Property preservation along embedding of biological regulatory networks[*]

Mbarka Mabrouki[12], Marc Aiguier[12],
Jean-Paul Comet[3], and Pascale Le Gall[12]

[1] École Centrale Paris
Laboratoire de Mathématiques Appliquées aux Systèmes (MAS)
Grande Voie des Vignes - F-92295 Châtenay-Malabry
`marc.aiguier@ecp.fr`
[2] Programme d'Épigénomique
523, Place des Terrasses de l'Agora - F-91025 Evry
`{mabrouki,pascale.legall}@epigenomique.genopole.fr`
[3] Laboratory I3S, UMR 6070 CNRS/UNSA
Algorithmes-Euclide-B, 2000 route des Lucioles
B.P. 121, F-06903 Sophia-Antipolis
`comet@unice.fr`

Abstract. In the course of understanding biological regulatory networks (BRN), scientists usually start by studying small BRNs that they believe to be of particular importance to represent a biological function, and then, embed them in the whole network. Such a reduction can lead to neglect relevant regulations and to study a network whose properties can be very different from the properties of this network viewed as a part of the whole. In this paper we study, from a logical point of view, the preservation of properties inherited from small BRNs. The signature of BRN, constituted by a graph, is one of the distinctive features on which embeddings can be defined which leads us to give a first condition on the subgraphs ensuring the preservation of properties of the embedded graphs.

Key words: Biological regulatory networks, network embedding, property preservation, mathematical modeling, computational tree logic

1 Introduction

To understand biological regulatory networks (BRN for short), modeling frameworks and simulation technics are often useful since the complexity of the interactions between constituents of the network (mainly genes and proteins) makes intuitive reasoning difficult [3]. Nevertheless, simulation technics are in practice difficult to manage for most of the systems because they are either large, complex or only partially known. Indeed, the lack of precise knowledge about the

---

[*] This work is performed within the European project GENNETEC *(GENetic NeT-works: Emergence and Complexity)* STREP 34952.

system (are all constituents/interactions taken into account? Which values are given to parameters? Which is the confidence on these parameters?...) is one of the more accurate difficulties to handle computationally all possible hypotheses on the system. Qualitative modeling frameworks have then arose [7, 13, 5]: they consist in abstracting continuous concentrations of constituents into qualitative ones (discrete and finite) although preserving qualitative observations (like presence/absence of a constituent, increasing of the concentration of a target when increasing the one of a regulator...).

We focus in this paper on the multivalued discrete approach developed by R. Thomas and co-workers [13], in which the concentrations of constituents are abstracted by integers to denote thresholds from which constituents can act on other ones in the network. In this formalism, biological systems are described by an interaction graph defining the static part of the system from which we can build a huge but finite set of state transition graphs defining all the possible dynamics of the system. However, given an interaction graph, just a few dynamic models meet the set of biological experiment observations bringing into play interactions between graph's constituents. To cut down in the class of dynamic models and just preserve the good candidates, some recent works expressed these biological experiment observations by temporal properties and used various model-checking technics to select suitable dynamic models [4, 2, 9]. From these works, two software tools have been developed: GNA [4] which automatically checks that a given dynamic model satisfies some biological experiment observations, and SMBioNet [2] which cuts down in the whole class of dynamic models to select the ones that satisfy some given biological experiment observations. In both cited tools, temporal properties denoting biological experiment observations have been expressed in Computation Tree Logic [6].

These logical approaches based on model-checking technics have been shown very efficient to study small BRNs but are not well-adapted for large BRNs. The well-known reason is because model-checking technics are time consuming. Indeed, we have to deal with the limit given by complexity theory: model-checking is based on NP-hard or exponential algorithms. In practice, human cleverness is used to find the situations for which model-checking may become tractable. This is what we propose to do in this paper. Indeed, BRNs are generally embedded into other ones. To allow to describe and study BRN behaviors in the large, we propose in this paper to study the consequences of the embedding. More precisely, we propose to study which are the conditions to impose on the embedding to preserve the dynamics of sub-BRNs. We show that, generally speaking, questioning temporal properties (i.e. biological experimental observations) leads us to study the dynamics of the global BRN "from scratch", i.e without taking benefit of the dynamics of the sub-BRNs, which can be unacceptable in running time. On the contrary, if all dynamics of sub-BRNs are preserved, this then leads us just to focus on the biological experiment observations linked with interactions of the sub-BRNs between them. We can then hope to be able to use both previous tools to automatically study the dynamics of the global BRN. Moreover, this approach corresponds to the classical method used by most of

biologists when they study a biological system. They start by studying small BRNs that they believe to be of particular importance to represent a biological function. The interactions of this BRN with the external genes, are studied only afterwards even if these external genes potentially could influence the behavior of the studied part. Of course, this bottom-up approach makes sense only if there is a complete preservation of sub-systems behaviors as this has been done in this paper up to some sufficient conditions. Otherwise, systems can only be studied globally because of the apparition of emergent properties. To comprehend this notion of emergent properties, we introduced in [1] an abstract mathematical denotation for complex systems.

