HAL
open science

Efficient parameter search for qualitative models of regulatory networks using symbolic model checking

Grégory Batt, Michel Page, Irene Cantone, Gregor Goessler, Pedro T. Monteiro, Hidde de Jong

To cite this version:

Grégory Batt, Michel Page, Irene Cantone, Gregor Goessler, Pedro T. Monteiro, et al.. Efficient parameter search for qualitative models of regulatory networks using symbolic model checking. [Research Report] RR-7284, INRIA. 2010. inria-00482569

HAL Id: inria-00482569

https://inria.hal.science/inria-00482569v1

Submitted on 10 May 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Efficient Parameter Search for Qualitative Models of Regulatory Networks using Symbolic Model Checking

Gregory Batt — Michel Page — Irene Cantone — Gregor Goessler — Pedro Monteiro — Hidde de Jong

N° 7284

April 2010

Computational Biology and Bioinformatics

Rapport de recherche

ISSN 0249-6399 ISRN INRIA/RR--7284--FR+ENG

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche PARIS - ROCQUENCOURT

Efficient Parameter Search for Qualitative Models of Regulatory Networks using Symbolic Model Checking*

Gregory Batt[†] , Michel Page[‡] , Irene Cantone[§] , Gregor Goessler[¶] , Pedro Monteiro[‖] , Hidde de Jong[**]

Theme : Computational Biology and Bioinformatics
Computational Sciences for Biology, Medicine and the Environment
Équipes-Projets Contraintes, Pop Art and Ibis

Rapport de recherche n° 7284 — April 2010 — 23 pages

Abstract: Investigating the relation between the structure and behavior of complex biological networks often involves posing the following two questions: Is a hypothesized structure of a regulatory network consistent with the observed behavior? And can a proposed structure generate a desired behavior? Answering these questions presupposes that we are able to test the compatibility of network structure and behavior.

We cast these questions into a parameter search problem for qualitative models of regulatory networks, in particular piecewise-affine differential equation models. We develop a method based on symbolic model checking that avoids enumerating all possible parametrizations, and show that this method performs well on real biological problems, using the IRMA synthetic network and benchmark experimental data sets. We test the consistency between the IRMA network structure and the time-series data, and search for parameter modifications that would improve the robustness of the external control of the system behavior.

GNA and the IRMA model are available at http://www-helix.inrialpes.fr/gna

Key-words: systems and synthetic biology, qualitative models of gene networks, symbolic model checking, model validation and system design

* Contact: gregory.batt@inria.fr
† INRIA Paris - Rocquencourt, Le Chesnay, France
‡ INRIA Grenoble - Rhône-Alpes, Montbonnot, France and Université Pierre Mendès France, Grenoble, France
§ Clinical Sciences Center, Imperial College, London, United Kingdom
¶ INRIA Grenoble - Rhône-Alpes, Montbonnot, France
‖ INRIA Grenoble - Rhône-Alpes, Montbonnot, France and INESC/Instituto Superior Técnico, Lisbon, Portugal
** INRIA Grenoble - Rhône-Alpes, Montbonnot, France

Centre de recherche INRIA Paris – Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 Le Chesnay Cedex
Téléphone : +33 1 39 63 55 11 — Télécopie : +33 1 39 63 53 30

Paramétrisation efficace de modèles qualitatifs de réseaux de régulation à l'aide de model checking symbolique

Résumé : L'étude du lien entre structure et fonctionnement de réseaux biologiques complexes revient souvent à poser l'une des deux questions suivantes: Est ce qu'une structure hypothétique est cohérente avec des comportements observés? Est-ce qu'un comportement désiré peut être obtenu avec une structure proposée? Nous formulons ce problème en un problème de recherche de paramè-tres pour des modèles qualitatifs de réseaux de régulations.

Nous développons une méthode basée sur des techniques de model checking symbolique qui évite de devoir énumérer toutes les paramétrisations possibles, et nous montrons que cette méthode est bien adaptée aux problèmes biologiques réels à l'aide du réseau synthétique IRMA et des données expérimentales de référence associées. Nous testons la cohérence entre la structure du réseau IRMA et les données de séries temporelles, et nous cherchons des modifications des paramètres qui rendent le système plus robuste à un contrôle externe par addition de galactose.

GNA et le modèle IRMA sont disponibles en ligne à l'adresse http://www-helix.inrialpes.fr/gna

Mots-clés : biologie systémique et synthétique, modèles qualitatifs de réseaux de gènes, model checking symbolique, validation de modèles et conception de réseaux

1 Introduction

A central problem in the analysis of biological regulatory networks concerns the relation between their structure and dynamics. This problem can be narrowed down to the following two questions: (a) Is a hypothesized structure of the network consistent with the observed behavior? (b) Can a proposed structure generate a desired behavior?

Qualitative models of regulatory networks, such as (synchronous or asynchronous) Boolean models and piecewise-affine differential equation (PADE) models, have been proven useful for answering the above questions. The models are coarse-grained, in the sense that they do not specify the biochemical mechanisms in detail. However, they include the logic of gene regulation and allow different expression levels of the genes to be distinguished. They are interesting in their own right, as a way to capture in a simple manner even complex dynamics induced by the structure of interactions, for example steady states and transient responses to external perturbations. They can also be used as a first step to orient the development of more fine-grained quantitative ODE models. Several applications of logical and PADE models have confirmed their interest for the study of large and complex regulatory networks [6, 13, 22, 26].

Qualitative models bring specific advantages over numerical models when studying the relation between structure and dynamics. In order to answer questions (a) and (b), one has to search the parameter space to check if for some parameter values the network can be consistent with the data or a desired control objective can be attained. In qualitative models the number of different parametrizations is finite and the number of possible values for each parameter is usually rather low. This makes parameter search easier to handle than in quantitative models, where exhaustive search of the continuous parameter space is in general not feasible. Moreover, much of the available data in biology is semi-quantitative rather than fully quantitative due to variability in the experimental conditions and biological material, imprecise and relative measurements, low sampling density, ... Qualitative models are more concerned with qualitative trends in the data rather than with precise quantitative values.

Nevertheless, the parametrization of qualitative models remains a complex problem. For large models, the state and parameter spaces are usually too large to test all combinations of parameter values using existing techniques. This makes it difficult to answer questions (a) and (b) for most networks of actual biological interest. The aim of this paper is to address this search problem for PADE models by treating it in the context of formal verification and symbolic model checking. More specifically, we formulate the dynamic properties in temporal logic and verify by means of a model checker if the network satisfies these properties [8, 14].

Our contributions are twofold. On the methodological side, we develop a method that in comparison with our previous work [3] makes it possible to analyze very efficiently models with a large state space, and even to analyze incompletely parametrized models without the need for exhaustive enumeration of all parametrizations. This is achieved by a symbolic encoding of the model structure, the constraints on parameter values (if available), and the transition rules describing the qualitative dynamics of the PADE models. We can thus take full advantage of symbolic model checkers for testing the consistency of the network structure with dynamic properties expressed in temporal logics. The

current version 8 of GNA [21] has been extended with export functionalities to generate the symbolic encoding of PADE models in the NuSMV language [7]. In comparison with related work [1, 4, 9, 15], our method applies to incompletely instead of fully parametrized models, provides more precise results, and the encoding is efficient without (strongly) simplifying the PADE dynamics.

On the application side, we show that the method performs well on real problems, by means of the IRMA synthetic network and benchmark experimental data sets [5]. More precisely, we are able to find parameter values for which the network satisfies temporal-logic properties describing observed expression profiles, both on the level of individual and averaged time-series. The method is selective in the sense that only a small part of the parameter space is found to be compatible with the observations. Analysis of these parameter values reveals that biologically-relevant constraints have been identified. Moreover, we make suggestions to improve the robustness of the external control of the IRMA behavior by proposing a rewiring of the network.

2 Qualitative model of IRMA network

2.1 IRMA network

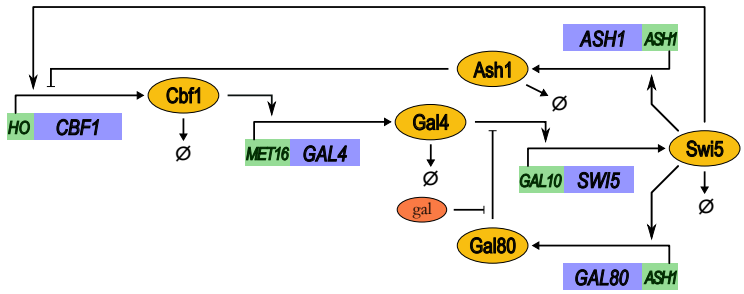
IRMA is a synthetic network constructed in yeast and proposed as a benchmark for modeling and identification approaches [5]. The network consists of five well-characterized genes that have been chosen so as to include different kinds of interactions, notably transcription regulation and protein-protein interactions. The endogenous copies of the genes were deleted, so as to reduce crosstalk of IRMA with the regulatory networks of the host cell. In order to further isolate the synthetic network from its cellular environment, the genes belong to distinct and non-redundant pathways. Moreover, they are non-essential, which means that they can be knocked out without affecting yeast viability.

