Genes, Genomes and Genomics ©2009 Global Science Books

The SMBioNet Method for Discovering Models of Gene Regulatory Networks

Zohra Khalis[1,2] • Jean-Paul Comet[1] • Adrien Richard[1] • Gilles Bernot[1,2*]

[1] Laboratory I3S, CNRS & University of Nice-Sophia Antipolis, F-06903 Sophia Antipolis cedex, France
[2] Epigenomics Project, University of Évry & Genopole, Tour Évry 2, Place des Terrasses, F-91034 Évry, France

Corresponding author: * **khalis@unice.fr**

ABSTRACT

To study gene regulatory networks, we work on an iterative approach that permits us via formal modelling to elaborate models *in silico* and to validate them *in vivo* and/or *in vitro*. An iterative approach consists in reviewing the gene regulatory networks at each cycle with novel biological predictions and new information brought by experimental methods. The cornerstone of the modelling process is the selection of parameter values that are consistent with the known properties of the system. In this article, the coherent parameter selection step of the iterative approach is illustrated in an extension of Thomas' discrete modelling framework. This extension encodes into multiplexes information about cooperative, concurrent or more complex molecular interactions. We emphasize how formal methods can be helpful to perform the selection step. We express the dynamic knowledge into a temporal logic formula and we confront it, via model checking algorithm, to each model consistent with the static knowledge. In this way, *all* the models consistent with both static and dynamic knowledge are selected. The software platform SMBioNet implements this approach. We illustrate its use with a biological system triggering the tail resorption during the metamorphosis of tadpole.

Keywords: biological regulatory networks, gene interaction, formal methods, model checking, temporal logic

CONTENTS

INTRODUCTION

To study complex biological systems, formal modelling is often mandatory since the complexity of the interleaved interactions between constituents makes intuitive reasoning error prone (De Jong 2002). Several mathematical modelling frameworks have been proposed to model gene regulatory networks (see for example Thomas and d'Ari 1990; Kauffman 1993; Smolen *et al*. 2000; De Jong 2002; Soule 2006; Wilkinson 2006). Common approaches are quantitative, based on differential or stochastic equations (Heinrich and Schuster 1998; Tyson 2002; Goldbeter and Pourquié 2008), and provide numerical simulations of the system. Nevertheless actual predictions often remain only qualitative because the parameter values of these systems are not precisely known (Thomas and Kaufman 2001; Ronen *et al*. 2002; Bernot *et al*. 2004; de Jong and Ropers 2006). Several other modelling frameworks are based on a qualitative view, see for example boolean networks and their generalizations (Kaufman 1969; Thomas 1973; Thomas and Kaufman 2001), Petri nets (Matsuno *et al*. 2000; Matsuno 2003; Chaouiya *et al*. 2004; Comet *et al*. 2005), hybrid modellings (Ahmad *et al*. 2007; Siebert and Bockmayr 2007), and stochastic π-calculus (Ciocchetta and Priami 2007). Each modelling framework highlights some views of models and allows one to detail or to abstract different biological aspects.

Modelling is useful to understand dynamics of the studied system. For a given interaction shape, the set of possible dynamics is huge (Richard *et al*. 2006; Bernot *et al*. 2008). So the search of dynamics coherent with the known dynamic properties is difficult and must be computer aided. In this article, we present an iterative modelling approach based on an exhaustive search of all models coherent with known or hypothetical dynamic properties. This iterative approach is done in discrete Thomas' framework (Thomas and d'Ari 1990) and illustrated with the system triggering the tail resorption during the metamorphosis of the tadpole (Brown and Cai 2007). There are many questions when we study a biological function at the gene interaction level. Which are the dynamics that can be associated with known interactions? What can be deduced from the known interac-

Received: 3 June, 2009. Accepted: 6 October, 2009

Invited Review

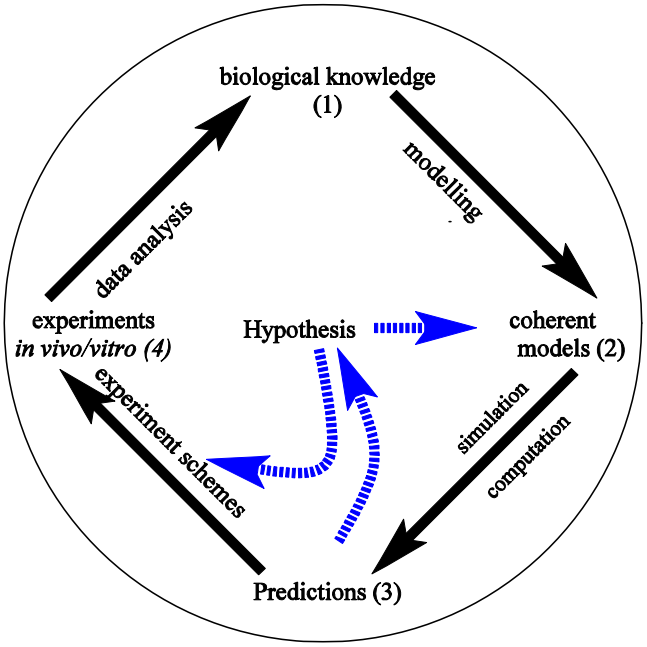

Fig. 1 Iterative approach for modelling complex biological systems. For formalization of biological knowledge (1), all the coherent models (2) are selected. These lead to some predictions (3) *in vivo or in vitro* experiments (4). These experiments enrich the biological knowledge (1).

tions? Are models still coherent with new knowledge? We need an iterative approach that permits us to progressively elaborate models *in silico* and to validate them *in vivo* and/or *in vitro*. We start from a first "coarse" biological knowledge, then we elaborate a first model which leads to predictions that can be confronted to the experimentation. Experimentation, in turn, provides new knowledge and we refine the model accordingly, then we start a new cycle. **Fig. 1** illustrates this approach.

