Spatial Information and Multivalued Genetic Regulatory Networks[1]

Matthieu Manceny[a,b], Marc Aiguier[a], Pascale Le Gall[a,b],
Joan Hérisson[b], Ivan Junier[b], François Képès[b]

[a] École Centrale Paris, MAS Laboratory, France
[b] Epigenomics Project, Genopole® and University of Evry, F-91000 Evry

Abstract

Modelling frameworks for biological networks are used to reason on the models and their properties. One of the main problems with such modelling frameworks is to determine the dynamics of gene regulatory networks (GRN). Recently, it has been observed in *in vivo* experiments and in genomic and transcriptomic studies, that spatial information is useful to better understand both the mechanisms and the dynamics of GRN. In this paper we propose to extend the modelling framework of R. Thomas in order to introduce such spatial information between genes, and we will show how these further informations allow us to restrict the number of dynamics to consider.

Keywords. Genetic Regulatory Networks, Spatial Information, Multivalued Dynamics, Discrete Mathematical Modelling.

1 Introduction

To understand Genetic Regulatory Networks (GRN), modelling frameworks and simulation techniques are often useful since the complexity of the interactions between constituents of the network (mainly genes and proteins) makes intuitive reasoning difficult. Most of the time, parameters of the model have to be inferred from a set of biological experiments. Formal methods, such as model checking or symbolic execution ([1, 12]), have been proved useful to determine values of parameters leading to valid dynamics of GRN, that is dynamics consistent with biological properties expressed using temporal logic. Nevertheless, these techniques are in practice difficult to manage because biological systems are either large, complex or incompletely known, resulting in a huge number of parameters to consider. Hence, in order to reduce this number, it seems relevant to embed within the model some biological knowledge such as spatial relation between genes.

Recent experiments have shown that both in eukaryotes [6] and in bacteria [2] gene transcription occurs in discrete foci where several RNA polymerases (the transcribing elements) are co-localized. This suggests that genes also tend to co-localize in space in order to optimize transcription rates. Such a scenario is supported by genomic and transcriptomic analysis [7, 3]. These have revealed that the genes which are regulated by a given transcription

[1]This work is performed within the European project GENNETEC (STREP 34952).

factor and the gene which codes for the transcription factor tend to be located periodically along the DNA [7]. In this way, the genes can be easily co-localized in the three-dimensional space according to a solenoidal structure of the DNA/chromatin, even in the presence of several kinds of transcription factors [8]. As a result, the effect of a transcription factor is enhanced due to the spatial proximity of the targets. This phenomenon is reminiscent of the local concentration effect that has been uncovered by Müller-Hill [13] a decade ago. Local concentration simply means that the interaction between molecules that are able to interact with each other is all the more efficient when molecules are close to each other. This straightforward statement is crucial to understand genome organization because genomes seem to have evolved in order to optimize the spatial proximity of reactive groups [8, 13, 9].

In this article, we propose to include spatial information into GRN and to study its effect upon the dynamics of the network. Our approach is based on the discrete modelling of GRN that has been introduced by René Thomas [14]. The spatial information concerns the gene proximity that results from a specific organization of DNA/chromatin. This proximity is modelled through two notions. The notion of *cluster* expresses the notion of co-regulation, that is a set of spatially closed genes that are expressed at the same time due to the expression of a single regulating gene (*i.e.* the presence of a single transcription factor). The notion of *privileged interaction* between genes is an ubiquitous concept in biology; for instance, specific interactions (e.g. between a transcription factor and DNA) in contrast to non-specific interactions, or local concentration phenomena are examples of privileged interactions. The use of privileged interaction is mainly based on the idea that if two interactions lead to contradictory effects, then the privileged interaction is preferred to the non privileged one.

This paper is an extension to multivalued dynamics of our previous work in [10] on Boolean dynamics. Main results of this work are recall, and we will see that whereas it is possible, in a Boolean approach, to determine constraints on the model of GRN to drastically reduce the number of dynamics to consider, this is usually not possible with a multivalued approach.

The paper is structured as follows. Section 2 presents our model of GRN including privileged interactions and clusters. In Section 3, we are interested in the multivalued dynamics of classical GRN. The dynamics is governed by a set of so called threshold and logical parameters, and we present how the structure of the GRN determines the possible values of these parameters. Nevertheless, the possible dynamics still remain too numerous, and so, Section 4 presents how to use privileged interactions and clusters to reduce the number of dynamics to consider. Section 5 presents a illustrative example, and some numerical simulations. Finally, Section 6 gives some concluding remarks.

Aknowledgments: Authors want thank all the members of the ”Groupe Observabilité” from Epigenomics Project for the discussions around distances between genes in genetic regulatory networks..

2 GRN with Privileged Interactions and Clusters (PCGRN)

Genetic Regulatory Networks are usually represented by an oriented graph, called *interaction graph*, whose nodes abstract the proteins or genes which play a role in the system and edges abstract the known interactions of the GRN. The model of this article is based on Multivalued GRN, that is GRN where gene have a finite set of *expression levels* which discretise their continuous concentration in the cell (see Section 3). An interaction ($a \to b$) can be either an activation or an inhibition: in an *activation*, the increase of the expression level of a leads to an increase of the expression level of b, the edge is labelled by the sign $+$ and a is an activator of b; in an *inhibition*, the increase of a leads to a decrease of b, the edge is labelled by the sign $-$ and a is an inhibitor of b. To this classic representation, we add the notion of *privileged interactions* as a subset of the interactions of the GRN. The notion of *clusters* defines groups of genes which are simultaneously activated or inhibited by a same gene.

