Qualitative Modeling and Simulation of Genetic Regulatory Networks: From Piecewise-Affine Differential Equations to Reporter Gene Data (and Back)

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

The functioning and development of living organisms is controlled by large and complex networks of genes, proteins, small molecules, and their interactions, so-called *genetic regulatory networks*. The concerted efforts of genetics, molecular biology, biochemistry, and physiology have led to the accumulation of enormous amounts of data on the molecular components of genetic regulatory networks and their interactions. Notwithstanding the advances in the mapping of the network structure, surprisingly little is understood about how the dynamic behavior of the system emerges from the interactions between the network components. This has incited an increasingly large group of researchers to turn from the structure to the behavior of genetic regulatory networks, against the background of a broader movement nowadays often referred to as systems biology (Kitano 2002).

In addition to powerful experimental tools, the study of the dynamic behavior of genetic regulatory networks also requires the support of mathematical and computational tools. Since most networks of biological interest consist of a large number of molecular components involved in complex feedback loops, predicting the behavior of the system by intuition alone quickly becomes unfeasible or fraught with error. The use of mathematical models in combination with computer tools allows for the precise and unambiguous description of a network and the systematic prediction of its behavior.

A variety of methods for the *modeling, analysis, and simulation of genetic regulatory networks* have been proposed in the past forty years (de Jong 2002; Hasty *et al.* 2001; McAdams & Arkin 1998; Smolen, Baxter, & Byrne 2000). The classical approach towards the modeling of genetic regulatory networks is based on the use of ordinary differential equations in combination with numerical simulation. For most networks of biological interest this approach is difficult to put to work in practice though. Numerical simulation requires quantitative data on kinetic parameters and molecular concentrations, which are generally absent, even for well-studied model systems. Moreover, for many purposes it is more important to have a qualitative understanding of the connection between the dynamics of the system and the network structure than to be able to make precise, quantitative predictions.

For all of these reasons, there has been a constant and growing interest in qualitative methods for the modeling and simulation of genetic regulatory networks and other networks of biological interactions (see (de Jong & Ropers 2006; Gagneur & Casari 2005) for recent reviews). From the end of sixties onwards, several approaches have been proposed in mathematical and theoretical biology, most notably Boolean networks (Kauffman 1969; 1993) and generalized logical models (Thomas 1973; Thomas & d'Ari 1990). Our own work has taken a similar point of view, but has been based on the use of a particular class of *piecewise-affine differential equation (PADE)* models proposed in the early seventies by Glass and Kauffman (1973). The PADE models make similar modeling abstractions as the logical models, but are closer to the kinetic models traditionally used. Moreover, the fields of qualitative reasoning (QR) and hybrid systems theory (HST) provide appropriate concepts and techniques to study the dynamics of PADE models.

During my presentation, I will give a rapid introduction to the mathematical properties of PADE models of genetic regulatory networks. In particular, I will show that the phase space can be subdivided into hyperrectangular regions in each of which the system reduces to a linear and uncoupled system of ordinary differential equations with an extremely simple local dynamics (Glass & Kauffman 1973). Since the reduced system will vary from one region to another, the global dynamics may be quite complex however. Moreover, discontinuities may occur at the boundaries between regions, which need to be handled with care in order not to lose dynamical phenomena of biological interest. We have shown how the latter problems can be treated in a mathematically rigorous and practically useful way by extending the differential equations to differential inclusions (Casey, de Jong, & Gouzé 2006; Gouzé & Sari 2002).

In order to describe the qualitative dynamics of the system, we have adapted traditional QR concepts like qualitative states, transitions between qualitative states, and state transition graphs (de Kleer & Brown 1984; Kuipers 1986; Forbus 1984) to the particular constraints of the PADE models. The relation between PADE models and state transition graphs can be formalized by means of the notion of dis-

crete abstraction, familiar from theoretical computer science and used for similar purposes in HST (Alur *et al.* 2000). We have shown that the resulting state transition graph, and thus the qualitative dynamics of the system, are invariant for large ranges of parameter values which can be expressed in the form of inequality constraints (de Jong *et al.* 2004b). Whereas exact numerical values for the parameters are usually not available, the weaker information required for the formulation of the inequality constraints can often be obtained from the experimental literature. Moreover, these constraints can be used for the actual computation of the state transition graph by means of simple, symbolic rules.

