BIOINFORMATICS

Vol. 21 Suppl. 1 2005, pages i19–i28
doi:10.1093/bioinformatics/bti1048

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] and Dominique Schneider[2]

[1]INRIA Rhône-Alpes, Montbonnot, France, [2]Laboratoire Adaptation et Pathogénie des Microorganismes, CNRS UMR 5163, Université Joseph Fourier, Grenoble, France and [3]Université Pierre Mendès France, Grenoble, France

Received on January 15, 2005; accepted on March 27, 2005

ABSTRACT

Motivation: The modeling and simulation of genetic regulatory networks have created the need for tools for model validation. The main challenges of model validation are the achievement of a match between the precision of model predictions and experimental data, as well as the efficient and reliable comparison of the predictions and observations.

Results: We present an approach towards the validation of models of genetic regulatory networks addressing the above challenges. It combines a method for qualitative modeling and simulation with techniques for model checking, and is supported by a new version of the computer tool Genetic Network Analyzer (GNA). The model-validation approach has been applied to the analysis of the network controlling the nutritional stress response in *Escherichia coli*.

Availability: GNA and the model of the stress response network are available at http://www-helix.inrialpes.fr/gna

Contact: Hidde.de-Jong@inrialpes.fr

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, the 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 (de Jong, 2002). 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. The approach extends our previous work on a method for the *qualitative modeling and simulation* of genetic regulatory networks, supported by the computer tool *Genetic Network Analyzer* (GNA) (de Jong *et al*., 2003, 2004). This method is based on a class of *piecewise-linear* (PL) *differential equations* that permits a coarse-grained, qualitative analysis of the network dynamics to be carried out. Instead of numerical values for the parameters, the method uses inequality constraints that can be inferred from the experimental literature. It yields 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. The novelty of the model-validation approach is that it integrates qualitative modeling and simulation with *model-checking* techniques (Clarke *et al*., 1999) to verify whether the predictions of the system behavior are consistent with experimental data.

[*] To whom correspondence should be addressed.

In particular, the measured evolution of the derivative sign pattern or other experimental observations can be formalized as properties in temporal logic, while model-checking techniques verify whether the predictions account for these properties. If they do not, then the 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.

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 (Koch *et al*., 2005), process algebras (Regev *et al*., 2001), concurrent transition systems (Chabrier-Rivier *et al*., 2004), rewriting logic (Eker *et al*., 2002), and Boolean networks and their generalizations (Bernot *et al*., 2004). 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 (Antoniotti *et al*., 2004; Ghosh *et al*., 2003; Shults and Kuipers, 1997), 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.

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 (Huisman *et al*., 1996). At the molecular level, this growth phase transition is controlled by a complex genetic regulatory network (Hengge-Aronis, 2000). We have constructed a model including key proteins and their interactions involved in the carbon starvation response, and validated this model by comparing the predicted temporal evolution of the protein concentrations with available experimental data, both during the transition from exponential to stationary phase, and during the reentry into exponential phase after a nutrient upshift. Although some of the predictions have thus been confirmed, one prediction has been refuted, suggesting model revisions. Another prediction concerns a surprising phenomenon that has not been experimentally investigated yet.

In the next section of the paper, we briefly outline the qualitative modeling and simulation method used to predict the behavior of genetic regulatory networks. Section 3 describes the model-checking approach towards model validation in some detail, as well as its computer implementation in GNA. The initial results of the validation of our model of the *E.coli* nutritional stress response are summarized in Section 4, followed by a discussion of the achievements in the final section.

2 QUALITATIVE SIMULATION

The method for the qualitative modeling and simulation of genetic regulatory networks that we use in this paper is a refinement of the method that we previously presented (de Jong *et al*., 2003, 2004). It is based on a qualitative abstraction that preserves stronger properties of the network dynamics, in particular the sign patterns of the derivatives of the concentration variables. This information is critical for the experimental validation of models of genetic regulatory networks, since experimental measurements of the system dynamics by means of quantitative RT–PCR, reporter genes and DNA microarrays usually result in observations of changes in the sign of the derivatives. We will provide an intuitive overview of the method, using a simple example. For technical details, the reader is referred to Batt *et al*. (2005).

Figure 1a shows a network consisting of two genes. When a gene (*a* or *b*) is expressed, the corresponding protein (A or B) is synthesized. Proteins A and B regulate the expression of genes *a* and *b*. More specifically, protein B inhibits the expression of gene *a* above a certain threshold concentration, whereas protein A inhibits the expression of gene *b* above a threshold concentration, and the expression of its own gene above a second, higher threshold concentration. The degradation of the proteins is not regulated.

The dynamics of genetic regulatory networks can be modeled by a class of *piecewise-linear* (PL) *differential equation* models originally introduced by Glass and Kauffman (1973). The example network gives rise to the following model:

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

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

where x_a and x_b denote the concentrations of proteins A and B, $\dot{x}_a$ and $\dot{x}_b$ their time derivatives, θ_a^1, θ_a^2 and θ_b threshold concentrations, κ_a and κ_b synthesis parameters, and γ_a and γ_b degradation parameters. The step function $s^-(x, \theta)$ evaluates to 1, if $x < \theta$, and to 0, if $x > \theta$. Step functions are approximations of the steep sigmoid functions often characterizing gene regulation, preserving their non-linear, switch-like character. As a consequence, PL models are coarse-grained models that abstract from the fine aspects of gene regulation, such as stochasticity, but have been shown adequate for a wide range of applications (see de Jong *et al*., 2004, for references).

