Validation of Qualitative Models of Genetic Regulatory Networks by Model Checking: Analysis of the Nutritional Stress Response in *Escherichia coli*

Grégory Batt [1], Delphine Ropers [1],
Hidde de Jong [1], Johannes Geiselmann [2],
Radu Mateescu [1], Michel Page [1,3], Dominique Schneider [2]
[1]INRIA Rhône-Alpes, Montbonnot, France,
[2]Laboratoire Adaptation et Pathogénie des Microorganismes,
CNRS UMR 5163, Université Joseph Fourier, Grenoble, France
[3]Université Pierre Mendès France, Grenoble, France

1 Introduction

The functioning and development of living organisms is controlled by large and complex networks of genes, proteins, small molecules, and their mutual interactions, so-called *genetic regulatory networks*. In order to gain an understanding of how the behavior of an organism – *e.g.*, the response of a bacterial cell to a physiological or genetic perturbation – emerges from such a network of interactions, we need mathematical and computational tools for modeling and simulation [10]. The predictions obtained through the application of these tools have to be confronted with experimental data. This gives rise to the problem of *model validation*, the assessment of the adequacy of a model by comparing its predictions with observations, either already available in the literature or obtained through novel experiments suggested by the model.

The main challenges of model validation are twofold. First of all, the precision of the model predictions and the experimental data need to be brought

in agreement. At present, quantitative information on kinetic parameters is usually absent, thus making traditional numerical models and analysis techniques difficult to apply. In addition, numerical predictions on the dynamics of the system are difficult to verify, because available data are mostly qualitative in nature. A second challenge is to ensure that the comparison of model predictions with experimental data is efficient and reliable. Models of genetic regulatory networks of biological interest may become quite large, as they include many genes and proteins, thus making manual verification of dynamical properties error-prone or even practically infeasible. In this paper, we propose an approach towards model validation addressing the above two challenges (see [4] for an extended description).

2 Model validation by model checking

The approach extends our previous work on a method for the *qualitative modeling and simulation* of genetic regulatory networks [12], supported by the computer tool *Genetic Network Analyzer (GNA)* [11]. This method is based on a class of piecewise-linear (PL) differential equation models originally proposed by Glass and Kauffman [15]. While abstracting from the precise biochemical reaction mechanisms involved, the PL models capture essential aspects of gene regulation. Moreover, their simple mathematical form permits a qualitative analysis of the dynamics of genetic regulatory systems to be carried out. More precisely, the dynamics of a PL system can be described by means of a so-called *qualitative transition system*. This qualitative transition system consists of a partition of the phase space into a set of so-called domains that are regions where the system behaves in a qualitatively-homogeneous way, a transition relation where the transitions between domains correspond to solution trajectories connecting adjacent domains, and a labeling function describing qualitatively the dynamical properties of the system in the domains, notably the sign pattern of the derivatives of the variables [5]. Given that the variables denote protein concentrations, the qualitative transition system provides predictions on the possible ways in which the sign pattern of the derivatives of the protein concentrations can evolve, a level of precision that is well-adapted to currently-available data. Note that we speak of the *sign pattern* of the derivatives. This comes from the way we deal with mathematical problems brought about by the piecewise nature of the differential equations we consider. Following an approach

widely-used in control theory, we extend differential equations into differential inclusions. Given that a differential inclusion may not have a unique solution, the sign of the solution derivatives may not be unique too. However, we have introduced the notion of derivative sign pattern, and proved that it is unique in each domain. We have shown that the qualitative transition system is invariant for sets of parameter values defined by inequality constraints on the parameters that can be easily inferred from the experimental literature. Moreover, the qualitative transition system can be computed by means of simple symbolic rules using these inequality constraints.

The model-validation approach integrates the above qualitative modeling and simulation method with *model-checking* techniques. These techniques allow for the verification of properties of the behavior of discrete transition systems, expressed as formulas in a *temporal logic* [9]. Highly-efficient algorithms have been developed and implemented in tools called model checkers for supporting this verification task. In order to verify whether the predictions of the system behavior are consistent with experimental data, we express the observed properties as temporal logic formulas, compute the qualitative transition system using qualitative simulation, and use model checkers to verify whether the qualitative transition system satisfies the temporal logic formulas. If it does not, then the PL model is inconsistent with the experimental data and may need to be revised or extended. The combination of qualitative modeling and simulation and model checking allows large and complex networks to be verified, with the guarantee that no model is falsely ruled out.