The structure of the IRMA network is shown in Fig. 1(a). The expression of the *CBF1* gene is under the control of the *HO* promoter, which is positively regulated by Swi5 and negatively regulated by Ash1. *CBF1* encodes the transcription factor Cbf1 that activates expression of the *GAL4* gene from the *MET16* promoter. The *GAL10* promoter is activated by Gal4, but only in the absence of Gal80 or in the presence of galactose which releases the inhibition of Gal4 by Gal80. The *GAL10* promoter controls the expression of *SWI5*, whose product not only activates the above-mentioned *HO* promoter, but also the *ASH1* promoter controling the expression of both the *GAL80* and *ASH1* genes.

Notice that the network contains both positive and negative feedback loops. Negative feedback loops are a necessary condition for the occurrence of oscillations [27], while the addition of positive feedback loops is believed to increase the robustness of the oscillations [28]. This suggests that the network structure, for suitable parameter values, might be able to function as a synthetic oscillator.

2.2 Measurements of IRMA dynamics

The behavior of the network has been monitored in response to two different perturbations [5]: shifting cells from glucose to galactose medium (switch-on experiments), and from galactose to glucose medium (switch-off experiments).



(a)

$$\dot{x}_{Gal4} = \kappa^0_{Gal4} + \kappa_{Gal4}\, s^+(x_{Cbf1}, \theta_{Cbf1}) - \gamma_{Gal4}\, x_{Gal4}$$
$$\dot{x}_{Swi5} = \kappa^0_{Swi5} + \kappa_{Swi5}\, s^+(x_{Gal4}, \theta_{Gal4})\,(1 - s^+(x_{Gal80}, \theta_{Gal80})\, s^-(u_{gal}, \theta_{gal})) - \gamma_{Swi5}\, x_{Swi5}$$
$$\dot{x}_{Ash1} = \kappa^0_{Ash1} + \kappa_{Ash1}\, s^+(x_{Swi5}, \theta^a_{Swi5}) - \gamma_{Ash1}\, x_{Ash1}$$
$$\dot{x}_{Cbf1} = \kappa^1_{Cbf1}\, s^+(x_{Swi5}, \theta^c_{Swi5}) + \kappa^2_{Cbf1}\, s^+(x_{Swi5}, \theta^c_{Swi5})\, s^-(x_{Ash1}, \theta_{Ash1}) - \gamma_{Cbf1}\, x_{Cbf1}$$
$$\dot{x}_{Gal80} = \kappa^0_{Gal80} + \kappa_{Gal80}\, s^+(x_{Swi5}, \theta^g_{Swi5}) - \gamma_{Gal80}\, x_{Gal80}$$

(b)

Figure 1: Synthetic IRMA network in yeast. (a) Schematic representation of the network constructed in [5]. The green and blue boxes are promoter and genes, and the yellow and red ovals are proteins and metabolites. (b) PADE model of IRMA, with state variables x, protein synthesis constants κ, decay constants γ, and thresholds θ. The input variable u_{gal} refers to the presence of galactose ($\dot{u}_{gal} = 0$). The subscripts $_{Gal4}$, $_{Swi5}$, $_{Ash1}$, $_{Cbf1}$, $_{Gal80}$ refer to the proteins with the same name.

The terms 'switch-on' ('switch-off') refer to the activation (inhibition) of *SWI5* expression during growth on galactose (glucose). For these two perturbations, the temporal evolution of the expression of all the genes in the network was monitored by qRT-PCR with good time resolution: samples every 10 min (switch-off) or 20 min (switch-on) over more than 3 h.

Fig. 2(a) represents the expression of all genes, averaged over 5 (switch-on) or 4 (switch-off) independent experiments. In the switch-off experiments (galactose to glucose), the transcription of all genes is shut down. In the switch-on experiments, a seemingly oscillatory behavior is present with Swi5 peaks at 40 and 180 min, while Swi5, Cbf1, and Ash1 are expressed at moderate to high levels [5]. We remark that the confidence intervals are large (not shown), which means that the data are essentially semi-quantitative.

The analysis of the individual time-series reveals that in some cases the gene expression profiles are indeed similar, at least qualitatively, whereas in other cases notable differences are observed (for example, the oscillatory behavior is not present in all switch-on time-series; Fig. 2(c)). In the latter case, average expression levels may be a misleading representation of the network behavior.

2.3 PADE model of IRMA network

We built a qualitative model of the IRMA dynamics using PADE models of genetic regulatory networks. The PADE models, originally introduced in [16],

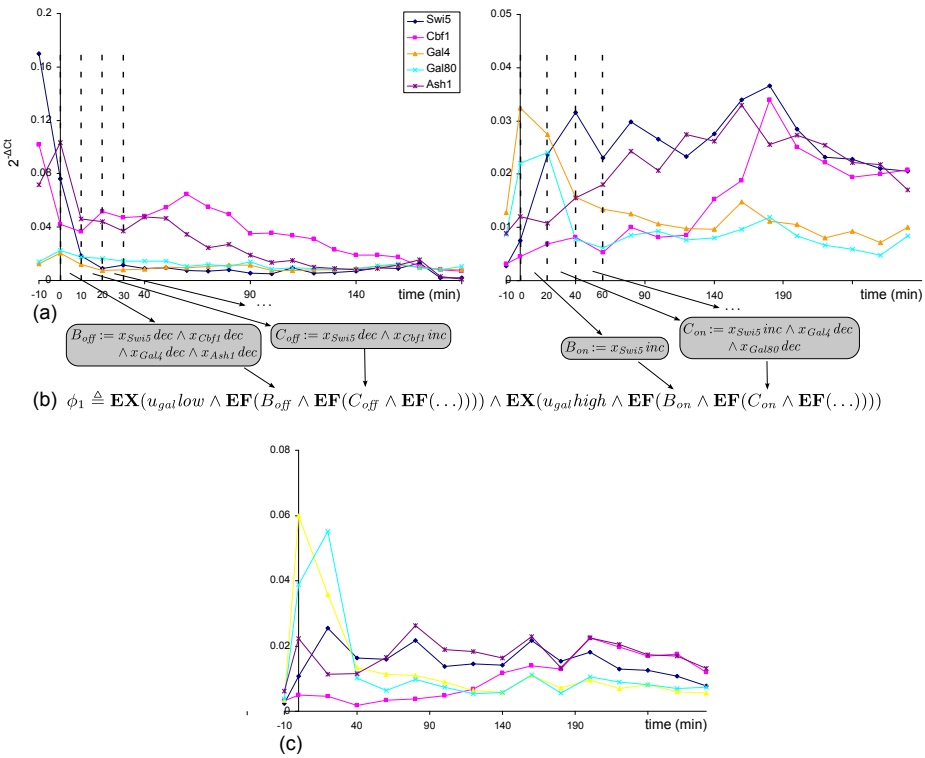


Figure 2: Dynamic behavior of the IRMA network in response to medium shift perturbations. (a) Temporal profiles of averaged gene expression measured with qRT-PCR during switch-off (left) and switch-on (right) experiments (data from [5]). (b) Temporal logic encoding of the switch-off and switch-on behaviors. The operator $\mathbf{EF}\,\phi$ expresses the possibility to reach a future state satisfying ϕ, whereas the operator $\mathbf{EX}\,\phi$ is used to require the existence of an initial state satisfying ϕ. u_{gal} *low* and u_{gal} *high* denote the absence and presence of galactose, respectively. See [8] for more details on the temporal logic CTL. (c) Temporal profile of gene expression in an individual switch-on experiment showing a switch-off-like behavior.

provide a coarse-grained picture of the network dynamics. They have the following general form:

$$\dot{x}_i = f_i(x) \triangleq \sum_{l \in L_i} \kappa_i^l \, b_i^l(x) - \gamma_i \, x_i, \;\; i \in [1, n] \tag{1}$$

where $x \in \Omega \subset \mathbb{R}^n_{\geq 0}$ represents a vector of n protein (or RNA) concentrations. The synthesis rate is composed of a sum of synthesis constants κ_i^l, each modulated by a regulation function $b_i^l(x) \in \{0, 1\}$. A regulation function is an algebraic expression of step functions $s^+(x_j, \theta_j)$ or $s^-(x_j, \theta_j)$ which formalizes the regulatory logic of gene expression. θ_j is a so-called threshold for the concentration x_j. The step function $s^+(x_j, \theta_j)$ evaluates to 1 if $x_j > \theta_j$, and to 0 if $x_j < \theta_j$, thus capturing the switch-like character of gene regulation ($s^-(x_j, \theta_j) = 1 - s^+(x_j, \theta_j)$). The degradation of a gene product is a first-order term, with a degradation constant γ_i that includes contributions of growth dilution and protein degradation. The models can be easily extended to account for proteolytic regulators, but we will omit this here as IRMA does not include such factors.