Equations (1) and (2) describe the rate of change of the protein concentrations. Equation (2) states that protein B is produced (at a rate κ_b), if and only if $s^-(x_a, \theta_a^1) = 1$, that is, if and only if $x_a < \theta_a^1$. This captures the inhibition of the expression of gene *b* by protein A. Equation (1) states that protein A is produced (at a rate κ_a), if and only if neither $x_a > \theta_a^2$ nor $x_b > \theta_b$. Both proteins are degraded at a rate proportional to their own concentration.

Mathematical analysis of this model reveals that mere knowledge of the relative order of the threshold parameter(s) and

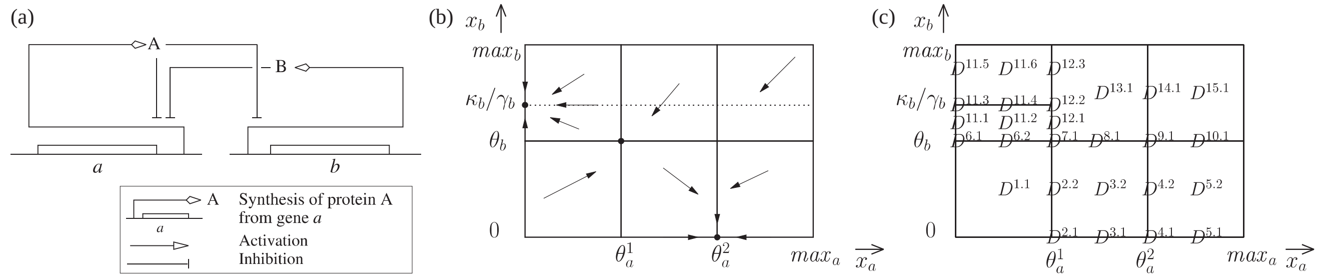


Fig. 1. (**a**) Simple genetic regulatory network consisting of two genes. (**b**) Sketch of the dynamics in the phase space of the two-gene network. The system has three equilibrium points, represented by dots. (**c**) Domain partition of the phase space.

the quotient of the synthesis and degradation parameter, for each of the two variables, is sufficient to sketch the flow in the phase space. This result has been shown to be generalizable to the whole class of PL models considered here. More particularly, assuming that

$$0 < \theta_a^1 < \theta_a^2 < \frac{\kappa_a}{\gamma_a} < max_a, \tag{3}$$

$$0 < \theta_b < \frac{\kappa_b}{\gamma_b} < max_b, \tag{4}$$

the phase space can be partitioned into hyperrectangular boxes, called *domains*, in which the flow is qualitatively identical, in the sense that either all solutions of the system traverse a domain instantaneously (*instantaneous* domain) or they have the same derivative sign pattern while remaining in the domain (*persistent* domain). Figures 1b and c represent the flow in the phase space and the domain partition of the phase space for the two-gene example. $D^{2.2}$ is an instantaneous domain, while $D^{1.1}$, $D^{4.2}$ and $D^{4.1}$ are persistent. Moreover, the latter domain coincides with an equilibrium point of the system. The domain partition is finer grained than the one used in our earlier work, for which the property that all solutions in a domain have the same derivative sign pattern does not generally hold.[1]

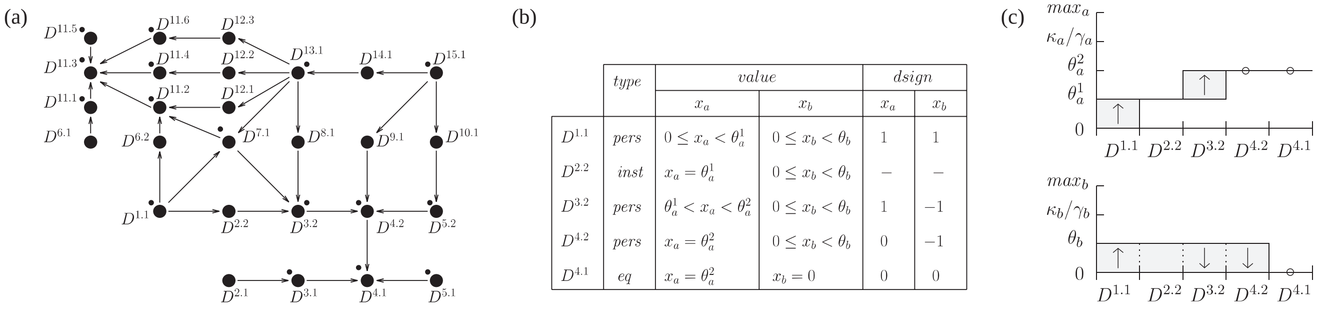
Using the domain partition of the phase space, together with the qualitative characterization of the dynamics in each of the domains, we can discretize the continuous dynamics. In the resulting abstract description, the state of the system is represented by a domain and its associated dynamical properties. There exists a transition from a domain D to another domain D', if and only if there exists a solution reaching D' from D, without leaving $D \cup D'$. This naturally leads to the introduction of a so-called *qualitative transition system*, consisting of the set of all domains, the set of all transitions between the domains and a labeling function that associates to every domain the sign of the derivatives of the concentration variables and an indication of whether the domain is persistent or instantaneous. The graph representation of the qualitative transition system is called a *state transition graph* and the domains are also called *qualitative states* (or qualitative *equilibrium* states, if the domains consist in equilibrium points). Figure 2 shows the qualitative transition system of the two-gene model.

