Constraints Programming for Unifying Gene Regulatory Networks Modeling Approaches

Damien Eveillard[1], Jonathan Fromentin[2] & Olivier Roux[2]

[1] Computational Biology (ComBi) group, LINA UMR 6241, CNRS & Université de Nantes
2, rue de la Houssinière - BP 92 208 - 44322 Nantes CEDEX 03
damien.eveillard@univ-nantes.fr
[2] IRCCyN UMR 6597, CNRS & École Centrale de Nantes
1, rue de la Noë - BP 92 101 - 44321 Nantes CEDEX 03
{jonathan.fromentin,olivier.roux}@irccyn.ec-nantes.fr

Abstract. Qualitative approaches, like Piecewise-Affine Differential Equations (PADEs) or those inspired from the R. Thomas formalism, represent one of the major recent improvements in biological modeling. We show herein that these approaches might be naturally represented within a unified theoretical framework using constraints. This result allows us to reason about biological models which is helpful for (i) passing from one qualitative formalism to another one and as well for (ii) building a constraints-based protocol that opens perspectives on modeling large genetic regulatory systems.

1 Introduction

Experimental approaches that study living system behaviors, focus on various and complementary aspects: (i) a set of genes that composes gene regulatory networks and (ii) a set of proteins that shapes metabolic networks. However, despite their clear experimental distinction, both components belong to the same system and interact between them for producing specific dynamical biological behaviors (see Fig. 1 for illustration). It hence remains interesting to mix up this distinct information through a unique modeling approach. It is achieved by various recent modeling techniques that focus on the dynamical biological behavior (see [1] for review) with a special emphasis on their qualitative behaviors. These approaches consider the gene interaction as the corner stone of an accurate macromolecular system modeling. Like this, each gene regulatory reaction summarizes a protein production that activates/represses the target gene. Among these modeling techniques, the approaches based on Piecewise-Affine Differential Equations (PADEs) [2, 3] and the R. Thomas formalism [4] showed astonishing achievements at investigating gene regulatory network properties and share as well common biological assumptions (i.e. discretizing the gene interaction impact). However, although these modeling techniques show at similar biological results, they focus on distinct theoretical features. We propose to present herein theoretical investigations that show that two modeling approaches might be

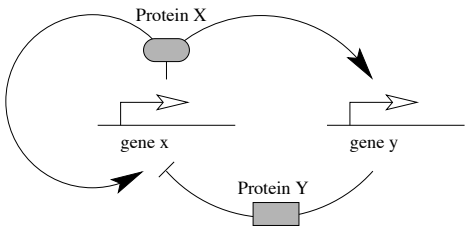


Fig. 1. Description of a two genes interaction network that resumes a system composed of genes x and y. The gene x produces the protein X that activates the transcription of genes x (i.e. auto-activation) and y. It implies a production of the protein Y that represses the transcription of the gene x.

unified within a unique framework using the constraints programming. It emphasizes a description of a novel biological modeling protocol that deals with assumptions used in either formalisms. It allows as well further investigations like a fine reasoning on constraints related to specific biological behaviors.

This paper will introduce first in Sec. 2 the unified constraints based framework. After a brief overview of the modeling approaches of interest (Sec. 2.1), we are going to show how to transform a PADEs model and a Thomas's model into a set of constraints (Sec. 2.2). This set of constraints describes the discrete dynamics that might be investigated for a better understanding of the biological behaviors. As a guideline, such a protocol will be illustrated on a simplistic system shown in Fig. 1. Second, Sec. 3 will present two kinds of analysis based on the previous constraints. In particular, for investigating large gene regulatory network, Sec. 3.1 will show a constraints based trimming approach that restricts the study of the model on the behaviors of interest (i.e. experimentally investigated genes). Such a refinement will allow a more precise reasoning using a symbolic model-checking (Sec. 3.2) that focuses on interesting behaviors for an experimental validation.