Definition 1 (PCGRN: GRN with privileged interactions and clusters) *A* genetic regulatory network with privileged interactions and clusters *(PCGRN) is a labelled directed graph* $G = (V, E, S, P, C)$ *where*

- (V, E, S) *is an* interaction graph *that is*
  - V *is a finite set whose elements are called* variables
  - $E \subseteq V \times V$ *is the set of* interactions
  - $S : E \to \{+, -\}$ *associates to each interaction its* sign *("+" for* activation *and "−" for* inhibition*)*
- $P \subseteq E$ *is the set of* privileged interactions
- C *represents the clusters of* G*, that is for each gene a partition of its target genes: for each* i *in* V*,* $C(i) = \{C_i^1, \ldots, C_i^{p_i}\}$ *where*
  - $\cup_{k=1}^{p_i} C_i^k = \{j | j \in V, (i, j) \in E\}$
  - *for all* k*,* k'*:* $k \neq k' \Rightarrow C_i^k \cap C_i^{k'} = \emptyset$

For any $i \in V$, $V^-(i)$ (resp. $V^+(i)$) denotes the set of predecessors (resp. successors) of i, that is elements of V which have an action on i (resp. on which i has an action): $V^-(i) = \{j | j \in V, (j, i) \in E\}$, $V^+(i) = \{j | j \in V, (i, j) \in E\}$; $P(i)$ denotes the set of privileged predecessors of i: $P(i) = \{j | j \in V^-(i), (j, i) \in P\}$.

Definition 2 (Activators and inhibitors) *Let* (V, E, S, P, C) *be a PCGRN, and let* $i \in V$ *be a gene. We denote by* $A(i)$ *(resp.* $I(i)$*) the set of* activators *(resp.* inhibitors*) of* i*:* $A(i) = \{j | j \in V^-(i), S(j, i) = +\}$ *and* $I(i) = \{j | j \in V^-(i), S(j, i) = -\}$.

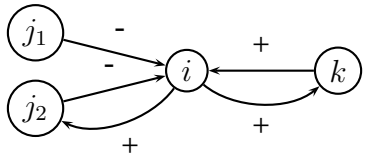

Figure 1: Example of interaction graph

In the following, a PCGRN will be represented as a graph where nodes are variables, arrows are interactions (dashed arrows for the privileged ones) and signs label arrows (see Fig. 3).

Example 1 (Interaction Graph) *Let us exemplify Definition 1 with the toy interaction graph (that is without any information on privileged interactions nor clusters) from Fig. 1 where a gene i is inhibited by j_1 and j_2 and activated by k, and activates genes j_1 and k.*

Section 3 will present the dynamics of classical interaction graphs (that is PCGRN without privileged interactions nor clusters); the influence of privileged interactions and clusters is presented in Section 4.

3 Multivalued Dynamics of Interaction Graphs

The *dynamics of an interaction graph* consists in the evolution of each gene expression level step by step. Several dynamics can be associated to an interaction graph, and the main problem is to reduce the number of dynamics we have to consider [1]. In reality, the evolution of a given gene's expression level does not depend on all the genes of the interaction graph, but only on the genes which have an action on the given gene, that is its predecessors. More precisely, not all the predecessors of a given gene have an effect on its expression level, but only the predecessors *with a sufficient expression level*, the interaction is then said to be *effective*.

3.1 Threshold Function and Multivalued Dynamic States

When a gene i acts on several targets, on j and k for example, it is often known that the level of i mandatory for an action on j to be is higher than the level necessary for the action of i on k. This knowledge is modelled through the notion of *thresholds*.

Definition 3 (Thresholds function) *Let $G = (V, E, S, P, C)$ be a PCGRN. A threshold function $T_G : E \to \mathbb{N}^*$ associates to each interaction of a GRN its threshold* thresholds parameters. *T_G is such that such that*

$$\forall (i,j) \in E, T(i,j) \neq 1 \Leftrightarrow \exists k \in E : T(i,k) = T(i,j) - 1$$

In other word, if an interaction outgoing from a variable i is labelled by a threshold α greater than 2, then there exist interactions outgoing from

i labelled by $1, \ldots, \alpha - 1$. This well represents the qualitative nature of thresholds in interaction graph, and an interaction (j, i) will be effective if and only if the expression level of j is above the threshold of (j, i). Obviously, several threshold parameters can be associated to a single interaction graph.

Example 2 (Threshold Functions) *In Fig. 1, because j_1, j_2 and k have only one successor, then the threshold of their unique outgoing interaction is* 1. *Because i has two successors, there are three possible threshold functions:* $T^1 : (i, k) \mapsto 1, (i, j_2) \mapsto 2$*;* $T^2 : (i, k) \mapsto 2, (i, j_2) \mapsto 1$*; and* $T^3 : (i, k) \mapsto 1, (i, j_2) \mapsto 1$.

In *multivalued dynamics*, genes can attain several levels, called *expression levels* which depend in both the interaction graph, and the associated threshold functions. Indeed, a gene can take as many values as the greatest outgoing threshold. The knowledge of the expression levels of all the genes define a *multivalued dynamic state*.

Definition 4 (Multivalued dynamic states) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. We denote for all $i \in V$: $b_i = max\{T_G(i, j) | j \in V^+(i)\}$). The set of possible level of expression for a gene i is $\mathbb{X}_i(G, T_G) = \{0, 1, ..., b_i\}$.*

We denote[2] *by $\mathbb{X}(G, T_G)$ the set of multivalued dynamic states of G, associated to T_G: $\mathbb{X}(G, T_G) = \prod_{i \in V} \mathbb{X}_i(G, T_G)$.*

For $x = (x_1, ..., x_{|V|}) \in \mathbb{X}(G, T_G)$, x_i is the expression level *of gene i in x.*

Example 3 (Multivalued dynamic states) *In Fig. 1, because j_1, j_2 and k have only one successor, then they have only two expression levels. Because i has two successors, there are three possible threshold parameters T^1, T^2 and T^3 (see example 2) leading to either two expression levels for i (with T^3) or three expression levels (with T^1 or T^2).*

3.2 Effective predecessors and Logical Parameters

The *dynamics of an interaction graph* consists in the evolution of each gene's expression level step by step. This evolution for a given gene does not depend on all the genes of the PGRN, but only on the genes which have an action on the given gene, that is its *effective predecessors*.