A sequence of qualitative states in the state transition graph is called a *path*. A path qualitatively describes a possible behavior of the system. In our two-gene example, ($D^{1.1}$, $D^{2.2}$, $D^{3.2}$, $D^{4.2}$, $D^{4.1}$) is a path leading to a qualitative equilibrium state (Fig. 2c). The qualitative transition system is defined such that it provides a *conservative approximation* of the dynamics of the original PL system, in the sense that to every solution of the model corresponds a path in the state transition graph. Note that the converse is not true: some paths may not correspond to any solution, and therefore represent spurious behaviors. The state transition graph has been shown to be invariant for all values of the parameters satisfying the parameter inequality constraints.

Simple rules have been formulated for the symbolic computation of the qualitative transition system from a PL model of the network. These rules exploit the favorable analytical properties of the class of PL models, thus allowing the qualitative states, the transitions between qualitative states, and the labeling function to be inferred from the parameter inequality constraints. The implementation of these rules has resulted in a new version of the computer tool GNA (de Jong *et al*., 2003). The new version of GNA, available at http://www-helix.inrialpes.fr/gna, has also been equipped with a strongly improved graphical user interface.

The paths in the state transition graph correspond to predicted qualitative behaviors of the system and can be compared with experimental data. The resulting model-validation problem is easy to solve for the simple two-gene example. For instance, the observation shown in Figure 3 is consistent with predictions, since there exists a path, ($D^{1.1}$, $D^{2.2}$, $D^{3.2}$, $D^{4.2}$, $D^{4.1}$), verifying the observed derivative sign pattern (Fig. 2c). However, the analysis of realistic models leads to large state transition graphs, which

[1] In this simple presentation of the method, we omit the problems raised by the discontinuities in the right-hand side of the PL differential equations, whose treatment goes beyond the scope of this article. See de Jong *et al*. (2004) and Gouzé and Sari (2002) for a detailed description.



	type	*value*		*dsign*	
		x_a	x_b	x_a	x_b
$D^{1.1}$	*pers*	$0 \leq x_a < \theta_a^1$	$0 \leq x_b < \theta_b$	1	1
$D^{2.2}$	*inst*	$x_a = \theta_a^1$	$0 \leq x_b < \theta_b$	–	–
$D^{3.2}$	*pers*	$\theta_a^1 < x_a < \theta_a^2$	$0 \leq x_b < \theta_b$	1	−1
$D^{4.2}$	*pers*	$x_a = \theta_a^2$	$0 \leq x_b < \theta_b$	0	−1
$D^{4.1}$	*eq*	$x_a = \theta_a^2$	$x_b = 0$	0	0

Fig. 2. Qualitative transition system of the two-gene model, with (**a**) the state transition graph and (**b**) the properties of some of the qualitative states in the graph. The following abbreviations have been used: *pers*, persistent state; *inst*, instantaneous state; *eq*, equilibrium state; *dsign*, derivative sign. The numbers −1, 0 and 1 denote the sign of the derivative of the protein concentrations. In instantaneous domains, the derivatives are not defined (Batt *et al*., 2005), indicated by a dash. The equilibrium states are $D^{4.1}$, $D^{7.1}$ and $D^{11.3}$, while dots next to states represent self-transitions. (**c**) Temporal evolution of the concentrations of proteins A and B in the path ($D^{1.1}$, $D^{2.2}$, $D^{3.2}$, $D^{4.2}$, $D^{4.1}$). Arrows indicate the sign of the derivatives for persistent states (up arrow for 1, down arrow for −1 and open circle for 0).

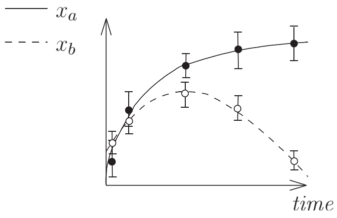


Fig. 3. Hypothetical experimental observation of the temporal evolution of the concentrations of proteins A and B.

make manual verification of dynamical properties error-prone or even practically infeasible. This has motivated the development of an automated, efficient method for model validation.

3 MODEL VALIDATION BY MODEL-CHECKING

Our model-validation approach combines the qualitative modeling and simulation method outlined above with techniques for *model checking* (Clarke *et al*., 1999). These techniques allow for the verification of properties of the behavior of discrete transition systems, expressed as formulas in some *temporal logic*. Using suitable model-checking algorithms and tools, it is possible to automatically and efficiently test whether the system satisfies the property. Model checking has been successfully applied to the verification of software, telecommunication systems, electronic circuits and other complex systems (for examples, see http://www.inrialpes.fr/vasy/cadp/case-studies/ and http://nusmv.irst.itc.it/).

Various model-checking frameworks exist, differing by their expressiveness, user-friendliness and computational efficiency. For the sake of simplicity, we focus here on one particular framework, in which the discrete transition system takes the form of a *Kripke structure*, and the behavioral properties are expressed in *Computation Tree Logic* (CTL) (Clarke *et al*., 1999). We describe the relation between qualitative simulation and model checking at the conceptual level, and briefly present an extension of GNA that connects the qualitative simulator with the model checker NuSMV. However, we emphasize that our approach is not restricted to CTL model-checking, and allows other more expressive temporal logics to be used as well (Section 3.3).