The first step (1) consists in formalizing biological knowledge as much as possible. There are two types of information: Static knowledge and dynamic properties. Static knowledge allows us the elaboration of the regulatory graph. Such a graph is a directed graph composed of two types of vertices. Variables abstract genes and their products, and "multiplexes" abstract actions or combined actions between variables. Dynamic properties consist in information about the observed behaviours of the system. For example, the DNA sequence of genes and promoters can provide static information about genes interactions whilst the existence of homeostasis is a dynamic property.

From a computer science point of view such dynamic information, which is often crucial, has to be formalized (Huth and Ryan 2000) in order to be handled automatically. We encode it in temporal logic formulas (Clarke and Emerson 1981; Huth and Ryan 2000). These formulas can specify stable states, the fact that the system is either multistationary or homeostatic, etc. Once biological knowledge is formalized, the following step (2) in **Fig. 1** consists in selecting models that are coherent with both static knowledge and dynamic properties. Thanks to selected models, predictions and new hypotheses can be proposed (3) and these help us to propose new experiments *in vitro* or *in vivo* (4). Finally from these experiments new biological knowledge is generated (1) and the process can be reiterated.

In the following, the steps (1,2) of this iterative approach of genetic regulatory network modelling are illustrated in Thomas' discrete modelling framework. We also present the tool SMBioNet that implements these steps in the context of Thomas' framework.

THOMAS' MODELLING WITH MULTIPLEXES

This section is devoted to the definition of our discrete modelling framework for regulatory networks. The following notation will be useful.

Notation 1 *Given a directed graph G and a node* v *of* G, G^{-1} (v) *is the set of all nodes* v′ *of* G *such that* (v′, v) *is an edge of* G (*set of predecessors of* v).

The static part of a regulatory network is represented by a directed graph composed of two types of vertices:

- Variables that abstract usually genes and their products (some variables may also abstract global or external conditions, such as presence of sugar, calcium depletion, etc.)
- Multiplexes that abstract combined actions, they represent biological phenomena such as the formation of complex to activate some genes.

The predecessors of a multiplex are either variables or other multiplexes brought into play in the (combined) action; the successors are called the targets of the multiplex.

Predecessors and successors of variables are multiplexes.

A multiplex is provided with a propositional logic formula that encodes the situations in which the interaction occurs. For example, if a complex composed of two proteins a and b is required and if the complex (a-b) is not activated in the presence of a protein c, then the corresponding formula looks like "$a \wedge b \wedge \neg c$," where the symbols "$\wedge$" and "$\neg$" stand for "and" and "not" respectively.

Definition 1 *A* gene regulatory graph with multiplexes*, RG for short, is a tuple* $G = (V, M, E_V, E_M)$ *such that:*

1. $(V \cup M, E_V \cup E_M)$ *constitutes a (labelled) directed graph whose set of nodes is* $V \cup M$ *and set of edges is* $E_V \cup E_M$*, with* $E_V \subset V \times IN \times M$ *and* $E_M \subset M \times (V \cup M)$.
2. V *and* M *are disjoint finite sets. Nodes of* V *are called variables and nodes of* M *are called multiplexes. An edge* (v, s, m) *of* E_V *is denoted* $v^s \rightarrow m$ *and* s *is called the threshold.*
3. *Each variable* v *of* V *is labelled with a positive integer by called the bound of* v.
4. *Each multiplex* m *of* M *is labelled with a formula belonging to the language* L_m *inductively defined by:*

- *If* $v^s \rightarrow m \in E_V$*, then* v_s *is an atom of* L_m*, and if* $(m' \rightarrow m) \in E_M$ *then* m′ *is an atom of* L_m.
- *If* φ *and* ψ *belong to* L_m *then* $\neg\varphi$, $(\varphi \wedge \psi)$, $(\varphi \vee \psi)$ *and* $(\varphi \Rightarrow \Psi)$ *belong to* L_m
- *All cycles of the underlying graph* $(V \cup M, E_V \cup E_M)$ *contain at least one node belonging to* V.

Note: *Condition 5 is necessary for the definition of dynamics (***Def. 3***).*

Fig. 2 provides graphical conventions. In this figure, a, b, c, d are variables; m, m′ are multiplexes; m and c are the predecessors of m′ and b and d are its successors; the cycle b, m, m′ contains the variable b.

A gene regulatory graph with multiplexes constitutes the static representation of the system. We have now to focus on the dynamics of the system, abstracted by the evolutions of expression levels of the variables. Let us first define the states of a system.

Definition 2 *A state of a RG* $G = (V, M, E_V, E_M)$ *is a map* η: $V \rightarrow IN$ *such that for each variable* v *belonging to* V *,* $\eta(v) \leq b_v$ *.* $\eta(v)$ *is called the expression level of* v *at state* η.

A multiplex does not have any expression level because it is just a logical composition of variables at a given state: From the expression level of the variables, we deduce if the multiplex is active or not via the interpretation of its propositional formula.

Definition 3 *Given a RG* $G = (V, M, E_V, E_M)$*, a state* η *of* G *and a multiplex* $m \in M$*, we say that* m *is active at state* η *iff the formula* φ_m *of* m *is satisfied at state* η*; the interpretation of* φ_m *at state* η *being inductively defined by:*

- *If* φ_m *is reduced to an atom* v_s *of* $G^{-1}(m)$ *then* φ_m *is satisfied iff* $\eta(v) \geq s$.