Definition 5 (Effective predecessors) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. Let $i \in V$ be a gene and let $x \in \mathbb{X}(G, T_G)$ be a dynamic state. We denote by $A^*(i, x)$ (resp. $I^*(i, x)$, $w^*(i, x)$) the set of* effective activators *(resp.* effective inhibitors, effective predecessors*) of i in the state x:*

$$A^*(i, x) = \{j | j \in V^-(i), S(j, i) = +, x_j \geq T_G(j, i)\}$$

[2] Let us recall that $|V|$ denotes the number of elements in the set V.

$$I^*(i,x) = \{j | j \in V^-(i), S(j,i) = -, x_j \geq T_G(j,i)\}$$

$$w^*(i,x) = A^*(i,x) \cup I^*(i,x)$$

Several dynamics can be associated to a given PGRN. These dynamics are described by a set of *logical parameters* which associates the future expression level of a given gene according to its effective predecessors.

Definition 6 (Logical parameters) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. For $i \in V$, we denote by $K_i^{T_G} : 2^{V^-(i)} \to \{0, \ldots, b_i\}$ (with $b_i = max\{T_G(i,j) | j \in V^+(i)\}$) the set of* logical parameters *associated to i, considering T_G.*

For any i in V, if the system is in the dynamic state $x \in \mathbb{X}(G, T_G)$, then i's next expression level is given by $K_i^{T_G}(w^*(i,x))$.

Example 4 (Logical parameters) *In Fig. 1, gene i has three predecessors. Thus, there is 8 logical parameters K_i to consider for any T in T^1, T^2 or T^3: $K_i^T(\emptyset)$, $K_i^T(\{j_1\})$, $K_i^T(\{j_2\})$, $K_i^T(\{k\})$, $K_i^T(\{j_1, j_2\})$, $K_i^T(\{j_1, k\})$, $K_i^T(\{j_2, k\})$ and $K_i^T(\{j_1, j_2, k\})$. We also have to consider $K_{j_2}^T(\emptyset)$, $K_{j_2}^T(\{i\})$, $K_k^T(\emptyset)$ and $K_k^T(\{i\})$. Since j_1 has no predecessor, it remains stable anytime.*

Let us now consider a dynamic state such that $x_i = 1$, $x_{j_1} = 0$, $x_{j_2} = 1$ and $x_k = 1$. Thus, because for any threshold parameters T in T^1, T^2 or T^3 we have $T(j_1, i) = T(j_2, i) = T(k, i) = 1$, we can state that i evolves toward $K_i^T(\{j_2, k\})$. The evolution of j_2 and k depends on the thresholds of (i, j_2) and (i, k). For example, if we consider the threshold function T^1, then, because $T^1(i, j_2) = 2$ and $T^1(i, k) = 1$, j_2's next expression level is given by $K_{j_2}^{T^1}(\emptyset)$ and k's next expression level is given by $K_i^{T^1}(\{i\})$.

Determining the dynamics of an interaction graph consists in the selection of possible threshold parameters, and then the attribution of values to the different logical parameters. The number of the possible attributions is huge: given a gene i with at least one predecessor, there are $2^{|V^-(i)|}$ logical parameters K_i, and each parameter can take at least two values. Thus, we have to consider $\prod_{i \in V} 2^{2^{|V^-(i)|}}$ possible attributions. For example, just for the interaction graph from Fig. 1, there are three possible set of threshold parameters, one leading to $2^{2^3} \times 2^{2^1} \times 2^{2^1} = 4096$ attributions for logical parameters (if i has two expressions levels), the two others leading to $3^{2^3} \times 2^{2^1} \times 2^{2^1} = 26244$ attribution (for i with three predecessors). Nevertheless, the structure of the interaction graph restricts the possible values of logical parameters.

3.3 Valid Logical Parameters

The values of logical parameters of an interaction graph must satisfy some constraints, linked to the graph structure and to the type of interaction. Logical parameters respecting the following constraints are said to be *valid*.

The *Definition constraint* is based on the definition of activation and inhibition. If a gene j which activates a gene i becomes effective, then we cannot be sure that i becomes itself effective (it may be inhibited by other genes), but the expression level of i cannot decrease.

Constraint 1 (Definition) *Let $G = (V, E, S, P, C)$ be a PGRN, and let T_G be an associated set of threshold function. Let i, j in V be two genes such that $j \in V^-(i)$. If $S(j,i) = +$ then $\forall \omega \subseteq V^-(i), K_i^{T_G}(\omega) \leq K_i^{T_G}(\omega \cup \{j\})$. If $S(j,i) = -$ then $\forall \omega \subseteq V^-(i), K_i^{T_G}(\omega) \geq K_i^{T_G}(\omega \cup \{j\})$.*

The *Observation constraint* expresses how we identify that a predecessor is an activator or an inhibitor. If j is an activator of i, then it exists at least one dynamic state where the effectiveness of j leads to an increase of the expression level of i. In other word, at least one of the previous inequalities is strict.

Constraint 2 (Observation) *Let $G = (V, E, S, P, C)$ be a PGRN, and let T_G be an associated threshold function. Let i, j in V be two genes such that $j \in V^-(i)$. If $S(j,i) = +$ then $\exists \omega \subseteq V^-(i), K_i^{T_G}(\omega) < K_i^{T_G}(\omega \cup \{j\})$. If $S(j,i) = -$ then $\exists \omega \subseteq V^-(i), K_i^{T_G}(\omega) > K_i^{T_G}(\omega \cup \{j\})$.*

Finally, the *Maximum constraint* expresses that in a dynamic state where all the activators of a gene are effective and simultaneously none of the inhibitors is effective, then the gene's expression level is maximum. Conversely, if none of the activators is effective, and all inhibitors are, then the logical parameter is minimum, that is equal to 0.