3.1 Translate qualitative transition system into Kripke structure

As a preliminary step, we introduce a set of *atomic propositions* to describe the state of the system. To be more precise, the set of atomic propositions we use consists of simple expressions describing the range of a protein concentration (e.g. $value\_x_a < \theta_a^1$), the sign of the derivative of a protein concentration (e.g. $dsign\_x_a = 1$) or the type of a state (e.g. $type = pers$). That is, in the example of Figure 2, the set of atomic propositions AP is given by

$$\begin{aligned} AP = \{&value\_x_a = 0, value\_x_a > 0, value\_x_a < \theta_a^1, \ldots, \\ &dsign\_x_a = -1, dsign\_x_a = 0, dsign\_x_a = 1, \ldots, \\ &type = pers, type = inst, type = eq\}. \end{aligned}$$

In general, a Kripke structure over a set of atomic propositions AP is a triple $\langle S, R, L\rangle$, where S is a set of states, $R \subseteq S \times S$ a total transition relation between the states, and $L{:}S \to 2^{AP}$ a labeling function that associates to each state, the set of atomic propositions true in that state (Clarke *et al*., 1999). The qualitative transition systems introduced in Section 2 are Kripke structures. As an illustration, the qualitative transition system of the two-gene network, graphically represented in Figure 2, can be alternatively represented as the triple $\langle S, R, L \rangle$, where,

$$S = \{D^{1.1}, D^{2.1}, D^{2.2}, \ldots, D^{15.1}\},$$

$$R = \{(D^{1.1}, D^{2.2}), (D^{1.1}, D^{6.2}), \ldots, (D^{15.1}, D^{14.1})\},$$

$$L : \begin{cases} L(D^{1.1}) = \{value\_x_a \geq 0, value\_x_a < \theta_a^1, \ldots, \\ \qquad value\_x_b \geq 0, value\_x_b < \theta_b, \ldots, \\ \qquad dsign\_x_a = 1, dsign\_x_b = 1, \\ \qquad type = pers\}, \\ L(D^{2.1}) = \{value\_x_a = \theta_a^1, \ldots, type = inst\}, \\ \ldots \\ L(D^{15.1}) = \{value\_x_a > \theta_a^2, \ldots, type = pers\}. \end{cases}$$

3.2 Express dynamical properties in temporal logic

A CTL formula is built upon atomic propositions. The usual operators from propositional logic, such as negation ($\neg$), logical or ($\vee$), logical and ($\wedge$), and implication ($\rightarrow$), can also be used. In addition, CTL provides two types of operators: *path quantifiers*, **E** and **A**, and *temporal operators*, such as **F** and **G**. Path quantifiers are used to specify that a property p is satisfied by some ($\mathbf{E}p$) or every ($\mathbf{A}p$) path starting from a given state. Temporal operators are used to specify that, given a state and a path starting from that state, a property p holds for some ($\mathbf{F}p$) or for every ($\mathbf{G}p$) state of the path. Each path quantifier must be paired with a temporal operator.[2]

Informally speaking, path quantifiers are used to quantify over the possible behaviors of the system, since $\mathbf{A}p$ means that p must hold for every behavior, and $\mathbf{E}p$ means that p must hold for at least one behavior. Temporal operators are used to specify, given a behavior, temporal constraints on the state of the system, since $\mathbf{F}p$ and $\mathbf{G}p$ can be interpreted as meaning that for some future state and for every future state, respectively, p must hold.

How can the properties of interest for model validation be expressed as CTL formulas? This can be illustrated by means of the hypothetical experimental observation in Figure 3. The observation allows us to infer that the system reaches a state in which the concentrations of proteins A and B are both increasing, and from that state onwards, a second state in which the concentration of protein A is increasing and that of B decreasing. The property can be formalized by the CTL formula

$$\begin{aligned} &\mathbf{EF}(dsign\_x_a = 1 \wedge dsign\_x_b = 1 \wedge \\ &\quad \mathbf{EF}(dsign\_x_a = 1 \wedge dsign\_x_b = -1)). \end{aligned} \tag{5}$$

The expression $\mathbf{EF}p$ means that there exists at least one path (**E**) leading to a future state (**F**) where p holds, thus expressing the *reachability* of that state. More generally, any time-series measurement of gene expression can be given as a combination of **EF** operators with conjunctions of atomic propositions describing the derivative sign patterns.

When understood in a broader sense, model validation does not just amount to the comparison of model predictions with time-series measurements of protein concentrations, but also involves the testing of other biologically meaningful properties (Bernot *et al*., 2004; Chabrier-Rivier *et al*., 2004). Suppose that we are interested in knowing whether every behavior of the system will eventually satisfy some property, for example, reach a specific state. We can investigate this by means of formulas using **AF** operators, which express the *inevitability* of a behavior. The following CTL formula expresses the conjecture that the two-gene network of Figure 1 will inevitably reach the equilibrium state $D^{11.3}$:

$$\mathbf{AF}(type = eq \wedge value\_x_a = 0). \tag{6}$$

As a second example, CTL can be used to express the sufficiency of certain conditions to cause the system to behave in a particular way. For example, one could ask, given that protein B is the only regulator of gene a, whether a high concentration of protein B guarantees the eventual disappearance of protein A. This *response* property can be expressed by the CTL formula

$$\mathbf{AG}(value\_x_b > \theta_b \rightarrow \mathbf{AF}value\_x_a = 0), \tag{7}$$

where $\mathbf{AG}p$ specifies that the property p must hold for every state.