In the case of IRMA, we define five variables that correspond to the total protein concentrations of Cbf1, Gal4, Gal80, Ash1, and Swi5, as well as an input variable denoting the concentration of galactose. Notice that the measurements of the network dynamics concern mRNA and not protein levels. We assume that the variations in mRNA and protein levels are the same, even though this may not always be the case. A similar approximation is made in [5], where protein and mRNA levels are considered to be proportional.

The PADE model of the IRMA network is shown in Fig. 1(b). Consider for example the equation for the protein Gal4. κ^0_{Gal4} is its basal synthesis rate, and $\kappa^0_{Gal4} + \kappa_{Gal4}$ its maximal synthesis rate when the *GAL4* activator Cbf1 is present (*i.e.*, $x_{Cbf1} > \theta_{Cbf1}$). Swi5 is regulated in a more complex way. The expression of its gene is activated by Gal4, but only when Gal80 is absent or galactose present (which prevents Gal4 inactivation by Gal80), that is, only when not both Gal80 is present and galactose absent. The step-function expression in Fig. 1(b) mathematically describes this condition. We remark that for the regulation of *CBF1*, we take into account that Ash1 can override the effect of Swi5, that is, inhibition dominates activation. Moreover, Swi5 is assumed to have three different thresholds, for the regulation of *CBF1*, *GAL80*, and *ASH1*.

The PADE model is a direct translation of the IRMA network into a simple mathematical format. The model resembles the ODE model in [5], but notably approximates the Hill-type kinetic rate laws by step functions. It thus makes the implicit assumption that important qualitative dynamical properties of the network are intimately connected with the network structure and the regulatory logic, independently from the details of the kinetic mechanisms and precise parameter values. Several studies have shown this assumption to be valid in a number of model systems [6, 11, 10], although care should be exercised in deciding exactly when modeling approximations are valid [25].

To investigate for the possible existence of unknown interactions between the synthetic network and the host, we would like to test given the PADE model above whether the network structure and the regulatory logic can account for the qualitative trends in the gene expression data observed in [5]. Because in

some experiments it has been observed that the addition of galactose does not always lead to an activation of the IRMA genes, we also search for parameter modifications that renders the network response to an addition of galactose more robust.

3 Search of parameter space using symbolic model checking

3.1 Qualitative analysis of PADE models

The advantage of PADE models is that the qualitative dynamics of high-dimensional systems are relatively easy to analyze, using only the total order on parameter values rather than exact numerical values [2, 12]. The main difficulty lies in treating the discontinuities in the right-hand side of the differential equations, at the threshold values of the step functions. Following [18], the use of differential inclusions based on Filippov solutions has been proposed in [2] and implemented in the computer tool GNA [3]. Here, we recast this analysis in a form that underlies the symbolic encoding of the dynamics below.

The key to our reformulation of the qualitative analysis of the PADE dynamics is the extension of step functions s^+ to interval-valued functions S^+ , where

$$S^+(x_j, \theta_j) = \begin{cases} [0,0] \text{ if } x_j < \theta_j, \\ [0,1] \text{ if } x_j = \theta_j, \\ [1,1] \text{ if } x_j > \theta_j \end{cases} \tag{2}$$

That is, because the step functions are not defined at their thresholds, we conservatively assume that they can take any value between 0 and 1 (see [6] for a similar idea). When replacing the step functions by their extensions, the regulation functions $b_i^l(x)$ become interval-valued functions $B_i^l : \mathbb{R}^n_{\geq 0} \to \{[0,0],[0,1],[1,1]\}$, and Eq. (1) generalizes to the following differential inclusion:

$$\dot{x}_i \in F_i(x) \triangleq \sum_{l \in L_i} \kappa_i^l \, B_i^l(x) - \gamma_i \, x_i, \;\; i \in [1,n] \tag{3}$$

As shown in Section B, for most models the solutions of this differential inclusion are the same as the solutions of the differential inclusions defined in [2].

The starting-point for our qualitative analysis is the introduction of a rectangular partition $\mathcal{D}$ of the state space Ω. This partition is induced by the union of the two sets Θ_i and Λ_i, $i \in [1,n]$, where $\Theta_i = \{\theta_i^j \mid j \in J_i\}$ and $\Lambda_i = \{\sum_{l \in B} \kappa_i^l / \gamma_i \mid B \subseteq L_i\}$. That is, the partition is a rectangular grid defined by the threshold parameters θ_i^j and the so-called focal parameters $\sum_{l \in B} \kappa_i^l / \gamma_i$. The focal parameters are steady-state concentrations towards which the PADE system locally converges in a monotonic way [16]. For the variable x_{Gal4}, we have $\Theta_{Gal4} = \{\theta_{Gal4}\}$ and $\Lambda_{Gal4} = \{0, \kappa^0_{Gal4}/\gamma_{Gal4}, (\kappa^0_{Gal4} + \kappa_{Gal4})/\gamma_{Gal4}\}$.

Interestingly, the partition has the property that in each domain $D \in \mathcal{D}$, the protein production rates are identical: for all $x, y \in D$, it holds that $B_i^l(x) = B_i^l(y) \triangleq B_i^l(D)$. As a consequence, the derivatives of the concentration variables have a unique sign pattern: for all $x, y \in D$, it holds that

$sign(F_i(x)) = sign(F_i(y)) \subseteq \{-1, 0, 1\}$ [2]. Notice that this property is not obtained for less fine-grained partitions used in related work [1, 4, 6, 9, 13, 15, 17]. It will be seen to be critical for the search of parametrized models of IRMA that satisfy the time-series data.

The above considerations motivate a discrete abstraction, resulting in a state transition graph. In this graph, the states are the domains $D \in \mathcal{D}$, and there is a transition from a domain D to another domain D', if there exists a solution of the differential inclusion in Eq. 3 that starts in D and reaches D', without leaving $D \cup D'$.

$$\begin{aligned} D \to D' \text{ iff } & \exists \xi \text{ solution of (3)}, \exists \tau \in \mathbb{R}_{\neq 0} \cup \{\infty\} \\ & \text{such that } \xi(0) \in D, \xi(\tau) \in D', \text{ and } \forall t \in [0, \tau], \xi(t) \in D \cup D' \end{aligned}$$

The state transition graph defines the qualitative dynamics of the system, in the sense that the states give a qualitative description of the state of the system (derivative sign patterns, threshold and focal parameters bounding the domain), while paths in the state transition graph describe how this state evolves over time (changes in derivative patterns, changes in bounds of domain) [2].

We reformulate here the transition rules using the interval extensions of the regulation functions. The existence of a transition depends on the sign of F at the boundary between the two domains. To capture this notion, we introduce an interval-valued function $F_i : \mathcal{D} \times \mathcal{D} \to 2^{\mathbb{R}}$, where $F_i(D, D') = \sum_{l \in L_i} \kappa_i^l \; B_i^l(D) - \gamma_i \, D_i'$, for $D, D' \in \mathcal{D}$. $F_i(D, D')$ represents the flow in D infinitely close to D'. In order to evaluate $F_i(D, D')$, we use interval arithmetic [23]. For instance, in a domain in which $x_{Swi5} > \theta^c_{Swi5}$ and $x_{Ash1} = \theta_{Ash1}$, we have $S^+(x_{Swi5}, \theta^c_{Swi5}) = [1, 1]$ and $S^-(x_{Ash1}, \theta_{Ash1}) = [0, 1]$, so that the differential inclusion for x_{Cbf1} becomes $[\kappa^1_{Cbf1}, \; \kappa^1_{Cbf1} + \kappa^2_{Cbf1}] - \gamma_{Cbf1} \cdot x_{Cbf1}$.

We distinguish three types of transitions, depending on whether the transition goes from a domain D to itself ($D = D'$, internal transition), from a domain D to another, higher-dimensional domain D' ($D \subseteq \partial D'$, dimension-increasing transition), or from a domain D to another, lower-dimensional domain D' ($D' \subseteq \partial D$, dimension-decreasing transition), where ∂D denotes the boundary of D in its supporting hyperplane. For dimension-increasing transition, we obtain the following rule:

Prop. 1 (Dimension-increasing transition)**.** Let $D, D' \in \mathcal{D}$ and $D \subseteq \partial D'$. $D \to D'$ is a dimension-increasing transition iff

1. $\forall i \in [1, n]$, such that D_i and D_i' coincide with a value in $\Theta_i \cup \Lambda_i$, it holds that $0 \in F_i(D', D)$, and
2. $\forall i \in [1, n]$, such that $D_i \neq D_i'$, it holds that $\exists \alpha > 0$ such that $\alpha \in F_i(D', D) \, (D_i' - D_i)$

The first condition guarantees that solutions can remain in domains located in threshold and focal planes, while the second condition expresses that the direction of the flow in the domains ($F_i(D', D)$) is consistent with their relative position ($D_i' - D_i$). The rules for other types of transitions and their proofs can be found in Section A.