Constraint 3 (Maximum) *Let $G = (V, E, S, P, C)$ be a PGRN, and let T_G be an associated threshold function. Let i in V be a gene. By denoting $b_i = max\{T_G(i,j) | (i,j) \in E\}$, we have: $K_i^{T_G}(A(i)) = b_i$, and $K_i^{T_G}(I(i)) = 0$.*

Example 5 (Valid parameters) *Let us consider the interaction graph from Fig. 1. The considerations are done for any threshold function T in T^1, T^2 or T^3. The Maximum constraint imposes that $K_i^T(\{k\}) = 1$ and $K_i^T(\{j_1, j_2\}) = 0$. Other relations between parameters are resumed in Fig. 2, where an arrow from a node K to a node K' means $K \geq K'$ (Definition constraint), and this inequality is strict (Observation constraint) for at least one arrow of each type (plain, dashed or doted arrows). All three constraints taking into account, there are only 9 valid sets of parameters.*

4 Toward a reduction of valid dynamics

PCGRN include two new notions within the definition of interaction graph. Clusters help us to reduce the number of threshold functions to consider whereas privileged interactions reduce the number of valid logical parameters.

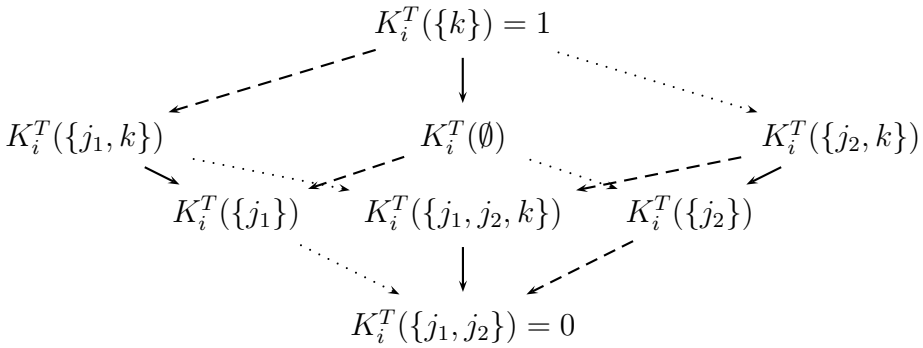

Figure 2: Relation among logical parameters of the interaction graph from Fig. 1 for any T in T^1, T^2 or T^3.

4.1 Clusters: Reduce the Number of Threshold Functions

The notion of clusters expresses the co-regulation of a set of genes, that is a set of spatially closed genes that are expressed at the same time due to the expression of a single regulating gene (*i.e.* the presence of a single transcription factor). Thus by definition, clusters allow us to reduce the set of threshold function to consider. Indeed, if two genes j and k are influenced by a gene i, and belonged to a same cluster of i, then the two interactions (i, j) and (i, k) have the same threshold.

Constraint 4 (Clusters and thresholds) *Let $G = (V, E, S, P, C)$ be a PC-GRN. Then the threshold functions T_G to consider are such that: for all i in V, for all k, k' in $V^+(i)$*

$$\exists p \in \mathbb{N}, k \in C_i^p, k' \in C_i^p \Rightarrow T_G(i, k) = T_G(i, k')$$

Example 6 (Clusters and thresholds) *Let us consider the interaction graph from Fig. 1. If j_2 and k belong to a same cluster of i, then there is only one threshold function to consider: T^3 such that $T^3(i, j_2) = T^3(i, k) = 1$. Otherwise, the three possible threshold functions must be considered.*

4.2 Conflicts and Dilemma

Despite the above constraints, valid dynamics of PGRN still remain too numerous. The different dynamics exist due to some dynamics states where the three constraints do not allow us to determine unique values for logical parameters: *Conflicts* occur when a gene is simultaneously activated and inhibited, *Dilemma* occur when all the activators (resp. inhibitors) of a gene are not effective.

Definition 7 (Conflicts and dilemma) *Let $G = (V, E, S, P, C)$ be a PC-GRN, and let T_G be an associated threshold function. Let $i \in V$ be a gene and let $x \in \mathbb{X}(G, T_G)$ be a dynamic state.*

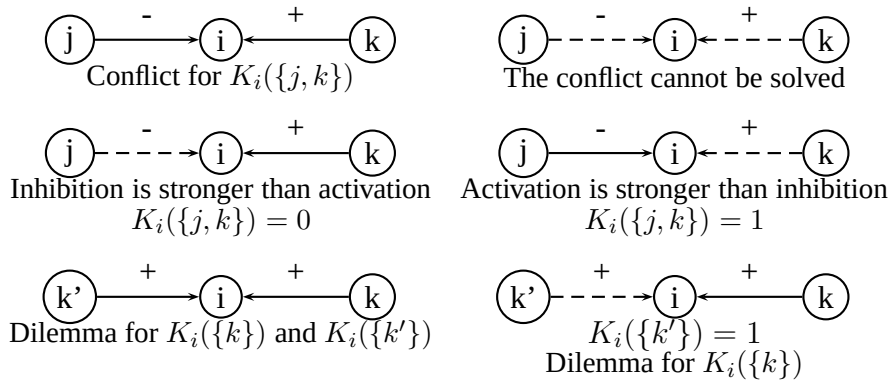

Figure 3: Solving conflicts and dilemma with privileged interactions

- x *is a* situation of conflict *for gene* i *iff* $A^*(i,x) \neq \emptyset$ *and* $I^*(i,x) \neq \emptyset$
- x *is a* situation of dilemma *for gene* i *iff (*$A^*(i,x) \neq \emptyset$ *and* $A^*(i,x) \neq A(i)$*) or (*$I^*(i,x) \neq \emptyset$ *and* $I^*(i,x) \neq I(i)$*)*