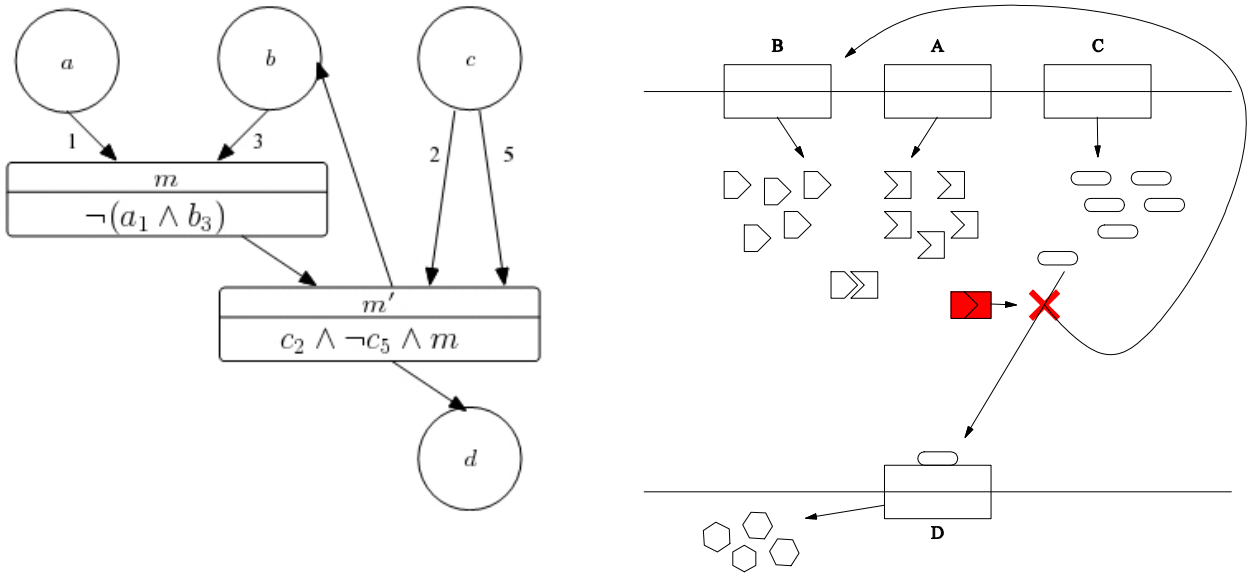

Fig. 2 Graphical conventions. This is a RG with 4 variables a,b,c, and d and 2 multiplexes m,m'. *c* activates both *b* and *d* when its concentration level is 2,3 or 4. The complex *a*-*b* inhibits activations.

- *If* φ_m *is reduced to an atom* $m' \in M$ *of* $G^{-1}(m)$ *then* φ_m *is satisfied iff* $\varphi_{m'}$ *is satisfied at state* η.
- *If* $\varphi_m \equiv \psi_1 \wedge \psi_2$ *then* φ_m *is satisfied iff* ψ_1 *and* ψ_2 *are satisfied; and we proceed similarly for all other connectives.*

An active multiplex helps its targets to increase their expression levels. If *a* is the target of an active multiplex *m*, we then say that *m* is a resource of *a*.

Notation 2 *We note* $\rho(v, \eta)$ *the set of resources of* v *at state* η*:* $\rho(v, \eta) = \{m \in G^{-1}(v) \mid \varphi_m$ *is satisfied at state* $\eta\}$.

Contrarily to the original framework of René Thomas, edges of regulatory graphs have no sign but negative actions are taken into account through multiplexes with the operator ¬. For example, in **Fig. 2** the multiplex *m* represents an inhibition (the complex *a*-*b* inhibits *b* and *d* via *m'*). In **Fig. 2**, we also see that in multiplex formulas the variables are indexed by their thresholds. This is useful when a given variable participates to a multiplex at several thresholds. The multiplex formula of *m'* means that the expression level of *c* must be both greater than 2 and lower than 5 in order to participate to the induction of *d*.

In addition to these standard graphical conventions, we simplify these conventions:

- If a variable is an input of a multiplex with only one threshold, the threshold is not necessary in the formula. For example, in **Fig. 2**, the formula of multiplex *m* can be simply written *"¬(a ∧ b)."* Of course, this light form is not possible for *m'*.
- Multiplexes, the formula of which is reduced to a unique atom, can be removed from the diagram. In **Fig. 3a**, removing the multiplex *m* allows us to retrieve the usual diagrammatic convention of R. Thomas for activations.

Similarly, in **Fig. 3b**, we retrieve usual inhibitions, either by adding the minus sign, or by using the "inhibition arrow" usual in biology.

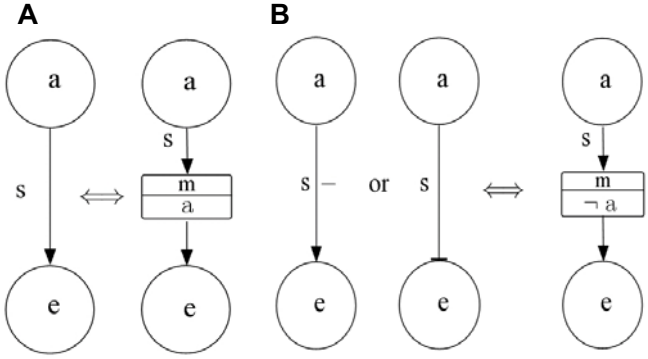

Fig. 3 Light graphical simplification for activations and inhibitions. There exists two ways to represent an activaction (**A**) and three ways to represent an inhibition (**B**).

TAIL RESORPTION IN TADPOLE METAMORPHOSIS

We illustrate this modelling framework on the biological system triggering the tail resorption during the metamorphosis of tadpole.

Anuran metamorphosis is an intense period of post embryonic development that is characterized by the extensive remodelling of the tadpole into a juvenile frog (Shi 2000; Veldhoen *et al.* 2002).

The metamorphosis of the tadpole starts with limb growth and differentiation and ends with tail and gill resorption. The organs like muscle, skin, intestine, pancreas, liver, brain, etc. are remodelled. During metamorphosis, thyroid hormones (*TH*) play a crucial role for these developmental changes in particular for the tail resorption. *TH* bring into play apoptotic mechanisms (Troncale *et al.* 2007) controlling the cellular death. We illustrate our modelling framework on a simplified model of the system regulating the tail resorption.

From biological knowledge, we establish the regulatory graph. Among the variables of the regulatory graph, two types of thyroid hormones (*TH*) have to be taken into account: The tri-iodothyronine (*T3*) and the thyroxine (*T4*) (Brown and Cai 2007). These hormones are regulated by two enzymes: Deiodinase of type 2 (*D2*) and deiodinase of type 3 (*D3*) that have different roles in the system. We also consider intermediate genes (*IG* for short) thatare activators of D2. Early genes (*EG*) and late genes (*LG*) are both responsible for tail resorption via apoptosis. Notice that *LG* are expressed after *EG*. Finally, the nuclear thyroid hormone receptor of *T3* isoform β, denoted TR_β, is explicitly represented even if it is an early gene because it has an important role (Wang and Brown 1993; Troncale 2007). In this system, *D2* is a catalyst that allows the transformation of *T4* into *T3* (Huang *et al.* 2001): *T4* loses iodine under the action of *D2*. So, the synthesis of *T3* from *T4* by *D2* does not consume *D2*. This catalysis is modelled, in the **Fig. 4**, by the multiplex named catalysis which is labelled by the formula $T4_1 \wedge D2_1$ acting on *T3*.

