

Qualitative Petri Net Modelling of Genetic Networks

Claudine Chaouiya¹, Elisabeth Remy², and Denis Thieffry¹

¹ Institut de biologie du Développement de Marseille Luminy
UMR 6216, Case 907 - Luminy, 13288 Marseille Cedex 9 - France
{chaouiya, thieffry}@ibdm.univ-mrs.fr

² Institut de Mathématiques de Luminy,
UMR 6206, case 907 - Luminy, 13288 Marseille Cedex 9 - France
remy@iml.univ-mrs.fr

Abstract. The complexity of biological regulatory networks calls for the development of proper mathematical methods to model their structures and to give insight in their dynamical behaviours. One qualitative approach consists in modelling regulatory networks in terms of logical equations (using either Boolean or multi-level discretisation). Petri Nets (PNs) offer a complementary framework to analyse large systems.

In this paper, we propose to articulate the logical approach with PNs. We first revisit the definition of a rigorous and systematic mapping of multi-level logical regulatory models into specific standard PNs, called *Multi-level Regulatory Petri Nets* (MRPNs). In particular, we consider the case of multiple arcs representing different regulatory effects from the same source. We further propose a mapping of multi-level logical regulatory models into Coloured PNs, called *Coloured Regulatory Petri Nets* (CRPNs). These CRPNs provide an intuitive graphical representation of regulatory networks, relatively easy to grasp.

Finally, we present the PN translation and the analysis of a multi-level logical model of the core regulatory network controlling the differentiation of T-helper lymphocytes into Th1 and Th2 types.

1 Introduction

Most biological processes are spatially and temporally regulated by networks of interactions between regulatory products and genes. To cope with the complexity and characterise the dynamical properties of these *genetic (regulatory) networks*, various formal approaches have been proposed (for a review, see [6]). The lack of precise, quantitative information about the shape of regulatory functions or about the values of involved parameters pleads for the development of qualitative approaches. One qualitative approach consists in modelling regulatory networks in terms of logical equations (using a Boolean or multi-level discretisation, [8,21]). The development of logical models for various biological networks has already induced interesting relationships between topological network features and specific dynamical properties (*e.g.* the crucial roles of regulatory feedback circuits [22]).

The Petri net (PN) formalism offers a complementary framework to analyse the dynamical properties of concurrent systems, either from a qualitative or a

quantitative point of view (see [17] for an extensive introduction to PNs). Petri nets have been successfully applied to the modelling and the analysis of metabolic and signal transduction networks (*e.g.* [18,13,12]). As emphasised in [23], one can draw extensive relationships between the traditional biochemical modelling and the PN theory. In particular, the stoichiometry matrix of a metabolic network corresponds to the PN incidence matrix. In this context, PN invariants are associated to conservation relations and flux modes (for a review of PN modelling of biochemical networks, see [11]).

Published PN models of genetic networks are mostly quantitative models written for particular systems, detailing the mechanisms related to transcription and translation (as, for example in [15], where the authors employ hybrid PNs). Logical regulatory networks consider a higher level of abstraction where the semantics associated with the interactions between components varies, and regulatory products are not consumed during the regulatory processes. We have previously introduced systematic procedures to obtain standard PN models from logical regulatory graphs (see [3] for the Boolean case, and [4] for the multi-level case with no multiple arcs). In [5], Comet *et al.* also proposed a rewriting of logical regulatory graphs into Coloured PNs comprising just one place and one transition. Their objective was to provide a mean to automatically generate correct sets of logical parameters, given the topology of a regulatory graph and temporal logic formulae expressing observed behaviours of the biological system.

Our proposal of a PN framework (CRPNs and MRPNs) for the modelling of regulatory networks opens the way for the use of standard PN tools, including model checking techniques, to identify interesting dynamical properties, or to confront models with available dynamical data (*e.g.* temporal gene expression profiles).

In this paper, after recalling the logical formalism basis, we reassess the mapping of multi-level logical regulatory models into standard PNs called Multi-level Regulatory Petri Nets (MRPNs). Here, the main novelty consists in taking into account non-monotonous regulatory effects through multiple arcs. Indeed, this situation can arise in regulatory networks as demonstrated in Section 4. Next, we briefly define a Coloured PN representation of logical regulatory graphs. The resulting Coloured Regulatory Petri Nets (CRPNs) constitute an intuitive graphical representation of regulatory networks, relatively easy to grasp by biologists. In Section 4, we illustrate our approach with the PN representation and the analysis of a multi-level logical model of the core regulatory network controlling the differentiation of T-helper lymphocytes into Th1 and Th2 types.

2 Logical Regulatory Graphs

In this section, we revisit the definition of logical regulatory graphs, introducing some additional notations which will be useful for the PN rewriting (see [21,22,2] for further detail on the logical formalism).

A *regulatory graph* is a labelled directed graph where nodes represent genes (or, more generally, regulatory components) and arcs (directed edges) represent interactions between genes. Let $\mathcal{G} = \{g_1, \dots, g_n\}$ be the set of genes (or

nodes of the regulatory graph). For each $g_i \in \mathcal{G}$, we define its *maximum expression level* Max_i ($\text{Max}_i \in \mathbb{N}^*$), and x_i denotes its current expression level ($x_i \in \{0, 1, \dots, \text{Max}_i\}$).

For each gene g_i , $\text{Reg}(i)$ denotes the set of its regulators, *i.e.*: $g_j \in \text{Reg}(i)$ iff there exists an interaction from g_j to g_i in the regulatory graph. Note that a regulatory graph may contain self-regulated genes.

Depending on its levels of expression, a gene $g_j \in \text{Reg}(i)$ may have distinct regulatory effects on gene g_i . This situation is represented by multiple arcs joining g_j to g_i . Therefore, the specification of an interaction comprises (in addition to its source and target) an interval of integers included in $[1, \text{Max}_j]$ defining the range of the levels of the source for which the interaction is *operating*. Consequently, a couple of integers $(S_i(j), m_i(j))$ is associated to each $g_j \in \text{Reg}(i)$, where:

- $S_i(j)$ is the lowest expression level for which g_j has a regulatory effect upon g_i . It verifies: $0 < S_i(j) \leq \text{Max}_j$;
- $m_i(j)$ is the multiplicity of the arc from g_j to g_i , *i.e.* the number of interactions from g_j towards g_i . It verifies: $0 < m_i(j) \leq \text{Max}_j - S_i(j) + 1$.

Summing-up, if g_j is a regulator of g_i , either it regulates g_i through a unique interaction ($m_i(j) = 1$), or through several interactions ($m_i(j) > 1$). In this case, a multiple arc connects g_j to g_i , composed by $m_i(j)$ simple arcs labelled by integer intervals $[s_k, s'_k]$, $k = 1, \dots, m_i(j)$, with,

- $s_1 = S_i(j)$, $s'_{m_i(j)} = \text{Max}_j$ and $\cup_{k=1}^{m_i(j)} [s_k, s'_k] = [S_i(j), \text{Max}_j]$,
- for any $k \neq l$, $1 \leq k, l \leq m_i(j)$, $[s_k, s'_k] \cap [s_l, s'_l] = \emptyset$.