3.3 Check if model satisfies dynamical properties

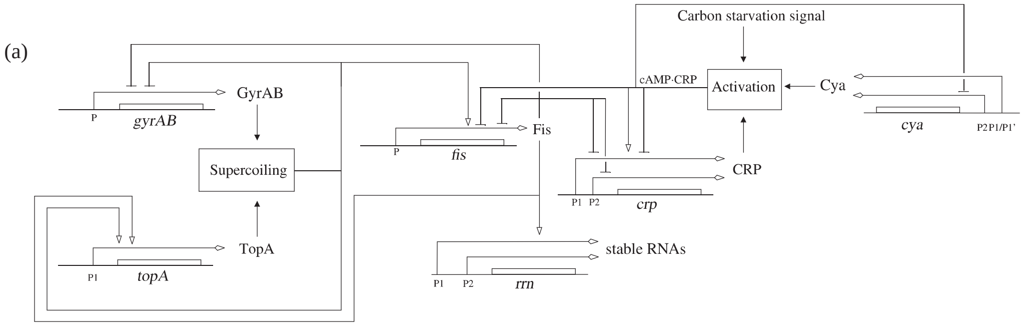
In order to test whether a discrete transition system satisfies a given temporal-logic formula, highly efficient algorithms have been developed and implemented in a range of model checkers. In addition to a yes/no answer, these tools return a diagnostic, either a witness or a counterexample, depending on whether the property holds or not. The diagnostic often provides valuable information for understanding why the property is satisfied or not.

In order to combine our qualitative simulator with model-checking tools, we have integrated export functionalities in the new version of GNA, allowing the user to generate text files describing the qualitative transition system in the format accepted by two widely used model checkers, NuSMV (Cimatti *et al*., 2002) and Evaluator, a component of the CADP toolbox (Mateescu and Sighireanu, 2003). NuSMV is an efficient, state-of-the-art model checker for CTL, whereas Evaluator is an on-the-fly model checker for the alternation-free μ-calculus, a temporal logic based on regular expressions. The text files generated by GNA can be imported in the model checkers, after which the verification of the properties of interest continues in the environment of the latter tools.

In this paper, we focus on the relation between GNA and NuSMV. Given a description of the Kripke structure, an initial state and a CTL formula, it is possible to check whether the

[2] For the formal syntax and semantics of CTL, see Clarke *et al*. (1999).

(a)



(b)

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

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

Fig. 4. (**a**) Network of key genes, proteins and regulatory interactions involved in the nutritional stress network in *E.coli*. The contents of the boxes labeled ‘Activation’ and ‘Supercoiling’ are detailed in Ropers *et al*. (2004). (**b**) PL differential equation and parameter inequality constraints for the topoisomerase TopA.

qualitative transition system in Figure 2 satisfies the property described by the formula. Provided that $D^{1.1}$ is the initial state, property (5) holds, and the path ($D^{1.1}$, $D^{2.2}$, $D^{3.2}$, $D^{4.2}$, $D^{4.1}$), shown in Figure 2c, is returned as a witness. Also, NuSMV shows that neither of the properties (6) and (7) hold.

Suppose that an experimentally-observed behavior does not correspond to any path in the state transition graph. Does this imply that the model must be rejected? Since the qualitative simulation method produces a conservative approximation of the dynamics of the original PL system (Section 2), one can be sure that a path corresponding to the experimentally-observed behavior must be present in the state transition graph, unless the model is invalid. As a consequence, the model can be safely rejected in the above case. On the other hand, if a path in the state transition graph corresponds to an experimentally-observed behavior, then the model is not necessarily corroborated by the observation, because the path may be a spurious behavior.

4 ANALYSIS OF NUTRITIONAL STRESS RESPONSE IN *E.COLI*

4.1 Model of nutritional stress response

In case of nutritional stress, an *E.coli* population abandons exponential growth and enters a non-growth state called *stationary phase*. This growth-phase transition is accompanied by numerous physiological changes in the bacteria, concerning among other things the morphology and the metabolism of the cells, as well as gene expression (Huisman *et al*., 1996). At the molecular level, the transition from exponential phase to stationary phase is controlled by a complex genetic regulatory network integrating various environmental signals.

Understanding the molecular basis of this essential developmental decision has been the focus of extensive studies for decades (Hengge-Aronis, 2000). However, notwithstanding the enormous amount of information accumulated on the genes, proteins and other molecules known to be involved in the stress adaptation process, there 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 methods for the dynamical modeling of genetic regulatory networks.

The above circumstances have motivated the qualitative analysis of the nutritional stress response network in *E.coli* by means of the method presented in this paper (Ropers *et al*., 2004). On the basis of literature data, we have decided to focus, as a first step, on a network of six genes that are believed to play a key role in the response of the cell to carbon starvation (Figure 4). The network includes genes involved in the transduction of the carbon starvation signal (the global regulator *crp* and the adenylate cyclase *cya*), metabolism (the global regulator *fis*), cellular growth (the *rrn* genes coding for stable RNAs) and DNA supercoiling, an important modulator of gene expression (the topoisomerase *topA* and the gyrase *gyrAB*).