2 Constraints for Modeling Genetic Regulatory Systems

2.1 Qualitative Approaches

The regulation of genetic system is achieved via macromolecular interactions that describe positive and negative feedback loops. Qualitative approach appeared quickly as an appropriate way for investigating such a complexity. We mention herein the formalisms that have been successful during the last decade and that might be expressed in a natural manner by a set of constraints.

The Biological Regulatory Graph (BRG) is widely applied for a discrete modeling of gene regulatory networks like in Fig. 1. A BRG is a labelled directed graph $G = (V, E)$ where V is the set of vertices and E is the set of edges (see Fig. 2(a)). Each edge $(i \to j) \in E$ is labelled with a couple $(\alpha_{ij}, \theta_{ij})$ where $\alpha \in \{+, -\}$ is

the sign of interactions (respectively activation and repression) and θ_{ij} is the concentration threshold beyond which the regulation is effective.

Notation 1 *We note L_i, the set of labels related to the regulatory functions of the gene i that we call the resources of i.*

The System of Piecewise-Affine Differential Equations (PADEs) represents as well the dynamic of a genetic regulatory network [5, 6]. The system follows the form:

$$\dot{x}_i = f_i(x) - \gamma_i x_i \quad with \quad 0 \leq x_i \quad and \quad 1 \leq i \leq n \tag{1}$$

where $x = (x_1, \ldots, x_n)$ is a vector of protein concentrations called the quantitative state of the system. (1) describes the variation of the concentration x_i as the difference between the rate of synthesis $f_i(x)$ and the rate of degradation $\gamma_i x_i$. Note that $f_i(x)$ expresses the dependency between the synthesis rate of i and its regulator concentrations. It can be defined as:

$$f_i(x) = k_i + \sum_{j \in L_i} k_{ij} b_{ij}(x) \tag{2}$$

where $k_i, k_{ij} \in \mathbb{R}^{+*}$ are the kinetic parameters and, b_{ij} is a sigmoidal function approximated by a combination of step functions s^+ and s^- such as for a regulator gene i' of i, we have:

$$s^+(x_{i'}, \theta_{i'}) = \begin{cases} 1, & x_{i'} > \theta_{i'} \\ 0, & x_{i'} < \theta_{i'} \end{cases} \quad \text{et} \quad s^-(x_{i'}, \theta_{i'}) = 1 - s^+(x_{i'}, \theta_{i'}) \tag{3}$$

where θ_i is a concentration threshold. For illustration, Fig. 2(b) shows the PADE system that models the biological behavior of the system in Fig. 1.

The Discrete Modeling Formalism of R. Thomas (Fig. 2(c) for illustration) is a natural discrete description of the BRG shown above and represents as well a discretization of a PADE system. The Thomas's formalism have to take into account two kinds of parameters.

Numbering thresholds and discretization state. The thresholds numbering keeps the order between the qualitative thresholds mentioned above. Therefore, for the thresholds of i like $\theta_i^1 < \theta_i^2 < \cdots < \theta_i^n$ then its qualitative thresholds are $t_i^1 < t_i^2 < \cdots < t_i^n$ with $\forall j \in [1, n]$, $t_i^j = j$. The states of the system are thus discretized into domains by the function $\mathcal{D}$ defines as:

$$\mathcal{D}_i(x_i) = \begin{cases} t_i^j, & \theta_i^j < x_i < \theta_i^{j+1} \\ 0, & x_i < \theta_i^1 \end{cases} \tag{4}$$

Parameters in discrete modeling. To each qualitative domain s is associated a qualitative focal point standing for the tendency of evolution in s. For each qualitative domain s within the discrete abstraction, the vector of