For each gene g_i , we define the set $\mathcal{I}(i)$, called *input of g_i* , which contains all the interactions towards g_i in the regulatory graph and their corresponding intervals: $\mathcal{I}(i) = \{(g_j, [s_k, s'_k]), g_j \in \text{Reg}(i), k = 1, \dots, m_i(j)\}$.

Remark 1. For all $g_i \in \mathcal{G}$ we have:

- all its levels of expression are pertinent: $\min_{g_k \in \mathcal{G} / g_i \in \text{Reg}(k)} S_k(i) = 1$,
- the indegree of the node g_i is given by: $\#\mathcal{I}(i) = \sum_{g_j \in \text{Reg}(i)} m_i(j)$,
- $\text{Reg}(i) = \mathcal{I}(i)$ if and only if all regulators of g_i are the sources of simple arcs towards g_i : $\text{Reg}(i) = \mathcal{I}(i) \Leftrightarrow m_i(j) = 1, \forall g_j \in \text{Reg}(i)$.

For sake of conciseness, considering a set of interactions $X \subseteq \mathcal{I}(i)$, we write “ $g_j \in X$ ” instead of “ $\exists [s_k, s'_k] \subseteq [1, \text{Max}_j]$ such that $(g_j, [s_k, s'_k]) \in \mathcal{I}(i)$ ”.

Admissible sets $X \subseteq \mathcal{I}(i)$ as defined below correspond to possible combination of interactions operating together upon gene g_i .

Definition 1. A subset X of $\mathcal{I}(i)$ is said to be admissible if it does not contain several interactions from the same regulator: $g_j \in X$ and $g_k \in X \implies g_j \neq g_k$.

Now, given X an admissible subset of $\mathcal{I}(i)$, it defines a partition of $\text{Reg}(i)$:

$$\text{Reg}(i) = \text{Reg}(i)^X \cup \overline{\text{Reg}(i)^X},$$

$Reg(i)^X$ being the set of regulators of g_i which are source of one interaction in X , and $\overline{Reg(i)^X}$ the set of the other regulators: $\overline{Reg(i)^X} = \{g_j \in Reg(i), g_j \notin X\}$.

When the expression levels of all the genes are given, one can determine the operating interactions of the network and, for each gene, its relevant admissible set of operating interactions. Then, the effects of these sets of interactions are represented by *logical parameters* defined as follows.

Definition 2. For $g_i \in \mathcal{G}$, the application K_i associates a value $K_i(X)$ in $[0, \text{Max}_i]$ to each admissible subset X of $\mathcal{I}(i)$. The value $K_i(X)$, called *logical parameter*, defines the level towards which g_i tends when X is the set of operating incoming interactions. We denote \mathcal{K} the set of all the applications K_i .

Thus, for each gene g_i , the $K_i(X)$'s constitute parameters of the model as they define the qualitative effects of all possible combinations of incoming interactions. They are said *logical* because $X \subseteq \mathcal{I}(i)$ can be described by a conjunction of conditions on the levels of expression of the regulators of g_i .

Summarising, a logical regulatory graph is defined by three components:

- a set of nodes $\mathcal{G} = \{g_1, \dots, g_n\}$ with the maximum level $\text{Max}_i \in \mathbb{N}$ of each g_i ;
- a set of labelled arcs \mathcal{I} defined by the union of the sets of interactions targeting the genes g_i of \mathcal{G} : $\bigcup_{i=1, \dots, n} \mathcal{I}(i)$;
- a set of parameters $\mathcal{K} = \{K_i(X), i = 1, \dots, n, X \subseteq \mathcal{I}(i), X \text{ admissible}\}$.

Note that the biologists commonly consider two types of interactions: *activation* (respectively *repression*, or inhibition) has a positive (resp. negative) effect on the targeted gene, *i.e.* induces an increase (resp. a decrease) of its level of expression. However, an effective activatory or inhibitory regulation generally depends on the level of cofactors. Indeed, as one gene may be regulated by several genes, the regulatory effect of one of them may depend on the state of the others. Therefore, in the sequel, we will not explicitly consider the *sign* of an interaction which could be derived from the values of the logical parameters.

A state \mathbf{x} of a regulatory graph $(\mathcal{G}, \mathcal{I}, \mathcal{K})$ is a n -tuple (x_1, \dots, x_n) of the expression levels of the n genes of \mathcal{G} : $\mathbf{x} \in \prod_{i=1}^n [0, \text{Max}_i]$.

The (discrete) dynamics of the system can be represented by *state transition graphs*, where nodes represent *states*, and arcs represent *transitions* between states. In the sequel, we consider only asynchronous elementary transitions (for each transition, only one variable value is changed by an unitary increase or decrease). Thus, a state differs from its predecessor by exactly one component.

For each gene $g_i \in \mathcal{G}$, the application \mathcal{T}_i associates to each state \mathbf{x} of the system the logical parameter $K_i(X)$ to be considered in state \mathbf{x} . The set X is determined by the set of interactions which are operating upon g_i in state \mathbf{x} :

$$\mathcal{T}_i(\mathbf{x}) = \sum_{X \subset \mathcal{I}(i)} K_i(X) \prod_{(g_j, [s_k, s'_k]) \in X} \mathbf{1}_{[s_k, s'_k]}(x_j) \prod_{g_j \in \overline{Reg(i)^X}} (1 - \mathbf{1}_{[S_i(j), \text{Max}_j]}(x_j)), \tag{1}$$

where $\mathbf{1}_A$ is the indicator function of A : $\mathbf{1}_A(x) = 1$ if $x \in A$, 0 otherwise.

Now, the applications \mathcal{T}_i ($i = 1, \dots, n$) allow us to formally define the dynamics of the system. For any state $\mathbf{x} = (x_1, \dots, x_n)$, if $x_i \neq \mathcal{T}_i(\mathbf{x})$ ($i = 1, \dots, n$), there is one transition (arc) in the state transition graph from state \mathbf{x} to state \mathbf{y} defined by:

$$y_j = x_j \text{ for all } j \neq i, \quad y_i = x_i + \text{sign}(\mathcal{T}_i(\mathbf{x}) - x_i),$$

with, $\text{sign}(a) = +1$ if $a > 0$ and $\text{sign}(a) = -1$ if $a < 0$, for all $a \in \mathbb{Z}^*$.

For a given initial state, the corresponding state transition graph defines all the possible trajectories of the system from the selected initial conditions. In this graph, terminal strongly connected components correspond to *attractors* of the system, *i.e.* sets of states in which the system dynamics is trapped (*e.g.* cyclic behaviour or stable states). Therefore, it is interesting to determine such structures, as well as, for each attractor, its basin of attraction (*i.e.* the maximal set of states S such that all paths containing a state in S reach the attractor).

