Qualitative Simulation of the Initiation of Sporulation in *B. 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
[2]PEGM, Université Joseph Fourier
460 rue de la Piscine, BP 53, 38041 Grenoble CEDEX 9
[3]ESA, Université Pierre Mendès France
{ Hidde.de-Jong, Celine.Hernandez, Michel.Page}@inrialpes.fr,
Hans.Geiselmann@ujf-grenoble.fr

March 26, 2001

Abstract

The modeling and simulation of regulatory systems are hampered by the size and complexity of regulatory networks and the absence of quantitative informations on regulatory interactions. In response to these problems, we have developed a method for the qualitative simulation of large and complex genetic regulatory networks. The method is here presented in the context of its application to a regulatory network of biological interest, namely the genes and interactions regulating the initiation of sporulation in *B. subtilis*. Simulations with a model of this network reveal that the salient features of sporulation initiation are reproduced, but that an additional interaction, hypothesized in the literature before, may be involved.

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 molecules. 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.

As regulatory systems of interest usually involve many components connected through interlocking positive and negative feedback loops, an intuitive understanding of the dynamics of the system is hard to obtain. Computer-supported *modeling* and *simulation* contribute to the elucidation of regulatory networks by enabling the biologist to systematically investigate the consequences of different hypotheses. Simulation of a regulatory network leads to predictions that can be verified by means of experimental data. If a conflict is observed, the biologist may suggest modifications of the mathematical model and test these through further experimentation.

The modeling and simulation of regulatory systems are hampered by two major difficulties [2]. A first problem is that a detailed description of the kinetics of a system of regulatory interactions requires large and complex mathematical models. For instance, in order to model the PER/TIM

feedback loop in the mechanism underlying circadian rhythms in *Drosophila*, Leloup and Goldbeter [9] proposed a system of ten nonlinear differential equations with about thirty parameters. Bearing in mind that just two genes and their products are involved, it is obvious that the analysis of large networks will be difficult to achieve on this level of detail. On the other hand, restricting attention to a small part of the network, without independent evidence that it is appropriate to do so, may cause one to miss behavioral features emerging from the global structure of interactions.

This first problem is aggravated by a second problem, namely the absence of quantitative information on the interactions. This severely complicates the application of traditional methods for numerical analysis. Although sometimes rough experimental values for the kinetic parameters in the model may be available, as in the case of λ phage growth control in *E. coli* [10], these examples form an exception to the rule.

In response to the above two problems, we have developed a method for the *qualitative simulation of large and complex genetic regulatory networks* [3, 4]. The method represents regulatory interactions between genes by a class of differential equations that abstract from biochemical details, while retaining a biological significance. Instead of giving a numerical value to the model parameters, qualitative constraints in the form of parameter inequalities are specified that allow predictions to be made about the possible qualitative behaviors of a regulatory system. The method has been implemented in Java 1.2, resulting in the computer tool *GNA (Genetic Network Analyzer)*.

The aim of this contribution is to introduce the above method in the context of its application to a regulatory network of biological interest, namely the genes and interactions regulating the *initiation of sporulation* in the Gram-positive soil bacterium *Bacillus subtilis* [6, 8]. Under conditions of nutrient deprivation, *B. subtilis* can decide not to divide and to form a dormant, environmentally-resistant spore instead. The decision to either divide or sporulate is controlled by a complex network integrating various environmental, cell-cyle, and metabolic signals. We have simulated the sporulation network using a model constructed from published reports of experiments. The simulations reveal that the salient features of the initiation of sporulation are reproduced, but that an additional interaction, hypothesized in the literature before, may be involved.

2 Modeling and simulation of initiation of sporulation in *B. subtilis*

On the basis of the extensive literature on *B. subtilis* sporulation (see [6, 8] for reviews), and information contained in the database Subtilist [12], we have constructed a mathematical model of the sporulation network. A graphical representation of this network, displaying key genes and their promoters, proteins encoded by the genes, and the regulatory action of the proteins, is shown in figure 1. For every interaction in the figure, genetic and usually molecular evidence is available.