Similarly, *D3* catalyses the transformations of *T3* and *T4* into inactive forms of *TH* (these inactive forms are not

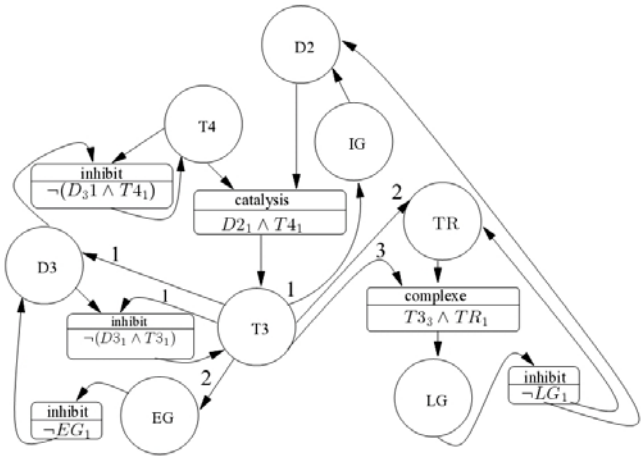

Fig. 4 RG of the TH-dependent regulation of the tadpole tail resorption. From biological knowledge, we establish this RG. The variables of the regulatory graph are the tri-iodothyronine (*T3*) and the thyroxine (*T4*). These hormones are regulated by two enzymes: Deiodinase of type 2 (*D2*) and deiodinase of type 3 (*D3*). We also consider intermediate genes (*IG*), early genes (*EG*) and late genes (*LG*). Finally, the nuclear thyroid hormone receptor of *T3* isoform β, denoted TR_β.

represented in this model). The catalysis of *D2* on *T4* does consume *T4* but less than the quantity brought by the sanguine flux.

The action of *D3* on *T4* and *T3* decreases significantly their concentrations. These catalyses are then modelled by multiplexes labelled by the formulas $\neg(T3_1 \wedge D3_1)$ and $\neg(T4_1 \wedge D3_1)$ acting respectively on *T3* and *T4*. Notice that these inhibitions are represented by the operator of the negation in the formulas. *TH* bind with its nuclear receptor TR_β to form complexes *TH/TR* which are transcription factors of the genes responsible for cellular death. We transcript this information into the multiplex labelled by the formula $T3_3 \wedge TR_1$ that acts on *LG*. **Fig. 4** represents the RG deduced from the previous information.

FROM REGULATORY GRAPHS TO REGULATORY NETWORKS

We call *network* a graph associated with the parameters that determine the dynamics.

Definition 4 *A gene regulatory network with multiplexes (RN for short) is a couple (G, K) where*

- $G = (V, M, E_V, E_M)$ *is a RG.*
- $\kappa = \{k_{v,\omega}\}$ *is a family of parameters indexed by* $v \in V$ *and* $\omega \in G^{-1}(v)$ *such that all* $k_{v,\omega}$ *are integers and* $0 \leq k_{v,\omega} \leq b_v$.
- *If* $\omega \subset \omega'$ *then* $k_{v,\omega} \leq k_{v,\omega'}$

Note that each variable *v* admits 2^n parameters of the form $k_{v,\omega}$ where *n* is the in-degree of *v* in *G*, and that we can always ignore the parameters $k_{v,\omega}$ such that the conjunction of the formulas associated with the multiplexes in ω is unsatisfied for all states.

The value of the parameter $k_{v,\rho(v,\eta)}$ (where ρ is defined in **Def. 3**), indicates how the expression level of *v* can evolve at state η. It can increase (respectively decrease) if the parameter value is greater (respectively less) than $\eta(v)$. The expression level must stay constant if both values are equal. The third item of **Def. 4** expresses that the decrease of the expression level of v cannot be induce by an effective resource of v. The tendency (increasing, decreasing, unchanging) of variables are given by the directional map associated with each state:

Notation 3 *Given a RN* $N = (G, K)$ *and a state* η *of* $G = (V, M, E_V, E_M)$*, the directional map* $d: V \rightarrow \{-1, 0, 1\}$ *is defined by:*

$$\forall v \in V,\ d(v) = \begin{cases} -1 & \text{if } \eta(v) > k_{v,\rho(v,\eta)} \\ 0 & \text{if } \eta(v) = k_{v,\rho(v,\eta)} \\ 1 & \text{if } \eta(v) < k_{v,\rho(v,\eta)} \end{cases}$$

The probability that two variables change their expression level at the same time is negligible *in vivo*; following the Thomas' approach a state transition of the model involves only one variable at a time. This procedure is called *asynchronous update in Thomas' framework.*

Definition 5 *Let* $N = (G, K)$ *be a RN, and let* η *be a state of* G. *A state* η' *of* G *is a successor of the state* η *if and only if:*

- *There exists a variable* v *such that* $\eta'(v) = \eta(v) + d(v)$ *and* $d(v) \neq 0$
- *For any other variable* $u \neq v$ *we have* $\eta'(u) = \eta(u)$

In each state transition, one variable is changed;

Definition 6 *The asynchronous state graph of a RN* $N = (G, \kappa)$ *is the graph* S *defined by:*

- *The set of vertices of* S *is the set of possible states of G (isomorphic to the Cartesian product* $\Pi_{v \in V}\ [0, b_v]$*).*
- *The set of edges of S is the set of couples* (η, η') *such that* η' *is a successor of* η.

Example *For lack of space, let us focus on the sub-graph of* **Fig. 5a**.