Based on data in the experimental literature, a PL model of seven variables has been constructed, one protein concentration variable for each of the six genes and one input variable representing the presence or absence of the carbon starvation signal (Ropers *et al*., 2004). Seven differential equations, one for each variable, and forty inequality constraints describe the dynamics of the system. As an illustration, the differential equation and the parameter inequality constraints for the state variable x_{topA} are given in Figure 4b. For instance, the constraints $0 < \kappa^1_{topA}/\gamma_{topA} < \theta^1_{topA}$ express that without stimulation of the *topA* promoter, the TopA

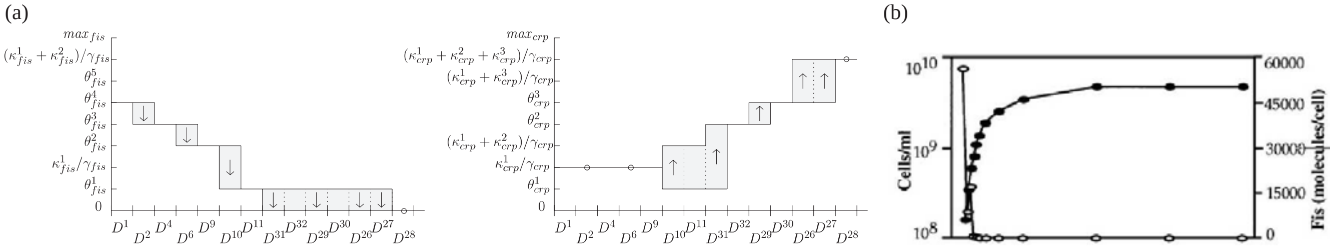


Fig. 5. Temporal evolution of the concentration of the proteins in the nutritional stress response network during the transition from exponential to stationary phase. (**a**) Predictions for Fis and CRP in a path in the state transition graph generated by qualitative simulation. (**b**) Observation for Fis (open circles) during the growth-phase transition, as indicated by cell density (closed circles) (Ali Azam *et al*., 1999).

concentration decreases towards a background level, below the threshold θ^1_{topA}.

Using the new version of the computer tool GNA, described in the previous sections, we have simulated two phenomena, namely the transition from exponential to stationary phase, and the reentry into exponential phase after a nutrient upshift. In order to validate the model, the simulation results have been compared with the available experimental data, using the export functionalities of GNA and the model checker NuSMV.

4.2 Validation of nutritional stress response model

In the absence of the carbon starvation signal, the system reaches a single qualitative equilibrium state that corresponds to the physiological conditions found in exponentially-growing *E.coli* cells. Starting from this equilibrium state, we perturb the system by switching on the carbon starvation signal and simulate the transition from exponential to stationary phase. This gives rise to a state transition graph of 66 states (27 of which are persistent), computed in less than one second on a PC (800 MHz, 256 MB). The graph contains a single equilibrium state corresponding to stationary-phase conditions. Figure 5 represents the temporal evolution of two of the protein concentrations in a path in the state transition graph. It shows that the concentration of Fis monotonically decreases to 0 and that of CRP monotonically increases to $(\kappa^1_{crp}+\kappa^2_{crp}+\kappa^3_{crp})/\gamma_{crp}$.

Are the predictions obtained from the model verified by the experimental data? Figure 5b shows the measured evolution of the Fis concentration (Ali Azam *et al*., 1999). Towards the end of the exponential phase, the concentration of Fis decreases and then becomes steady in stationary phase, which is characterized by a low concentration of stable RNAs x_{rrn}, that is, a concentration below the threshold θ_{rrn}. This observation can be translated into the following CTL formula:

$$\begin{aligned}&\mathbf{EF}(dsign\_x_{fis}=-1\wedge\\&\qquad\mathbf{EF}(dsign\_x_{fis}=0\wedge value\_x_{rrn}<\theta_{rrn})).\qquad(8)\end{aligned}$$

The qualitative transition system has been exported to the model checker, in order to verify the property. Verification takes a fraction of a second to complete and shows that the observed temporal evolution of the Fis concentration is reproduced by the model, i.e. there exists a path in the state transition graph satisfying the property (8).

Figure 5b suggests that we could be even more precise in our temporal-logic formulation of the experimental data. Not only $dsign\_x_{fis}=0$ in stationary phase, but in addition it would seem that $value\_x_{fis}=0$. However, since the precision of the measurements is limited, there may remain some small amount of Fis in the cell in stationary phase. The description $value\_x_{fis}=0$ is therefore too strong and might falsely rule out the model. Also, in this and similar examples, we use the temporal operator **F** instead of **G**, which would allow us to express that a property holds all of the time. The use of **G** is compromised by the fact that the usually low sampling frequency may cause us to miss phenomena predicted by simulation (e.g. a transient increase in a protein concentration) and thus, falsely rule out the model.

It would be interesting to put the predictions of the nutritional stress response model to more severe experimental tests. Unfortunately, time-series measurements of the evolution of the concentration of the other proteins in the network in Figure 4 during the transition from exponential to stationary phase are currently not available. However, even from the weak data that are available today, some interesting conclusions for model validation can be drawn. For instance, from the data in Balke and Gralla (1987) it can be inferred that the level of DNA supercoiling decreases during and after the transition to stationary phase. Since the level of DNA supercoiling is determined by the ratio of the concentration of GyrAB (which introduces supercoils into the DNA molecule) and the concentration of TopA (which removes supercoils from the DNA molecule) (Drlica, 1990), we require the following property to be satisfied by our model:

$$\begin{aligned}&\mathbf{EF}((dsign\_x_{gyrAB}=-1\vee dsign\_x_{topA}=1)\\&\qquad\wedge\ value\_x_{rrn}<\theta_{rrn}).\qquad(9)\end{aligned}$$

That is, during stationary phase, the concentration of GyrAB must decrease or the concentration of TopA must increase. Interestingly, the model does not satisfy the property (9), as revealed by model checking: in all paths in the state transition graph, the TopA concentration remains constant, while the GyrAB concentration increases! The inconsistency between the model and the observed level of DNA supercoiling indicates a flaw in the model. It demonstrates that our picture of the nutritional stress response is incomplete, in the sense that the network of Figure 4 may need to be extended with interactions not yet identified or with regulators not yet considered. In Ropers *et al*. (2004) we propose experiments and model extensions to further investigate these possibilities.