The network is centered around a phosphorylation pathway integrating a variety of environmental, cell-cycle, and metabolic signals [6, 8]. This pathway, called a *phosphorelay*, transfers phosphates to the Spo0A regulator *via* a chain of kinases and phosphotransferases. Several regulatory proteins, detecting relevant physiological conditions, interact with the phosphorelay and stimulate or inhibit the flux of phosphate across the chain. When a sufficient number of inputs in favor of sporulation accumulate, the concentration of Spo0A$\sim$P reaches a threshold value above which it activates various genes that commit the bacterium to sporulation. In order to produce a critical level of Spo0A$\sim$P, signals arriving at the phosphorelay need to be amplified and stabilized. This is achieved by a number of positive and negative feedback loops controlling the activity of the phosphorelay by transcriptional regulation of its components. The decision to enter sporulation is therefore not determined by a single gene, but emerging from a complex network of genes and regulatory interactions that integrates a variety of external stimuli.

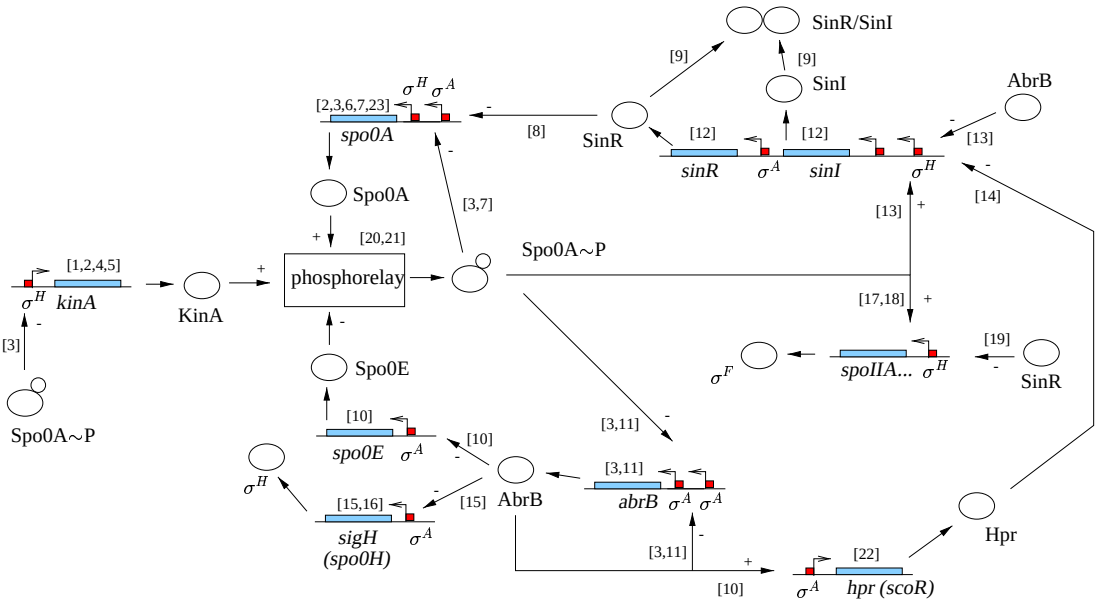

Figure 1: Genetic regulatory network underlying the inititation of sporulation in *B. subtilis.* For every gene, the coding region and the promoters are shown. Promoters are distinguished by the specific σ factor directing DNA transcription. The regulatory action of a protein tending to activate (inhibit) expression is indicated by a '+' ('-'). The numbers in the figure are references to the literature describing the structure of genes and the regulatory interactions: [1] Jaacks *et al.* (1989), *J. Bacteriol.*, 171(8):4121; [2] Predich *et al.* (1992), *J. Bacteriol.*, 174(9):2771; [3] Fuyita and Sadaie (1998), *J. Biochem.*, 124:98; [4] LeDeaux *et al.* (1995), *J. Bacteriol.*, 177(3):861; [5] Jiang *et al.* (2000), *Mol. Microbiol.*, 38(3):535; [6] Chibazakura *et al.* (1991), *J. Bacteriol.*, 173(8):2625; [7] Strauch *et al.* (1992), *Biochimie*, 74:619; [8] Mandić-Mulic *et al.* (1995), *J. Bacteriol.*, 177(16):4619; [9] Bai and Mandić-Mulic (1993), *Genes Dev.*, 7:139; [10] Strauch *et al.* (1989), *EMBO J.*, 8(5):1615; [11] Strauch (1993), In: Sonenshein *et al.*, *Bacillus Subtilis and other Gram-Positive Bacteria*, ASM Press, 757; [12] Gaur *et al.* (1988), *J. Bacteriol.*, 170(3):1046; [13] Strauch and Hoch (1993), *Mol. Microbiol.*, 7(3):337; [14] Kallio *et al.* (1991), *J. Biol. Chem.*, 266(20):13411; [15] Weir *et al.* (1991), *J. Bacteriol.*, 173(2):521; [16] Healy *et al.* (1991), *Mol. Microbiol.*, 5(2):477; [17] Trach *et al.* (1991), *Res. Microbiol.*, 142:815; [18] Wu *et al.* (1991), *Gene*, 101(1):113; [19] Mandić-Mulic *et al.* (1992), *J. Bacteriol.*, 174(11):3561; [20] Burbulys *et al.* (1991), *Cell*, 64:545; [21] Hoch (1993), *Annu. Rev. Microbiol.*, 47:441; [22] Perego and Hoch (1988), *J. Bacteriol.*, 170(6):2560. [23] Yamashita *et al.* (1989), *J. Gen. Microbiol.*, 135:1335.

The logic of the regulatory interactions is precisely specified in a mathematical model. The model has the form of a system of *piecewise-linear differential equations (PLDEs)* with nonnegative, real variables corresponding to the concentrations of the proteins encoded by the genes in figure 1. Differential equations of this form, and their logical abstractions, have been used to analyze gene regulation before [5, 11, 16]. The regulatory interactions are modeled by means of step functions mapping the concentration of a protein to a value 1 or 0, depending on whether this value is above or below a certain threshold concentration.