Let us assume for instance that parameter values are the following: $k_{D3} = 0$, $k_{D3,T3} = 1$, $k_{T3} = 0$, and $k_{T3,inhibit} = 1$. To build the asynchronous state graph, we first construct the table associating with each state, the directional map (see **Table 1**).

Then previous definition allows us to construct the asynchronous state graph (see **Fig. 5b**).

In Thomas' modelling framework with multiplexes, we have to give a value to each parameter in order to deduce the dynamics of the system. Because parameter values are not *a priori* known this leads us to consider by default an enormous number of parameterizations. Indeed, each variable *v* admits 2^n parameters of the form $k_{v,\omega}$ where *n* is the in-degree of *v* in *G* ($\omega \in G^{-1}(v)$). Each of these parameters can take $b_v + 1$ different values (b_v is the bound of *v*). The number of parameterizations is consequently $\Pi_{v \in V}(b_v+1)^{2n}$ where *n* is the in-degree of *v*. For the *TH*-dependent regulation of the tadpole tail resorption, the number of parameterizations is on the order of 2.6×10^8 so we need a tool for selecting interesting parameterizations. In reality, it is less than that because here we do not take account of item 3 of **Def. 4** which corresponds to Snoussi's conditions (Snoussi 1998). Even if we consider Snoussi's conditions, the number of parameterizations still huge so we develop the software SMBioNet.

Table 1 Construction of the asynchronous state graph **Fig. 5B** for the interaction graph of **Fig. 5A**. Columns D3 and T3 give the different levels of concentration to consider. Columns $k_{D3,\omega}$ and $k_{T3,\omega}$: according the levels of D3 and T3, the parameter k determine the evolution (increase, decrease, stable) of D3 and T3, and columns d(D3) and d(T3) determine the tendency.

D3	T3	$k_{D3,\omega}$	$k_{T3,\omega}$	d(D3)	d(T3)
0	0	$k_{D3,\varnothing} = 0$	$k_{T3,inhibit} = 1$	0	1
0	1	$k_{D3,T3} = 1$	$k_{T3,inhibit} = 1$	1	0
1	0	$k_{D3,\varnothing} = 0$	$k_{T3,inhibit} = 1$	−1	1
1	1	$k_{D3,T3} = 1$	$k_{T3,\varnothing} = 0$	0	-1

COMPUTATIONAL TREE LOGIC AND MODEL CHECKING

For using SMBioNet, dynamic knowledge on the biological system to model has to be translated into a formal language interpretable by a computer. For SMBioNet, this formal language is a classical temporal logic called Computational Tree Logic (CTL) (Pérés and Comet 2003).

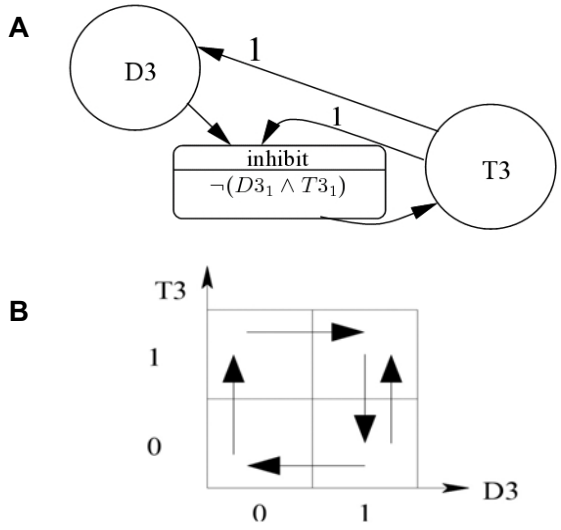

Fig. 5 A sub-graph of the graph of Fig. 4 and its asynchronous state graph according to given parameters (see text). (**A**) A RG. (**B**) An asynchronous state graph.

CTL is well suited for the formulation of properties present in indeterministic state graphs, such as the asynchronous state graphs considered here (a state graph is indeterministic if some states have several successors). It permits us to express, for example, that some events occur before other ones, that a specific event has to take place in order to reach a given state, or that it is impossible to reach another state or that an event is always possible.

Once the known biological properties are expressed into CTL, we can check the consistency between a dynamical model and these properties via model checking (Kwiatkowska 2003). In our case, model checking takes as input an asynchronous state graph and a set of temporal properties and returns yes (no) if the properties are (are not) satisfied by the state graph.

SMBIONET

In the context of R. Thomas' modelling, SMBioNet (see **Fig. 6**) allows one to select the models that are coherent with the regulatory graph and the dynamic properties expressed in CTL. For each parameterization, SMBioNet constructs the corresponding asynchronous state graph and check if the CTL temporal formula is satisfied by this state graph. This verification step is performed by the model checker NuSMV (Cimatti *et al.* 2002).

Only parameterizations leading to dynamics coherent with the behavioural properties are retained. If none of them are retained it is necessary to reassess either the regulatory graph (it can be too simple to be able to lead to a state graph that expresses the specified properties) or the temporal properties.

Input file for the tadpole example

SMBioNet is fed through an input file that is divided into 4 parts, called VAR, REG, PARA and CTL. The parts VAR and REG define the regulatory graph. The (optional) part PARA constraints the value of some selected parameters, and the part CTL describes the temporal properties of the system. In the following, we illustrate these four parts with the TH-dependent regulation of the tadpole tail resorption.

Part VAR defines the set of variables (V) with their associated bounds. For the TH-dependent regulation of the tadpole tail resorption, the part VAR is the following:

```
VAR
T3 = 0 3 ;
T4 = 0 1 ;
d2 = 0 1 ;
d3 = 0 1 ;
gi = 0 1 ;
gp = 0 1 ;
gt = 0 1 ;
trb = 0 1 ;
```

There are 8 variables: *T3* evolves in the interval {0, 1, 2, 3}, and all other variables evolve into the interval {0, 1}. So *T3* has four possible expression levels, and the other variables have two possible expression levels. The number of states is thus $4 \times 2^7 = 512$.