3 Multi-level Regulatory Petri Nets (MRPNs)

In this section, we define a systematic rewriting of logical regulatory networks into PNs called Multi-level Regulatory Petri Nets (MRPNs). The previously defined Boolean Regulatory Petri Nets (cf. [3]) are a special case of MRPNs. In [4], we have introduced MRPNs and their application to the genetic switch controlling the lysis-lysogeny decision in the bacteriophage lambda. But this definition of MRPNs did not take into account the possible presence of multiple arcs in the regulatory graph. Here, we reconsider the definition of MRPNs, to handle multiple regulatory effects of one gene upon another.

Consider a regulatory graph where each gene g_i has $\text{Max}_i + 1$ significant levels of expression and a current level $x_i \in \{0, \dots, \text{Max}_i\}$. Recall that $\mathcal{I}(i)$ is the set of all possible incoming interactions towards g_i .

- For all gene $g_i \in \mathcal{G}$, two (complementary) places are defined, denoted g_i and \tilde{g}_i . The sum of their marking must always equal Max_i . More precisely, the marking of place g_i represents the current expression level of the corresponding gene, and then g_i has x_i tokens, while \tilde{g}_i has $\text{Max}_i - x_i$ tokens.
- For all parameter $K_i(X)$ and all admissible set $X \subseteq \mathcal{I}(i)$ ($i = 1, \dots, n$), two transitions are defined, denoted $t_{i,X}^+$ and $t_{i,X}^-$, corresponding to an increase or a decrease of the expression level of g_i . Indeed, only three situations are possible. When the current expression level is greater (*resp.* smaller) than $K_i(X)$, we allow an increase (*resp.* a decrease) of the gene level by one unit at a time. The case where the current level equals $K_i(X)$ is omitted in the PN representation as it implies no change in the gene expression. Transitions $t_{i,X}^+$ and $t_{i,X}^-$ are enabled when the interactions in X are operating and those in $\mathcal{I}(i) \setminus X$ are not.

Therefore, $\#Reg(i)$ places are connected to $t_{i,X}^+$ and $t_{i,X}^-$ by test arcs (double arcs which test the presence of a given number of tokens):

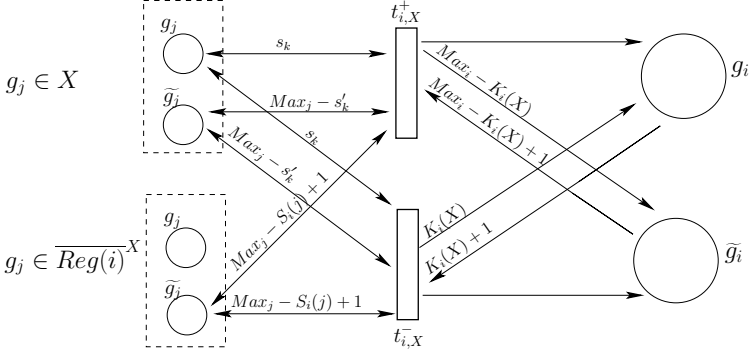


Fig. 1. Weighted arcs and places connected to transitions $t_{i,X}^+$ and $t_{i,X}^-$

- for all $(g_j, [s_k, s'_k]) \in X$, the condition $s_k \leq x_j \leq s'_k$ must hold (to ensure that the interaction is operating): places g_j and \tilde{g}_j are connected to both transitions with test arcs weighted s_k and $Max_j - s'_k$, respectively;
 - for all $g_j \in \overline{Reg(i)}^X$, the condition $x_j \leq S_i(j) - 1$ must hold (recall that $S_i(j)$ denotes the lowest level for which g_j has an effect upon g_i): place \tilde{g}_j is connected to both transitions with a test arc weighted $Max_j - S_i(j) + 1$.
- Transitions $t_{i,X}^+$ and $t_{i,X}^-$ are connected to g_i and \tilde{g}_i in the following way,
- transition $t_{i,X}^-$ is enabled if $x_i \geq K_i(X) + 1$ and its firing decreases the level of g_i . Therefore place g_i is connected to $t_{i,X}^-$ by an output arc weighted $K_i(X) + 1$ and an input arc weighted $K_i(X)$ (when enabled, $t_{i,X}^-$ removes one token from g_i). Moreover, \tilde{g}_i is connected to $t_{i,X}^-$ by an input arc (the firing of $t_{i,X}^-$ adds the token removed from g_i to \tilde{g}_i);
 - transition $t_{i,X}^+$ is enabled if $x_i \leq K_i(X) - 1$ and its firing increases the level of gene g_i . Therefore \tilde{g}_i is connected to $t_{i,X}^+$ by an output arc weighted $Max_i - K_i(X) + 1$, and an input arc weighted $Max_i - K_i(X)$ (when enabled, $t_{i,X}^+$ removes one token from \tilde{g}_i). Moreover, g_i is connected to $t_{i,X}^+$ by an input arc (the firing of $t_{i,X}^+$ adds one token in g_i).

The Figure 1 illustrates the connections between transitions $t_{i,X}^+$ and $t_{i,X}^-$ and places g_i , \tilde{g}_i , g_j and \tilde{g}_j (for $g_j \in Reg(i)$).

Remark 2. In the case of a self-regulator, $(g_i, [s_k, s'_k]) \in \mathcal{I}(i)$, we have also: $s_k \leq x_i \leq s'_k$ if $(g_i, [s_k, s'_k]) \in X$, or $x_i \leq S_i(i) - 1$ if $g_i \in \overline{Reg(i)}^X$.

3.1 MRPNs, Definition and Properties

The following definition provides the re-writing rules which define the MRPN corresponding to a regulatory graph.

Definition 3. Given a multi-level logical regulatory graph, $\mathcal{R} = (\mathcal{G}, \mathcal{I}, \mathcal{K})$ and an initial state $\mathbf{x}^0 = (x_1^0, \dots, x_n^0)$, the associated **Multi-level Regulatory Petri Net** $\mathbf{N}(\mathcal{R}) = (P, T, Pre, Post, \mathcal{M}_0)$ is defined as follows:

$$\begin{aligned}
 Pre(g_i, t_{i,X}^-) &= K_i(X) + 1 & Pre(\tilde{g}_i, t_{i,X}^+) &= \text{Max}_i - \nu_i + 1 \\
 Pre(\tilde{g}_i, t_{i,X}^-) &= \text{Max}_i - S_i(i) + 1 & Post(t_{i,X}^+, \tilde{g}_i) &= \text{Max}_i - \nu_i \\
 Post(t_{i,X}^-, \tilde{g}_i) &= \text{Max}_i - S_i(i) + 2 & Post(t_{i,X}^+, g_i) &= 1 \\
 Post(t_{i,X}^-, g_i) &= K_i(X).
 \end{aligned} \tag{7}$$

In the absence of self-regulation, equations (2) define the test arcs: for all $(g_j, [s_k, s'_k]) \in X$, the number of tokens in place g_j must lay between s_k and s'_k ; and for all $g_j \in \overline{Reg(i)}^X$ the number of tokens in g_j must be less than $S_i(j) - 1$ (or the number of tokens in \tilde{g}_j must be as least $\text{Max}_j - S_j(i) + 1$).