Figure 2(a) shows the differential equation corresponding to *spo0E*, a gene encoding a protein phosphatase involved in the phosphorelay. The differential equation states that *spo0E* is transcribed at a rate κ_{se} from a σ^A-promoter when the concentration x_{ab} of its repressor AbrB is below the threshold concentration θ_{ab_1} (i.e., $s^-(x_{ab}, \theta_{ab_1}) = 1$) [15]. 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., $s^+(x_{sa}, \theta_{sa_1}) = 1$). The differential equations also states that the degradation of Spo0E is proportional to its own concentration ($-\gamma_{se}\, x_{se}$). The differential equations for the other genes can be formulated in a similar way and may involve complex expressions of step functions. The 10 state equations comprise a total of 45 step functions expressing the logic of gene regulation.

Instead of specifying numerical values for the threshold and rate parameters in the PLDEs, the simulation method requires the model to be supplemented by qualitative constraints on the parameters in the form of threshold and equilibrium inequalities [3, 4]. The *threshold inequalities* order the different threshold concentrations of a protein. The threshold inequalities divide the concentration space into regions where the concentration of every protein lies between consecutive thresholds. Within such a region, the concentration will move towards a target equilibrium indicating the strength of gene expression in the region. The *equilibrium inequalities* order the possible target equilibria of the protein in different regions of the phase space with respect to the threshold concentrations.

For the proteins in the sporulation model, 1 to 4 threshold concentrations and 1 to 3 equilibrium concentrations need to be ordered. The threshold and equilibrium inequalities for Spo0E are shown in figure 2(b)-(c). To a large extent, the choice of appropriate threshold and equilibrium inequalities is constrained by biological considerations. For instance, the choice for the equilibrium inequalities $\theta_{se_3} < \kappa_{se}/\gamma_{se} < max_{se}$ is motivated by the fact that when *spo0E* is expressed, its concentration should be able to reach a level above which it can block the phospor flux of phosphate through the phosphorelay. If the threshold and equilibrium inequalities cannot be unambiguously determined, one has to repeat the simulation for different parameter constraints.

By means of the algorithm described in [4], the qualitative behavior of a genetic regulatory network can be simulated from a specified initial qualitative state. A *qualitative state* of the network is composed of qualitative values for each of the protein concentrations, where a qualitative value is defined as a concentration range bounded by threshold concentrations. That is, the possible qualitative values for Spo0E are $0 \leq x_{se} < \theta_{se_1}$, $\theta_{se_1} < x_{se} < \theta_{se_2}$, $\theta_{se_2} < x_{se} < \theta_{se_3}$, and $\theta_{se_3} < x_{se} \leq max_{se}$ (compare figure 2(b)).