Part REG allows the definition of the set of multiplexes (M), and the sets of edges (E_V and E_M). For instance, the multiplexes of the TH-model are described in the following way:

```
REG
catalyse1 [((d2>=1)&(T4>=1))]=> T3 ;
catalyse3 [((d3<1)&(T3>=1))]=> T3 ;
catalyse2 [((d3<1)&(T4>=1))]=> T4 ;
factransc [((T3>=3)&(trb>=1))]=> gt ;
acti [(gi>=1)]=> d2 ;
inhi [(gt<1)]=> d2 trb ;
acti [(T3>=1)]=> d3 gi ;
inhi [(gp<1)]=> d3 ;
acti [(T3>=2)]=> gp trb ;
```

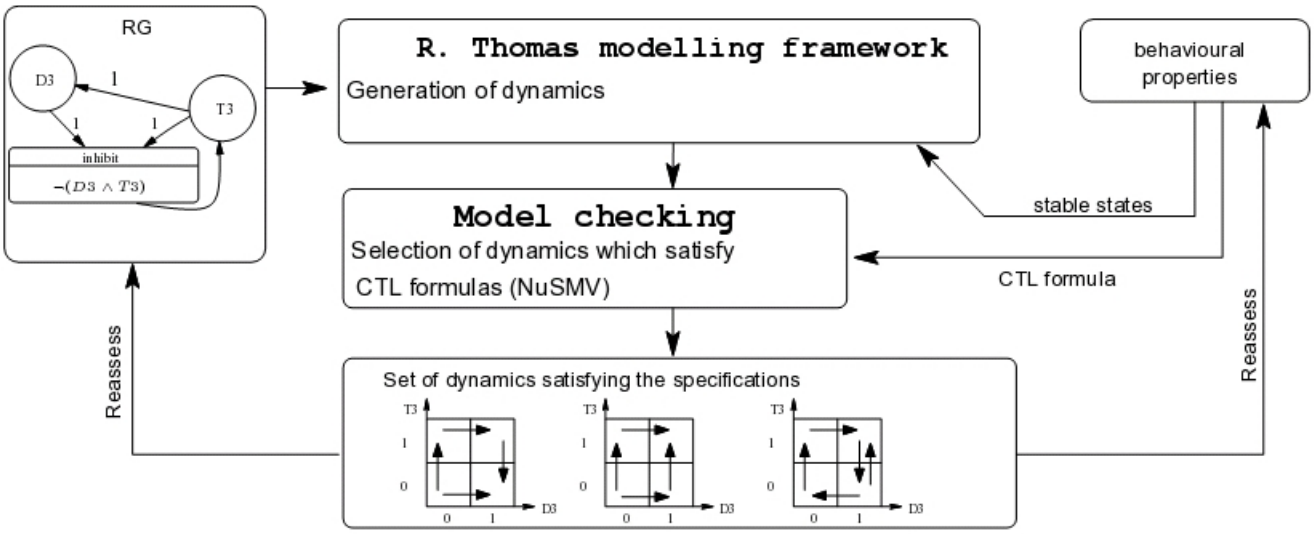

Fig. 6 SMBioNet. From a genetic regulatory graph and some behavioural properties, the tool SMBioNet is able to enumerate each possible parametricization and to select only those which lead to a state graph which is coherent with temporal properties.

This code indicates, for instance, that *catalyse1* is a multiplex whose the associated formula is *((d2 >= 1)* and *(T4 >= 1))* and whose unique target is *T3*.

Part PARA permits us to constrain parameter values. For example, if we want to constrain the parameter $k_{T3}(\{\varnothing\})$ in the integer interval [0, 2], we write:

```
PARA
K_T3 = 0 2 ;
```

Parameter constraints are important because they reduce the number of all possible parameterizations to consider.

Part CTL contains formulas expressing known dynamical properties (syntax and semantics of CTL formulas are given in the appendix).

We show in this section that the RG introduced in section Tail resorption in tadpole metamorphosis is able to explain the observed variation of the TH hormones (Leloup and Buscaglia 1977). It means that there should exist at least one parameterization for which the model exhibits the successive increasing and decreasing phases.

So, for the TH-dependent regulation of the tadpole tail resorption, we know that there is growing phase during which the concentrations of *T4* and *T3* increase, and *D2* is so that *GI* are not activated (Leloup and Buscaglia 1977). We know also that, during this growing phase, *D3* is present and *GP*, *TR* and *GT* are absent. The end of this growing phase is called the climax: the concentrations of *T4* and *T3* are at their maximum, *D2* is over-regulated so *GI* are activated, *D3* is under-regulated, *GP* and *TR* are present, and *GT* are absent. This information can be translated into this temporal formula: *(T4=0 & T3=0 & D2=0 & D3=1 & GI=0 & GP=0 & TR=0 & GT=0) → AF (T3=3 & T4=1 & D2=1 & D3=0 & GI=1 & GP=1 & TR=1 & GT=0)*

From the climax, we attend the resorption of tadpole tail that corresponds to the presence of *GT*. This information can be translated into this temporal formula:

(T3=3 & T4=1 & D2=1 & D3=0 & GI=1 & GP=1 & TR=1 & GT=0) → (AF(GT=1))

The CTL part of the input file with the conjunction of the two previous formulas is:

```
CTL
(T4=0 & T3=0 & D2=0 & D3=1 & GI=0 & GP=0 & TR=0 & GT=0) → AF (T3=3 & T4=1 & D2=1 & D3=0 & GI=1 & GP=1 & TR=1 & GT=0) & (T3=3 & T4=1 & D2=1 & D3=0 & GI=1 & GP=1 & TR=1 & GT=0) → (AF(GT=1))
```

Output file for the tadpole example

The output file contains all the parameterizations of the regulatory graph leading to dynamics verifying the CTL formulas (and that are consistent with the constraints on parameters). For the previous described input file, the output file contains 18 parameterizations that are by construction fully relevant. This means that the established regulatory graph is consistent with the known properties of the system at your knowledge[1]. These 18 models are coherent with biological knowledge at disposal in the literature. One of them is:

```
K_T3 = 1
K_T3+catalyse1 = 1
K_T3+catalyse3 = 2
K_T3+catalyse1+catalyse3 = 3
K_T4 = 1
K_T4+catalyse2 = 1
K_d2 = 0
K_d2+acti = 0
K_d2+inhi = 0
K_d2+acti+inhi = 1
K_d3 = 0
K_d3+acti = 0
K_d3+inhi = 0
K_d3+acti+inhi = 0
K_gi = 0
K_gi+acti = 1
K_gp = 0
K_gp+acti = 1
K_gt = 0
K_gt+factransc = 0
K_trb = 0
K_trb+acti = 0
K_trb+inhi = 0
K_trb+acti+inhi = 1
```

If we had found no parameterization at all, it would definitely have proved that RG of section Tail resorption in tadpole metamorphosis is inconsistent and the RG has to be recalling in question.