The paper is then structured as follows: after some reminders on the temporal logic CTL in Section 2, Section 3 presents a logical characterization for BRNs. Section 4 presents the main result of this paper: the preservation of properties through the embedding of BRNs into larger networks. In Section 5, we give a counter-example to justify the constraints we put on graph embedding to ensure property preservation. Finally in Section 6 we give some concluding remarks.

Let us notice the particularity of the logic for BRN presented in this paper: signatures are not simple sets of symbols but are interaction graphs (the static part of BRN). This is what makes tough the definition of the embedding (see Definition 2) as well as the definitions of the consequences of the embedding both on biological experiment observations expressed over sub-BRNs (see Definition 4) and on the dynamics of sub-BRNs embedded into a larger one (see Definition 8). This is what makes also nontrivial the proof of the preservation of temporal properties through embedding.

2   Preliminaries

Computational tree logic (CTL) [6] is a branching-time temporal logic where the structure representing all possible executions is *tree-like* rather than linear. It is well-adapted to specify and reason about non-deterministic and/or concurrent processes. Here, we consider actually a restriction of CTL by removing the next operator X, noted CTL-X [14, 15]. The reason is for biological applications, the logical connector X is not of big relevance. The reason is twofold. First, the time mandatory for a biological system to change of qualitative state is not deterministic and the elapsed time between two consecutive states has a large variance. Secondly, the discretization of the dynamical system abstracts the quantitative time (represented by $t \in \mathbb{R}^+$) into a qualitative time ($n \in \mathbb{N}$). Then the real time necessary for a NEXT transition of the biological system depends also on the number of intermediate states used for the discretization step.

When dealing with propositional fragment of logics, a signature *Atom* is only a set of propositional variables which are the atomic formulas.

Given a signature *Atom*, a model over *Atom*, so-called Kripke frame, is a transition system (S, T) where:

- S is a set whose elements are usually called *states*;
- $T \subseteq S \times S$ is a binary relation satisfying: $\forall s \in S, \exists s' \in S, \ (s, s') \in T$;

and (S, T) is equipped with a total function $L : S \rightarrow 2^{Atom}$ called *labeling function*.

Therefore, models over *Atom* are labeled transition systems where T denotes the transition relation and L is the labeling associating for each state s of S the set of propositional variables true at s.

Formulas over *Atom* are well-formed formulas whose syntactical rules are given by:

$$For ::= ATOM \mid For \Rightarrow For \mid For \wedge For \mid For \vee For \mid \neg For$$
$$AG \ For \mid EG \ For \mid AF \ For \mid EF \ For \mid A[For \ U \ For] \mid E[For \ U \ For]$$

The intuitive meaning of modal operator $F\varphi$ (resp. $G\varphi$) means that φ will be finally (F) (resp. is globally (G)) true. The prefix A (resp. E) means that the formula is true for all possible futures (resp. there exists a future for which the following property is true). Finally, formulas of the form $\varphi \ U \ \psi$ mean that φ has to be true until (U) ψ becomes true. They are also preceded by the prefixes A or E.

The validity of formulas is expressed *via* a binary relation usually denoted by $\models$ between models and formulas over a set of atomic formulas *Atom*. A path is any sequence $\sigma = (s_0, s_1, \ldots, s_n, \ldots)$ such that for every $i \in \mathbb{N}$ we have $(s_i, s_{i+1}) \in T$. Then, $(S, T) \models \varphi$ if for any state $s \in S$, (S, T) *satisfies* φ, denoted by $((S, T), s) \models \varphi$, according to the following inductive definition:

- $((S, T), s) \models p$ iff $p \in L(s)$ for $p \in Atom$;
- $((S, T), s) \models AG\varphi$ (resp. $((S, T), s) \models EG\varphi$) iff for every (resp. there exists a) path $(s_0, s_1, \ldots, s_n, \ldots)$, for every $i \in \mathbb{N}$, $((S, T), s_i) \models \varphi$;
- $((S, T), s) \models AF\varphi$ (resp. $((S, T), s) \models EF\varphi$) iff for every (resp. there exists a) path $(s_0, s_1, \ldots, s_n, \ldots)$, there exists $i \in \mathbb{N}$, $((S, T), s_i) \models \varphi$;
- $((S, T), s) \models A[\varphi \ U \ \psi]$ (resp. $((S, T), s) \models E[\varphi \ U \ \psi]$) iff for every (resp. there exists a) path $(s_0, s_1, \ldots, s_n, \ldots)$, there exists $i \in \mathbb{N}$ such that $((S, T), s_i) \models \psi$ and for every $j < i$, $((S, T), s_j) \models \varphi$;
- Boolean connectives are handled as usual.

In the sequel, to prove the preservation of properties through the embedding of biological regulatory networks, we will use a standard equivalence relation on the states of transition systems, the so-called *divergence blind stuttering equivalence* (dbs), which have been proved to preserve CTL-X formulas, i.e. the transition system and its quotient, with respect to the dbs equivalence relation, are elementary equivalent [10].

Let us recall the definition of a dbs relation R on a transition system (S, T).