The simulation algorithm produces a state transition graph by determining the qualitative states that are reachable from the initial qualitative state through one or more transitions. The paths in the graph starting from the initial state, and ending in an attractor (a state without successors or a state cycle), form the possible *qualitative behaviors* of the system. Figure 3 shows an example of a state transition graph generated by GNA, as well as the temporal evolution of the qualitative value of AbrB and KinA for a selected qualitative behavior in the graph.

State equation for Spo0E:

$$\frac{dx_{se}}{dt} = \kappa_{se}\, s^{-}(x_{ab}, \theta_{ab_1}) s^{+}(x_{sa}, \theta_{sa_1}) - \gamma_{se} x_{se}, \tag{1}$$

with

$$s^{-}(x_{ab}, \theta_{ab_1}) = \begin{cases} 1, & x_{ab} < \theta_{ab_1} \\ 0, & x_{ab} > \theta_{ab_1} \end{cases}, \qquad s^{+}(x_{sa}, \theta_{sa_1}) = \begin{cases} 0, & x_{sa} < \theta_{sa_1} \\ 1, & x_{sa} > \theta_{sa_1} \end{cases}$$

(a)

Four threshold concentrations are defined for Spo0E, $\theta_{se_1}, \ldots, \theta_{se_4}$, indicating four levels of activation of the phosphorelay. Assuming that $\theta_{se_1} < \ldots < \theta_{se_4}$, we have the *threshold inequalities*:

$$0 < \theta_{se_1} < \theta_{se_2} < \theta_{se_3} < \theta_{se_4} < max_{se},$$

where max_{se} is a constant denoting a maximum concentration.

(b)

The target equilibria for SpoOE are determined by $\frac{dx_{se}}{dt} = 0$. With (2), the possible equilibrium concentrations of Spo0E, for different values of $s^{-}(x_{ab}, \theta_{ab_1})$ and $s^{+}(x_{sa}, \theta_{sa_1})$, are given by $x_{se} = 0$ and $x_{se} = \frac{\kappa_{se}}{\gamma_{se}}$. The *equilibrium inequalities* for $x_{se} = \frac{\kappa_{se}}{\gamma_{se}}$ are:

$$\theta_{se_4} < \frac{\kappa_{se}}{\gamma_{se}} < max_{se}$$

(c)

Figure 2: (a) Differential equation and (b)-(c) threshold and equilibrium inequalities for Spo0E (see [3, 4] for details). The concentrations of Spo0E, AbrB, and SigA are denoted by x_{se}, x_{ab}, and x_{sa}, respectively.

3 Simulation results

GNA has been used to simulate the network underlying the initiation of sporulation from initial states reflecting a perturbation of the vegetative growth conditions.[1] The perturbations consist in an external signal indicating a state of nutritional deprivation, which causes KinA to autophosphorylate. The simulated behavior of our network should reflect the essential biological characteristics of the sporulation initiation process. In particular, we expect to observe two stable steady states, corresponding to vegetative growth and sporulation, and a controlled, all-or-non transition between these states in response to external stimuli [1].

After the modification to be discussed below, our model of the regulatory system reproduces exactly these properties. The simulations result in two stable steady states, V2 and V24 in Fig. 3, corresponding to vegetative growth (low concentrations of σ^H and Spo0A$\sim$P, high concentration of AbrB) and sporulation (high concentrations of KinA and Spo0A$\sim$P, activation of *spoIIA*). For a wide range of initial conditions, the network settles into either one of these states. A second important aspect of the model is that it shows how an external stimulus can lead to a transition between the stable steady states. Low-amplitude signals from the environment, giving rise to the autophosphorylation of KinA, are amplified and stabilized through feedback loops in order to provoke the dramatic change in the pattern of gene expression accompanying the switch from vegetative growth to sporulation.