Equations (3, left), state that if g_i contains at least $K_i(X) + 1$ tokens, then $t_{i,X}^-$ is enabled and one token is removed from g_i to be added to \tilde{g}_i . Symmetrically, equations (3, right) state that if the number of tokens in g_i is less than $K_i(X) - 1$, then $t_{i,X}^+$ is enabled (its firing removes one token from g_i and adds one to \tilde{g}_i).

In the case of a self-regulator, if $(g_i, [s_k, s'_k]) \in X$ equations (5, left) state that if the marking of g_i belongs to $[s_k, s'_k]$ (the self-regulation is operating), and is also greater than $K_i(X) + 1$, then $t_{i,X}^-$ is enabled. Whereas if the marking of g_i belongs to $[s_k, s'_k]$ and is smaller than $K_i(X) - 1$, then $t_{i,X}^+$ is enabled (equations (5, right)). The case where $g_i \in \overline{Reg(i)}^X$ is symmetrical (cf. equations 7).

Recall that markings of complementary places together represent levels of expression. They thus satisfy the condition introduced in the following definition.

Definition 4. *Given a MRPN $\mathbf{N}(\mathcal{R})$, a **valid marking** \mathcal{M} corresponds to a state of the multi-level regulatory graph $\mathcal{R} = (\mathcal{G}, \mathcal{I}, \mathcal{K})$ and verifies:*

$$\forall g_i \in \mathcal{G}, \mathcal{M}(g_i) = \text{Max}_i - \mathcal{M}(\tilde{g}_i). \tag{8}$$

Property 1. Given a MRPN $\mathbf{N}(\mathcal{G}, \mathcal{I}, \mathcal{K})$ and a valid initial marking, any reachable marking is still valid. Therefore, the MRPN is bounded: places g_i and \tilde{g}_i are Max_i -bounded (for all g_i in \mathcal{G}).

The proof is straightforward. Moreover, the marking graph of the MRPN is isomorphic to the state transition graph of the corresponding regulatory graph. This property has been formally stated for the Boolean case in [3].

The MRPN associated to a logical regulatory graph $\mathcal{R} = (\mathcal{G}, \mathcal{I}, \mathcal{K})$ has $2\#\mathcal{G}$ places and $2\sum_{g_i \in \mathcal{G}} 2^{\#\mathcal{I}(i)}$ transitions. In most of the cases, this last number can be significantly reduced applying the rules discussed in the following remarks.

Remark 3. The first reduction consists in avoiding all transitions which are never enabled by construction (their enabling markings are not valid). As illustrated in the Figure 2, for a gene g_i and an admissible set X of incoming interactions, transitions $t_{i,X}^+$ and $t_{i,X}^-$ are enabled for exclusive ranges of g_i values ($[inf_i, sup_i]$ denotes the interval of possible values of the expression levels of g_i). These ranges can reduce to the empty set in some cases. This is easily seen in the absence of self-loops, when parameter values are extremal (*i.e.* 0 or Max_i). Indeed in this case $[inf_i, sup_i] = [0, \text{Max}_i]$ and,

- if $K_i(X) = 0$, we can omit transition $t_{i,X}^+$ as equations (3,right) state that $t_{i,X}^+$ is enabled if \tilde{g}_i contains $\text{Max}_i - K_i(X) + 1 = \text{Max}_i + 1$ tokens, and this marking is not valid;
- if $K_i(X) = \text{Max}_i$, we can omit transition $t_{i,X}^-$ as equations (3,left) state that $t_{i,X}^-$ is enabled if g_i contains $K_i(X) + 1 = \text{Max}_i + 1$ tokens, and this marking is not valid.

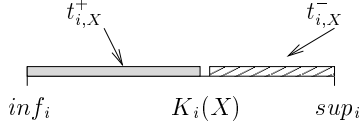


Fig. 2. Transitions $t_{i,X}^+$ and $t_{i,X}^-$ defined for each parameter $K_i(X)$ are exclusive: if $x_i \leq K_i(X) - 1$, then $t_{i,X}^+$ is enabled, whereas if $x_i \geq K_i(X) + 1$, then $t_{i,X}^-$ is enabled

The case where g_i is a self-regulator is somewhat fastidious to describe, but it follows the same principles. Let consider gene g_i self-regulated. Then,

- if $(g_i, [s_k, s'_k]) \in X$, then $[inf_i, sup_i] = [s_k, s'_k]$ (with $\mu_i = \max\{s_k, K_i(X) + 1\}$, $\lambda_i = \min\{s'_k, K_i(X) - 1\}$ as in Definition 3), then,
 - if $K_i(X) = 0$ then $\lambda_i = -1$ and we can omit transition $t_{i,X}^+$ as equations (5,right) state that $t_{i,X}^+$ is enabled if \tilde{g}_i contains $\text{Max}_i - \lambda_i = \text{Max}_i + 1$ tokens, and this marking is not valid;
 - if $K_i(X) = \text{Max}_i$ then $\mu_i = \text{Max}_i + 1$ and we can omit transition $t_{i,X}^-$ as equations (5,left) state that $t_{i,X}^-$ is enabled if g_i contains $\mu_i = \text{Max}_i + 1$ tokens, and this marking is not valid;
- if $g_i \in \overline{\text{Reg}(i)}^X$, then the set of possible values of g_i is $[0, \text{Max}_i] \setminus [s_k, s'_k]$, so that the interval in Figure 2 should be divided in two parts. Considering $\nu_i = \min\{S_i(i), K_i(X)\}$ as in Definition 3,
 - if $K_i(X) = 0$ then $\nu_i = 0$ and we can omit transition $t_{i,X}^+$ as equations (7,right) state that $t_{i,X}^+$ is enabled if \tilde{g}_i contains $\text{Max}_i - \nu_i + 1 = \text{Max}_i + 1$ tokens, and this marking is not valid;
 - if $K_i(X) = \text{Max}_i$, we can omit transition $t_{i,X}^-$ as equations (7,left) state that $t_{i,X}^-$ is enabled if g_i contains $K_i(X) + 1 = \text{Max}_i + 1$ tokens, and this marking is not valid.

Moreover, if g_i is a self-regulator, both $t_{i,X}^+$ and $t_{i,X}^-$ can be omitted in the two following cases:

- if $(g_i, [s_k, s'_k]) \in X$ and $K_i(X) = s_k = s'_k = \text{Max}_i$ (in this case $inf_i = \text{Max}_i$, the range of values enabling both transitions is empty),
- if $g_i \in \overline{\text{Reg}(i)}^X$ and $K_i(X) = S_i(i) - 1 = 0$.