It can be shown that exact parameter values are not needed for the analysis of the qualitative dynamics of a PADE model: it is sufficient to know the ordering

of the threshold and focal parameters [2]. This comes from the fact that the sign of F_i, and hence the transitions and the state transition graph, are invariant for regions of the parameter space defined by a particular total order on $\Theta_i \cup \Lambda_i$ [2]. We call such a total order a parametrization of the PADE model.

3.2 Search of parameter space: a model-checking approach

Verifying the compatibility of the network structure with an observed or desired behavioral property (Section 2.2) can be achieved by comparing the state transition graph with qualitative trends in the data. For large graphs like that obtained for IRMA (which has about 50000 states) this becomes quickly impossible to do by hand. This has motivated the use of model-checking tools (*e.g.*, [1, 3, 4, 14]). For PADE models, each state in the graph is described by atomic propositions whose truth-value are preserved under the discrete abstraction, such as the above-mentioned derivative sign patterns. The atomic propositions are used to formulate observed or desired properties in a temporal-logic formula ϕ and model checkers test if the state transition graph T satisfies the formula ($T \models \phi$).

Because the number of possible parametrizations and the size of state transition graphs rapidly grow with the number of genes, the naive approach consisting in enumerating all parametrizations of a PADE model, and for each of these generating the state transition graph and testing whether $T \models \phi$, is only feasible for the simplest networks. We therefore propose an alternative approach, based on the symbolic encoding of the above search problem, without explicitly generating the possible parametrizations of the PADE models and the corresponding state transition graphs. This enables one to exploit the capability of symbolic model checkers to efficiently manipulate implicit descriptions of the state and parameter space.

3.3 Symbolic encoding of PADE model and dynamics

In this section we summarize the main features of the encoding . We particularly focus on the discretization of the state space, which connects the symbolic encoding to the mathematical analysis of PADE models, and the use of the discretization for the computation of $F_i(D', D)$ in Prop. 1, which is essential for state transition computations.

The symbolic encoding is based on a discretization of the state space implied by the partition $\mathcal{D}$. We call $\mathcal{C}$ a discretization function that maps $D \in \mathcal{D}$ to a set of unique integer coordinates, and $\mathcal{C}(D) = \mathcal{C}(D_1) \times \ldots \times \mathcal{C}(D_n)$. Let m_i be the number of non-zero parameters in $\Theta_i \cup \Lambda_i$, $i \in [1, n]$. Then $\mathcal{C}(D_i) \in \{0, 1, \ldots, 2m_i + 1\}$, and more specifically, $\mathcal{C}(D_i) \in \{0, 2, \ldots, 2m_i\}$ if D_i coincides with a threshold or focal plane, and $\mathcal{C}(D_i) \in \{1, 3, \ldots, 2m_i + 1\}$ otherwise. More generally, $\mathcal{C}(S) = \{\mathcal{C}(D) \mid D \subseteq S\}$, for any set of domains S. Obviously, $\mathcal{C}$ can also be used for the discretization of parameter values. For example, in the case of the variable x_{Gal4}, we have one threshold and three focal parameters. Now, let D be a domain in the state space and D_{Gal4} its component in the x_{Gal4}-dimension. Given the following total order on the threshold and focal parameters, $0 < \kappa^0_{Gal4}/\gamma_{Gal4} < \theta_{Gal4} < (\kappa^0_{Gal4} + \kappa_{Gal4})/\gamma_{Gal4}$, we find $\mathcal{C}(0) = 0$ (by definition), $\mathcal{C}(\kappa^0_{Gal4}/\gamma_{Gal4}) = 2$, $\mathcal{C}(\theta_{Gal4}) = 4$, and $\mathcal{C}(\kappa^0_{Gal4} + \kappa_{Gal4})/\gamma_{Gal4}) = 6$.

The above discretization motivates the introduction of symbolic variables $\hat{D}_i$, $\hat{D}'_i$, $\hat{\theta}^j_i$, $\hat{\lambda}^j_i$ that encode $\mathcal{C}(D_i)$, $\mathcal{C}(D'_i)$, $\mathcal{C}(\theta^j_i)$, $\mathcal{C}(\lambda^j_i)$, respectively. The different conditions in Prop. 1 can be expressed in terms of this encoding. For instance, the sign of $D'_i - D_i$ simply becomes $\hat{D}'_i - \hat{D}_i$. The translation is less evident for the encoding of $F_i(D', D)$, the sign of which needs to be computed in the transition rules. Multiplication by $1/\gamma_i$ does not change the sign, but gives the more convenient expression

$$F_i(D, D')/\gamma_i = \sum_{l \in L_i} (\kappa^l_i/\gamma_i)\ B^l_i(D) - D'_i \tag{4}$$

Recall that the first term in the righthand side is simply an interval whose upper and lower bound are focal parameters, determined by the regulation functions $B^l_i(D)$. By redefining the step functions in terms of the symbolic variables:

$$S^+(D_j, \theta_j) = \begin{cases} [0,0] & \text{iff } \hat{D}_j < \hat{\theta}_j \\ [0,1] & \text{iff } \hat{D}_j = \hat{\theta}_j \\ [1,1] & \text{iff } \hat{D}_j > \hat{\theta}_j \end{cases} \tag{5}$$

each $B^l_i(D)$ can be simply computed by means of interval arithmetic. Evaluating the expression $\sum_{l \in L_i} (\kappa^l_i/\gamma_i)\ B^l_i(D)$ leads to an interval with focal parameters as bounds, and which can therefore be represented by $\hat{\lambda}^j_i$. From this interval we subtract $\hat{D}'_i$ to symbolically define $F_i(D, D')/\gamma_i$. The sign of the latter expression allows one to check the conditions of Prop. 1, and thus to derive the transitions in the state transition graph. The specification of transitions in a symbolic way is the main stumble block for the efficient encoding of the PADE dynamics, especially when D is located on a threshold plane. In our previous work [2], the computation of transitions required the enumeration of an exponential number of domains surrounding D [1]. The interval-based formulation proposed here avoids this inefficient approach and allows $F_i(D, D')/\gamma_i$ to be computed in one stroke.

The implementation in a model checker like NuSMV [7] is straightforward with the above encoding. In particular, we apply invariant constraints on the symbolic variables to exclude all valuations of $\hat{D}_i$, $\hat{D}'_i$, $\hat{\theta}^j_i$, $\hat{\lambda}^j_i$ that do not correspond to a valid transition from D to D' for a given parameterization. We apply three types of invariants. The first one constrains parameters to remain constant. The second one constrains D and D' to be neighbors in the state space (*e.g.*, $D \subseteq \partial D'$ for dimension-increasing transitions). The final invariant constrains the relative position of D and D' and the parameter order as stated in the transitions conditions. For comparison with experimental data, we also need to know the variations of concentrations of gene products in each state. Formally, it is defined as the derivative sign pattern, and simply corresponds to the sign of $F_i(D, D')$ as computed above.

The initial states of our symbolic structure correspond to each possible parametrization and transitions towards all states D. A CTL property ϕ holds for a symbolic structure if all initial states satisfy ϕ. Therefore, by testing whether $\neg\phi$ holds, we verify the absence of a parametrization satisfying ϕ. A counterexample to $\neg\phi$ thus returns a valid parametrization. The current version 8 of GNA [21] has been extended with export functionalities to generate the symbolic encoding of PADE models in the NuSMV language.

4 Validation: consistency of IRMA network with experimental data

Are the observations of the IRMA dynamics consistent with the network structure? At first sight this question may seem incongruous as one expects this to be the case by definition (each genetic construct was tested before integration in the yeast cell). However, in practice it is far from trivial, even if the design and construction have been carried out with great care, to avoid interactions between the synthetic network and the host.

4.1 Temporal-logic encoding of observations

To test the consistency between our PADE model of the IRMA network and the experimental data, we express that for each condition, switch-on and switch-off, there must exist an initial state of the system and a path starting from this state along which the gene expression changes correspond to the observed time-series data. For example, for the switch-off time-series we encode that there exists an initial state where in absence of galactose the expression of *SWI5*, *CBF1*, *GAL4* and *ASH1* decreases (in the interval $[0, 10]$ min), and from which a state can be reached where the expression of *SWI5* decreases and the expression of *CBF1* increases (in the interval $[10, 20]$ min), *etc.* The generation of this property, called ϕ_1, from the experimental data leads to the temporal-logic formula shown in Fig. 2(b). The property is automatically generated from the experimental data using a Matlab script.

To disregard small fluctuations due to biological and experimental noise, we considered that changes of magnitude less that $5 \cdot 10^{-3}$ units are not significant. This smooths out, for example, Gal4 expression levels in switch-off conditions after 40 min. In [5] it was demonstrated by glucose-to-glucose shift experiments that the mere resuspension of cells into fresh medium has a network-independent effect: the expression of *GAL80* and *GAL4* is strongly increased in the first 10 min after resuspension. Therefore, we did not incorporate in our specification the very first measurements (in the interval $[-10, 0]$) made just before shifting cells to a new medium.

