Formal methods for modelling biological regulatory networks

Adrien Richard, Jean-Paul Comet and Gilles Bernot

La.M.I UMR 8042, CNRS & Université d'Évry
Boulevard François Mitterrand, 91025 Évry Cedex-France
{arichard,comet,bernot}@lami.univ-evry.fr

Abstract:

This chapter presents how the formal methods can be used to analyse biological regulatory networks which are at the core of all biological phenomena as, for example, cell differentiation or temperature control. The dynamics of such a system, *i.e.* its semantics, is often described by an ordinary differential equation system, but has also been abstracted into a discrete formalism due to R. Thomas. This second description is well adapted to state-of-the-art measurement techniques in biology, which often provide qualitative and coarse-grained descriptions of biological regulatory networks. This formalism permits us to design a formal framework for analysing the dynamics of biological systems. The verification tools, as model checking, can then be used not only to verify if the modelling is coherent with known biological properties, but also to help biologists in the modelling process. Actually, for a given biological regulatory network, a large class of semantics can be automatically built and model checking allows the selection of the semantics which are coherent with the biological requirement, *i.e.* the temporal specification. This modelling process is illustrated with the well studied genetic regulatory network controlling immunity in bacteriophage lambda.

Table of Contents

1 Introduction

Biological systems are one of the most fascinating aspects in biology. They control such diverse dynamics phenomena as temperature control in warm-blooded animals; differentiation of a zygote into the various specialized organs, tissues and cells of the mature organism; the fate of certain viruses, called temperate bacteriophages, which upon infection of a bacterial population can behave in two extremely different ways. Most of these infected cells display a response called *lytic*: virus multiplies and kills cells. But, a fraction of the cells become *lysogenic bacteria* and carry the viral genome in a dormant form making the host immune towards infection of other virus.

Computational systems biology tries to establish methods and techniques that enable us to understand such systems as systems, including their robustness, design and manipulation. It means to understand : the structures and the dynamics of systems, methods to control, design and modify systems to cope with desired properties. The modelling contributes in a major way to reach these aims by introducing methods for understanding, simulating and predicting the behaviour of the systems. However, the modelling of biological systems is currently subject to two major difficulties: the biochemical reaction mechanisms underlying the interactions of systems are usually not or incompletely known and quantitative information on kinetic parameters or molecular concentration is rarely available. Thus the modelling activity needs an interaction with the experimental biology to confront models to biological objects. Consequently as in the design of large computing systems, two activities can be distinguished in the modelling step:

1. build a rigorous model of the system satisfying the assumed behaviour corresponding to biological knowledge,
2. design experiments to verify *a posteriori* the model predictions.

Here we would like to show that some methods from computer science can be reused in the context of system biology, as, for example, formal methods for validation and verification used for the design of large computing systems. For designing experiments, we just mention that the test methods *via* test generation from model theories may be an efficient way to propose experiments permitting biologists to validate or refute models. For building a rigorous model, the model checking verification tool is particularly suited. In this chapter we present an application of this formal method to build qualitative models of biological regulatory networks.

A biological regulatory network describes interactions between the biological entities, often macromolecules or genes, of a given system. It is statically represented by an interaction graph whose vertices abstract biological entities and arcs their interactions. For describing the evolution of the system, the concentration level of each entity is represented by a value associated to the corresponding vertex. The temporal evolution of these levels constitutes the dynamics of the system.

Ordinary differential equation systems have been first used for describing the dynamics of networks. They are powerful tools particularly to model metabolic processes [33]. However, due to the non-linearity of biological regulations, these differential equation systems cannot often be solved analytically. They can be solved numerically to any desired precision, but this precision itself may be misleading because the values of the parameters and the shape of interactions often have to be guessed for lack of experimental data. This remark led Thomas to simplify the models: he introduced in the 70's a Boolean approach to capture the qualitative nature of the dynamics and he proved its usefulness in the context of immunity in

bacteriophages [30, 27]. Later on, he generalized his formalism to multi-valued levels of concentration (the so called multi-valued logic or ``generalized logical approach" [32]) since the Boolean idealization may be too caricatured to correctly model biological systems. It has been proved that this qualitative description allows the representation of the essential qualitative features of an ordinary differential equation system provided that the differential equations are piece-wise linear [22]. The underlying parameters of the qualitative description can be deduced from the kinetic parameters of the continuous system but can take only a finite number of values. Consequently, all possible qualitative features of the system can be reduced to a finite number of models *i.e.* parameterisations.

Certainly the most important concepts of the generalized logical analysis are those of positive and negative feedback circuits. If an entity tends to favour (resp. decrease) its own production *via* the feedback circuit, the circuit is said positive (resp. negative). It has been conjectured by Thomas [28] and then proved in different contexts [18, 23, 6, 5, 25] that at least one positive circuit is necessary to generate multi-stationarity whereas at least one negative circuit is necessary to obtain a stable oscillatory behaviour. These concepts are especially important since when modelling biological systems, differentiation and homeostasis have often to be taken into consideration. In such cases, these biological constraints can reduce drastically the set of models to consider. These properties can be reinforced by introducing some more complex properties on the dynamics of the system extracted from the biological knowledge or hypotheses. It becomes necessary to construct models which are coherent not only with the previous conditions of multi-stationarity/homeostatis but also with the additional ones. Formal methods from computer science should be able to help modeller to automatically perform this verification [17, 3] and to select exhaustively all suitable models.

The chapter is organized as follows. Section 2 introduces the formalism due to Thomas for modelling the dynamics of a biological regulatory network. The resulting dynamics corresponds to a Kripke structure which can be deduced easily from the interaction graph. Section 3 describes the link between this transition system and the dynamics obtained with the classical modelling using piece-wise linear ordinary differential equation systems. Section 4 explains how formal methods can improve the modelling process of regulatory networks. The temporal properties have first to be translated into a temporal specification language. Then one has to answer automatically the question: does a given model satisfy the given temporal specifications ? Model checking makes this stage automatic and its principle is also presented in this section. Section 5 illustrates the use of model checking to model the well studied genetic regulatory network of temperate bacteriophage lambda rapidly described before. A model of this system has already been proposed by Thieffry and Thomas in [26]. We show that our approach, using model checking, automatically selects this model as well as other models satisfying the same criteria of validation.

2 Qualitative dynamics of biological regulatory networks

The multi-valued modelling of Thomas is able to represent the qualitative dynamics of biological regulatory networks whose entities can be molecules, macromolecules, cells, organs, or organisms, if no societies. In fact, all systems whose regulations have a sigmoid shape can be modelled in this formalism. The regulations of genetic regulatory networks have almost always a sigmoidal nature, that explains why this formalism has been introduced in this context and why its main application domain remains the genetic regulatory networks. In

such systems, the concentration of a protein encoded by a gene *u* may activate or inhibit the synthesis of proteins encoded by other genes or itself (figure 1).

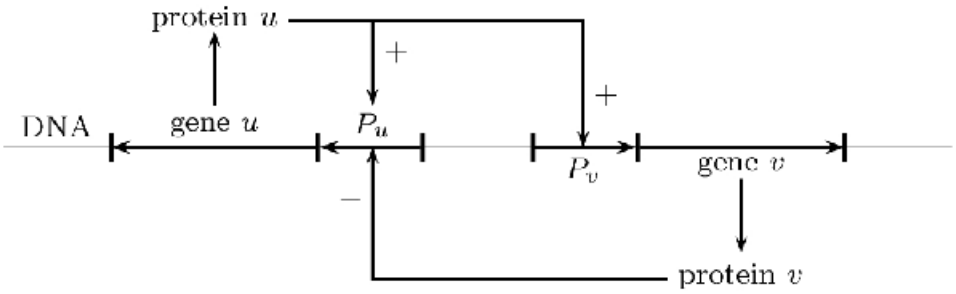

Figure 1: A genetic regulatory network. The gene *u* synthesizes a protein which activates the expression of gene *v* and itself by binding the promoters P_v and P_u respectively. In turn, the protein of gene *v* inhibits the expression of gene *u* when it binds P_u. Then, the arrow from a gene to its protein represents the transcription and translation processes and the arrow from a protein to a promoter abstracts the diffusion and the fixation of the protein on the promoter.

If the protein of *u* activates (resp. inhibits) the expression of a gene *v*, we said that *u* is a positive (resp. negative) regulator of *v*. In such situation an increasing of the concentration of the protein encoded by *u* induces an increasing of the rate of synthesis of the protein encoded by *v*. Generally, the relation between the concentration of a regulator and the rate of synthesis of its target is, as we have seen before, sigmoidal. When the sigmoid is steep, as in figure 2-(a), *u* has a little effect on *v* if it is below the concentration threshold θ_{uv} and at higher concentration a plateau is reached representing the maximal rate of synthesis of *v* under the effect of *u*. Naturally, for an negative regulator, the sigmoid is decreasing.

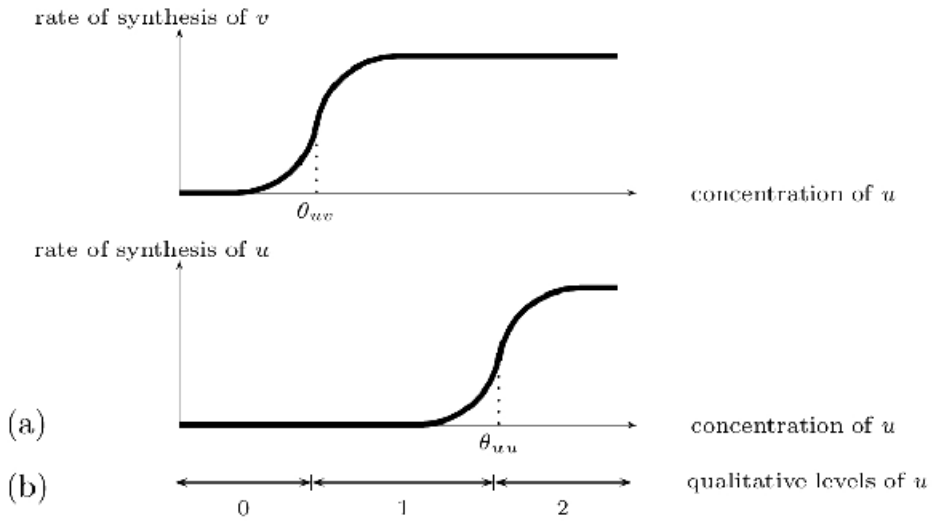

Figure 2: (a) Sigmoid relations between the concentration of *u* and the rate of synthesis of *v* and itself. As *u* is an activator of *v* and itself, see figure 1, both sigmoids are increasing. (b) Resulting qualitative levels of *u*.

This section presents successively how the regulations can be summarized into a regulatory graph corresponding to the static part of the modelling, then introduces the parameters which encode the effects of regulators on their targets in all possible situations, and finally presents how the dynamics can be deduced from these parameters.

2.1 Biological regulatory graphs

To formally define the static part of biological regulatory networks, we use labelled directed graphs. Vertices represent the biological entities of the network and arcs their regulations.

Definition 1 [Biological regulatory graph] *A* biological regulatory graph *is a labelled directed graph G=(V,E) where*

- *each vertex v of V, called* variable*, is provided with a* boundary $b_v \in \mathbb{N}$ *less or equal to the out-degree of v in G.*
- *each arc* $u \to v$ *of E is labelled with a couple* (t_{uv},α_{uv}) *where* t_{uv} *is an integer between 1 and* b_v *, called* qualitative threshold *and where* $\alpha_{uv} \in \{+,-\}$ *is the sign of the regulation.*

Moreover it is required that for any variable u with $b_u > 0$, $\forall\, i \in \{1,2,...,b_u\}$, *there exists a successor v of u such that* $t_{uv} = i$.

In a biological regulatory graph G, the set of the regulators of a variable v corresponds to the set of its predecessors in G, denoted by $G^-(v)$, and the set of its targets corresponds to the set of its successors in G, denoted by $G^+(v)$. Each regulation $u \to v$ is labelled by a sign α_{uv} which indicates if u is an activator or an inhibitor of v, and by a qualitative threshold t_{uv}. Thresholds t are integers and do not correspond to biological thresholds $\theta \in \mathbb{R}$, most often difficult to measure, but they give the order of the continuous thresholds: if $t_{uv} = i$ then the corresponding continuous threshold θ_{uv} is the i^{th} lowest threshold among $\{\theta_{uv} \mid v \in G^+(u)\}$. That explains the requirement on qualitative thresholds of the previous definition which implies that b_v is the number of different thresholds "outgoing" from v.
Figure 3-(a) gives an example of biological regulatory graph which can be deduced from the genetic regulatory networks described in figure 1. Figure 2 assumes that $\theta_{uv} < \theta_{uu}$, and consequently $t_{uv} = 1$ and $t_{uu} = 2$.