A binary relation R on S is called a *divergence blind stuttering* (dbs) relation if, and only if it is symmetric and

$$r \ R \ s \Longleftrightarrow \begin{cases} L(r) = L(s) \\ (r, r') \in T \Rightarrow \exists s_0, s_1, \ldots, s_n \text{ finite path }, n \geq 0, (s_0 = s) \\ \qquad \wedge (\forall i < n, r \ R \ s_i) \wedge r' \ R \ s_n \end{cases}$$

It is obvious to show that every dbs relation is transitive. Moreover, as the case $n = 0$ is allowed in the second condition, the empty relation is a dbs relation. Finally, the diagonale relation on S is also a dbs relation, and it is easy to show that dbs relations are closed under union. Hence, the largest dbs relation exists and is an equivalence relation noted $\simeq_{dbs}$.

Given a transition system (S, T), its quotient by $\simeq_{dbs}$, denoted $(S, T)_{/\simeq_{dbs}}$, is defined by:

- the set of states $S_{/\simeq_{dbs}}$ is the set of equivalence classes of $\simeq_{dbs}$, $[s]$ denoting the equivalence class of s for s state of S
- the set of transitions $T_{/\simeq_{dbs}}$ defined by $([s], [t]) \in T_{/\simeq_{dbs}}$ iff there exists $s' \in [s]$ and $t' \in [t]$ such that $(s, t) \in T$
- $(S, T)_{/\simeq_{dbs}}$ is provided with the labeling function $L_{/\simeq_{dbs}}$ defined by $L_{/\simeq_{dbs}}([s]) = L(s)$

3   BRN logic

In this section, we will present the multivalued discrete approach developed by R. Thomas [13] for genetic regulatory networks as a logic built over the logic CTL-X. We will follow the standard approach for presenting a logic, *i.e.* syntax (signatures and formulas) and semantics (models and the satisfaction relation).

3.1   Syntax

Signatures. A biological regulatory network is represented by a labeled directed graph, called interaction graph. Vertices abstract biological entities, as genes or proteins, and will be called *variables*. Edges abstract interactions between variables. When a variable i activates a variable j, variable i can act positively on j, then there exists an edge from i to j labeled by the sign "+". On the contrary, when a variable i inhibits a variable j, variable i can act negatively on j, then there exists an edge from i to j labeled by the sign "-". Moreover, the action, activation or inhibition, between two variables becomes efficient only when the level of concentration of the regulator reaches a given threshold. In the discrete modeling framework of R. Thomas, the concentration levels for the variable i can take a finite number of values $\{0, 1, \ldots, b_i\}$ and thresholds related to the actions of i are numbered from 1 to b_i: the action of i on j is triggered only if the concentration of i crosses its concentration level. Thus, each interaction $i \longrightarrow j$ is labeled by a sign and a threshold. The knowledge of interactions between variables, including signs and thresholds, is called the static part of BRNs and constitutes the elements of signatures for a logic dedicated to BRNs.

Definition 1 (Signature) *A BRN-signature is a labeled directed graph $G = \langle V, F, Sn, Th \rangle$ where:*

1. *V is a finite set whose the elements are called* variables.
2. *$F \subseteq V \times V$ denotes the set of edges.*

3. Sn is a mapping from F to $\{+,-\}$.

4. Th is a mapping from F to $\mathbb{N}^*$ such that:

$$\forall (i,j) \in F \; \forall l \in \mathbb{N}^* \; \exists k \in V \; (Th(i,j) = l \wedge l \neq 1 \Rightarrow (i,k) \in F \wedge Th(i,k) = l-1)$$

Point 4. gives some restrictions on the way the edges are labeled. If an edge outgoing from a variable i is labeled by $l \geq 2$, then there exist edges outgoing from i labeled by $1, \ldots, l-1$. This well represents the qualitative nature of thresholds in BRN as used in this paper.

Notation 1 *Let $G = \langle V, F, Sn, Th \rangle$ be a BRN-signature and i be a variable in V. G_i^+, resp. G_i^-, denotes the set of successors, resp. predecessors, of i in G, and b_i denotes the cardinal of the set of thresholds for i. Formally, we have:*

$$G_i^+ = \{j \in V | (i,j) \in F\}$$

$$G_i^- = \{j \in V | (j,i) \in F\}$$

$$b_i = |\{l \in \mathbb{N}^* \mid \exists j \in G_i^+, Th(i,j) = l\}|$$

Example 1 *To illustrate Definition 1, we take as running example a model inspired from the one of control of immunity in temperate bacteriophage lambda. This model, proposed by Thieffry and Thomas in [11], contains genes cI and cro: cI inhibits cro and activates its own synthesis whereas the variable cro inhibits the expression of both variables, see Figure 1. The associated BRN-signature, denoted G_1 in the sequel, is simply given by:*

$\langle \; \{cI, cro\}, \{(cI, cI), (cI, cro), (cro, cI), (cro, cro)\},$
$\quad Sn : \{(cI, cI) \mapsto +, (cI, cro) \mapsto -, (cro, cI) \mapsto -, (cro, cro) \mapsto -\},$
$\quad Th : \{(cI, cI) \mapsto 1, (cI, cro) \mapsto 1, (cro, cI) \mapsto 1, (cro, cro) \mapsto 2\} \; \rangle$
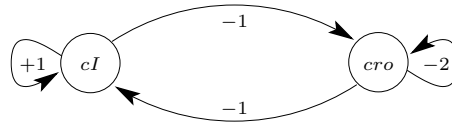
Fig. 1. Interaction graph for the $cI - cro$ system

Signature embedding. Biologists can identify small parts issued from a BRN involving a large number of genes. These parts are assimilated to a biological function insofar as it can be proven that the biological function is essentially related to the concentration levels of the variables occurring in the considered subpart.