In addition to simulating the transition from exponential to stationary phase, we have also studied the reentry into exponential phase after a nutrient upshift, i.e. when cells in stationary phase have been put into fresh medium. Using the same model as above, but starting the simulation from the qualitative state characterizing stationary-phase conditions and with the carbon starvation signal switched off, qualitative simulation results in a state transition graph of 1143 states (202 of which are persistent), generated in 1.7 s. The graph is more complex than that generated for the transition from exponential to stationary phase, in the sense that it contains several cyclic paths. From all states in the graph, one of these cyclic paths can be reached, which we have shown to be attractive. To be more precise, the qualitative transition system satisfies the property

$$\mathbf{AG}(\mathit{statesInCycle} \rightarrow \mathbf{AG}\mathit{statesInCycle}), \qquad (10)$$

where the predicate *statesInCycle* is satisfied by all and only states in the cyclic path. That is, if the system has reached this path, it always remains in the path (testing this property takes NuSMV 9.1 s). Further mathematical analysis has revealed that the cyclic path arises from solutions spiraling inwards to an equilibrium point (Ropers *et al*., 2004). In other words, during the reentry into stationary phase, the concentrations of some of the proteins oscillate towards a new equilibrium level. This is a surprising result, which has not been subject to investigation so far. We are currently carrying out experiments in our laboratory to measure the temporal evolution of the protein concentrations in the nutritional stress response network, directly after a nutrient upshift, in order to verify this prediction and continue the validation of our model.

5 DISCUSSION

We have presented an approach for the validation of models of genetic regulatory networks, which combines a method for qualitative modeling and simulation with techniques for model checking. The qualitative modeling and simulation method, exploiting favorable mathematical properties of a class of coarse-grained models of genetic regulations, is a refinement of our previous work (de Jong *et al*., 2003). The method yields predictions on the derivative sign patterns of the concentration variables that are particularly well adapted to the currently available experimental methods. The methodological novelty of this paper is that we use model-checking techniques to deal with the problem that the state transition graphs generated by qualitative simulation may become prohibitively large for biologically-interesting networks. They permit observed dynamical properties of the system to be reliably and efficiently verified. Moreover, due the fact that the state transition graphs are conservative approximations of the dynamics of the underlying PL models, the latter are guaranteed not to be ruled out falsely. The model-validation approach is supported by a new version of the computer tool GNA.

The applicability of our model-validation approach has been illustrated by the analysis of the complex regulatory network underlying the nutritional stress response of *E.coli*. We have constructed a model of a part of this network, consisting of key proteins and their interactions involved in the carbon starvation response, and validated this model by the available experimental data in the literature. Although most predictions on the entry into stationary phase are consistent with the observations, in one case they contradict the experimental data, i.e. the observed decrease of the DNA supercoiling level, and necessitate revisions of the model. In addition, we have used model checking to further analyze the surprising prediction of the model that some of the protein concentrations oscillate after a nutrient upshift. This involves verifications that would be difficult to achieve by visual inspection.

Several applications of model checking and other formal verification techniques for the analysis and validation of biochemical network models have been proposed recently. Most approaches apply to discrete models, such as Petri nets (Koch *et al*., 2005), process algebras (Regev *et al*., 2001), concurrent transition systems (Chabrier-Rivier *et al*., 2004), rewriting logic (Eker *et al*., 2002) and Boolean networks and their generalizations (Bernot *et al*., 2004). For instance, in Bernot *et al*. (2004), a logical modeling approach is used in combination with CTL model checking to analyze models of mucus production in *Pseudomonas aeruginosa*, while the validation of a Petri net model of the sucrose breakdown pathway is investigated in Koch *et al*. (2005). The work presented in this paper shows that model checking can also be used for more conventional continuous models, like differential equation models. However, this requires a preliminary discretization of the dynamics of the system using abstractions. Several other approaches taking this direction can be mentioned (Antoniotti *et al*., 2004; Ghosh *et al*., 2003; Shults and Kuipers, 1997), based on qualitative differential equations (Shults and Kuipers, 1997) or hybrid automata (Antoniotti *et al*., 2004; Ghosh *et al*., 2003). However, contrary to our approach, these methods either do not result in a conservative approximation of the dynamics of the underlying continuous models (Antoniotti *et al*., 2004) or they are based on general purpose analysis techniques (Ghosh *et al*., 2003; Shults and Kuipers, 1997). The conservative approximation that we obtain is critical for preventing that models are unnecessarily rejected. The particular mathematical form of the PL models allows simple, tailor-made algorithms to be used, which promote the upscalability of our approach to large and complex networks, but at the same time limits its generality.