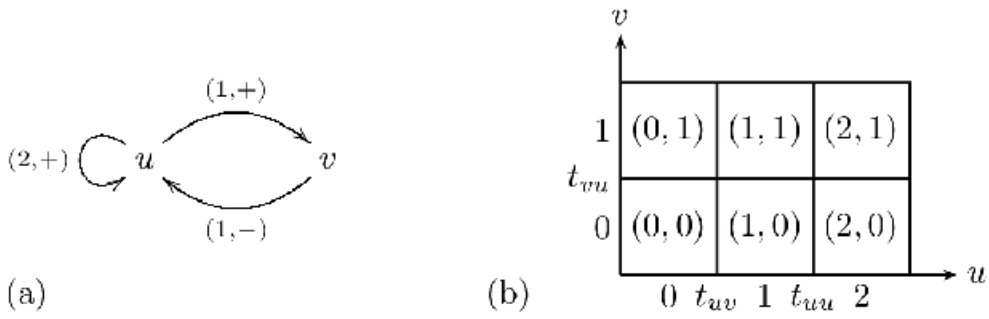

Figure 3: (a) Biological regulatory graph deduced from the genetic regulatory network of figure 1. (b) States of the biological regulatory graph.

Obviously concentration levels are associated to variables. For describing the evolution of the concentration level of each variable, it is necessary to know which regulators have an effect on the variable. Only the position of the regulator concentration with regard to their thresholds is sufficient. The concentrations are then discretized according to thresholds and each variable can take a finite number of values called abstract qualitative levels. For example, in figure 2-(b), the variable u has three different behaviours with regard to its targets:

- In the first region (the concentration of u is less than θ_{uv}), u acts neither on v nor on itself.
- In the second region (the concentration of u is between θ_{uv} and θ_{uu}), u acts on v but it still not act on itself.
- In the last region (the concentration of u is greater than θ_{uu}), u acts both on v and on itself.

Three qualitative levels emerge, 0, 1 and 2, corresponding to the three previous regions and constitute the only relevant information from a qualitative point of view. More generally, a variable v can take b_v+1 qualitative levels, from 0 to b_v, and the qualitative level q means that v acts on all targets v' such that $t_{vv'} \leq q$. A state of the system is then defined as a vector constituted by qualitative levels of variables.

Definition 2 [Qualitative state] *Let $G=(V,E)$ be a biological regulatory graph. A* qualitative state *of G is a vector $q=(q_v)_{v\in V}$ such that for all $v\in V$, $q_v\in\{0,1,\ldots,b_v\}$. The set Q of states of G is then defined by $Q=\prod_{v\in V}\{0,1,...,b_v\}$.*

In the sequel, we write $v = l$ for denoting $q_v = l$ if it does not cause confusion. Figure 3-(b) shows the possible states of the biological regulatory graph of figure 3-(a).

2.2 Models of biological regulatory graphs

Remember that the sigmoid nature of a regulation $u \to v$ leads to distinguish two different situations: if $u \geq t_{uv}$, then the regulation is active and if $u < t_{uv}$, it is not. We said that u is a *resource* of v when u induces an increasing of v:

- if the regulation $u\to v$ is positive, u is a resource of v when the regulation is active,
- if the regulation $u\to v$ is negative, u is a resource of v when the regulation is not active.

From the point of view of resources, the absence of an inhibitor acts as the presence of an activator.

Definition 3 [Resources] *Let $G=(V,E)$ be a biological regulatory graph, $v\in V$ and $q\in Q$. The set $\omega_v(q)$ of* resources *of v at the state q is the subset of $G^-(v)$ defined by:*

$$\omega_v(q)=\{u\in G^-(v) \mid (q_u \geq t_{uv} \text{ and } \alpha_{uv} = +) \text{ or } (q_u < t_{uv} \text{ and } \alpha_{uv} = -)\}.$$

At the state q, the evolution of variable v depends on its resources $\omega_v(q)$. It remains to define in which direction evolves v at the state q. The parameter $K_{v,\omega_v(q)}$, called the attractor of v when the resources are $\omega_v(q)$, denotes the level towards which v is attracted:

- if $q_v < K_{v,\omega_v(q)}$ then v tends to increase,

- if $q_v = K_{v,\omega_v(q)}$ then v does not evolve and
- if $q_v > K_{v,\omega_v(q)}$ then of v tends to decrease.

Different values for these parameters are possible and we call *model* of a biological regulatory graph, a possible instantiation of these parameters.

Definition 4 [Model of a biological regulatory graph] *Let $G=(V,E)$ be a regulatory graph. A* model *of G, denoted $M(G)$ by abuse of notation, is a family of natural numbers $K_{v,\omega}$ indexed by the set of couples (v,ω) such that*

- $v \in V$,
- $\omega \subseteq G^-(v)$,
- $K_{v,\omega} \leq b_v$.

It is often additionally required that:

$$K_{v,\omega} \leq K_{v,\omega'} \text{ for all } v \in V, \text{ and for all } \omega,\omega' \subseteq G^-(v) \text{ such that } \omega \subseteq \omega'. \quad (1)$$

These constraints, called Snoussi's constraints in the remainder, mean that the more a variable has resources the greater is the level towards which it is attracted. In other words, neither the presence of an activator nor the absence of an inhibitor can induce a decrease of the considered target (see the following section for the mathematical grounds of these constraints). This property, as well as signs of regulations, can often be deduced from biological knowledge and when it can be used, the number of models to consider for a given biological regulatory graph decreases drastically.

Model		u	v	Attractors				Tendencies	
$K_{u,\{\}}$	$= 0$	0	0	$K_{u,\{v\}}$	$= 2$	$K_{v,\{\}}$	$= 0$	↗	→
$K_{u,\{u\}}$	$= 2$	0	1	$K_{u,\{\}}$	$= 0$	$K_{v,\{\}}$	$= 0$	→	↘
$K_{u,\{v\}}$	$= 0$	1	0	$K_{u,\{v\}}$	$= 2$	$K_{v,\{u\}}$	$= 1$	↗	↗
$K_{u,\{u,v\}}$	$= 2$	1	1	$K_{u,\{\}}$	$= 0$	$K_{v,\{u\}}$	$= 1$	↗	→
$K_{v,\{\}}$	$= 0$	2	0	$K_{u,\{u,v\}}$	$= 2$	$K_{v,\{u\}}$	$= 1$	→	↗
$K_{v,\{u\}}$	$= 1$	2	1	$K_{u,\{u\}}$	$= 2$	$K_{v,\{u\}}$	$= 1$	→	→

Table 1: One possible model for the biological regulatory graph of figure 3. The table gives for each state the corresponding attractors and tendencies deduced from the model.

Models $M(G)$ of the biological regulatory graph G of figure 3 are all possible instantiations of six parameters: $K_{u,\{\}}$, $K_{u,\{v\}}$, $K_{v,\{\}}$, $K_{v,\{u\}}$, $K_{v,\{v\}}$, $K_{v,\{u,v\}}$. Because $b_u = 1$ (resp. $b_v = 2$) each $K_{u,\ldots}$ (resp. $K_{v,\ldots}$) can take 2 (resp. 3) different values. So $2^2 \times 3^4 = 324$ different models can be *a priori* associated to G, but only 60 of them respect the Snoussi's constraints. Table 1 gives the tendencies of variables resulting from such a model.

More generally there are $\prod_{v \in V} (b_v+1)^{2^{|G^-(v)|}}$ models associated to a biological regulatory graph G. This number increases exponentially with the number of predecessors of each variable and even if static constraints on parameters are used, as the Snoussi's constraints, it

remains huge. Moreover, since parameters K are most often not measurable *in vivo*, additional properties deduced from biological experiments are needed to eliminate the models whose dynamics do not satisfy them.

2.3 Dynamics of models

The classical approach to describe the dynamics of models is to define the state of the system at time t +1 from its state at time t. One possibility is to consider that the next state is directly the attractor of the current state: if q is the current state then $q' = (K_{v,\omega_v(q)})_{v\in V}$ is the next one and we said that there is a transition from q to q' (figure 5-(a)). This description raises serious problems for its application to biological systems:

1. From any initial state, the system will follow a well-defined path, without any branching or possibility of choice whereas biological systems typically include choices among several pathways (as illustrated for example by the numerous different pathways leading to various cell lines from a zygote during embryonic development).
2. Suppose that v is a gene which can take two values ($b_v = 1$) and that the current state is q. If $q_v = 0$, then $K_{v,\omega_v(q)} = 1$ means that resources of v induce the production of the corresponding protein. This protein will appear after a time delay corresponding, for example, to the time of diffusion of its regulators (figure 4). Similarly the same phenomenon is observed when $q_v = 1$ and $K_{v,\omega_v(q)} = 0$ with an *a priori* different delay. However, when q differs from $q' = (K_{v,\omega_v(q)})_{v\in V}$ by at least two components, the corresponding variables change simultaneously (dashed arrow in figure 5-(a)). This synchronous description thus assumes that time delays are equal which is unlikely.

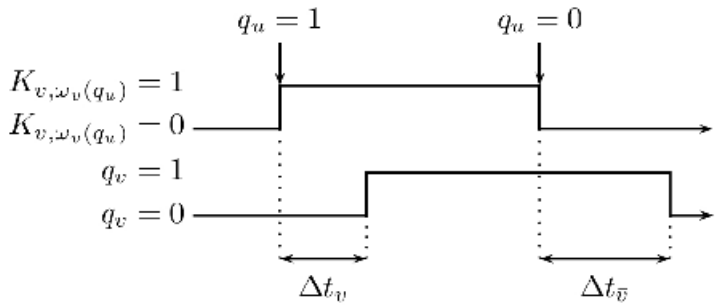

Figure 4: Time delays. Gene v has a unique regulator u which is an activator. Initially, both u and v are absent. Then, protein u appears and stimulates the expression of v ($K_{v,\omega_v(q_u)} = K_{v,\{u\}} = 1$). The resulting protein appears after the delay Δt_v. Finally, the protein u disappears, the gene v is no more stimulated ($K_{v,\omega_v(q_u)} = K_{v,\{\}} = 0$), and the protein v disappears after the different delay $\Delta t_{\bar{v}}$.

3. If the attractor of a variable is sufficiently away from its current value, one can have $|q_v - K_{v,\omega_v(q)}| > 1$. In such cases the qualitative level increases abruptly and jumps several thresholds (dotted arrow in figure 5-(a)). Since the dynamics of the model abstracts a continuous phenomenon, during a transition, each variable can pass through at most one threshold.

These points lead us to introduce the following asynchronous description.

Definition 5 [Asynchronous state graph] *Let G=(V,E) be a biological regulatory graph and M(G) be a model of G. The* asynchronous state graph *of M(G) is a directed graph whose set of vertices is the set Q of states of G, and such that there is an edge from q to q' if:*

- *for all variables* $v \in V$, $q_v = q'_v = K_{v,\omega_v(q)}$ *or*
- *there exist a variable* $v \in V$ *such that:*
  - *for any variable* $u \neq v$, $q_u = q'_u$ *and*
  - $q_v < K_{v,\omega_v(q)}$ *and* $q'_v = q_v + 1$ *or* $q_v > K_{v,\omega_v(q)}$ *and* $q'_v = q_v - 1$.

In this definition, a state q which has itself as successor, is a *stable steady state* of the asynchronous state graph: $q_v = K_{v,\omega_v(q)}$ for all $v \in V$. Otherwise, if q is a state for which n variables tend to evolve (n variables v such that $q_v \neq K_{v,\omega_v(q)}$), q has n successors and each of them differs from q by only one component corresponding to one of these n variables. Thus, when time delays are unknown, the asynchronous state graph contains all the *a priori* possible transitions. Some of them can be removed when time delays are taken into consideration.

Figure 5 shows the synchronous and asynchronous dynamics of the model of table 1. The attractors are the same in both descriptions but paths differ: the asynchronous state graph contains a circuit $(0,0) \to (1,0) \to (1,1) \to (0,1) \to (0,0)$ which is absent in the synchronous description.

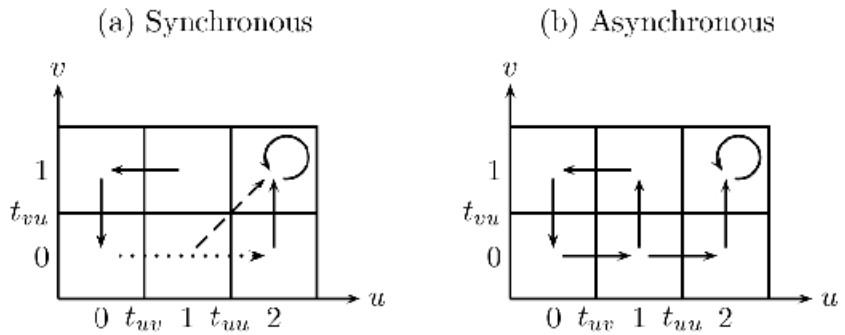

Figure 5: Synchronous and asynchronous dynamics for the model, given in table 1, of the biological regulatory graph of figure 3.

3 Differential modelling

We have seen that the asynchronous dynamics is more suited than the synchronous one for describing biological regulatory networks. This section proves how this asynchronous description can be deduced from a discretization of a particular class of ordinary differential equation systems classically used for describing biological regulatory networks.

