Modeling and Simulation of Genetic Regulatory Networks

Hidde de Jong

Institut National de Recherche en Informatique et en Automatique (INRIA), Unité de recherche Rhône-Alpes, 655 avenue de l'Europe, Montbonnot, 38334 Saint Ismier Cedex, France, Hidde.de-Jong@inrialpes.fr

Abstract. The analysis of genetic regulatory networks will much benefit from the recent upscaling to the genomic level of experimental methods in molecular biology. In addition to high-throughput experimental methods, mathematical and bioinformatics approaches are indispensable for the analysis of genetic regulatory networks. Given the size and complexity of most networks of biological interest, an intuitive comprehension of their behavior is often difficult, if not impossible to obtain. A variety of methods for the modeling and simulation of genetic regulatory networks have been proposed in the literature. In this tutorial, the two principal approaches that have been used will be reviewed: methods based on differential equation models and stochastic models. In addition, we will indicate some alternative methods that have emerged in response to the difficulties encountered in applying the classical approaches.

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.

Genetic regulatory networks control the spatiotemporal expression of genes in an organism, and thus underlie complex processes like cell differentiation and development in prokaryotic and eukaryotic organisms. Genetic regulatory networks consist of genes, proteins, metabolites, and other small molecules, as well as their mutual interactions. Their study 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, mathematical methods supported by computer tools are indispensable for the analysis of genetic regulatory networks. 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 allow the behavior of large and complex systems to be predicted in a systematic way.

A variety of methods for the modeling and simulation of genetic regulatory networks have been proposed in the literature [3, 12, 15, 25]. In this tutorial, the two principal approaches that have been used in the literature will be briefly reviewed: differential equation models and stochastic models (section 2 and 3). The networks described by these models are examples of *positive systems* [7], in the sense that the state and output variables remain nonnegative on a time-interval T, if the input variables are positive on T. In fact, the variables in the models represent positive quantities, in particular the concentrations or numbers of molecules of proteins, mRNA, metabolites, and other constituents. In section 4, we will discuss the difficulties encountered in applying the classical approaches and point at alternative approaches that have emerged.

2 Differential equation models

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 mRNAs, proteins, and other molecules by time-dependent variables having non-negative real values. Regulatory interactions take the form of functional and differential relations between the concentration variables. More specifically, gene regulation is modeled by nonlinear equations expressing the rate of production or degradation of a component of the system as a function of the concentrations of other components. The equations have the mathematical form

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

where $\boldsymbol{x} = [x_1, \ldots, x_n]' \geq \mathbf{0}$ is the vector of concentrations of proteins, mRNAs, or small molecules, and $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ a usually nonlinear function. The rate of synthesis of i is seen to be dependent upon the concentrations $\boldsymbol{x}$, possibly including x_i.

Figure 1 shows a simple example of a genetic regulatory network. Genes a and b, transcribed from separate promoters, encode proteins A and B, each of which independently 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.

Figure 2(a) shows how the regulatory network in figure 1 can be modeled in terms of differential equations. The model consists of four variables denoting

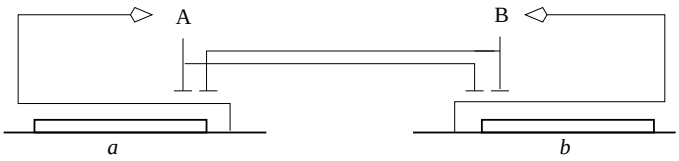


Fig. 1. Example of a simple regulatory network, consisting of the genes a and b, proteins A and B, and their mutual interactions. The notation follows, in a somewhat simplified form, the graphical conventions proposed by Kohn [14].

the concentration of mRNA and protein for genes a and b. The transcriptional inhibition of these genes is described by means of sigmoidal functions $h^- : \mathbb{R}^2 \to [0,1]$, which is motivated by the usually nonlinear, switch-like character of gene regulation. The translation of mRNA and the degradation of mRNA and proteins are assumed to be non-regulated and proportional to the substrate concentration. Due to the nonlinearity of f_i, analytical solution of the rate equations (1) is not normally possible. In special cases, qualitative properties of the solutions, such as the number and the stability of steady states and the occurrence of limit cycles, can be established. Most of the time, however, one has to take recourse to numerical techniques. In figure 2(b) the results of a numerical simulation of the example network are shown. As can be seen, the system reaches a steady state in which protein A is present at a high concentration, whereas protein B is nearly absent. For different initial conditions, but the same parameter values, a steady state may be reached in which the concentrations of A and B are reversed

Differential equations of the form (1) do not take into account the spatial dimension of regulatory processes, essential though in multicellular organisms. The equations can be generalized by defining compartments that correspond to cells or nuclei, by introducing concentration variables specific to each compartment, and by allowing diffusion between the compartments to take place. In the limit of the number of compartments, the resulting equations can be approximated by *partial differential equations (PDEs)*. Partial differential equations are even more difficult to solve analytically than ordinary differential equations, and in almost every situation of practical interest their use requires numerical techniques.