In order to arrive at the above results, we had to modify the equilibrium inequalities for the phosphatase Spo0E specified in Fig. 2. In fact, for initial conditions in which *B. subtilis* is expected to sporulate, the model reproduces this behavior only when the target equilibrium concentrations of Spo0E are placed below the lower threshold concentrations of the protein ($0 < \kappa_{se}/\gamma_{se} < \theta_{se_1}$ or $\theta_{se_1} < \kappa_{se}/\gamma_{se} < \theta_{se_2}$). This is troublesome, because it implies that Spo0E cannot exert any influence on the decision to sporulate, as its concentration will not reach the threshold levels above which it can block the phosphate flux through the phosphorelay. 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 [13].

4 Discussion

We have presented a generic method for the qualitative simulation of large and complex genetic regulatory networks. The method has been implemented in the computer tool GNA and applied to the analysis of the network of interactions controlling the initiation of sporulation in *B. subtilis*. The simulations reveal that the essential features of the initiation of sporulation can be reproduced by means of a model constructed from the experimental literature. However, we also conclude that an additional interaction regulating the activity of the phosphatase Spo0E may be necessary for the decision to continue vegetative growth or enter sporulation. This example demonstrates the potential of our method to discover new and missing interactions and guide further experimentation.

Qualitative methods for the analysis of genetic regulation systems have been developed in artificial intelligence and mathematical biology before (reviewed in [2]). Applications of qualitative simulation methods developed in artifical intelligence (e.g., [7]) have revealed serious upscaling problems. The method presented in this paper is able to deal with large and complex networks by focusing on a class of differential equations that put strong constraints on the local dynamics of the

[1] Both GNA and the sporulation model are available from the authors upon request.

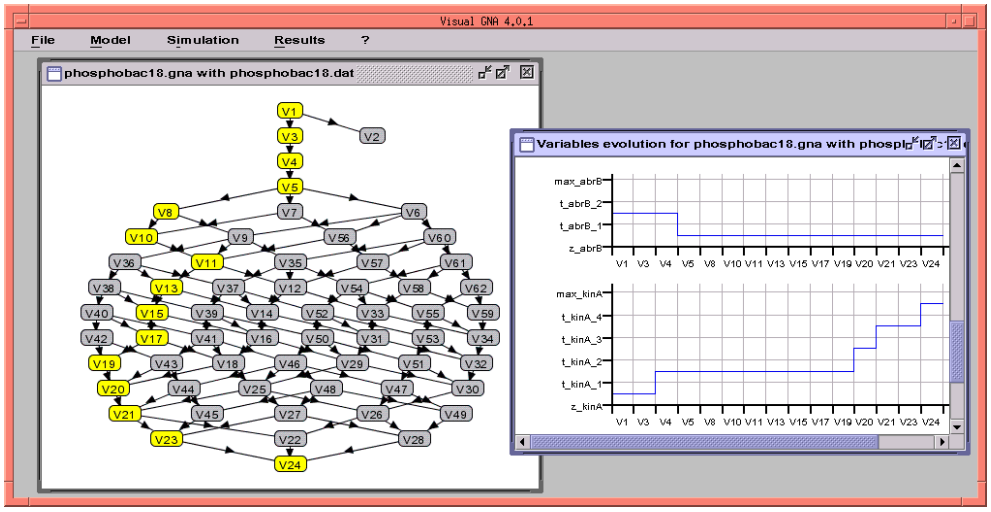

Figure 3: GNA output for the simulation of the sporulation model in Fig. 1. The left window shows the state transition graph with a qualitative behavior running from the initial qualitative state V1 to the attractor state V24. The right window shows the qualitative temporal evolution of the concentration of two of the proteins, AbrB and KinA. The simulation takes less than one second to complete on a SUN Ultra 10 workstation.

system [3, 4]. In fact, the simulations of the sporulation network take a few seconds to complete on an average PC, which allows alternative network topologies to be rapidly evaluated.

In comparison with the logical method developed by Thomas and colleagues [16], which is based on comparable abstractions of regulatory interactions, we have opted for differential equation models. We believe that the latter formalism is intuitively clear and of large generality. Moreover, it facilitates the integration of any quantitative data becoming available through improvements of current measurement technologies. Our approach has been influenced by the work of Snoussi [14] and others, who have demonstrated that the class of logical equations in [16] can be reformulated as a class of differential equations.