3.1 Ordinary differential equation systems

Classically, the dynamic of a biological regulatory graph $G=(V,E)$ is modelled by ordinary differential equation systems [10, 22, 29], called here *underlying differential systems* (UDSs for short) of G, whose form is

$$\dot{x}_v = f_v(x) - ?_v x_v, \quad \text{for all } v \in V,$$

where $x = (x_v)_{v \in V}$ is a vector, whose components $x_v \in \mathbb{R}^+$ give the concentrations of variables v. The vector x is called the quantitative state of G. The previous equations define the rate of change of each concentration x_v as the difference of the synthesis rate $f_v(x)$ and the degradation rate $\lambda_v x_v$ of v. The function f_v expresses how the synthesis rate of v depends on the concentrations x_u of its regulators $u \in G^-(v)$. It can be defined as,

$$f_v(x) = k_v + \sum_{u \in G^-(v)} k_{uv}\, s^{\alpha_{uv}}(x_u, \theta_{uv}),$$

where:

- $k_v \in \mathbb{R}^+$ and $k_{uv} \in \mathbb{R}^{+*}$ are kinetic parameters,
- the function $s^{\alpha_{uv}}$ gives the effect of a regulator u and its target v. This function is usually a sigmoid depending on the sign α_{uv} and on the quantitative threshold $\theta_{uv} \in \mathbb{R}^{+*}$ of the interaction.

Since the qualitative thresholds t of G give the order of the continuous thresholds θ (see the section 2.1), it is required that for all targets $v' \in G^+(u)$ of u different than v, $\theta_{uv} < \theta_{uv'}$ if $t_{uv} < t_{uv'}$ and $\theta_{uv} > \theta_{uv'}$ if $t_{uv} > t_{uv'}$. By denoting θ_u^l the threshold(s) θ_{uv} such that $v \in G^+(u)$ and $t_{uv}=l$ we then have:

$$\theta_u^1 < \theta_u^2 < \ \ldots < \theta_u^1 < \theta_u^{b_u}$$

The sigmoidal function $s^{\alpha_{uv}}$ is often approximated by a step function in order to make possible the analytical analysis of the system. $s^{\alpha_{uv}}$ is then defined as a Boolean function which indicates if u is or not a resource of v:

$$s^+(x_u, \theta_{uv}) = \begin{cases} 1, \text{ if } x_u > \theta_{uv} \\ 0, \text{ if } x_u < \theta_{uv} \end{cases} \quad \text{and} \quad s^-(x_u, \theta_{uv}) = 1 - s^+(x_u, \theta_{uv})$$

Notice that these functions are not defined for $x_u = \theta_{uv}$ and that the system becomes a piecewise linear equation system. Figure 6 gives an example of an UDS of the biological regulatory graph of figure 3.

$$\begin{cases} \dot{x}_u = 20 + 35 \times s^-(x_v, 10) + 40 \times s^+(x_u, 20) - 5 \times x_u \\ \dot{x}_v = 25 \times s^+(x_u, 16) - 2 \times x_v \end{cases}$$

Figure 6: Example of UDS of the biological regulatory graph of figure 3. Parameters are: $k_u = 20$, $k_{vu} = 35$, $k_{uu} = 40$, $\theta_{vu} = 10$, $\theta_{uu} = 20$, $\lambda_u = 5$ for the first equation and $k_v = 0$, $k_{uv} = 25$, $\theta_{uv} = 16$, and $\lambda_v = 2$ for the second. Notice that $\theta_{uv} < \theta_{uu}$ since $1 = t_{uv} < t_{uu} = 2$.

3.2 Discretization map and domains

Since the step functions $s^{\alpha_{uv}}$ are not defined for $x_u = \theta_{uv}$, the differential equation system is not defined for the states x for which at least one component x_v equals a threshold $\theta_{vv'}$, $v' \in G^+(v)$. Such states are called *singular states*. Consequently, the properties of the system can be analysed in the $|V|$-dimensional phase space Ω defined by

$$\Omega = \prod_{v \in V} \Omega_v \quad \text{with} \quad \Omega_v = \mathbb{R}^+ \setminus \{\theta_{vv'} \mid v' \in G^+(v)\} \text{ for all } v \in V.$$

Ω corresponds to the set of *regular states*. We are now in position to define the *discretization map* $d : \Omega \to Q$ by $d(x) = (d_v(x_v))_{v \in V}$ with, for every $v \in V$, $d_v : \Omega_v \to Q_v$ defined by

$$d_v(x_v) = |\{\theta_{vv'} \mid v' \in G^+(v) \text{ and } \theta_{vv'} < x_v\}|.$$

This discretization map gives directly the cardinal of the set of thresholds less than the concentration of v. If $d_v(x_v) = l$, then x_v is greater than the l smallest thresholds and less than others. For all $v \in G^+(u)$ we have $x_u > \theta_{uv} \Rightarrow d_u(x_u) \geq t_{uv}$ and $x_u < \theta_{uv} \Rightarrow d_u(x_u) < t_{uv}$. Consequently, for all state $x \in \Omega$:

$$s^{\alpha_{uv}}(x_u, \theta_{uv}) = 1 \Rightarrow u \in \omega_v(d(x)).$$

Then, for all $x \in \Omega$, f_v can be rewritten as

$$f_v(x) = k_v + \sum_{\substack{u \in \omega_v \\ (d(x))}} k_{uv}. \qquad (2)$$

The infinite set of continuous states whose discretization gives $q \in Q$ is an hyper-rectangular region $D(q)$ of Ω, called *domain*, defined by:

$$D(q) = \prod_{v \in V} D_v(q_v) \quad \text{with} \quad D_v(q_v) = \{x_v \in \Omega_v \mid d_v(x_v) = q_v\} \text{ for all } v \in V.$$

A domain $D(q)$ is bounded by hyperplanes corresponding to thresholds: for all variable v, if $q_v > 0$ then $\theta_v^{q_v}$ is the lower bound of $D_v(q_v)$ and if $q_v < b_v$ then $\theta_v^{q_v+1}$ is the upper bound of $D_v(q_v)$ (see the figure 7).

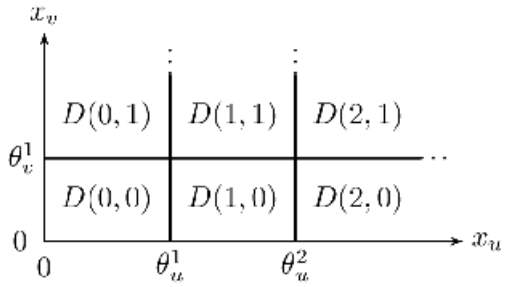

Figure 7: Domains of the phase space Ω of the UDS of figure 6 ($\theta_u^1 = \theta_{uv}$ and $\theta_u^2 = \theta_{uv}$).

3.3 Dynamics of differential equation systems

In a domain $D(q)$, each function f_v reduces to the constant $k_v+\sum_{u\in\omega_v(q)} k_{uv}$ (see equation 2). The system thus simplifies to a linear and uncoupled differential equation system whose solution in $D(q)$, starting at $x^0\in D(q)$, is given by

$$x_v(t) = \kappa_v(x^0) - (\kappa_v(x^0) - x_v^0)\, e^{-\lambda_v t}, \quad \text{for all } v\in V,$$

with $\kappa_v(x) = f_v(x)/\lambda_v$ for all $x\in\Omega$. Function κ_v is also reduced to a constant $(k_v+\sum_{u\in\omega_v(q)} k_{uv})/\lambda_v$ in $D(q)$, denoted $\kappa_v(q)$ by abuse of notation. The state $\kappa(q) = (\kappa_v(q))_{v\in V}$ acts as an attractor in $D(q)$. Indeed, it is easy to verify that in $D(q)$, $x(t)$ has the following properties:

1. if $\kappa_v(q)\in D_v(q_v)$ then, $x_v(t)$ monotonically converges from x_v^0 towards $\kappa_v(q)$ and reaches $\kappa_v(q)$ in infinite time. Thus, if $\kappa(q)\in D(q)$, $x(t)$ does not leave $D(q)$ and the state $\kappa(q)$ is the unique stable steady state in $D(q)$.
2. if $\kappa_v(q)\notin D_v(q_v)$, then, if $x_v^0 < \kappa_v(q)$ (resp. $x_v^0 > \kappa_v(q)$), $x_v(t)$ monotonically increases (resp. decreases) from x_v^0 until to reach the threshold value $\theta_v^{q_v+1}$ (resp. $\theta_v^{q_v}$). The threshold is reached in a finite time iff $\kappa_v(q)$ is different from it. Throughout this section, we suppose that the parameters k and λ are taken such that $\kappa(q)\in\Omega$ for all $q\in Q$. Consequently, if $\kappa(q)\notin D(q)$, $x(t)$ leaves in a finite time $D(q)$ by reaching a threshold hyperplane.

If $\kappa_v(q)\notin D_v(q_v)$, $x_v(t)$ reaches its corresponding threshold at time t_v given by

$$t_v = -\frac{1}{\lambda_v}\ln\left(\frac{\kappa_v(q)-\theta_v^{q_v+\alpha}}{\kappa_v(q)-x_v^0}\right)$$

with $\alpha = 1$ if $x_v^0 < \kappa_v(q)$ and $\alpha = 0$ if $x_v^0 > \kappa_v(q)$. If at least two components $x_u(t)$ and $x_v(t)$ reach their thresholds simultaneously, one can deduce that x^0 belongs to an at most $(|V|-1)$-dimensional surface of zero Lebesgue measure in $D(q)$. Therefore, we do not consider this case and reason now as for almost every $x^0\in D(q)$.

Suppose that t_v is the smallest value in $\{t_u \mid \kappa_u(q)\notin D_u(q_u)\}$, in other words, suppose that v is the variable whose concentration first leaves the domain. The component $x_v(t)$ reaches $\theta_v^{q_v+\alpha}$ at the singular state x^1 given by

$$x_v^1 = \theta_v^{q_v+\alpha} \quad \text{and} \quad x_u^1 = x_u(t_v) \quad \text{for all} \quad u \neq v.$$

At this time, the trajectory exits from the domain $D(q)$ and enters into $D(q')$ defined by:

- $D_u(q'_u) = D_u(q_u)$ for all $u \neq v$, since only v reached its threshold,
- $D_v(q'_v) = D_v(q_v+1)$ if $\alpha = 1$ and $D_v(q'_v) = D_v(q_v-1)$ if $\alpha = 0$.

But at the singular state x^1, the differential equation system is not defined as well as $\kappa_v(x^1)$. The linear differential equation system of $D(q')$ is then extended by continuity to the hyperplane $x_v = \theta_v^{q_v+\alpha}$. Thus $\kappa_v(x^1)$ is defined, and the trajectory is extended with the solution of the differential system of $D(q')$ from the new starting point x^1.

3.4 Discrete and differential modellings are coherent

Let *M(G)* be a model of a biological regulatory graph *G=(V,E)*. The UDSs of *G* such that :

$$d_v(\kappa_v(q)) = K_{v,\omega_v(q)} \quad \text{for all } v \in V \text{ and } q \in Q$$

are called *underlying differential systems of M(G)*. A model *M(G)* has UDSs if and only if it satisfies the Snoussi's constraints (equation 1) since we have:

$$d_v\left(\left(k_v + \sum_{u\in\omega} k_{uv}\right)/\lambda_v\right) = K_{v,\omega}, \text{ for all } v \in V \text{ and } \omega \subseteq G^-(v)$$

and thus $\omega \subseteq \omega'$ implies $\sum_{u\in\omega} k_{uv} \leq \sum_{u\in\omega'} k_{uv}$ which implies $K_{v,\omega} \leq K_{v,\omega'}$. For example the UDS of figure 6 is an UDS of the model described in Table 1. The following propositions show the coherence between the asynchronous dynamics of *M(G)* and the dynamics of its UDSs.

Proposition 1

- *If there is an UDS of M(G) such that $x \in D(q)$ is a stable steady state, then q is a stable state of the asynchronous state graph S of M(G).*
- *Conversely, if q is a stable state of S then, for all UDSs of M(G), there is a stable steady state in the domain $D(q)$.*

Proof. A state $x \in D(q)$ is a stable steady state iff $x_v = \kappa_v(q)$ for all $v \in V$. That implies $d_v(x_v) = d_v(\kappa_v(q)) \Rightarrow q_v = K_{v,\omega_v(q)}$ for all $v \in V$ and thus, q is a stable state of S. Conversely, if $q \in Q$ is a stable state, then $q_v = K_{v,\omega_v(q)} = d_v(\kappa_v(q))$ for all $v \in V$. Thus, $\kappa_v(q) \in D_v(q_v)$ for all $v \in V$ and consequently, $\kappa(q) \in D(q)$ is a stable steady state.

We define now the *boundary* of a domain as the set of singular states whose distance to the domain is null.

Proposition 2

- *If there is an UDS of M(G) for which there is a trajectory starting in $D(q)$ which reaches directly from $D(q)$ the hyperplane separating $D(q)$ and an adjacent domain $D(q')$, then $q \to q'$ is a transition of the asynchronous state graph S of M(G).*
- *Conversely, there exist UDSs of M(G) such that, for each successor q' of q in S, there is a trajectory starting in $D(q)$ which reaches directly from $D(q)$ the hyperplane separating $D(q)$ and $D(q')$.*

Proof. We have seen in section 3.3 that if a trajectory starting at $x^0 \in D(q)$ reaches the hyperplane separating $D(q)$ and an adjacent domain $D(q')$, then there is a unique variable $v \in V$ such that $q'_v \neq q_v$ and we have $q'_v = q_v + 1$ if $x^0_v < \kappa_v(q)$ or $q'_v = q_v - 1$ if $x^0_v > \kappa_v(q)$. Moreover, $\kappa_v(q) \notin D_v(q_v)$ thus $x^0_v < \kappa_v(q)$ iff $d_v(x^0_v) < d_v(\kappa_v(q))$ which is equivalent to $q_v < K_{v,\omega_v(q)}$. Similarly $x_v{}^0 > \kappa_v(q)$ iff $d_v(x^0_v) > d_v(\kappa_v(q))$ which is equivalent to $q_v > K_{v,\omega_v(q)}$. According to definition 5, $q \to q'$ is a transition of S.