3 Stochastic models

An implicit assumption underlying (1), and differential equations more generally, is that concentrations of substances vary continuously and deterministically. Both of these assumptions may be questionable in the case of gene regulation, due to the usual small number of molecules of certain components [13, 16]. Instead of taking a continuous and deterministic approach, some authors have proposed to use discrete and stochastic models of gene regulation.

$$\frac{dx_{ra}}{dt} = \kappa_{ra}\, h^-(x_{pb}, \theta^1_{pb})\, h^-(x_{pa}, \theta^2_{pa}) - \gamma_{ra}\, x_{ra}$$

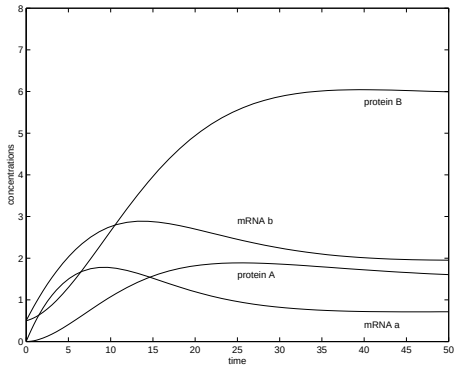
$$\frac{dx_{pa}}{dt} = \kappa_{pa}\, x_{ra} - \gamma_{pa}\, x_{pa}$$

$$\frac{dx_{rb}}{dt} = \kappa_{rb}\, h^-(x_{pa}, \theta^1_{pa})\, h^-(x_{pb}, \theta^2_{pb}) - \gamma_{rb}\, x_{rb}$$

$$\frac{dx_{pb}}{dt} = \kappa_{pb}\, x_{rb} - \gamma_{pb}\, x_{pb}$$

$$h^-(x, \theta) = \frac{\theta^2}{x^2 + \theta^2}$$

(a)



(b)

Fig. 2. (a) ODE model of the regulatory network in figure 1. The variables x_{pa} and x_{pb} denote the concentration of protein A and B, the variables x_{ra} and x_{rb} the concentration of the corresponding mRNA, the parameters κ_{ra}, κ_{pa}, κ_{rb}, and κ_{pb} production rates, the parameters γ_{ra}, γ_{pa}, γ_{rb}, and γ_{pb} degradation rates, and the parameters θ^1_{pa}, θ^2_{pa}, θ^1_{pb}, and θ^2_{pb} threshold concentrations. The variables are non-negative and the parameters positive. (b) Time-concentration plot resulting from a numerical simulation of the system described in (a), given specified values for the parameters.

Discrete amounts $\boldsymbol{X}$ of molecules are taken as state variables, and a joint probability distribution $p(\boldsymbol{X}, t)$ is introduced to express the probability that at time t the cell contains X_1 molecules of the first species, X_2 molecules of the second species, *etc.* The time evolution of the function $p(\boldsymbol{X}, t)$ can then be specified as follows:

$$p(\boldsymbol{X}, t + \Delta t) = p(\boldsymbol{X}, t)(1 - \sum_{j=1}^{m} \alpha_j \Delta t) + \sum_{j=1}^{m} \beta_j \Delta t, \tag{2}$$

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 $[t, t+\Delta t]$ [8, 9]. Rearranging (2), and taking the limit as $\Delta t \to 0$, gives the *master equation* [30]:

$$\frac{\partial}{\partial t} p(\boldsymbol{X}, t) = \sum_{j=1}^{m} (\beta_j - \alpha_j p(\boldsymbol{X}, t)). \tag{3}$$

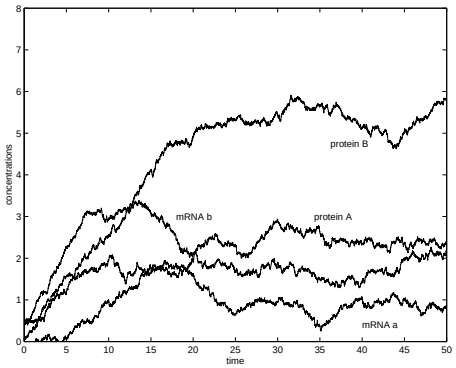
Compare this equation with the rate equations (1) above. Whereas the latter determine how the state of the system changes with time, the former describes how the probability of the system being in a certain state changes with time.