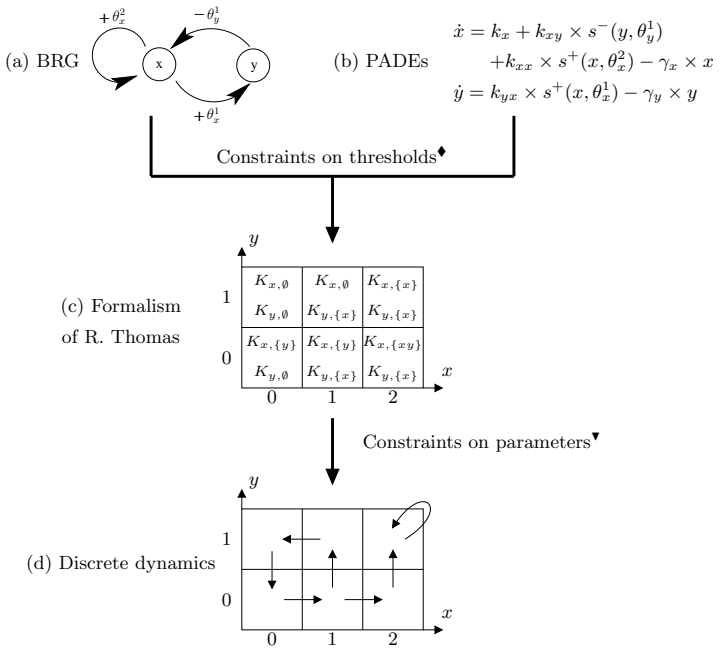


Fig. 2. Constraints-based protocol applied on the two genes system, where ♦ are $0 < \theta_x^1 < \theta_x^2 < max_x$ and $0 < \theta_y^1 < max_y$; ▼ are $0 < \frac{k_x}{\gamma_x} < \theta_x^1$ and $\theta_y^1 < \frac{k_{yx}}{\gamma_y} < max_y$ and $\theta_x^2 < \frac{k_x + k_{xy}}{\gamma_x}, \frac{k_x + k_{xx}}{\gamma_x}, \frac{k_x + k_{xy} + k_{xx}}{\gamma_x} < max_x$.

discrete parameters $(K_{1,\omega_1(s)}, \ldots, K_{n,\omega_n(s)})$ gives the position of the qualitative focal point. The focal point is the abstract region containing the steady state for the PADEs in each domain. The concentration evolution is continuous in a domain and the system state tends toward the focal point of the domain. The discrete parameter of the gene i in the domain s is obtained with $\dot{x}_i = 0$ in (1) and $b_{ij} = 1$ (due to the presence of the resource j) in (2):

$$K_{i,\omega_i(s)} = \mathcal{D}_i \left(\frac{k_i + \sum_{j \in \omega_i(s)} k_{ij}}{\gamma_i} \right) \tag{5}$$

where $\omega_i(s) \subseteq L_i$ represents the resources of i in the domain s. The valuation of these discrete parameters gives a discrete dynamics where each transition between two contiguous domains is asynchronous. For illustration, it gives the discrete dynamics shown in Fig. 2(d).

2.2 Using Constraints for Building a Discrete Dynamics Model

Based on previous descriptions, we propose to transform one formalism into another using an automatic approach that integrates the qualitative formalisms (see Fig. 2).

Transforming a BRG or a PADEs system into the Thomas's formalism (Fig. 2(a) and Fig. 2(b) to Fig. 2(c)) is achieved by reasoning on the knowledge associated with the thresholds. There can be simple equality or inequality constraints that allow the numbering of thresholds. These constraints are of the form $\theta_i = \theta'_i$ or $\theta_i < \theta'_i$ where i is a gene.

Transforming the R. Thomas formalism into discrete dynamics (Fig. 2(c) to Fig. 2(d)) is achieved by two distinct approaches.