Embedding of BRN signatures can formalize such an approach. However, signature embedding cannot be simple graph embeddings (which is defined by Conditions 1 and 2 of Definition 2 just below). Indeed, as well as preserving edge signs (see Condition 2), as the thresholds on edges depend on the properties of

the graph (a threshold cannot be greater than the number of outgoing edges), it matters to pay attention to the preservation of the conditions on the thresholds (Conditions 3 and 4). In fact, as thresholds are taken into consideration in signatures, the key point to carry through the embedding is the preservation of the equality between thresholds and the numerical order between them. New intermediate thresholds for a given variable can be introduced when including a BRN in another one, but relationships between existing thresholds have to be preserved in the larger one. Finally, a supplementary condition (Condition 5) has to be added. This condition means the preservation of predecessors in interaction graphs. This condition can seem very restrictive. However, it is useful to ensure the preservation of properties inherited from the small BRN to the large BRN (see the counter-example given in Section 5) which makes fail the preservation when Condition 5 does not hold. This leads to the following definition:

Definition 2 (Signature embedding) *Let G and G' be BRN-signatures such that $G = \langle V, F, Sn, Th \rangle$ and $G' = \langle V', F', Sn', Th' \rangle$. A signature embedding $G \to G'$ is an injective mapping $\sigma : V \to V'$ such that:*

1. $\forall i, j \in V, (i, j) \in F \Leftrightarrow (\sigma(i), \sigma(j)) \in F'$
2. $\forall i, j \in V, (i, j) \in F, Sn(i, j) = Sn'(\sigma(i), \sigma(j))$
3. $\forall i \in V, \forall j, k \in G_i^+, Th(i, j) = Th(i, k) \Leftrightarrow Th'(\sigma(i), \sigma(j)) = Th'(\sigma(i), \sigma(k))$
4. $\forall i \in V, \forall j, k \in G_i^+, Th(i, j) < Th(i, k) \Leftrightarrow Th'(\sigma(i), \sigma(j)) < Th'(\sigma(i), \sigma(k))$
5. $\forall j \in V, \forall k' \in V', (k', \sigma(j)) \in F' \Rightarrow \exists i \in V, (i, j) \in F \wedge \sigma(i) = k'$

Notation 2 *Let $\sigma : G \to G'$ be a signature embedding where $G = \langle V, F, Sn, Th \rangle$ and $G' = \langle V', F', Sn', Th' \rangle$ and let ω a set of variables in V, $\sigma(\omega)$ denotes the set $\{\sigma(i) \mid i \in \omega\}$.*

Example 2 *Figure 2 presents the BRN-signature G_2, sharing with G_1 both variables cI and cro, and containing a new variable N. According to Definition 1, a signature embedding σ_{id} between $\{cI, cro\}$ and $\{cI, cro, N\}$ can be defined: $\sigma_{id}(cI) = cI$ and $\sigma_{id}(cro) = cro$. Conditions 1 and 2 are clearly verified (all edges of G_1 are in G_2 labeled with the same sign). Condition 3 requires that the equality between thresholds for outgoing edges in G_1 is preserved in G_2, it is verified since only $Th(cI, cI) = Th(cI, cro)$ in G_1 and we have $Th'(cI, cI) = Th'(cI, cro)$ in G_2. Condition 4, which requires that the order between thresholds for outgoing edges in G_1 is preserved in G_2, is also verified. For instance, in G_1, cro has two outgoing edges (cro, cI) and (cro, cro) with $Th(cro, cI) < Th(cro, cro)$. In G_2, we have $Th'(cro, cI) < Th'(cro, cro)$. Condition 5 is also verified (cI and cro in G_2 have no new predecessors with respect to G_1).*

Roughly speaking, we can link two BRN-signatures by a signature embedding when the addition of new variables has only the effect of shifting the thresholds issued from the inherited variables.

Formulas. Formulas for BRN are simply CTL-X formulas whose atomic formulas describe comparisons between a concentration level of a variable with some threshold values.
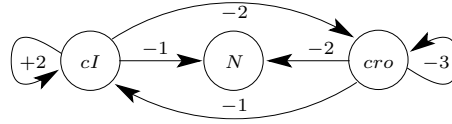
Fig. 2. BRN-signature G_2

Definition 3 (BRN Formulas) *Let $G = \langle V, F, Sn, Th \rangle$ be a BRN-signature. Formulas over G are CTL-X formulas whose atomic formulas are of the form $(i \sim l)$ where $i \in V$, $l \in \{0, \ldots, b_i\}$ and $\sim \in \{=, <, >\}$.*

We denote by $Atom(G)$ the set of atomic formulas built on G and by $Sen(G)$ the set of formulas over G.