In the following, we will focus on the determination of logical parameters. Thus, conflicts and dilemma will refer to parameters, that is $K_i(w^*(i,x))$ is a conflict (resp. a dilemma) if and only if x is a situation of conflict (resp. dilemma) for gene i. In other words, if $w^*(i,x) = \omega$, then $K_i(\omega)$ is a conflict iff $\omega \cap A(i) \neq \emptyset$ and $\omega \cap I(i) \neq \emptyset$; $K_i(\omega)$ is a dilemma iff $A(i) \not\subseteq \omega \not\subseteq I(i)$ or $I(i) \not\subseteq \omega \not\subseteq A(i)$.

Note that, in this model, $K_i(\emptyset)$ is neither a conflict nor a dilemma, but corresponds to the basal situation, where a gene i is not activated or inhibited.

Example 7 (Conflicts and dilemma) *Let us consider the* 8 *possible dynamic states and the associated logical parameters for gene* i *for the interaction graph from fig. 1:* $K_i(\{j_1\})$ *and* $K_i(\{j_2\})$ *are dilemma;* $K_i(\{j_1,j_2,k\})$ *is a conflict;* $K_i(\{j_1,k\})$, $K_i(\{j_2,k\})$ *are both conflicts and dilemma.* $K_i(\{k\})$ *and* $K_i(\{j_1,j_2\})$ *are neither conflict nor dilemma: the former correspond to a situation where* i *is fully activated and is not inhibited, the latter corresponds to the reverse situation.*

4.3 Privileged Interactions: Reduce values of Logical Parameters

By definition, privileged interactions are such that their force is higher than the force of non privileged interactions. Figure 3 illustrates how to solve conflicts and dilemma using the privileged interactions: for conflicts, if two interactions occur simultaneously, then the privileged one is preferred; a dilemma is solved if one of the present gene is a privileged one.

This idea is captured through two constraints on logical parameters. The first constraint, called *Direct influence* indicates that if none of privileged activators (resp. inhibitors) is effective, and some privileged inhibitors (resp. activators) of the considered gene are effective, then the expression level cannot be maximum (resp. minimum).

Constraint 5 (Direct influence) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. Let $i \in V$ be a gene and $x \in \mathbb{X}(G, T_G)$ be a dynamic state. By denoting $b_i = max\{T_G(i,j)|(i,j) \in E\}$, we have:*

- *if $A^*(i,x) \cap P(i) \neq \emptyset$ and $I^*(i,x) \cap P(i) = \emptyset$ then $K_i^{T_G}(w^*(i,x)) > 0$*
- *if $I^*(i,x) \cap P(i) \neq \emptyset$ and $A^*(i,x) \cap P(i) = \emptyset$ then $K_i^{T_G}(w^*(i,x)) < b_i$*

The second constraint, called *Relative influence*, states that expression levels of non privileged predecessors is not important compared to the presence or absence of privileged ones. In other words, the value of a logical parameter for a set of effective genes, whose at least one is a privileged predecessor, remains the same whatever non privileged predecessors becoming effective.

Constraint 6 (Relative influence) *Let $G = (V, E, S, P, C)$ be a PCGRN, and let T_G be an associated threshold function. Let $i \in V$ be a gene and let $\omega \subseteq V^-(i)$ be a set of predecessors of i such that $\omega \cap P(i) \neq \emptyset$. Let $j \in V^-(i)$ be a gene such that $j \notin P(i)$. By denoting $b_i = max\{T_G(i,j)|(i,j) \in E\}$, we have:*

- *if $K_i^{T_G}(\omega) < b_i$ then $K_i^{T_G}(\omega \cup \{j\}) < b_i$*
- *if $K_i^{T_G}(\omega) > 0$ then $K_i^{T_G}(\omega \cup \{j\}) > 0$*

Example 8 (Influence of privileged interactions) *Let us suppose that j_1 is the only privileged predecessor in Fig. 1. Then, as soon as j_1 is ineffective, conflict and dilemma appears between other genes, but when j_1 is effective, they are solved. The 9 valid sets of parameters are reduced to 2. If we now suppose that k is the only privileged predecessor, there is no conflict, but some dilemma remains, which reduced the number of dynamics to consider to 2. If j_1 and k are privileged predecessors, there are still conflict and dilemma, but the number of dynamics to consider is to reduced to 2. Finally, if we suppose that both j_1 and j_2 are privileged predecessors, then there is neither conflict nor dilemma, and the dynamics is unique.*

In [10], we study the case of Boolean dynamics, that is interaction graphs where genes have only two levels of expression. In that case, constraints on direct or relative influences are far more restrictive than in multivalued approach. Indeed, for the direct influence, the statement $K_i(w^*(i,x)) > 0$ is equivalent to $K_i(w^*(i,x)) = 1$ (and $K_i(w^*(i,x)) < b_i$ equivalent to $K_i(w^*(i,x)) = 0$); and the formulation of relative influence becomes $K_i(\omega) = K_i(\omega \cup \{j\})$. But, even if these constraints are not constructive in a multivalued approach, they reduce the number of dynamics to consider, and can be added to other systems of constraints, such as the ones we developed in [11] to search GRN with a dynamics verifying a given temporal property.