Using the inequality constraints on the kinetic parameters like

$$\theta_i < \frac{k_i + \sum_{j\in\omega} k_{ij}}{\gamma_i} \quad \text{or} \quad \theta_i > \frac{k_i + \sum_{j\in\omega} k_{ij}}{\gamma_i} \tag{6}$$

where $\omega \subseteq L_i$, θ_i is a threshold of i and, where $\left(k_i + \sum_{j\in\omega} k_{ij}\right)/\gamma_i$ is a component of a focal point that gives the tendency of the evolution of i with the ressources ω. Both constraints are directly extracted from the PADEs formalism [7] and provide the discrete parameters values. The number of these constraints is proportional to the number of kinetic parameters. Both inequality constraints indicate the localization of the focal point within a domain. Therefore the number of these constraints is twice the number of components of focal points.

Using temporal qualitative specification when inequality constraints are difficult to obtain. Among studies that propose such an approach, two use the constraints programming. Both approaches chosen the use of *reified constraints*[1], because the formalism of R. Thomas produces graphs that might contain domains with multiple successors. The set of constraints are based on simple inequality constraints on the discrete parameters, which give the notion of successors. These constraints are usually not sufficient for depicting a unique discrete dynamic but give a set of possible discrete dynamics. In this purpose, F. Corblin *et al.* [8] uses constraint logic programming for analysis of GRN by knowing qualitative pathways or stable qualitative domains (with no out-going transitions). The number of constraints is a linear function of the number of qualitative domains in the pathway. And, the number of equations expressing a transition between qualitative domains is also a linear function of the number of component of the qualitative focal

[1] by adding boolean parameters such that the parameter is true iff the linked constraints are true

points. On the other hand, J. Fromentin *et al.* [9] uses constraint programming and the CTL language to find the discrete parameters. For example, we consider the CTL formula $x = 0 \Rightarrow EF(x = 1)$ in Fig. 2(c) to force the discrete dynamics in Fig. 2(d) in order to have a pathway from $x = 0$ to $x = 1$. For any operator $\diamond$, we associate a Constraint $C^{s_i}_{\diamond}$ at each discrete domain s_i. In addition for a CTL operator $\triangle$, we associate a boolean variable $B^{s_i}_{\triangle}$ at each discrete domain s_i. This boolean variable indicates if the related discrete domain validates or not the operator constraints. Therefore, the principle is to propagate the information given by the possible successors of the discrete domains *via* their reified constraints and their boolean variables. In this case, the reified constraints are more complex than those explained in [8] because they must be equivalent to those applied within the CTL formulae. Nevertheless, the basic constraints for the transitions are the same: similar equality or inequality constraints on the discrete parameters. For illustration, we consider the sub-graph of Fig. 2 (c) that includes two domains $s_1 = (0, 0)$ and $s_2 = (1, 0)$. The application of $x = 0 \Rightarrow EF(x = 1)$ on this sub-graph implies this following decomposition for s_1:

- $x = 0$ implies the constraint $C^{s_1}_{x=0} \equiv true$
- $x = 1$ implies $C^{s_1}_{x=1} \equiv false$
- $EF(x = 1)$ implies $C^{s_1}_{EF(x=1)} \equiv B^{s_1}_{EF(x=1)}$ for which
  $B^{s_1}_{EF(x=1)} \Leftrightarrow C^{s_1}_{x=1} \vee \left(B^{s_2}_{EF(x=1)} \wedge K_{x,\{y\}} > 0\right)$
- $x = 0 \Rightarrow EF(x = 1)$ implies $C^{s_1}_{x=0 \Rightarrow EF(x=1)} \equiv C^{s_1}_{x=0} \Rightarrow C^{s_1}_{EF(x=1)}$

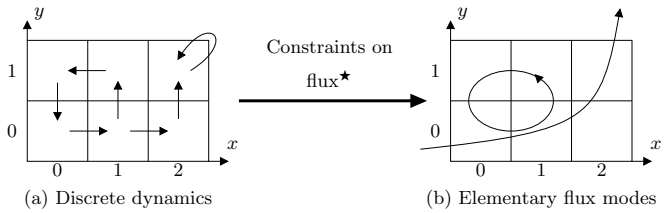
and this decomposition for the domain s_2:

- $x = 0$ implies the constraint $C^{s_2}_{x=0} \equiv false$
- $x = 1$ implies $C^{s_2}_{x=1} \equiv true$
- $EF(x = 1)$ implies $C^{s_2}_{EF(x=1)} \equiv B^{s_2}_{EF(x=1)}$ for which
  $B^{s_2}_{EF(x=1)} \Leftrightarrow C^{s_2}_{x=1} \vee \left(B^{s_1}_{EF(x=1)} \wedge K_{x,\{y\}} < 1\right)$
- $x = 0 \Rightarrow EF(x = 1)$ implies $C^{s_2}_{x=0 \Rightarrow EF(x=1)} \equiv C^{s_2}_{x=0} \Rightarrow C^{s_2}_{EF(x=1)}$

Note herein that the constraint that satisfies $x = 0 \Rightarrow EF(x = 1)$ in s_2 is a tautology whereras $K_{x,\{y\}} > 0$ (i.e. the transition $s_1 \rightarrow s_2$) have to be true for satisfying $C^{s_1}_{x=0 \Rightarrow EF(x=1)}$. Thus, and according to [9], the number of constraints is related to the number of domains and CTL operators.

3 Reasoning on the Biological Constraints Based Model

Biological knowledge is obviously – by nature – incomplete. Only specific behaviors related to genes of interest are experimentally investigated. On the other hand, biological models are often too large and/or complex for using standard constraints reasoning approaches. Among several solving or analysis techniques, we propose to use the constraints based framework for refining the model on distinct biological components, i.e. genes or gene products, in order (i) to validate a complex model based on specific behaviors, (ii) to emphasize behaviors that might be experimentally studied.



(a) Discrete dynamics (b) Elementary flux modes

Fig. 3. Reasoning on the discrete dynamics given by a result of the constraints-based protocol where ★ allow the flux balance analysis.

3.1 Qualitative Behaviors Trimming

Previously shown constraints mainly describe qualitative behaviors. However analyzing such behaviors remains difficult for large gene regulatory networks. The use of our theoretical framework overcomes in a natural manner this weakness by allowing us to trim qualitative behaviors using an additional constraints-based approach: the flux balance analysis using the Minimal Metabolic Behavior (MMB) technique [10] that is an elegant extension of the Elementary Flux Modes (EFM) approach (see [11] for flux balance analysis overview). This technique is already well-known for analyzing the metabolic flux of a balanced (steady-state) system. It decomposes the flux constraints into minimal elementary pathways. Combinations of these pathways describe multiple paths that material can follow through the system. For applying these techniques on discrete dynamics graphs, we assume (i) a specific qualitative behavior as a combination of qualitative pathways and (ii) the discrete abstraction depicts transient behaviors between stable domains which implies flux between initial and stable domains (or set of domains that produce stable dynamics). Mathematically, the constraints that describe the discrete dynamics graph have the form:

$$Sv = 0, v_i \geq 0, \text{ for } i \in Irr \tag{7}$$

where Irr is the set of the transitions (i.e. irreversible transitions), S is the $s \times m$ stoichiometric matrix of the discrete dynamical network, with s domains (rows) and m transitions (columns), and $v \in \mathbb{R}^m$ is the flux vector. As explained in [10], the set of all possible flux distribution through the discrete dynamics graph at steady state (i.e. all possible solutions of the constraints system described in (7)), defines a polyhedral cone, named the steady state flux cone.