3 Analysis of nutritional stress response in *E. coli*

The model validation approach proposed in this paper has been applied to the analysis of the network controlling the *nutritional stress response* in *Escherichia coli*. In case of nutritional stress, an *E. coli* population abandons exponential growth and enters a non-growth state called stationary phase [17]. On the molecular level, this growth-phase transition is controlled by a complex genetic regulatory network integrating various environmental signals [16]. Understanding the molecular basis of this essential developmental decision has been the focus of extensive studies for decades. However, there

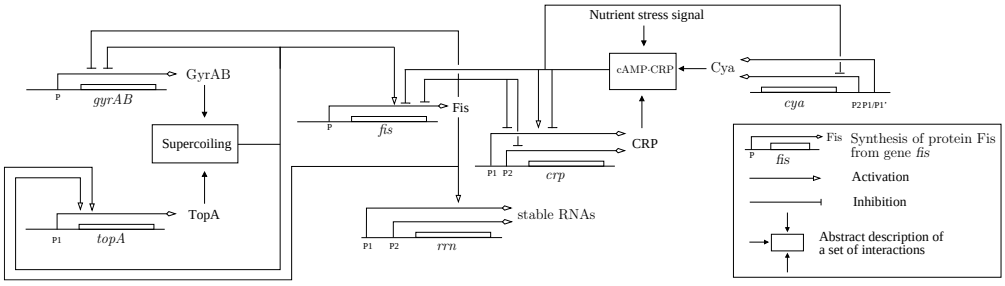

Figure 1: Network of key genes, proteins, and regulatory interactions involved in the carbon starvation response in *E. coli*.

$$\dot{x}_{topA} = \kappa^1_{topA} + \kappa^2_{topA}\ s^+(x_{gyrAB}, \theta^3_{gyrAB})\ s^-(x_{topA}, \theta^1_{topA})\ s^+(x_{fis}, \theta^4_{fis}) - \gamma_{topA}\ x_{topA}$$

$$0 < \kappa^1_{topA}/\gamma_{topA} < \theta^1_{topA} < \theta^2_{topA} < \theta^3_{topA} < (\kappa^1_{topA} + \kappa^2_{topA})/\gamma_{topA} < max_{topA}$$

Figure 2: PL differential equation and parameter inequalities for the TopA protein in the network of Figure 1.

is currently no global understanding of how the response of the cell emerges from the network of molecular interactions. Moreover, with some exceptions, numerical values for the parameters characterizing the interactions and the molecular concentrations are absent from the literature, which makes it difficult to apply traditional simulation methods. Based on data in the experimental literature, we have constructed a PL model including key proteins and their interactions involved in the response to a particular nutritional stress, namely, carbon starvation [20]. The model includes genes involved in the transduction of the carbon starvation signal (*crp* and *cya*), metabolism (*fis*), cellular growth (*rrn* genes), and DNA supercoiling, an important modulator of gene expression (*topA* and *gyrAB*) (see Figures 1 and 2).

Using a new version of GNA supporting our model-validation approach, we have simulated two phenomena, namely the transition from exponential to stationary phase, and the reentry into exponential phase after a carbon upshift. In order to validate the model, the simulation results have been compared with the available experimental data, using the model checker NuSMV [8]. The predictions for the entry into stationary phase were found to be consistent with most of the observed properties, such as the observed decrease of the concentration the protein Fis [1], but not with the observed decrease of

Biological property	Temporal logic formula	Result
Fis concentration decreases and becomes steady in stationary phase	$\mathbf{EF}(\dot{x}_{fis} < 0 \wedge \mathbf{EF}(\dot{x}_{fis} = 0 \wedge x_{rrn} < \theta_{rrn}))$	True
DNA supercoiling decreases during transition to stationary phase	$\mathbf{EF}((\dot{x}_{gyrAB} < 0 \vee \dot{x}_{topA} > 0) \wedge x_{rrn} < \theta_{rrn})$	False

Figure 3: Experimentally-observed properties of the network, their translation into temporal logic formulas (not detailed) and the result of the verification of the formula by means of NuSMV.

the DNA supercoiling level [3] (see Figure 3). Consequently, the model has to be revised. In [20], we propose to extend the model with interactions not yet identified or with regulators not yet considered. Another prediction, the occurrence of damped oscillations in some of the protein concentrations after a carbon upshift, is a surprising result, and is currently subject to experimental investigation in our laboratory.