The above approach towards the qualitative simulation of genetic regulatory network has been implemented in the computer tool *Genetic Network Analyzer (GNA)* (Batt *et al.* 2005b; de Jong *et al.* 2003).[1] The state transition graphs for networks with more than fifteen genes usually consist of several hundreds or even thousands of states, which make them too large to be analyzed by visual inspection alone. We have therefore coupled GNA with model-checking tools for the automated verification of dynamical properties expressed in temporal logic (Batt *et al.* 2005a; 2005b), following similar ideas in QR (Shults & Kuipers 1997) and in HST (Alur *et al.* 2000).

GNA has been used for the analysis of several genetic regulatory networks, such as the networks controlling initiation of sporulation in the soil bacterium *Bacillus subtilis* (de Jong *et al.* 2004a) and quorum sensing in the pathogenic bacterium *Pseudomonas aeruginosa* (Usseglio Viretta & Fussenegger 2004). We currently focus on its application in the context of the nutritional stress response of *Escherichia coli*, a well-known model bacterium. In particular, we have developed qualitative models of the adaptation of the growth of an *E. coli* cell upon depletion of the carbon sources in its environment. This carbon starvation response is controlled by a large and complex regulatory network that we are modelling in a step-wise, modular fashion. This has resulted in PADE models of about a dozen genes that account for some well-known properties of the response of the cell to a lack of carbon sources (Ropers *et al.* 2006). In addition, we have made novel predictions that are currently being tested experimentally, by measuring the time-course of gene expression with the help of so-called gene reporter systems. These experimental tools allow the dynamics of the state variables to be monitored with high precision and high sampling density.

The experimental data obtained in this way will not only allow us to validate the PADE models of the carbon starvation response network, but also provide input for the revision and completion of the models. Along these lines, we have started to work on the identification of PADE models of genetic regulatory network from time-course gene expression data. Adapting existing approaches in HST (Ferrari-Trecate *et al.* 2003) to the particular constraints of the class of models we consider, an algorithm for the identification of regulatory interactions between genes in the network has been developed (Drulhe *et al.* 2006). Ideally, future extensions of this algorithm should take into account *a priori* knowledge on existing, experimentally validated interactions and propose minimal extensions of the model to account for conflicts between predictions and observations. Such a network identification approach would be an important step in closing the empirical cycle running from models to data, and back.

In summary, in my presentation I will show how ideas that originated in QR and HST have been put to use in a particular application domain, the analysis of genetic regulatory networks. Instead of reusing standard QR and HST approaches, which usually apply to generic classes of models, we have developed tailored algorithms that maximally exploit the favorable mathematical properties of the class of PADE models we consider (see, *e.g.*, (Nishida & Doshita 1995; Sacks 1990) for ideas similar in spirit). In combination with model-checking tools for the automated verification of model properties, this strategy has allowed us to improve the scalability of qualitative simulation and make biologically relevant predictions.

[1] GNA is currently distributed by the company Genostar SA, but remains freely available for non-profi t academic research purposes (`http://www-helix.inrialpes.fr/gna`).

References

Alur, R.; Henzinger, T.; Lafferriere, G.; and Pappas, G. 2000. Discrete abstractions of hybrid systems. *Proc. IEEE* 88(7):971–984.

Batt, G.; Ropers, D.; de Jong, H.; Geiselmann, J.; Mateescu, R.; Page, M.; and Schneider, D. 2005a. Analysis and verification of qualitative models of genetic regulatory networks: A model-checking approach. In Kaelbling, L., ed., *Proc. IJCAI-05*, 370–375. San Francisco, CA: Morgan Kaufmann.

Batt, G.; Ropers, D.; de Jong, H.; Geiselmann, J.; Mateescu, R.; Page, M.; and Schneider, D. 2005b. Validation of qualitative models of genetic regulatory networks by model checking: Analysis of the nutritional stress response in *Escherichia coli*. *Bioinformatics* 21(Suppl 1):i19–i28.