In the sequel, $i \geq l$ (resp. $i \leq l$) will denote the formula $i = l \vee i > l$ (resp. $i = l \vee i < l$).

Signature embeddings obviously rename variables and thresholds occurring in atomic formulas. However, the threshold renaming is not so simple. Indeed, the presence of new variables makes side effects on the thresholds by shifting them. This gives rise to the following definition:

Definition 4 (Formula renaming) *Let $\sigma : G \to G'$ be a signature embedding with $G = \langle V, F, Sn, Th \rangle$ and $G' = \langle V', F', Sn', Th' \rangle$. For all $i \in V$, let us note $\sigma_i : \{0, 1, \ldots, b_i\} \to \{0, 1, \ldots, b_{\sigma(i)}\}$ the mapping defined by :*

- *$\sigma_i(0) = 0$*
- *For all $l \neq 0$, $\sigma_i(l) = Th'(\sigma(i), \sigma(j))$ with j any arbitrary variable such that $j \in G_i^+$ and $Th(i, j) = l$*

Let us note $\overline{\sigma} : Atom(G) \to Sen(G')$ the mapping defined by:

- *For all $(i = l) \in Atom(G)$ with $l \neq b_i$, $\overline{\sigma}(i = l) = \sigma(i) \geq \sigma_i(l) \wedge \sigma(i) < \sigma_i(l+1)$*
- *For all $(i = b_i) \in Atom(G)$, $\overline{\sigma}(i = b_i) = \sigma(i) \geq \sigma_i(b_i)$*
- *For all $(i > l) \in Atom(G)$, $\overline{\sigma}(i > l) = \sigma(i) \geq \sigma_i(l+1)$*
- *For all $(i < l) \in Atom(G)$, $\overline{\sigma}(i < l) = \sigma(i) < \sigma_i(l)$*

Let us note $\sigma^\sharp$ the canonical extension of the signature embedding σ on formulas in $Sen(G)$ defined as follows:

- *For $p \in Atom(G)$, $\sigma^\sharp(p) = \overline{\sigma}(p)$,*
- *For other formulas, Boolean connectives and temporal operators are preserved.*

The definition explains how to convert formulas in $Sen(G)$ into formulas in $Sen(G')$ by following the simple idea of translating a threshold into an interval of possible values.

3.2   Semantics

Models. Each variable i in a BRN-signature G is a genetic entity which is characterized at a given point in time by a concentration level. Dealing with regulatory networks with thresholds whose the set of nodes is finite, the state space generated from G is finite and defined by:

Definition 5 (State) *Let $G = \langle V, F, Sn, Th \rangle$ be a BRN-signature. The state space S_G of G is the set of mappings $s : V \to \mathbb{N}$ such that for every $i \in V$, $s(i) \in \{0, \ldots, b_i\}$.*

Example 3 *In the BRN-signature $G1$ of Example 1, Variables cI and cro have, respectively 2 and 3 possible concentration levels: 0 or 1, and 0, 1 or 2. Therefore, The state space for G_1 is $S_{G_1} = \{(0,0), (0,1), (0,2), (1,0), (1,1), (1,2)\}$.*

The concentration level of each variable $i \in V$ of a given BRN-signature G, evolves over time depending on the concentration level of its resources (*i.e.* sets of i's predecessors in G which have reached a concentration level to affect i's one by making it increase or decrease). However, neither G nor the concentration level of i's resources gives clues to decide the concentration level that i can reach. This is a degree of freedom of BRN-signatures which gives rise to a class of possible G-models, so-called dynamics of G. All these possible G-models do not correspond to actual biological functions. This is by biological knowledge described by CTL-X properties that we can cut down in the class of all possible G-models. Formally, G-models are defined as follows:

Definition 6 (Resources) *Let G be a BRN-signature. The set of resources $R_{G,i}(s)$ of a variable i at the state $s \in S_G$ is defined by:*

$$R_{G,i}(s) = \begin{cases} \{j \in G_i^- | (Sn(j,i) = +\ and\ s(j) \geq Th(j,i))\} \\ \qquad\qquad \cup \\ \{j \in G_i^- | (Sn(j,i) = -\ and\ s(j) < Th(j,i))\} \end{cases}$$

Hence, a resource is the presence of an activator or the absence of an inhibitor.

Example 4 *Figure 3 gives the sets of resources for the three variables cI, cro and N in S_{G_1} and S_{G_2}.*

Definition 7 (G-models) *Let $G = \langle V, F, Sn, Th \rangle$ be a BRN-signature and let $\kappa = \{(i,w) \mid i \in V \wedge w \subseteq G_i^-\}$ be the set of all subsets of predecessors in G for every variable $i \in V$. A G-model is a mapping $p : \kappa \to \mathbb{N}$ such that: $\forall (i,w) \in \kappa,\ p((i,w)) \in \{0, \ldots, b_i\}$.*