$$C = \{v \in \mathbb{R}^m \mid Sv = 0, v_i \geq 0, i \in Irr\} \tag{8}$$

As illustration, we depict this flux analysis applied on the discrete dynamics model in Fig. 3(a) that represents the behaviors of the simplistic system in Fig. 1. It is obtained by adding an input transition to the domain $(0, 0)$: the initial domain. We consider as well the stable domain $(2, 1)$ as a natural output of the system that finally consists of nine domains and eight transitions (six

regular plus two added transitions). The steady state cone can be represented by two minimal proper faces (Fig. 3(b)) named MMB:

$$\text{MMB}_1 : \rightarrow (0,0) \rightarrow (1,0) \rightarrow (2,0) \rightarrow (2,1) \rightarrow$$
$$\text{MMB}_2 : (0,0) \rightarrow (1,0) \rightarrow (1,1) \rightarrow (0,1) \rightarrow (0,0)$$

where MMB_1 and MMB_2 show respectively a linear and a circular qualitative pathway that passes through the qualitative domains. Interestingly, these two pathways represent the two characteristic behaviors of the system. Note that the lineality space $\text{lin.space}(C) = \{v \in C \mid v_i = 0, i \in Irr\}$ has dimension 0 due to the absence of reversible transitions in the discrete dynamics graph. Such an approach is particularly helpful for trimming a large biological model. Indeed, focusing on a specific gene, we consider only the pathways that possess domains and transitions related to the gene investigated. A linear combination of these MMBs hence produces a subgraph that describes all qualitative behaviors of interest.

3.2 Symbolic Model-Checking

The constraint-based protocol shown above is a natural unified theoretical framework for the qualitative modeling approaches. Furthermore, it achieves to combine additional constraints-based techniques usually applied for analyzing metabolic networks. Beyond these qualitative applications, our framework provides the opportunity to extend the modeling towards quantitative aspects by adding delays on the discrete transitions, hence producing a hybrid model. Several studies were done for analyzing hybrid models of genetic regulatory networks: [3, 12–15]. The common assumption is to partition the qualitative domains. This partition provides a finer transition system such that the sign patterns of the derivatives of concentrations levels are preserved. The methods for partitioning differ according to the different works, and the aim for each one is to give raise to executions that have to be compared with the experimental knowledge. For this purpose, different kinds of *symbolic model checking* techniques are applied, i.e. verify biological temporal properties (e.g. CTL formulae, reachability). It is either classical model checking ([3, 13]) or timed model checking ([14]) or hybrid (parametric) model checking ([12, 15]), and properties are either chronological ([3, 13]) or chronometrical ([14, 15]). Our constraint-based protocol integrates a rather different approach [9] since it implements CTL-model checking algorithms by the mean of constraints programming, which reinforces our unified approach on biological system modeling.

4 Discussion

This study shows that PADEs and Thomas based approaches are convergent. Indeed, we emphasize that both formalisms might be expressed using constraints. In practice, our modeling approach allows to choose between both formalizations according to the experimental knowledge at disposal, i.e. known constraints on

kinetic or discrete parameters. Moreover, our unified framework achieves a novel hybrid description of biological systems that exploits the advantages of both formalisms.

As a natural extension of the biological problem formalization, our unified framework allows several constraints based analyses that focus on distinct goals. Comparing our formalization with different solving frameworks (CP, CSP or SAT solvers [16]) represents by itself an interesting investigation area. However, we consider that one of the advantages of our approach is to produce a well-formalized and/or biologically certified problem that might be suitable for further constraints based investigations.

Our protocol aims at reasoning on more realistic biological networks. As illustration, we applied it on the gene regulatory network of the carbon starvation response in E. coli formalized using a PADEs system (following the description given in [17]). Six genes compose the gene regulatory network that might be represented using 37 constraints (constraints on inequalities and thresholds used in the PADEs system). They produce a discrete dynamics graph with 912 qualitative domains. This problem is hence formalized with simple constraints. However, although the constraints formalization is a relatively easy task, the problem remains difficult to analyze with standard techniques due to the complexity of the discrete dynamics graphs. It confirms the interest of dedicated constraints based techniques for investigating the biological properties of the complete genes interaction networks.

Acknowledgements D.E. thanks Olivier Bernard for long-term discussions on the qualitative modeling approaches. The authors alsos thanks Jamil Ahmad for fruitful discussions during this work.