The data presented in [5] for switch-on and switch-off conditions are the average of 5 and 4 individual experiments, respectively. As noticed in Section 2.2, expression profiles obtained in similar conditions may differ significantly. In the case of such heterogeneous behavior, properties capturing the average gene expression profile may be misleading. Consequently, asking for consistency between our model and the result of each individual experiment might be more appropriate. This leads us to define a second property ϕ_2 similar to ϕ_1 but requiring the existence of 9 paths in the graph, one for each of the observed behaviors in the 5 switch-on and 4 switch-off experiments. Although the information we extract from the experimental data is purely qualitative, only concerning trends in gene expression levels, the accumulation of these simple observations leads to fairly complex constraints. Property ϕ_2 involves nearly 160 constraints on derivative signs.

4.2 Testing consistency of network with observations

We use our symbolic encoding of the PADE dynamics and verify the existence of a valid parameter ordering. We do this by testing the negation of ϕ_1 or ϕ_2, such that a negative answer from the model checker proves the existence of at least one valid parametrization, as explained in Section 3.3. Moreover, the counterexample returned provides one such parametrization. By means of this approach, we can prove the existence of a parametrization satisfying the averaged time-series data (ϕ_1). The result was obtained in 49 s on a laptop (PC, 2.2 Ghz, 1 core, 2 Gb RAM). The state space contains nearly 50000 discrete states and the parameter space is discretized into nearly 5000 different parameter orderings. The counterexample of $\neg\phi_1$, obtained in 100 s, provides a valid parametrization (Table 1).

When analyzing the corresponding parametrization, the thresholds are mostly higher than the focal parameter for basal expression and lower than the focal parameter for upregulated expression, *e.g.*, $\kappa^0_{Ash1}/\gamma_{Ash1} < \theta_{Ash1} < (\kappa^0_{Ash1} + \kappa_{Ash1})/\gamma_{Ash1}$. This is not surprising as the focal parameters correspond to the lowest and highest possible expression levels. The threshold at which Ash1 controls *CBF1* expression is expected to lie between the two extremes. The only exception in the parameters found by the model checker is Gal80, for which it holds $(\kappa^0_{Gal80} + \kappa_{Gal80})/\gamma_{Gal80} < \theta_{Gal80}$. According to this constraint, Gal80 plays no role in the system, since it cannot exceed the threshold concentration above which it inhibits Swi5. This is interesting because it suggests that the switch-off behavior may occur even without any inhibition by Gal80, and consequently, in a galactose-independent manner.

The dynamic properties of the PADE model can be analyzed in more detail by means of GNA. This shows the existence of an asymptotically stable steady state corresponding to switch-off conditions, with low Swi5, Gal4, Cbf1, Ash1, and Gal80 concentrations. In addition, GNA finds strongly connected components (SCCs) consistent with the observed damped oscillations observed in galactose media. However, the attractors co-exist irrespectively of the presence or absence of galactose, revealing that galactose does not necessarily drive the system to a single attractor for this particular parametrization.

We also tested whether the above parametrization is consistent with time-series data from the individual experiments. In 3 s the model checker shows that it does not satisfy the more constraining property ϕ_2. However, we do find another parametrization for which ϕ_2 holds. In this case, all thresholds are situated between the basal and upregulated focal parameters (237 s, including counterexample generation).

4.3 Detailed analysis of valid parameter set

As stated above, our consistency tests only confirm that a parametrization exists for which the structure of the network is consistent with the observed behavior. However, it does not say if this is trivially the case (when most parametrizations are) or if the properties are selective (when most parametrization are not). To investigate this we exhaustively generated all possible parametrizations, and tested for each of them property ϕ_1 (averaged time-series) and ϕ_2 (individual time-series). Although the total number of parameter orderings (4860) is fairly large, the exhaustive analysis is still manageable for networks of this size.

	Symbolic state space and symbolic parameter space		Symbolic state space and explicit parameter space	
Property	Existence of parametrization	Parametrization*	Number of parametrizations	Parametrization*
ϕ_1: averaged time-series	Yes (49 s)	$\frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} < \theta^g_{Swi5} < \theta^c_{Swi5} < \theta^a_{Swi5} < \frac{\kappa^0_{Swi5}+\kappa_{Swi5}}{\gamma_{Swi5}} \wedge \frac{\kappa^0_{Gal80}}{\gamma_{Gal80}} < \frac{\kappa^0_{Gal80}+\kappa_{Gal80}}{\gamma_{Gal80}} < \theta_{Gal80}$	12 (925 s)	$\frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} < \theta^c_{Swi5} < \theta^a_{Swi5} < \frac{\kappa^0_{Swi5}+\kappa_{Swi5}}{\gamma_{Swi5}} \wedge$ $(\theta_{Gal80} < \frac{\kappa^0_{Gal80}}{\gamma_{Gal80}} \wedge \frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} < \theta^g_{Swi5} < \frac{\kappa^0_{Swi5}+\kappa_{Swi5}}{\gamma_{Swi5}}$ $\vee \frac{\kappa^0_{Gal80}}{\gamma_{Gal80}} < \theta_{Gal80} < \frac{\kappa^0_{Gal80}+\kappa_{Gal80}}{\gamma_{Gal80}} \wedge \frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} < \theta^g_{Swi5}$ $\vee \frac{\kappa^0_{Gal80}+\kappa_{Gal80}}{\gamma_{Gal80}} < \theta_{Gal80})$
ϕ_2: individual time-series	Yes (131 s)	$\frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} < \theta^c_{Swi5} < \theta^a_{Swi5} < \theta^g_{Swi5} < \frac{\kappa^0_{Swi5}+\kappa_{Swi5}}{\gamma_{Swi5}} \wedge \frac{\kappa^0_{Gal80}}{\gamma_{Gal80}} < \theta_{Gal80} < \frac{\kappa^0_{Gal80}+\kappa_{Gal80}}{\gamma_{Gal80}}$	4 (2021 s)	$\frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} < \theta^c_{Swi5} < (\theta^a_{Swi5}, \theta^g_{Swi5}) < \frac{\kappa^0_{Swi5}+\kappa_{Swi5}}{\gamma_{Swi5}}$ $\wedge (\frac{\kappa^0_{Gal80}}{\gamma_{Gal80}}, \theta_{Gal80}) < \frac{\kappa^0_{Gal80}+\kappa_{Gal80}}{\gamma_{Gal80}}$
ϕ_3: single attractor	Yes (126 s)	$\theta^c_{Swi5} < \frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} < \theta^g_{Swi5} < \theta^a_{Swi5} < \frac{\kappa^0_{Swi5}+\kappa_{Swi5}}{\gamma_{Swi5}} \wedge \theta_{Gal80} < \frac{\kappa^0_{Gal80}}{\gamma_{Gal80}} < \frac{\kappa^0_{Gal80}+\kappa_{Gal80}}{\gamma_{Gal80}}$	7 (1300 s)	$\theta^c_{Swi5} < \frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} < \theta^a_{Swi5} < \frac{\kappa^0_{Swi5}+\kappa_{Swi5}}{\gamma_{Swi5}}$ $\wedge \theta_{Gal80} < \frac{\kappa^0_{Gal80}+\kappa_{Gal80}}{\gamma_{Gal80}}$ $\wedge (\theta^g_{Swi5} < \frac{\kappa^0_{Swi5}}{\gamma_{Swi5}} \vee \theta_{Gal80} < \frac{\kappa^0_{Gal80}}{\gamma_{Gal80}})$

*All parametrizations additionally include $\kappa^1_{Cbf1}/\gamma_{Cbf1} < \theta_{Cbf1} < (\kappa^1_{Cbf1} + \kappa^2_{Cbf1})/\gamma_{Cbf1} \wedge \kappa^0_{Gal4}/\gamma_{Gal4} < \theta_{Gal4} < (\kappa^0_{Gal4} + \kappa_{Gal4})/\gamma_{Gal4} \wedge \kappa^0_{Ash1}/\gamma_{Ash1} < \theta_{Ash1} < (\kappa^0_{Ash1} + \kappa_{Ash1})/\gamma_{Ash1}$.

Table 1: Summary of parametrizations found by checking the consistency of the IRMA structure with the observed and desired behaviors, expressed as temporal-logic properties ϕ_1, ϕ_2, and ϕ_3. The table shows the parametrization returned when testing the truth-value of the property on the symbolically encoded PADE model and dynamics (left) and summarizes all parametrizations satisfying the properties (right).

Out of the 4860 completely parametrized PADE models, we found that only a surprisingly small subset is consistent with the observations. For the averaged time-series, only 12 parametrizations are consistent, while for the individual time-series this subset is further reduced to 4 parametrizations (Table 1). The properties extracted from the data are thus seen to be quite selective.