Notice that the state variables in the stochastic formulation can be reformulated as concentrations by dividing the number of molecules X_i by a volume factor.

Although the master equation provides an intuitively clear picture of the stochastic processes governing the dynamics of a regulatory system, it is even more difficult to solve by analytical means than the deterministic rate equation. In order to approximate the solution of the master equation, *stochastic simulation methods* have been developed [8, 21]. Given a set of possible reactions, the temporal evolution of the state $\boldsymbol{X}$, the number of molecules of each species, is predicted. The evolution of the state is determined by stochastic variables τ and ρ, representing the time interval between two successive reactions and the type of the next 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 (2).

$$
\begin{aligned}
\mathrm{A} + \mathrm{A} &\longleftrightarrow \mathrm{A}_2 \\
\mathrm{A}_2 + \mathrm{DNA}_b &\longleftrightarrow \mathrm{A}_2 \cdot \mathrm{DNA}_b \\
\mathrm{RNAP} + \mathrm{DNA}_b &\longleftrightarrow \mathrm{RNAP} \cdot \mathrm{DNA}_b \\
\mathrm{RNAP} \cdot \mathrm{DNA}_b &\longrightarrow \mathrm{RNAP} + \mathrm{DNA}_b \\
&\quad + \mathrm{RNA}_b \\
\mathrm{B} + \mathrm{B} &\longleftrightarrow \mathrm{B}_2 \\
\mathrm{B}_2 + \mathrm{DNA}_b &\longleftrightarrow \mathrm{B}_2 \cdot \mathrm{DNA}_b \\
\mathrm{B}_2 \cdot \mathrm{DNA}_b + \mathrm{A}_2 &\longleftrightarrow \mathrm{A}_2 \cdot \mathrm{B}_2 \cdot \mathrm{DNA}_b \\
\mathrm{A}_2 \cdot \mathrm{DNA}_b + \mathrm{B}_2 &\longleftrightarrow \mathrm{A}_2 \cdot \mathrm{B}_2 \cdot \mathrm{DNA}_b
\end{aligned}
$$

(a)



(b)

Fig. 3. (a) Some of the reactions involved in the expression of gene b in the regulatory network of figure 1. The following abbreviations are used: A and B (protein A and B), A_2 and B_2 (homodimer of A and B), RNAP (RNA polymerase), DNA_b (promoter region of gene b), and RNA_b (mRNA b). (b) A typical time-concentration plot resulting from stochastic simulation of the reaction system described in (a).

In figure 3(a) a few examples of reactions occurring in the network of figure 1 are shown: dimerization of the repressor A, binding of the repressor complex A·A to the promoter region, fixation of DNA polymerase to the promoter in the absence of the repressor complex, transcription of the gene b, *etc.* Typical results of a stochastic simulation of the example network are shown in figure 3(b). Notice the noisy aspect of the time evolution of the protein and mRNA concentrations. This effect, reflecting the stochastic nature of the initiation of transcription and the number of protein molecules produced per

transcript, may have important consequences. More particularly, fluctuations in the rate of gene expression may lead to phenotypic variation in an isogenic population [16, 18]. Indeed, starting from the same initial conditions, two different simulations may lead to qualitatively different outcomes. Whereas in one simulation protein A may be ultimately present at a high concentration and B at a low concentration like in figure 3(b), another simulation could lead to the opposite result.

4 Discussion

In summary, differential equation and stochastic models provide detailed descriptions of genetic regulatory networks, down to the molecular level. In addition, they can be used to make precise, numerical predictions of the behavior of regulatory systems. Many excellent examples of the application of these methods to prokaryote and eukaryote networks can be found in the literature. McAdams and Shapiro [17] have simulated the choice between lytic and lysogenic growth in bacteriophage λ using nonlinear differential equations, while Arkin and colleagues have studied the same system by means of a detailed stochastic model [1]. In a series of publications, the groups of Novak and Tyson have developed ODE models of the kinetic mechanisms underlying cell cycle regulation in *Xenopus* [2] and in yeast [22] (see [29] for a review). Differential equation models for the segmentation of *Drosophila* have been studied, focusing on the formation on the expression patterns of the gap, the pair-rule, and the segment polarity gene products in the trunk of the embryo [23, 24, 31].

In many situations of biological interest, however, the application of differential equation and stochastic models is seriously hampered. In the first place, the biochemical reaction mechanisms underlying regulatory interactions are usually not or incompletely known. This means that it is difficult to specify the rate functions f_i in (1) and the reactions j in (3). In the second place, quantitative information on kinetic parameters and molecular concentrations is only seldom available, even in the case of well-studied model systems. As a consequence, the numerical simulation methods mentioned above are often difficult to apply.