The second kind of reduction which can be performed is presented next.

Remark 4. All admissible set $X \subseteq \mathcal{I}(g_i)$ defines a logical formula which is a conjunction of literals of the form $[x_j \geq s_k]$ and $\text{not}[x_j \geq s'_k]$ for all $(g_j, [s_k, s'_k]) \in X$ and $\text{not}[x_j \geq S_j(i)]$ for all $g_j \in \overline{\text{Reg}(i)}^X$ ($i = 1, \dots, n$). In general two transitions are associated to such formula (excepting the reductions derived from the previous remark).

Now, consider all the logical parameters having the same value $x \in [0, \text{Max}_i]$. They define a disjunction of conditions (the corresponding admissible sets $X \subseteq \mathcal{I}(g_i)$ such that $K_i(X) = x$) under which g_i should tend to level x . This disjunctive formula can often be simplified. In a work in progress, we consider extensions of Binary Decision Diagrams (introduced in [1]) to represent, for a given gene and a given value, the set of logical parameters taking this value. This representation can lead to a simplified disjunctive formula expressing the condition under which g_i should tend to x . This reduction implies a lower number of transitions in the MRPN. Indeed, it is easy to verify that the number of transitions is at most twice the number of terms in the reduced disjunctive formula.

Remarks 3 and 4 can lead to a significant reduction of the MRPN corresponding to a logical regulatory graph. An illustration is provided in the Figure 6 in Section 4. Observe that, in any case, the number of transitions is related to the (reasonably low) indegrees of the genes in the logical regulatory graph.

3.2 Coloured Regulatory Petri Nets (CRPNs)

One drawback of the MRPNs is that they are not easily readable. This point has motivated the use of Coloured PN for the modelling of logical regulatory networks (for an introduction to Coloured PNs, see [14]). In addition to readability, Coloured PN are well suited for model checking techniques.

In the following definition, we specify the rewriting of a logical regulatory graph \mathcal{R} into a *Coloured Regulatory Petri Net* (CRPN).

Definition 5. *Given a regulatory graph $\mathcal{R} = (\mathcal{G}, \mathcal{I}, \mathcal{K})$ and an initial state \mathbf{x}^0 , we define the Coloured Regulatory Petri Net $\mathbf{C}(\mathcal{R}) = (\Sigma, P, T, A, \mathcal{C}, G, E, \mathbf{x}^0)$ as follows:*

- Σ the finite set of colour sets: $\Sigma = \{[0, \text{Max}_i], i = 1, \dots, n\}$.
- $P = \{g_1, \dots, g_n\}$ the set of places.
- \mathcal{C} the color function associates to each place its expression domain (or colour set): $\mathcal{C} : P \rightarrow \Sigma$, $\mathcal{C}(g_i) = [0, \text{Max}_i]$.
- $T = \{T_1, \dots, T_n\}$ is the set of transitions.
- $A \subseteq (P \times T \cup T \times P)$ is the set of arcs linking places and transitions; $\forall T_i \in T$,

$$\forall g_j \in \text{Reg}(i), g_j \neq g_i, (g_j, T_i) \in A \text{ and } (T_i, g_j) \in A, \quad (9)$$

$$(g_i, T_i) \in A, (T_i, g_i) \in A. \quad (10)$$

Let denote $\bullet T_i = \text{Reg}(i) \cup \{g_i\}$ the set of input places of T_i .

- E the arc expression function defined as follows: $\forall T_i \in T$,

$$\forall g_j \in \bullet T_i \setminus \{g_i\}, \quad E(g_j, T_i) = E(T_i, g_j) = x_j, \quad x_j \in \mathcal{C}(g_j), \quad (11)$$

$$E(g_i, T_i) = x_i, \quad x_i \in \mathcal{C}(g_i), \quad (12)$$

$$E(T_i, g_i) = x_i + \text{sign}(T_i(\mathbf{x}) - x_i), \quad \mathbf{x} \in \prod_{g_k \in P} \mathcal{C}(g_k). \quad (13)$$

- $G = \{G_1, \dots, G_n\}$ is the set of guards; to each transition T_i is associated a guard G_i , a Boolean function defined as follows:

$$\forall \mathbf{x} \in \prod_{g_j \in P} \mathcal{C}(g_j), \quad G_i(\mathbf{x}) = [T_i(\mathbf{x}) \neq x_i]. \quad (14)$$

- The initial marking $\mathbf{x}^0 = (x_1^0, \dots, x_n^0)$ assigns to each place g_i one token $x_i^0 \in \mathcal{C}(g_i)$ with x_i^0 being the value of gene g_i in the initial state \mathbf{x}^0 .

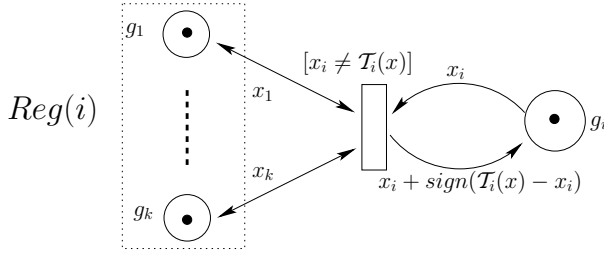


Fig. 3. The CRPN structure representing the regulation of gene g_i by its regulators

The arcs defined in (9) are test arcs, labelled by the current marking of the places in $Reg(i) \setminus \{g_i\}$ (Equation 11); the input arc (g_i, T_i) defined in (10) has also x_i as a label (Equation 12). The enabling of the transition T_i depends on the current state of its input places $\bullet T_i$ and on the Boolean value of its guard G_i . When enabled, the firing of T_i modifies the marking of place g_i according to Equation (1) through the arc (T_i, g_i) , g_i expression takes the value $x_i + 1$ or $x_i - 1$, depending on whether $T_i(\mathbf{x})$ is greater or smaller than x_i (Equation 13).

In (13) and (14), we could consider a projection of the state vector \mathbf{x} on the subset of places $Reg(i) \cup \{g_i\}$, as it is only necessary to consider the markings of places in $\bullet T_i$.

Note that when $T_i(\mathbf{x}) = x_i$, the guard of T_i is false. Hence, the stable states of the regulatory network correspond to dead markings in the CRPN (in contrast with [5] where a stable state is a marking which is its unique successor).

While Coloured PNs are more compact than the corresponding standard PNs, these are amenable to more powerful analyses. Consequently, depending on the questions to be addressed, one may preferably represent logical regulatory graphs using one or the other formalism.

A procedure to recover a MRPN from a CRPN can be easily defined. It first consists in applying the usual method to unfold Coloured PNs. The resulting ordinary net can then be reduced in order to obtain a MRPN complying the Definition 3.