Now, we prove the second part of the proposition. Consider the UDSs of *M(G)* such that $\lambda_u = \lambda$ for all $u \in V$ and an initial state $x^0 \in D(q)$. The trajectory starting at x^0 describes the part of the segment connecting x^0 to $\kappa(q)$ which belongs to $D(q)$.

Let q' be a successor of q in S. We have $\kappa(q) \notin D(q)$. Let us choose a point x^1 of the boundary of $D(q)$ belonging to the hyperplane separating $D(q)$ from the domain $D(q')$ and whose only one component equals a threshold. The trajectories starting at a point of the line connecting x^1 and $\kappa(q)$ which belongs to $D(q)$, reach x^1.

We deduce from the previous propositions that all the regular stable steady states of an UDS of *M(G)* are represented in its asynchronous state graph S. Moreover if a trajectory of an UDS of *M(G)* passes successively through the domains $D(q^0)$, $D(q^1)$, ..., $D(q^n)$ then $q^0 \to q^1 \to ... \to q^n$ is a path of S. But if $q^0 \to q^1 \to ... \to q^n$ is a path of S, it does not mean that there is a trajectory passing successively through the domains $D(q^0)$, $D(q^1)$, ..., $D(q^n)$. Using the terminology of [14], the qualitative modelling is said sound. A graphical comparison between the asynchronous dynamics of a model and a trajectory of one of its UDS is given in figure 8.

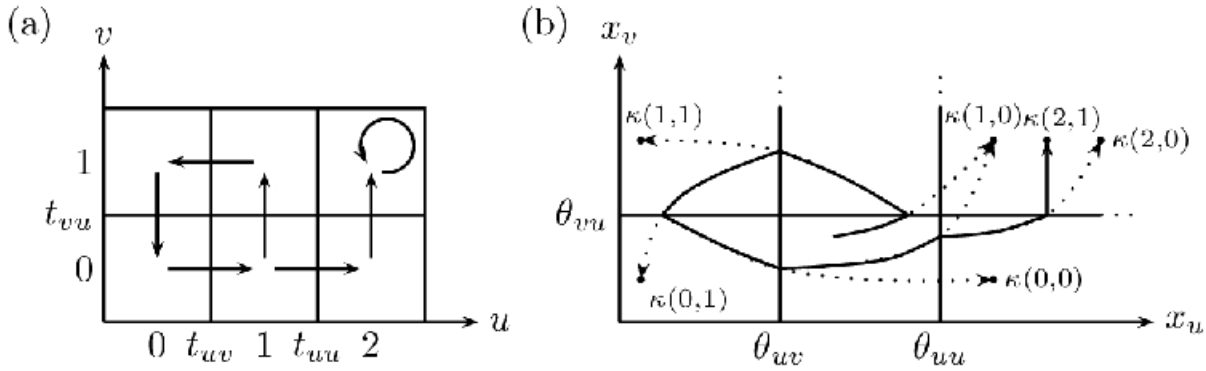

Figure 8: (a) The asynchronous state graph of the model *M(G)* of Table 1. (b) A trajectory of an UDS of *M(G)*. The dotted arrows represent the extensions of solutions towards the attractors.

Any UDS of a biological regulatory graph G is an UDS of a model of G satisfying the Snoussi's constraints. Thus the trajectories of the infinite set of UDSs of G are summarized by a finite set of asynchronous state graphs (for the biological regulatory graph of figure 3, we have 42 different state graphs deduced from the 60 different models satisfying the Snoussi's constraints).

3.5 Feedback circuit functionality

Feedback circuits play a major role for the dynamics of systems since they can generate multi-stationarity or homeostasis. A positive (resp. negative) circuit is said *functional* if it generates multi-stationarity (resp. homeostasis). The functionality of circuits is strongly related to the the stationarity of particular singular states and to discontinuities of the UDS. To deal with them, we first introduce the differential inclusion systems.

3.5.1 Differential inclusion systems

To deal with ordinary differential equation systems with discontinuous right-hand sides, Filippov [9] proposed to extend them to systems of differential inclusions. For the regulatory networks, the UDSs can be extended to the following differential inclusions systems:

$$x_v \in H_v(x), \text{ for all } v \in V, \tag{3}$$

where H_v is a set-valued function, defined as follow:

- for all regular state x, $H_v(x) = \{ f_v(x) - \lambda_v x_v \}$. For all $x \in D(q)$, since $f_v(x)/\lambda_v = \kappa_v(q)$, $H_v(x)$ can be rewritten as $H_v(x) = \{\lambda_v(\kappa_v(q) - x_v)\}$.
- for all singular state x,

$$H_v(x) = \overline{co}\ (\{\lambda_v(\kappa_v(q) - x_v) \mid q \in N(x)\}).$$

- where $\overline{co}(E)$ designs the smallest closed convex set of a set E which is the intersection of all closed convex sets containing E, and where $N(x)$ is the set of qualitative states which correspond to domains whose boundary contains x:

$$N(x) = \left\{ q \in Q \;\middle|\; \forall u \in V,\ q_u = \begin{cases} d_u(x_u), \text{ if } x_u \in \Omega_u \\ t_{uv} - 1 \text{ or } t_{uv} \text{ if } x_u = \theta_{uv} \text{ with } v \in G^+(u) \end{cases} \right\}.$$

- Obviously we have

$$H_v(x) = \left[\min_{q \in N(x)} \lambda_v\left(\kappa_v(q) - x_v\right),\ \max_{q \in N(x)} \lambda_v\left(\kappa_v(q) - x_v\right) \right]$$

Consider the example of figure 7. For x such that $x_u = \theta_u^2 = \theta_{uu}$ and $x_v > \theta_u^2$, we have $N(x) = \{(1,1), (2,1)\}$ and for these states, $\omega_u(1,1) = \{\}$ is included in $\omega_u(2,1) = \{u\}$ and $\omega_v(1,1) = \omega_v(2,1) = \{u\}$. We deduce that $H_u(x) = [\lambda_u(\kappa_u(1,1) - x_u),\ \lambda_u(\kappa_u(2,1) - x_u)]$ and $H_v(x) = \{\lambda_v(\kappa_v(1,1) - x_v)\}$. Intuitively, at the singular state x, the regulation $u \to v$ is clearly defined: $s^+(x_u, \theta_{uv}) = 1$. This is why the set $H_v(x)$ of the possible derivatives of x_v is single-valued. However, as $x_u = \theta_{uu}$ the self regulation of u remains undefined and $H_u(x)$ is *a priori* not single-valued: the derivative of x_u is comprised between the derivatives obtained with $s^+(\theta_{uu}, \theta_{uu}) = 0$ and $s^+(\theta_{uu}, \theta_{uu}) = 1$.

An absolutely continuous function $x(t)$ is solution of the system (3) in the sense of Filippov if $\dot{x}_v(t) \in H_v(x(t))$ for all $v \in V$ and for almost all $t \geq 0$. The qualification ``for almost all $t \geq 0$" means that the set time-points for which the condition does not holds if of measure 0. In particular, the condition is not satisfied at time-points when the solution arrives or leaves a threshold hyperplane.

We do not analyse the solutions in the sense of Filippov in this section (see [11, 7] for a detailed analysis), but the previous formalism will be useful for analysis of the steadiness of singular states.

3.5.2 Steadiness of singular states

It is not surprising that a state x, regular or singular, is an equilibrium point (in the sense that there is a solution $x(t)$ such that $x(t) = x$ for all $t \geq 0$) when $0 \in H_v(x)$ for all $v \in V$. For a regular state $x \in D(q)$, we have, as for differential equation systems:

$$0 \in H_v(x) \ \Rightarrow\ 0 \in \{\ \lambda_v(\kappa_v(q) - x_v)\ \} \ \Rightarrow\ x_v = \kappa_v(q).$$

In this case, x is a regular stable steady state. For a singular state, the inclusion can be written as an inequality:

$$0 \in H_v(x) \quad \Leftrightarrow \quad \min_{q \in N(x)} \lambda_v\left(\kappa_v(q) - x_v\right) \leq 0 \leq \max_{q \in N(x)} \lambda_v\left(\kappa_v(q) - x_v\right)$$
$$\Leftrightarrow \quad \min_{q \in N(x)} \kappa_v(q) \leq x_v \leq \max_{q \in N(x)} \kappa_v(q)$$

and if $x_v \notin \Omega_v$ the inequality becomes strict:

$$0 \in H_v(x) \quad \Leftrightarrow \quad \min_{q \in N(x)} \kappa_v(q) < x_v < \max_{q \in N(x)} \kappa_v(q)$$

because $\kappa_v(q) \in \Omega_v$ for all $q \in N(x)$. Among all singular equilibrium points, those for which we have $\min_{q \in N(x)} \kappa_v(q) = x_v = \max_{q \in N(x)} \kappa_v(q)$ for $x_v \in \Omega_v$, are *singular steady states* [24, 7]. Figure 9 shows a graphical representation of the conditions for the steadiness of singular states.

Proposition 3 *Let x be a singular state and v a variable. If for all $u \in G^-(v)$ $x_u \neq \theta_{uv}$, then $\kappa_v(q)$ is constant for all $q \in N(x)$.*

Proof. For all $u \in G^-(v)$ we have $x_u \in \Omega_u$ or $x_u = \theta_{uv'} \neq \theta_{uv}$ with $v' \in G^+(u)$. In the first case, it is evident that $q_u = q'_u$ for all q and q' in $N(x)$. In the second case, for all q and q' in $N(x)$, q_u and q'_u belong to $\{t_{uv'}-1, t_{uv'}\}$ and $t_{uv'} \neq t_{uv}$. Then q_u and q'_u are on the same side of t_{uv}. Consequently, for all q and q' in $N(x)$ we have $\omega_v(q) = \omega_v(q')$ which implies $\kappa_v(q) = \kappa_v(q')$.

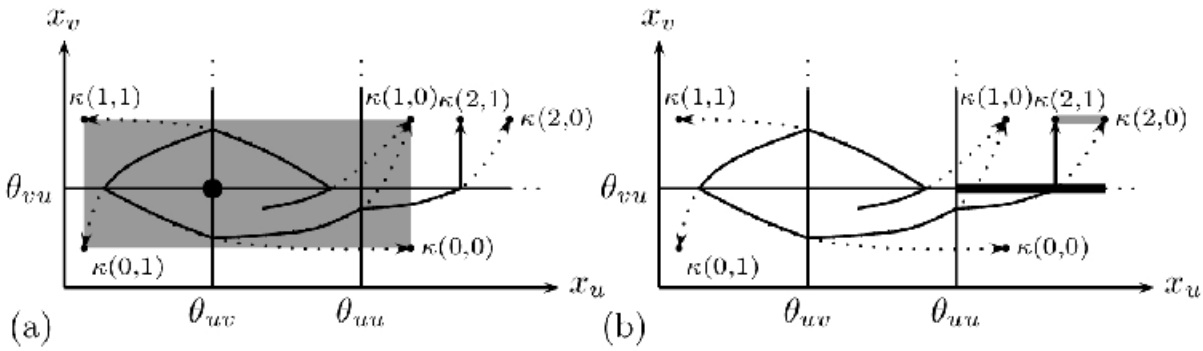

Figure 9: Equilibrium points and their steadiness. Grey regions, a rectangle in (a) and a segment in (b), correspond to the Cartesian product $\Psi(x) = [\min_{q \in N(x)} \kappa_u(q), \max_{q \in N(x)} \kappa_u(q)] \times [\min_{q \in N(x)} \kappa_v(q), \max_{q \in N(x)} \kappa_v(q)]$ for a singular state x. In (a) $x = (\theta_{uv}, \theta_{vu})$ is an equilibrium point ($x \in \Psi(x)$) and since all variables are singular, it is steady. In (b) the singular state is such that $x_u > \theta_{uu}$ and $x_v = \theta_{vu}$ and it is not an equilibrium point ($x \notin \Psi(x)$).

3.5.3 Circuit characteristic states

Definition 6 [Circuit] *Let $G=(V,E)$ be a biological regulatory graph. A circuit of G is a finite sequence of distinct elements of V, denoted $C = v_1, v_2, \ldots, v_n$, such that $v_n \to v_1 \in E$ and $v_i \to v_{i+1} \in E$ for all $i \in \{1,\ldots,n-1\}$.*

In the sequel, $\{C\}$ denotes the set of variables of a circuit C and $i+1$ (resp. $i-1$) is always computed modulo n: v_{i+1} (resp. v_{i-1}) denotes the successor (resp. predecessor) of v_i in C. Two circuits C and C' are disjointed if they have no variable in common. In a pedagogical objective, we focus here on the properties of a single circuit, but all results can be extended to

a union of disjointed circuits [24]. Moreover we take into consideration only regulatory graphs where for any variable the out-thresholds are distinct ($b_v = |G^+(v)|$, $\forall\ v \in V$).

A singular state x is said *characteristic* of a circuit $C = v_1,v_2,...,v_n$ if the concentration x_{v_i} of each variable v_i of the circuit is equal to the threshold $\theta_{v_iv_{i+1}}$ and if the concentrations of other variables are regular: $x_u \in \Omega_u$, $\forall\ u \notin \{C\}$.