4.4 Unique Boolean Dynamics

We present here conditions to obtain, given a PCGRN, a unique set of parameters leading to a unique dynamics. We reduce the considered dynamics to Boolean dynamics and recall the result we present in [10]. Such a situation is obtain when every threshold is equal to 1, which correspond to situations where any gene has only one cluster among its target. For that reason, we do not precise the chosen threshold function in this section. The theoretical results for any threshold function are more difficult to obtain, since we cannot control values of parameters with the constraints on direct or relative influence.

Obviously, if some genes have no predecessor, we cannot determine their expression levels, which in fact do not evolve along the time. A necessary and sufficient condition to have *no conflict* is that the set of privileged predecessors is either equal to activators or inhibitors.

Theorem 1 (No conflict) *Let $G = (V, E, S, P, C)$ be a Boolean PCGRN. The conflict situations of G can be solved iff for all $i \in V$, $P(i) = A(i)$ or $P(i) = I(i)$*

Proof 1 Sufficient. *Let x be a situation of conflict for gene i: $A^*(i, x) \neq \emptyset$ and $I^*(i, x) \neq \emptyset$. Let us suppose that $P(i) = A(i)$ (the proof is similar for $P(i) = I(i)$). Then we have $I^*(i, x) \cap P(i) = \emptyset$ and $A^*(i, x) \cap P(i) = A^*(i, x)$. Thus, due to the constraint of direct influence, $K_i(w^*(i, x)) = 1$ and the conflict is solved.*

Necessary. *Let us suppose that the condition is not verified for a given gene i, that is $P(i) \neq A(i)$ and $P(i) \neq I(i)$. $P(i) \neq A(i)$ iff either it exists $k \in A(i) \setminus P(i)$ or it exists $j \in I(i) \cap P(i)$; $P(i) \neq I(i)$ iff either it exists $j' \in I(i) \setminus P(i)$ or it exists $k' \in A(i) \cap P(i)$. If it exists $k \in A(i) \setminus P(i)$ and it exists $j' \in I(i) \setminus P(i)$, then the situation x where the only effective genes are k and j' is a situation of conflict. If it exists $k \in A(i) \setminus P(i)$ and it exists $k' \in A(i) \cap P(i)$, then two cases must be considered: if $I(i) \cap P(i) = \emptyset$ then, with $j'' \in I(i)$, the situation x where the only effective genes are k and j'' is a situation of conflict; if $I(i) \cap P(i) \neq \emptyset$ then, with $j'' \in I(i) \cap P(i)$, the situation x where the only effective genes are k' and j'' is a situation of conflict.*

Nevertheless, if all privileged predecessors are ineffective, then a situation of dilemma may occur. Dilemmas occur when two genes having the same action (either activation or inhibition) are not effective simultaneously. Thus, a necessary and sufficient condition to have *no dilemma* is that either there is only one gene for a given action, or each predecessor having this type of action is a privileged predecessor of the target.

Theorem 2 (No dilemma) *Let $G = (V, E, S, P, C)$ be a Boolean PCGRN. The dilemma situations of G can be solved iff for all $i \in V$, ($A(i) \subseteq P(i)$ or $|A(i)| = 1$) and ($I(i) \subseteq P(i)$ or $|I(i)| = 1$).*

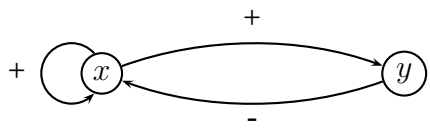

Figure 4: Interaction graph for the mucus production system in *P. aeruginosa*

Proof 2 Sufficient. *Let us consider the case of activation (the proof is similar for inhibition). Obviously, if* $|A(i)| = 1$*, then there is no dilemma. If* $A(i) \subseteq P(i)$*, then: for all* $\omega \subseteq A(i)$*, if* $\omega \neq \emptyset$ *then* $K_i(w) = 1$ *due to the constraint of direct influence; for all* $\omega_a \subseteq A(i)$*, for all* $\omega_i \subseteq I(i) \setminus P(i)$*, if* $\omega_a \neq \emptyset$ *then* $K_i(\omega_a \cup \omega_i) = 1$*, due to the constraint of relative influence; the remaining cases correspond to situations of conflict where both activators and predecessors are privileged predecessors of* i*.*

Necessary. *Let us suppose that the condition is not verified. Let us suppose we have* $|A(i)| > 1$ *and* $A(i) \not\subseteq P(i)$ *(the proof is similar for the inhibition). Then it exists* $a \in A(i) \setminus P(i)$*, and the situation* x *where* a *is the only effective predecessor of* i *is a situation of dilemma.*

Theorem 3 (No conflict nor dilemma) *Conflict and dilemma situations of a Boolean PCGRN* (V, E, S, P, C) *can be solved iff for all* $i \in V$*, (*$A(i) = P(i)$ *and* $|I(i)| = 1$*) or (*$|A(i)| = 1$ *and* $I(i) = P(i)$*)*

Proof 3 *The theorem is a direct consequence of theorems 1 and 2.*

Under the conditions of this theorem, only one dynamics is consistent with all constraints. Obviously, these conditions are difficult to state in practice. Section 5 will nevertheless illustrate that in any case, the consideration of privileged interactions allows us to reduce the set of consistent dynamics.

5 Influence of Clusters and Privileged Interactions on Dynamics

5.1 From a Biological Case Study

Pseudomonas aeruginosa are bacteria that secrete mucus (alginate) in lungs affected by cystic fibrosis, but not in common environment. As this mucus increases respiratory defficiency, this phenomenon is a major cause of mortality. Details of the regulatory network associated with the mucus production by *Pseudomas aeruginosa* are described by Govan and Deretic [4] but a simplified genetic regulatory network has been proposed by Guespin and Kaufman [5], see Fig.4.