4 Qualitative Dynamical Modelling of Th-lymphocyte Differentiation

4.1 Introducing Th-cell Differentiation

When human are challenged by a microbial infection, various cell types are mobilized to protect vital functions and eliminate the pathogen. Among these different cell populations, the Th lymphocytes play a crucial role in the regulation of the immune response through the integration and excretion of specific molecular signals (lymphokines) (the letters Th refer to the thymus, the organ where Th cells mature, and to the *helping* function of this cell population). Depending on the challenge and on the activity of other cell lines, the virgin T lymphocytes may differentiate into different subtypes (Th1 and Th2) characterised by specific gene expression and lymphokine excretion patterns. On the basis of an extensive analysis of the literature, L. Mendoza has recently proposed a logical model encompassing the most crucial regulatory components and the cross-interactions involved in these differentiative decisions [16].

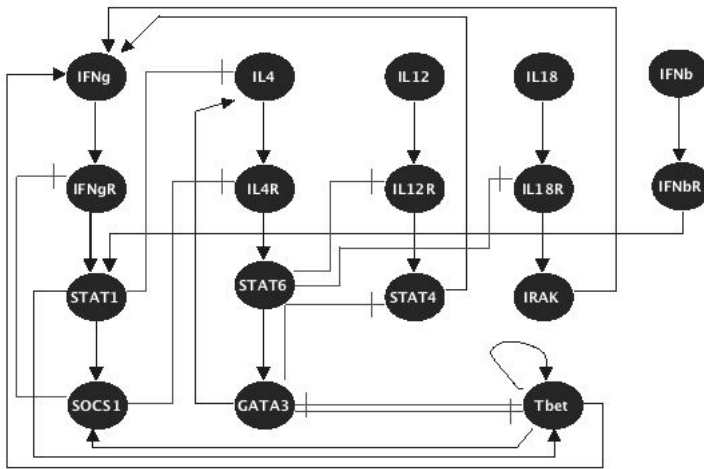


Fig. 4. Logical regulatory graph for the network controlling Th differentiation described in [16]. The nodes represent transcription regulatory factors (T-bet, GATA-3), signaling transduction factors (STAT1, STAT4, STAT6, SOCS1, IRAK), lymphokines (IFN- β , IFN- γ , IL-4, IL-12, IL-18), and lymphokine receptors (IFN- β R, IFN- γ R, IL-4R, IL-12R, IL-18R). Normal arrows represent activations, blunt arrows inhibitions.

The model of Mendoza is briefly described in the supplementary material provided on the GINsim web site [25], including an XML (GINML) file containing the definition of the logical regulatory graph, interaction intervals and parameter values. This logical model can be analysed using the software suite GINsim, which can be downloaded from the same url. GINsim provides an interface to define logical regulatory graphs and to construct state transition graphs [9]. In

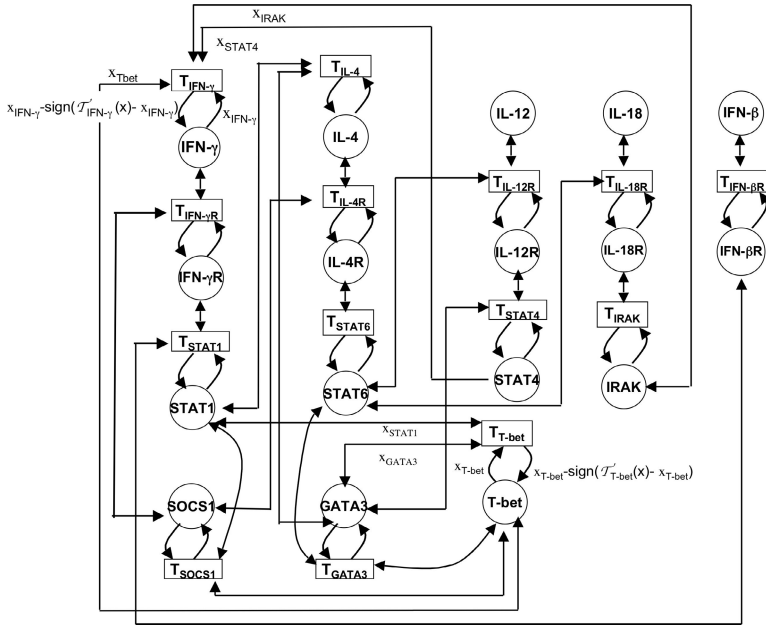
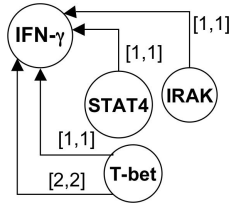


Fig. 5. CRPN translation of the logical model of the network controlling Th differentiation. Note that the topology of the regulatory graph given in Figure 4 can be easily recovered from the CRPN.

particular, GINSim can be used to identify all the stable states of a system (considering all possible initial conditions) and to check whether these states can be reached from specific initial conditions.

Using the Definition 5, we first construct the CRPN associated with the logical model defined by Mendoza (including parameter values and considering an initial state corresponding to virgin Th cells). The resulting CRPN is shown in the Figure 5. Each place of the CRPN of the Figure 5 corresponds to one element of the original logical regulatory graph. The marking of a place represents the level of expression (or protein activity) of the corresponding regulatory element. Each place is fed by one transition, which encodes the logical function of the corresponding regulatory element. A transition is further linked by test arcs to each regulator of the corresponding component.

Alternatively, using the Definition 3, the same logical regulatory graph can be rewritten as a MRPN, where each regulatory node is represented by two complementary places. The logical level of a gene is then represented by the marking of the reference place, while the number of tokens is constant for each pair of complementary places. The transitions then correspond to relevant logical parameters. As mentioned in the previous section, this MRPN can also be obtained by a proper deployment of the CRPN just described. The Figure 6 shows the subnet of the MRPN corresponding to the regulation of IFN- γ and illustrates the application of the reduction rules of Remarks 3 and 4.



0	1	2
$K_{IFN-\gamma}(\emptyset)$ $K_{IFN-\gamma}(\{IRAK,1\})$	$K_{IFN-\gamma}(\{STAT-4,1\})$ $K_{IFN-\gamma}(\{STAT-4,1\},(T-bet,1))$ $K_{IFN-\gamma}(\{IRAK,1\},(T-bet,1))$ $K_{IFN-\gamma}(\{T-bet,1\})$	$K_{IFN-\gamma}(\{STAT-4,1\},(IRAK,1))$ $K_{IFN-\gamma}(\{STAT-4,1\},(IRAK,1),(T-bet,1))$ $K_{IFN-\gamma}(\{STAT-4,1\},(IRAK,1),(T-bet,2))$ $K_{IFN-\gamma}(\{STAT-4,1\},(T-bet,2))$ $K_{IFN-\gamma}(\{IRAK,1\},(T-bet,2))$ $K_{IFN-\gamma}(\{T-bet,2\})$
$(STAT-4=0) \wedge (T-bet=0)$	$(STAT-4=1) \wedge (T-bet < 2) \wedge (IRAK=0)$ \vee $(STAT-4=0) \wedge (T-bet=1)$	$(T-bet=2)$ \vee $(STAT-4=1) \wedge (IRAK=1)$
$t_{IFN-\gamma}^{1-}$	$t_{IFN-\gamma}^{2+}, t_{IFN-\gamma}^{2-}, t_{IFN-\gamma}^{3+}, t_{IFN-\gamma}^{3-}$	$t_{IFN-\gamma}^{4+}, t_{IFN-\gamma}^{5+}$