The model-validation approach of this paper has been illustrated in the context of CTL model checking. While CTL allows a variety of biologically meaningful properties to be expressed, some properties fall outside its scope. For instance, in Section 4.2 we would have liked to express the occurrence of oscillations in some of the protein concentrations after a nutrient upshift. That is, we would have liked to state that there exists a path in the qualitative transition system, such that from a state satisfying p it is always possible to reach a state satisfying $\neg p$, and from a state satisfying $\neg p$, it is always possible to reach a state satisfying p, where p might express that the concentration of some protein is above a threshold and $\neg p$ that it is below this threshold. The formula $\mathbf{EG}(p \rightarrow \mathbf{F}\neg p \wedge \neg p \rightarrow \mathbf{F}p)$ expresses this property, but unfortunately it is not a CTL formula (because **F** is not paired with a path quantifier) and it does not admit any CTL equivalent (Clarke and Draghicescu, 1988). However, the above property can be expressed in the μ-calculus and evaluated using XTL, a component of the CADP toolbox (Mateescu and Garavel, 1998). The capability of GNA to generate export files for different model checkers, allows one to take advantage from the specific strengths of each of them.

A problem encountered in the validation of our model is that time-series measurements of the concentrations of the proteins in the model are currently rare and usually have a low sampling frequency. In addition, the measurements for different proteins are difficult to combine, because they have been carried out under different conditions (using different strains, different culture media, etc.). This has the practical consequence that many interesting predictions obtained through qualitative simulation cannot currently be tested. In order to validate the model more rigorously, we are currently working on fine-grained measurements of gene expression in wild-type and mutant strains during growth-phase transitions. More generally, as systems biology takes hold, we expect such model-driven experiments to become more prominent.

REFERENCES

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

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

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

Batt,G. Ropers,D., de Jong,H., Geiselmann,J., Page,M. and Schneider,D. (2005) Qualitative analysis and verification of hybrid models of genetic regulatory networks. In Morari,M. and Thiele,L. (eds), *HSCC' 05*, Lecture Notes in Computer Science Vol. 3414, Springer, Berlin, pp. 134–150.

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

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

Cimatti,A., Clarke,E.M., Giunchiglia,E., Giunchiglia,F., Pistore,M., Roveri,M., Sebastiani,R. and Tacchella,A. (2002) NuSMV2: an opensource tool for symbolic model checking. In Brinksma,E. and Larsen,K.G. (eds), *CAV'02*, Lecture Notes in Computer Science, Vol. 2404, Springer, Berlin, pp. 359–364.

Clarke,E.M. and Draghicescu,I.A. (1988) Expressibility results for linear-time and branching-time logics. In de Bakker,J.W., de Roever,W.P. and Rozenberg,G. (eds), *REX Workshop*, Lecture Notes in Computer Science Vol. 354, Springer, Berlin, pp. 428–437.

Clarke,E.M., Grumberg,O. and Peled,D.A. (1999) *Model Checking*. MIT Press, Cambridge, MA.

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

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

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

Drlica,K. (1990) Bacterial topoisomerases and the control of DNA supercoiling. *Trends Genet.*, **6**, 433–437.

Eker,S., Knapp,M., Laderoute,K., Lincoln,P., Meseguer,J. and Sönmez,M.K. (2002) Pathway logic: symbolic analysis of biological signaling. In Altman,R.B., Dunker,A.K., Hunter,L., Jung,T., and Klein,T.C. (eds), *PSB'02*, World Scientific Publishing, Singapore, pp. 400–412.

Ghosh,R., Tiwari,A. and Tomlin,C.J. (2003) Automated symbolic reachability analysis, with application to Delta-Notch signaling automata. In Maler,O. and Pnueli,A. (eds), *HSCC' 03*, Lecture Notes in Computer Science Vol. 2623, Springer, Berlin, pp. 233–248.

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

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

Hengge-Aronis,R. (2000) The general stress response in *E.coli*. In Storz,G. and Hengge-Aronis,R. (eds), *Bacterial Stress Responses*. ASM Press, Washington, DC, pp. 161–177.

Huisman,G.W., Siegele,D.A., Zambrano,M.M. and Kolter,R. (1996) Morphological and physiological changes during stationary phase. In Neidhardt,F.C., Curtiss III,R., Ingraham,J.L., Lin,E.C.C., Low,K.B., Magasanik,B., Reznikoff,W.S., Riley,M., Schaechter,M. and Umbarger,H.E. (eds), *Escherichia coli and Salmonella: Cellular and Molecular Biology*. ASM Press, Washington, DC, pp. 1672–1682.

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

Mateescu,R. and Sighireanu,M. (2003) Efficient on-the-fly model-checking for regular alternation-free mu-calculus. *Sci. Comput. Program.*, **46**, 255–281.

Mateescu,R. and Garavel,H. (1998) XTL: a meta-language and tool for temporal logic model-checking. In Margaria,T. and Steffen,B. (eds) STTT'98. Brics, Aalborg, pp. 33–42.

Regev,A., Silverman,W. and Shapiro,E. (2001) Representation and simulation of biochemical processes using the pi-calculus process algebra. In Altman,R.B., Dunker,A.K., Hunter,L. and Klein,T.E. (eds), *PSB'01*. World Scientific Publishing, Singapore, pp. 459–470.

Ropers,D., de Jong,H., Page,M., Schneider,D. and Geiselmann,J. (2004) Qualitative simulation of nutritional stress response in *Escherichia coli. Technical Report INRIA RR-5412*.

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