These results indicate that to be consistent with the experimental data, the activation threshold of *CBF1* by Swi5, θ^c_{Swi5}, must be smaller than the activation thresholds of *ASH1* and *GAl80* by Swi5, θ^a_{Swi5} and θ^g_{Swi5}. Interestingly, this result is corroborated by independent experimental studies. Fitting of experimental data on promoter activities to Hill functions showed that the activation threshold for the *ASH1* promoter, controlling *ASH1* and *GAL80* expression, is nearly twice as high as the one for the *HO* promoter controlling *CBF1* expression (Table S1 of [5]).

A second finding is that the dynamics of the system is consistent with the experimental data even if $\theta_{Gal80} < \kappa^0_{Gal80}/\gamma_{Gal80}$, that is when *GAL80* is constitutively expressed above its inhibition threshold. This indicates that an effective regulation of *GAL80* expression by Swi5 is of little importance for the functioning of the network. And indeed, it was found that *GAL80* is not much responsive to changes in Swi5 availability: Cantone *et al* observed that a 6-fold increase of *SWI5* expression leads to only a negligible (1.08-fold) increase in *GAL80* expression levels (Fig. 4A in [5]).

5 Re-engineering: improving external control by galactose

As stated above, it has been experimentally observed that the system response to an addition of galactose is not always identical. In one experiment at least, the addition of galactose does not significantly changes the system's behavior: a switch-off like response is observed in switch on conditions (Fig. 2(c)). To obtain a more robust external control of the system, we would like to ensure that the addition of galactose drives the system out of the low-Swi5 state.

5.1 Temporal-logic specification of design objective

We start by specifying that two attractors can be reached, one in switch-off conditions, and one in switch-on conditions. In switch-off conditions, the Swi5 concentration must eventually remain low, that is, equal to its basal expression level $\kappa^0_{Swi5}/\gamma_{Swi5}$. This is expressed in CTL as $\mathbf{AF}\,\mathbf{AG}\,x_{Swi5}\ low$. In switch-on conditions, an oscillatory behavior in the concentration of Swi5 is expected. It can be formulated by means of the formula $\mathbf{AG}\,\mathbf{AF}\,(x_{Swi5}\ inc \wedge \mathbf{AF} x_{Swi5}\ dec)$, requiring that an increase in x_{Swi5} is observed infinitely often and is necessarily followed by a decrease in x_{Swi5}. In addition to these two basic requirements, we impose that in presence of galactose, the Swi5 concentration cannot indefinitely stay low: $u_{gal}\ high \rightarrow \mathbf{AF}\,\neg x_{Swi5}\ low$. We prefix these specifications so as to express the possibility (**EX**) to reach the appropriate attractor from some initial state, and the necessity (**AX**) to leave the switch-off steady state for all initial states in switch-on conditions. This gives rise to the following property:

$$\begin{aligned}\phi_3 = \quad & \mathbf{EX}\,(u_{gal}\,high \wedge \mathbf{AG}\,\mathbf{AF}\,(x_{Swi5}\,inc \wedge \mathbf{AF}x_{Swi5}\,dec)) \\ & \wedge \mathbf{EX}\,(u_{gal}\,low \wedge \mathbf{AF}\,\mathbf{AG}\,x_{Swi5}\,low) \\ & \wedge \mathbf{AX}\,(u_{gal}\,high \rightarrow \mathbf{AF}\,\neg x_{Swi5}\,low)\end{aligned}$$

5.2 Parametrizations consistent with design objective

Using symbolic model checking, we test the feasibility of ϕ_3. In about 2 min, we find that parametrizations exist for which the system presents the desired behavior (Table 1). Using GNA, we can analyze the proposed parameter ordering. In the presence of galactose, several SCCs are found, with two terminal SCCs attracting the major part of the state space that includes notably the switch-off state: from a switch-off initial state, oscillations necessarily happen. In the absence of galactose, a unique stable steady-state where all genes are off is attracting the entire state space. Indeed, although SCCs are present, they are non-terminal and one can show that the switch-off steady state is eventually always reached.

As explained above one of the time-series in the switch-on conditions contradicts our specification. It is consequently not surprising that none of the parametrizations consistent with the experimental data satisfies our design requirements, suggesting that changes are needed. We therefore tried to find other parametrizations, consistent with ϕ_3. Our method indeed finds an order on the threshold and focal parameters satisfying the property (proven in 126 s), while the enumeration of all 4860 parametrizations shows that only 7 are valid (1300 s; Table 1).

A first surprising feature is that $\theta^c_{Swi5} < \kappa^0_{Swi5}/\gamma_{Swi5}$: Swi5 must always activate *CBF1*. Stated differently, this constraint simply suggests to remove the regulation of *CBF1* by Swi5. This can be explained by a qualitative analysis of the system dynamics. In the presence of galactose, we expect oscillations for Swi5. However, the presence of Swi5 is required for the expression of *CBF1* since the *HO* promoter functions like an AND gate: *HO* is on if and only if Swi5 is present *and* Ash1 is absent. So, if Swi5 is not permanently present, Cbf1 and then Gal4 might diseappear, causing the system to converge to the switch-off state.

A second surprising feature is that the regulation of *GAL80* by Swi5 should not be effective. Indeed $\theta^g_{Swi5} < \kappa^0_{Swi5}/\gamma_{Swi5}$ or $\theta_{Gal80} < \kappa^0_{Gal80}/\gamma_{Gal80}$ means that either the *GAL80* promoter is always activated, or that the Gal80 concentration is always sufficient to repress *SWI5*. As above, this suggests to remove an interaction, namely the regulation of *GAL80* by Swi5. Interestingly, the demand for increased external control of the system leads us to a simplified design in which two out of the three feedback loops are removed.

6 Discussion

We proposed a method for efficient search of the parameter space of qualitative models of genetic regulatory networks. This allows us to test whether a hypothesized structure of the network is consistent with the observed behavior, or whether a proposed structure can generate a desired behavior.

On the methodological side, the main novelty is that we develop a symbolic encoding of the dynamics of PADE models, enabling the use of highly

efficient model-checking tools for analyzing incompletely parametrized models. The symbolic encoding avoids the explicit generation of the state space, and the enumeration of possible parametrizations. Although developed for PADE models, the main ideas underlying the approach carry over to logical models [27].

On the biological side, we show the practical relevance of our approach by means of an application to the IRMA network. The parameter constraints we obtained are precise, have a clear biological interpretation, and are consistent with independent experimental observations. Even when considering complex dynamical properties, the search of the parameter space takes at most a few minutes. Our results seem to confirm the intended separation of IRMA from the host network, and suggest that to obtain a more robust response to the addition of galactose, an effective rewiring of the network would be needed.

In comparison with traditional quantitative modeling approaches, the results we obtain are quite general, since they do not depend on a specific representation of the molecular details of the interactions and on specific parameter values. Moreover, the analysis is exhaustive in the sense that the entire parameter space is scanned. These two features are particularly interesting for negative results, such as showing that a given design is not likely to present a desired behavior. In contrast, quantitative ODE models like those developed in [5] do not predict a range of possible behaviors but rather single out one likely behavior with quantitative traits. Qualitative and quantitative approaches provide complementary information on system dynamics.

In comparison with other analysis and verification methods developed for similar modeling formalisms [1, 4, 9, 15], our approach is original in two respects. First, it applies to incompletely parametrized models and can handle any dynamical property of the network expressible in the temporal logic supported by the model checker. Second, we reason at a finer abstraction level, in that we take into account dynamics on the thresholds and work with a partition of the state space preserving derivative sign patterns. The latter feature is particularly well-suited for the comparison of model predictions with time-series data in IRMA.

An interesting direction for further research would be to consider even more general problems in which not only parameters but also regulation functions are incompletely specified. This would make a connection with work on the reverse engineering of Boolean models (*e.g.*, [19, 20, 24].

Acknowledgment

We would like to thank Delphine Ropers, Maria Pia Cosma, and Diego di Bernardo for helpful discussions and contributions to this work. We acknowledge financial support by the European Commission COBIOS FP6-2005-NEST-PATH-COM/043379.