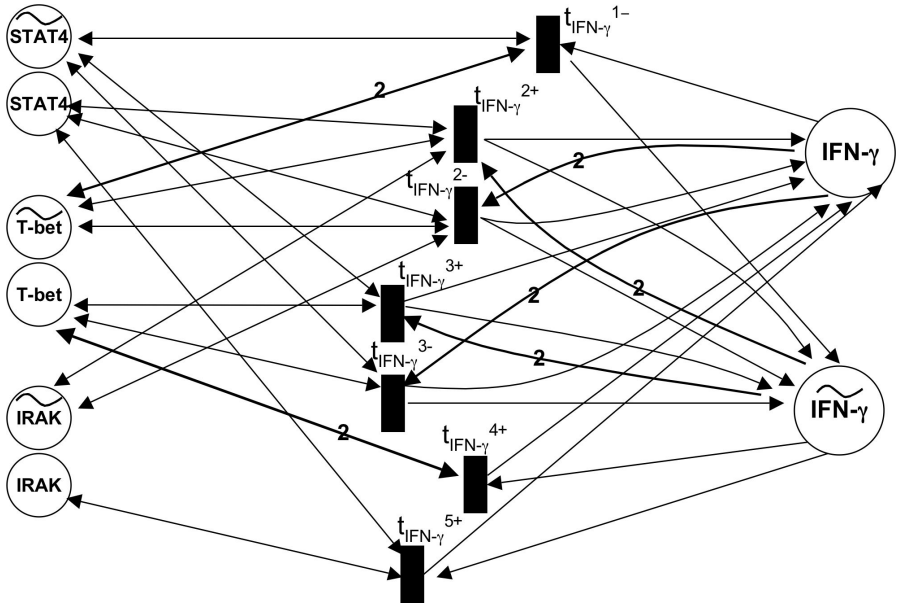


Fig. 6. Top: graph of the Figure 4 restricted to the INF- γ regulators. Middle: each column corresponds to a specific parameter value given in the first row (0, 1, 2); row 2, logical parameters for all admissible sets of interactions; row 3, simplified conditions under which IFN- γ tends to the given values (cf. remark 4); row 4, transitions representing these conditions. The reduction rules of remarks 3 and 4 result in lowering the number of transitions from 24 to 7. Bottom: the MRPN representing the regulations of IFN- γ .

At this point, we thus have two formally equivalent PN representations of a single parametrised logical regulatory graph. The CRPN representation has the advantage of graphical simplicity. Furthermore, standard CPN tools can be used to perform simulations or even develop model checking approaches. In contrast, although more difficult to grasp visually, the MRPN representation is amenable to more extensive algebraic analyses. In the sequel, we briefly discuss the analysis of the MRPN of the regulatory network controlling the Th differentiation. This analysis has been performed using INA [24] (the INA and PNML files can be downloaded from [25]).

4.2 Stable States and Their Biological Interpretation

Using a logical simulation tool (e.g. GINsim [9]) or constraint programming [7], one can identify all existing stable states of a logical model (considering all possible initial conditions). In the case of Mendoza's model, four different logical stable states have been readily identified, each corresponding to a specific cellular differentiation state:

- the first stable state has all nodes at the level zero and corresponds to the *naive* (or *virgin*) Th lymphocytes;
- the second stable state encompasses four nodes at level one: the interferon- γ (IFN- γ) and its receptor, the signal transduction factors STAT1 and SOCS1, and the transcription factor T-bet (all the other nodes are at level zero);
- the third stable state is identical to the second one, excepting that IFN- γ and T-bet are at their highest levels (two);
- finally, the fourth stable state encompasses four nodes at level one: the interleukin-4 and its receptor, as well as the signal transduction factor STAT6 and the transcription factor GATA3 (all the other nodes are at level zero).

The last stable state clearly corresponds to the Th2 differentiation state, whereas the second and third stable states correspond to Th1 variants. The co-existence of these two Th1 states accounts for different lymphokine dose effects (IFN- γ) and synergic effects of IL12 and IL18, which favor Th1 polarisation.

This multistability property can be related to the presence of specific positive regulatory circuits (each involving an even number of negative interactions) found in the original regulatory graph. As shown in [16], four circuits are playing a crucial role in this process:

- the INF- γ pathway, including its receptor, STAT1, and T-bet, which in turn regulates INF- γ expression;
- the IL-4 pathway, including its receptor, STA6, and GATA3, which in turn regulates IL-4 expression;
- the positive circuit made of the cross-regulation between T-bet and GATA3;
- and finally the self-regulation of T-bet.

The two first positive circuits ensure a cohesive expression of all the regulatory elements characteristic of the Th1 or Th2 cell populations, respectively. The third

circuit ensures the mutual exclusion between these expression patterns. Finally, the last circuit enables the differentiation of Th1 subtypes, characterised by different qualitative levels of T-bet and IFN- γ .

Using INA, it is easy to check the stability of the corresponding markings in our MRPN translation of Mendoza's model. Furthermore, one can check their reachability from specific initial conditions, e.g. for proper combinations of lymphokines (initially and transiently). To illustrate this point, let us consider three situations reported in experimental articles (cf. citations in [16]):

- starting from an initial condition with all nodes at zero but in the presence of a medium level of IFN- γ , the system can reach the virgin state (early extinction of IFN- γ signal) or the Th1 state characterized by medium level of IFN- γ and T-bet;
- starting from the same initial condition but in the presence of a high level of IFN- γ , the system can reach the virgin state or two Th1 states, depending on the duration of the IFN- γ signal;
- starting from the virgin state plus a combination of IL-12 and IL-18, the system can reach the same three stable states as in the preceding situation;
- finally, starting from the virgin state in the presence of IL-4, the system can reach the virgin state (early extinction of IL-4 signal) or the Th2 state.

5 Conclusions, Discussion and Prospects

We have presented a combined modelling approach encompassing two main steps. First, the model specification is done in terms of a generic regulatory graph, followed by its parameterisation, taking advantage of the flexibility of the definition of the logical parameters. Next, the resulting parameterised regulatory graph is translated into the Petri net formalism. In this respect, we have proposed two formally equivalent rewritings, the first based on standard Petri nets (MRPNs), the second based on Coloured Petri nets (CRPNs). Our approach has been illustrated through the PN translation of a logical model of the core regulatory network controlling the differentiation of T lymphocytes into Th1 and Th2 subtypes. We have shown that the derived CRPN and MRPN models allow to fully recover the salient dynamical properties delineated in the original logical model analysis, in particular the stable states and their reachability from given initial conditions. Note that the current version of GINsim supports the definition of logical models and the construction of the state transition graphs [25]. A forthcoming version of the software, currently in development, will provide a PN export functionality based on the rules defined in this paper.