The results of the simulation of sporulation initiation in *B. subtilis* confirm the importance of the method's capability to handle large networks. In fact, the binary decision between vegetative growth and sporulation is an emerging property of the regulatory interactions between a large number of genes that cooperatively switch to a new genetic program. Further work is aimed at embedding the decision module structured around the phosphorelay into the larger network of sporulation and other responses to nutrient deprivation (e.g., competence development [6]). Analysis of these models will give us a clearer picture of the molecular mechanisms underlying the adaptation of *B. subtilis* to its environment.

Acknowledgements The authors thank A. Morgat, F. Rechenmann, A. Viari, and two anonymous referees for their comments on a draft of this paper.

References

[1] J. CHUNG, G. STEPHANOPOULOS, K. IRETON, AND A. GROSSMAN, (1994), *Gene expression in single cells of Bacillus subtilis: Evidence that a threshold mechanism controls the initiation of sporulation*, J. Bacteriol., 176, pp. 1977–1984.

[2] H. DE JONG, *Modeling and simulation of genetic regulatory systems: A literature review*, Tech. Rep. RR-4032, INRIA Rhône-Alpes, Montbonnot, (2000).

[3] H. DE JONG, J. GEISELMANN, C. HERNANDEZ, AND M. PAGE, (2001), *Qualitative simulation of genetic regulatory networks: Method and application*, in Proc. 17th Int. Joint Conf. Artif. Intell. (IJCAI-01), B. Nebel, ed., Morgan Kaufmann. to appear.

[4] H. DE JONG AND M. PAGE, (2000), *Qualitative simulation of large and complex genetic regulatory systems*, in Proc. 14th Europ. Conf. Artif. Intell. (ECAI 2000), W. Horn, ed., IOS Press, pp. 141–145.

[5] L. GLASS, (1975), *Classification of biological networks by their qualitative dynamics*, J. Theor. Biol., 54, pp. 85–107.

[6] A. GROSSMAN, (1995), *Genetic networks controlling the initiation of sporulation and the development of genetic competence in Bacillus subtilis*, Annu. Rev. Genet., 29, pp. 477–508.

[7] K. HEIDTKE AND S. SCHULZE-KREMER, (1998), *Design and implementation of a qualitative simulation model of λ phage infection*, Bioinformatics, 14, pp. 81–91.

[8] J. HOCH, (1993), *Regulation of the phosphorelay and the initiation of sporulation in Bacillus subtilis*, Annu. Rev. Microbiol., 47, pp. 441–465.

[9] J.-C. LELOUP AND A. GOLDBETER, (1998), *A model for the circadian rhythms in Drosophila incorporating the formation of a complex between the PER and TIM proteins*, J. Biol. Rhythms, 13, pp. 70–87.

[10] H. MCADAMS AND L. SHAPIRO, (1995), *Circuit simulation of genetic networks*, Science, 269, pp. 650–656.

[11] T. MESTL, E. PLAHTE, AND S. OMHOLT, (1995), *A mathematical framework for describing and analysing gene regulatory networks*, J. Theor. Biol., 176, pp. 291–300.

[12] I. MOSZER, P. GLASER, AND A. DANCHIN, (1995), *SubtiList: A relational database for the Bacillus subtilis genome*, Microbiology, 141, pp. 261–268.

[13] K. OHLSEN, J. GRIMSLEY, AND J. HOCH, (1994), *Deactivation of the sporulation transcription factor Spo0A by the Spo0E protein phosphatase*, Proc. Natl Acad. Sci. USA, 91, pp. 1756–1760.

[14] E. SNOUSSI, (1989), *Qualitative dynamics of piecewise-linear differential equations*, Dyn. Stab. Syst., 4, pp. 189–207.

[15] M. STRAUCH, *AbrB, a transition state regulator*, in *Bacillus subtilis* and other Gram-Positive Bacteria: Biochemistry, Physiology, and Molecular Genetics, A. S. *et al.*, ed., ASM Press, Washington, DC, (1993), pp. 757–764.

[16] R. THOMAS AND R. D'ARI, *Biological Feedback*, CRC Press, (1990).