Proposition 4 *A singular steady state is a characteristic state of a circuit of the biological regulatory graph G.*

Proof. Let x be a singular state and $S = \{v \mid x_v \notin \Omega_v\}$ be the set of variables equal to a threshold at the state x. If x is steady, we have for all $v \in S$:

$$\min_{q \in N(x)} \kappa_v(q) < x_v < \max_{q \in N(x)} \kappa_v(q).$$

According to the proposition 3, if for all $u \in G^-(v)$ we have $x_u \neq \theta_{uv}$ then $\min_{q\in N(x)} \kappa_v(q) = \max_{q\in N(x)} \kappa_v(q)$ and x is not steady. Thus v has at least one predecessor u such that $x_u = \theta_{uv}$, which implies that $u \in S$. Moreover, because $\theta_{uv'} \neq \theta_{uv}$ for all $v' \in G^+(u)$, the successor v of u is the only one such that $x_u = \theta_{uv}$. Each variable v of S has then a unique predecessor u in S such that $x_u = \theta_{uv}$.

Now, we prove that all the steady singular states can be identified in the qualitative modelling.

Proposition 5 *Let G be a biological regulatory graph containing a circuit $C = v_1,...,v_n$. Consider a UDS of a model M(G) and a characteristic state x of C. Let $q \in N(x)$. If x is steady, then M(G) is such that*

$$\begin{cases} K_{v,\omega_v(q)} = q_v & \text{for all } v \notin \{C\} \\ K_{v_i,\omega_{v_i}(q)\setminus\{v_{i-1}\}} < t_{v_iv_{i+1}} \leq K_{v_i,\omega_{v_i}(q)\cup\{v_{i-1}\}} & \text{for all } i \in \{1,\ldots,n\} \end{cases}$$

Proof. Let $v \notin \{C\}$. Since x is characteristic of C, we have $x_v \in \Omega_v$. If x is steady, then $\min_{q\in N(x)} \kappa_v(q) = x_v = \max_{q\in N(x)} \kappa_v(q)$. That means that $\kappa_v(q)$ is constant for all $q \in N(x)$ and we have $d_v(x_v) = d_v(\kappa_v(q))$ which is equivalent to $q_v = K_{v,\omega_v(q)}$.

Let $v_i \in \{C\}$. As x is characteristic of C, v_{i-1} is the unique predecessor of v_i such that $x_{v_{i-1}} = \theta_{v_{i-1}v_i}$. Thus, $\omega_{v_i}(q) \setminus \omega_{v_i}(q')$ equals $\{\}$ or $\{v_{i-1}\}$ for all q and q' in $N(x)$. Moreover there is at least one state $q \in N(x)$ such that $q_{v_{i-1}} = t_{v_{i-1}v_i}$ and another one such that $q_{v_{i-1}} = t_{v_{i-1}v_i}-1$. Thus, there is a state $q^+ \in N(x)$ such that $v_{i-1} \in \omega_{v_i}(q^+)$ and a state q^- with $v_{i-1} \notin \omega_{v_i}(q^-)$. We deduce that for all $q \in N(x)$, $\omega_v(q)\cup\{v_{i-1}\} = \omega_{v_i}(q^+)$ and $\omega_v(q)\setminus\{v_{i-1}\} = \omega_{v_i}(q^-)$. So $\max_{q\in N(x)} \kappa_{v_i}(q) = \kappa_{v_i}(q^+)$ and $\min_{q\in N(x)} \kappa_{v_i}(q) = \kappa_{v_i}(q^-)$. Since $x_{v_i} = \theta_{v_iv_{i+1}}$, if x is steady, we have for all $q \in N(x)$:

$$\begin{aligned} \kappa_{v_i}(q^-) < \theta_{v_iv_{i+1}} < \kappa_{v_i}(q^+) \quad &\Leftrightarrow \quad d_{v_i}(\kappa_{v_i}(q^-)) < \theta_{v_iv_{i+1}} < d_{v_i}(\kappa_{v_i}(q^+)) \\ &\Leftrightarrow \quad K_{v_i,\omega_{v_i}(q^-)} < t_{v_iv_{i+1}} \leq K_{v_i,\omega_{v_i}(q^+)} \\ &\Leftrightarrow \quad K_{v_i,\omega_{v_i}(q)\setminus\{v_{i-1}\}} < t_{v_iv_{i+1}} \leq K_{v_i,\omega_{v_i}(q)\cup\{v_{i-1}\}} \end{aligned}$$

Definition 7 [Quasi-characteristic qualitative states] *Let G=(V,E) be a biological regulatory graph containing a circuit $C = v_1,...,v_n$. A state $q \in Q$ is quasi-characteristic of C if $q_{v_i} = t_{v_iv_{i+1}}$ for all $v_i \in \{C\}$.*

The quasi-characteristic states are useful to locate the singular characteristic states of the UDS.

Proposition 6 *Let G be a biological regulatory graph containing a circuit $C = v_1,\ldots,v_n$ and a quasi-qualitative characteristic state q of C. If a model $M(G)$ satisfies the Snoussi's constraints and if*

$$\begin{cases} K_{v,\omega_v(q)} = q_v & \text{for all } v \notin \{C\} \\ K_{v_i,\omega_{v_i}(q)\setminus\{v_{i-1}\}} < t_{v_i v_{i+1}} \leq K_{v_i,\omega_{v_i}(q)\cup\{v_{i-1}\}} & \text{for all } i \in \{1,\ldots,n\} \end{cases} \quad (4)$$

then, for all the UDSs of $M(G)$, there exists a unique steady characteristic state x of C such that $d_u(x_u) = q_u$ for all $u \notin \{C\}$.

As the proof is quite similar to the previous one, it is omitted.

The previous proposition makes easy the determination of all steady singular states underlying of a qualitative model. Let us consider for instance the model *M(G)* of Table 1. The corresponding biological regulatory graph G (Figure 3-(a)) contains two circuits, $C^1 = u,v$ and $C^2 = u$. The unique quasi-characteristic state of C^1 is (t_{uv},t_{vu}). It satisfies

$$K_{u,\omega_u(t_{uv},t_{vu})\setminus\{v\}} < t_{uv} = 1 \leq K_{u,\omega_u(t_{uv},t_{vu})\cup\{v\}} \quad \text{and} \quad K_{v,\{\}} < t_{vu} = 1 \leq K_{v,\{u\}}.$$

Indeed the first inequality is verified because $\omega_u(t_{uv},t_{vu}) = \{\}$, $K_{u,\{\}} = 0$ and $K_{u,\{v\}} = 2$, the second is also verified since $K_{v,\{\}} = 0$ and $K_{v,\{u\}} = 1$. Consequently, the characteristic state $(\theta_{uv},\theta_{uv})$ is steady in all the UDSs of *M(G)*.
For circuit C^2, there are two quasi-characteristic states: $(t_{uu},0)$ and $(t_{uu},1)$.

- The first one, $(t_{uu},0)$, does not satisfy

  $K_{u,\omega_u(t_{uu},0)\setminus\{u\}} < t_{uu} = 2 \leq K_{u,\omega_u(t_{uv},0)\cup\{u\}}$ and $K_{v,\omega_v(t_{uu},0)} = 0$.

  since $\omega_v(2,1) = \{u\}$ and $K_{v,\{u\}} = 1$. Thus there is not any steady characteristic state of C^2 such that $x_v < \theta_{vu}$.

- The second quasi-characteristic state, $(t_{uu},1)$, satisfies

  $K_{u,\omega_u(t_{uu},1)\setminus\{u\}} < t_{uu} = 2 \leq K_{u,\omega_u(t_{uv},1)\cup\{u\}}$ and $K_{v,\omega_v(t_{uu},1)} = 1$.

  since $\omega_v(2,1) = \{u\}$, $K_{v,\{u\}} = 1$, $\omega_u(2,1) = \{u\}$, $K_{u,\{\}} = 0$ and $K_{u,\{u\}} = 2$. For all UDSs of *M(G)* there is a unique steady characteristic state x of C^2 such that $x_v > \theta_{vu}$.

The detected singular states are represented in the asynchronous state graph of *M(G)* in figure 10.

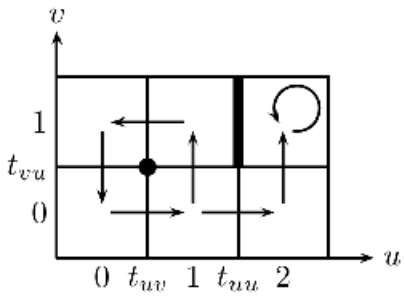

Figure 10: Representation of the steady singular states of model of Table 1.

3.5.4 Circuit functionality

Each variable of a feedback circuit has an influence on its target but also an indirect effect on all following variables including itself. A circuit is said positive (resp. negative) if each variable has a positive (resp. negative) influence on itself. The sign of a circuit is determined by the number of inhibitions: if it is odd, the circuit is negative and otherwise, the circuit is positive. Negative and positive circuits have different typical behaviours.

- In a negative circuit, a high level of a variable tends to make decrease itself and conversely. Thus the circuit makes the level of each variable to tend to (or oscillate around) an equilibrium concentration. It generates stable oscillation behaviour corresponding to homeostasis in biology.
- In a positive circuit, a high (resp. low) level of a variable tends to make it increase (resp. decrease). Thus each variable stays either at a low or high concentration and the positive circuit generates multi-stationarity corresponding to differentiation in biology.

A circuit which presents a typical behaviour is said functional. Several authors have proved that at least one positive circuit is necessary to generate multi-stationarity whereas at least one negative circuit is necessary to obtain a stable oscillatory behaviour [18, 23, 6, 5, 25]. Snoussi and Thomas realized that when a characteristic state is steady, the corresponding circuit is functional [24]. In the qualitative formalism, the circuit functionality is then defined as follow.

Definition 8 [Functional circuit] *Let M(G) be a model of a biological regulatory graph G containing a circuit C. If there is a quasi-characteristic state q of C satisfying the constraint (4) then C is functional.*

We deduce from the proposition 6 that if a circuit C is functional, there is, for all underlying differential systems, a steady characteristic state x of C such that $x_u = d_u(q_u)$ for all $u \notin \{C\}$. In the model of Table 1, both circuits $u \to u$ and $u \to v \to u$ are functional. As a result, multi-stationarity and homeostasis are present in the corresponding asynchronous state graph (figure 10).

Summing up, homeostasis and/or multi-stationarity are dynamical properties almost always present in biological systems. Circuit functionality is then useful for modelling such systems. For example, it has been used to model immunity control in lambda phage [26], pattern formation during the embryonic development of *Drosophilae* [20, 19] and flower morphogenesis in *Arabidopsis thaliana* [16].

4 Formal methods

To study the behaviour of the genetic regulatory network, the ordinary differential equation systems are well adapted if all the parameters are well known. Unfortunately they are most often unknown and are difficultly measurable *in vivo*. The discrete approach of Thomas and co-workers simplifies the problem of determining the suitable parameters since the number of possible models is finite. Indeed finding suitable classes of those parameters constitutes a major issue of the modelling activity. Even if the Snoussi's constraints on parameters are used, the number of remaining models is too large to analyse them by hand. Then biological knowledge or hypotheses on the behaviour of the system can be used as an indirect criterion to constrain the parameters. For example homeostasis (resp. multi-stationarity) is experimentally observable and it indicates that a negative (resp. positive) feedback circuit is functional, this functionality leading to some constraints on the parameters (see section 3).

To go further, conditions of multi-stationarity and homeostasis can be reinforced by introducing other conditions on the dynamics of the system. The available knowledge on the evolution of the system, as temporal properties, can be taken into consideration for constraining the values of parameters. Among all suitable models only a part of them are coherent with these temporal properties. Since numerous models have to be checked against those properties, a formal language is needed to perform automatically these checkings.

4.1 Temporal logic

The properties as the deadlock can be easily checked by exploring the transition system, called *asynchronous state graph* in section 2. For more complex properties on the dynamics of the system it is necessary to use a well adapted formal language: a temporal language which allows the specification of properties along the execution paths of the transition system. The step of the specification of the properties can then be distinguished from the specification of the system since it is not necessary to know the dynamic structure of the system to be checked for specifying the properties.

Expressing temporal properties on a transition system needs to define the atomic propositions which depends of the considered regulatory graph $G=(V,E)$. Generally the set of atomic propositions is denoted by AP. The subset of AP containing all the atomic propositions which are true in a state q, is given by the labelling function L:

$$L(q) = \{ (v = q_v) \mid v \in V \}$$

where $(v = q_v)$ signifies that the variable v has the concentration level q_v. The pair composed of a transition system and a labelling function is called a Kripke structure.

Execution traces of the transition system model implicitly a discrete time: if an execution passes from the state s_0 to s_1, the instant associated to the state s_1 follows the one corresponding to the state s_0. The temporal logics allow one to specify dynamical properties referring to this discrete time [8]. The Linear Temporal Logic, LTL, is used to specify properties on an execution of the system. If the system is determinist, from any initial state there is a unique execution, LTL is appropriated to specify properties of the system. Nevertheless the qualitative behaviour of a biological regulatory network is represented by an asynchronous state graph which is non determinist: the current state can have several possible futures. Since time has a tree structure, we prefer the Computation Tree Logic, CTL, in which it is possible to express properties of the form "*it is possible in the future that...*".