References

1. de Jong, H.: Modeling and simulation of genetic regulatory systems: a literature review. J Comput Biol **9**(1) (Jan 2002) 67–103
2. de Jong, H., Gouzé, J.L., Hernandez, C., Page, M., Sari, T., Geiselmann, J.: Qualitative simulation of genetic regulatory networks using piecewise-linear models. Bull Math Biol **66**(2) (Mar 2004) 301–40
3. Batt, G., Ropers, D., de Jong, H., Geiselmann, J., Mateescu, R., Page, M., Schneider, D.: Validation of qualitative models of genetic regulatory networks by model checking: analysis of the nutritional stress response in escherichia coli. Bioinformatics **21 Suppl 1** (Jun 2005) i19–28
4. Thomas, R., Thieffry, D., Kaufman, M.: Dynamical behaviour of biological regulatory networks–i. biological role of feedback loops and practical use of the concept of the loop-characteristic state. Bull Math Biol **57**(2) (Mar 1995) 247–76
5. Glass, L., Kauffman, S.: The logical analysis of continuous non linear biochemical control networks. J. Theor. Biol. **39**(1) (1973) 103–129

6. Snoussi, E.: Qualitative dynamics of a piecewise-linear differential equations : a discrete mapping approach. Dynamics and stability of Systems **4** (1989) 189–207
7. de Jong, H., Page, M., Hernandez, C., Geiselmann, J.: Qualitative simulation of genetic regulatory networks: Method and application. In: IJCAI. (2001) 67–73
8. Corblin, F., Fanchon, E., Trilling, L.: Modélisation de réseaux biologiques discrets en programmation logique par contraintes. Technique et science informatiques **26**(numéro spécial "Modélisation et simulation pour la post-génomique") (2007) 73
9. Fromentin, J., Comet, J., Gall, P.L., Roux, O.: Analysing gene regulatory networks by both constraint programming and model-checking. In: EMBC07, 29th IEEE EMBS Annual Int. Conf. (2007) 4595–4598
10. Larhlimi, A., Bockmayr, A.: A new approach to flux coupling analysis of metabolic networks. In Berthold, M.R., Glen, R., Fischer, I., eds.: CompLife. Volume LNBI 4216., Berlin Heidelberg, Springer-Verlag (2006) 205–215
11. Gagneur, J., Klamt, S.: Computation of elementary modes: a unifying framework and the new binary approach. BMC Bioinformatics **5** (Nov 2004) 175
12. Adélaïde, M., Sutre, G.: Parametric analysis and abstraction of genetic regulatory networks. In: Proc. 2nd Workshop on Concurrent Models in Molecular Biol. (BioCONCUR'04), London, UK, Aug. 2004. Electronic Notes in Theor. Comp. Sci., Elsevier (2004)
13. Bernot, G., Richard, A., Comet, J.P., Guespin-Michel, J.: Application of formal methods to biol. regulatory networks: Extending thomas' asynchronous logical approach with temporal logic. J. Theor. Biol. **229**(3) (2004) 339–347
14. Siebert, H., Bockmayr, A.: Incorporating time delays into the logical analysis of gene regulatory networks. In: Int. Conf. on Comput. Methods in Systems Biol. (CMSB'06), Trento, Italy. LNBI 4210, Springer (2006) 169–183
15. Ahmad, J., Bernot, G., Comet, J.P., Lime, D., Roux, O.: Hybrid modelling and dynamical analysis of gene regulatory networks with delays. ComPlexUs **3**(4) (2007) 231–251
16. Tamura, N., Taga, A., Kitagawa, S., Banbara, M.: Compiling finite linear csp into sat. In: Constraints Programming (CP). (2006)
17. Ropers, D., de Jong, H., Page, M., Schneider, D., Geiselmann, J.: Qualitative simulation of the carbon starvation response in escherichia coli. BioSystems **84**(2) (May 2006) 124–52