Appendix

A Transition rules

Prop. 2 (Internal transition)**.** Let $D, D' \in \mathcal{D}$ and $D = D'$. $D \to D'$ is a internal transition iff

> $\forall i \in [1, n]$, such that D_i coincides with a value in $\Theta_i \cup \Lambda_i$, it holds that $0 \in F_i(D, D)$

Sketch of the proof.
(Necessity): If for some $i \in [1, n]$, D_i coincides with a value in $\Theta_i \cup \Lambda_i$ and $0 \notin F_i(D, D)$, then because $\forall x \in D$, $F_i(x) = F_i(D, D)$, any solution ξ of (3) starting in D satisfies $\dot{\xi}_i(0) \neq 0$, and consequently instantaneously leaves D.
(Sufficiency): We only need to show that for some $x \in D$, there exists a solution ξ of (3) that remains in D for some (possibly small) time interval $[0, \tau]$. Let x_0 be any point in D. For all dimensions i where D_i coincides with a value in $\Theta_i \cup \Lambda_i$, choose $\xi_i(t) = x_{0_i}$ for $t \geq 0$. For all other dimensions i, choose any solution of the differential inclusion $\dot{x}_i \in F_i(x)$. Then for a sufficiently small $\tau > 0$, $\xi(t) = (\xi_1(t), \ldots \xi_n(t))$ remains in D and is a solution of (3) on $[0, \tau]$. □

Prop. 3 (Dimension-increasing transition)**.** Let $D, D' \in \mathcal{D}$ and $D \subseteq \partial D'$. $D \to D'$ is a dimension-increasing transition iff

1. $\forall i \in [1, n]$, such that D_i and D'_i coincide with a value in $\Theta_i \cup \Lambda_i$, it holds that $0 \in F_i(D', D)$, and
2. $\forall i \in [1, n]$, such that $D_i \neq D'_i$, it holds that $\exists \alpha > 0$ such that $\alpha \in F_i(D', D)\,(D'_i - D_i)$

Sketch of the proof.
(Necessity): Condition 1 expresses that if D_i and D'_i are singletons, then any solution ξ remaining in D' should satisfy $\dot{\xi}_i(t) = 0$. The proof is made as in Proposition 2. For condition 2, assume that for some i such that $D_i \neq D'_i$, all $\alpha \in F_i(D', D)\,(D'_i - D_i)$ are non-positive. Moreover, assume without loss of generality that $D'_i - D_i > 0$. Then denoting $\underline{\lambda}_i$ and $\overline{\lambda}_i$ the bounds of the interval $\sum_{l \in L_i} (\kappa_i^l / \gamma_i)\; B_i^l(D')$, we have for any $x \in D$, $F_i(D', D)/\gamma_i = [\underline{\lambda}_i, \overline{\lambda}_i] - x_i \leq 0$. We deduce that for all $x' \in D'$, $F_i(x')/\gamma_i = [\underline{\lambda}_i, \overline{\lambda}_i] - x'_i < 0$. Consequently, given the relative positions of D and D', no solution can enter D' from D.

(Sufficiency): We show that there exists a solution ξ of (3) that starts in some $x_0 \in D$ ($\xi(0) = x_0$) and enters and remains in D' for some (possibly small) time interval ($\xi(t) \in D', t \in]0, \tau]$). Let x_0 be any point in D. For all dimensions $i \in [1, n]$ where D_i and D'_i are singletons, choose $\xi_i(t) = x_{0_i}$ for $t \geq 0$. For all dimensions $i \in [1, n]$ where $D'_i - D_i > 0$ (the case $D'_i - D_i < 0$ being symmetrical), D_i is a singleton, and for any $x \in D$, $\max F_i(D', D)/\gamma_i = \overline{\lambda}_i - x_i > 0$ implies that $\max F_i(x') = \gamma_i(\overline{\lambda}_i - x'_i) > 0$ for all x' in a (possibly small) neighborhood of x in D'. Then choose for ξ_i the solution of the differential equation $\dot{x}_i = \gamma_i(\overline{\lambda}_i - x_i)$ with $\xi_i(0) = x_{0_i}$. For all other dimensions, choose any solution of the differential inclusion $\dot{x}_i \in F_i(x)$, with $\xi_i(0) = x_{0_i}$. Then for a sufficiently small $\tau > 0$, $\xi(t) = (\xi_1(t), \ldots \xi_n(t))$ starts in D, remains in D' on $]0, \tau]$ and is a solution of (3) on $[0, \tau]$. □

Prop. 4 (Dimension-decreasing transition)**.** Let $D, D' \in \mathcal{D}$ and $D' \subseteq \partial D$. $D \to D'$ is a dimension-decreasing transition iff

A) 1. $\forall i \in [1, n]$, such that D_i and D'_i coincide with a value in $\Theta_i \cup \Lambda_i$, it holds that $0 \in F_i(D, D')$, and

2. $\forall i \in [1, n]$, such that $D_i \neq D'_i$, it holds that $\exists \alpha > 0$ such that $\alpha \in F_i(D, D')\,(D'_i - D_i)$

B) or $\forall i \in [1, n]$, $0 \in F_i(D, D')$

Sketch of the proof.
(Necessity): Assume that A1 does not hold. Then B does not hold either. This means that for some $i \in [1, n]$ such that D_i and D'_i coincide with a value in $\Theta_i \cup \Lambda_i$, it holds that $0 \notin F_i(D, D')$. Then it can be shown as in the proof of Prop. 2 that no solution can remain in D. Now, assume that neither A2 nor B holds. As in the proof of Prop. 3 and using the same notations, we can show that not A2 implies that if $D'_i - D_i > 0$, then $\overline{\lambda}_i - x'_i \leq 0$. If $\overline{\lambda}_i - x'_i < 0$, there cannot be transitions from D to D'. If $\overline{\lambda}_i - x'_i = 0$, one can show that the solutions of the differential equation $\dot{x}_i = \max F_i(x)$ reach $\overline{\lambda}_i = x'_i$ after an infinite amount of time. Then, obviously no solution of (3) can reach $D'_i = \{x'_i\}$ in lesser time. But then, the asymptotic convergence towards some point $x' \in D'$ implies that for all $i \in [1, n]$, $0 \in \lim_{x \to x'} F_i(x)$, and hence $0 \in F_i(D, D')$. Indeed, if for some $i \in [1, n]$ and $\epsilon > 0$, $F_i(D, D') > \epsilon$ (or equivalently if $F_i(D, D') < \epsilon$), any solution would leave in finite time any neighborhood in D of any point $x' \in D'$.

(Sufficiency): Assume that conditions A1 and A2 hold. Then one can construct as in the proof of Prop 3 a solution ξ that starts and remains in D some time interval $[0, \tau[$ ($\xi(t) \in D, t \in [0, \tau[$) and enters in D' at time τ ($\xi(\tau) \in D'$). Alternatively, assume that condition B holds. Then, $\forall i \in [1, n]$, $0 \in F_i(D, D')$ implies that for some $x^* \in D'$, $0 \in \sum_{l \in L_i} (\kappa_i^l / \gamma_i)\, B_i^l(D) - x_i^*$, $i \in [1, n]$. Let x_0 be any point in D and $\xi_i(t)$ be the solution of $\dot{x}_i = \gamma_i(x_i^* - x_i)$ on $[0, \infty[$ with $\xi_i(0) = x_{0i}$, $i \in [1, n]$. One can check that $\xi = (\xi_1, \ldots, \xi_n)$ is a solution of (3) such that $\forall t \geq 0$, $\xi(t) \in D$, and $lim_{t \to \infty} \xi(t) = x^* \in D'$. □

B Comparison with previous definition of dynamics

In [2], we introduced a different definition of the dynamics in regions of step function discontinuities -that is, threshold hyperplanes- also based on differential inclusions. The goal of this section is to show that in most cases the differential inclusions, and hence the set of solutions, are the same for both definitions.

The definition proposed in [2] makes use of the notions of regular and singular domains. Intuitively speaking, singular domains are located on threshold or focal planes, contrary to regular domains. Moreover, one defines for any singular domain D, $R(D)$ as the set of regular domains surrounding D. We refer to [2] for the precise definition of these notions.

For simplicity of notations, we set $R(D) = D$ for a regular domain D. Also, for having more compact notations, we introduce $g_i(D, x)$ as the value of f in D infinitely close to x: $g_i(D, x) = \lim_{y \to x, y \in D} f_i(y)$. Naturally, $g_i(D, x)$ is well defined only if x is in $\overline{D}$, the closure of D, and $g_i(D, x) = f_i(x)$ if $x \in D$. Then

in [2], the dynamics is defined as

$$\dot{x}_i \in G_i(x) \triangleq [\min_{D' \in R(D)} g_i(D', x), \max_{D' \in R(D)} g_i(D', x)] \tag{6}$$

The definition we propose in this paper is more conservative than the one given in [2].

