kappa_b \, s^-(x_a, \theta_a^1) \, s^-(x_b, \theta_b^2) - \gamma_b \, x_b. \tag{3}$$

Gene a is expressed at a rate $\kappa_a > 0$, if the concentration of protein A is below its threshold θ_a^2 and the concentration of protein B below its threshold θ_b^1, that is, if $s^-(x_a, \theta_a^2) \, s^-(x_b, \theta_b^1) = 1$. Recall that $s^-(x, \theta)$ is a step function evaluating to 1, if $x < \theta$, and to 0, if $x > \theta$. Protein A is spontaneously degraded at a rate proportional to its own concentration ($\gamma_a > 0$ is a rate constant). The state equation of gene b is interpreted analogously.

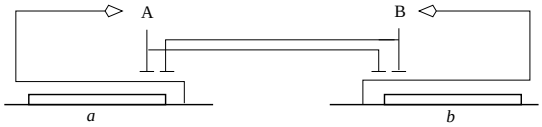

Fig. 1. Example of a genetic regulatory network of two genes (a and b), each coding for a regulatory protein (A and B). The notation follows, in a somewhat simplified form, the graphical conventions proposed by Kohn [9].

Most of the time, precise numerical values for the threshold and rate parameters in the differential equations are not available. Rather than numerical values, we specify qualitative constraints on the parameter values. These constraints, having the form of algebraic inequalities, can usually be inferred from biological data. The first type of constraint is obtained by ordering the p_i threshold concentrations of gene i, yielding the *threshold inequalities*. The second type of constraint, the *equilibrium inequalities*, are obtained by ordering the quotients of production and degradation parameters with respect to the thresholds. In the example, we specify the constraints:

$$0 < \theta_a^1 < \theta_a^2 < max_a, \qquad \theta_a^2 < \kappa_a/\gamma_a < max_a, \tag{4}$$

$$0 < \theta_b^1 < \theta_b^2 < max_b, \qquad \theta_b^2 < \kappa_b/\gamma_b < max_b. \tag{5}$$

On the one hand, the parameter inequalities divide the phase space into regions where the systems behaves in a qualitatively distinct way. These regions correspond to *qualitative states* of the system. On the other hand, the

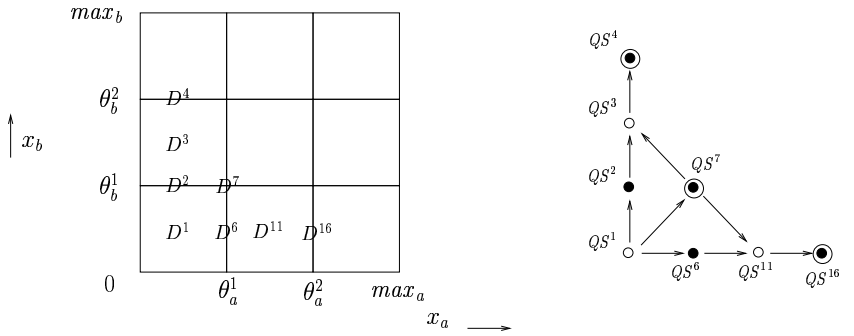

Fig. 2. The left figure shows the subdivision of the phase space into regions corresponding to qualitative states, for the example network in figure 1. The right figure shows the transition graph resulting from a simulation of the network starting in the qualitative state QS^1, corresponding to the domain D^1. Qualitative states corresponding to an equilibrium of the differential equations are circled [5].

parameter inequalities allow possible transitions between qualitative states to be determined by exploiting the mathematical properties of the PL models. A *qualitative simulation* consists of the generation of all qualitative states reachable through one or more transitions from a given initial qualitative state. A qualitative simulation results in a transition graph, consisting of qualitative states and transitions between qualitative states. The paths in the transition graph represent the possible qualitative behaviors predicted by the simulator [5] (figure 2).

The qualitative simulation method has been implemented in Java 1.3 in the program *Genetic Network Analyzer (GNA)* [4]. GNA is available for non-profit academic research purposes at `http://www-helix.inrialpes.fr/gna`. The core of the system is formed by the simulator, which generates a transition graph from a PL model, parameter inequalities, and an initial qualitative state. The input of the simulator is obtained by reading and parsing text files specified by the user. A graphical user interface (GUI), named *VisualGNA*, assists the user in specifying the model of a genetic regulatory network as well as in interpreting the simulation results.

3 Initiation of sporulation in *B. subtilis*

The use of GNA can be illustrated in the context of a large and complex regulatory network of biological interest, consisting of the genes and interactions regulating the initiation of sporulation in the Gram-positive soil bacterium *Bacillus subtilis* [1,7,8]. Under conditions of nutrient deprivation, *B. subtilis* cells may not divide and form a dormant, environmentally-resistant spore instead. The decision to either divide or sporulate is controlled by a regulatory network integrating various environmental, cell-cyle, and metabolic signals.