Example 5 *From the BRN-signature G_2 of Figure 2, we have the following set κ:*

$$\kappa = \begin{cases} \{(cI, \emptyset), (cI, \{cI\}), (cI, \{cro\}), (cI, \{cI, cro\})\} \\ \qquad\qquad \cup \\ \{(cro, \emptyset), (cro, \{cI\}), (cro, \{cro\}), (cro, \{cI, cro\})\} \\ \qquad\qquad \cup \\ \{(N, \emptyset), (N, \{cI\}), (N, \{cro\}), (N, \{cI, cro\})\} \end{cases}$$

cI	cro	N	$R_{G',cI}$	$R_{G',cro}$	$R_{G',N}$
0	0	0	$\{cro\}$	$\{cI,cro\}$	$\{cI,cro\}$
0	1	0	$\emptyset$	$\{cI,cro\}$	$\{cI,cro\}$
0	2	0	$\emptyset$	$\{cI,cro\}$	$\{cI\}$
0	3	0	$\emptyset$	$\{cI\}$	$\{cI\}$
1	0	0	$\{cro\}$	$\{cI,cro\}$	$\{cro\}$
1	1	0	$\emptyset$	$\{cI,cro\}$	$\{cro\}$
1	2	0	$\emptyset$	$\{cI,cro\}$	$\emptyset$
1	3	0	$\emptyset$	$\{cI\}$	$\emptyset$
2	0	0	$\{cI,cro\}$	$\{cro\}$	$\{cro\}$
2	1	0	$\{cI\}$	$\{cro\}$	$\{cro\}$
2	2	0	$\{cI\}$	$\{cro\}$	$\emptyset$
2	3	0	$\{cI\}$	$\emptyset$	$\emptyset$

cI	cro	$R_{G,cI}$	$R_{G,cro}$
0	0	$\{cro\}$	$\{cI,cro\}$
0	1	$\emptyset$	$\{cI,cro\}$
0	2	$\emptyset$	$\{cI\}$
1	0	$\{cI,cro\}$	$\{cro\}$
1	1	$\{cI\}$	$\{cro\}$
1	2	$\{cI\}$	$\emptyset$

Fig. 3. Resources of cI, cro and N in S_{G_1} (left) and in S_{G_2} (right)

From the value of the concentration levels for cI, cro and N, a possible G_2-model p_2 is given in Figure 4 (left).

Signature embeddings $\sigma : G \to G'$ have a counterpart on models which is expressed by a classic forgetful mapping. Here also, some difficulties occur due to some restrictions to make on thresholds from the "richer" model defined on G' to the "poorer" one defined on G. This then leads to the following definition:

Definition 8 (Reduced model) *Given a signature embedding $\sigma : G \to G'$ and a G'-model p', the reduced G-model p from p', denoted $p'_{|\sigma}$, is defined as follows: $\forall (i,w) \in \kappa$,*

$$p((i,w)) = \begin{cases} Th(i,j) \ \textit{if it exists } j \textit{ in } V \textit{ such that} \\ \qquad Th'(\sigma(i),\sigma(j)) = max_{(i,k)\in F}\{Th'(\sigma(i),\sigma(k)) \mid \\ \qquad\qquad\qquad Th'(\sigma(i),\sigma(k)) \leq p'((\sigma(i),\sigma(w)))\} \\ 0 \qquad \textit{otherwise} \end{cases}$$

Example 6 *Figure 4 (right) gives the reduced G_1-model p_1 of p_2 along the signature embedding given in Example 2.*

resource ω'	$p_2((cI,\omega'))$	$p_2((cro,\omega'))$	$p_2((N,\omega'))$
$\emptyset$	0	0	0
$\{cI\}$	2	2	0
$\{cro\}$	2	1	0
$\{cI,cro\}$	2	3	0

resource ω	$p_1((cI,\omega))$	$p_1((cro,\omega))$
$\emptyset$	0	0
$\{cI\}$	1	1
$\{cro\}$	1	1
$\{cI,cro\}$	1	2

Fig. 4. A G_2-model p_2 (left) and its reduced G_1-model p_1 (right)

From a G-model p, a transition system (S_G, T) can be generated where the transitions in T give the state evolution as described in p. Here, two possibilities can occur. We make evolve either many variables directly to their concentration level specified by p, or one variable i and only by one unit in the direction of $p((i, \omega))$ where ω is the set of resources of i at the current state. These two possibilities are respectively called *synchronous* and *asynchronous* description of the G-model p. Here, we follow the asynchronous description because in the nature, it is unlikely that, *in vivo*, several variables cross a threshold simultaneously [12].
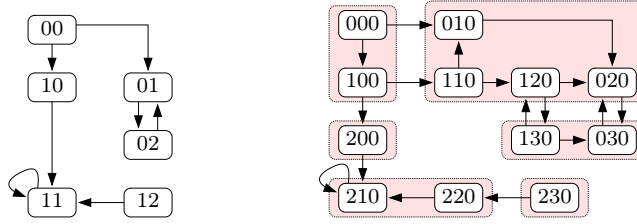
Definition 9 (Asynchronous transition system) *Let $G = \langle V, F, Sn, Th \rangle$ be a BRN-signature and let p be a G-model. The* asynchronous transition system *generated from p is a directed graph $GTA((G, p)) = (S_G, T)$ such that:*