It has been observed that mucoid *P. aeruginosa* can continue to produce mucus isolated from infected lungs. It is commonly thought that the mucoid state of *P. aeruginosa* is due to a mutation which cancels the inhibition of gene x. An alternative hypothesis has been made: this mucoid state can occur by reason of an epigenetic modification, *i.e.* without mutation [5]. The models compatible with this hypothesis are constructed in [1].

5.1.1 Boolean Dynamics

The logical parameters to consider are $K_y(\emptyset)$ and $K_y(\{x\})$ for the gene y and $K_x(\emptyset)$, $K_x(\{x\})$, $K_x(\{y\})$ and $K_x(\{x,y\})$ for gene x, which leads without further consideration, to $2^2 \times 2^4 = 64$ possible dynamics. Obviously, this number is decreased considering the constraints previously presented. $K_y(\emptyset) = 0$ and $K_y(\{x\}) = 1$ due to the observation rule. The maximum rule leads to $K_x(\{x\}) = 1$ and $K_x(\{y\}) = 0$, and then the observation rule leads to two possible dynamics: either ($K_x(\emptyset) = 1$ and $K_x(\{x,y\}) = 1$) or ($K_x(\emptyset) = 0$ and $K_x(\{x,y\}) = 0$).

The two possible dynamics are due to the conflict between x and y, and then the knowledge of privileged interactions among the activation of x by itself or the inhibition of x by y would lead to the determination of a unique dynamics. If both the interactions are privileged ones (or conversely are not privileged ones) then the two dynamics remain valid. If the inhibition is privileged and not the activation, then $K_x(\emptyset) = 0$ and $K_x(\{x,y\}) = 0$. If the activation is privileged and not the inhibition, then $K_x(\emptyset) = 1$ and $K_x(\{x,y\}) = 1$.

5.1.2 Multivalued Dynamics

Given that x has two predecessors, and y only one, there are three threshold functions to consider. Obviously, for each one $T(y,x) = 1$. The first threshold function is such that $T^1(x,y) = T^1(x,x) = 1$, and may seems similar to the Boolean situation, but in fact because the constraints on direct and relative influence are not constructive in multivalued approach, they do not allow us to choose between the different model. The two others are such that $T^2(x,y) = 2$, $T^2(x,x) = 1$ and $T^3(x,y) = 1$, $T^3(x,x) = 2$. The known logical parameters are given in the following table:

x	y	$K_x^{T^2}$	$K_y^{T^2}$	$K_x^{T^3}$	$K_y^{T^3}$
0	0	$K_x^{T^2}(\emptyset)$	0	$K_x^{T^2}(\emptyset)$	0
0	1	0	0	0	0
1	0	2	0	$K_x^{T^2}(\emptyset)$	1
1	1	$K_x^{T^2}(\{x,y\})$	0	0	1
2	0	2	1	2	1
2	1	$K_x^{T^2}(\{x,y\})$	1	$K_x^{T^2}(\{x,y\})$	1

Because of the observation constraint, we cannot have ($K_x^T(\{x,y\}) = 2$ and $K_x^{T^2}(\emptyset) = 0$) or ($K_x^T(\{x,y\}) = 0$ and $K_x^{T^2}(\emptyset) = 2$), which leads to seven valid dynamics.

5.2 From Artificial PGRN

In order to estimate the reduction in number of models induced by the introduction of privileged interactions, we have randomly generated PGRN, that is

PCGRN without any cluster information. The generation is parameterized by three values: n the number of genes, p the number of predecessors of a gene and r a ratio to determine which interactions are privileged. We first generate n genes; for each gene we then randomly select p predecessors among the n genes, each one being a privileged predecessor with a probability r. For each gene, we finally randomly select a maximum threshold (that is a random number between 1 and its number of successors), and define for each outgoing interaction its threshold between 1 and this maximum threshold, verifying that every value between 1 and the maximum threshold is selected at least one time.

Fig. 5 presents some results on artificial PCGRN composed of $n = 10$, 25, 50 and 100 genes. We give one table by hypothesis on the considered number of predecessors: the first two tables correspond to situations where each gene has exactly $p = 2$ or 3 predecessors, and the last table to a situation where each gene has a random number of predecessors between 1 and 3. We chose these rather small values for the number of predecessors per gene to fit a realistic ratio between number of genes and number of interactions.

For each PCGRN we evaluate the number of dynamics without any constraint (row named "*Total*" in each table). We then compute the number of dynamics when all the constraints (definition, observation, maximum, direct and relative influence) are applied, for several ratios of privileged interactions: when there is no privileged interaction (row "0"), when one interaction out of ten is privileged (row "1/10"), one out of five (row "1/5"), one out of two (row "1/2") and when all interactions are privileged ones (row "1"). Let us note that results between row "1" and row "0" may be largely different, since when all predecessors are privileged (row "1"), then the effectiveness of only one of them allows us to solve dilemma unsolved in row "0". All the values in the different tables given in Fig. 5 are the result of an arithmetic mean over 100 tests. The column "100 genes" for the hypothesis "3 predecessors per gene" is left empty, due to the excessive required computation time.

Obviously, the number of dynamics we have to deal with is huge (at least 10^{16}, see row "*Total*"), When considering the constraints of definition, observation and maximum, the number of dynamics is already significantly reduced (see row "0" where none of the interactions is privileged). With the constraints induced by the introduction of privileged interactions (direct and relative influence), the number of dynamics still decreases and the best results are obtained when half of interactions are privileged ones (row "1/2"). Nevertheless, let us point out that the improvement is clearly observed even with small information. For example, when only one interaction out of ten is privileged (row "1/10"). we can observe that in the third table, the number of dynamics is divided by 100 for a ten genes network, by 10^8 for 25 genes, and by 10^8 for 100 genes.