Prop. 5.

$$G_i(x) \subseteq F_i(x), \qquad x \in \Omega$$

Sketch of the proof.
As was done for $g_i(D, x)$, we define $s^+(D, x_j, \theta_j)$ and $b_i^l(D, x)$ as $\lim_{y \to x, y \in D} s^+(y_j, \theta_j)$ and $\lim_{y \to x, y \in D} b_i^l(y)$, respectively. For any $x \in D$, $D \subseteq \Omega$, one can easily show that

$$\forall D' \in R(D),\ s^+(D', x_j, \theta_j) \in S^+(x_j, \theta_j)$$

If $D_j \neq \{\theta_j\}$, then this obviously holds, as $S^+(x_j, \theta_j) = [0, 0]$ (or $[1, 1]$) and for all $D' \in R(D)$, $s^+(D', x_j, \theta_j) = 0$ (or 1). If $D_j = \{\theta_j\}$, then $S^+(x_j, \theta_j) = [0, 1]$, and necessarily, $s^+(D', x_j, \theta_j) \in S^+(x_j, \theta_j)$ for all $D' \in R(D)$. Then, from interval arithmetics, we have

$$\forall D' \in R(D),\ b_i^l(D', x) \in B_i^l(x)$$

and finally

$$\forall D' \in R(D),\ g_i(D', x) \in F_i(x)$$

From the definition of $G_i(x)$, we conclude that $G_i(x) \subseteq F_i(x)$. □

In most cases, *the inclusion of Prop 5 is an equality.* In fact strict inclusion can arise only in two cases. The first one occurs when a protein has a dual role (activator and inhibitor) with a same activity threshold on a single promoter. For example, this is the case if some b_i^l term equals $s^+(x_j, \theta_j) \cdot (1 - s^+(x_j, \theta_j))$. Indeed, for $x_j = \theta_j$, it holds that $B_i^l(x) = [0, 1]$ whereas $\max_{D' \in R(D)} b_i^l(D', x) = 0$. The second case occurs when a protein has a dual role with a same activity threshold on two different promoters of a gene. For example, this is the case if $f_i(x)$ equals $\kappa_i^1 s^+(x_j, \theta_j) + \kappa_i^2(1 - s^+(x_j, \theta_j)) - \gamma_i x_i$. Indeed, for $x_j = \theta_j$ it holds that $F_i(x) = [0, \kappa_i^1 + \kappa_i^2] - \gamma_i x_i$, whereas $G_i(x) = [\kappa_i^1, \kappa_i^2] - \gamma_i x_i$, assuming $\kappa_i^1 < \kappa_i^2$. These rare cases appear in none of the PADE models developed so far and distributed with GNA, and might be considered as modeling problems.

References

[1] J. Barnat, L. Brim, I. Cerna, S. Drazan, J. Fabrikova, and D. Safranek. On algorithmic analysis of transcriptional regulation by LTL model checking. *Theor. Comput. Sci.*, 410(33-34):3128–48, 2009.

[2] G. Batt, H. de Jong, M. Page, and J. Geiselmann. Symbolic reachability analysis of genetic regulatory networks using discrete abstractions. *Automatica*, 44(4):982–9, 2008.

[3] 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. *Bioinformatics*, 21(Suppl 1):i19–28, 2005.

[4] G. Bernot, J.-P. Comet, A. Richard, and J. Guespin. Application of formal methods to biological regulatory networks. *J. Theor. Biol.*, 229(3):339–48, 2004.

[5] I. Cantone, L. Marucci, F. Iorio, M.A. Ricci, V. Belcastro, M. Bansal, S. Santini, M. di Bernardo, D. di Bernardo, and M.P. Cosma. A yeast synthetic network for *in vivo* assessment of reverse-engineering and modeling approaches. *Cell*, 137(1):172–81, 2009.

[6] M. Chaves, A. Sengupta, and E.D. Sontag. Geometry and topology of parameter space: Investigating measures of robustness in regulatory networks. *J. Math. Biol.*, 59(3):315–58, 2009.

[7] 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 *CAV'02*, volume 2404 of *LNCS*, pages 359–64. Springer, 2002.

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

[9] F. Corblin, S. Tripodi, E. Fanchon, D. Ropers, and L. Trilling. A declarative constraint-based method for analyzing discrete genetic regulatory networks. *Biosystems*, 98(2):91–104, 2009.

[10] H. de Jong D. Ropers, V. Baldazzi. Model reduction using piecewise-linear approximations preserves dynamic properties of the carbon starvation response in *Escherichia coli*. *ACM/IEEE Trans. Comput. Biol. Bioinform.*, 2010. In press.

[11] M. Davidich and S. Bornholdt. The transition from differential equations to Boolean networks: a case study in simplifying a regulatory network model. *J. Theor. Biol.*, 255(3):269–77, 2008.

[12] R. Edwards and L. Glass. A calculus for relating the dynamics and structure of complex biological networks. In R.S. Berry and J. Jortner, editors, *Adventures in Chemical Physics*, volume 132, pages 151–78. Wiley, 2006.

[13] A. Fauré, A. Naldi, C. Chaouiya, and D. Thieffry. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics*, 22(14):e124–31, 2006.

[14] J. Fisher and T.A. Henzinger. Executable cell biology. *Nat. Biotechnol.*, 25(11):1239–50, 2007.

[15] J. Fromentin, J.-P. Comet, P. Le Gall, and O. Roux. Analysing gene regulatory networks by both constraint programming and model-checking. In *EMBC07*, pages 4595–4598, 2007.

[16] L. Glass and S.A. Kauffman. The logical analysis of continuous non-linear biochemical control networks. *J. Theor. Biol.*, 39(1):103–29, 1973.

[17] G. Goessler. Component-based modeling and reachability analysis of genetic networks. *ACM/IEEE Trans. Comput. Biol. Bioinform.*, 2010. In press.

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

[19] S. Liang, S. Fuhrman, and R. Somogyi. REVEAL, A general reverse engineering algorithm for inference of genetic network architectures. In *PSB'98*, volume 3, pages 18–29, 1998.

[20] S. Martin, Z. Zhang, A. Martino, and J.-L. Faulon. Boolean dynamics of genetic regulatory networks inferred from microarray time series data. *Bioinformatics*, 23(7):866–74, 2007.

[21] P.T. Monteiro, E. Dumas, B. Besson, R. Mateescu, M. Page, A.T. Freitas, and H. de Jong. A service-oriented architecture for integrating the modeling and verification of genetic regulatory networks. *BMC Bioinform.*, 10:450, 2009.

[22] P.T. Monteiro, D. Ropers, R. Mateescu, A.T. Freitas, and H. de Jong. Temporal logic patterns for querying dynamic models of cellular interaction networks. *Bioinformatics*, 24(16):i227–33, 2008.

[23] R.E. Moore. *Methods and Applications of Interval Analysis*. SIAM, 1979.

[24] T.J. Perkins, M. Hallett, and L. Glass. Inferring models of gene expression dynamics. *J.Theor. Biol.*, 230(3):289–99, 2004.

[25] A. Polynikis, S.J. Hogan, and M. di Bernardo. Comparing different ODE modelling approaches for gene regulatory networks. *J. Theor. Biol.*, 261(4):511–30, 2009.

[26] J. Saez-Rodriguez, L.G. Alexopoulos, J. Epperlein, R. Samaga, D.A. Lauffenburger, S. Klamt, and P.K. Sorger. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol.*, 5:331, 2009.

[27] R. Thomas and R. d'Ari. *Biological Feedback*. CRC Press, 1990.

[28] T.Y.-C. Tsai, Y.S. Choi, W. Ma, J.R. Pomerening, C. Tang, and J.E. Ferrell Jr. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science*, 321(5885):126–9, 2008.

Contents

1 Introduction **3**

2 Qualitative model of IRMA network **4**
2.1 IRMA network 4
2.2 Measurements of IRMA dynamics 4
2.3 PADE model of IRMA network 5

3 Search of parameter space using symbolic model checking **8**
3.1 Qualitative analysis of PADE models 8
3.2 Search of parameter space: a model-checking approach 10
3.3 Symbolic encoding of PADE model and dynamics 10

4 Validation: consistency of IRMA network with experimental data **12**
4.1 Temporal-logic encoding of observations 12
4.2 Testing consistency of network with observations 13
4.3 Detailed analysis of valid parameter set 13

5 Re-engineering: improving external control by galactose **15**
5.1 Temporal-logic specification of design objective 15
5.2 Parametrizations consistent with design objective 16

6 Discussion **16**

A Transition rules **18**

B Comparison with previous definition of dynamics **19**

Centre de recherche INRIA Paris – Rocquencourt
Domaine de Voluceau - Rocquencourt - BP 105 - 78153 Le Chesnay Cedex (France)

Centre de recherche INRIA Bordeaux – Sud Ouest : Domaine Universitaire - 351, cours de la Libération - 33405 Talence Cedex
Centre de recherche INRIA Grenoble – Rhône-Alpes : 655, avenue de l'Europe - 38334 Montbonnot Saint-Ismier
Centre de recherche INRIA Lille – Nord Europe : Parc Scientifique de la Haute Borne - 40, avenue Halley - 59650 Villeneuve d'Ascq
Centre de recherche INRIA Nancy – Grand Est : LORIA, Technopôle de Nancy-Brabois - Campus scientifique
615, rue du Jardin Botanique - BP 101 - 54602 Villers-lès-Nancy Cedex
Centre de recherche INRIA Rennes – Bretagne Atlantique : IRISA, Campus universitaire de Beaulieu - 35042 Rennes Cedex
Centre de recherche INRIA Saclay – Île-de-France : Parc Orsay Université - ZAC des Vignes : 4, rue Jacques Monod - 91893 Orsay Cedex
Centre de recherche INRIA Sophia Antipolis – Méditerranée : 2004, route des Lucioles - BP 93 - 06902 Sophia Antipolis Cedex

Éditeur
INRIA - Domaine de Voluceau - Rocquencourt, BP 105 - 78153 Le Chesnay Cedex (France)
http://www.inria.fr
ISSN 0249-6399