- $\forall s \in S_G, (s, s) \in T \Leftrightarrow \forall i \in V, s(i) = p((i, R_{G,i}(s)))$
- $\forall s \neq s' \in S_G, (s, s') \in T$ *if, and only if:*
  - *there exists $i \in V$, such that*

$$s'(i) = \begin{cases} s(i) + 1 \text{ and } s(i) < p((i, R_{G,i}(s))) \\ s(i) - 1 \text{ and } s(i) > p((i, R_{G,i}(s))) \end{cases}$$

  - *and $s'(j) = s(j)$ for every $j \in V \setminus \{i\}$.*

Example 7 *Figure 5 gives from the left to the right, the asynchronous transition systems $GTA((G_1, p_1))$ and $GTA((G_2, p_2))$ generated from p_1 and p_2.*

Fig. 5. Asynchronous transition systems $GTA((G_1, p_1))$ and $GTA((G_2, p_2))$. Colored boxes represent the $\simeq_{dbs}$ equivalence classes of $GTA((G_2, p_2))$ – see Section 4

Satisfaction relation. The asynchronous transition system (S_G, T) generated from a G-model p is a transition system following the definition in Section 2. However, to satisfy CTL formulas, we have to manipulate Kripke frames and then we need to precise the labeling function $L : S_G \to 2^{Atom(G)}$. Given a state s in S_G,

$$L(s) = \{i > l, i < l', i = l'' \mid i \in V, l \in \{0, 1, \ldots, b_i - 1\}, l' \in \{1, 2, \ldots, b_i\},$$

$$l'' \in \{0, 1, \ldots, b_i\}, s(i) > l, s(i) < l', s(i) = l''\}$$

Therefore, the satisfaction relation of a formula φ over a BRN-signature G for a G-model p is then defined by: $p \models \varphi \iff GTA((G, p)) \models \varphi$ following the definitions given in Section 2.

4   Property preservation along signature embeddings

In this section, we show that given a signature embedding $\sigma : G \to G'$ and a G'-model p', p' and $p'_{|_\sigma}$ are elementary equivalent on formulas in $Sen(G)$ up to σ. This is stated by the following result.

Theorem 1. *For every signature embedding $\sigma : G \to G'$, for every G'-model p' and for every formula $\varphi \in Sen(G)$,*

$$p' \models \sigma^\sharp(\varphi) \quad \Longleftrightarrow \quad p'_{|_\sigma} \models \varphi$$

Proof (Sketch). Let us consider a signature embedding $\sigma : G \to G'$, a G'-model p' for the BRN-signature G', its associated asynchronous transition system $(S_{G'}, T') = GTA((G', p'))$ and a formula $\varphi \in Sen(G)$. Let us note $(S_G, T) = GTA(G, p'_{|_\sigma})$. Start by defining the mapping $B : S_G \to 2^{S_{G'}}$ as follows: for every $s \in S_G$, $B(s)$ is the set of states s' in $S_{G'}$ verifying for every i in V:

- if $s(i) = b_i$, then
$$s'(\sigma(i)) \geq \sigma_i(b_i)$$

- else,
$$s'(\sigma(i)) \geq \sigma_i(s(i)) \land s'(\sigma(i)) < \sigma_i(s(i) + 1)$$

The proof of Theorem 1 rests on the following intermediate propositions. The proofs of these propositions can be found in [1].

Proposition 1 *The mapping B makes a partition of $S_{G'}$, i.e.*

1. $\forall s, s' \in S, \ B(s) \cap B(s') = \emptyset$, and
2. $\bigcup_{s \in S_G} B_s = S_{G'}$.

Note $P = \{B(s) | s \in S_G\}$. Then, we have:

Proposition 2 *P is a dbs equivalence.*

It then remains to prove:

Proposition 3 *$(S_{G'}, T')_{/\simeq_{dbs}}$ and (S_G, T) are isomorphic.*

It is well known that isomorphic models are elementary equivalent (i.e. they satisfy the same set of properties). Therefore, by applying the result of [10], we can conclude that $GTA((G', p'))$ and $GTA((G, p))$ are elementary equivalent on formulas over G up to formula renaming resulting from σ.

Some readers will recognize the so-called satisfaction condition of the institution framework [8]. Hence, this BRN-logic based on signature embeddings is then an institution.

Example 8 *For the current example, the equivalence classes of $\simeq_{dbs}$ in G_2 are highlighted in Figure 5 by colored boxes.*
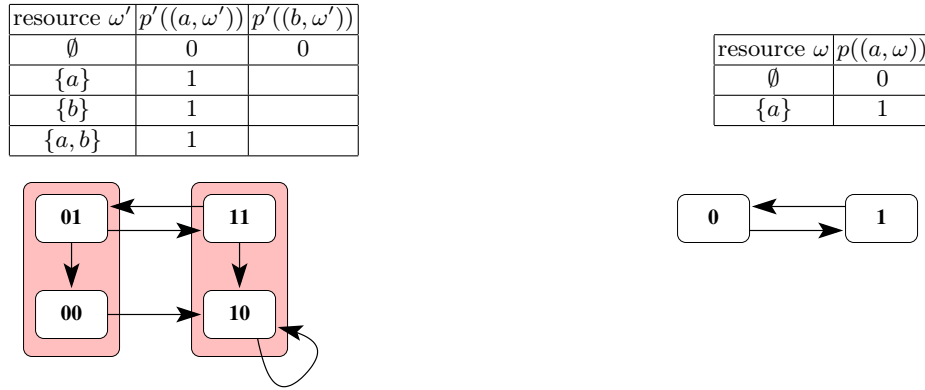
5 Counter-example justifying our restrictive notion of signature embeddings