Definition 9 [Syntax of CTL] *A CTL formula on the set of atomic propositions AP is inductively defined by:*

- $\overline{\ }$, $\bot$ *and any atomic proposition of AP are formulae*
- *if* ϕ *and* ψ *are formulae, then* $(\neg\phi)$, $(\phi\wedge\psi)$, $(\phi\vee\psi)$, $(\phi\Rightarrow\psi)$, $(\phi\Leftrightarrow\psi)$, $AX\phi$, $EX\phi$, $A[\phi U\psi]$, $E[\phi U\psi]$, $AG\phi$, $EG\phi$, $AF\phi$, $EF\phi$ *are formulae.*

The semantics of CTL is defined on the execution trees of the transition system which are completely defined by their initial state and the transition relation. The semantics is given by the definition of the satisfaction relation $s = \phi$ meaning that the formula ϕ is satisfied on the execution tree starting at *s*.

Definition 10 [Semantics of CTL] *Let* s_0 *be a state. The semantics of CTL is defined inductively by:*

- $s_0 = \overline{\ }$ *and* $s_0 \neq \bot$,
- $\forall\, p\in AP$, $s_0 = p$ *iff* $p\in L(s_0)$,
- $s_0 = \neg\varphi$ *iff* $s_0 \neq \varphi$,
- $s_0 = \varphi_1\wedge\varphi_2$ *iff* $s_0 = \varphi_1$ *and* $s_0 = \varphi_2$,
- $s_0 = \varphi_1\vee\varphi_2$ *iff* $s_0 = \varphi_1$ *or* $s_0 = \varphi_2$,
- $s_0 = \varphi_1\Rightarrow\varphi_2$ *iff* $s_0 \neq \varphi_1$ *or* $s_0 = \varphi_2$,
- $s_0 = \varphi_1\Leftrightarrow\varphi_2$ *iff* $s_0 = (\varphi_1\Rightarrow\varphi_2)\wedge(\varphi_2\Rightarrow\varphi_1)$,
- $s_0 = AX\varphi$ *iff for all successors* s_1 *of* s_0, *we have* $s_1 = \varphi$,
- $s_0 = EX\varphi$ *iff for any successor* s_1 *of* s_0, *we have* $s_1 = \varphi$,
- $s_0 = AG\varphi$ *iff for all paths* $s_0,s_1...s_i...$, *and for all* s_i *along the path we have* $s_i = \varphi$,
- $s_0 = EG\varphi$ *iff for a particular path* $s_0,s_1...s_i...$ *we have for all* s_i *along the path* $s_i = \varphi$,
- $s_0 = AF\varphi$ *iff for all paths* $s_0,s_1...s_i...$, *there exists* s_i *along the path such that* $s_i = \varphi$,
- $s_0 = EF\varphi$ *iff for a particular path* $s_0,s_1...s_i...$, *there exists* s_i *along the path such that* $s_i=\varphi$,
- $s_0 = A[\varphi_1 U\varphi_2]$ *iff for all paths* $s_0,s_1...s_i...$, *there exists* s_i *along the path such that* $s_i = \varphi_2$ *and for each* $j<i$ *we have* $s_j = \varphi_1$,
- $s_0 = E[\varphi_1 U\varphi_2]$ *iff for a particular path* $s_0,s_1...s_i...$, *there exists* s_i *along the path such that* $s_i = \varphi_2$ *and for each* $j<i$ *we have* $s_j = \varphi_1$.

$\overline{\ }$ is the always true formula; $\bot$ is the always false formula; a state *s* satisfies all the atomic formulae of $L(s)$; $\neg$, $\wedge$, $\vee$, $\Rightarrow$, $\Leftrightarrow$ are the usual connectives (respectively *not*, *and*, *or*, *implication*, *equivalence*). All the temporal connectives are pairs of symbols: the first element is A or E followed by X, F, G or U whose meanings are given in the next table and illustrated in Figure 11.

A	for **A**ll paths choices	X	ne**X**t state
E	for at least one path choice (**E**xist)	F	some **F**uture state
		G	all future states (**G**lobally)
		U	**U**ntil

Consider the example of Figure 5-(b) where variables are *u* and *v*. The atomic proposition are $AP=\{(u = 0),(u = 1),(u = 2),(v = 0),(v = 1)\}$. $AX(v = 1)$ means that in all next states accessible

from the current state in the asynchronous state graph, the concentration level of v is 1. This formula is true iff the current state is (1,1), (2,0) or (2,1). EG(¬(u = 2)) means that there exists at least one path starting from the current state where the concentration level of u is constantly strictly less than 2. In Figure 5-(b), all states for which u is strictly less than 2 satisfy the formula. Then ¬(u = 2) ⇒ EG(¬(u = 2)) is satisfied for all states. A[(v = 1)U(v = 0)] means that for any possible path from the current state there exists a future state where v = 0 and in between v remains equal to 1. Note that (2,1) is the only state which does not satisfy the formula. And so on for other temporal connectives.

It is now possible to translate a biological temporal property into a CTL formula. Classically a biological system can have several steady states corresponding to distinct phenotypes. Let us suppose that two distinct stable states, ss_1 and ss_2, are possible and that formulae ψ_1 and ψ_2 characterize the states ss_1 and ss_2 respectively. If the system is able to go from state s_0, characterized by the formula φ_0, either to state ss_1 or to state ss_2, these temporal properties can be translate into formulae:

$\psi_1 \Rightarrow AG\ \psi_1$	stability of state ss_1
$\psi_2 \Rightarrow AG\ \psi_2$	stability of state ss_2
$(\varphi_0 \Rightarrow EF\psi_1)\ (\varphi_0 \Rightarrow EF\psi_2)$	reachability of ss_1 and ss_2 from s_0

Such formulae are used in the concrete example of section 5 for expressing biological knowledge on the immunity control in bacteriophage lambda.

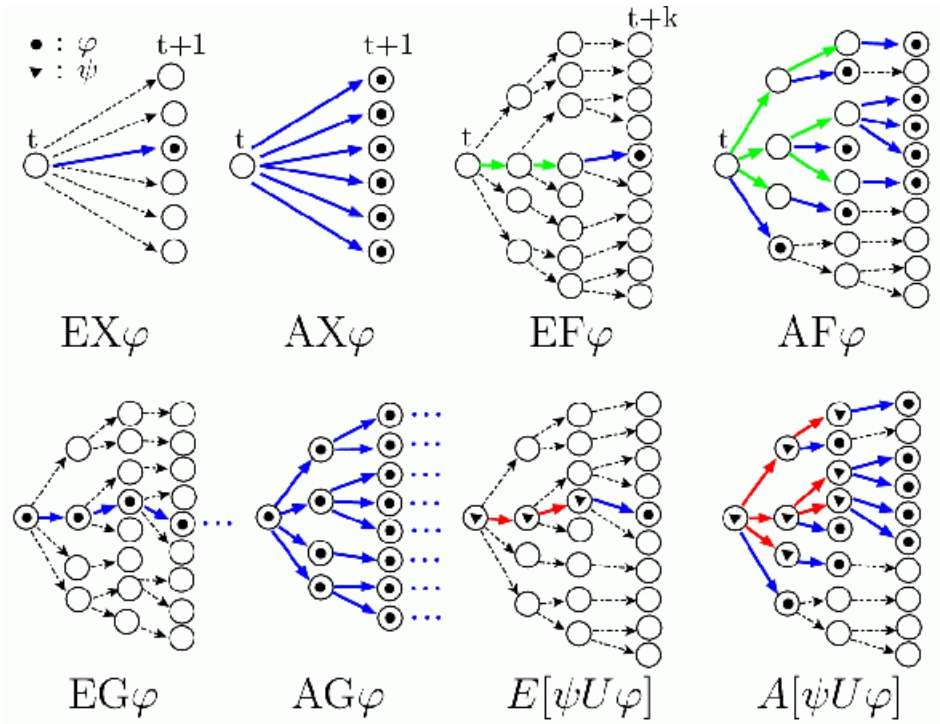

Figure 11: Semantics of temporal connectives of CTL.

4.2 Model checking

The model checking is a verification method that proves automatically if a Kripke structure satisfies a temporal formula [13]. We briefly present the basic algorithm of model checking

for a CTL formula. Since the connectives $\vee$, $\Rightarrow$ and $\Leftrightarrow$ can be rewritten in term of $\neg$ and $\wedge$, and since we have the following equivalence:

$$\begin{aligned} AX\varphi &\equiv \neg EX(\neg\varphi) \\ EG\varphi &\equiv \neg AF(\neg\varphi) \\ EF\varphi &\equiv E(\overline{\ } U\ \varphi) \\ AG\varphi &\equiv \neg EF(\neg\varphi) \\ A[\varphi_1 U \varphi_2] &\equiv \neg\ (\ E[\neg\varphi_2 U(\neg\varphi_1 \wedge \neg\varphi_2)] \vee EG(\neg\varphi_2)\ ) \end{aligned}$$

we consider in the sequel formulae containing only the connectives: $\neg$, $\wedge$, EX, AF and EU. Obviously any CTL formula can be transformed into a semantically equivalent CTL formula which uses only those connectives.

The model checking for a CTL formula φ consists in labelling each state s of the transition system with sub-formulae of φ which are satisfied at the state s. These sub-formulae are added to $L(s)$ containing initially the atomic propositions true in s. Suppose that ψ is a sub-formula of φ and that states satisfying all the immediate sub-formulae of ψ have already been labelled. The labelling algorithm for ψ uses a case analysis to label states with ψ:

- if $\psi \in AP$, then the labelling is given directly by $L(s)$
- if $\psi = p \wedge q$, then $L(s) = L(s) \cup \{p \wedge q\}$ for all s such that $p, q \in L(s)$
- if $\psi = \neg p$, then $L(s) = L(s) \cup \{\neg p\}$ for all s such that $p \notin L(s)$
- if $\psi = EXq$, then $L(s) = L(s) \cup \{EXq\}$ for all predecessors s of a state t such that $q \in L(t)$
- if $\psi = AFq$, then
    1. $L(s) = L(s) \cup \{AFq\}$ for all s such that $q \in L(s)$
    2. Repeat: $L(s) = L(s) \cup \{AFq\}$ for all states s such that all successors are labelled with AF q, until there is no change.
- if $\psi = E[qUr]$, then
    1. $L(s) = L(s) \cup \{E[qUr]\ \}$ for all s such that $r \in L(s)$,
    2. Repeat: $L(s) = L(s) \cup \{E[qUr]\}$ for all states s such that $q \in L(s)$ and which have a successor labelled with $E[qUr]$, until there is no change.

It can be proved that this labelling algorithm ends and that states are labelled with all sub-formulae of φ that they satisfy. Thus $s \models \varphi$ if the state s is labelled with φ. By extension if all states are labelled with φ, we say that the considered Kripke structure satisfies φ.

The model checking algorithm is linear with the size of the system and the size of the formula. Unfortunately, practical applications lead to transition systems with an enormous number of states, and the previous algorithm is often inefficient. To push back these limits, *symbolic* model checking [15] has been developed. It consists in computations on symbolic representation of subspaces of states.

To sketch the symbolic model checking, let us introduce the operator *Pre*. Let S be the set of states and x be a subset of S. $Pre(x)$ gives the set of states which have a successor in x. The set $sat(\varphi)$ of states satisfying φ can then be defined inductively:

- if $\varphi \in AP$, $sat(\varphi) = \{s \in S \mid \varphi \in L(s)\}$
- $sat(\neg\varphi) = S \setminus sat(\varphi)$
- $sat(\varphi \vee \psi) = sat(\varphi) \cup sat(\psi)$
- $sat(\varphi \wedge \psi) = sat(\varphi) \cap sat(\psi)$
- $sat(EX\varphi) = Pre(sat(\varphi))$

- $sat(\mathrm{AX}\varphi) = S \setminus Pre(S \setminus sat(\varphi))$
- The connectives $\mathrm{AF}\varphi$ and $\mathrm{E}[\varphi_1 \mathrm{U} \varphi_2]$ are more difficult to define. Let us remark that we have the following equivalence:

$$\begin{aligned} \mathrm{AF}\ \varphi &\equiv \varphi \vee (\mathrm{AX}\ (\mathrm{AF}\varphi)) \\ \mathrm{E}[\varphi_1 \mathrm{U} \varphi_2] &\equiv \varphi_2 \vee (\ \varphi_1 \wedge \mathrm{EX}(\mathrm{E}[\varphi_1 \mathrm{U} \varphi_2])\ ). \end{aligned}$$

  Then $sat(\mathrm{AF}\varphi)$ and $sat(\mathrm{E}[\varphi_1 \mathrm{U} \varphi_2])$ can be defined as the smallest fixed points of equations:

$$f_1(x) = sat(\varphi) \cup sat(\mathrm{AX}\ x)$$
$$f_2(x) = sat(\varphi_2) \cup (sat(\varphi_1) \cap sat(\mathrm{EX}\ x)).$$

  Since functions f_1 and f_2 are monotone and that the set of states is finite, the iterative computation of the smallest fixed point ends.

The *Binary Decision Diagrams*, or BDD for short, are data structures allowing the representation of Boolean expressions in a very compact way. Then subsets of states can be coded with such Boolean expressions and necessary operations for computing *sat* can be defined on these structures. Numerous works detail utilization of BDDs for the verification of systems, see for example [*15*, *13*].