The combination of the logical approach with the PN framework offers promising prospects for the modelling and analysis of complex regulatory systems. Consequently, logical models become amenable to the numerous tools developed by the PN community, including model checking techniques. MRPNs are somewhat difficult to grasp but readily appropriate for algebraic analyses. In particular, it is possible to identify dead markings (enabling no transition, they correspond to

stable states of regulatory networks). The identification of livelocks (*i.e.* cycles of transition firings in which the dynamics is trapped) is also relevant. In marking graphs, livelocks correspond to terminal strongly connected components of more than one node. They represent cyclical attractors of the biological system and can denote periodic behaviours (e.g. the cell cycle) or homeostasis (e.g. the control of temperature in the cell). Attractors and their basins of attraction can be identified applying classical graph theory algorithms on the marking graphs. But such methods may be intractable when facing complex systems. Moreover, it can be useful to derive general properties (independently of initial conditions). Thus, one challenge consists in delineating structural properties of MRPNs (and CRPNs), in order to derive specific theorems on induced dynamical features. Taking inspiration from the logical approach (cf. [22]), we presently focus on the characterization of the dynamical roles of regulatory circuits (cf. [19] for some preliminary results in the Boolean case). This should ease the analysis of large and complex regulatory systems, which remain difficult to explore through systematic simulations. In addition, CRPNs offer an intuitive graphical representation, and can still be readily used to perform simulations or to perform model checking.

At this stage, the resulting marking graphs cover various (and often incompatible) temporal behaviours. In principle, the distinction between alternative pathways can be forced through assumptions on transition delays or on priorities. In this context, Stochastic PNs enable the representation of such assumptions taking into account experimental noise.

In the future, this combined approach will be challenged through its application to more complex regulatory networks, eventually combining genetic and metabolic interactions (cf. [20] for a first step in this direction).

Acknowledgements. We thank A. Naldi for his active participation in the software developments, as well as L. Mendoza and P. Ruet for insightful discussions. We acknowledge financial support from the French Research Ministry through the *ACI IMPbio LumImDynNet project*, the *ANR project JC05-53969*, and the European Commission (contract LSHG-CT-2004-512143).

References

1. Bryant, R.: Graph-Based Algorithms for Boolean Function Manipulation. *IEEE Transactions on Computers*, **C-35** (1986): 677–691.
2. Chaouiya, C., Remy, E., Mossé, B., Thieffry, D.: Qualitative analysis of regulatory graphs: a computational tool based on a discrete formal framework. *LNCIS*, **294** (2003): 119–126.
3. Chaouiya, C., Remy, E., Ruet, P., Thieffry, D.: Qualitative Modelling of Genetic Networks: From Logical Regulatory Graphs to Standard Petri Nets. *LNCIS*, **3099** (2004): 137–156.
4. Chaouiya, C., Remy, E., Thieffry, D.: Petri Net Modelling of Biological Regulatory Networks. *Proc. of CompBioNets'04*, KCL publications, London, 2004.

5. Comet, J.-P., Kludel, H., Liauzu, S.: Modeling Multi-valued Genetic Regulatory Networks Using High-Level Petri Nets. LNCS **3536** (2005): 208-227.
6. de Jong, H.: Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. J. Comput. Biol. **9** (2002): 67-103.
7. Devloo, V., Hansen, P., Labbe M.: Identification of all steady states in large networks by logical analysis. Bull. Math. Biol. **65** (2003): 1025-1051.
8. Glass, L., Kauffman, S.A.: The logical analysis of continuous, non-linear biochemical control networks. J. theor. Biol. **39** (1973): 103-129.
9. Gonzalez, A.G., Naldi A., Sánchez, L., Thieffry, D., Chaouiya, C.: GINsim: a software suite for the qualitative modelling, simulation and analysis of regulatory networks. Biosystems **84-2** (2006): 91-100.
10. Goss, P.J.E., Peccoud, J.: Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets. Proc. Nat. Acad. Sci. USA. **95** (1998): 6750-6755.
11. Hardy, S., Robillard, P. N.: Modeling and simulation of molecular biology systems using petri nets: modeling goals of various approaches. J. Bioinform. Comput. Biol. **2** (2004): 595-613.
12. Heiner, M., Koch, I.: Petri Net Based Model Validation in Systems Biology. LNCS, **3099** (2004): 216-237.
13. Hofestädt, R., Thelen, S.: Quantitative Modeling of Biochemical Networks. In Silico Biol. **1** (1998): 39-53.
14. Jensen, K.: An introduction to the theoretical aspects of coloured Petri nets. LNCS **803** (1993): 230-272.
15. Matsuno, H., Doi, A., Nagasaki, M., Miyano, S.: Hybrid Petri net representation of gene regulatory networks. Proc. Pac. Symp. Biocomput. (2000): 341-352.
16. Mendoza, L.: A network model for the control of the differentiation process in Th cells. Biosystems **84-2** (2006): 101-14.
17. Murata, T.: Petri Nets: Properties, Analysis and Applications. Proc. IEEE **77** (1989): 541-580.
18. Reddy, V.N., Liebman, M.N., Mavrovouniotis, M.L.: Qualitative analysis of biochemical reaction systems. Comput. Biol. Med. **26** (1996): 9-24.
19. Remy, E., Ruet, P., Mendoza, L., Thieffry, D. and Chaouiya C.: From Logical Regulatory Graphs to Standard Petri Nets: Dynamical Roles and Functionality of Feedback Circuits. Proceedings of BioConcur 2004, London, England, 2004.
20. Simao, E., Remy, E., Thieffry, D. and Chaouiya C.: Qualitative Modelling of Regulated Metabolic Pathways: Application to the Tryptophan Biosynthesis in *E. Coli*. Bioinformatics **21** (2005): ii190-196.
21. Thomas, R.: Regulatory networks seen as asynchronous automata: a logical description, J. theor. Biol. **153** (1991): 1-23.
22. Thomas, R., Thieffry, D., Kaufman, M.: Dynamical behaviour of biological regulatory networks-I. Biological role of feedback loops and practical use of the concept of the loop-characteristic state. Bull. Math. Biol. **57** (1995): 247-276.
23. Zevedei-Oancea, I., Schuster, S.: Topological analysis of metabolic networks based on Petri net theory. In Silico Biol. **3** (2003): 0029.
24. INA: Integrated Net Analyzer URL: <http://www.informatik.hu-berlin.de/lehrstuehle/automaten/ina/>.
25. GINsim (Gene Interaction Network simulation) URL: <http://gin.univ-mrs.fr/GINsim>.