In this section we give a counter-example to show the significance of Condition 5 (preservation of predecessors) of Definition 2. Let us consider both BRN-signatures G and G' of figure 6. It is possible to construct an injective mapping $\sigma : V \longrightarrow V'$ satisfying Conditions 1, 2, 3, and 4 of Definition 2 with $\sigma(a) = a$. For the signature G' we consider the model p' given in Figure 7 (left) from

Fig. 6. Counter-example: we consider an embedding not satisfying Condition 5 in Definition 2

which we deduce the reduced G-model $p = p'_{|\sigma}$ from p' (see Figure 7-right), and we consider the asynchronous transition systems generated from p and p'.

resource ω'	$p'((a, \omega'))$	$p'((b, \omega'))$
$\emptyset$	0	0
$\{a\}$	1	
$\{b\}$	1	
$\{a, b\}$	1	

resource ω	$p((a, \omega))$
$\emptyset$	0
$\{a\}$	1

Fig. 7. a G'-model p' and its reduced G-model p.

It is then easy to see that models p' and p do not satisfy the same formulas of CTL-X. For example the formula $AG(AF(a = 0))$ which means that the system will infinitely often pass through a state where $a = 0$ is true, is satisfied by p but not by p'.

6 Conclusion

We have presented the multivalued discrete approach for biological regulatory networks under the classical form of a logical formalism. BRN-signatures are

made of graphs, denoting the static part of BRN. Formulas are CTL-X formulas over atoms expressing comparisons between concentration levels of gene products with some abstract discrete values. Models are asynchronous transition systems deduced from the knowledge of parametrization explicating towards which concentration level tends a variable when it is under the influence of other ones. Lastly, the satisfaction relation is simply deduced from the one defined for the CTL-X formalism. In order to study how properties expressed on a small BRN are preserved or not when embedding it within a larger one, we have equipped our BRN formalism with signature embeddings. Their main particularity is that they capture the fact that a concentration level or threshold relative to a network is converted into an interval of concentration levels. We have proved that CTL-X properties are preserved along such signature embeddings.

We plan to pursue our work by investigating some other conditions which will allow us to go further in the property preservation while keeping a meaning for biological experts. In particular, we are currently investigating under which weaker embedding condition, a weaker set of CTL properties can be still preserved.

References

1. M. Aiguier, P. Le Gall, and M. Mabrouki, *A formal denotation of complex systems: how to use algebraic refinement to deal with complexity of systems*, Tech. report, http://www.epigenomique.genopole.fr/~aiguier, 2008.
2. G. Bernot, J.-P. Comet, A. Richard, and J. Guespin, *Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic*, Journal of Theoretical Biology **229** (2004), no. 3, 339–347.
3. H. de Jong, *Modeling and simulation of genetic regulatory systems: a literature review.*, J. Comput. Biol. **9** (2002), no. 1, 67–103.
4. H. de Jong, J. Geiselmann, C. Hernandez, and M. Page, *Genetic network analyzer: qualitative simulation of genetic regulatory networks.*, Bioinformatics **19** (2003), no. 3, 336–44.
5. H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann, *Qualitative simulation of genetic regulatory networks using piecewise-linear models*, Bulletin of Mathematical Biology **66** (2004), no. 2, 301–340.
6. E.A. Emerson, *Handbook of theoretical computer science, volume b : formal models and semantics*, ch. Temporal and modal logic, pp. 995–1072, MIT Press, 1990.
7. L. Glass and S.A. Kauffman, *The logical analysis of continuous non-linear biochemical control networks*, J. Theor. Biol. **39** (1973), 103–129.
8. J.A. Goguen and R.-M. Burstall, *Institutions: Abstract model theory for specification and programming*, Journal of the ACM **39** (1992), no. 1, 95–146.
9. D. Mateus, J.-P. Gallois, J.-P. Comet, and P. Le Gall, *Symbolic modeling of genetic regulatory networks*, Journal of Bioinformatics and Computational Biology (2007).
10. Rocco De Nicola and Frits Vaandrager, *Three logics for branching bisimulation*, J. ACM **42** (1995), no. 2, 458–487.
11. D. Thieffry and R Thomas, *Dynamical behaviour of biological regulatory networks - ii. immunity control in bacteriophage lambda.*, Bull. Math. Biol. **57** (1995), no. 2, 277–297.

12. R. Thomas, *Logical analysis of systems comprising feedback loops.*, J. Theor. Biol. **73** (1978), no. 4, 631–56.
13. R. Thomas and R. d'Ari, *Biological feedback*, CRC Press, 1990.
14. R. van Glabbeck and W.P. Weijland, *Refinement in branching time semantics.*, Proc. IFIP Conference, 1989, pp. 613–618.
15. Heike Wehrheim, *Inheritance of temporal logic properties*, FMOODS, 2003, pp. 79–93.