4.3 A tool for the selection of models: SMBioNet

We have designed a software for a computer aided modelling based on the previous described formal methods [*3*]. This software, SMBioNet[1], helps the biologist and/or the modeller to verify systematically the coherence of models of a given biological system, and to select suitable models which satisfy the temporal properties extracted from knowledge or hypothesis. More precisely inputs of SMBioNet consist in

- a biological regulatory graph representing the interactions of the biological system and
- a CTL formula expressing its known or hypothetical dynamical properties.

Then it generates all the models of the biological regulatory graph and gives as output those satisfying the CTL formula. For each generated model, SMBioNet calls the model checker NuSMV [*4*] and selects it if the formula is satisfied. For each selected model, the asynchronous state graph and the steady states (regular and singular) are given. Depending on the available biological knowledge, the user can

- reduce the domain of variation of some parameters,
- apply general constraints on parameters as, for example, the Snoussi's and observability[2] constraints,
- specify a set of steady states (regular and singular) and a set of functional circuits.

These direct constraints on parameters decrease significantly the number of models to generate and consequently increase the efficiency of the selection. However, one can test directly the coherence of the regulatory graph (*i.e.* is there at least one suitable model ?), without enumeration of models by using a symbolic description of the set of all models.

[1] Selection of Models for Biological Networks, see `http://smbionet.lami.univ-evry.fr`
[2] Presented in the next section.

In the next section, we shows how SMBioNet can be used for modelling the immunity control in bacteriophage lambda.

5 Immunity control in bacteriophage lambda

One of the most studied genetic regulatory networks is probably the one controlling immunity in temperate bacteriophage lambda which is a temperate virus. As described in figure 12, after infection of a bacterial population, many bacteria soon lyse and produce new phages but some survive and carry lambda genome in a dormant form. The first response is called *lytic* and the second *lysogenic*. In the lysogenic bacteria, viral DNA has integrated into the bacterial chromosome and will be faithfully transmitted to the bacterial progeny. In this condition, the viral gene cI, produces a repressor which blocks the expression of all the other genes of the phage, thus making the viral genome harmless for the bacterium. Moreover, cI makes lysogenic bacteria *immune* towards other infections. Lysogenization necessitates two events, integration of the viral DNA into the bacterial chromosome and development of immunity due to the expression of the repressor. The choice between the lytic and lysogenic pathways is very similar to cell differentiation, in the sense that a given virus, infecting apparently identical cells, can behave in two extremely different ways.

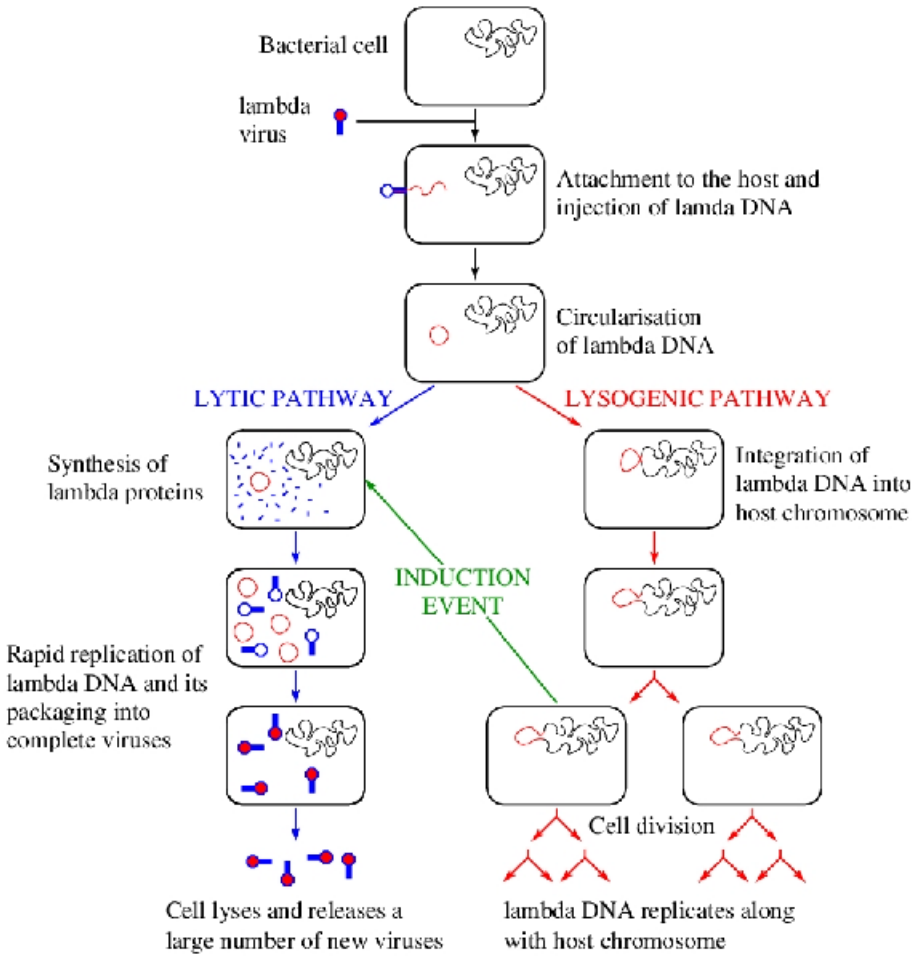

Figure 12: The life cycle of bacteriophage lambda.

It is actually in the context of this biological system that Thomas started to develop his formalism. Although he proposed various models of the immunity control [30, 29, 26], we focus in this section on the model developed by Thieffry and Thomas in [26], which is denoted $M(G)$ in the sequel. We will show that SMBioNet allows one to select, automatically and with very few biological knowledge, a set of models containing $M(G)$ and satisfying the validation criteria given by Thieffry. All models of this set have to be considered since they have *a priori* the same prediction capacity than $M(G)$.

5.1 Biological regulatory graph

The biological regulatory graph G summarizes the main regulations of the immunity control (Figure 13). Obviously it contains gene cI, but also three others (cro, cII, and N) which play a predominant role. Gene cI is activated by cII. Once on, gene cI remains on because its product activates its own synthesis, but at the same time, gene cI switches off the other lambda genes, including cII which had just switched it on. In addition gene cro exerts a negative control on cI, directly and indirectly, by repressing gene cII. Finally, gene N exerts a positive control on cII and is itself under negative control of cI and cII. According to the thresholds fixed by Thieffry, variables cI, cro, cII and N are 3-,4-,2- and 2-valued respectively, leading to 48 possible states. In the remainder, the state of the system is represented by the vector (cI,cro,cII,N).

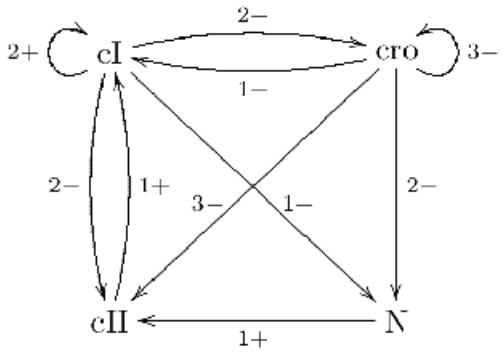

Figure 13: Biological regulatory graph G for immunity control.

5.2 Temporal properties

When the viral genome integrates a cell, all the viral proteins are initially absent. Thus (0,0,0,0) corresponds to the initial state of the system. The existence of both responses, lytic and lysogenic, implies that there exist two paths starting from the initial state leading respectively to the lytic state and to the immune one. The lytic state is known to be characterized by high concentration of cro and a low concentration of cI, cII and N whereas immune state is characterized by high concentration of cI and low concentration of cro, cII and N. In [26], both states (0,2,0,0) and (0,3,0,0) correspond to the lytic state and (2,0,0,0) is the only state corresponding to the immunity. Without change of the environment, the choice between the lytic and the lysogenic pathways is irreversible, thus the lytic and immune states are steady. Then if the system reaches one state of the sets A={(0,2,0,0),(0,3,0,0)} or B={(2,0,0,0)}, then it will never leave it. These sets of states are said steady sets.

Summing up, dynamics of models to consider have to contain paths from (0,0,0,0) to the steady sets of states A and B. These properties are translated into the CTL formula Φ as follow:

$$
\begin{aligned}
\mathit{init} &= ((\mathrm{cI} = 0) \wedge (\mathrm{cro} = 0) \wedge (\mathrm{cII} = 0) \wedge (\mathrm{N} = 0)) \\
\mathit{lytic} &= ((\mathrm{cI} = 0) \wedge (\mathrm{cro} \geq 2) \wedge (\mathrm{cII} = 0) \wedge (\mathrm{N} = 0)) \\
\mathit{immune} &= ((\mathrm{cI} = 2) \wedge (\mathrm{cro} = 0) \wedge (\mathrm{cII} = 0) \wedge (\mathrm{N} = 0)) \\
\Phi_A &= \mathit{lytic} \Rightarrow \mathrm{AG}(\mathit{lytic}) \\
\Phi_B &= \mathit{immune} \Rightarrow \mathrm{AG}(\mathit{immune}) \\
\Phi_r &= \mathit{init} \Rightarrow (\mathit{EF}(\mathit{lytic}) \wedge \mathit{EF}(\mathit{immune})) \\
\Phi &= \Phi_A \wedge \Phi_B \wedge \Phi_r
\end{aligned}
$$

The sub-formulae *init*, *lytic* and *immune* characterize the initial state, and the sets A and B. The steadiness of A and B is translated by Φ_A and Φ_B. The formula Φ_r expresses reachability of A and B from the initial state and Φ represents the temporal properties to use for the selection of models.

5.3 Selected models

There is near 7 thousands of millions of models associated to G leading to about 3 millions of different asynchronous state graphs. If we consider the Snoussi's constraints (equation 1) as Thieffry and Thomas did, it remains 151200 models. Moreover, we use the activity constraints [2]:

$$\text{for each regulation } u \to v \text{ there is a set } \omega \subset G^-(v) \text{ such that } K_{v,\omega} \neq K_{v,\omega \cup \{u\}}$$

which stands for the observability of any regulation. If $u \to v$ does not satisfy the constraints, the attractor of v does not depend on the level of u. It seems then quite obvious that any model should satisfy this property in order that all regulations play a role in the dynamics. Taking into account these constraints, SMBioNet selects among the 882 remaining models, 88 models satisfying the formula Φ. The model $M(G)$ proposed by Thieffry and Thomas is one of them. Table 2 shows the possible values of parameters for the selected models. 17 parameters among 24 are fixed by formula Φ (in particular, all the parameters associated to N).

$K_{\mathrm{cI},\{\}}$	= **0**	$K_{\mathrm{cII},\{\}}$	= **0**
$K_{\mathrm{cI},\{\mathrm{cI}\}}$	= **1** or 2	$K_{\mathrm{cII},\{\mathrm{cI}\}}$	= **0**
$K_{\mathrm{cI},\{\mathrm{cro}\}}$	= 0,1 or **2**	$K_{\mathrm{cII},\{\mathrm{cro}\}}$	= **0**
$K_{\mathrm{cI},\{\mathrm{cII}\}}$	= 0,1 or **2**	$K_{\mathrm{cII},\{\mathrm{N}\}}$	= **0**
$K_{\mathrm{cI},\{\mathrm{cI,cro}\}}$	= **2**	$K_{\mathrm{cII},\{\mathrm{cI,cro}\}}$	= **0** or 1
$K_{\mathrm{cI},\{\mathrm{cI,cII}\}}$	= 1 or **2**	$K_{\mathrm{cII},\{\mathrm{cI,N}\}}$	= **0** or 1
$K_{\mathrm{cI},\{\mathrm{cro}\}}$	= **2**	$K_{\mathrm{cII},\{\mathrm{cro,N}\}}$	= **0** or 1
$K_{\mathrm{cI},\{\mathrm{cI,cro,cII}\}}$	= **2**	$K_{\mathrm{cII},\{\mathrm{cI,cro,N}\}}$	= **1**
$K_{\mathrm{cro},\{\}}$	= **0**	$K_{\mathrm{N},\{\}}$	= **0**
$K_{\mathrm{cro},\{\mathrm{cI}\}}$	= **2**	$K_{\mathrm{N},\{\mathrm{cI}\}}$	= **0**
$K_{\mathrm{cro},\{\mathrm{cro}\}}$	= **0**	$K_{\mathrm{N},\{\mathrm{cro}\}}$	= **0**
$K_{\mathrm{cro},\{\mathrm{cI,cro}\}}$	= 2 or **3**	$K_{\mathrm{N},\{\mathrm{cI,cro}\}}$	= **1**

Table 2: Possible values of parameters for the selected models. Bold numbers correspond to the model $M(G)$.

5.4 Validation of models

Thieffry and Thomas exhibited one model whose coherence is analysed through the likelihood of some paths of the asynchronous state graph of *M(G)* and through the pertinence of predictions on the dynamics of some mutants. Our approach leads to select 88 models which have to be evaluated with the same biological criteria of validation.

- Although 4 positive feedback circuits are present in the regulatory graph, the 88 selected models present only two steady states (regular or singular): (2,0,0,0) is always steady and the other one is either (0,2,0,0) or a singular state adjacent to (0,2,0,0) and (0,3,0,0). These steady states correspond to the lytic and immune states, and no other stable behaviour (phenotype) can be observed.

- Even if several pathways are possible from the initial state to immune state, all selected models present the most likely pathway in *M(G)* from initial state to *A* (see Figure 14).