These few simulations illustrate that as soon as spatial information is known, the set of all possible dynamics is really restricted. To go further in this restriction, one can express temporal properties to characterise some

Privileged ratio r	Number of genes n			
	10	25	50	100
0	10^{6}	10^{15}	10^{31}	10^{60}
$1/10$	10^{5}	10^{14}	10^{28}	10^{53}
$1/5$	10^{5}	10^{13}	10^{25}	10^{50}
$1/2$	10^{4}	10^{10}	10^{21}	10^{40}
1	10^{4}	10^{11}	$.10^{22}$	10^{42}
$Total$	10^{16}	10^{42}	10^{82}	10^{162}

Each gene has $p = 2$ predecessors

Privileged ratio r	Number of genes n			
	10	25	50	100
0	10^{21}	10^{51}	10^{98}	–
$1/10$	10^{19}	10^{48}	10^{82}	–
$1/5$	10^{19}	10^{42}	10^{70}	–
$1/2$	10^{15}	10^{38}	10^{48}	–
1	10^{17}	10^{44}	$.10^{65}$	–
$Total$	10^{41}	10^{100}	10^{182}	–

Each gene has $p = 3$ predecessors

Privileged ratio r	Number of genes n			
	10	25	50	100
0	10^{11}	10^{30}	10^{48}	10^{82}
$1/10$	10^{9}	10^{22}	10^{39}	10^{74}
$1/5$	10^{9}	10^{22}	10^{37}	10^{67}
$1/2$	10^{8}	10^{18}	10^{26}	10^{57}
1	10^{10}	10^{19}	10^{32}	10^{66}
$Total$	10^{27}	10^{63}	10^{110}	10^{211}

Each gene has between 1 and 3 predecessors

Figure 5: Number of Dynamics for Artificial PCGRN

knowledge about the behaviour of the GRN. Formal techniques, most of them based on model checking [1], have been applied to select valid dynamics, that is dynamics consistent with biological experiments expressed by temporal properties. The problem is that these formal techniques rapidly become intractable because dynamics associated to the GRN are most of the time very numerous. Thus, from a general point of view, the set of PCGRN dynamics is all the more reduced than all biological knowledge, including spatial information, is taken into account.

6 Concluding Remarks

In this article we have presented a simple way to include spatial information within the René Thomas' framework of GRN. This supplementary information is described as a property of interactions: an interaction is privileged when the source and target genes are known to be spatially close. In the framework of Boolean dynamics, values of logical parameters are weakly constrained, leading to situations of conflicts or dilemmas where several dynamics are possible. With the notion of privileged interactions, we have determined conditions to solve some of these situations.

The spatial oriented framework we have defined is based on René Thomas' Boolean dynamics and presents the two following advantages. Firstly, since the dynamics for our spatial framework are chosen among classical René Thomas' Boolean dynamics associated to the underlying GRN without privileged interaction, then our dynamics are clearly included in the usual dy-

namics of GRN. Secondly, since spatial information allows us to solve some conflicts and dilemmas, and thus to determine some logical parameters, the number of dynamics is in practice considerably reduced.

In the goal of validating our approach, we are facing to the fact that, although spatial information seams to be central in order to apprehend the complexity of biological networks, experimental data are rare. Indeed, available data mainly concern large GRN, which are for the moment hardly attainable with our approach due to the high number of parameters to consider. Nevertheless our approach seems particularly adapted, since the first results appear even with few information on spatial relation.

References

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

[2] Julio E Cabrera and Ding J Jin. The distribution of rna polymerase in escherichia coli is dynamic and sensitive to environmental cues. *Mol Microbiol*, 50(5):1493–1505, 2003.

[3] Anne-Sophie Carpentier, Bruno Torresani, Alex Grossmann, and Alain Henaut. Decoding the nucleoid organisation of bacillus subtilis and escherichia coli through gene expression data. *BMC Genomics*, 6(1):84, 2005.

[4] J.R. Govan and V. Deretic. Microbial pathogenesis in cystic fibrosis: mucoid pseudomonas aeruginosa and burkholderia cepacia. *Microbiol rev.*, 60(3):539–74, 1996.

[5] J. Guespin-Michel and M. Kaufman. Positive feedback circuits and adaptive regulations in bacteria. *Acta Biotheor.*, 49(4):207–218, 2001.

[6] D. A. Jackson, A. B. Hassan, R. J. Errington, and P. R. Cook. Visualization of focal sites of transcription whitin human nuclei. *J. Cell Biol.*, 164:515–526, 2004.

[7] François Képès. Periodic transcriptional organization of the e.coli genome. *J Mol Biol*, 340(5):957–964, 2004.

[8] François Képès and Cédric Vaillant. Transcription-based solenoidal model of chromosomes. *Complexus*, 1(4):171–180, 2003.

[9] Christian Lanctot, Thierry Cheutin, Marion Cremer, Giacomo Cavalli, and Thomas Cremer. Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat Rev Genet*, 8(2):104–115, 2007.

[10] M. Manceny, M. Aiguier, P. Le Gall, J. Hérisson, I. Junier, and F. Képès. Spatial information and boolean genetic regulatory networks. *BioInformatics and Computational Biology (BICoB)*, LNBI 5462:270 – 281, 2009.

[11] M. Manceny, P. Le Gall, J.-P. Comet, and J.-P. Gallois. Model-checking for parametric transition systems based on symbolic execution. Technical report, Ecole Centrale Paris, MAS Laboratory, 2008.

[12] D. Mateus, J.-P. Gallois, J.-P. Comet, and P. Le Gall. Symbolic modeling of genetic regulatory networks. *Journal of Bioinformatics and Computational Biology*, 2007.

[13] B Muller-Hill. The function of auxiliary operators. *Mol Microbiol*, 29(1):13–18, 1998 Jul.

[14] R. Thomas. Logical analysis of systems comprising feedback loops. *J. Theor. Biol.*, 73(4):631–56, 1978.