4 Related work

Model-checking or other formal verification techniques have been used before in systems biology for analyzing genetic, metabolic, signal-transduction, and cell-cycle networks. Most approaches start from discrete models, such as Petri nets [18], process algebras [19], concurrent transition systems [7], rewriting logic [13], and Boolean networks and their generalizations [6]. In this paper we show that model-checking techniques can also be used for more conventional continuous models, in particular differential equation models, when using qualitative abstractions to discretize the dynamics of the system. In comparison with ideas along the same line [2, 14, 21], our approach is adapted to a particular class of PL differential equations with favorable mathematical properties, allowing the development of tailored algorithms that scale up well to models of large and complex genetic regulatory networks.

References

[1] T. Ali Azam, A. Iwata, A. Nishimura, S. Ueda, and A. Ishihama. Growth phase-dependent variation in protein composition of the *E. coli* nucleoid. *J. Bacteriol.*, 181(20):6361–6370, 1999.

[2] M. Antoniotti, C. Piazza, A. Policriti, M. Simeoni, and B. Mishra. Taming the complexity of biochemical models through bisimulation and collapsing: Theory and practice. *Theor. Comput. Sci.*, 325(1):45–67, 2004.

[3] V.L. Balke and J.D. Gralla. Changes in the linking number of supercoiled DNA accompany growth transitions in *Escherichia coli*. *J. Bacteriol.*, 169(10):4499–4506, 1987.

[4] G. Batt, D. Ropers, H. de Jong, J. Geiselmann, R. Mateescu, M. Page and D. Schneider. 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, 2005.

[5] G. Batt, D. Ropers, H. de Jong, J. Geiselmann, M. Page and D. Schneider. Qualitative analysis and verification of hybrid models of genetic regulatory networks: Nutritional stress response in *Escherichia coli*. In M. Morari and L. Thiele, editors, *HSCC'05*, *LNCS* 3414, 134–150. Springer, 2005.

[6] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. A fruitful application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. *J. Theor. Biol.*, 229(3):339–347, 2004.

[7] N. Chabrier-Rivier, M. Chiaverini, V. Danos, F. Fages, and V. Schächter. Modeling and querying biomolecular interaction networks. *Theor. Comput. Sci.*, 325(1):25–44, 2004.

[8] A. Cimatti, E.M. Clarke, E. Giunchiglia, F. Giunchiglia, M. Pistore, M. Roveri, R. Sebastiani, and A. Tacchella. NuSMV2: An opensource tool for symbolic model checking. In E. Brinksma and K.G. Larsen, editors, *CAV'02*, *LNCS* 2404, 359–364. Springer, 2002.

[9] E.M. Clarke, O. Grumberg, and D.A. Peled. *Model Checking*. MIT Press, 1999.

[10] H. de Jong. Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol.*, 9(1):69–105, 2002.

[11] 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.

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

[13] S. Eker, M. Knapp, K. Laderoute, P. Lincoln, J. Meseguer, and M.K. Sönmez. Pathway logic: Symbolic analysis of biological signaling. In R.B. Altman *et al.*, editors, *PSB'02*, 400–412. World Scientific Publishing, 2002.

[14] R. Ghosh and C.J. Tomlin. Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-Notch protein signalling. *Systems Biology*, 1(1):170–183, 2004.

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

[16] R. Hengge-Aronis. The general stress response in *E. coli*. In G. Storz and R. Hengge-Aronis, editors, *Bacterial Stress Responses*, 161–177. ASM Press, 2000.

[17] G.W. Huisman, D.A. Siegele, M.M. Zambrano, and R. Kolter. Morphological and physiological changes during stationary phase. In F.C. Neidhardt *et al.*, editors, *Escherichia coli and Salmonella: Cellular and Molecular Biology*, 1672–1682. ASM Press, 1996.

[18] I. Koch, B.H. Junker, and M. Heiner. Application of Petri net theory for modelling and validation of the sucrose breakdown pathway in the potato tuber. *Bioinformatics*, 21(7):1219–1226, 2005.

[19] A. Regev, W. Silverman, and E. Shapiro. Representation and simulation of biochemical processes using the π-calculus process algebra. In R.B. Altman *et al.*, editors, *PSB'01*, 459–470. World Scientific Publishing, 2001.

[20] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann. Qualitative simulation of the carbon starvation response in *Escherichia coli*. *BioSystems*, 2005. To appear.

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