The above two constraints call for methods based on coarse-grained models that, while abstracting from the precise molecular mechanisms involved, capture essential aspects of gene regulation. Moreover, these methods should allow a qualitative analysis of the dynamics of the genetic regulatory systems to be carried out. A number of such methods have been proposed, such as the qualitative analysis of genetic regulatory networks described by piecewise-linear (PL) differential equations [4, 6, 10, 11, 20, 26], and the analysis of genetic regulatory networks by means of asynchronous, multivalued logic [19, 27, 28]. Although the methods are based on different formalisms, differential and log-

ical equations, they share important biological intuitions, in particular the description of gene activation in terms of on/off-switches.

The above-mentioned methods have been used to study a variety of prokaryotic and eukaryotic model systems, such as the choice between vegetative growth and sporulation in *B. subtilis* and the genetic control of the segmentation in the early *Drosophila* embryo (see [5] for a review). The applications show that, in order to understand the functioning of an organism in terms of the interactions in regulatory networks, it is not always necessary to model the process down to individual biochemical reactions. In fact, when a global understanding of the evolution of spatiotemporal patterns of gene expression is sought, coarse-grained and qualitative models might be profitably employed. However, when a more detailed and quantitative view of the dynamics of a regulatory system is required, the qualitative approaches need to be supplemented by conventional methods of the type discussed in sections 2 and 3.

References

1. 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.
2. 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.
3. H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *Journal of Computational Biology*, 9(1):69–105, 2002.
4. H. de Jong, J. Geiselmann, C. Hernandez, and M. Page. Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics*, 19(3):336–344, 2003.
5. H. de Jong, J. Geiselmann, and D. Thieffry. Qualitative modeling and simulation of developmental regulatory networks. In S. Kumar and P.J. Bentley, editors, *On Growth, Form, and Computers*. Academic Press, London, 2003. In press.
6. H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann. Hybrid modeling and simulation of genetic regulatory networks: A qualitative approach. In A. Pnueli and O. Maler, editors, *Hybrid Systems: Computation and Control (HSCC 2003)*, Lecture Notes in Computer Science. Springer-Verlag, Berlin, 2003.
7. L. Farina and S. Rinaldi. *Positive Linear Systems: Theory and Applications*. Wiley, New York, 2000.
8. D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
9. D.T. Gillespie. A rigorous derivation of the chemical master equation. *Physica D*, 188:404–425, 1992.
10. L. Glass and S.A. Kauffman. The logical analysis of continuous non-linear biochemical control networks. *Journal of Theoretical Biology*, 39:103–129, 1973.
11. J.-L. Gouzé and T. Sari. A class of piecewise linear differential equations arising in biological models. *Dynamical Systems*, 2003. To appear.

12. J. Hasty, D. McMillen, F. Isaacs, and J.J. Collins. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nature Review Genetics*, 2(4):268–279, 2001.
13. M.S.H. Ko. Induction mechanism of a single gene molecule: Stochastic or deterministic? *BioEssays*, 14(5):341–346, 1992.
14. K.W. Kohn. Molecular interaction maps as information organizers and simulation guides. *Chaos*, 11(1):1–14, 2001.
15. H.H. McAdams and A. Arkin. Simulation of prokaryotic genetic circuits. *Annual Review of Biophysics and Biomolecular Structure*, 27:199–224, 1998.
16. H.H. McAdams and A. Arkin. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics*, 15(2):65–69, 1999.
17. H.H. McAdams and L. Shapiro. Circuit simulation of genetic networks. *Science*, 269:650–656, 1995.
18. 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.
19. 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.
20. 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.
21. C.J. Morton-Firth and D. Bray. Predicting temporal fluctuations in an intracellular signalling pathway. *Journal of Theoretical Biology*, 192:117–128, 1998.
22. B. Novak, A. Csikasz-Nagy, B. Gyorffy, K.C. Chen, and J.J. Tyson. Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions. *Biophysical Chemistry*, 72:185–200, 1998.
23. 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.
24. 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.
25. 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.
26. 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.
27. R. Thomas and R. d'Ari. *Biological Feedback*. CRC Press, Boca Raton, FL, 1990.
28. 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. *Bulletin of Mathematical Biology*, 57(2):247–276, 1995.
29. J.J. Tyson, K. Chen, and B. Novak. Network dynamics and cell physiology. *Nature Reviews Molecular Cell Biology*, 2(12):908–916, 2001.
30. N.G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier, Amsterdam, revised edition, 1997.
31. 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.