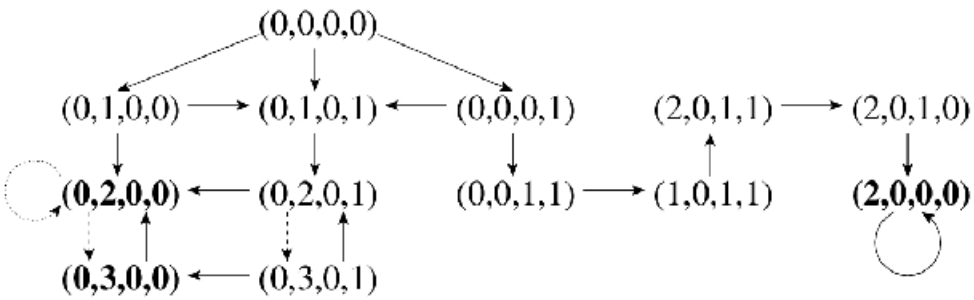

Figure 14: Likely paths from the initial state to the lytic and immune states (in bold). The dotted arrow is absent for the 44 models such that $K_{cro,\{cro,cI\}} = 3$, *M(G)* included, whereas the dashed ones are absent for others.

Similarly the pattern of dynamics present in *M(G)* allowing the system to evolve from initial state to lytic state, is also present in all selected models.

- Biological knowledge on mutants is available and can be used for validating models. The considered mutations correspond to the inactivation of different combinations of genes. Then simulations of the behaviour of these mutants can be performed and confronted to the biological knowledge. For example, the dynamics of the mutant $\lambda_{cI\text{-}cro}$-, where genes cI and cro are inactivated, is obtained from *M(G)* by setting to 0 all parameters associated to cro or cI. Consequently, from an initial state where cI and cro are absent, they will never appear. The dynamics of this mutant is given in Figure 15.

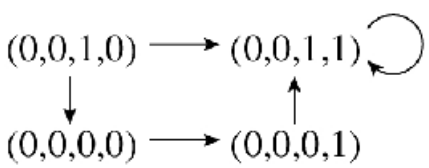

Figure 15: Dynamics of the mutant $\lambda_{cI\text{-}cro}$- obtained from *M(G)*.

The dynamics of mutants obtained from *M(G)* are coherent from a biological point of view, since the remaining basins of attraction allow the prediction of the behaviour of mutants. For any selected model, results are the same and are given in table 3.

Among the 88 selected models, some differences can be highlighted. For example, 2 states are unreachable from the initial state in *M(G)* whereas for some selected models 15 states are unreachable. In such models, all states with cI = 2 and cro = 3 are not reachable, which is reasonable because high concentration of cI and cro is rarely observed. Moreover, such models do not contain the path (0,0,0,0) → (1,0,0,0) → (2,0,0,0) present in the dynamics of *M(G)* and which is unlikely in view of the low expression of cI when cII is absent.

Mutants	Basins of attraction
λ_{cI}-	A
λ_{cro}-	B
λ_{cII}-	A and B
λ_{N}-	A and B
$\lambda_{cI\text{-}cro}$-	{(0,0,1,1)}
$\lambda_{cII\text{-}N}$-	A and B
$\lambda_{cro\text{-}N}$-	A
$\lambda_{cro\text{-}cII}$-	A
$\lambda_{cro\text{-}cII\text{-}N}$-	A

Table 3: Basins of attraction for a collection of mutants.

In conclusion, these 88 selected models satisfy the same criteria of validation that *M(G)* and have also to be considered. These models have been selected using a formula Φ expressing the well known properties of the system. Thieffry and Thomas have exhibited their model with the circuit functionality and some hypothesis. We can notice that the used constraints for functionality are not necessary to reproduce the biological properties (expressed by Φ) because some of the models selected by Φ do not satisfy these functionality constraints. Moreover some parameters are valuated according to hypotheses ($K_{cII,\{cI,N\}} = 2$ for example) which have to be slacken since some models selected by Φ propose different values for these parameters.

6 Conclusion

We have defined a *formal* description of biological regulatory networks which allows a computer aided manipulation of the semantics of the discrete modelling of Thomas, this manipulation being proved correct by construction. Our approach allows biology to take advantage of the whole corpus of formal methods from computer science. Model checking is a first powerful tool offered by the formalization of biological regulatory networks. In

particular, temporal properties can be added into the specifications of the system, and the modelling task consists in exhibiting one or more generally all models that are coherent with the previous specifications expressing a part of the biological knowledge concerning the dynamics of the system. All potential models have to be checked against temporal formulae, and this task can be done automatically using model checking. This *brute force* approach permits one to exhibit exhaustively all suitable models, *i.e.* all models satisfying the temporal formulae. Information provided by a new experiment or a new theoretical point of view will refine the set of selected models.

The available temporal properties concern generally the homeostasis, the multi-stationarity, stable steady states and the accessibility of some stable steady states from a partially specified initial state. Unfortunately the stable steady states are some time singular and not formally represented in the asynchronous state graph of Thomas. Then the specifications cannot easily contain temporal properties concerning such singular states. This would necessitate to rewrite these temporal properties with only atomic propositions of regular states, and this task is generally difficult.

De Jong et al. [7] introduced the singular states into their qualitative dynamics. Their qualitative modelling of genetic regulatory networks is also based on piecewise-linear differential equations. Authors propose a mathematically well founded method to deal with singular states using differential inclusions [9, 11]. Our approach consisting in adding temporal properties into the specifications for determining the suitable parameter values, would allows in this context to treat regular states as well as singular states.

More generally the formal methods can be applied in the field of biological regulatory networks and systems biology in order to explicit some behaviours or to take into account biological knowledge which have been ignored for the moment. The cooperation between biology and formal methods from computer science opens a large horizon of research perspectives.

- The introduction of transitions in the regulatory graph could help to specify how the different regulators cooperate for inducing or repressing their common target [1]. One can also separate inhibitors from activators [2] to increase the expressivity of the approach, or take into account time delays [31] between the beginning of the activation order and the synthesis of the product and conversely for the turn-off delays.
- Automatic generation of experiment schema from suitable models. In order to reduce again the set of suitable models, we would like to propose the biologist to perform an determining experiment. The result is then confronted to each model and only those which are coherent with the experiment, have to be kept. An experiment often consists to put the system in a particular state (partially specified) and to observe after a while if one or several gene products are present or not. This implies to extract the specificities of the biological application domain in order to define patterns of formulae expressing feasible experiments.
- The modelling of a regulatory network concerns generally only a small part of the global regulatory network of the cell. It becomes crucial to prove that the dynamical properties of this sub-network are preserved when it is embedded into the global network. This is correlated to the treatment of knock-out mutants, identification of functional patterns [21] as well as the structure of huge regulatory networks.

To achieve such development several directions have to be considered. High-level Petri nets are graphical oriented languages for design, specification, simulation and verification of systems. They are in particular well-suited for systems in which communication,

synchronization and resource sharing are important. Clearly, biological systems present these characteristics, and modelling by such nets would allows us to take advantage of all results and tools in the field of high-level Petri nets.

Hybrid automata can take into account the continuous aspects of a regulatory network: it is a mathematical model for hybrid systems, which combines, in a single formalism, automaton transitions for capturing discrete changes with differential equations for capturing continuous changes. Symbolic model checkers, as HyTech [*12*], have been developed for the subclass of linear hybrid automata. It becomes possible to perform parametric analysis, *i.e.* to determine the values of parameters for which a linear hybrid automaton satisfies a temporal-logic requirement.

These research perspectives aim to link modelling and experiments together, by furnishing to biologists model structuring methods and model validation tools from current researches in theoretical computer science. The resulting formal models are not only *a posteriori* explanations of biological results, they are guides for biological experiments whose success will be *in fine* the discriminating criterion.

Acknowledgements

The authors thank *genopole*®-research in Evry (H. Pollard and P. Tambourin) for constant supports. We gratefully acknowledge the members of the *genopole*® working groups *observability* and G^3 for stimulating interactions.

References

[1] V. Bassano and G. Bernot. Marked regulatory graphs: a formal framework to simulate biological regulatory networks with simple automata. In *14'th International Workshop on Rapid System Prototyping*, pages 93-99, San Diego, 2003.

[2] G. Bernot, F. Cassez, J.-P. Comet, F. Delaplace, C. Müller, O. Roux, and O.H. Roux. Semantics of biological regulatory networks. In *Proceedings of the Workshop on Concurrent Models in Molecular Biology (BioConcur'2003)*, 2003.

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

[4] A. Cimatti, E. Clarke, F. Giunchiglia, and M. Roveri. NuSMV: a reimplementation of SMV. In *Proceeding of the International Workshop on Software Tools for Technology Transfer (STTT-98)*, BRICS Notes Series, NS-98-4, pages 25-31, 1998.

[5] O. Cinquin and J. Demongeot. Positive and negative feedback: striking a balance between necessary antagonists. *J. Theor. Biol.*, 216(2):229-241, 2002.

[6] O. Cinquin and J. Demongeot. Roles of positive and negative feedback in biological systems. *C.R.Biol.*, 325(11):1085-1095, 2002.

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

[8] E.A. Emerson. *Handbook of theoretical computer science, Volume B : formal models and semantics*, chapter Temporal and modal logic, pages 995-1072. MIT Press, 1990.

[9] A.F. Filippov. *Differential equations with discontinuous right-hand sides*. Kluwer Academic Publishers, 1988.

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

[11] J.-L. Gouzé and S. Tewfik. A class of piecewise linear differential equations arising in biological models. *Dynamical Syst.*, 17:299-316, 2003.
[12] T.A. Henzinger, P.-H. Ho, and H. Wong-Toi. HyTech: A model checker for hybrid systems. *Software Tools for Technology Transfer*, 1:110-122, 1997.
[13] M. Huth and M. Ryan. *Logic in Computer Science: Modelling and reasoning about systems*. Cambridge University Press, 2000.
[14] B.J. Kuipers. *Qualitative reasoning: modeling and simulation with incomplete knowledge*. MIT Press, 1994.
[15] K. McMillan. *Symbolic Model Checking*. Kluwer Academic Publishers, 1993.
[16] L. Mendoza, D. Thieffry, and E.R. Alvarez-Buylla. Genetic control of flower morphogenesis in arabidopsis thaliana: a logical analysis. *Bioinformatics*, 15(7-8):593-606, 1999.
[17] S. Pérès and J.-P. Comet. Contribution of computational tree logic to biological regulatory networks: example from pseudomonas aeruginosa. In *International workshop on Computational Methods in Systems Biology*, volume 2602 of *LNCS*, pages 47-56, February 24-26, 2003.
[18] E. Plathe, T. Mestl, and S.W. Omholt. Feedback loops, stability and multistationarity in dynamical systems. *J. Biol. Syst.*, 3:569-577, 1995.
[19] L. Sánchez and D. Thieffry. A logical analysis of the drosophila gap-gene system. *J. Theor. Biol.*, 211(2):115-141, 2001.
[20] L. Sánchez, J. van Helden, and D. Thieffry. Establishment of the dorso-ventral pattern during embryonic development of drosophila melanogaster: a logical analysis. *J. Theor. Biol.*, 189(4):377-389, 1997.
[21] S.S. Shen-Orr, R. Milo, S. Mangan, and U. Alon. Network motifs in the transcriptional regulation network of Escherichia coli. *Nat. Genet.*, 31(1):64-68, 2002.
[22] E.H. Snoussi. Qualitative dynamics of a piecewise-linear differential equations : a discrete mapping approach. *Dynamics and stability of Systems*, 4:189-207, 1989.
[23] E.H. Snoussi. Necessary conditions for multistationarity and stable periodicity. *J. Biol. Syst.*, 6:3-9, 1998.
[24] E.H. Snoussi and R. Thomas. Logical identification of all steady states : the concept of feedback loop characteristic states. *Bull. Math. Biol.*, 55(5):973-991, 1993.
[25] C. Soulé. Graphical requirements for multistationarity. *ComPlexUs*, 1:123-133, 2003.
[26] D. Thieffry and R. Thomas. Dynamical behaviour of biological regulatory networks - II. Immunity control in bacteriophage lambda. *Bull. Math. Biol.*, 57(2):277-297, 1995.
[27] R. Thomas. Logical analysis of systems comprising feedback loops. *J. Theor. Biol.*, 73(4):631-656, 1978.
[28] R. Thomas. On the relation between the logical structure of systems and their ability to generate multiple steady states or sustained oscillations. *Springer Series in Synergies 9*, pages 180-193, 1980.
[29] R. Thomas and R. d'Ari. *Biological Feedback*. CRC Press, 1990.
[30] R. Thomas, A.M. Gathoye, and L. Lambert. A complex control circuit. Regulation of immunity in temperate bacteriophages. *Eur. J. Biochem.*, 71(1):211-227, 1976.
[31] R. Thomas and M. Kaufman. Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback circuits. *Chaos*, 11:180-195, 2001.
[32] R. Thomas, D. Thieffry, and M. Kaufman. Dynamical behaviour of biological regulatory networks - I. Biological role of feedback loops an practical use of the concept of the loop-characteristic state. *Bull. Math. Biol.*, 57(2):247-276, 1995.
[33] E.O. Voit. *Computational Analysis of biochemical systems: a practical guide for biochemists and molecular biologists*. Cambridge University Press, 2000.