Because we have more than one possible parameterization (18) several models are compatible with biological information. According to the biological objectives:

- If the final motivation is to check the consistency of the interaction graph, the methodological iteration can be stop.
- If on the contrary, we want to refine the model we have to design a new discrimination property which is not satisfied by all models.
- If we want to explore further the biological system then we have to take into account new biological knowledge or biological hypotheses. These new knowledge or hypotheses will had been translated into temporal logic and then the methodological iteration will restart. Notice that the knowledge of the 18 models can be used to design relevant biological experiment in order to suggest these new biological hypotheses.

CONCLUSION

When René Thomas, in the 1970's, introduced the idea of "logical" modelling of gene regulatory networks, it had a great impact on the understanding of cell behaviours. The advantage of this approach is its qualitative nature that allows us to better handle the complexity of such networks. When we have formalized the approach into a computer science discrete framework, we have introduced the idea of using temporal logic in order to automatically and exhaustively extract all the sensible parameter values of the models (Bernot *et al.* 2004).

This framework has been implemented in a first version of SMBioNet and we have been able to treat bigger networks. Nevertheless, the needed computation time of SMBioNet asks for modelling methods that better encodes biological knowledge in order to drive our algorithms. For this reason, we have introduced multiplexes and this article has shown how they are defined and how it works on an example. Let us emphasize the complexity that is inherent to the parameter selection, whatever the underlying modelling framework. We believe that our iterative way to introduce logical methods in discrete Thomas' approach provides an elegant solution.

The encoding of biological knowledge is crucial because it allows us to always manage the exhaustive set of parameters compatible with known properties. Consequently, we provide the biologists with incremental, iterative, approach where additional behavioural information is easily taken into account because it simply reduces the exhaustive set of pertinent parameters. The software platform SMBioNet implements this parameter set management and uses efficient model checking algorithms.

Mainly, SMBioNet exhaustively enumerates in an optimal manner the successive parameter sets in order to treat the corresponding state graphs. Multiplexes have the advantage to considerably reduce the number of parameter sets to consider, but they also have the advantage to describe them with symbolic formulas that could be used to express con-

[1] The authors want to acknowledge Nicolas Pollet

straints. So, our future works tend to introduce symbolic constraint manipulations and solving methods into SMBioNet, which, with respect to model checking, would constitute an efficient complementary framework.

ACKNOWLEDGEMENTS

This work was supported by grants from Région Ile-de-France. The authors thank to the members of the observability working group of Genopole for helpful discussions on the Tadpole tail resorption model, especially N. Pollet.

APPENDIX

The syntax of ctl formulas is:

```
ctl := bool_expr        |
       "!" ctl          |
       ctl "&" ctl      |
       ctl "|" ctl      |
       ctl "->" ctl     |
       "EX" ctl         |        # Exists neXt state
       "AX" ctl         |        # for All neXt state
       "EF" ctl         |        # Exists Finally
       "AF" ctl         |        # for All Finally
       "EG" ctl         |        # Exists Globally
       "AG" ctl         |        # for All Globally
       "E" ctl "U" ctl  |        # Exists (ctl) Until (ctl)
       "A" ctl "U" ctl  |        # for All (ctl) Until (ctl)

bool_expr := atome | "!" bool_expr | bool_expr "&" bool_expr | bool_expr "|" bool_expr |
atome := id "<" id    |
         id ">" id    |
         id "<=" id   |
         id ">=" id   |
         id "=" id

   id := variable_name | integer
```

The semantic of ctl formulas is defined as follows. A state η of the state graph satisfies a formula like:

- EX ctl if η admits a successor that satisfies the ctl formula.
- AX ctl if all successors of η satisfy the ctl formula.
- EF ctl if there is a path from η that goes through a state satisfying the ctl formula.
- AF ctl if all elementary paths of maximum length from η go through a state satisfying the ctl formula (a path is elementary when it does not pass several times through the same states).
- E ctl1 U ctl2 if there is a path starting from η that goes through a state η′ such as: η′ satisfies the formula ctl2, and the states of the path located between η and η′ excluded satisfies the formula ctl1.
- A ctl1 U ctl2 if all elementary paths of maximum length starting from η go through a state η′ such as: η′ satisfies the formula ctl2, and the states of the path located between η and η ′ excluded satisfies the formula ctl1.

The interpretation of formulas like !ctl, ctl & ctl, ctl | ctl and ctl -> ctl is usual. A formula is satisfied by an asynchronous state graph if the formula is satisfied by all the states of this state graph.

REFERENCES

Ahmad J, Bernot G, Comet J-P, Lime D, Roux O (2006 Cover Date: November 2007) Hybrid modelling and dynamical analysis of gene regulatory networks with delays. *ComPlexUs* **3**, 231-251

Bernot G, Comet J-P, Richard A, Guespin J (2004) Application of formal methods to biological regulatory networks: Extending Thomas' asynchronous logical approach with temporal logic. *Journal of Theoretical Biology* **229 (3)**, 339-347

Bernot G, Comet J-P, Khalis Z (2008) Gene regulatory networks with multiplexes. In: *Proceedings of European Simulation and Modelling Conference*, ESM'2008, October 27-29, Le Havre, France, pp 423-432

Brown DD, Cai L (2007) Amphibian metamorphosis. *Developmental Biology* **306**, 20-33

Chaouiya C, Remy E, Ruet P, Thieffry D (2004) Qualitative modelling of genetic networks: From logical regulatory graphs to standard Petri nets. In: *ICATPN 2004*, LNCS 3099, Springer-Verlag, pp 137-156