A graphical representation of the network is shown in figure 1, displaying key genes and their promoters, proteins encoded by the genes, and the regulatory action of the proteins (see [3] for details and references to the experimental literature).

The graphical representation of the network can be translated into a PL model supplemented by qualitative constraints on the parameters. The resulting model consists of nine state variables and two input variables. The 49 parameters are constrained by 58 parameter inequalities, the choice of which is largely determined by biological data. Simulation of the sporulation network by means of GNA reveals that essential features of the initiation of sporulation in wild-type and mutant strains of *B. subtilis* can be reproduced by means of the model [3]. In particular, the choice between vegetative growth and sporulation is seen to be determined by competing positive and negative feedback loops influencing the accumulation of the phosphorylated transcription factor Spo0A. Above a certain threshold, Spo0A∼P activates various genes whose expression commits the bacterium to sporulation, such as genes coding for sigma factors that control the alternative developmental fates of the mother cell and the spore.

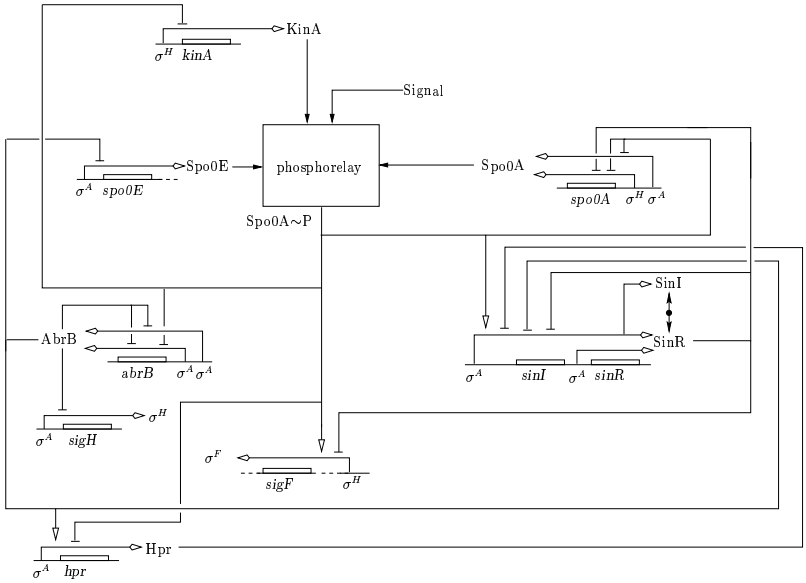

Fig. 3. Key genes, proteins, and regulatory interactions making up the network involved in *B. subtilis* sporulation. In order to improve the legibility of the figure, the control of transcription by the sigma factors σ^A and σ^H has been represented implicitly, by annotating the promoter with the sigma factor in question.

4 Conclusions

We have presented the computer tool GNA for the qualitative simulation of genetic regulatory networks and illustrated its use in the analysis of the network of interactions controlling the initiation of sporulation in *B. subtilis*. GNA implements a simulation method that is based on a class of piecewise-linear (PL) differential equation models described in mathematical biology [5]. Instead of giving numerical values to the parameters and initial conditions, which are usually not available, we use qualitative constraints in the form of algebraic inequalities. These are obtained by directly translating biological data into a mathematical formalism.

References

1. W.F. Burkholder and A.D. Grossman. Regulation of the initiation of endospore formation in *Bacillus subtilis*. In Y.V. Brun and L.J. Shimkets, editors, *Prokaryotic Development*, chapter 7, pages 151–166. ASM, 2000.
2. H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.*, 9(1):69–105, 2002.
3. H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page. Qualitative simulation of the initiation of sporulation in *B. subtilis*. Technical Report RR-4527, INRIA, 2002. Submitted for publication.
4. H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics*, 2002. In press.
5. H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and H. Geiselmann. Qualitative simulation of genetic regulatory networks using piecewise-linear models. Technical Report RR-4407, INRIA, 2002. Submitted for publication.
6. L. Glass and S.A. Kauffman. The logical analysis of continuous non-linear biochemical control networks. *J. Theor. Biol.*, 39:103–129, 1973.
7. A.D. Grossman. Genetic networks controlling the initiation of sporulation and the development of genetic competence in *Bacillus subtilis*. *Ann. Rev. Genet.*, 29:477–508, 1995.
8. J.A. Hoch. Regulation of the phosphorelay and the initiation of sporulation in *Bacillus subtilis*. *Ann. Rev. Microbiol.*, 47:441–465, 1993.
9. K.W. Kohn. Molecular interaction maps as information organizers and simulation guides. *Chaos*, 11(1):1–14, 2001.
10. 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.
11. E.H. Snoussi. Qualitative dynamics of piecewise-linear differential equations: A discrete mapping approach. *Dyn. Stabil. Syst.*, 4(3-4):189–207, 1989.
12. R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks: I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.*, 57(2):247–276, 1995.