Casey, R.; de Jong, H.; and Gouzé, J.-L. 2006. Piecewise-linear models of genetic regulatory networks: Equilibria and their stability. *J. Math. Biol.* 52(1):27–56.

de Jong, H., and Ropers, D. 2006. Qualitative approaches towards the analysis of genetic regulatory networks. In Szallasi, Z.; Periwal, V.; and Stelling, J., eds., *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. Cambridge, MA: MIT Press. 125–148.

de Jong, H.; Geiselmann, J.; Hernandez, C.; and Page, M. 2003. Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics* 19(3):336–344.

de Jong, H.; Geiselmann, J.; Batt, G.; Hernandez, C.; and Page, M. 2004a. Qualitative simulation of the initiation of sporulation in *B. subtilis*. *Bull. Math. Biol.* 66(2):261–299.

de Jong, H.; Gouzé, J.-L.; Hernandez, C.; Page, M.; Sari, T.; and Geiselmann, J. 2004b. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull. Math. Biol.* 66(2):301–340.

de Jong, H. 2002. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.* 9(1):67–103.

de Kleer, J., and Brown, J. 1984. A qualitative physics based on confluences. *Artif. Intell.* 24(1-3):7–83.

Drulhe, S.; Ferrari-Trecate, G.; de Jong, H.; and Viari, A. 2006. Reconstruction of switching thresholds in piecewise-affine models of genetic regulatory networks. In Hespanha, J., and Tiwari, A., eds., *Proc. HSCC 2006*, volume 3927 of *LNCS*. Berlin: Springer-Verlag. 184–199.

Ferrari-Trecate, G.; Muselli, M.; Liberati, D.; and Morari, M. 2003. A clustering technique for the identification of piecewise affine systems. *Automatica* 39(1):205–218.

Forbus, K. 1984. Qualitative process theory. *Artif. Intell.* 24(1-3):85–168.

Gagneur, J., and Casari, G. 2005. From molecular networks to qualitative cell behavior. *FEBS Lett.* 579(8):1867–1871.

Glass, L., and Kauffman, S. 1973. The logical analysis of continuous non-linear biochemical control networks. *J. Theor. Biol.* 39(1):103–129.

Gouzé, J.-L., and Sari, T. 2002. A class of piecewise linear differential equations arising in biological models. *Dyn. Syst.* 17(4):299–316.

Hasty, J.; McMillen, D.; Isaacs, F.; and Collins, J. 2001. Computational studies of gene regulatory networks: *In numero* molecular biology. *Nat. Rev. Genet.* 2(4):268–279.

Kauffman, S. 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.* 22(3):437–467.

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

Kitano, H. 2002. Systems biology: A brief overview. *Science* 295(5560):1662–1664.

Kuipers, B. 1986. Qualitative simulation. *Artif. Intell.* 29(3):289–388.

McAdams, H., and Arkin, A. 1998. Simulation of prokaryotic genetic circuits. *Ann. Rev. Biophys. Biomol. Struct.* 27:199–224.

Nishida, T., and Doshita, S. 1995. Qualitative analysis of behavior of systems of piecewise linear differential equations with two state variables. *Artif. Intell.* 75(1):3–29.

Ropers, D.; de Jong, H.; Page, M.; Schneider, D.; and Geiselmann, J. 2006. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *BioSystems*. In press.

Sacks, E. 1990. Automatic qualitative analysis of dynamic systems using piecewise linear approximations. *Artif. Intell.* 41(3):313–364.

Shults, B., and Kuipers, B. 1997. Proving properties of continuous systems: Qualitative simulation and temporal logic. *Artif. Intell.* 92(1-2):91–130.

Smolen, P.; Baxter, D.; and Byrne, J. 2000. Modeling transcriptional control in gene networks: Methods, recent results, and future directions. *Bull. Math. Biol.* 62(2):247–292.

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

Thomas, R. 1973. Boolean formalization of genetic control circuits. *J. Theor. Biol.* 42:563–585.

Usseglio Viretta, A., and Fussenegger, M. 2004. Modeling the quorum sensing regulatory network of human-pathogenic *Pseudomonas aeruginosa*. *Biotechnol. Progr.* 20(3):670–678.