Cimatti A, Clarke EM, Giunciglia E F, Pistore M, Roveri M, Sebastiani R, Tacchella A (2002) Nusmv 2: An open source tool for symbolic model checking. In: *Proceeding of International Conference on Computer-Aided Verification(CAV 2002)*, pp 27-31

Ciocchetta F, Priami C (2007) Biological transactions for quantitative models. *Electronic Notes in Theoretical Computer Science* **171 (2)**, 55-67

Clarke EM, Emerson EA (1981) Design and syntheses of synchronization skeletons using branching time temporal logic. In: *Logics of Programs Workshop*, LNCS 131, Springer, pp 52-71

Comet J-P, Klaudel H, Liauzu S (2005) Modeling multi-valued genetic regulatory networks using high-level Petri nets. In: *ICATPN 2005*, LNCS 3536, pp 208-227

de Jong H (2002) Modeling and simulation of genetic regulatory systems: a literature review. *Journal of Computational Biology* **9 (1)**, 67-103

de Jong H, Ropers D (2006) Strategies for dealing with incomplete information in the modeling of molecular interaction. *Briefings in Bioinformatics* **7 (4)**, 354-363

Fujita S, Matsui M, Matsuno H, Miyano A (2004) Modeling and simulation of fission yeast cell cycle on hybrid functional Petri net. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences* **E87-A (11)**, 2919-2928

Goldbeter A, Pourquié O (2008) Modeling the segmentation clock as a network of coupled oscillations in the Notch, Wnt and FGF signaling pathways. Journal of Theoretical Biology **252**, 574-585

Heinrich R, Schuster S (1998) The modelling of metabolic systems. Structure, control and optimality. *Biosystems* **47 (1-2)**, 61-77

Huang H, Cai L, Remo BF, Brown DD (2001) Timing of metamorphosis and the onset of the negative feedback loop between the thyroid gland and the pituitary is controlled by type II iodothyronine deiodinase in *Xenopus laevis*. *Proceedings of the National Academy of Sciences* **98 (13)**, 7348-7353

Huth M, Ryan M (2000) *Logic in Computer Science: Modelling and Reasoning About Systems*, Cambridge University Press, 427 pp

Leloup J, Buscaglia M (1977) La triiodothyronine, hormone de la métamorphose des amphibiens. *Comptes Rendus de l'Académie des Sciences* **284**, 2261-2263

Kauffman S (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology* **22**, 437-467

Kauffman S (1993) *Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford, 740 pp

Kwiatkowska M (2003) Model checking for probability and time: from theory to practice. In: *Proceedings of the 18th Annual IEEE Symposium on Logic in Computer Science*, LICS'03 22-25 June 2003, Ottawa, Canada, IEEE Computer Society Press, pp 351-360

Matsuno H, Doi A, Nagasaki M, Miyano S (2000) Hybrid Petri net representation of gene regulatory network. In: *Proceedings of the Pacific Symposium on Biocomputing 2000*, 4-9 January 2000, Honolulu, Hawaii, pp 341-352

Pérès S, Comet J-P (2003) Contribution of computational tree logic to biological regulatory networks: example from *Pseudomonas æruginosa*. In: *International Workshop on Computational Methods in Systems Biology*, LNCS 2602, February 24-26, 2003, Rovereto, Italy, pp 47-56

Richard A, Comet J-P, Bernot G (2006) Formal methods for modeling biological regulatory networks. In: Gabbar, Hossam A (Eds) *Modern Formal Methods and Applications*, Springer, The Netherlands, pp 83-122

Ronen M, Rosenberg R, Shraiman BI, Alon U (2002) Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. *Proceeding of the National Academy of Sciences USA* **99 (16)**, 10555-1560

Shi Y-B (2000) Amphibian metamorphosis: From morphology to molecular biology. *Cell Biology International* **24 (12)**, 909

Siebert H, Bockmayr A (2007) Context sensitivity in logical modelling with time delays. In: *Computational Methods in Systems Biology 2007*, LNBI 4695, Springer, Berlin, pp 64-79

Smolen P, Baxter DA, Byrne JH (2000) Modeling transcriptional control in gene networks: Methods, recent results, and future directions. *Bulletin of Mathematical Biology* **62 (2)**, 247-292

Snoussi EH (1998) Necessary conditions for multistationarity and stable periodicity. *Journal of Biological Systems* **6**, 3-9

Soulé C (2006) Mathematical approaches to differentiation and gene regulation. *Comptes Rendus Biologies* **329**, 13-20

Thomas R (1973) Boolean formalization of genetic control circuits. *Journal of Theoretical Biology* **42**, 563-585

Thomas R, d'Ari R (1990) *Biological Feedback*, CRC Press, Boca Raton, 316 pp

Thomas R, Thieffry D, Kaufman M (1995) Dynamical behaviour of biological regulatory networks. *Bulletin of Mathematical Biology* **57 (2)**, 247-276

Thomas R, Kaufman M (2001) Multistationarity, the basis of cell differentiation and memory. II. Logical analysis of regulatory networks in terms of feedback. *Chaos* **11**, 80-195

Troncale S, Thuret R, Ben C, Pollet N, Comet J-P, Bernot G (2007) Modelling of the TH-dependent regulation of tadpole tail resorption. *Journal of*

Biological Physics and Chemistry **7 (2)**, 45-50

Tyson JJ, Csikasz-Nagy A, Novak B (2002) The dynamics of cell cycle regulation. *Bioessays* **24 (12)**, 1095-1109

Veldhoen N, Crump D, Werry K, Helbing CC (2002) Distinctive gene profiles occur at key points during natural metamorphosis in the *Xenopus laevis* tadpole tail. *Developmental Dynamics* **225**, 457-468

Wang Z, Brown DD (1993) Thyroid hormone-induced gene expression program for amphibian tail resorption. *Journal of Biological Chemistry* **268 (22)**, 16270-16278

Wilkinson DJ (2006) *Stochastic Modelling for Systems Biology*, CRC press, Boca Raton, 288 pp